Universidad Champagnat - Universidad Nacional de San Juan

CICCSI 2019

Congreso Internacional de Ciencias de la Computación
y Sistemas de Información

CICCSI 2019
Libro de Actas

[Universidad Champagnat], [Universidad Nacional de San Juan]

Universidad Champagnat

Anales del Congreso Internacional de Ciencias de la Computación y Sistemas de Información-CICCSI 2019 / compilado por Fernando Pinciroli ... [et al.]. - 1a edición para el alumno - Godoy Cruz : FUSMA Ediciones, 2020.

Libro digital, PDF

Archivo Digital: descarga

ISBN 978-987-45683-7-3

1. Computación. 2. Sistemas de Información. I. Pinciroli, Fernando, comp. II. Título.

CDD 004.071

Diseño de portada y logo: Pinciroli Braceli, María de los Ángeles.

Congreso Internacional de Ciencias de la Computación y Sistemas de Información CICCSI 2019

Mendoza, Argentina, 19 y 20 de noviembre de 2019.
San Juan, Argentina, 21 y 22 de noviembre de 2019.

Instituciones Organizadoras:

Universidad Nacional de San Juan, San Juan Argentina.
Universidad Champagnat, Mendoza, Argentina.

Fecha de Catalogación: julio de 2020

Comité Científico

Juan Iglesia, Universidad Champagnat, Argentina
Miguel Méndez Garabetti, Universidad Champagnat, Universidad Atlántida Argentina, Argentina
José Luis, Barros Justo, Universidade de Vigo, España
Laura Zeligueta, Universidad Champagnat, Argentina
María Inés Lund, Universidad Nacional de San Juan, Argentina
Nelson Rodríguez, Universidad Nacional de San Juan, Argentina
Laura, Nidia Aballay, Universidad Nacional de San Juan, Argentina
Manuel Ortega, Universidad Nacional de San Juan, Argentina
Elisa Galdame, Universidad Champagnat, Argentina
Mario Mirallas, Universidad Champagnat, Argentina
Carolina Canessa, Universidad Champagnat, Argentina
Alejandra Malberti, Universidad Nacional de San Juan, Argentina
Raúl Klenzi, Universidad Nacional de San Juan, Argentina
Juan Aranda, Universidad Nacional de San Juan, Argentina
María Murazzo, Universidad Nacional de San Juan, Argentina
Daniel Díaz, Universidad Nacional de San Juan, Argentina
Fabiana Píccoli, Universidad Nacional de San Luis, Argentina
Sandra Rodríguez, Universidad Nacional de La Rioja, Argentina
Rosana Giménez, Universidad del Aconcagua, Argentina
Patricia Ontiveros, Universidad Tecnológica Nacional-FRM, Argentina
Fabiane Benitti, Universidade Federal de Santa Catarina, Brasil
Ania Cravero, Universidad de La Frontera, Chile
Samuel Sepúlveda, Universidad de La Frontera, Chile
Alonso Toro Lazo, Universidad Católica de Pereira, Colombia
Juan Carlos Blandón Andrade, Universidad Católica de Pereira, Colombia
Álex Armando Torres Bermúdez. Corporación Universitaria Comfacauca, Colombia
José Luis Arciniegas Herrera, Universidad del Cauca, Colombia
Martín Monroy Ríos, Universidad de Cartagena, Colombia
Sonia Mora, Universidad Nacional de Costa Rica, Costa Rica
Santiago Matalonga, University of the West of Scotland, Escocia
Gilberto Borrego Soto, Universidad Autónoma de Baja California, México
Gisela Torres, Universidad Tecnológica de Panamá, Panamá
Gerardo Maturro, Universidad ORT, Uruguay
Lucía Cortés, Facultad de Ingeniería, Universidad Nacional de Cuyo, Argentina
Jorge Guerra Torres, Facultad de Ingeniería, Universidad Nacional de Cuyo, Argentina
Martin Marchetta, LABSIN, Facultad de Ingeniería, Universidad Nacional de Cuyo, Argentina
Pablo Godoy, ITIC, Facultad de Ingeniería, Universidad Nacional de Cuyo, Argentina
Facundo Bromberg, DHARMA, Facultad Regional Mendoza, Universidad Tecnológica Nacional, Argentina
Gabriel Caffaratti, LABSIN, Facultad de Ingeniería, Universidad Nacional de Cuyo, Argentina
Osvaldo Marianetti, Universidad de Morón, Argentina
Yisel Gari, ITIC, Facultad de Ingeniería, Universidad Nacional de Cuyo, Argentina

Sebastian Perez, DHARMA, Facultad Regional Mendoza, Universidad Tecnológica Nacional, Argentina
Carlos Diaz, DHARMA, Facultad Regional Mendoza, Universidad Tecnológica Nacional, Argentina
Elina Pacini, CONICET, Universidad Nacional de Cuyo, Argentina
Pablo Vidal, CONICET, Universidad Nacional de Cuyo, Argentina
Josefina Huespe, Facultad de Ingeniería, Universidad Nacional de Cuyo, Argentina
Ana Diedrichs, DHARMA, Facultad Regional Mendoza, Universidad Tecnológica Nacional, Argentina
Lucas Iacono, Facultad de Ingeniería, Universidad Nacional de Cuyo, Argentina
Marisa Haderne, Facultad de Ingeniería - UNCUYO, Argentina
Carlos Taffernaberry, GRIDTICS, Facultad Regional Mendoza, Universidad Tecnológica Nacional, Argentina
Ana Olivera, CONICET, Facultad de Ciencias Médicas, Universidad Nacional de Cuyo, Argentina
Rodrigo Gonzalez, Universidad Nacional de Cuyo, Argentina
Paula Martinez, Universidad Nacional de Cuyo, Argentina
Carlos Garcia Garino, ITIC, Universidad Nacional de Cuyo, Argentina

Comité Académico

General Chair:

Raymundo Forradellas, Universidad Nacional de San Juan, Argentina

Program Chairs:

José Luis Barros Justo, Universidade de Vigo, España
Javier Rosenstein, Universidad Champagnat, Argentina
Enzo Completa, Universidad Champagnat, Argentina
María Inés Lund, Universidad Nacional de San Juan, Argentina

Workshops Chairs:

Fernando Pinciroli, Universidad Champagnat, Argentina
Laura Nidia Aballay, Universidad Nacional de San Juan, Argentina

Research Meeting Chairs:

Nelson Rodríguez, Universidad Nacional de San Juan, Argentina
Carlos Catania, Universidad Nacional de Cuyo - Universidad Champagnat, Argentina

Professors Meeting Chairs:

Manuel Ortega, Universidad Nacional de San Juan, Argentina
Laura Zeligueta, Universidad Champagnat, Argentina

Comité Organizador

Fernando Pinciroli, Universidad Champagnat
Susana Dueñas, Universidad Champagnat
Javier Rosenstein, Universidad Champagnat
María Inés Lund, Universidad Nacional de San Juan
Carlos Catania, Universidad Champagnat
Verónica Miguez, Universidad Champagnat
Laura Nidia Aballay, Universidad Nacional de San Juan
Manuel Ortega, Universidad Nacional de San Juan
Miguel Méndez-Garabetti, Universidad Champagnat
Sandra Oviedo, Universidad Nacional de San Juan
Elisa Galdame, Universidad Champagnat
Marcelo Palma, Universidad Champagnat

Prólogo

El Congreso Internacional de Ciencias de la Computación y Sistemas de Información, CICCSI, es la actividad académica de referencia de la región y se ubica entre las más importantes de la Argentina. Su objetivo es reunir en un evento de periodicidad anual a investigadores, profesionales y alumnos de diferentes países, al promover un espacio para la difusión de los conocimientos generados en los ámbitos académico y de la industria. Con esta actividad se busca potenciar el intercambio entre investigadores, profesionales y estudiantes de informática, como así también de todos los participantes al evento.

En esta tercera edición, el CICCSI 2019, que fue organizado por la Universidad Champagnat y coorganizado por la Universidad Nacional de San Juan, se logró un hito importantísimo en la historia de los congresos internacionales de informática del interior de la Argentina, al ser el primer evento que alcanza tres ediciones consecutivas.

Esto pone de manifiesto el compromiso de las instituciones organizadoras en seguir potenciando el nivel académico con el que están comprometidos y también la aceptación de todos los participantes, tanto de aquellos que nos acompañaron como protagonistas en las conferencias, presentaciones de artículos y reuniones, como así también de parte de quienes asistieron a las diferentes actividades, con cuyos comentarios y aporte de ideas nos permite seguir mejorando CICCSI año tras año.

CICCSI 2019 contó con una docena de conferencias de expositores internacionales de la academia y la industria, una conferencia magistral, un "open space" de agilidad, cursos de primer nivel con certificaciones internacionales, reuniones de profesores de las áreas de Ciencias de la Computación y de Sistemas de Información, encuentros de investigadores, encuentro de graduados de la Universidad Champagnat, presentación de artículos científicos y posters, encuentro de jóvenes estudiantes y stands de empresas.

Agradecemos la participación de todos los que hicieron y nos acompañaron en CICCSI 2019. Noviembre de 2019.

Fernando Pinciroli
Susana Dueñas
Javier Rosenstein
María Inés Lund
Carlos Catania
Verónica Miguez
Laura Nidia Aballay
Manuel Ortega
Miguel Méndez-Garabetti
Sandra Oviedo
Elisa Galdame
Marcelo Palma

ÍNDICE

Trabajos presentados **1**

2 Modelamiento de interacciones de GitHub con un Sistema de Base de datos de Grafos: Propuesta de Evaluación
Silvina Migani, Cristina Vera, Sergio Zapata

14 Comparación de Herramientas de Accesibilidad Web: ARWeb vs TAW
Pablo Pandolfo, Adrián De Armas, Bibiana Rossi

28 Redes sociales y privacidad
Santiago Serra, Rosana Giménez, Pablo Sandez, Miguel Méndez-Garabetti

40 Modelo para la generación de datamarts en microempresas
Gabriel Dávila, Silvina Migani, María Inés Lund

50 Diseño y Modelado de un Sistema Web para el Registro Único de Acciones de Vinculación Académica y Científica de Unidades de Docencia e Investigación de la U.N.S.J. con sus pares del extranjero, en el marco de la Internacionalización de la Educación Superior (RUAVIN)
Isabel Castro, María Inés Lund

62 Integrando Afectividad en Test UX: Evaluación de una Aplicación mHealth
Laura Aballay, Mathias Díaz, María Murazzo, Silvana Aciar

75 Deserción Académica y Emociones en Ambientes E-learning: Percepciones de los estudiantes de un curso de posgrado
Edith Lovos, Laura Aballay

83 Modeling the Static View in Aspect-Oriented Software Development
Fernando Pinciroli

93 Calidad Universitaria. Una mirada exploratoria del año 2018
María Inés Lund, Myriam Herrera

107 API REST extensible para Robots no Industriales
César Omar Aranda

117 Performance evaluation of data base for big data in cloud
María Murazzo, Nelson Rodríguez, Pablo Gómez

127 Orquestación de microservicios para realizar cómputo de grandes volúmenes de datos
Miguel José Guevara Atencio, María Murazzo, Nelson Rodríguez.

137 Integración de una BCI mediante la implementación de VRPN en escenarios de Serious Game para rehabilitación de problemas neurológicos
Javier Rosenstein, Rodrigo Gonzalez, Nicolás Zárate Alvarez, Matías Campos

141 DNS Tunneling: A Deep Learning based Lexicographical Detection Approach
Franco Palau, Carlos Catania, Jorge Guerra, Sebastian Garcia, Maria Rigaki

153 Action Recognition in Videos: LSTM Network Parameter Setting
Carlos Ismael Orozco, María Elena Buemi, Julio Jacobo Berlles

159 Incorporar Análisis de Dominio (Joda) al Proceso de Desarrollo Uml Components (DSBC)
Ezequiel Moyano, Adriana Urciuolo, Matías Gel, Rodolfo Iturraspe, Emilio Izarra, Martín Villareal

163 Uso del Estándar IEEE-830 en la Especificación de Requerimientos para el Manejo Sostenible de Cuencas en TDF
Ezequiel Moyano, Adriana Urciuolo, Matías Gel, Rodolfo Iturraspe, Emilio Izarra, Martín Villareal, Jhonatan Vieira

167 Criterios para el Análisis Preliminar de Datos Extraídos Mediante Web Scraping
Roxana Martínez, Christian Parkinson, Rocío Rodríguez, Pablo Vera

177 Mantenimiento del software, aspectos aplicados en sitios web accesibles
Mariño Sonia, Juan Gallardo, Pedro Alfonzo

185 Integración de tecnologías de minería de datos y GIS. Generación de conocimiento en torno a la identificación y caracterización de clusters de robos y hurtos en una ciudad Argentina
Mariño Sonia, Lorena Flores, Sebastián Martins

195 Importancia de la Innovación en los Sistemas de Salud Pública de San Juan
Alicia Aballay, Adriana Sarmiento, María R. Romagnano, María del Carmen Becerra

205 Detección de enfermedades o plagas en cultivos de papas andinas del norte argentino utilizando procesamiento de imágenes digitales
Fabiola Patricia Paz, Javier Izetta

215 De los autómatas finitos a las expresiones regulares: Un análisis áulico de identificación de expresiones equivalentes
Mathias Diaz Ogás, Elisa S. Oliva, Ana Molina, Nancy Alonso

225 Una aproximación para construir sistemas de programación de la producción utilizando servicios en la nube
Daniel Díaz, Sandra Oviedo, Fernando Guardia, Francisco Ibañez, Juan Aranda, Nancy Alonso, Juan Cúneo

237 Criterios de calidad del software. Un enfoque de Accesibilidad Web vinculado a la gestión de recursos físicos
José M. Bordón, Pedro L. Alfonzo, Sonia I. Mariño

245 Técnicas de Análisis y Visualización de Minería de Datos para la Reducción de Dimensiones en Tablas de Datos
Laura Viviana Gutiérrez, Manuel Oscar Ortega, Maria Isabel Masanet, Facundo De La Jara

255 Una Estrategia para la Evaluación y Selección de Ideas en la Plataforma Ideas en Tiempo Real
Sandra Oviedo, Sandra Rodriguez, Daniel Díaz, Alejandra Otazú, Juan M. Cúneo, Marta Cabrera Villafañe

265 Aplicación de Internet de las Cosas a la administración y control de un robot autónomo móvil
Amaro, M. , Azcurra, D.

275 Privacidad en Datos de Salud con Tecnología Blockchain
María del Carmen Becerra, Pedro Zarate

285 Sistema de información Farmacéutica para la dispensación de Medicamentos e Insumos médicos en el Hospital Escuela y de Clínicas Virgen María de Fátima
Leonardo Garay, Enrique Mamani, Gabriel Pacheco, Ariel Rivadulla, Soledad Ciccarelli

289 Propuesta de Análisis Descriptivo y Prescriptivo en Alumnos Aspirantes e Ingresantes de la Licenciatura en Ciencias de la Comutación de la Universidad Nacional de Cuyo
Marisa Haderne, María Alejandra Marín, Cristian Gamba, Raquel Zumel, Rosana Giménez, Eleonora Luconi, Eleonora Valdivieso

293 Análisis de procesos Maps en Hadoop
Rosana Giménez, Miguel Méndez-Garabetti, Martín Eraso

297 Transformation of Models as Key Element for Increasing the Productivity in Software Development: A Case Study
Castro Rojas Luis Fernando, Esperanza Espitia Peña, Edwin Moreno Cuero

311 Representation of TROPOS by using Essence Kernel
Castro Rojas Luis Fernando, Esperanza Espitia Peña, Andrés Montilla

321 Repensando Software que requiere alto rendimiento Paradigma Reactivo
Susana Chavez, Nelson Rodríguez, Adriana Martin

331 Diseño web responsive para el registro, control y difusión de donantes voluntarios de sangre del proyecto “Vida por vidas” en la provincia de La Rioja
Priscila Orona, Sandra Rodriguez, Emilio Rearte, Diego Candelero

335 Realidad Virtual y Realidad Aumentada aplicadas al Museo Histórico del Colegio Provincial Joaquín V. González - La Rioja
Raúl Esteban Luna Mercado, Marta Cabrera Villafañe

339 Grupos Focales en la Elicitación de Requisitos Desarrollo de Videojuego en Realidad Virtual destinado a Usuarios con Parálisis Cerebral
Cintia Ferrarini Oliver, Mónica Gonzalez, Alicia Aballay, Luís Olguín

347 Modelo de Proceso incorporando Requisitos de Usabilidad y Experiencia de Usuario en el Desarrollo de Software
Iván Balmaceda Castro, Luis Fuentes, Juan Magaquian, Carlos Salgado, Mariela Fernandez, Sandra Rodriguez, Monica Gimenez

355 Data Mining aplicado a la deserción estudiantil en la Facultad de Ciencias Exactas de la UNSa
Claudia Ibarra, Cristian Martínez

Lista de autores **365**

,,

Modelamiento de interacciones de GitHub con un Sistema de Base de datos de Grafos: Propuesta de Evaluación

Silvina Migani[1], Cristina Vera, Sergio Zapata

[1]Departamento e Instituto de Informática, FCEFN-UNSJ
{silvina.migani,civerados}@gmail.com, szapata@iinfo.unsj.edu.ar

Resumen. La sencillez, el lenguaje de consulta estándar y las propiedades ACID, son algunas de las características de los sistemas de gestión de bases de datos relacionales que provocaron su exitoso y masivo uso hace ya muchas décadas. Sin embargo, la continua evolución de la tecnología ha generado cambios, y con ellos las aplicaciones se encuentran ante nuevos desafíos que, en muchas ocasiones, no pueden ser abordados adecuadamente por los populares y robustos sistemas de bases de datos relacionales. Consecuentemente surgieron nuevos tipos de sistemas de bases de datos, dentro de los cuales se encuentran los basados en grafos. Ellos prometen ser convenientes para gestionar datos altamente conectados de manera más natural y eficiente. Este trabajo presenta algunos ensayos realizados con Neo4j, un sistema de gestión de bases de datos de grafos, orientados a la experimentación y evaluación del comportamiento de estos sistemas de gestión en el dominio del análisis de interacciones dentro de una comunidad de desarrollo de software Open Source; y así poder corroborar o no las bondades proclamadas por este nuevo paradigma.

Palabras claves: Bases de Datos basadas en Grafos – Neo4j – Ingeniería Global de Software

1 Introducción

Los sistemas de gestión de bases de datos relacionales han sido usados ampliamente y con óptimos resultados hace ya muchas décadas. Sin embargo, la continua evolución de la tecnología generó cambios, que en muchas ocasiones no han podido ser abordados adecuadamente por los populares y robustos sistemas de bases de datos relacionales. Como solución a esta problemática surgieron los llamados sistemas de bases de datos NoSQL, dentro de los cuales se encuentran los sistemas basados en grafos. Ellos prometen ser convenientes para gestionar datos altamente conectados de manera más natural y eficiente que los sistemas de bases de datos tradicionales. Otra característica distintiva es que no verifican las propiedades ACID (por su denominación en inglés de *Atomicity, Consistency, Isolation and Durability*).

Por otro lado, a partir de la Web 2.0 el desarrollo colaborativo de software open source creció considerablemente, y GitHub es una plataforma ampliamente utilizada en este contexto. Es de interés encontrar la relación existente entre las características de las interacciones de los desarrolladores y el éxito de un proyecto de software. En este contexto, la naturaleza de los datos es eminentemente vincular, es decir, la importancia de las vinculaciones sobrepasa a la de las propiedades de los datos en sí mismos.

El presente trabajo expone los resultados de un ensayo preliminar del modelamiento de interacciones registradas en GitHub usando el motor de bases de datos basado en grafos Neo4j.

2 Marco Teórico

2.1 Bases de Datos basadas en grafos

Un SGBDG, es un sistema de base de datos específicamente diseñado para poseer las siguientes capacidades [1]:

- Administrar datos de tipo grafo. Es decir, su modelo de datos lógico está basado en alguna de las variantes de la definición matemática básica de grafo, como por ejemplo, grafos dirigidos o no dirigidos, con vértices y arcos etiquetados o no etiquetados, con propiedades en nodos y arcos, hipergrafos e hipernodos [2][3]. Así, las operaciones CRUD (Create-Read-Update-Delete, operaciones básicas de los SGBDs) trabajan sobre grafos y sus elementos. Sin embargo esto no significa que en el almacenamiento subyacente efectivamente se encuentren grafos. Algunos SGBDGs utilizan almacenamiento nativo [4]; es decir, sus estructuras de datos físicas están diseñadas y optimizadas para almacenar y administrar grafos. También existen gestores que mapean los grafos a otras estructuras [4], como por ejemplo tablas u objetos. En cuanto a la implementación de las conexiones entre nodos, también se presentan diferentes enfoques, algunos que eligen la adyacencia sin índices, y otros que no.

- Satisfacer los principios básicos de todo SGBD, es decir, el almacenamiento persistente, la independencia física/lógica, la integridad y la consistencia de los datos.

Hoy en día existen numerosos SGBDs basados en grafos. A continuación se mencionan en orden de popularidad según [5] algunos: Neo4j, Microsoft Azure Cosmos DB, OrientDB, ArangoDB, Virtuoso y JanusGraph.

2.2 Desarrollo de software basado en Pull-Request

Promovido por el auge del desarrollo distribuido o global de software y por la comunidad Open Source los repositorios o sistemas de control de versión (VCS por su denominación en inglés de *version control systems*), tales como GitHub, Bitbucket, JIRA, StackOverflow, etc, se han transformado en un estándar en la industria del software.

Estos VCS, también denominados *software forge*, son plataformas colaborativas web que proporcionan herramientas para facilitar el desarrollo distribuido o global de software. Estos repositorios de software contienen grandes volúmenes de información sobre código fuente, personas y procesos que intervienen en el desarrollo de un producto de software [6].

El desarrollo de software depende fuertemente de las personas [8]. El desarrollo de software es, en gran medida, un esfuerzo de equipo, y por lo tanto la interacción entre las personas y la complementariedad en la experiencia son requisitos previos para tener éxito [9]. Las interacciones y tareas de colaboración quedan registradas en la plataforma de software VCS, brindando la posibilidad de analizar esos datos desde la perspectiva de las relaciones sociales y profesionales entre desarrolladores [6]. Esas vinculaciones registradas podrían ser evidencias de confianza interpersonal, cohesión grupal, empatía, etc. Aspectos valiosos que redundarán en la efectividad del equipo de desarrollo de software.

GitHub es una de los sistemas VCS más usados actualmente y representa la última generación de *software forge*, ya que combina las capacidades tradicionales que ofrecen tales sistemas (por ejemplo, capacidades de alojamiento gratuito o sistema de control de versiones) con rasgos sociales [7].

3 Trabajo Realizado

Con el fin de evaluar los beneficios de Neo4j en el contexto de GitHub se han efectuado un conjunto de tareas contenidas en las siguientes etapas:

3.1 Obtención de datos desde la plataforma elegida.

Para concebir el ejemplo que concreta el experimento se tomó el Pull-Request 640 de la plataforma GitHub (https://github.com/cdr/code-server/pull/640). La Figura 1 lo muestra abajo.

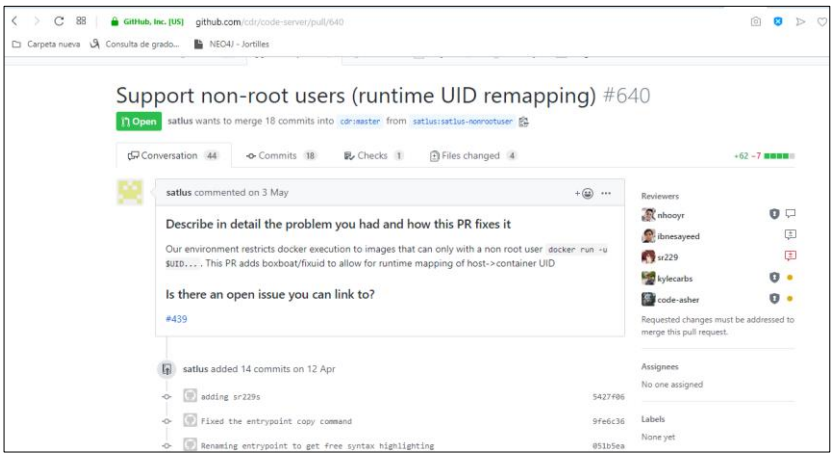


Figura 1: Pull-Request 640 en GitHub

A través del GitHub API Explorer (https://developer.github.com/v4/explorer/) y el lenguaje de consulta GraphQL se especificó una consulta que devolviera algunos datos de interés para el ensayo. A continuación se presenta el código de la consulta en la Figura 2.

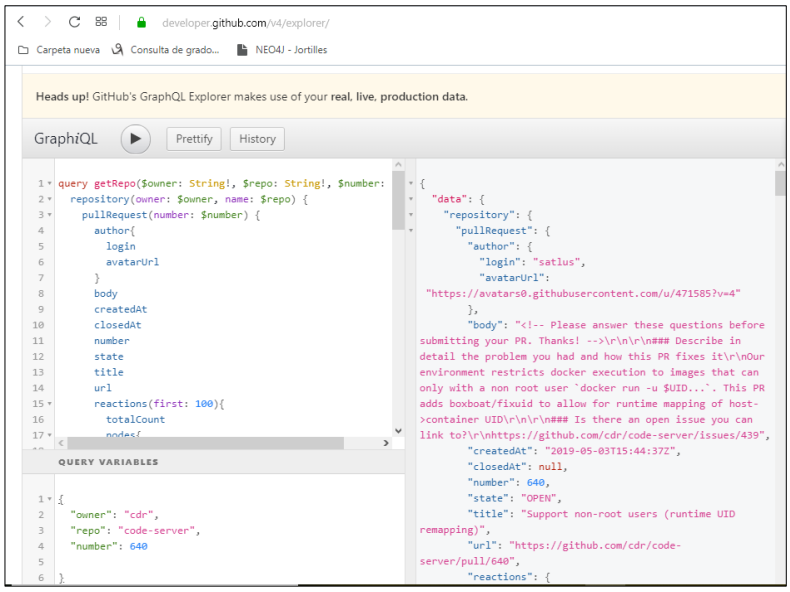


Figura 2: Consulta en GraphiQL

3.2 Transformación de los datos a grafos.

El formato de los datos que devuelve GraphQL es JSON. Se elaboró un proceso de conversión para que los datos contenidos en el documento JSON resultante se transformarán en un grafo para poder gestionarlos con Neo4j. En la **Tabla 1** se muestra el código desarrollado a tal fin.

Tabla 1: Código Fuente Desarrollado

- **Creación de nodos Pullrequest y Participante**

```
CALL apoc.load.json('file:///github.json') YIELD value
UNWIND value.data as q
MERGE (pull:Pullrequest {id:q.repository.pullRequest.number}) ON CREATE
  SET pull.autor = q.repository.pullRequest.author.login, pull.estado =
q.repository.pullRequest.state,pull.titulo=q.repository.pullRequest.title
FOREACH (p IN q.repository.pullRequest.participants.nodes | MERGE
(pa:Participante {id:p.login}) ON CREATE
  SET pa.id = p.login
 MERGE (pull)-[:PARTICIPA]->(pa))
```

- **Creación de nodos comentario**

```
CALL apoc.load.json('file:///github.json') YIELD value
UNWIND value.data.repository.pullRequest.reviewThreads as q
UNWIND q.nodes AS c1
unwind c1.comments as c
unwind c.nodes as c2
CREATE (co:Comentario {autor:c2.author.login,fecha:c2.createdAt,
texto:c2.body})
```

- **Creación de la arista (relación) ha_comentado**

```
MATCH (au:Comentario {autor: 'satlus'})
MATCH (pa:Participante {id: 'satlus'})
MERGE (au)-[:ha_comentado]->(pa);
MATCH (au:Comentario {autor: 'nhooyr'})
MATCH (pa:Participante {id: 'nhooyr'})
MERGE (au)-[:ha_comentado]->(pa);
MATCH (au:Comentario {autor: 'ibnesayeed'})
MATCH (pa:Participante {id: 'ibnesayeed'})
MERGE (au)-[:ha_comentado]->(pa);
MATCH (au:Comentario {autor: 'ptoulouse'})
MATCH (pa:Participante {id: 'ptoulouse'})
MERGE (au)-[:ha_comentado]->(pa);
MATCH (au:Comentario {autor: 'carlos-sarmiento'})
MATCH (pa:Participante {id: 'carlos-sarmiento'})
MERGE (au)-[:ha_comentado]->(pa);
MATCH (au:Comentario {autor: 'sr229'})
MATCH (pa:Participante {id: 'sr229'})
MERGE (au)-[:ha_comentado]->(pa);
```

La Figura 3de abajo presenta el grafo resultante en la interfaz de Neo4j.

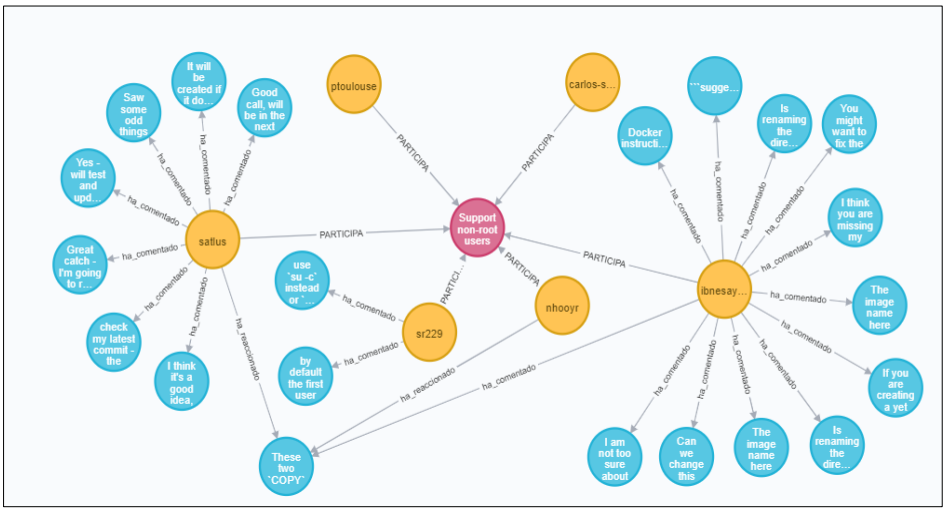


Figura 3: Grafo generado en Neo4j

El grafo resultante cuenta con nodos y aristas descriptas a continuación:

- El nodo Pullrequest (color violeta) tiene cuatro propiedades a saber: id (número), autor, estado y título.
- Los nodos participantes (color amarillo) tiene la propiedad id que está dada por el nombre o login de la persona.
- Los nodos comentarios (color celeste) tiene tres propiedades a saber: texto del mismo, fecha en la que se realizó y autor.
- Las aristas participa, vinculan el Pullrequest con los participantes.
- Las aristas ha_comentado vinculan los participantes con los comentarios realizados.
- Las aristas ha_reaccionado vinculan los participantes con los comentarios a los que ha reaccionado.

3.3 Consultas en Neo4j

A modo de ejemplo se detallan a continuación dos consultas con el fin de comenzar a descubrir el potencial del lenguaje Cypher de Neo4j en este contexto:

- Mostrar los 5 primeros participantes en función a la cantidad de intervenciones realizadas.

```
MATCH (u:Participante)
OPTIONAL MATCH (u)<-[:ha_comentado|PARTICIPA]-()
RETURN u,count(*)
ORDER BY count(*) DESC
LIMIT 5
```

El resultado obtenido se muestra a continuación en la Figura 4.

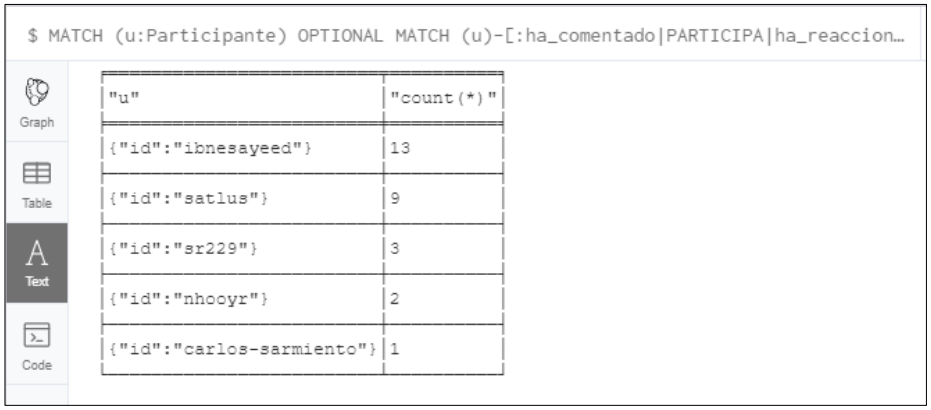


Figura 4: Resultado Consulta

- Obtener el/los comentarios a los que el participante Satlus reaccionó, concretamente los comentarios realizados por otros participantes.

```
MATCH (p:Participante{id:'satlus'})-[r:ha_reaccionado]-(co:Comentario)<-[r2:ha_comentado]-(au)
RETURN p.id,co.texto,au.id
```

La consulta devuelve el resultado de la Figura 5.

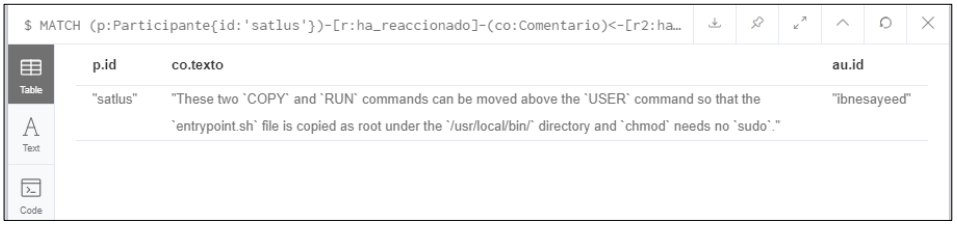


Figura 5: Resultado Consulta

- Obtener las reacciones del participante Satlus, concretamente los comentarios, junto a la fecha en que los realizó.

```
MATCH (p:Participante{id:'satlus'})-[r:ha_reaccionado]-(co:Comentario)
RETURN co.texto,co.fecha
```

El resultado de la consulta anterior se observa en la Figura 6.

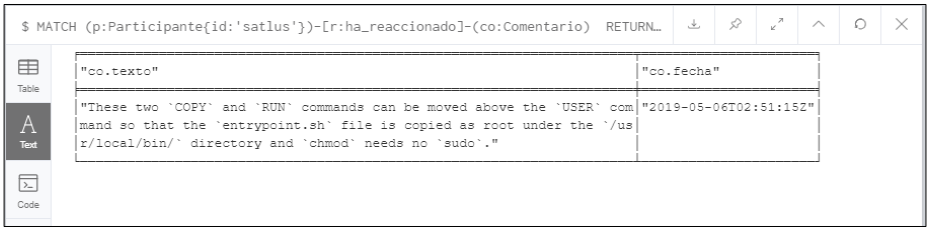


Figura 6: Resultado Consulta

4 Conclusiones

La experiencia realizada permitió advertir, en forma preliminar, las virtudes del lenguaje Cypher que provee Neo4j. Es un lenguaje declarativo, potente e intuitivo, que hace muy sencilla la tarea de codificación. En cuanto a las aristas, ellas brindan la posibilidad de navegar relaciones complejas (recursivas), a cualquier nivel de profundidad, incluso no fijo. Capacidad no ofrecida convenientemente por los sistemas SQL.

Por otra parte, la estructura de los datos en Neo4j, es muy flexible. Tanto los nodos como las aristas no necesitan ajustarse a una estructura predeterminada. Además, otro aspecto valioso se desprende de la capacidad que tienen los grafos para estructurar y visualizar los datos, permitiendo así un análisis más claro y rápido.

Todas estas características abonan, en forma preliminar, la conveniencia de abordar el análisis de los datos, relativos a las interacciones realizadas por equipos de desarrollo de software, mediante un SGBDG.

Este trabajo debe ser continuado para arribar a conclusiones definitivas. Por ello es que se propone incorporar otros pull-requests al análisis como también otros tipos proyectos de desarrollo, otros contextos globales, etc.

Contar con una herramienta que facilite la medición de aspectos humanos en equipos de desarrollo de software será un paso importante para mejorar los desempeños y calidad de los resultados de dichos equipos.

Referencias

[1] Evans, E. (2009). NoSQL: What's in a name. Eric Evans's Weblog. http://blog. sym-link. com/2009/10/30/nosql_whats_in_a_ name. html. [Ultimo Acceso: Diciembre de 2017].

[2] Cattell, R. (2011). Scalable SQL and NoSQL data stores. ACM Sigmod Record, 39(4), 12-27.

[3] Brewer, E., & Fox, A. (1999). Harvest, yield, and scalable tolerant systems. In HOTOS Proceedings of the The Seventh Workshop on Hot Topics in Operating Systems.

[4] Brewer, E. (2010). A certain freedom: Thoughts on the CAP theorem. 29th ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing PODC, 335–335.

[5] Abadi, D. (2012). Consistency tradeoffs in modern distributed database system design: CAP is only part of the story. Computer, 45(2), 37-42.

[6] Munaiah, N., Kroh, S., Cabrey, C., & Nagappan, M. (2017). Curating GitHub for engineered software projects. Empirical Software Engineering, 22(6), 3219-3253.

[7] Cosentino, V., Izquierdo, J. L. C., & Cabot, J. (2017). A systematic mapping study of software development with GitHub. IEEE Access, 5, 7173-7192.

[8] DeMarco, T., & Lister, T. (2013). Peopleware: productive projects and teams. Addison-Wesley.

[9] Wohlin, C., Šmite, D., & Moe, N. B. (2015). A general theory of software engineering: Balancing human, social and organizational capitals. Journal of Systems and Software, 109, 229-242.

,,

Comparación de Herramientas de Accesibilidad Web: ARWeb vs TAW

Mag. Pablo Pandolfo[1], Mag. Adrián De Armas[1], Mag. Bibiana Rossi[1]

[1]Instituto de Tecnología (INTEC) de la Universidad Argentina de la Empresa, UADE, Buenos Aires, Argentina
{ppandolfo, adearmas, birossi}@uade.edu.ar

Resumen. Las herramientas de accesibilidad web son aplicaciones software que permiten identificar problemas de accesibilidad web en forma automática de acuerdo con determinadas normas. Este artículo presenta la comparación de las herramientas de accesibilidad web ARWeb y TAW, ambas detectan errores de accesibilidad web, sin embargo, cada una de ellas calcula puntajes y presenta sus resultados en forma diferente.

Palabras clave: accesibilidad web, WCAG 2.0, CSS, HTML, herramientas de evaluación automática de accesibilidad, ARWeb, TAW.

1 Introducción

Normalmente la accesibilidad web se asocia con la idea de desarrollar páginas web que puedan ser utilizadas por las personas con discapacidad. Sin embargo, hoy en día la accesibilidad web no está orientada exclusivamente a las personas con discapacidad. La accesibilidad web es única porque no diferencia entre dispositivos desde donde se visualiza contenido web y es universal porque prevé la utilización de la web independientemente de las características del usuario. La accesibilidad web es el acceso universal a la web, independientemente del tipo de hardware, software, infraestructura de red, idioma, cultura, localización geográfica y capacidades de los usuarios. [REVILLA MUÑOZ, 2013]

La evaluación, revisión o análisis de la accesibilidad web tiene por finalidad analizar, estudiar y validar las páginas web para verificar el cumplimiento de las pautas y directrices de accesibilidad existentes. La evaluación de la accesibilidad web se puede realizar de forma automática y manual. Una herramienta de evaluación de la accesibilidad web es un programa informático que analiza automáticamente el nivel de accesibilidad de una página web.

La evaluación automática permite realizar una evaluación rápida, ayuda a tener una primera impresión de la accesibilidad de una página web, pero no proporciona un análisis definitivo, ya que puede no detectar errores o señalar falsos positivos. Se requiere un análisis manual complementario por parte de un experto. [SEGOVIA, 2006]

El Word Wide Web Consortium (W3C) es la organización internacional que trabaja en el desarrollo de estándares web [W3C, 1994]. Uno de los grupos de trabajo permanente del W3C desde el año 1997 es la Web Accessibility Initiative (WAI) quien se dedica a promover soluciones de accesibilidad en la Web [WAI, 1998]. La WAI publicó en el año 2008 la versión 2.0 de las Pautas de Accesibilidad al Contenido Web (WCAG 2.0), las cuales son consideras estándares internacionales de accesibilidad web.

La WCAG 2.0 se organiza en cuatro principios [WCAG, 2008]:

- *Perceptible*: la información y los componentes de la interfaz de usuario deben ser presentados a los usuarios de modo que ellos puedan percibirlos.
- *Operable*: los componentes de la interfaz de usuario y la navegación deben ser operables.
- *Comprensible*: la información y el manejo de la interfaz de usuario deben ser comprensibles.
- *Robusto*: el contenido debe ser suficientemente robusto como para ser interpretado de forma fiable por una amplia variedad de aplicaciones de usuario.

Los principios están conformados por pautas, las pautas por criterios de conformidad y los criterios por técnicas que pueden ser suficientes o recomendables. [WCAG-TECH, 2016]

En Argentina, en noviembre de 2010 se aprobó la Ley 26.653 de accesibilidad de la información en las páginas web.

La Ley garantiza y amplía el acceso a la información pública a las personas con discapacidad. Mediante esta ley, el Estado Nacional deberá respetar en los diseños de sus páginas web las normas y requisitos sobre accesibilidad de la información que faciliten el acceso a sus contenidos para todas las personas con discapacidad. [HONORABLE CONGRESO DE LA NACIÓN ARGENTINA, 2010]

La Disposición 02/2014 de la Oficina Nacional de Tecnologías de Información (ONTI) de agosto de 2014 aprueba las WCAG 2.0 y establece que el nivel mínimo de conformidad a ser cumplimentado por los organismos, durante el primer período evaluatorio, vigente en octubre del 2018, deberá ser de 50 puntos del nivel A. LA disposición prevé un segundo período evaluatorio dónde el umbral de aprobación se establecerá en 80 puntos. [OFICINA NACIONAL DE TECNOLOGÍAS DE INFORMACIÓN, 2014]

2 Herramientas para la validación de la accesibilidad web

Las herramientas de evaluación son aplicaciones de escritorio que se pueden descargar e instalar en la computadora del usuario o aplicaciones web que se pueden acceder y emplear a través de un navegador. [LUJAN MORA, 2006]

En la Tabla 1 se pueden observar los validadores de código HTML/CSS más usados. Un validador de código verifica si la sintaxis del lenguaje empleado es correcta, por ejemplo, nombre de las etiquetas, valores de las propiedades, etc. [HTML, 2003] [CSS, 2003]

Caracteristicas	W3C Unicorn	W3C Markup Validation Service	W3C CSS Validation Service
Validación.	(X)HTML, CSS (diferentes perfiles medios), Mobile OK y RSS/Atom	(X)HTML	CSS
Tipo de herramienta.	Online (http://validator.w3.org/unicorn/)	Online (http://validator.w3.org/)	Online (https://jigsaw.w3.org/css-validator/)
Opciones.	URL, upload de archivo e inclusión directa de código.	URL, upload de archivo e inclusión directa de código.	URL (HTML con CSS o CSS), upload de archivo e inclusión directa de código.
Idioma.	Castellano	Inglés	Castellano

Tabla 1. Validadores de código HTML/CSS.

En la Tabla 2 se pueden observar los validadores de accesibilidad (WCAG 2.0) más usados. [TAW, 2009][HERA, 2003][WAVE, 2008][EXAMINATOR, 2015] Un validador de accesibilidad comprueba si se cumplen los puntos de verificación de las pautas de accesibilidad que se pueden evaluar de forma automática. Una herramienta de este tipo no sustituye la revisión que realiza un experto, se aplican para una primera revisión. [CARRERAS MONTOTO, 2007]

Caracteristicas	TAW	Hera FFX 2.2	WAVE	eXaminator
Validación.	WCAG 2.0 (A, AA, AAA) y mobileOK	WCAG 2.0 (A, AA, AAA)	WCAG 2.0 (A, AA)	Usa como referencia técnicas de las WCAG 2.0 para otorgar una puntuación a la página.
Tipo de herramienta.	Online (https://www.tawdis.net/) La versión local y la extensión para Firefox validan de acuerdo a las WCAG 1.0.	Extensión para Firefox. La versión online valida de acuerdo a las WCAG 1.0 (http://www.sidar.org/recur/aplica/heraffx.php)	Online (http://wave.webaim.org/	Online, sólo permite revisar una cantidad limitada de páginas por día.
Opciones.	URL. No permiten validación de un sitio completo salvo con su herramienta "TAW Monitor" que es pago.	Valida la página a la que se ha navegado.	URL. No admite validación de sitio completo.	URL, upload de archivos e inclusión directa de código.
Idioma.	Castellano,Inglés y Portugués	Castellano	Inglés	Castellano

Tabla 2. Validadores de accesibilidad.

3 Comparación de las herramientas ARWeb y TAW

Las herramientas más utilizadas, existentes hoy en día, son: TAW (creada por miembros de la W3C) [TAW, 2009] y HERA (creada por la fundación SIDAR) [HERA, 2003], dado que ambas reportan menor cantidad de falsos positivos y negativos que las otras, generando en el evaluador mayor confianza en estas herramientas. [ACCESS IQ, 2018]
En relación a las herramientas nombradas no se tiene pleno conocimiento de qué y cómo hacen la evaluación de las páginas web, siendo normal en un proceso de evaluación, obtener distintos resultados sobre una misma web evaluada. Es por ello, que se creó la herramienta ARWeb [ARWEB, 2017], en el marco de un proyecto de investigación del Instituto de Tecnología (INTEC) de la Universidad Argentina de la Empresa (UADE). [INTEC, 2018]

ARWeb es una aplicación de escritorio desarrollada con Swing del lenguaje Java. Es una herramienta libre y abierta que cuenta con toda la documentación para que otros desarrolladores puedan continuar extendiendo el framework [ARWEB, 2017]. En la Figura 1, puede observarse la ventana principal de la herramienta que presenta cuatro solapas:

- Página Web, donde se solicita la dirección URL del recurso a verificarse.
- Archivo, donde se solicita examinar en el sistema de archivos, un archivo de extensión html a evaluar.
- Código, donde se solicita la edición directa del código HTML.
- URLs., se habilita una lista con todos los links identificados en la dirección URL solicitada en la solapa Página Web.

Figura 1. Ventana principal del Sistema ARWeb.

ARWeb comprueba la accesibilidad web basándose en las pautas de las WCAG 2.0 [WCAG, 2008] y el sistema de puntaje establecido en la disposición 02/2014 de la ONTI para calcular el nivel mínimo de conformidad. [OFICINA NACIONAL DE TECNOLOGÍAS DE INFORMACIÓN, 2014]. En la Figura 2, se muestra la ventana

de resultados de la aplicación ARWeb, donde se observa la información de análisis de accesibilidad, el histórico de verificación del recurso y los resultados generales.

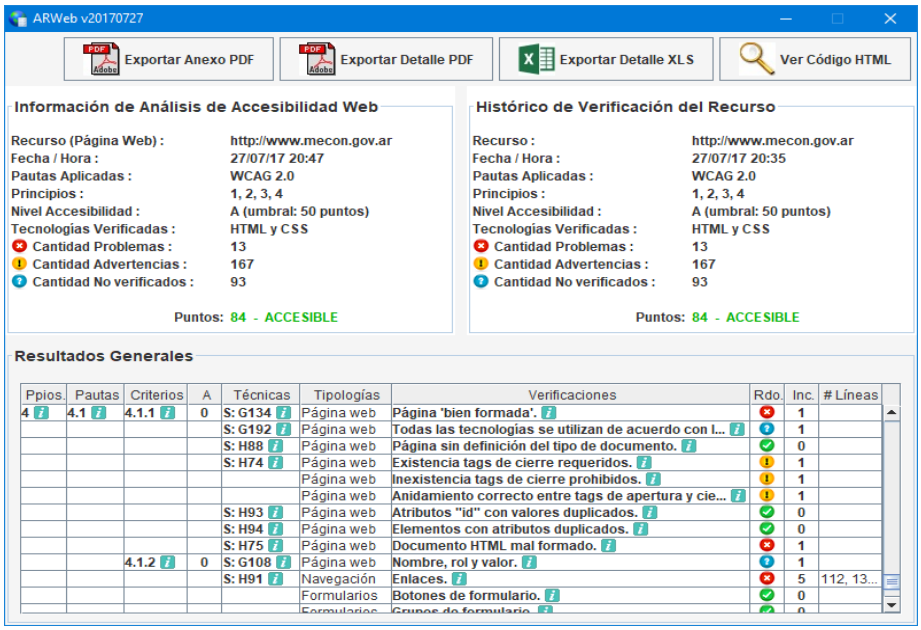


Figura. 2. Ventana de resultados del Sistema ARWeb.

El puntaje de accesibilidad se calcula sumando 4 puntos por cada criterio de aceptación si y sólo si todas las técnicas suficientes asociadas son superadas.

Los cuatro puntos surgen de la siguiente regla de tres simple aplicada:
En el nivel de accesibilidad A existen 25 criterios de conformidad que, de cumplirse todos sobre un recurso analizado, se obtiene 100 (cien) puntos. Por lo tanto:

25 criterios de conformidad --------> 100 puntos
1 criterio de conformidad ---------> ?

? = (1 criterio de conformidad x 100 puntos) / 25 criterios de conformidad
= 4 puntos

El usuario que desea verificar qué tan accesible puede llegar a ser una página HTML en particular, puede guiarse con el valor que arroje los Puntos de Accesibilidad. Si el valor es igual o supera los 50 puntos, el recurso es Accesible, caso contrario, el recurso es No Accesible. [OFICINA NACIONAL DE TECNOLOGÍAS DE INFORMACIÓN, 2014]

Al código HTML del recurso se accede a través del botón “Ver Código HTML”, que muestra al usuario el código fuente analizado destacando las incidencias de Problemas, Advertencias y No verificados por cada línea.

Los reportes que se pueden generar desde ARWeb son:

- *Resultados de Niveles Mínimos de Conformidad en formato PDF:* resultados conforme el modelo del Anexo II de la Disposición 2/2014 de la ONTI.
- *Resultados del Detalle de Análisis en formato PDF:* resultados detallados del análisis de Accesibilidad Web; es decir, en el caso de falla de alguna técnica, muestra el código HTML donde falló y las recomendaciones de accesibilidad para arreglarlo.
- *Resultados del Detalle de Análisis en formato XLS:* el archivo contiene puntajes y fórmulas para que un evaluador ajuste el análisis web y la planilla los recalcule automáticamente.

En las Figuras 3.a, 3.b y 3.c se muestran los listados generados automáticamente por ARWeb.

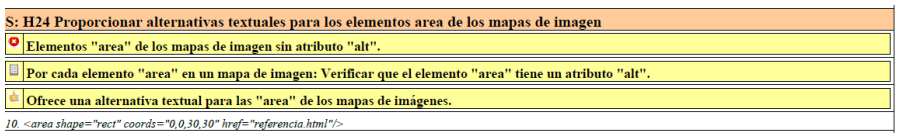
…

Principios	Pautas	Criterios de Conformidad	sub	A
1 Perceptibilidad.	1.1 Alternativas textuales.	1.1.1 Contenido no textual.		0
	1.2 Contenido multimedia dependiente del tiempo.	1.2.1 Sólo audio y sólo vídeo (pregrabado).		4
		1.2.2 Subtítulos (pregrabados).		4
		1.2.3 Audiodescripción o alternativa multimedia (pregrabada).		4
	1.3 Adaptabilidad.	1.3.1 Información y relaciones.		4
		1.3.2 Secuencia significativa.		4
		1.3.3 Características sensoriales.		4
	1.4 Distinguible.	1.4.1 Uso del color.		4
		1.4.2 Control de audio.		4
		subtotal	32	32
2 Operabilidad.	2.1 Accesible a través del teclado.	2.1.1 Teclado.		4
		2.1.2 Sin trampa de teclado.		4
	2.2 Tiempo suficiente.	2.2.1 Límite de tiempo ajustable.		4
		2.2.2 Pausar, detener, ocultar.		4
	2.3 Ataques.	2.3.1 Tres destellos o por debajo del umbral.		4
	2.4 Navegable.	2.4.1 Saltar bloques.		4
		2.4.2 Página titulada.		4
		2.4.3 Orden de foco.		4
		2.4.4 Propósito de un vínculo (en su contexto).		4
		subtotal	36	36
3 Comprensibilidad.	3.1 Legible.	3.1.1 Idioma de la página.		0
	3.2 Predecible.	3.2.1 Con foco.		4
		3.2.2 Con entrada de datos.		4
	3.3 Ayuda a la entrada de datos.	3.3.1 Identificación de errores.		4
		3.3.2 Instrucciones o etiquetas.		4
		subtotal	16	16
4 Robustez.	4.1 Compatible.	4.1.1 Interpretación.		4
		4.1.2 Nombre, rol, valor.		4
		subtotal	8	8
		TOTAL	92	

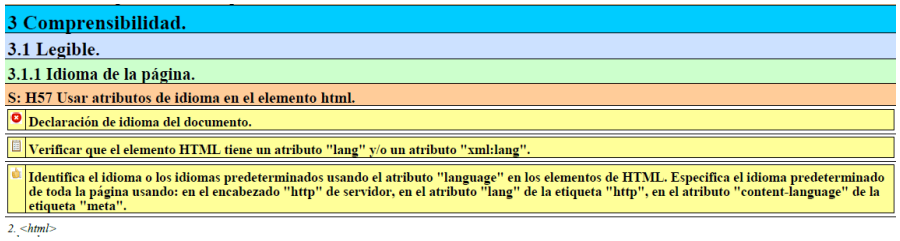
Información expuesta conforme el modelo del Anexo II de Disposición 2/2014 de la ONTI.

Figura 3. a. Resultados de Niveles Mínimos de Conformidad en formato PDF.

...

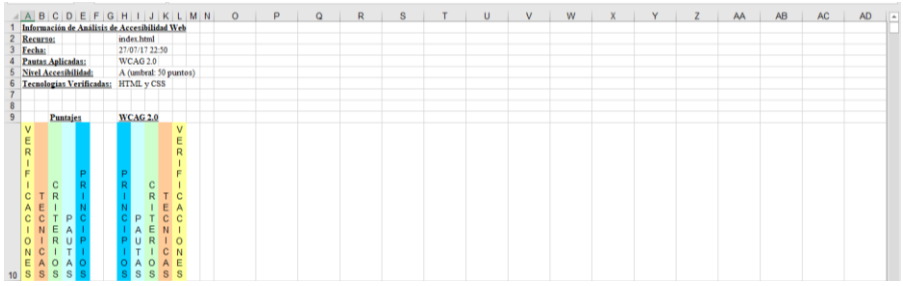


...

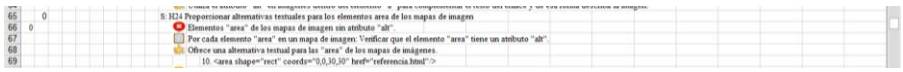


...

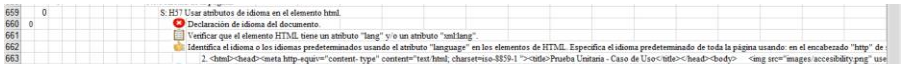
Figura 3. b. Resultados del Detalle de Análisis en formato PDF.



...



...



...

Figura 3. c. Resultados del Detalle de Análisis en formato XLS.

TAW es un analizador basado en las Pautas de Accesibilidad al Contenido Web 2.0 (WCAG 2.0). TAW es la herramienta propuesta por la W3C. Se la puede encontrar en la siguiente dirección http://www.tawdis.net/ [TAW, 2009]

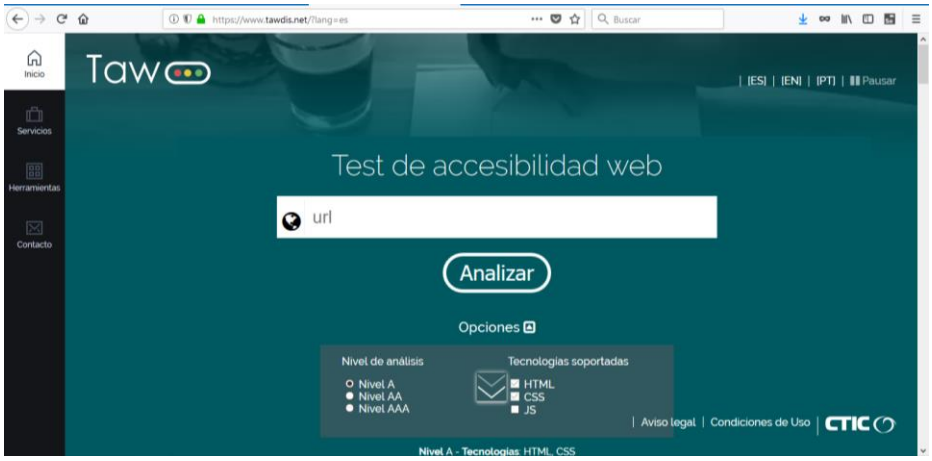


Figura 3. Ventana principal del Sistema TAW.

Luego de introducir una URL y pulsar el botón Analizar, se muestra un resumen de los resultados del análisis (Ver Figura 4)

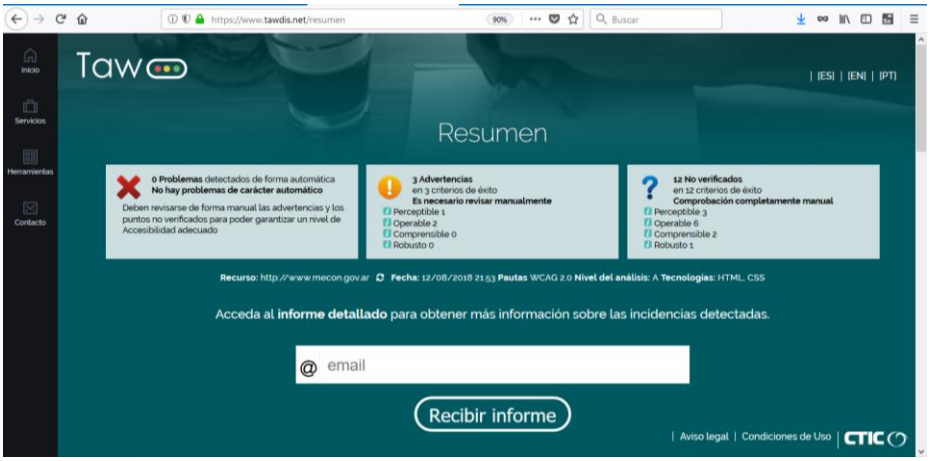


Figura 4. Ventana de resultados (resumen) del Sistema TAW.

En la columna de la izquierda, la X en rojo, indica los problemas automáticos que ha detectado la herramienta y que se deben corregir.

En la columna central, el signo de admiración, indica las advertencias, es decir, posibles errores de accesibilidad que se deben de comprobar en forma manual si realmente son un error o no.

En la columna de la derecha, el signo de interrogación, indica los criterios que no se han podido comprobar de forma automática y por lo tanto se deben hacer de forma manual.

Además del resumen, se puede continuar leyendo y se accede al detalle del análisis, es decir, se pueden ver los errores y advertencias detectados de forma automática por la herramienta, de acuerdo a los 4 principios (perceptible, operable, comprensible y robusto), pautas y criterios de conformidad enunciados por la WCAG 2.0 [WCAG, 2008], como se muestra en la Figura 5.

Figura 5. Ventana de resultados (detalle) del Sistema TAW.

En la Tabla 3 se puede observar un cuadro comparativo respecto de las características de ambas herramientas:

Caracteristicas	**TAW**	**ARWeb**
Pautas.	WCAG 1.0, WCAG 2.0, mobileOK	WCAG 2.0
Tecnologías soportadas.	HTML, CCS, Javascript	HTML y CSS
Técnicas HTML.	Parcial	Completa
Opciones.	URL. No permite validación de un sitio completo.	URL, upload de archivo e inclusión directa de código.
Selección del Nivel de Análisis.	Si	No
Selección de Principios de Accesibilidad	No	Si
Agrupación de resultados por Principios.	Si	Si
Niveles de Conformidad WAI.	A, AA, AAA	A, AA, AAA
Posibilidad de extender herramienta.	No	Si

Consejos de desarrollo.	No	Si
Indicación del Grado de Accesibilidad.	No	Si, en función de la Legislación Argentina.
Indicación del problema en el código fuente.	Si	Si
Visualización en formato PDF y XLS.	No	Si
Registro histórico de verificaciones.	No	Si
Multi-idiomas	Si (ES, EN, PT)	No (ES)

Tabla 3. Cuadro comparativo de TAW y ARWeb.

4 Conclusiones

La evaluación de accesibilidad es un paso importante para lograr una web accesible. Que una web sea accesible beneficia a todos los usuarios, mejorando el acceso web en general.

Tras comparar las dos herramientas de análisis de accesibilidad web, pudo observarse que ARWeb es más detallada y brinda mayor información y herramientas al evaluador, lo que le permite generar una evaluación más precisa. Conforme al nivel A, máximo nivel a evaluar según la norma vigente en Argentina, ARWeb calcula el puntaje necesario para establecer nivel mínimo de conformidad. Esta información es visualizada a través de diferentes reportes.

La característica más lograda de ARWeb es la transparencia de la evaluación ya que expone en sus distintos reportes el resultado de cada técnica, incluyendo una propuesta de solución (en caso de ser pertinente) y una explicación del contexto donde se desarrolla la técnica. La evaluación realizada por ARWeb es fácilmente comprobable.

TAW es una herramienta que sirve de ayuda a los desarrolladores de páginas web y consigue que su trabajo cumpla las normas de accesibilidad WCAG 2.0 en todos sus niveles (A, AA y AAA), no brinda ningún tipo de valoración del grado de accesibilidad, tampoco se tiene certeza de qué y cómo evalúa las páginas ni permite exportar reportes de análisis.

Ambas herramientas contribuyen a la detección de los problemas de accesibilidad para que un sitio web pueda ser evaluado, son de gran ayuda, pero no sustituyen la evaluación manual.

Realizar un análisis automático, permite una evaluación más rápida y así tener un resultado preliminar o primera impresión de la accesibilidad de una página. Sin embargo, las herramientas deben entenderse como una ayuda, no como un análisis completo, ya que pueden no detectar problemas importantes que sólo pueden ser

detectados a través de una revisión manual. Por esta razón, después de haber validado el código a través de una o varias herramientas automáticas (la WAI recomienda el uso de al menos dos herramientas para la evaluación de las páginas), es necesario complementarlo con una evaluación manual.

Referencias

ACCESS IQ. Useful accessibility tools. [en línea]. 2018. [consulta 06 marzo 2018]. <http://www.accessiq.org/create/content/useful-accessibility-tools>

ARWEB. Herramienta de Accesibilidad Web. [en línea]. 2017. [consulta 09 marzo 2018]. <https://sites.google.com/site/uadeticarweb/>

CARRERAS MONTOTO, Olga. Usable y Accesible. [en línea]. 2007. [consulta 01 agosto 2018]. <http://olgacarreras.blogspot.com.ar/>

CSS Validator. [en línea]. 2003. [consulta 8 febrero 2018]. <http://www.css-validator.org/>

EXAMINATOR. Servicio en línea para evaluar de modo automático la Accesibilidad de una Página Web [en línea]. 2015. [consulta 12 enero 2018]. <http://examinator.ws/>

HERA. Hojas de Estilo para la Revisión de la Accesibilidad. [en línea]. 2003. [consulta 16 agosto 2018]. <http://www.sidar.org/hera/>

HONORABLE CONGRESO DE LA NACIÓN ARGENTINA. Ley 26.653 de Accesibilidad de la Información en Páginas Web. Buenos Aires. Argentina. [en línea]. 2010. [consulta 03 enero 2018]. <http://www.infoleg.gob.ar/infolegInternet/anexos/175000-179999/175694/norma.htm>

HTML Validator.Markup Validation Service. [en línea]. 2003. [consulta 7 enero 2018]. <https://validator.w3.org/>

INTEC. Instituto de Tecnología UADE. Buenos Aires. Argentina. [en línea]. 2018. [consulta 24 mayo 2018]. < https://www.uade.edu.ar/investigacion/instituto-de-tecnologia-intec/proyectos>

LUJAN MORA, Sergio. Accesibilidad en la Web. [en línea]. 2006. [consulta 10 julio 2018]. <http://accesibilidadenlaweb.blogspot.com.ar/>

OFICINA NACIONAL DE TECNOLOGÍAS DE INFORMACIÓN. Decreto 2/2014: Norma de Accesibilidad Web 2.0. Buenos Aires. Argentina [en línea]. 2014. [consulta 15 julio 2018]. <http://servicios.infoleg.gob.ar/infolegInternet/anexos/230000-234999/233667/norma.htm>

REVILLA MUÑOZ, Olga. *WCAG 2.0 de forma sencilla*. 1a. ed. Madrid: Itákora Press, 2013. 170 p. ISBN 978-84-614-6410-4.

SHAWN LAWTON, Henry. Easy Checks – A First Review of Web Accessibility. [en línea]. 2014. [consulta 10 abril 2018]. <https://www.w3.org/WAI/eval/preliminary>

SEGOVIA, Claudio. *Accesibilidad en Internet*. 1a. ed. California: Creative Commons, 2006. 212 p.

TAW. Servicios de Accesibilidad y Movilidad Web [en línea]. 2009. [consulta 06 enero 2018]. <http://www.tawdis.net/index.html?lang=es>

W3C. World Wide Web Consortium. [en línea]. 1994. [consulta 03 enero 2018]. <https://www.w3.org/>

WAI. Web Accessibility Initiative. [en línea]. 1998. [consulta 03 enero 2018]. <https://www.w3.org/WAI/>

WAVE. Web Accesibility Evaluation Tool. [en línea]. 2008. [consulta 14 febrero 2018]. <http://wave.webaim.org/>

WCAG. Web Content Accessibility Guidelines 2.0. [en línea]. 2008. [consulta 03 enero 2018]. <https://www.w3.org/TR/WCAG20>

WCAG-TECH. Techniques for WCAG 2.0. [en línea]. 2016. [consulta 03 enero 2018]. <https://www.w3.org/TR/WCAG20-TECHS/>

,,,

Redes sociales y privacidad

Santiago Serra, Rosana Giménez, Pablo Sandez y Miguel Méndez-Garabetti

Laboratorio de Investigación en Ciencia y Tecnología (LabICyT-UDA), Facultad de Ciencias Sociales y Administrativas, Universidad del Aconcagua, Mendoza, Argentina
santiser.1999@hotmail.com

Resumen Actualmente la sociedad está inmersa en el uso y consumo de las redes sociales, esto ha provocado nuevos problemas de privacidad y comunicación. Permanentemente se crean perfiles y cuentas con gran cantidad de información personal en estos tipos de software, subiendo datos ya sean sobre actividades que realizan los usuarios, dónde se encuentran geograficamente, cuáles son sus deseos y gustos, entre otros. A partir de estos comportamientos y actividades surgen las siguientes preguntas: ¿sabe el usuario qué sucede con sus datos una vez cargados en redes sociales? Además: ¿son los datos utilizados de manera incorrecta? Y, de ser así: ¿es responsabilidad de las empresas de software creadoras y dueñas de las redes sociales o es responsabilidad de la falta de informacion que el usuario tiene sobre el cuidado e importancia de sus datos? La investigación tiene como objetivo principal estudiar los términos y condiciones que imponen las redes sociales para su uso y los datos que son solicitados al usuario, analizando en qué nivel el público que las utiliza tiene conocimiento y conciencia sobre esta información. Se utilizarán encuestas para reunir datos y opiniones de usuarios de redes sociales, se confeccionarán estadísticas y, dependiendo del resultado, estará la posibilidad de formular una solución o una mejoría a la situación observada.

Palabras clave redes sociales - privacidad - comunicación - datos personales

1 Introducción

Utilizando internet, la base de las redes sociales, en la actualidad las personas consumimos y subimos gran cantidad de datos en la red, mayoritariamente en redes sociales. Según el estudio “Digital 2019: Argentina” de We are Social y Hootsuite [1] los usuarios activos en Internet han aumentado un 20 por ciento, es decir, aproximadamente siete millones de personas y han aumentado los usuarios en el uso de redes sociales en dispositivos móviles un 3.3 por ciento, es decir, aproximadamente un millón de personas. Teniendo de ejemplo este estudio y el estudio de los mismos autores pero en su versión global [2] se observa claramente el constante crecimiento en el uso de internet y por consiguiente en aplicaciones y redes sociales.

Esto lleva a pensar en la gran cantidad de información que es cargada, utilizada, eliminada, etcétera de las redes sociales. Actualmente los datos son el

recurso más valioso en la tierra y por la misma razón las empresas relacionadas a ellos son las más acaudaladas, esto justifica su importancia desde el apartado económico para las empresas pero también debería resaltar la importancia para el usuario desde el apartado de su seguridad y privacidad.

Ya se han observado casos mediáticos de grandes compañías utilizando datos de sus usuarios de manera incorrecta o sin la responsabilidad necesaria como por ejemplo Cambridge Analytica [3], Facebook [4], entre otros. Aún así, la prioridad que debería tener este tipo de información está infravalorada. Por lo tanto, al usuario de redes sociales, ¿le interesa lo que se realice con sus datos? o , simplemente, ¿no le da importancia y utiliza las redes sociales por gusto y/o necesidad?

2 Texto principal

La hipótesis de la investigación se basa en que el usuario de redes sociales no tiene conocimiento suficiente o ninguno sobre el uso de sus datos y los términos y condiciones que las redes sociales imponen para su respectivo uso. Para corroborar esta hipótesis se realizará una encuesta breve y concisa a personas usuarias o no usuarias de redes sociales.

La encuesta es la siguiente:

¿Edad?*

- Menor de 14 años
- Menor de 14 años
- Entre 14 y 18 años
- Entre 19 y 24 años
- Entre 25 y 34 años
- Entre 35 y 44 años
- Entre 45 y 55 años
- Mayor de 55 años

¿Qué redes sociales utiliza?*

- Instagram
- Facebook
- Twitter
- Snapchat
- No utilizo redes sociales
- Otro:

Aproximadamente, ¿cuánto tiempo diría que utiliza redes sociales al día?*

- Menos de 1 hora
- Entre 1 y 3 horas
- Entre 3 y 5 horas
- Más de 5 horas
- No utilizo redes sociales

¿Con qué fin utiliza redes sociales?*

- Entretenimiento
- Social
- Negocios
- Noticias
- Otro:

¿Considera que es necesario saber utilizar redes sociales en la actualidad?*

- Totalmente en desacuerdo
- En desacuerdo
- Neutral
- De acuerdo
- Totalmente de acuerdo

¿Sube o publica datos personales como fotos, ubicaciones, gustos, entre otros en las redes sociales? *

- Si
- No

¿Considera que es seguro publicar estos tipos de datos?*

- Totalmente en desacuerdo
- En desacuerdo
- Neutral
- De acuerdo
- Totalmente de acuerdo

¿Por qué?

...

Al crear una cuenta en alguna red social, ¿lee los términos y condiciones?*

- Si
- No

Una vez subidos los datos, ¿sabe qué sucede con ellos?*

- Si
- No
- Parcialmente

¿Cuál es su opinión general sobre las redes sociales?

...

3 Resultados de la encuesta

Se utilizó la herramienta de Google Forms para reunir los datos y, filtrando ciertos resultados, mostrarlos de manera práctica y concisa.

Primer pregunta

La edad es un factor importante en lo relacionado a las redes sociales, según la encuesta realizada un 66% son jóvenes de entre 19 a 24 años y esto demuestra que en la actualidad la mayor´ıa de usuarios de redes sociales son personas de aproximadamente esa edad. Esto podría relacionarse a cómo ha cambiado la comunicación en los últimos años siendo estas personas las principales razones de ese mismo cambio.

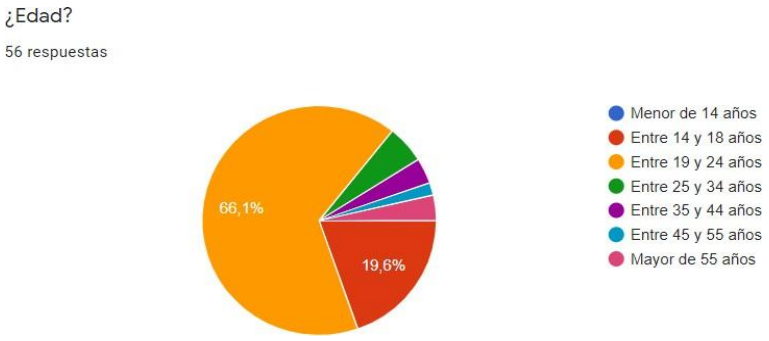


Fig. 1. Edad de los encuestados.

En la Fig. 1 se observa que un 66% de encuestados tienen entre 19 y 24 años y un 19% tiene entre 14 y 18 años.

Segunda pregunta

Las redes sociales más utilizadas son las siguientes:

En la Fig. 2 se observa que la red social más utilizada es Instagram, siguiendole Facebook y Twitter. Se realizó esta pregunta para conocer qué redes sociales son las más utilizadas y poder relacionar siguientes preguntas referidas al fin con que se utilizan las mismas.

Tercera pregunta

El tiempo aproximado que las personas utilizan redes sociales es:

En la Fig. 3 se observa que el 42.9% utiliza redes sociales entre 1 y 3 horas al día, un 32.1% las utiliza entre 3 y 5 horas al día. Se realizó esta pregunta ya que era importante ahondar sobre la cantidad de tiempo que las personas están conectadas y usando estas redes sociales. Además concuerda con el estudio previamente citado de DIGITAL 2019: ARGENTINA en donde el tiempo promedio es de 3 horas 18 minutos diarios.

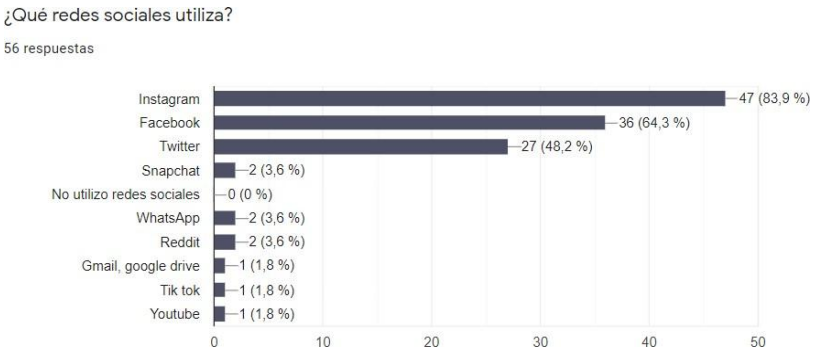


Fig. 2. Redes sociales más utilizadas.

Fig. 3. Tiempo aproximado que las personas utilizan redes sociales.

Cuarta pregunta

El fin con que se usan las redes sociales es:

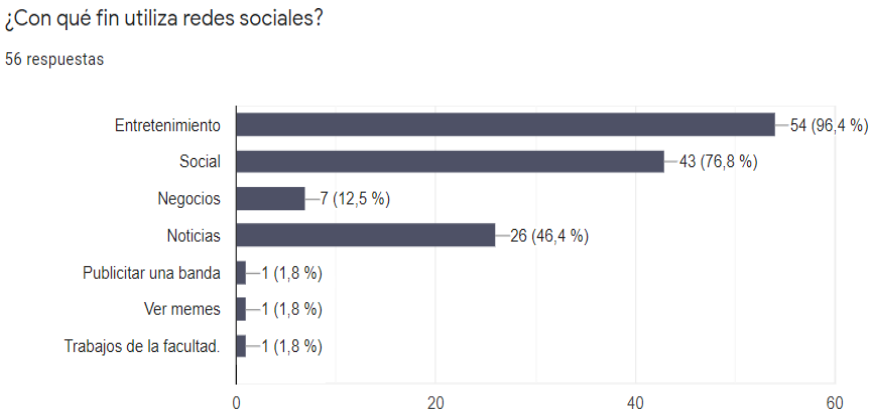


Fig. 4. Fin con que se utilizan redes sociales.

En la Fig. 4 se observa que el entretenimiento es el fin para el que más se utilizan redes sociales siguiéndole un fin social.

Quinta pregunta

Esta pregunta se realizó con la intención de conocer la opinión de los encuestados sobre si las redes sociales se han vuelto una herramienta necesaria o solo una herramienta de uso específico.

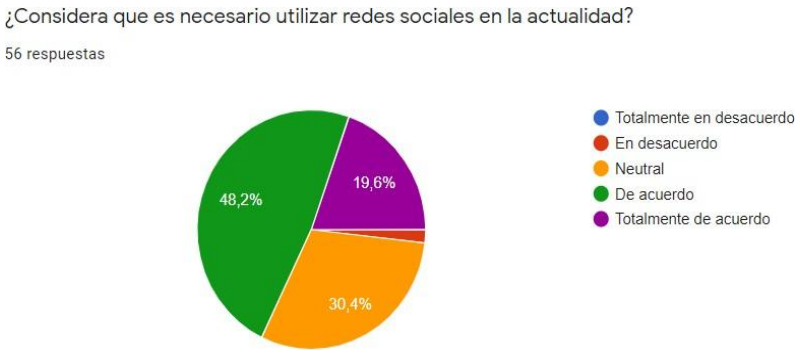


Fig. 5. Personas que consideran necesarias o no a las redes sociales.

En la Fig. 5 se observa que el 48.2% coincide con que es necesario utilizar redes sociales mientras que un 30.4% se mantiene en una postura neutral. Los resultados demuestran que la mayor´ıa de personas actualmente considera a las redes sociales como una herramienta necesaria para realizar sus actividades pero no imprescindible.

Sexta pregunta

Los resultados son:

Fig. 6. Cantidad de personas que suben datos personales a redes sociales.

En la Fig. 6 se observa que el 66.1% de las personas suben datos personales a sus redes sociales. El resultado demuestra que gran cantidad de personas suben datos personales a redes sociales y hace reflexionar sobre la cantidad de información que circula en Internet.

Séptima pregunta

Se realizó esta pregunta para conocer qué opinan los usuarios sobre la seguridad que ofrecen las redes sociales en cuanto al uso de los datos que se suben a las redes.

En la Fig. 7 se observa que un 44.6% de personas se encuentra en una postura neutral sobre si considera seguro subir este tipo de datos.Según los resultados se demuestra un desconocimiento o cierta inseguridad sobre el uso de sus datos. Mientras que el resto concuerda con la pregunta y afirma que no considera que sea seguro subir datos personales a las redes sociales.

Octava pregunta

Se realizó como pregunta no obligatoria para que el encuestado pudiera dar una opinión abierta sobre el tema. Algunas respuestas fueron:

Refiriéndose a una respuesta negativa sobre seguridad:

“Porque lamentablemente nunca uno sabe si los datos o información que sube a las redes sociales en verdad está protegida, cualquier persona puede descargar

Fig. 7. Personas que no considera seguro publicar datos personales o si a redes sociales.

una foto de facebook y utilizarla como quiera, lo cual puede resultar muy peligroso."

"Se corre peligro al subir cosas personales a una red que básicamente es un público masivo, pero si uno sabe manejar la privacidad de su cuenta puede compensar ese peligro aunque no del todo."

Novena pregunta

Se realizó esta pregunta para conocer si los términos y condiciones de las redes sociales son un impedimento para la creación de cuentas y uso de las mismas o si son ignoradas y/o aceptadas con tal de poder utilizar el servicio.

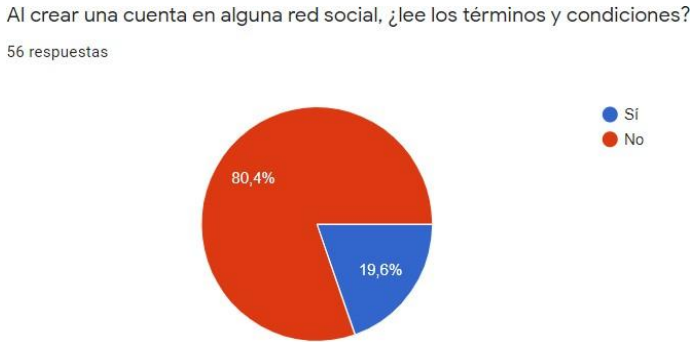


Fig. 8. Cantidad de personas que leen los términos y condiciones al crear una cuenta en redes sociales.

En la Fig. 8 se observa que un 80.4% de personas no leen los términos y condiciones al crear una cuenta en algún tipo de red social. Según los resultados

la mayoría de personas no leen los términos y condiciones. Para solucionar esto se podría actualizar o cambiar de cierta manera la forma en que estos términos son presentados al usuario para que la documentación sea más fácil y amena de revisar y posteriormente aceptar o rechazar.

Décima pregunta

Se realizó esta pregunta para conocer si los usuarios tienen conocimientos sobre la traza digital[5] de sus datos.

Fig. 9. Cantidad de personas que tienen conocimiento o no sobre lo que sucede con sus datos una vez subidos a redes sociales.

En la Fig. 9 se observa que un 41.1% de personas tienen conocimientos parciales sobre lo que sucede sobre sus datos una vez subidos a redes sociales.

Décimo primer pregunta

Como conclusión a la encuesta se solicitó una opinión general de los usuarios sobre las redes sociales. Cómo las siguientes:

"Son una forma de estar conectados/as con otros/as, y estar al tanto de otras cosas que suceden en el mundo, nos sirven de entretenimiento y demás. Pero también es cierto que al no ser totalmente seguras, el/la usuario/a debe tener cierto grados de responsabilidad a la hora de compartir cualquier tipo de información con otros/as. También considero que se deben cambiar algunas normas por parte de los creadores/as de las mismas, es decir, que proporcionen mayor seguridad."

"Es una buena forma de conectar a las personas y no perder contacto con ellas, pero al ser mal usado tiende a causar el efecto contrario."

4 Conclusiones

A partir de las 57 encuestas realizadas, 37 (66%) personas suben datos personales como fotos, gustos, ubicaciones, entre otros, a redes sociales. De igual manera

51 (91%) personas no consideran seguro subir ese tipo de datos a redes sociales, y además, 45 (80%) personas no leen los términos y condiciones que se imponen para utilizarlas.

Con esta información se observa una clara falta de información y concientización sobre seguridad e importancia sobre datos privados en redes sociales. En la actualidad la mayoría de la población se está acostumbrando al uso de redes sociales e internet en general, pero aún así, el uso no justifica el conocimiento y/o aprendizaje suficiente sobre las herramientas que se están usando.

En este trabajo de investigación de cátedra se demuestra como correcta la hipótesis y da la justificación para seguir avanzando en el futuro con la investigación, ampliándola, y así lograr un mayor entendimiento de por qué sucede esto y, si es posible, formular una mejora, alternativa o solución a lo observado.

Referencias

1. DIGITAL 2019: ARGENTINA - We are Social - HootSuite, https://datareportal.com/reports/digital-2019-argentina
2. DIGITAL 2019: GLOBAL DIGITAL YEARBOOK - We are Social - HootSuite, https://datareportal.com/reports/digital-2019-global-digital-yearbook?utm_source=Reports&utm_medium=PDF&utm_campaign=Digital_2019&utm_content=Global_Overview_Promo_Slide
3. The Great Hack - Netflix
4. Facebook–Cambridge Analytica data scandal, https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal
5. Big Data: el Valor de la Información Personal y la Privacidad - Carlos Fernando Arenas Fonseca, https://www.jdc.edu.co/revistas/index.php/rciyt/article/view/75/71

,,

Modelo para la generación de datamarts en microempresas

Gabriel Dávila[1], Silvina Migani[2] y María Inés Lund[2]

[1]Graduado Lic. en Sistemas de Información, FCEFN-UNSJ
[2]Departamento e Instituto de Informática, FCEFN-UNSJ
{gabriel.davila32, silvina.migani}@gmail.com, mlund@iinfo.unsj.edu.ar

Resumen. Desde hace ya muchas décadas las organizaciones poseen sistemas automatizados que dan soporte a sus operaciones diarias y consecuentemente generan una gran cantidad de datos digitales propios que contienen información "silenciosa" de considerable valor. Por otra parte, hoy en día se encuentran inmersas en un entorno complejo y altamente competitivo, donde un factor clave para el éxito es contar con información oportuna, que las asista en la toma de decisiones. La inteligencia de negocios es el área que comprende un conjunto de tecnologías y herramientas, dentro de las cuales se encuentran los almacenes de datos y los datamarts. Ellos permiten aprovechar esos datos operacionales generando información y/o conocimiento, imprescindibles para afrontar una toma de decisiones segura. Este trabajo propone un modelo simplificado para la construcción de datamarts en microempresas, y así poder replicar dicho proceso de manera sencilla y ordenada.

Palabras claves: Inteligencia de Negocios, Datamarts, Microempresas, Modelo.

1 Introducción

Desde la década del 1970, las bases de datos relacionales[1] (BD) fueron adoptadas por la mayoría de las organizaciones para almacenar sus datos operacionales, es decir, los datos generados por las transacciones diarias conocidas como Procesamiento de Transacciones En Línea (OLTP por sus siglas en inglés de On Line Transaction Processing). El principal objetivo de las BD fue soportar un gran número de accesos concurrentes, en un tiempo razonable; además de garantizar consistencia de los datos, manejo de transacciones, control de concurrencia y técnicas de recuperación. Todos estos requerimientos fueron resueltos eficientemente por los Sistemas de Gestión de Bases de Datos transaccionales (SGBD) [1].

[1]Base de datos relacionales: también llamadas transaccionales, operacionales, operativas, organizacionales y tradicionales.

Sin embargo, las necesidades fueron evolucionando, y hoy en día, además de gestionar las transacciones del día a día, la alta competitividad del mercado hace necesario contar con información oportuna para poder tomar decisiones correctas y poder alcanzar los objetivos del negocio [2]. Las BD no son apropiadas para dar respuesta a esta nueva demanda debido a que todos los datos necesarios no siempre están en línea, y además están diseminados en diferentes sistemas, más allá que al trabajar en línea los tiempos de respuesta pueden no ser aceptables. Consecuentemente surgieron tecnologías tales como los Almacenes de Datos (AD), en inglés, conocidos como Data Warehouse y el Procesamiento Analítico en Línea (OLAP, por sus siglas en inglés de On Line Analytical Processing), que conjuntamente conforman un nuevo paradigma llamado Inteligencia de Negocios (IN). Estos sistemas están optimizados para consultas de alto rendimiento ya que a menudo este tipo de consultas requieren que se analicen cientos o cientos de miles de transacciones [3].

Los sistemas de soporte a las operaciones diarias dentro de las organizaciones, en general están automatizados y maduros hace ya mucho tiempo. Esto ha provocado una creciente generación de datos digitales propios que contienen información "silenciosa" de considerable valor. Por lo tanto, es necesario extraer y hacer visible esa información/conocimiento que posibilite emprender una toma de decisiones basada en hechos concretos, y no en la mera intuición.

En el mundo de los negocios a gran escala ya está instalado el concepto de IN como una herramienta fundamental, les permite sacar provecho de todos los datos que poseen: de clientes, de proveedores, y hasta de sus competidores; todo esto con el propósito de alcanzar ventajas competitivas en un mercado dinámico y hostil. Un caso de éxito ampliamente conocido es la corporación multinacional de tiendas de origen estadounidense, Walmart. Las grandes empresas han podido acceder más rápidamente a estas nuevas tecnologías, esto se debe por un lado porque su presupuesto se los permite, pero también porque se han dado cuenta que tienen mucho que ganar al aplicarlas.

Las Pequeñas y Medianas Empresas (PYMEs), que representan el 98% del mapa empresarial argentino, suelen tomar sus decisiones en base a un conocimiento empírico de su negocio [4]. Además, los desafíos son cada vez más complejos y difíciles de abordar si no se cuenta con herramientas adecuadas. Sin embargo, hoy en día las PYMEs también pueden abordar sus estrategias de negocio de manera más eficaz, incorporando tecnologías de IN que se ajusten a sus necesidades y limitaciones [5]. En este trabajo se presenta una propuesta de modelo para construir artefactos de IN en una micro empresa del medio, concretamente Data Mart de ventas, que permitan apoyar la toma de decisiones, usando tecnologías open source, al alcance de cualquiera de ellas.

2 Conceptos Generales

2.1 Inteligencia de Negocios (IN)

Las organizaciones de hoy enfrentan desafíos cada vez más complejos en términos de gestión y resolución de problemas para alcanzar sus objetivos. Esta situación obliga a utilizar herramientas de análisis que puedan apoyar mejor sus decisiones. La Inteligencia de Negocios comprende una colección de metodologías, procesos, arquitecturas y tecnologías que transforman los datos en información útil y significativa para la toma de decisiones. Los sistemas de inteligencia de negocios y de apoyo a la decisión brindan asistencia a los gerentes en diversos niveles organizacionales para analizar información estratégica. Estos sistemas recopilan grandes cantidades de datos y los reducen a una forma que se puede usar para analizar el comportamiento organizacional. Esta transformación de datos comprende un conjunto de tareas que toman los datos de las fuentes y, a través de procesos de extracción, transformación, integración y limpieza, almacenan los datos en un repositorio común llamado almacén de datos [3].

2.2 Almacén de Datos (AD)

Los almacenes de datos se han desarrollado y desplegado como parte integral de los sistemas de soporte de decisiones para proporcionar una infraestructura que permita a los usuarios obtener respuestas eficientes y precisas a consultas complejas.

El término Almacén de Datos, Data Warehouse en inglés, fue mencionado por primera vez por Bill Inmon [6] en la década de los 90', y lo definió de la siguiente manera:

"Un almacén de datos es una colección de datos orientados por temas, integrados, no volátiles y variables en el tiempo, que se utilizan para dar soporte a la toma de decisiones estratégicas…"

De la definición se desprenden los siguientes conceptos [1]:

- Orientados por tema: Las BD están orientadas a los procesos, es decir, dan soporte a los procesos de una organización. Por el contrario, los AD están enfocados a temas de interés para la organización, por ejemplo, al análisis de ventas, la gestión de inventario, etc.
- Integrado: Los datos proceden de distintas fuentes, en general heterogéneas, que pueden ser internas o externas a la organización. Esto hace que se deban resolver problemas de incompatibilidad en cuanto a distintos formatos, codificaciones, sinónimos (campos con diferente nombre, pero igual significado), homónimos (campos con el mismo nombre, pero con distinto significado), etc.
- No volátiles: La durabilidad de los datos está garantizada al no permitir la modificación y eliminación de los datos. Un AD reúne datos que abarcan varios años, generalmente entre 5-10. Los nuevos datos se añaden periódicamente para aumentar el repositorio de datos, en lugar de reemplazarlos.

- Variables en el tiempo: Los datos en el AD siempre se cargan con una referencia temporal. Esta referencia puede ser puntual y concreta, es decir en un instante determinado, o ser válidos en un intervalo de tiempo. Esto indica la posibilidad de tener diferentes valores para la misma información, así como el momento en el que ocurrieron los cambios.

2.3 Data Mart

Un concepto importante a considerar dentro de los AD es el de Data Mart. Un Data Mart es una versión reducida del AD orientado a un tema específico, es decir, satisface las necesidades de un departamento o sección dentro de una organización. Por ejemplo, marketing, clientes, productos, ventas, etc [1].

Otra definición, similar a la anterior: Un DM es un repositorio de información, similar a un AD, pero orientado a un área o departamento específico de la organización (por ejemplo, Compras, Ventas, RRHH, etc.). La diferencia fundamental con un AD es su alcance, éste último cubre toda la organización [7].

3 Modelo Propuesto

Este modelo permite obtener DM en forma ordenada y sencilla en casos de pequeña envergadura. A continuación, en la Fig. 1, se muestra el modelo propuesto utilizando Notación y Modelado de Procesos de Negocios (BPMN, por sus siglas en inglés Business Process Modeling and Notation) [8].

El modelo propuesto fue desarrollado tomando características de los métodos tradicionales y agiles conocidos; corresponde a una primera versión del mismo, se espera que este evolucione y se perfeccione con su uso. El objetivo es brindar una alternativa a las metodologías/métodos de desarrollo propuestos por diversos autores [9], las cuales, en la práctica, muchas veces son difíciles de implementar por pequeños grupos de trabajo en ambientes de micro empresas (industria), como así también en el ámbito académico. Si bien las metodologías existentes se pueden aplicar en estos ambientes, requieren cierto conocimiento y/o experiencia previa difícil de obtener, y generalmente se enfocan en algunos aspectos y no en el "proceso completo" de diseño y desarrollo [9].

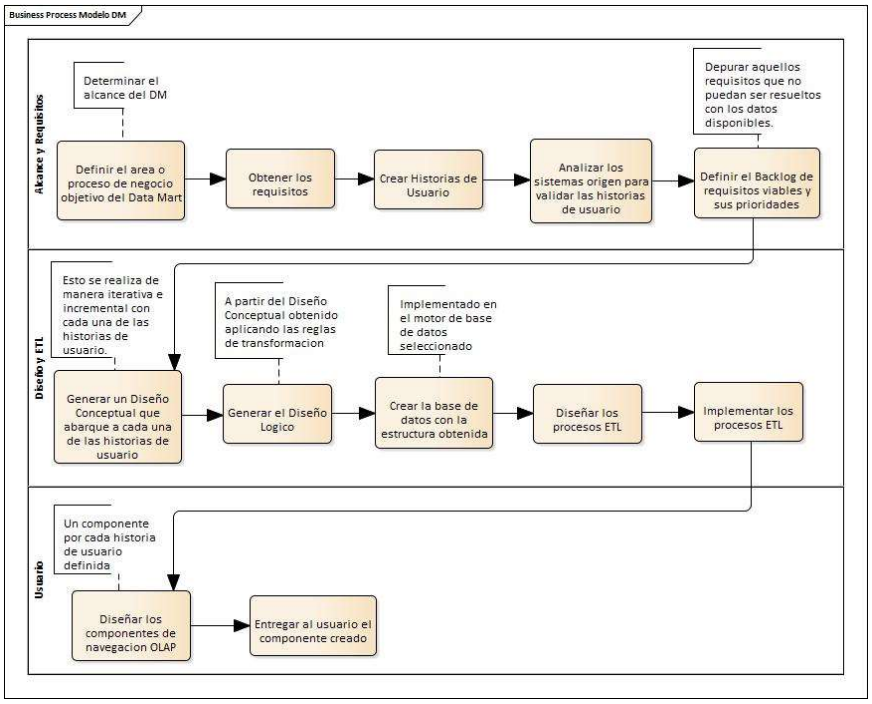


Fig. 1. Modelo para la generación de DM (elaboración propia).

A continuación, una breve explicación del modelo simplificado:

Paso 1: Determinar el alcance/objetivo del DM. Este es el punto de partida de todo el proyecto. Es muy importante tener en claro desde el comienzo, cual o cuales son los objetivos del DM que se desea construir. Generalmente los objetivos están vinculados a algún área de la cual se requiere obtener información, como por ejemplo el área de ventas, compras, recursos humanos, producción, etc. Es importante aclarar que cada iteración de este modelo es aplicable a un solo objetivo. En este paso, si se ha identificado más de uno, se deberá seleccionar uno en cada iteración.

Paso 2: Obtener los requisitos de usuario. Luego de haber definido el objetivo, se procede a relevar las necesidades del usuario en relación al objetivo planteado. Se puede utilizar cualquier técnica conocida para la obtención de requisitos como reuniones, entrevistas, lluvia de ideas, etc. Lo importante, es poder redactarlos en forma de "historias de usuario" [10] para una mejor comprensión. Es muy probable que se llegue a una amplia lista de requerimientos expuestos por el usuario al fin de esta etapa, pero es muy necesario que se identifiquen aquellos requerimientos que están en línea con el objetivo y dejar en "espera" aquellos que no lo estén, para una futura iteración con un objetivo distinto.

En esta etapa, el objetivo puede cambiar debido a que los requisitos planteados por el usuario pueden no coincidir con él, y de esa manera, redefinir el objetivo planteado de manera temprana y no en los pasos siguientes.

Paso 3: Validar las historias de usuarios definidas. Una vez obtenido un listado preliminar de historias de usuarios (alineadas con el objetivo) a desarrollar, se procede a analizar las fuentes de datos disponibles por la microempresa. Es muy importante poder tener una visión detallada (tecnología, documentación, funcionamiento, estructuras de datos, diagramas, etc.) de los orígenes de datos, para poder seleccionar las historias de usuarios viables y definir el backlog de requisitos con el que se va a continuar. El usuario es el encargado de asignarles prioridades a cada historia de usuario. Si al definir el backlog de requisitos se llega a un listado vacío, es muy probable que se tenga que redefinir el objetivo y/o las historias de usuario, pero esta vez con la ventaja de conocer los detalles de los orígenes de datos disponibles, por lo que debería ser un proceso más corto que el anterior.

Paso 4: Generar el Diseño Conceptual del DM. Una vez definido el backlog de requisitos viables, se procede a desarrollar el diseño conceptual del DM. Esta etapa tiene tantas iteraciones como historias de usuarios se obtenga en el backlog. El objetivo acá es obtener un diseño conceptual que sea capaz de responder a todas las historias de usuario definidas en el backlog, siendo capaz de reconocer los elementos propios de los modelos multidimensionales como dimensiones, jerarquías y hechos. Una buena práctica es ir más allá de los requisitos del backlog, y agregar elementos que nos puedan llegar a servir en futuras iteraciones. Pero, con sólo responder al backlog, esta etapa está terminada.

Paso 5: Generar el Diseño Lógico del DM. Una vez obtenido el diseño conceptual, a continuación, se desarrolla el diseño lógico del DM. Para ello, se utilizan las reglas de transformación para mapear los objetos multidimensionales definidos en el diseño conceptual a tablas relacionales. El resultado es un diseño de esquema de base de datos, en estrella o copo de nieve, listo para ser implementado en algún SGBD.

Paso 6: Generar el Diseño Físico del DM. Una vez desarrollado el diseño lógico, lo que sigue es implementar el diseño obtenido en el SGBD elegido. Se procede a la creación de la base de datos, tablas con sus tipos de datos, relaciones y restricciones de integridad correspondientes. El resultado es una BD disponible para cargar los datos del proceso de Extracción, Transformación y Carga (ETL, por sus siglas en inglés Extract, Transform and Load).

Paso 7: Diseñar los procesos ETL. En esta fase, se diseña el proceso ETL de manera conceptual. Esta etapa es importante para poder identificar cuáles son los tratamientos (transformaciones) que se le deben hacer a los datos origen para que lleguen "limpios" al DM y plasmarlos en un diagrama. Para una mayor claridad y disminuir la probabilidad de cometer errores, se recomienda realizar un diagrama por cada tabla de dimensión y tabla de hechos.

Paso 8: Implementar los procesos ETL. Luego de obtener los diseños del proceso ETL, se procede a su implementación en una herramienta ETL. Al igual que en el paso anterior, se realiza una implementación por cada diseño obtenido. Una vez desarrolladas las implementaciones, se planifica su ejecución para cargar los datos en el DM. El resultado de esta etapa es obtener los datos cargados y disponibles para su utilización en el DM.

Paso 9: Diseñar los componentes de navegación OLAP. A continuación, se procede a desarrollar los componentes de navegación OLAP para responder a cada una de las historias de usuario definidas en el backlog. De acuerdo a la herramienta seleccionada, se podría definir una configuración OLAP por cada historia, con sus respectivos gráficos, informes, observaciones, recomendaciones, etc. El resultado de esta etapa consiste en obtener los elementos OLAP, en respuesta a las historias de usuario, disponibles para ser entregados al usuario. Es muy importante, en el ambiente de desarrollo, la participación del usuario, para poder validar y ajustar los componentes que se están desarrollando.

Paso 10: Entregar al usuario los componentes OLAP. Una vez obtenido los componentes OLAP en el paso anterior, sólo resta poner a disposición del usuario los componentes creados. Dependiendo de las necesidades y de la infraestructura con la que cuenta el usuario, se realiza el despliegue (paso a producción) de dichos componentes. Además, se hace entrega de toda la documentación generada (gráficos, informes, observaciones, recomendaciones, etc.) al usuario. Con este paso, se cierra un ciclo de desarrollo y entrega, para luego poder realizar otras iteraciones para dar respuesta a otros objetivos.

4 Resultados y conclusiones

El modelo fue aplicado en una micro empresa sanjuanina, que se dedica a la venta de productos para el agro. Esta micro empresa brinda a sus clientes un servicio integral, desde el asesoramiento técnico, hasta la venta y entrega de los productos en finca, logrando cubrir las más exigentes necesidades del agricultor, y teniendo especial cuidado en la preservación del medio ambiente. Cuenta con cuatro empleados dedicados a tareas de asesoramiento, venta, depósito y entrega de productos. El propietario es quien conduce y toma las decisiones del negocio.

Se utilizó Lenguaje de Modelado Unificado (UML, por sus siglas en inglés Unified Modeling Language) [11], para realizar el diseño conceptual del DM basado en el modelo MutiDim [3]. Se utilizaron las reglas de transformación al modelo lógico relacional [3] que permite convertir los objetos multidimensionales del diseño conceptual a tablas relacionales. La Plataforma Analítica de Negocios Pentaho (Pentaho Business Analytics Platform) [12] que por un lado provee las funcionalidades básicas necesarias (ETL y OLAP), y además posee una edición comunitaria que es de código abierto (open source). Para los procesos ETL se realizó el diseño con BPMN y la implementación con la herramienta Integración de Datos Pentaho (PDI, por sus siglas en inglés Pentaho Data Integration) [13]. Como SGBD se utilizó MariaDB [14] para alojar los datos y hacerlos disponibles para su explotación. Para ello, se utilizó JPivot [15] que permite navegar cubos OLAP y mostrar esa información en forma de tablas multidimensionales y gráficos. Estas herramientas permitieron obtener resultados adecuados y que dieron respuesta a los requerimientos planteados como historias de usuario.

En este trabajo se ha concretado la generación y explotación de un DM de gran utilidad para una microempresa sanjuanina que le ha posibilitado, por medio de los

resultados y análisis realizados, mantener y mejorar su competitividad en el medio empresarial regional.

El éxito alcanzado residió en seguir un proceso simple y ordenado, alineado a las metas del negocio establecidas inicialmente por el gerente de la empresa. Dicho proceso fue abstraído y esquematizado en el modelo presentado en la Fig. 1. Este MODELO SIMPLIFICADO PARA GENERACION DE DATA MARTS constituye un marco general de desarrollo a seguir en el proceso de construcción de cualquier DM pequeño en empresas de recursos limitados.

En base al proceso de generación del DM, a los resultados y los análisis realizados, se puede concluir que el gerente del caso de estudio presentado, puede hacer cambios, tomar decisiones, prever stock según la estacionalidad, anticiparse a las necesidades de los clientes, seguir cuidando y manteniendo la fidelidad de sus principales compradores, ofrecerles otros servicios vinculados tanto a los productos que suele comprar como a la forma de hacérselos llegar (como un área de servicio al cliente o call-center), donde se ofrezca antes que el cliente lo pida, o algún servicio vía web o app de celular.

5 Referencias

1. Trujillo, J. C., Mazón, J. N., & Pardillo, J.: Diseño y Explotacion de Datos: Conceptos Basicos de Modelado Multidimensional.Editorial Club Universitario, Alicante (2009)
2. El nivel de madurez de las organizaciones en el análisis de datos. https://www.lainnovacionnecesaria.com/el-nivel-de-madurez-de-las-organizaciones-en-el-analisis-de-datos/
3. Vaisman, A., & Zimányi, E.:Sistemas Data Warehouse. Diseño e Implementacion. Springer, New York (2014)
4. Business Intelligence la clave para que las pymes decidan con inteligencia.https://www.cronista.com/pyme/negocios/Business-Intelligence-la-clave-para-que-las-pymes-decidan-con-inteligencia-20181205-0005.html
5. La inteligencia de negocio, clave para el desarrollo de las PYMES. https://www.marketingdirecto.com/digital-general/digital/la-inteligencia-negocio-clave-desarrollo-las-pymes
6. Inmon, W.: ConstruyendoData Warehouse, Fourth Edition.Wiley Publishing, Inc, Indianapolis, Indiana.(2005).
7. Kimball, R., Ross, M.: El kit de herramientas del data warehouse: la guia completa para el modelado dimensional - 2nd ed. Wiley Publishing, Inc, EEUU.(2002)
8. Modelo y Notación de Procesos de Negocios. http://www.bpmn.org/
9. Cravero, A., Sepúlveda, S.: Multidimensional design paradigms for data warehouses: a systematic mapping study. Journal of Software Engineering and Applications, 56-61 (2014).
10. Historias de Usuario. https://www.scrumalliance.org/learn-about-scrum/agile-atlas/agile-atlas-common-practices/planning/october-2014/user-stories
11. Lenguaje de Modelado Unificado. https://www.uml.org/

12. Plataforma Analítica de Negocios Pentaho. https://www.hitachivantara.com/en-us/pdf/datasheet/pentaho-business-analytics-platform-datasheet.pdf
13. Integración de Datos Pentaho. https://www.hitachivantara.com/en-us/products/big-data-integration-analytics/pentaho-data-integration.html
14. Acerca de MariaDB. https://mariadb.org/about/
15. Tutorial JPivot.http://ucumari.unc.edu.ar/repositorio/archivos/Tutorial-Jpivot.pdf

,

"Diseño y Modelado de un Sistema Web para el Registro Único de Acciones de Vinculación Académica y Científica de Unidades de Docencia e Investigación de la U.N.S.J. con sus pares del extranjero, en el marco de la Internacionalización de la Educación Superior (RUAVIN)."

Isabel Castro[1], María Inés Lund[2]

[1] Graduado Lic. en Sistemas de Información, FCEFN-UNSJ. Secretaría de Relaciones Internacionales y Posgrado de la UNSJ.
[2] Departamento e Instituto de Informática, FCEFN-UNSJ
isacastro.489@gmail.com;mlund@iinfo.unsj.edu.ar

Resumen. Este proyecto busca dar solución a una de las grandes problemáticas que enfrenta la Secretaría de Posgrado y Relaciones Internacionales - SEPRI - de la Universidad Nacional de San Juan – UNSJ. La SEPRI no cuenta con un registro sistematizado que les brinde información total en cuanto a las actividades internacionales que en la UNSJ se realizan.
A tal fin, se ha elaborado el diseño de un sistema web que da solución a las necesidades planteadas. En este trabajo se presenta el proceso de análisis, modelado y diseño del sistema.
Este sistema web Registro Único de Acciones de Vinculación Internacional - RUAVIN - será de suma importancia en el ámbito de la UNSJ, para dar solución a una necesidad urgente, y también como procedente, ya que no existe en ninguna institución universitaria del país uno semejante.

Palabras Claves: Sistema Web, Internacionalización, UNSJ, Secretaria de Posgrado y Relaciones Internacionales, Movilidad Internacional.

1 Introducción

Sistematizar la información en una organización, de cualquier índole, es un aspecto sumamente importante hoy en día. Cuando se sistematiza, se entiende aún más lo que se está haciendo y sobre lo que se está trabajando o investigando; la sistematización produce conocimiento.

Este proyecto busca dar solución a una de las problemáticas que desde hace tiempo enfrenta la SEPRI de la UNSJ.

La Secretaría tiene como misión promover y expandir las vinculaciones internacionales de la Universidad, con el objeto de fortalecer el desarrollo de sus funciones sustantivas, a través del desarrollo e implementación de políticas y programas que vinculen a la UNSJ con sus pares de otros países, centros de investigación y organismos internacionales, por medio de actividades de docencia, investigación y extensión universitaria.

Esto se logra a través de diversos convenios y redes de cooperación con instituciones de todo el mundo, para realizar actividades conjuntas y en su asociación para el beneficio mutuo, tales como la movilidad docente, estudiantil y de investigadores, en búsqueda de una mayor calidad académica de las carreras de grado y de posgrado, investigación y extensión.

La SEPRI no cuenta con un sistema que les permita registrar y consultar en tiempo real y de forma rápida, fácil, segura y certera todas las actividades de internacionalización, llevadas a cabo en su oficina, como las producidas en cada Unidad Académica- UA- de la UNSJ.

Cuando hablamos de actividades de internacionalización hacemos referencia a: movilidades de docentes, no docentes, investigadores y alumnos, al exterior, de la misma forma, aquel extranjero académico que llega a la UNSJ a realizar una tarea o actividad. Y la producción científica internacional tanto personal como las conjuntas surgidas de dichas movilidades.

Todas estas actividades de movilidad se realizan bajo convenios, programas y proyectos de movilidad existentes. Es decir que el personal académico puede realizar una movilidad dentro del marco formal como los mencionados anteriormente, como también pueden realizar movilidad por medio de invitaciones personales de otras instituciones, inscripciones a convocatorias internacionales, asistencia y participación a eventos académicos y científicos, entre otros.

1.1 Justificación

Actualmente recolectar estos datos implica, por parte la SEPRI, un gran esfuerzo. Ya que la misma no tiene conocimiento de todas las actividades de internacionalización producidas en otras UA de la UNSJ, o bien la recolección de las mismas no es completa. Esto se ha convertido en un problema significativo debido a que debe manejar información que viene en diferentes formatos y sin ningún estándar. La información que debe recolectar la SEPRI es crucial en el desarrollo de sus funciones, y depender de información que no es completa, o que se obtiene por medios no formales y, además, no se encuentra sistematizada, es un factor que impide realizar una gestión óptima.

El diseño y modelado de dicho sistema web, dará solución a esta necesidad, ya que una vez implementado se podrá obtener información útil y estratégica para la toma de decisiones, no solamente en el ámbito de la SEPRI sino en todas las UA de la UNSJ.

La UNSJ podrá conocer efectivamente las actividades internacionales que su personal realiza, como así también de los académicos extranjeros que llegan. Se logrará una total transparencia y exhaustividad de la información de internacionalización desarrollada en la UNSJ.

Mantener un sistema único descentralizado que permita administrar esta información en tiempo real, es una herramienta crucial para el desarrollo efectivo de las funciones de la SEPRI, y de todas las UA de la UNSJ.

1.2 Objetivo General

Brindar, a la Secretaría de Posgrado y Relaciones Internacionales de la UNSJ, la solución a una de sus problemáticas más urgentes e importantes, para su función sustantiva, el Diseño y Modelado de un Sistema Web (RUAVIN) para el Registro Único de Acciones de Vinculación Académica y Científica de Unidades de Docencia e Investigación de la UNSJ con sus pares del extranjero, en el marco de la Internacionalización de la Educación Superior.

1.3 Objetivos Específicos

- Relevar los requerimientos de la SEPRI para desarrollar una gestión efectiva y óptima de sus funciones.
- Identificar y especificar requisitos del Sistema.
- Diseñar y Modelar el Sistema Web RUAVIN, brindando una solución a la problemática identificada, y que sea revisada y aprobada por las autoridades de la SEPRI.
- Diseñar y elaborar el prototipo del Sistema, por medio de interfaces.

2 Marco Teórico

2.1 Metodologías de desarrollo

Desarrollar un buen software depende de un sin número de actividades y etapas, donde el impacto de elegir la mejor metodología para un equipo, en un determinado proyecto es trascendental para el éxito del producto. El papel preponderante de las metodologías es sin duda esencial en un proyecto y en el paso inicial, que debe ajustarse al equipo, guiar y organizar actividades que conlleven a las metas trazadas en el grupo.

Existen dos grandes enfoques, las metodologías tradicionales y metodologías ágiles, las primeras están pensadas para el uso exhaustivo de documentación durante todo el ciclo del proyecto mientras que las segundas ponen vital importancia en la capacidad de respuesta a los cambios, la confianza en las habilidades del equipo y al mantener una buena relación con el cliente. La elección de cuál metodología implementar será la que mejor se adapte al equipo de desarrollo[1].Entre las principales metodologías tradicionales tenemos los ya conocidos Proceso Unificado (UP) y Microsoft Solutions Framework (MSF) [2] entre otros. UP es uno de los métodos más usados dentro de los métodos tradicionales, y el cual se usará para el desarrollo de este trabajo.

2.2 Lenguaje Unificado de Modelado – UML

Para documentar el sistema web RUAVIN se ha utilizado UML, el cual es un lenguaje de modelado de sistemas de software más conocido y utilizado en la actualidad; está respaldado por el Object Management Group (OMG) [3]. Es un lenguaje gráfico que permite visualizar, especificar, construir y documentar un sistema. UML ofrece un estándar para describir un "plano" del sistema (modelo), incluyendo aspectos conceptuales tales como procesos, funciones del sistema, y aspectos concretos como expresiones de lenguajes de programación, esquemas de bases de datos y compuestos reciclados.

2.3. Generación de Prototipos RUAVIN

Para la realización de prototipos del sistema RUAVIN se utilizó una herramienta de software del estilo de AXURE pro 7.0 [4].

Axure es una herramienta estándar de la industria de software, orientada a diseñar wireframes y generar prototipos básicos o avanzados de forma fácil. Esta herramienta está dirigida tanto a la creación de aplicaciones web como de escritorio. Es muy fácil comenzar a trabajar con la misma, la herramienta es muy intuitiva y podemos empezar a producir casi inmediatamente sin necesidad de seguir un instructivo.

El programa tiene por defecto una amplia gama de widgets con los cuales se pueden crear interfaces de una forma rápida y sencilla, convirtiéndose en una herramienta muy útil para el prototipado. Una de sus características más interesantes es que genera prototipos HTML de sitios web, sin necesidad de ningún tipo de codificación, así como documentos en Microsoft Word para los procesos de documentación. La posibilidad de exportar el prototipo en HTML resulta ideal para documentar el funcionamiento y para realizar pruebas de usuario.

3 Proceso de Desarrollo de Software

3.1 Elicitación de Requisitos

Para comenzar el desarrollo se debió relevar, con distintos actores, las funciones de la SEPRI y todo lo que necesitaban para una buena gestión de la misma. La elicitación de requisitos es la primera fase en el desarrollo de software. Para ello se realizaron varias entrevistas al personal de la SEPRI, a la Secretaria General y al Coordinador de Relaciones Internacionales de la SEPRI.

Se realizó el Análisis de requisitos con la información obtenida y se pudo determinar las características que debería contemplar dicho sistema. Posteriormente se identificaron y definieron los requisitos funcionales y no funcionales del software a desarrollar, aclarando que un requisito funcional define una función que el sistema deberá realizar o alguna transformación que el sistema realiza sobre las entradas para producir salidas, a diferencia de los requisitos no funcionales, que tienen que ver con características que pueden limitar el diseño del sistema, como por ejemplo, el rendimiento, interfaz de usuario, fiabilidad, seguridad, etc. [5].

3.2. Diagrama de paquetes de Casos de Uso

Los diagramas de casos de uso se organizaron en paquetes y estos en un diagrama de paquetes de casos de uso, donde se especifican paquetes que contienen un grupo de casos de uso relacionados y paquetes que agrupan los actores del sistema. De esta manera se puede observar cómo estará estructurado el sistema.

En la Figura 1 se muestra el diagrama de paquetes del sistema RUAVIN, y en la Figura 2 se expone el detalle del paquete informes, que contiene a su vez sub-paquetes, detallando la variedad de informes que ofrece el sistema.

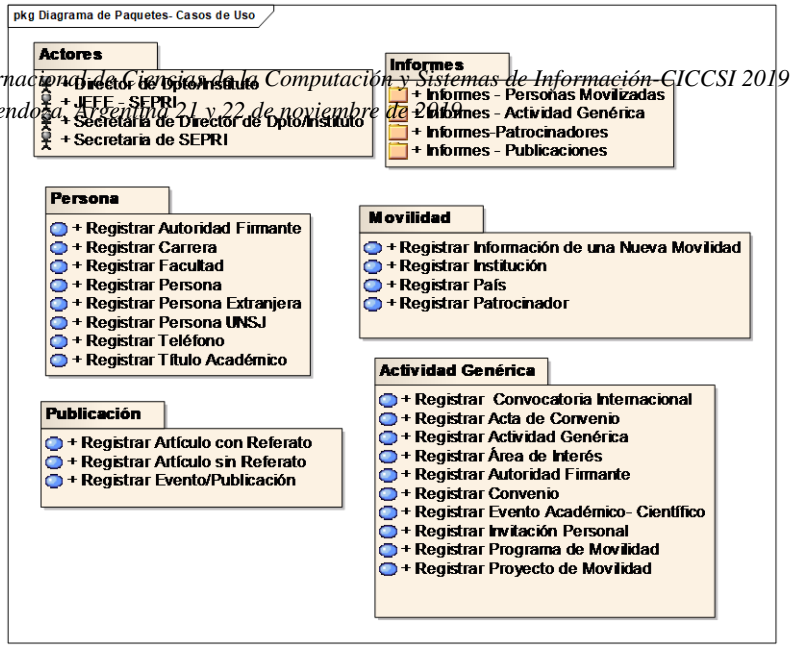


Fig. 1: Diagrama de Paquetes de Casos de Uso

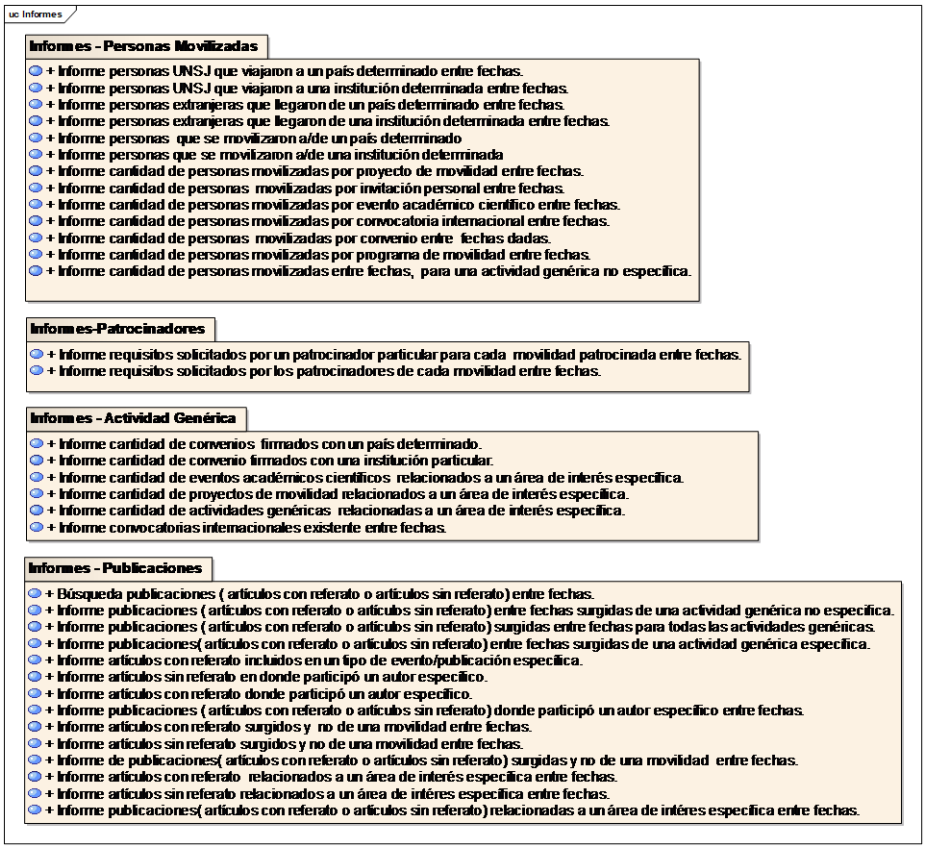


Fig. 2: Diagrama de Paquete Informes

3.2. Interface de Usuario – Prototipo

Menú Principal

Es la primera página que se visualiza luego de ser validado el usuario. A través de ella se observa el menú principal del sistema, el cual contiene distintas opciones a elegir, según la operación o tarea que se desee efectuar. El menú será acorde al nivel de permiso de acceso que tenga cada usuario y sus opciones contienen seis categorías diferentes:

- Movilidades
- Personas
- Publicaciones
- Actividades Genéricas
- Informes
- Otros

Los usuarios son dos: Usuario Director y Usuario Secretaria. Cuando el nivel de acceso sea **"Director"** estarán visibles las seis opciones del menú principal (categorías), en cambio cuando sea **"Secretaria"** tendrá acceso a todas las opciones excepto la categoría Informes. A continuación, en la Figura 3 se puede observar la pantalla que contiene el menú principal que visualizará el usuario "Director".

Fig. 3: Menú Principal - Usuario Director

4 Resultados

La ejecución del presente trabajo permitió obtener el modelo y diseño de una aplicación web, además de un prototipo dinámico, que permite registrar y consultar las movilidades realizadas tanto por el personal de la UNSJ (Docentes, No Docentes, Investigadores y Alumnos) al exterior, como de sus pares académicos del extranjero que llegan a la UNSJ. Esta aplicación también permitirá el registro y consulta de información en cuanto a Publicaciones Científicas realizadas, ya sean producto o no de una movilidad, de co-autoría compartida, dando así una solución a la problemática expresada por la SEPRI.

El uso de la herramienta AXURE para elaboración dinámica de prototipos permitió construir de manera sencilla y eficiente el prototipo de la aplicación web. La aplicación web propuesta es un producto que contribuye al desarrollo tecnológico de la Secretaria de Posgrado y Relaciones Internacionales y de todo el ámbito de la UNSJ. Es un aporte de gran beneficio ya que con los resultados e información que brinda este sistema se podrán tomar decisiones estratégicas y dar respuestas a consultas en tiempo real, mejorando así el desarrollo de sus funciones.

La aplicación RUAVIN ha sido aprobada por la Secretaria de Posgrado y Relaciones Internacionales y el Rector de la UNSJ para su implementación, habiendo quedado satisfechos y complacidos con la propuesta de su modelado y diseño. Por lo tanto, restaría concluir la fase de elaboración, construcción y transición propuesta por la metodología de desarrollo de software, para la implementación de la aplicación.

5 Conclusiones

Se pretende que la propuesta planteada evolucione, para su implementación y despliegue real en la SEPRI y en todo el ámbito de la UNSJ, adaptándose en forma progresiva y constante al quehacer diario de la misma, satisfaciendo así las necesidades de crear un canal de comunicación más dinámico entre alumnos, docentes, no docentes e investigadores y la SEPRI, como con cada UA relacionada al ámbito de las relaciones internacionales de la UNSJ. Esto permitirá mejorar el accionar de las funciones de la SEPRI y ofrecer información estadística relevante para el posicionamiento de la internacionalización de la educación superior en la UNSJ como para la toma de decisiones, entre otras.

Mantener un sistema único descentralizado que permita administrar esta información en tiempo real, es una herramienta crucial para el desarrollo efectivo de las funciones de la SEPRI, y de todas las UA de la UNSJ, ya que tanto la SEPRI como las UA podrán conocer, efectivamente, las actividades internacionales que se realizan, tanto de movilidades como de producción académico-científico. Se logrará una total transparencia y completitud en la información de internacionalización desarrollada en la UNSJ. El acceso al sistema web RUAVIN será a través de la página web de la UNSJ por cada UA y la SEPRI. Dicho acceso se ve representado en la Figura 4.

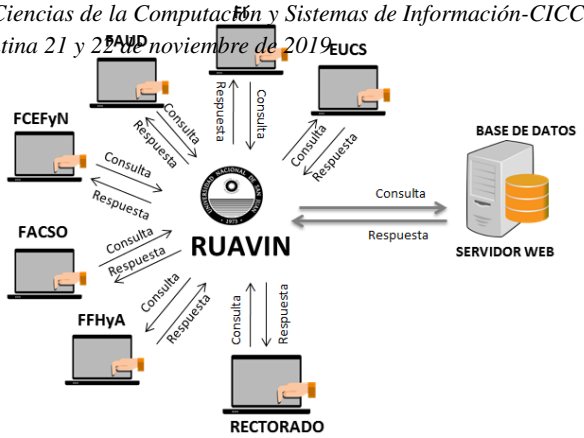


Fig. 4: Representación del Acceso al Sistema Web RUAVIN

El sistema web RUAVIN será de suma importancia en el ámbito de la UNSJ, para dar solución a una necesidad urgente. Es también una aplicación procedente, ya que no se ha encontrado alguna solución semejante en las instituciones universitarias del país.

El diseño, modelado y prototipado de este sistema web colaborará fuertemente a la misión de la SEPRI, que es promover y expandir las vinculaciones internacionales de la universidad, con el objeto de fortalecer el desarrollo de sus funciones sustantivas, a través del desarrollo e implementación de políticas y programas que vinculen a la UNSJ con sus pares de otros países, centros de investigación y organismos internacionales, a través de la docencia, la investigación y la extensión universitaria.

Se pretende a futuro, continuar con el desarrollo y la implementación de éste sistema. Una vez que RUAVIN se encuentre en uso, se podrá dar cumplimiento al objetivo planteado, al inicio del trabajo final, de brindar, a la Secretaría de Posgrado y Relaciones Internacionales de la UNSJ, la solución a una de sus problemáticas más urgentes e importantes, para su función sustantiva: el "Diseño y Modelado de un Sistema Web para el Registro Único de Acciones de Vinculación Académica y Científica de Unidades de Docencia e Investigación de la UNSJ con sus pares del extranjero, en el marco de la Internacionalización de la Educación Superior".

Los resultados de este artículo fueron validados por el mismo personal de la Secretaría de Posgrado y Relaciones Internacionales de la UNSJ, a través del manejo de la herramienta de prototipado, e incorporando, en su momento, algunas sugerencias que ellos plantearon.

6 Posibles Ampliaciones

Como se sabe, el Sistema de Integración Universitaria- SIU- tiene como fin dotar a cada una de las Instituciones Universitarias que componen el sistema universitario nacional argentino de aplicaciones informáticas que colaboren a mejorar la gestión y la calidad de los datos que día a día se producen en su ámbito. Para ello se trabaja con las instituciones con el objetivo de lograr que los módulos que lo conforman se adapten a sus necesidades y sean capaces de responder a cada uno de sus requerimientos [6].

Actualmente existen diversos módulos desarrollados, que se utilizan por distintas universidades nacionales y cada uno de ellos responde a necesidades específicas de la institución, a continuación, se nombran algunos de ellos:

- **SIU-Guaraní**: Registra las actividades de la gestión académica dentro de la universidad desde que un alumno se inscribe hasta que egresa. El objetivo de SIU-Guaraní es la administración de las tareas académicas en forma óptima y segura, con la finalidad de obtener información consistente para los niveles operativos y directivos [6].
- **SIU-Pilagá**: Permite realizar en forma integrada la gestión de presupuesto, la ejecución del gasto y de recaudación. [6].
- **SIU-Mapuche**: Recoge toda la información de los RRHH de una institución en un Legajo Electrónico Único (sistema integrado), está diseñado para brindar al operador, ya sea al trabajador del área o al encargado de liquidaciones, todos los servicios necesarios para disminuir las posibilidades de error y hacer más sencillo su trabajo, teniendo en cuenta los cambios en la legislación laboral vigente. [6].

Considerando que todas las Universidades Nacionales del país tienen las mismas necesidades o parecidas características que la Secretaría de Posgrado y Relaciones Internacionales de la UNSJ, la solución planteada en la presente proyecto, podría ser de utilidad en todas ellas.

Además, analizando los diferentes módulos desarrollados por SIU, no existe uno que contemple satisfacer la problemática planteada y solucionada en este trabajo final, con lo cual podría ser un gran aporte que el Sistema Web RUAVIN forme parte de los sistemas SIU, y que su uso pueda ser compartido por todos, dando visibilidad a las actividades internacionales de cada Universidad, en el marco de la **Internacionalización de la Educación Superior** [7].

Referencias

1. C. J. S. Roberth G. Figueroa, A. A. Cabrera, and E. de C. en C. Universidad Técnica Particular de Loja, “METODOLOGÍAS TRADICIONALES VS. METODOLOGÍAS ÁGILES.” .
2. “Microsoft Solution Framework INGENIERÍA DE SOFTWARE.”
3. “Acerca de la especificación del lenguaje de modelado unificado versión 2.5.1.” [Online]. Available: https://www.omg.org/spec/UML/About-UML/. [Accessed: 14-Jan-2019].
4. “Herramientas de diseño » AXURE, Diseño de wireframes y Prototipos.” [Online]. Available: http://multimedia.uoc.edu/blogs/labeines/es/prototipat/axure-disseny-de-wireframes-i-prototips/. [Accessed: 22-Dec-2018].
5. Michael Arias chaves, *La ingeniería de requerimientos y su importancia en el desarrollo de proyectos de software*. .
6. “SIU – Sistema de Información Universitaria.” [Online]. Available: https://www.siu.edu.ar/. [Accessed: 22-Jun-2019].
7. J. Knight, “Internacionalización de la educación superior,” . [Online]. Available: http://www.ugto.mx/internacional/images/pdf/4a.pdf. [Accessed: 22-Jun-2019].

,,,

Integrando Afectividad en Test UX: Evaluación de una Aplicación mHealth

Laura N. Aballay[1], Mathias G. Díaz[2], María A. Murazzo[2], Silvana V. Aciar[1]

[1] Instituto de Informática. FCEFyN – UNSJ. San Juan, Argentina.
[2] Departamento de Informática. FCEFyN – UNSJ. San Juan, Argentina
lnaballay@gmail.com; silvanav.aciar@gmail.com;
mathiasgdiaz@gmail.com; marite@unsj-cuim.edu.ar

Abstract— El uso de aplicaciones puede suscitar emociones, agradables y desagradables en las personas. Los aspectos emocionales juegan un papel fundamental en la interacción del usuario con las aplicaciones, que además de estar en función del estado del ánimo y la contextualización del usuario, son dependientes de los niveles de usabilidad y accesibilidad que tienen las mismas. Generalmente estos tres aspectos no son tenidos en cuenta a la hora de desarrollar apps de cualquier tipo, sin embargo son de gran importancia. En el presente trabajo se diseñó un instrumento que permite la evaluación UX integrando usabilidad, accesibilidad y afectividad para aplicaciones móviles. Para ello se eligieron 23 personas para testear la aplicación y luego completar el cuestionario. Los participantes elegidos fueron de ambos sexos, en un rango de edades va desde los 10 años hasta más de 60, y de diferentes países: Argentina, Colombia y Brasil.

Keywords— Evaluación de Experiencia de Usuario, Accesibilidad, Cuestionario, Emociones, Usabilidad, IoT, mHealth.

1 Introducción

El diseño y la evaluación de la experiencia del usuario (UX) son la clave del éxito para los sistemas interactivos, y un elemento clave de esta interacción son las emociones. El objetivo de UX no se limita a mejorar el rendimiento del usuario en la interacción, sino que intenta analizar la utilidad del producto, el problema psicológico el placer y la satisfacción de su uso. Las emociones afectan la capacidad de atención y memorización, el rendimiento del usuario y su valoración del producto [1]. Tan importante como conocer las consecuencias de los estados emocionales del usuario durante la interacción, es conocer cuáles pueden ser sus causas, principalmente aquellas dependientes del diseño de la aplicación [2]. De acuerdo con el enfoque de diseño centrado en el usuario, un aspecto importante para la evaluación de diseño es cómo los usuarios experimentan la interacción con un sistema [3]. Existe una diversidad de normas y estándares que se proponen facilitar adecuados niveles de usabilidad y accesibilidad sobre diferentes contextos tales como: aplicaciones web, aplicaciones móviles y más recientemente se empezó a considerar la usabilidad y accesibilidad en

un contexto de Internet de las cosas (IoT), como por ejemplo en aplicaciones mHealth donde cada categoría posee requerimientos y funcionalidades muy diferentes, además el grupo de usuarios a los que va dirigido son muy específicos [4]. Una característica muy importante de estas aplicaciones, es que crean nuevas técnicas de interacción por lo que deben ser modeladas y prototipadas para que la experiencia que tenga el usuario al interactuar con ella sea adecuada [5].

En el presente trabajo se diseñó un instrumento que permite la evaluación UX integrando usabilidad, accesibilidad y afectividad para aplicaciones móviles. Se aplicó a una aplicación mHealth, más precisamente de mMonitoring y se exponen los resultados.

El artículo está organizado de la siguiente manera, en la sección 2 se exponen los conceptos teóricos que abarcan la temática en cuestión. En la sección 3, se detallan algunos trabajos relacionados. Posteriormente en la sección 4, se explica el caso de aplicación. Luego en la sección 5, se explica el cuestionario propuesto en este trabajo. En la sección 6, se hace el análisis de resultados. Por último, en sección 7 se dejan las conclusiones y trabajo a futuro.

2 Conceptos Teóricos

2.1 Usabilidad

En la norma ISO 25010 [6] (que sustituye o amplía a la ISO 9241 [7]) se detalla un modelo de calidad (de producto/sistema) compuesto por 8 características, donde una de ellas es usabilidad (la que interesa en el presente trabajo), y se define Usabilidad como: "Grado en que un producto o sistema puede ser utilizado por usuarios específicos para lograr objetivos específicos con eficacia, eficiencia y satisfacción en un contexto específico de uso" [6]; y sus sub-características, donde algunas de ellas toman su definición basándose en ISO/IEC 9421. Con base en estas normas, se destaca la importancia de la usabilidad, porque permite a los usuarios cumplir con los objetivos de uso en una aplicación específica.

2.2 Accesibilidad

La accesibilidad es aplicada a contenidos web, los cuales deben ser accesibles para los usuarios independientemente de sus capacidades físicas, sensoriales o tecnológicas [8]. Con el objetivo de mantener y promover la accesibilidad web, la organización internacional World Wide Web Consortium (W3C), ha impulsado la Web Accessibility Initiative (WAI) [9], encargada de proporcionar pautas y recursos que ayudan a hacer la web accesible. Un diseño accesible implica la necesidad de eliminar las barreras de acceso y diseñar para la diversidad y heterogeneidad de necesidades de acceso.

2.3 Experiencia de Usuario (UX)

La capacidad de medir el grado de usabilidad, accesibilidad y las emociones que despierta en el usuario el uso de la aplicación, lleva a analizar la experiencia de usuario [10]. La experiencia de usuario (UX) se define como "la percepción de una persona y las respuestas subjetivas de ésta como resultado de la utilización y/o el uso de un producto, sistema o servicio" [6]. En la Experiencia del Usuario (UX) toman protagonismo los aspectos subjetivos o hedónicos de la calidad de uso debido a la propia evolución del concepto de satisfacción.

2.4 Internet de las cosas (IOT)

Los avances en computación móvil, redes inalámbricas, dispositivos móviles y sistemas embebidos han dado lugar al paradigma de IoT ("Internet of things"-Internet de las cosas) el cual consiste en una infraestructura de red global y dinámica de nodos (things) interconectados, inteligentes y auto configurables. IoT permite no solo que las personas se interconecten y se comuniquen, sino que ahora también lo hacen los procesos, los datos y los objetos, convirtiéndose en sus cuatro pilares. Esta interrelación permite transformar la información en acciones que crean nuevas capacidades y experiencias. IoT se refiere a una amplia gama de objetos cotidianos que adquieren habilidades de conectividad y sensado.

Las app de IoT se orientan a tres categorías:

- Seguimiento y control del rendimiento de hogares y edificios,
- Automatización del transporte,
- Auto seguimiento y control del entorno personal

De estas tres categorías, en los últimos años se han popularizado las tecnologías "wearables" (usables o "ponibles") que facilitan el "Auto seguimiento y control del entorno personal", gracias a la aparición de una nueva categoría de dispositivos conectados, con capacidad de procesamiento y cuya característica principal es que se integran en el cuerpo del usuario. Dentro de esta categoría de dispositivos, actualmente se encuentran los relojes inteligentes (smartwatches), las pulseras conectadas (popularizadas anteriormente por marcas deportivas para medición de la actividad física), los lentes (smartglasses), los teléfonos inteligentes (smartphones con apps que permitan sensar al usuario) y, en sentido amplio, todos aquellos objetos con capacidad para integrar algún tipo de conexión como la ropa o el calzado para la actividad deportiva [11].

2.5 Aplicaciones mHealth

El "Global Observatory for Electronic Health" (www.who.int/goe/en/) define a mHealth como todas aquellas prácticas de salud soportadas en dispositivos móviles. Estas aplicaciones permiten realizar un control y seguimiento de los parámetros sanitarios del usuario, tales como pulso sanguíneo, niveles de glucosa, frecuencia cardiaca, saturación de oxígeno en sangre, temperatura corporal, etc. Este tipo de aplicaciones deben tener en cuenta el contexto de las personas a las que debe servir,

aprendiendo de dicho contexto y realizando acciones de forma proactiva de acuerdo con su situación y expectativas en cada momento [12]. Las aplicaciones de mHealth, se pueden categorizar según el objetivo que persigue el usuario al interactuar con ellas.

- **mAssistance:** Para ayudar a personas con discapacidad de algún tipo, soluciones para una gestión más adecuada dentro y fuera de casa.
- **mMonitoring:** Seguimiento en tiempo real de las funciones del cuerpo, hábitos de salud del usuario, seguimiento y prevención de su deterioro.
- **mHospital:** Para brindar mayor cobertura al paciente para una mayor eficiencia de los servicios de salud a largo plazo.
- **mVisiting:** Para citas remotas virtuales con profesionales de la salud.
- **mEmergencies:** Para asistencia de emergencia del especialista (accidentes, zonas remotas, desastres, etc.) para los profesionales de la salud.
- **mPrescription:** Para soportar procesos con recetas establecida a través de móvil.

De todas estas aplicaciones, en este trabajo se evaluó una de tipo mMonitoring, que se asocia a periféricos (por ejemplo, electrocardiógrafos, pulsómetros, básculas, etc.) para personas de diferentes edades para conocer el nivel de satisfacción de los usuarios no sólo en cuanto a lo funcional sino también en lo afectivo.

3 Trabajos Relacionados

Se llevó a cabo una breve revisión de trabajos para analizar el estado del arte respecto a la evaluación UX de aplicaciones mHealth. A continuación se detallan algunas investigaciones halladas.

3.1 Cuestionarios para Evaluación

Para medir la Usabilidad se utilizan cuestionarios para conocer el uso y grado de satisfacción de los usuarios de un sistema determinado. Se diseña un cuestionario con preguntas que deben responderse escogiendo una opción única entre un rango de respuestas preestablecidas. La principal ventaja de utilizar cuestionarios es que se consigue recoger respuestas concretas que aportan datos discretos comprobables mediante, por ejemplo, técnicas estadísticas. Dentro de los cuestionarios más relevantes en este ámbito se encuentra QUIS (Questionnaire for User Interface Satisfaction) [13]**.** Es una técnica de valoración subjetiva enfocada a medir el grado de satisfacción de los usuarios mientras interactúan con la interfaz. Estos cuestionarios evalúan sólo usabilidad. Otro ejemplo de cuestionario es SUMI (Software Usability Measurement Inventory) [14] que es un inventario de medidas de usabilidad, utilizado para valorar la calidad de uso de un sistema o de un prototipo. Este cuestionario está referenciado en estándares de calidad ISO como el ISO 9241-10. Es un cuestionario no gratuito, disponible en varios lenguajes (español entre ellos). WAMMI (Web site Analysis and MeasureMent Inventory) [15] es un cuestionario que surge como extensión de SUMI

al intentar orientarlo hacia la medición de la usabilidad en la Web. Está disponible en varios idiomas pero es un cuestionario de pago.

Para determinar cómo los usuarios experimentan afectivamente la interacción con un sistema, se pueden emplear varios cuestionarios. Por ejemplo, AttrakDiff [16] y el User Experience Questionnaire (UEQ) [17] miden las percepciones y la aceptación del producto. Por ejemplo AttrakDiff, evalúa un sistema usando un cuestionario con formato diferencial semántico, es decir, coloca características opuestas y el evaluador debe marcar a cuál de ellas está más cercana su valoración. Sin embargo este tipo de herramienta no sirve para evaluar las emociones del usuario. Mientras que SAM [18], LEM-Tool [19] y PrEmo [1] se utilizan para capturar las reacciones emocionales del usuario. Estos cuestionarios contienen diferentes escalas y estructuras para obtener una imagen holística e integral de la experiencia del usuario.

Para ofrecer una alternativa a los instrumentos existentes, el cuestionario propuesto en este trabajo aborda componentes clave de UX con afectividad juntos.

3.2 Evaluación de aplicaciones mHealth

En el artículo [20] se presenta un modelo para evaluar la usabilidad de aplicaciones de mHealth. Este método, denominado Health-ITUEM (Health IT Usability Evaluation Model) se aplicó en la evaluación de 10 aplicaciones y se usaron 400 evaluadores. Los evaluadores solo pueden ponderar cada característica con dos valores posibles, sólo evalúan usabilidad y no tienen en cuenta factores de accesibilidad ni afectividad.

En [21] se cuantifica la usabilidad de aplicaciones para el control de enfermedades crónicas (por ejemplo, la diabetes). El análisis de usabilidad se realizó usando el estándar ISO 9241-11 sobre parámetros de efectividad, eficiencia y satisfacción. Se usó una muestra de 10 usuarios con conocimiento en tecnologías móviles a los cuales se les distribuyó una encuesta con las tareas que debería realizar la aplicación y si cumplió o no con esas tareas. El mayor inconveniente es la poca representatividad de la muestra usada y la falta de evaluación de otros parámetros.

Por su parte, en [22] se realiza la evaluación de usabilidad en aplicaciones para usuarios con degradación sensorial donde la accesibilidad debe ser basada en “touch screen”. La evaluación se realizó mediante dos métodos: recorrido cognitivo y HIMMS (Health Care Information and Management Systems Society) y se usaron 5 testers que recibieron instrucción previa en el uso de la aplicación. Los aspectos a ser evaluados se centraron en: simplicidad, perdón de errores, feedback y aproximación a la realidad, mediante cuestionarios donde se mapeaban criterios de usabilidad a directrices HIMMS, que el usuario debía responder con valores del 1 al 4. Los resultados sirven para re diseñar las aplicaciones cuya evaluación de usabilidad no fue satisfactoria.

Uno de los aspectos que no han sido evaluados con rigor es la interactividad y cuan satisfecho queda el usuario con ella. En [23] se pone especial énfasis en la necesidad de gamificación y empatía que debe brindar la aplicación al usuario, para lograr esto, los autores ofrecen un conjunto de recomendaciones para diseñar centrados en las necesidades del bienestar emocional del usuario. Estas recomendaciones no son evaluadas pero son sugeridas como mejores prácticas en el diseño de aplicaciones de mHealth.

Debido a que la accesibilidad es un aspecto de suma importancia en [24] se evalúa la capacidad de las aplicaciones a adaptarse a las necesidades de usuarios cuya enfermedad atrofia alguno de sus sentidos, este problema hace que diferentes usuarios posean diferentes habilidades con el dispositivo y con la aplicación en función del problema que posean. Este trabajo sólo tiene en cuenta la accesibilidad.

Estas contribuciones relevadas, nos permiten afirmar que la evaluación UX es considerada como una buena opción para medir la calidad de los sistemas y la satisfacción de los usuarios al interactuar con ellos. Sin embargo se encontró poca evidencia de estudios que realicen evaluación UX el punto de vista afectivo. El trabajo relacionado respalda la idea de que la evaluación integral UX aportará beneficios al momento de medir la calidad de las aplicaciones y satisfacción de usuarios.

4 Caso de Aplicación

Como se mencionó anteriormente, en el presente trabajo se busca evaluar la UX de forma integral, analizando patrones de usabilidad, accesibilidad y afectividad en el uso de una aplicación móvil del área de mHealth. Este análisis permitirá detectar si existen falencias en la aplicación, de manera de poder determinar en qué medida ésta es usable y accesible para los usuarios de la misma, y además poder captar que emociones experimentan los usuarios durante su uso para poder valorar la satisfacción de los mismos.

La aplicación seleccionada como caso de estudio se denomina “Pulsómetro único” (Disponible en: https://play.google.com/store/apps/details?id=com.supersimpleapps.heart_rate_monitor_newui). Es adecuada para personas que están pendientes de su estado de salud, que necesitan controlar su frecuencia cardíaca con regularidad, deportistas que quieran medir la intensidad de su entrenamiento, o para uso cotidiano. Dicha app corre bajo el sistema operativo Android, y es del tipo mHealth. La función principal de este aplicativo es detectar el pulso de una persona en diferentes situaciones mediante el análisis de la información del flujo sanguíneo en la punta de su dedo haciendo uso de la cámara de fotos del teléfono celular. Se aclara que esta app no es un dispositivo médico y que los resultados obtenidos pueden variar con algún dispositivo específico para la funcionalidad mencionada. Además posee la función de ver el historial de pulsaciones registradas a medida que se utilice la misma (ver Figura 1), dicha información puede visualizarse de forma textual o mediante un gráfico de barras.

Al realizar la detección del pulso la persona debe colocar un dedo en la cámara del dispositivo móvil e indicar en qué situación se encuentra ya sea en “REPOSO”, “EJERCICIO”, “POSTERIOR AL EJERCICIO” o simplemente puede utilizar el estado “GENERAL”, en pocos segundos la aplicación le indicará las pulsaciones que posee en ese momento y guardará dicha información para el historial.

Fig. 1. Pantallas de la aplicación Android "Pulsómetro único"

5 Cuestionario Propuesto

Se diseñó un cuestionario, con el cual los usuarios pudieron evaluar una aplicación mHealth. La propuesta intenta que el usuario pueda fácil y rápidamente hacer un test integral de UX, sobre usabilidad y accesibilidad de cualquier aplicación, permitiendo además conocer qué emociones experimenta el usuario mientras la utilizó. Para este caso, se pidió a 28 personas elegidas al azar que participaran de la experiencia. A estas personas se les indicó que instalaran y usaran la aplicación "Pulsómetro único" y luego de utilizarla respondieran el cuestionario on-line por medio del enlace brindado de Google forms. Se obtuvieron 25 cuestionarios completos.

De esta manera, el cuestionario quedó diseñado con los siguientes ítems con los correspondientes enunciados que lo miden:

- Efectividad: correcto funcionamiento de la aplicación. ¿Pudo utilizar la aplicación correctamente desde la primera vez que la instaló?
- Eficiencia: tiempo de demora en que la aplicación arroja un resultado. ¿Tardó más de lo deseado en tomarse el pulso?
- Facilidad: dificultad para usar la aplicación. ¿Le resultó difícil usar la aplicación?
- Intención de uso: propósito de utilizar la aplicación en el futuro. ¿Seguiría utilizando esta aplicación en el futuro?
- Errores: equivocaciones al usar la aplicación y soporte en caso de cometerlos. ¿Cometió errores al utilizar la aplicación? Si cometió errores ¿la aplicación le ayudó a solucionarlos?

- Contenido: cantidad de texto expuesto en la interface de la aplicación. ¿La cantidad de texto en la pantalla le pareció adecuado?
- Portabilidad: posibilidad de instalar la misma aplicación en diferentes sistemas u otros dispositivos. ¿Pudo instalar la aplicación sin dificultad?
- Contexto de uso: lugar, momento, ubicación, posición permitida durante el uso de la aplicación. ¿Se pudo tomar el pulso estando de pie? ¿Se pudo tomar el pulso estando sentado? ¿Se pudo tomar el pulso caminando? ¿Se pudo tomar el pulso mientras corría? ¿Se pudo tomar el pulso aún en la oscuridad?
- Accesibilidad: la aplicación brinda opciones para personalizarla según las necesidades del usuario. ¿Pudo utilizar la aplicación mediante la indicación de gestos faciales o movimientos de la mano, sin tocar la pantalla? ¿La app le indicó en algún momento que posee indicación por voz para sus funciones? ¿Está conforme con el tamaño de los botones? ¿Le pareció adecuada la elección de colores de la app? ¿La aplicación le permite aumentar el tamaño de todos los controles para hacerlos más visibles? ¿Las respuestas de las acciones de la aplicación, poseen un alerta de tipo visual o auditiva? Si en la app hay botones que cambian de función en algún momento ¿La descripción o etiqueta de los mismos también cambia? Si en la app hay mensajes o notificaciones temporales (que desaparecen transcurrido un tiempo) ¿Alcanza el tiempo para que el usuario pueda leer o escuchar dicha notificación?

Cada enunciado posee una escala de respuestas “Si/No/No Sabe”, permitiendo elegir sólo una de ellas para valorar cada ítem. En el caso de elegir “Si” significa que si está conforme con el ítem planteado, en el caso de elegir “No” quiere decir que no está de acuerdo con el ítem evaluado, y las respuestas “No sabe” indican que el encuestado duda o desconoce la respuesta a ese ítem evaluado.

Además se añadió a cada ítem de la encuesta, consultas acerca de las emociones que percibió el usuario en cada situación. Dichas emociones pueden ser negativas o positivas y forman parte de la lista básica de emociones de la encuesta PrEmo [18], en la que se elicitan catorce emociones para determinar si el diseño de un producto es agradable para los usuarios, donde siete emociones son agradables/positivas (por ejemplo: deseo, agradable sorpresa, inspiración, diversión, admiración, satisfacción, fascinación), y siete son desagradables (es decir, indignación, desprecio, disgusto, sorpresa desagradable, insatisfacción, decepción y aburrimiento). El encuestado podía elegir una positiva o una negativa respecto a cada ítem evaluado.

Del total de participantes efectivos, 13 fueron mujeres y 12 hombres. El rango de edades de los participantes va desde los 10 años hasta más de 50, resultando que 3 participantes pertenecían al rango de 10 a 20 años, 6 de 21 a 30 años, 9 de 31 a 40 años, 5 de 41 a 50 años y 2 con más de 50 años. Esta distribución se puede observar en la figura 2 Otro rasgo de los participantes es que fueron de diferentes países: 22 de Argentina, 2 de Colombia y 1 de Brasil. De manera que la heterogeneidad de los participantes (representatividad de la muestra) respalda la validez de los resultados.

Figura 2. Rangos de edades de los participantes

6 Análisis de Resultados

Se recopilaron los datos de las encuestas, contabilizando las respuestas según la escala para cada ítem y calculando su porcentaje respecto del total.

Haciendo un resumen de las respuestas "Si" (conformidad con el ítem) y las emociones con mayoría de votos, se obtuvieron los siguientes resultados (Tabla 1):

Tabla 1. Resumen de Respuestas

	% Conformidad	**% Emociones**
Usabilidad		
Efectividad	96%	"Satisfacción" 88%
Eficiencia	70%	"Inspiración" 52%
Facilidad de uso	78%	"Fascinación" 84%
Utilizará en el futuro	52%	"Deseo" 48%
Cometió errores	25%	"Satisfacción" 48%
Si cometió errores la aplicación le ayudó a solucionarlos	52%	
Texto presentado por la aplicación adecuado	87%	"Agradable" 56%
Fácil de instalar	92%	"Satisfacción" 86%
Contexto de uso		
"estando sentado"	96%	"Satisfacción" 64%
"estando de pie"	88%	
Accesibilidad		
Tamaño de los botones	92%	"Satisfacción" 60%
Color de aplicación	72%	

Se consiguió medir la aplicación en cuanto a parámetros de usabilidad accesibilidad y afectividad de manera conjunta, obteniendo valores altos respecto a la conformidad, todos mayores al 50% (salvo al indicar que cometió errores durante el uso de la

aplicación) y en todos los ítems la mayoría de las emociones elegidas por los encuestados fueron positivas o favorables.

7 Conclusiones

La encuesta que se sugiere en el presente trabajo, permitió inferir sobre el estado de una aplicación en particular, tomada como caso de estudio.

Como conclusión a la propuesta planteada en el artículo, se puede destacar que la encuesta realizada en el mismo presenta una alternativa a otros cuestionarios existentes, ya que permite realizar un test de usabilidad, accesibilidad y afectividad de manera integrada. Generalmente estos tres aspectos no son tenidos en cuenta a la hora de evaluar aplicaciones de cualquier tipo y sería de gran importancia, sobre todo en el tipo de aplicación analizada mHealth. Esto permitirá conocer cuales son las sensaciones que tienen los usuarios al utilizar las funcionalidades implementadas en el aplicativo, y de esta forma poder corregir aquellos criterios de usabilidad y accesibilidad que manifiesten un efecto negativo en las emociones de las personas que realicen el test.

La evaluación es un aspecto fundamental en el desarrollo de sistemas y aplicaciones, ya que sin hacer ningún tipo de evaluación es imposible conocer si una aplicación cumple las expectativas de los usuarios y si se adapta a su contexto social. Las percepciones que puedan tener los usuarios que testeen un aplicativo modifica la idea de diseño original acercándolo más a las necesidades de la población para la que fue desarrollada, es por ello que es muy importante contar con una herramienta de evaluación integral de tales aplicaciones.

Resulta una forma sencilla, económica y directa para poder obtener una devolución de una muestra de usuarios, al utilizar una plataforma web para realizarla se disminuyeron las limitaciones geográficas para poder llegar así a personas de otras nacionalidades, tal que, como se mencionó anteriormente, participaron usuarios de Argentina, Colombia y Brasil.

El cuestionario implementado puede tomarse como base para generar uno de tipo genérica que pueda ser utilizado como test de UX, integrando usabilidad, accesibilidad y afectividad para cualquier tipo de aplicación, ya sean del tipo mHealth, o de cualquier otra área de función.

Como propuesta de mejora se sugiere agregar al cuestionario preguntas que refieran a la retroalimentación que la aplicación pueda poseer, es decir si la app realiza indicaciones o consejos para los usuarios en respuesta al accionar de las funciones de la misma. Por ejemplo, para el caso en estudio, una vez tomadas las pulsaciones al usuario la aplicación podría indicar de acuerdo a los valores obtenidos en ciertos contextos que es conveniente visitar a un médico; o en caso de que no haya podido tomarse el pulso correctamente, alentar a la personas que lo intente de nuevo alejando el fastidio que pueda generar ese problema. En general, por el tipo de aplicación de la que se trata (una aplicación de monitoreo personal de salud), es necesario que cuente con sistemas de alarmas para realizar un seguimiento sistematizado de los usuarios, este aspecto la aplicación evaluada no lo tiene en cuenta, dejando al usuario la tarea de recordar tomarse el pulso. Ya que la aplicación ofrece valores predefinidos, sería bueno

que cuando el pulso no sea el adecuado para la categoría seleccionada, muestre al usuario un mensaje con algún tipo de consejo, por ejemplo, que visite al médico o que haga un poco de reposo cuando los valores no son normales.

El cuestionario hace preguntas acerca de la edad, género y nacionalidad del encuestado, lo que permitirá realizar análisis de las respuestas según estos parámetros en el futuro.

Incluso se podría hacer un análisis de correlación entre los valores de usabilidad y accesibilidad de la aplicación respecto a las emociones, para determinar qué grado de influencia tienen estos aspectos en la parte emotiva del usuario.

8 Referencias

[1] P. M. A. Desmet, M. A. Blythe, A. F. Monk, K. Overbeeke, and P. C. Wright, "Measuring Emotions Development and application of an instrument to measure emotional responses to products."

[2] Coulton, "Empathy in the Internet of Things," in *27th International British Computer Society HCI Conference: The Internet of Things: Interacting with Digital Media in Connected Environments*, 2013.

[3] E. Mor, M. Garreta, and M. Galofré, "Diseño Centrado en el Usuario en Entornos Virtuales de Aprendizaje, de la Usabilidad a la Experiencia del Estudiante," 2004.

[4] Olla and Shimskey, "mHealth taxonomy: a literature survey of mobile health applications. Health and Technology, , . .," vol. 4(4), pp. 299–308, 2015.

[5] S. Queirós, R. Alvarelhão, and Teixeira, "Usability, accessibility and ambient-assisted living: a systematic literature review," *Univers. Access Inf. Soc.*, vol. 14(1), pp. 57–66, 2015.

[6] ISO/IEC, "ISO/IEC 25010:2011 - Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models," 2011. [Online]. Available: https://www.iso.org/standard/35733.html. [Accessed: 03-Aug-2017].

[7] ISO, "ISO 9241-210," 2010. [Online]. Available: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=52075.

[8] W3C, "Guía Breve de Accesibilidad Web." [Online]. Available: http://w3c.es/Divulgacion/GuiasBreves/Accesibilidad.

[9] WAI, "Web Accessibility Initiative." [Online]. Available: http://www.w3.org/WAI/.

[10] E. L.-C. Law and E. Lai-Chong, "The measurability and predictability of user experience," in *Proceedings of the 3rd ACM SIGCHI symposium on Engineering interactive computing systems - EICS '11*, 2011, p. 1.

[11] A. Miraz and P. Excell, "A review on Internet of Things (IoT), Internet of Everything (IoE) and Internet of Nano Things (IoNT)," *Internet Technol. Appl.*, no. IEEE, pp. 219–224, 2015.

[12] Natarajan, Prasath, and Kokila, "Smart Health Care System Using Internet of Things," *J. Netw. Commun. Emerg. Technol. www. jncet. Org*, vol. 6(3), 2016.

[13] B. D. Harper and K. L. Norman, "IMPROVING USER SATISFACTION: THE QUESTIONNAIRE FOR USER INTERACTION SATISFACTION VERSION 5.5."

[14] "Questionnaire online SUMI." .

[15] "Questionnaire WAMMI." [Online]. Available: http://sumi.uxp.ie/.

[16] V. Roto, M. Obrist, and K. Väänänen-Vainio-Mattila, "User Experience Evaluation Methods in Academic and Industrial Contexts."

[17] "User Experience Questionnaire Handbook."

[18] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *J. Behav. Ther. Exp. Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[19] G. Huisman, M. van Hout, E. van Dijk, T. van der Geest, and D. Heylen, "LEMtool," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, 2013, p. 351.

[20] S. Househ, J. Yunus, and Aldossari, "The Use of an Adapted Health IT Usability Evaluation Model (Health-ITUEM) for Evaluating Consumer Reported Ratings of Diabetes mHealth Applications: Implications for Diabetes Care and Management," *Acta Inform. Medica*, vol. 23(5), p. 290, 2015.

[21] Georgsson and Staggers, "Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics," *J. Am. Med. Informatics Assoc.*, vol. 23(1), pp. 5–11, 2016.

[22] Wildenbos, Peute, and Jaspers, "A framework for evaluating mHealth tools for older patients on usability," *Stud. Health Technol. Inform.*, pp. 210–783, 2015.

[23] Heffernan *et al.*, "Guidelines and Recommendations for Developing Interactive eHealth Apps for Complex Messaging in Health Promotion. and uHealth," *JMIR mHealth*, vol. 4(1), 2016.

[24] Daihua, Parmanto, Dicianno, and Pramana, "Accessibility of mHealth self-care apps for individuals with Spina Bifida," *Perspect. Heal. Inf. Manag.*, vol. 12(Spring), 2015.

Deserción Académica y Emociones en Ambientes E-learning: Percepciones de los estudiantes de un curso de posgrado

Edith Lovos[1], Laura N. Aballay[2]

[1] Centro de Estudios Interdisciplinario sobre Derechos Inclusión y Sociedad (CIEDIS - UNRN), Viedma, Río Negro, Argentina
elovos@unrn.edu.ar
[2] Instituto de Informática – FCEFyN - Universidad Nacional de San Juan, San Juan, Argentina
lnaballay@gmail.com

Resumen. Existen investigaciones que asocian la deserción escolar con el estado afectivo de los estudiantes. Es por ello que es importante analizar y comprender el estado anímico de los alumnos, tratando de detectar qué emociones se manifiestan durante el cursado. Las emociones pueden expresarse y detectarse a través de diversos medios, tales como el habla, expresiones faciales, gestos y datos textuales. En este artículo se presenta el resultado de una breve revisión de entornos de aprendizaje virtuales donde se implementan diferentes formas de detectar emociones de los estudiantes. Adicionalmente, se presentan resultados de una encuesta sobre percepción de emociones que influyen en la deserción y abandono de un curso en modalidad e-learning. La misma fue realizada a 20 estudiantes de un curso de posgrado vinculado a la temática. Las evidencias indican que la frustración es considerada la emoción que influye más negativamente en el aprendizaje y puede ser causante de abandono o deserción de los estudiantes.

Palabras claves e-learning; deserción; abandono; plataformas educativas; computación afectiva; emociones.

1 Introducción

Un tema que preocupa a docentes, profesores y tutores es la deserción escolar. Este es un problema a nivel global [1] y es notable el aumento en la tasa de fracaso y abandono de estudiantes en cualquier modalidad de educación. Un informe de la UNESCO citado en [2], señala que en el caso de la educación universitaria, la deserción representa un problema económico con fuerte impacto en las naciones y en el que incluyen factores de orden social, económico y pedagógico. En relación a las consecuencias que la misma tiene sobre los estudiantes, en [2] se presentan entre otros: la baja de los niveles de autoestima y confianza, trastornos en el comportamiento a nivel laboral y/o familiar y trastornos que se reflejan en el plano físico (insomnio, falta de

apetito, etc) y que se vinculan a los sentimientos de frustración y fracaso. En este sentido, D'Errico y otros [3], afirman que las emociones de un individuo, se construyen como estados internos multifacéticos, que abarcan sentimientos y aspectos cognitivos, fisiológicos, expresivos y motivacionales, que se activan cuando éste logra alcanzar su objetivo o no. Siguiendo esta línea, durante el proceso de enseñanza y aprendizaje, entran en juego diferentes variables ambientales que pueden influir de forma negativa en los estudiantes en dos aspectos, por una parte la motivación hacia el estudio y por otra el resultado académico alcanzado. Así, resulta importante analizar y comprender el estado anímico de los estudiantes, tratando de detectar qué emociones presenta éste durante el cursado [1] [4].

Las emociones pueden expresarse y detectarse a través de diversos medios, tales como el habla, expresiones faciales, gestos y datos textuales [5]. Sin embargo, en un entorno virtual de aprendizaje, donde no existe el contacto presencial con los estudiantes, que le permite al docente inferir sobre su estado anímico y actuar en consecuencia, resulta de importancia contar con recursos que permitan recuperar y analizar las emociones que allí se suceden, para ofrecer asistencia y de esta manera influir positivamente en el aprendizaje.

En este trabajo se presenta el resultado de una breve revisión de entornos de aprendizaje virtuales, donde se implementan diferentes mecanismos de detección de emociones de los estudiantes. Adicionalmente, se presentan resultados de una encuesta realizada a 20 estudiantes de un curso de posgrado relacionado a la temática, que permite conocer sus percepciones en relación a la influencia de las emociones en situaciones de abandono o deserción en ambientes e-learning, cuáles consideran que son los canales de entrada apropiados y menos invasivos para recolectar emociones en entornos educativos de nivel superior bajo la modalidad e-learning.

Este trabajo forma parte de las actividades de investigación que se realizan en el marco del proyecto PI-UNRN 40-C- 750 de la Universidad Nacional de Río Negro y de un trabajo de tesis de Doctorado de la Universidad Nacional de San Juan.

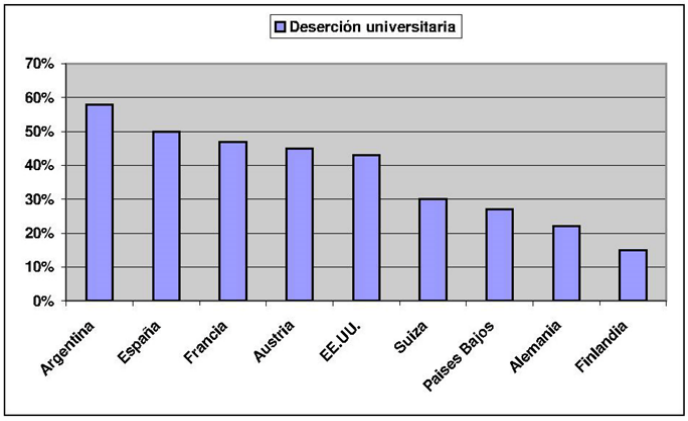


Fig. 1. Porcentajes de deserción escolar universitaria por países [2]

2 Modelos de Emociones

Existen tantos modelos y definiciones como escuelas y disciplinas psicológicas hay en el mundo. Sin embargo, las investigaciones en el campo de las ciencias informáticas [6] diferencian dos enfoques para la representación de las emociones: el modelo categórico y el modelo dimensional. El modelo categórico, asume que las emociones son discretas y define categorías emocionales, tal como Ekman [7] que se centra en un conjunto de seis emociones básicas que asocia a las expresiones faciales: ira, repugnancia, miedo, alegría, tristeza y sorpresa.

El segundo enfoque de identificación de las emociones es el dimensional, que representa los afectos en forma tridimensional. Los estados emocionales en este modelo están relacionados entre sí por un conjunto común de dimensiones y se define generalmente en un espacio de dos o tres dimensiones. Cada emoción ocupa un lugar en este espacio. En este enfoque, la variabilidad de los afectos es contemplada por tres dimensiones: valencia, entusiasmo y potencia (dominancia). La dimensión de valencia se refiere al grado de positividad o negatividad de la emoción, y va desde sensaciones desagradables a sentimientos de felicidad agradables. La dimensión de la excitación se refiere a cuán emocionada o apática es la emoción, y abarca desde la somnolencia o el aburrimiento a la excitación frenética. La dimensión de poder se refiere al grado de poder o de sensación de control sobre la emoción.

3 Emociones y Aprendizaje

Diferentes estudios indican que la motivación y la emoción inciden en la tasa de abandono que se observa en los sistemas de educación a distancia mediados por tecnologías [8], además de las características propias de los entornos tecnológicos. Arguedas Lafuente [11], en relación a tomar en cuenta las emociones en espacios de aprendizaje virtuales, señala que un cambio en el estado emocional de los estudiantes

permitiría reorientar su atención e inducir cambios en la forma en que piensan, actúan e interactúan con los demás, así como regular su comportamiento en una situación de aprendizaje. En este sentido, surge la necesidad de diseñar y desarrollar espacios de aprendizaje personalizados que tengan en cuenta las emociones de manera de influir positivamente en el aprendizaje [9] [10].

3.1 Entornos Virtuales de Aprendizaje Afectivos

Se realizó una recopilación de trabajos que incluyen la detección de emociones en entornos de aprendizaje por diversos métodos:

Un estudio [11] sobre conciencia (awareness) de emociones en ambientes educativos e-learning, propone el uso de análisis de discurso (texto y conversaciones) que tienen lugar en el marco de actividades de colaboración particularmente, WIKIs usando análisis de sentimientos o minería de opinión, para determinar cuáles son las emociones que se presentan y cuál es su influencia en el aprendizaje. La autora sostiene, que en estos sistemas es necesario poder etiquetar no solo las emociones académicas como positivas y negativas (alegría, orgullo, ansiedad, y vergüenza), sino también los comportamientos que surgen durante la interacción entre los estudiantes, como la solidaridad, sugerencia, opinión entre otros. Y propone el uso de un sistema de reglas Evento-Condición-Acción (ECA), basado en la herramienta WEKA. En un estudio posterior [11] presenta los resultados de una experiencia en un ambiente e-learning, que incluye la posibilidad de tomar conciencia de las emociones por parte de los estudiantes y el uso de un tutor pedagógico afectivo que pueda dar un feedback (a través de expresiones emocionales y consejos) teniendo en cuenta las emociones de los estudiantes. Como conclusiones de ese estudio, se destacan que cuando los estudiantes pueden tomar conciencia de sus emociones y de esta forma de su situación en el marco de un trabajo en equipo, esto les permite cambiar su actitud de manera que resulte beneficioso para el grupo.

Para la detección de emociones a partir del texto, existen diversas técnicas, entre las que se destaca la comparación con bases de datos léxicas, como son: WordNet y su derivada WordNet Affect, sin embargo como se señala en [12], una de las limitaciones de estos léxicos es el idioma, y agregan que hasta 2015 no se ha encontrado ninguna base de datos léxica en idioma español latinoamericano.

Otros estudios [13], proponen la detección de emociones a través del análisis de las expresiones faciales. La expresión facial de una persona, se determina a partir de la forma y la posición que ocupan los componentes faciales (cejas, ojos, boca, nariz, etc). Entre los algoritmos que permite realizar detección de rostros y extraer detalles de regiones faciales se encuentran: Viola Jones, Local Binary Pattern (LBP), Ada Boost and Neural Network. En [14] se presenta el diseño de un sistema denominado SERS (Sistema de reconocimiento de las emociones de los estudiantes), cuyo objetivo consiste en detectar el nivel de concentración del estudiante, a través del monitoreo constante de la rotación de la cabeza y el movimiento de los ojos. Sus autores señalan que para detectar la rotación de la cabeza es necesario determinar si ambos ojos están visibles o no. Es decir si no es posible visibilizar ambos ojos, eso implica que se ha producido una rotación de la cabeza, que permite inferir falta de interés por parte del

estudiante en lo que está visualizando. SERS permite detectar tres niveles de concentración a saber: bajo, medio y alto. Así mismo, SERS resulta eficiente para la detección de emociones negativas como el aburrimiento y la falta de interés. Sin embargo como se afirma en un estudio sobre computación afectiva y entornos de enseñanza y aprendizaje virtuales [15] el reconocimiento facial en escenarios naturales (hogares, escuela, etc) debe combinarse con otro medio no invasivo que permita confirmar la información recolectada. Ya que ésta queda sujeta a cuestiones vinculadas por ejemplo a la calidad de la cámara web y/o a movimientos del usuario durante la sesión, que dificultan la detección facial.

Otra forma de detección de emociones es a través del habla, en [16] se propone el desarrollo de un software usando el framework FILTWAM que busca a través de la detección, recolección y análisis de las emociones presentes en el habla, ofrecer comentarios en línea que resulten oportunos y permitan ayudar a mejorar la comunicación del estudiante en un entorno de educación mediado por TIC. La recolección de los datos, se realiza a través de una cámara web y un micrófono, cuando los estudiantes interactúan con materiales educativos del entorno de enseñanza y aprendizaje. Para la clasificación de las emociones, estos autores siguen el modelo propuesto por [17] que distingue seis emociones y agregan la emoción neutral. Para ello utilizan el clasificador de optimización mínima secuencial (SMO) del software WEKA.

Volviendo sobre la necesidad de contar con métodos combinados para reconocimiento de emociones, en [18] se presentan los resultados alcanzados con el uso de la Plataforma de Reconocimiento Multimodal de Emociones (PREMOC) [19], en una situación de aprendizaje gamificada de algoritmos y programación. PREMOC es un servicio web que puede usar imágenes faciales, audio, texto y opcionalmente señales electroencefalográficas (EEG), las cuales procesa para extraer sus características y alimentar una red neuronal que genera la emoción asociada utilizando un integrador de lógica difusa.

Estas contribuciones relevadas, nos permiten afirmar que las emociones se están teniendo en cuenta al momento de diseñar plataformas educativas.

4 Encuesta

Se presentan resultados de una encuesta realizada a estudiantes de un curso de posgrado relacionado a la temática, que permite conocer sus percepciones en relación a la influencia de las emociones en situaciones de abandono o deserción en ambientes e-learning, cuáles consideran que son los canales de entrada apropiados y menos invasivos para recolectar emociones en entornos educativos de nivel superior bajo la modalidad e-learning.

La encuesta, constó de 6 ítems y se implementó en forma online a través de un formulario de Google. La misma se destinó a un grupo de 20 estudiantes de un curso de posgrado sobre Computación Afectiva, desarrollado en la Universidad Nacional de La Plata (septiembre 2018). Los participantes, fueron docentes universitarios, de ambos sexos, de Argentina y Ecuador y de diferentes rangos etarios.

En la figura 2, se puede observar que la mayoría de los encuestados (85%) consideraron a la frustración como la emoción más influyente en los estudiantes al momento de decidir abandonar el curso. Y ésta se combina con la falta de acompañamiento. Esto coincide con estudios, que indican para el caso de propuestas formativas en modalidad e-learning, la importancia de generar sentimientos de comunidad que ayuden a contrarrestar el sentimiento de soledad [20].

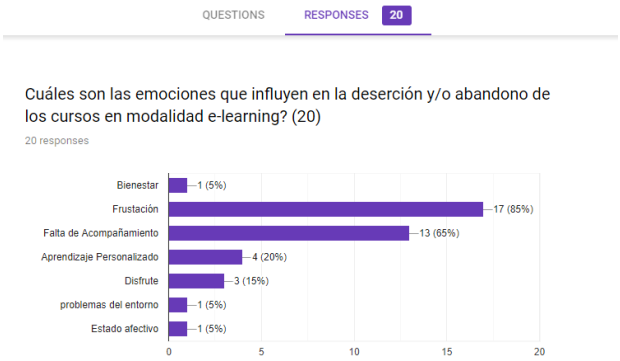


Fig. 2. Respuestas sobre emociones que influyen en abandono

De la misma manera, se consultó a los encuestados acerca de cuál consideraban el medio más adecuado para detectar la emociones en los estudiantes desde un entorno virtual de aprendizaje. Los resultados se pueden ver en la figura 3.

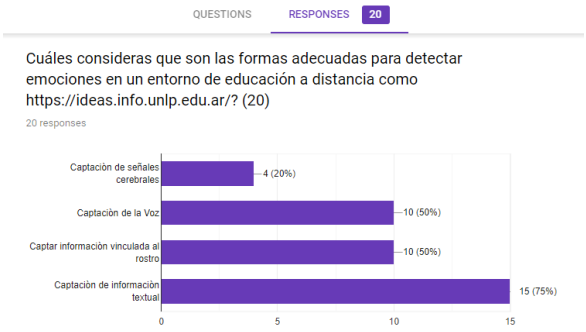


Fig. 3. Respuestas sobre modos de detección adecuados

Un 75% de los encuestados opinaron que la “captación de información textual” es el modo menos invasivo y que más se adecua en aprendizaje virtual a distancia. La “captación de la voz” y “la información vinculada al rostro” fue elegida por el 50% de

los participantes de la encuesta. Y un 20% opinó que la mejor manera de detectar emociones en estos ambientes es por medio de "captación de señales cerebrales".

5 Conclusiones

La deserción y abandono en espacios de formación, se presentan como un problema que preocupa a docentes y autoridades educativas de cualquier nivel. En este sentido, se vienen investigando y diseñando diversas estrategias que apuntan a la motivación y al acompañamiento de los estudiantes. En este trabajo se presenta la importancia de las emociones en los estudiantes, específicamente en el aprendizaje a distancia mediado por tecnología. Se hizo un recorrido por algunas plataformas educativas afectivas, analizando la solución empleada en cada una de ellas. Sobre este punto, las investigaciones resaltan la importancia de contar con métodos combinados para reconocimiento de emociones, que permitan dar certeza y resulten adecuados a los contextos educativos específicos y menos invasivos para los estudiantes.

Con la intención de conocer las percepciones de un grupo de estudiantes de posgrado en Computación Afectiva sobre influencia de las emociones en la deserción o abandono de actividades educativas en modalidad e-learning, se llevó adelante una encuesta online. Las evidencias muestran que la frustración es considerada la emoción que influye más negativamente en el aprendizaje y puede ser causante de abandono o deserción de los estudiantes.

La recolección de información sobre las emociones del estudiantado en el desarrollo de un curso en modalidad e-learning resulta entonces de utilidad y debiera ser tenida en cuenta por los profesores y los propios entornos de enseñanza y aprendizaje, para actuar en consecuencia, modificando estrategias y/o recursos de enseñanza de sus cursos, en pos de lograr una empatía emocional entre el curso y los estudiantes.

Se planea a futuro replicar esta encuesta a otras personas del ámbito educativo, para ampliar la cantidad de opiniones acerca de cuáles son las principales emociones que influyen en la deserción de los estudiantes en entornos de aprendizaje mediados por tecnología.

Referencias

[1] H. H. Binali, C. Wu, and V. Potdar, "A new significant area: Emotion detection in E-learning using opinion mining techniques," in 3rd IEEE International Conference on Digital Ecosystems and Technologies, 2009. DEST '09, 2009, pp. 259–264.

[2] "La Deserción Universitaria | Proyecto R.A.M.O.N." [Online]. Available: https://proyectoramon.wordpress.com/la-desercion-universitaria/. [Accessed: 14-Nov-2018].

[3] D'Errico, F., Paciello, M., De Carolis, B., Vattanid, A., Palestra, G., & Anzivino, G. (2018). Cognitive Emotions in E-Learning Processes and Their Potential Relationship with Students' Academic Adjustment. International Journal of Emotional Education, 10(1), 89-111.

[4] S. N. Shivhare, S. Garg, and A. Mishra, "EmotionFinder: Detecting emotion from blogs and textual documents," in 2015 International Conference on Computing, Communication Automation (ICCCA), 2015, pp. 52–57.

[5] R. Cowie, E. Douglas-Cowie, K. Karpouzis, G. Caridakis, M. Wallace, and S. Kollias, "Recognition of Emotional States in Natural Human-Computer Interaction," in Multimodal User Interfaces, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 119–153.

[6] G. Lescano, P. Santana Mansilla, and R. Costaguta, "Análisis de emociones en situaciones de aprendizaje colaborativo soportado por computadora : Estado del arte," Terc. Congr. Argentino la Interacción-Persona Comput. Telecomunicaciones, Informática e Inf. Científica (IPCTIIC 2014), no. December, pp. 1–16, 2014.

[7] P. Ekman, W. V. Friesen, and P. Ellsworth, Emotion in the human face: Guidelines for research and an integration of findings, vol. xii. Oxford, England: Pergamon Press, 1972.

[8] L. Rebollo Catalán, Ma. A., García Pérez, R., Barragán Sánchez, R., Buzón García, O. y Vega Caro, "Las emociones en el aprendizaje online," pp. 1–23, 2008.

[9] A. M. Hernández-Sánchez and J. A. Ortega, "Aprendizaje Electrónico Afectivo: un modelo Innovador para Desarrollar una Acción Aprendizaje Electrónico Afectivo: un modelo Innovador para Desarrollar una Acción Tutorial Virtual de Naturaleza Inclusiva Affective e-Learning: an Innovative Model for Developing a Mentoring Virtual Action in an Inclusive Context," Form. Univ., vol. 17, no. 2, pp. 19–26.

[10] A. E. Guzmán, A. L. Agüero, and E. N. Campazzo, "Las emociones en los mundos virtuales. Caso: EMO-SLEVA en la Universidad Nacional de la Rioja. En: Memorias XIV Congreso de Informática en la Educación."

[11] M. Arguedas Lafuente, "Emotions awareness on virtual learning students". Thesis submitted for the degree of Master in Multimedia Applications. July 2015. Open University of Catalonia

[12] L. Aballay, S. Aciar, and E. Reategui, "Propuesta de un Método para Detección de Emociones en E-Learning," in Argentine Symposium on Artificial Intelligence (ASAI 2015) - JAIIO 44 (Rosario, 2015), 2015, pp. 121–128.

[13] S. Duo and L. X. Song, "Physics Procedia Physics Procedia 00 (2011) 000-000 An E-learning System based on Affective Computing peer-review under responsibility of [name organizer]," Phys. Procedia, vol. 24, pp. 1893–1898, 2012.

[14] L.B Krithika and GG Lakshmi Priya, "Student Emotion Recognition System (SERS) for e-learning Improvement Based on Learner Concentration Metric," Procedia Comput. Sci., vol. 85, pp. 767–776, Jan. 2016.

[15] V. A. Bosquez Barcenes, C. Sanz, S. Baldassarri, E. Ribadeneira Ramos, G. V. Mendoza, R. Barragan Merino, W. Fierro "La Computación Afectiva: Emociones, Tecnologías y su Relación con la Educación Virtual". Revista de Investigación Talentos V. (1) Enero - Junio 2018.

[16] K. Bahreini, R. Nadolski, and W. Westera, "Towards real-time speech emotion recognition for affective e-learning," Educ. Inf. Technol., vol. 21, no. 5, pp. 1367–1386, Sep. 2016.

[17] P. Ekman and W. Friesen, "Facial Action Coding System," 1977.

[18] R. Z. Cabada, (2018). Reconocimiento afectivo y gamificación aplicados al aprendizaje de Lógica algorítmica y programación. REDIE: Revista Electrónica de Investigación Educativa, 20(3), 115-125.

[19] R. Zatarain, M. L. Barrón-Estrada y G. Muñoz, (2016). PREMOC: Plataforma de Reconocimiento Multimodal de Emociones. Trabajo presentado en el 8o. Congreso Mexicano de Inteligencia Artificial (COMIA), Tonanzintla, Puebla, México.

[20] Cabero Almenara, Julio, Bases pedagógicas del e-learning. RUSC. Universities and Knowledge Society Journal [en linea] 2006, 3 (abril): [Fecha de consulta: 15 de noviembre de 2018] Disponible en:http://www.redalyc.org/articulo.oa?id=78030102.

Modeling the Static View in Aspect-Oriented Software Development

Fernando Pinciroli

Research Institute, Champagnat University, Mendoza, Argentina
pincirolifernando@uch.edu.ar

Abstract. The amount of aspect-oriented software development techniques and tools have been increasing for the last years [1] [2], but still they have not enough maturity and are not sufficiently spread to convince a project leader to include them in his or her toolbox [3] [4]. There are many proposals on aspect-oriented techniques, notations, tools, etc., but they have not yet been unified on a common body of knowledge, and none of them has become the most popular approach. That's why we have developed AOP4ST [5], acronym that stands for Aspect-Oriented Process for a Smooth Transition, a framework process from the business phase of the software development life-cycle (SDLC) to the software development phase, trying to cover this gap. In this article we will explain how to manage the static view of AOP4ST, specially focusing on concern composition and conflict resolution.

Keywords: aspect-oriented software development, early aspects, static modeling, concern composition, conflict resolution.

1 Introduction

The static view is mainly composed by classes, which are detailed through their members and relationships among them. It is obtained by a realization from the use cases model [6].

By means of the static view we describe the internal view of the problem domain, since the use case model describes, at the same time, the external view, with no detailed mention of specific classes.

Aspect-oriented software development was mainly designed to successfully solve the scattering and tangling problem present in software build from traditional to object-oriented methodologies [7]. But the very nature of classes makes them crosscut to all concerns. Aspect-oriented techniques have their own way to drive this issue, but always making present the same class in different concerns. This solution triggers conflicts in the system composition, where the classes representing the same concept must be composed in only one class. In this article we will see how to realize requirement specifications made up with standard and auxiliary use cases (sections 2 and 3), how to compose the classes employed to compose the use cases (sections 4 and 5), how to solve the possible conflicts (section 6), to, finally, end with our conclusions and future work (section 7).

2 Realizing standard use cases with classes

We have been separating concerns through the previous models, from de business model to this one, the static view realizing software requirements. The approach offered by the use case model allow us to maintain perfectly separated our concerns until software requirements, because the very nature of use cases force it. But the use cases realization with a class model produces a many-to-many relationship among classes and use cases, returning to the scattering and tangling resolved on the previous models. That's why we need to include some novelties in order to have scattering and tangling controlled and out of our models. Ivar Jacobson says that aspect-orientation was like a "gift from Heaven" for use cases [8].

Jacobson give us the concept of "use case module" to keep separated and encapsulated the specific constituent elements for every use case. But there are non-specific parts of the use cases that need to contain the global system elements. Jacobson calls them "element structure" [4].

Use case realizations will exclusively contain the necessary members to realize that unique use case, what is called "class extension" by Ivar Jacobson [4]. Due to the many-to-many relationships among use cases and classes, several class extensions will have classes with the same relative names, presents in every use case slice, and those classes could have the same or different members depending on the needs of every single use case realization. We say "relative names" because the containing packages are namespaces in UML, so classes (and all elements) with the same names but located in different packages are considered completely different elements, independently if they are representing the same concept exiting in the real life. In aspect-oriented programming, the weaver weaves all the elements sharing the same name into only one with that name. Of course, there are a lot of unique classes that are present in only one realization.

All the elements used to realize a use case, from its specification and on, can be located into a package that it is called "use case module" [4]. Then, it is possible to separately manage every use case module, except the contact point that every system has because of its very nature. These contact points can be related with system architecture, the base where all the system functionalities are supported.

Dependencies among use cases modules are managed with UML package diagrams, since these modules, like Jacobson's use case slices, are UML stereotyped packages. We can find the same correspondence among UML packages and "themes" of Clarke and Baniassad's "The theme approach", since they have no conceptual differences, according with the authors of this method [9].

3 Realizing auxiliary use cases with classes

We have all the system functional requirements completely covered by use cases. These use cases are called "main use cases". But we also have non-functional requirements in the system that we can specify with use cases, called "infrastructure use cases". Finally, there are another kind called "auxiliary use cases", that corresponds to use cases derived from inclusion and generalization relationships,

employed to improve the model understandability. Main and infrastructure use cases are concrete ones, but auxiliary use cases are abstract ones.

There are two ways to manage the auxiliary use cases. The first one consists on collapsing them into the main use cases, having the content of the first ones included into the last ones. The second way is having them separately modeled, contained in specific packages. It this case, it will be necessary to model the relationships among the packages containing the main and the auxiliary use cases. We prefer the first alternative since it is easier to model and to maintain the models. However, it is also possible to employ the second approach, following the recommendations provided by Jacobson [4].

Besides, Jacobson proposes "non-specific use cases slices" to manage common elements belonging to the problem domain but not to a specific use case. These slices must be extended by specific use case slices. We don't consider necessary to employ this approach, with the only exception of modeling the packages containing auxiliary use cases.

4 Composing peer concerns in the static view

Once we have realized the whole use case model, we have a lot of separated concerns made of classes, some of them unique and other repeated along many packages. But we always have the need to be sure that the whole model will be able to be correctly composed, solving the potential conflicts that could occur. It is important to have the complete model, because it allows to review it, to validate it and, for instance, to apply automated tools to transform it into an entity-relationship model to generate the repository for some Relational Database Management System.

All these packages correspond to peer concerns, since all of them belong to the problem domain. The resulting model got by the composition of pair concerns will be a single model with unique classes with all the members, attributes and operations, existing in all the classes in all the packages.

The easiest composition is when we have unique classes in all the packages, but this situation never happens when we are composing a system instead of just separated functionalities.

There are different ways to compose classes. Clarke and Baniassad [9] give us two alternatives, but we will add some variations to them. These two alternatives are merging and overriding.

Composition by merging: consists on adding into a unique class all the members, relationships, constraints, etc., present in all the classes sharing the same relative name along the whole static model. We must always consider the following rule: when we compose two elements, the resulting element must have the characteristics of the more comprehensive one, since the resulting element cannot be less able than the original ones.

Composing class elements: when composing classes, or classifiers in general, the composition is made on elements sharing the same relative name, and we must consider:

- In case of a class is concrete and other is abstract, the resulting class must be concrete.
- All the common constraints must remain; the rest must be eliminated (for instance: {leaf}, {root}, etc.).
- All the stereotypes must remain in the resulting class.

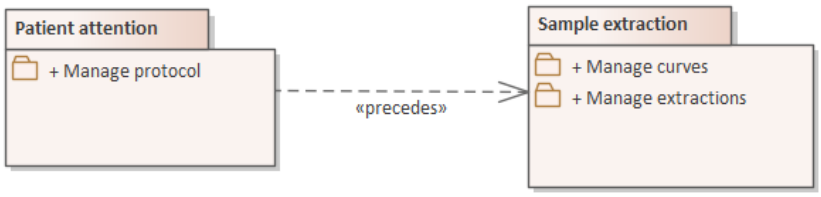
Composing attributes: all the attributes with different names are added to the resulting class. In case of attributes sharing the same name, we must include only one of them, but since they have several characteristics, we have to manage them individually. In UML [6] we specify the attributes according this pattern:

[visibility] [/] name [multiplicity] [:type] [= initial value] [{properties}]

so, we should consider the following criteria:

- In case of different visibilities, prevails the wider one, for instance, public over private.
- If an attribute is indicated as derived ([/]), this mark must remain.
- The lower side of every multiplicity must be the lowest of both, and the upper side must be the greater one.
- When the attributes have different types, the result must be a supertype generalizing both.
- The initial value will be the more comprehensive one, being blank in case of one of the attributes has no initial value.
- The properties will follow the recommendations for the «merge» relationship that UML uses among packages.
- All the stereotypes will remain on the resulting attribute.

Composing operations: the operations must follow a more complex procedure. Let's suppose we have two classes with the same name in two concerns (Concern1 and Concern2) and both have an operation called operation1(). We consider "the same operation" when both have identical signature. The resulting composition will include only one operation1() and will add two methods: concern1_operation1() and concern2_operation1(). Operation1() will call to the two new methods in the order indicated before the composition by means of a dependency relationship stereotyped by «precedes», as it is shown in Figure 1. In Figure 2 we present operacion1() in the composed class and in Figure 3 its resulting sequence diagram. We could use a numbering schema for the relationship precedence among the concerns as proposed in [9], but we prefer to indicate it by UML dependency relationships and applying a transitivity criterion.



Figur3 1: Precedence relationship for concern merging.

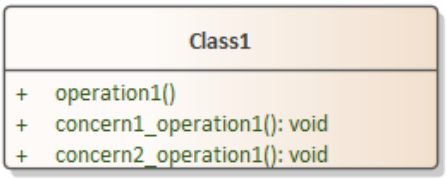


Figure 2: Operations after composition by merging.

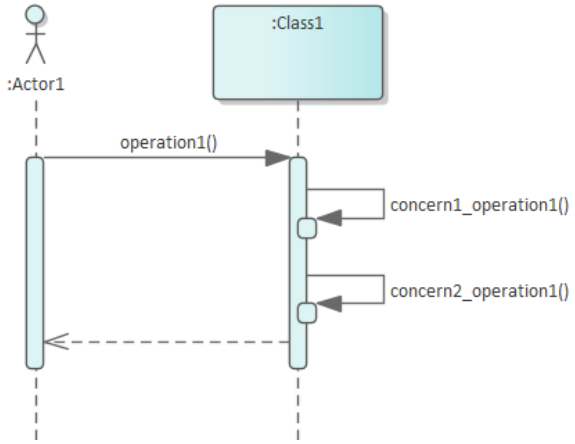


Figure 3: Operation sequence after composition by merging.

Beside this, in UML [6] operations are specified with the following characteristics:

[visibility] name [(list of parameters)] [:return type] [{properties}]

Since name, list of parameters and return type are the operation signature, and this is the base for the composition, we will just take in consideration these issues:

- In case of different visibilities, prevails the wider one, for instance, public over private.
- The properties will follow the recommendations for the «merge» relationship that UML uses among packages.

Composing associations: associations are merged considering the matching of name and meta-type. Since they are classifiers, we have to apply on them the same criteria mentioned for classes. Beside this, all the kind of associations relating objects (they are the associations where exists multiplicity) have the following attributes in UML:

- Order (how many elements are associated, being unary, binary ternary or n-ary associations).
- Association name.
- Name direction.

Associations have ends, which individually have properties:

- Multiplicity.
- Role.
- Navigability.
- Qualification.

Since name and meta-type are the base for composition, the criteria to be applied to the rest of the properties are as follows:

- The higher order must be taken, to avoid losing associations.
- The name direction remains just if it is the same, otherwise, indicate no direction.
- The lower side of every multiplicity must be the lowest of both, and the upper side must be the greater one.
- The role remains just if it is the same.
- Navigability remains just if it is the same.
- Qualification remains just if it is the same.

Jacobson call this composition by merging as “UML package merging operation” [4] by using the relationship stereotyped with «merge» [6]. We don’t use this approach since UML merge the elements from the merged package into the destination package (source package in the «merge» association), while in our approach we merge both packages into a new resulting one.

Composition by overriding: in this kind of composition we have to indicate which is the concern that overrides the other one. So, in the resulting the classes will prevail the members, constraints, stereotypes, etc., belonging to the overriding concern over the other ones. We indicate this prevalence among concerns by means of a relationship stereotyped with «overrides».

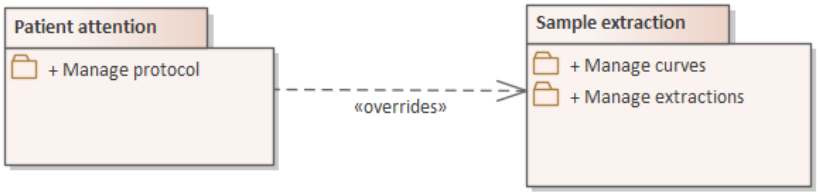


Figure 4: Overriding relationship for concern merging.

Let's continue with our previous example to explain this kind of composition. After composition we obtain a new class including operation1(), but we don't have to create two methods as we did previously, since this operation will only contain the functionality belonging to this operation as it has it in the overriding concern. In the example shown in Figure 4, we indicated that a package overrides to another one, so when we have classes with the same relative name in different packages, in the resulting class after composition will exist just the methods coming from the operations existing in the source package of the relationship when we had the same operation in both sides.

Independently of the composition method applied, all the classes modified after the composition procedure, we should maintain and turn false a tagged value "revised" ({revised=false}) in order to alert and to let the modeler analyze every change in the model and to act in consequence.

5 Composing crosscutting concerns in the static view

Just as we have peer concern among those belonging to the problem domain, we have other concern called crosscutting, because they transversally cut the peer concerns and they don't belong to the problem domain; what's more, they can be found in different domains, for instance: security, access control, authentication, etc.

In "The Theme Approach" [9], the authors propose parameterized packages to model the composition of crosscutting concerns, but we prefer to use parameterized classes because they belong to the standard UML and because they are supported by automated tools present in real-time settings.

We will use the example shown in Figure 5 to explain this kind of composition. Let's suppose that a crosscutting concern must be composed into a peer concern. The peer concern contains a class called Class1 that has several operations, and the crosscutting concern must be composed into one operation called operation1(). On the other hand, the crosscutting concern has a class called Clase2 implementing a "before" type advice that has to be composed into operation1(). This class is a template class and it has a parameter called operation3, that implies the trigger of the method called operation3(). Class1 is bound to Class2 through a dependency relationship stereotyped with «bind», which is a standard UML stereotype, and that indicates which is the operation that will be passed as parameter to Class2; in our example is operation1(). At the same time, operation3() calls to the methods operation4() and proceed() in that order, because it is a "before" type advice . We use

proceed() as a reference of operation1() passed as parameter (and, of course, equals any other operation that could be passed as parameter). We employ proceed() in order to conserve the concept that we presented in the business model of AOP4ST [10]. The resulting classes and sequence diagram got after the composition procedure are as follows:

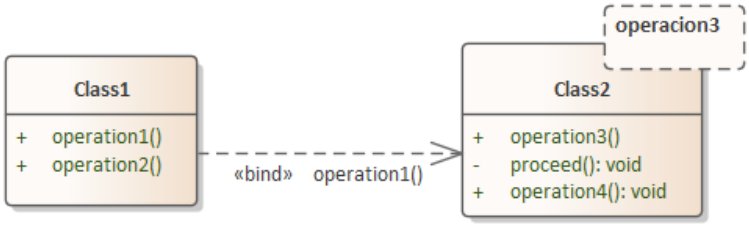


Figure 5: Class diagram of the crosscutting concern composition.

It is possible to use wildcards in «bind» associations, in order to reference all or part of the operations: asterisk "*" replace all the operations, and it is possible to use it combined with other characters. For instance, "abc*" will consider all the operations with a name beginning with "abc".

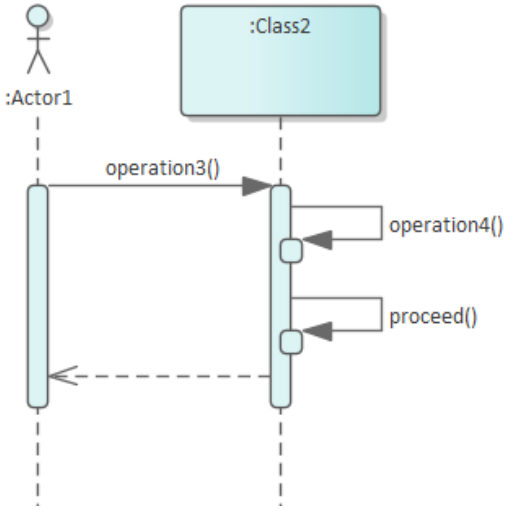


Figure 6: Sequence diagram of the crosscutting concern composition.

6 Conflict resolution

We said that composition is made by name and meta-type. In case of composition by merging we can find difficulties at the moment of merging elements. In composition by overriding the amount of conflicts decrease, but they can also appear. Conflicts can be of two types: formals or semantics.

Formal conflicts are easier to manage and solve, because they are related to the UML grammar, for instance:

- Impossible relationships, like cyclic generalizations or aggregations.
- Associations among elements not allowed.
- Merged concerns belong to the same hierarchical structure, this is, one of the packages contains directly or indirectly the others.

It is also possible to set some criteria for conflict resolution, just as listed below, that could be easily implemented with an automated tool:

- Definition by default: it could be possible to set the criteria that should employ the weaver, for instance, if it should take an element from a particular concern, or if which value should be taken if there are several values, etc.
- Definition by role: it could be specified to take the values for the attributes on the source or destination roles in a «precedes» association.
- Definition by selection: the weaver could present every conflict, allowing to select the alternative to choose.

Semantic conflict resolution is more complicated, because there could be collateral effects that could be difficult to discover. That's why every change must be perfectly identified in order to allow the modeler to be informed and to let him or her evaluate the possible impact. This doesn't warrant an accurate solution, but it will collaborate with conflict resolution. Automated functionalities to compare differences among models, beside conflict resolution by selection can be very valuable helps.

7 Conclusions

A lot of ways to apply use cases in aspect-oriented software development are proposed in scientific literature [11] [12] [13] [14] [15]. However, we decided to follow the creator of use cases, Ivar Jacobson. We have also selected Clarke and Baniassad approach, since we have found that it is he most cited proposal in scientific literature [16]. Of course, we have complemented both techniques and we have improved them by means of AOP4ST. Despite this, we clearly know that there is a lot of work pending to do.

We have finished two research projects about AOP4ST in our university, we are currently conducting the third one, and we have planned another project to specifically work on the issue addressed in this article. Conflict resolution is a very wide research area since it is very difficult predict all the possible conflicts that could appear on concern composition.

References

1. C. V. Lopes, “Aspect-Oriented Programming: An Historical Perspective (What’s in a Name?),” pp. 1–5, 2002.
2. A. Moreira, R. C. Jo, and A. Rashid, Aspect-Oriented Requirements Engineering. 2013.
3. E. Baniassad, S. Clarke, and S. C. Elisa Baniassad, “Theme: An approach for aspect-oriented analysis and design,” Proc. 26th Int. Conf. Softw. Eng. ICSE, pp. 158–167, 2004.
4. I. Jacobson and P. Ng, Aspect-oriented software development with use cases. Addison-Wesley, 2005.
5. F. Pinciroli, “AOP4ST - Aspect-Oriented Process for a Smooth Transition,” in 39th International Conference on Software Engineering, 2017.
6. OMG, “OMG Unified Modeling Language (OMG UML),” Language (Baltim)., vol. 05, no. November, pp. 1–212, 2007.
7. G. Kiczales et al., “Aspect-Oriented Programming,” ACM Comput. Surv., vol. 28, no. June, pp. 220–242, 1997.
8. I. Jacobson, “Use Cases and Aspects Working Seamlessly Together”.
9. S. Clarke and E. Baniassad, Aspect-oriented analysis and design. The Theme Approach. Boston: Addison-Wesley, 2005.
10. F. Pinciroli, “Aspect-oriented business process composition rules in AOP4ST,” in 35th International Conference of the Chilean Computer Science Society (SCCC 2016) held in conjuntion with the 42th Latin American Computing Conference (CLEI 2016)., 2016, pp. 1–6.
11. R. Pawlak and H. Younessi, “On Getting Use Cases and Aspects to Work Together,” J. Object Technol., vol. 3, no. 1, pp. 15–26, 2004.
12. R. Lopez-Herrejon and D. Batory, “Modeling Features in Aspect-Based Product Lines with Use Case Slices: An Exploratory Case Study,” Model. Softw. Eng., pp. 6–16, 2007.
13. K. Bittner, “Why Use Cases Are Not "Functions" Here’s Why,” 2001.
14. P. P. Dey, B. R. Sinha, M. Amin, and H. Badkoobehi, “Augmenting Use Case View for Modeling,” vol. 6, no. 12, pp. 1303–1306, 2012.
15. A. Rago, E. S. Abait, C. Marcos, and A. Díaz-Pace, “Early aspect identification from use cases using NLP and WSD techniques,” 15th Work. Early Asp. - EA ’09, pp. 19–26, 2009.
16. F. Pinciroli, L. Zeligueta, M. I. Lund, and J. L. Barros Justo, “Systematic Mapping Protocol - Coverage of Aspect-oriented Methodologies for the Early Phases of the Software Development Life Cycle,” 2017.

CALIDAD UNIVERSITARIA. Una mirada exploratoria del año 2018

María Inés Lund[1] y Myriam Herrera[1]

[1]Departamento e Instituto de Informática. Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de San Juan, Argentina
{mlund, mherrera}@iinfo.unsj.edu.ar

Resumen. Se presenta la metodología usada y los resultados obtenidos de aplicar los pasos de un mapeo sistemático para explorar, en todo el año 2018, que se ha dicho o investigado en relación a la calidad universitaria, especialmente en un contexto latinoamericano. Responder preguntas en relación a cuales son los objetos de estudio y que métodos o instrumentos se utilizan para recolección de datos y su posterior análisis estadístico, cuales son los tipos de publicación elegidos y qué países son los más abocados o están más dedicados al estudio de la calidad universitaria .

Palabras clave: calidad universitaria, calidad académica, mapeo sistemático, estudio exploratorio sistemático.

1 Introducción

Desde hace años que las universidades y sus claustros están abocados a mejorar la "calidad universitaria". En el caso de Argentina en particular, más formalmente con el proceso de acreditación de las carreras exigidas por CONEAU (Comisión Nacional de Evaluación y Acreditación Universitaria), de la SPU (Secretaria de Políticas Universitarias) del MEN (Ministerio de Educación de la Nación), que nace en 1995 para dar cumplimiento a la Ley Nº 24.521 de Educación Superior [1].

La acreditación es un mecanismo de aseguramiento de la calidad de la educación superior. Ese mecanismo de aseguramiento, que define la acreditación por estándares mínimos, se constituyó en un instrumento central para el diseño de políticas de Estado [2].

La búsqueda de la calidad o como medirla o evaluarla es una constante desde hace muchos años. Existe numerosa bibliografía que da fe de ello, no sólo a nivel nacional, sino también internacional.

Ante las preguntas de investigación: ¿Qué se analiza para medir la calidad universitaria?, ¿Qué instrumento o método de recolección de datos se usa? ¿Dónde se publican estas investigaciones?, ¿Quiénes estudian o analizan la calidad universitaria?, se realizó una búsqueda bibliográfica en Google Scholar [3], siguiendo un criterio específico y sistemático de búsqueda, con parámetros definidos. Es decir, se siguieron los lineamientos de un mapeo sistemático [4], [5] usando como fuente únicamente a este buscador académico, que se especializa en búsqueda de contenido y literatura científico-académica, para responder las preguntas de investigación.

Este trabajo de investigación se enmarca en el proyecto E-1050 "Calidad Educativa Universitaria mediante Técnicas de Clasificación", Res Nº 21/18-CS-UNSJ (2018-2019).

2 Marco Teórico

2.1 Educación Superior

En la Declaración Final de la Primera Conferencia Mundial de Educación Superior de 1998 (CMES 1998) se planteó que la educación es un bien público y un derecho, donde el Estado tiene un papel importante en su financiamiento, que se debe democratizar su acceso y que la investigación tiene un papel crucial en la contribución a los sistemas de innovación, así como en la sociedad. Esta referencia fue el concepto basal de educación superior. Diez años después, durante la Conferencia Regional de Educación Superior de 2008 (CRES 2008), la Declaración de Cartagena sostiene de manera contundente que *"la educación es un bien público social, un derecho humano y universal y un deber del Estado". Jugando un papel estratégico en los procesos de desarrollo sustentable de los países de la región"* [6].

En América Latina, la firma de convenios de integración económica entre los países de nuestra región, como el MERCOSUR, abrió un campo amplias acciones para la educación superior y particularmente para las políticas de internacionalización, se constituyó el Sector Educativo del MERCOSUR (SEM), se promovió la creación del Núcleo de Estudios e Investigaciones en Educación Superior del Mercosur (NEIES) con el propósito de contribuir y cooperar interinstitucionalmente entre los sistemas universitarios, impulsando la reflexión y producción de conocimiento sobre la Educación Superior vinculada a la integración [6].

2.2 Calidad universitaria

Mejorar la calidad universitaria implica un cambio, si se sigue haciendo lo mismo, las cosas seguirán igual. Para mejorar la educación se debe innovar. *"La innovación se convierte en el camino que se sigue cuando se asume una evaluación orientada a la mejora de calidad universitaria, ya sea a nivel de sistema, a nivel de institución, a nivel de unidad funcional o a nivel del docente"*[7]. La educación tiene que transformarse para ajustarse a un contexto que es cambiante por naturaleza, con el objeto de conseguir mejores resultados.

En [8] García-Peñalvo et.al presentan una ontología de indicadores para la innovación educativa, organizado en 4 regiones no disjuntas:

1. Perspectiva institucional. tendencias más relacionadas con la toma de decisiones, planificación estratégica, gestión de la tecnología y gestión de la propia innovación. Para que una universidad se vea inmersa en un proceso de innovación, la institución debe ser capaz de gestionar la innovación, involucrando a los distintos miembros de la comunidad universitaria.

2. Perspectiva del profesorado. Relacionada al contexto del profesor y su docencia, abarca todas aquellas innovaciones que tienen un carácter más ligado a la impartición de los contenidos curriculares.
3. Desarrollo de competencias transversales: se potencian las competencias transversales, muy ligadas a las denominadas habilidades blandas (soft skills) que tanta importancia tienen en el mercado laboral y en el desarrollo más humanista de los estudiantes.
4. Perspectiva de extensión institucional: se desarrollan los aspectos más novedosos que tienen que ver con la labor de extensión hacia la sociedad y la formación permanente.

Se puede concluir que la innovación educativa tiene múltiples perspectivas que se complementan. Se deben mantener los propósitos de innovar en los procesos universitarios para mejorar el aprendizaje de las personas y por ende la calidad en la educación.

En [9] se plantea que en el proyecto institucional universitario se evalúa: 1) docencia, investigación, desarrollo y creación; 2) extensión, producción de tecnología y transferencia; 3) gestión y gobierno; 4) recursos humanos; 5) infraestructura y recursos materiales; 6) servicios de biblioteca, de información y de carácter informático; 7) integración de la institución universitaria.

En el trabajo [10] se expone la guía del Plan Nacional de Evaluación (España), articulando el contenido en tres grandes bloques: enseñanza, investigación y gestión. Estos bloques se dividen en indicadores y sub-indicadores, para facilitar la operatividad, pudiendo cada universidad introducir modificaciones, para tener en cuenta su propio contexto. En ese artículo se exponen detalladamente los indicadores y conceptos que cada uno de ellos involucra en cada bloque de análisis.

2.3 Rankings universitarios como evaluadores de la calidad de la ES

Los rankings universitarios muestran el impacto que tienen las universidades en el ámbito social y científico, generando un impacto en la opinión pública y una evaluación y análisis hacia el interior de las mismas universidades. Llegando a condicionar las políticas universitarias. En el "El influjo de los rankings sobre la calidad en las universidades argentinas" [11] se hace una síntesis de la evolución y características de los rankings en Argentina y su adopción como evaluador de la calidad de la educación superior.

Claramente, los rankings tienen un propósito diferente a la evaluación promovida como política pública para el sistema de ES, además de la noción que hay detrás de la "calidad" para ellos y del "negocio" ligado a las evaluaciones universitarias. Existe una gran variedad de rankings internacionales y hay varios trabajos al respecto que han desmenuzado las historias que hay detrás de cada una de las agencias de rankings, entre ellos el muy completo estudio de Barsky [12], en donde la docencia, en este aparece claramente subvalorada por los rankings. Los indicadores se ponderan entre sí y arrojan un índice de calidad que permite comparar y jerarquizar a las universidades, bajo los cuales sólo el 2% de universidades del mundo caen entre las 500 rankeables. *"El sólo intento de rankear en orden descendente a miles de instituciones universitarias de diferentes países implica ocultar la principal característica de las*

universidades: que son normalmente grandes, complejas y que articulan una extraordinaria diversidad de tradiciones universitarias y disciplinarias. A su vez, las universidades pueden tener algunas disciplinas con fuerte tradición en investigación y otras en transferencia, o en calidad de la docencia" [12].

Por su propia naturaleza institucional, las universidades agrupan carreras disímiles que integran disciplinas con tradiciones diferenciadas en relación a los campos del conocimiento. Someterlas a un tratamiento similar y a escala mundial no es un buen criterio para comparar, clasificar y ordenar. No todas las universidades tienen los mismos objetivos ni la misma historia institucional y nacional; por lo tanto, no pueden jugar el mismo juego, ni competir directamente entre sí [13].

Uno de los principales motivos del rezago latinoamericano en esos rankings internacionales, es que la mayor parte de la investigación (mayor peso en los ranking) se hace en el ámbito académico estatal y está divorciada de las necesidades del mercado [14]. Este autor realiza un análisis detallado de los posicionamientos de algunos países en los rankings internacionales y fundamenta las razones de ello, haciendo especial hincapié en las universidades latinoamericanas.

3 Método de trabajo

Para poder satisfacer las preguntas de investigación surgidas en el proyecto antes citado, se realizó un mapeo sistemático [4], [5], siguiendo básicamente lo propuesto en la guía de mapeo sistemático [15]. Los estudios de mapeos sistemáticos también son conocidos como estudios de alcance y son diseñados para dar una vista general de un área de investigación, por medio de la clasificación, contando contribuciones en relación a las categorías de clasificación.

La búsqueda bibliográfica se realizó en Google Scholar [3]. Se usó este buscador debido a que en sondeos previos, no aparecieron resultados significativos respecto a calidad universitaria en las bibliotecas digitales de uso común en estos casos, como ACM [16], IEEExplore [17], ó Scopus [18], ente otros. Google Scholar se especializa en búsqueda de contenido y literatura científico-académica, en el cual, siguiendo un criterio específico y sistemático de búsqueda, y con parámetros definidos, se pueden obtener resultados que permiten responder las preguntas de investigación.

3.1 Preguntas de investigación

El objetivo que se persigue en esta investigación es conocer o identificar ¿qué se entiende por calidad universitaria o a qué se refieren los estudios cuando hablan de calidad universitaria o qué aspectos se analizan para considerar la calidad en la educación superior?, pensando principalmente en estudios con base estadística, realizados o propuestos. Esto lleva a generar las siguientes preguntas de investigación:

P1. ¿Cuáles son los objetos de estudio en relación a la calidad universitaria?,
P2. ¿Cuáles son los instrumentos/métodos de recolección de datos?
P3. ¿Qué tipos de publicaciones son las usadas para exponer resultados?
P4. ¿Qué países o quienes están abocados al estudio de la calidad universitaria?

3.2 Búsqueda

En ésta etapa, se debe definir el protocolo de búsqueda, para ello se aplica la técnica PICO [15] (siglas en inglés de Population, Intervention, Comparison and Outcomes), para identificar palabras claves y formular strings de búsqueda, a partir de las preguntas de investigación.

En este caso, Population o Población: son todas las publicaciones de calidad universitaria, que satisfagan las características requeridas. Intervention o Intervención: identificar qué cual es el objeto de estudio que se mide o evalúa. Respecto a Comparasion o Comparación: Si bien no es una comparación estadística, se analizan las distintas características identificadas en el estudio de mapeo, en base a la intervención. En cuanto a Outcomes o Resultados: Son los resultados del proceso del mapeo sistemático ya analizado.

Para la identificación de la literatura relevante se definieron cadenas de búsqueda, conteniendo las palabras relevantes para considerar en la selección del buscador. La cadena resultante, teniendo en cuenta las reglas del buscador Google Scholar es: (medición OR dimensión OR medida OR indicador OR métrica OR índice OR variable) AND estadística AND intitle:("calidad educación superior" OR "calidad universitaria").

El resultado de la búsqueda fue de 381 hallazgos, al 18/09/2019.

3.3 Selección de hallazgos pertinentes

Los criterios de selección, también denominados criterios de inclusión/exclusión, tienen por propósito identificar, del total de hallazgos de la búsqueda, aquellos que proporcionan evidencias directas acerca de las preguntas de investigación. Los criterios de inclusión/exclusión definidos para este estudio son:

- Criterios de inclusión: -trabajos publicados exclusivamente en el año 2018, -artículos de revistas, de conferencias, libros, capítulos de libros y proceedings de conferencias, informes, tesis, -relacionados directamente a la calidad universitaria, -que realice o proponga mediciones con base estadística.
- Criterios de exclusión**:** -los que no cumplan el criterio de inclusión, -aquellos hallazgos que no tengan relación directa a la calidad universitaria, -repetidos, -informes de avance o final de proyectos, -reseñas, editoriales o prefacios de libros, revistas o proceedings.

A cada uno de los resultados obtenidos en la búsqueda se aplicaron los criterios precedentes en base a la lectura de título, abstract, y palabras claves. Si bien el buscador permite seleccionar por intervalo (en este caso el 2018) se excluyeron 24 hallazgos que no correspondían al intervalo, 24 repetidos, Se descartaron 240 resultados que, aunque contuvieran alguna de las palabras de la cadena de búsqueda, sus contenidos estaban poco o nada relacionados con el tema central de la investigación y 4 que cayeron en la categoría de opiniones, reseñas o informes de avance.

En total se excluyeron 292 hallazgos que no cumplieron con los criterios definidos. La lista final se redujo a 89 estudios.

3.4 Extracción de los datos, análisis y clasificación

Los hallazgos incluidos en el presente estudio se volcaron en una planilla, almacenando los siguientes datos: Nombre del estudio, tipo de estudio (tesis, artículo de journal, o de conferencia, libro, capítulo de libro, libro proceeding de conferencia), objeto de estudio (rankings, acreditación, autoevaluación, docentes, alumnos, egresados, gestión, etc), palabras claves, país que realiza el estudio (en los artículos la institución del primer autor, en las tesis el que corresponde a la universidad que otorga el título) y el resumen. A su vez cada uno de estos fue almacenado en el gestor de referencias bibliográficas Mendeley, para su posterior cita y recuperación del hallazgo.

3.5 Evaluación de la validez

En cuanto a la validez descriptiva, teórica, generalizable e interpretativa, en este trabajo no se siguieron todas las sugerencias propuestas por [15]. Todos los hallazgos fueron analizados por una sola persona y no se aplicó la técnica de snowball. Por lo tanto, trataremos este trabajo de investigación como un estudio exploratorio, realizado en forma sistemática, siguiendo los lineamientos de un mapeo.

Los resultados de la búsqueda en google scholar tampoco es repetible en forma exacta. La cantidad de hallazgos varía, aunque el intervalo sea el mismo.

4 Resultados de aplicar el mapeo para el estudio exploratorio

4.1 P1. ¿Cuál es el objeto de estudio en relación a la calidad universitaria?

Los objetos de estudio fueron clasificados como sigue:

- Acreditación/Evaluación/Autoevaluación: estos tres tipos se unifican debido a que para toda acreditación se requiere un proceso de Autoevaluación y se realiza una Evaluación para definir la Acreditación de una Institución. Se considera que una institución acreditada tiene un nivel satisfactorio de calidad universitaria/académica/de servicio.
- Docente. Calidad universitaria en base al desempeño del profesor, nivel de estudio, si el clima laboral influye en su desempeño, productividad, etc.
- Calidad del servicio Académico/Educativo: son los estudios más relacionados con el proceso de enseñanza-aprendizaje.
- Calidad de la gestión universitaria: incluye los estudios relacionados a la gestión en general (incluida infraestructura, comunicaciones, administración y otros servicios universitarios de soporte).
- Rendimiento académico: son los estudios que direccionan al estudiante y su rendimiento académico, como medida de calidad universitaria.
- Rankings: analizan los rankings como indicador de calidad universitaria.

- Modelos educativos de calidad: se refieren a modelos o propuestas de modelos, en forma explícita, que ayudan a mejorar la calidad educativa o de gestión de calidad.

La tabla 1 presenta los hallazgos según estas categorías mencionadas. Como se observa, las investigaciones muchas veces analizan más de un objeto de estudio y aparecen en diferentes renglones.

Tabla 1. Hallazgos según objeto de estudio.

Objeto de Estudio	Cant.	Hallazgos
Acreditación/ Evaluación/ Autoevaluación	25	[19][20][21][22][23][24][25][26][27][28][29][30][31][32] [33][34][35][36][37][38][39][40][41][42][43]
Docente	13	[19][44][45][46][47][48][49][50][51][52][53][54][55]
Calidad del Servicio Académico/ Educativo	22	[56][57][58][59][60][61][62][63][64][65][66][67][68][69] [70][71][72][73][74][75][76][77]
Calidad de la gestión universitaria	19	[73][74][78][79][80][81][82][83][84][85][86][87][88][89] [90][91][92][93][94]
Rendim. Académico	5	[61][72][95][96][97]
Rankings	6	[11][13][98][99][100][101]
Modelos educativos/ gestión de calidad	13	[30][31][34][35][62][81][82][85][92][102][103][104][105]

4.2 P2. ¿Cuál es el instrumento/método de recolección de datos?

En la tabla 2 se muestran los instrumentos o formas de recolección de datos, los cuales fueron categorizados en:

- Análisis documental: es el análisis bibliográfico, estudio de casos, revisión bibliográfica, etc,
- Encuestas a docentes: entrevistas, cuestionarios respondidos por docentes.
- Encuestas a estudiantes: entrevistas, cuestionarios respondidos por estudiantes.
- Encuestas a administración: entrevistas, cuestionarios respondidos por personal administrativo, directivos y autoridades del ámbito universitario.
- Encuestas a graduados: entrevistas, cuestionarios respondidos por graduados.
- Encuestas a externos de la Universidad: acá se consideran aquellas entrevistas o cuestionarios que son respondidos por la comunidad o ciudadanos que no pertenecen al ámbito universitario, empleadores o autoridades gubernamentales.
- Seguimiento de procesos: son aquellos estudios que manifiestan el control y medición de los procesos de acreditación.
- Observaciones: observación directa de la situación y toma de datos.
- Varios/Otros/Cuestionarios sin especificar: todos aquellos trabajos que incluyen varios métodos o cuestionarios sin especificarlos u otros especiales.
- BD académicas nacionales.

Tabla 2. Hallazgos según instrumento de recolección o medición.

Instrumento de recolección	Cant	Hallazgos
Análisis documental	22	[11][13][26][27][28][29][32][38][41][42][52][58][70][76] [89][90] [91][98][99][100][101][102]
Encuestas a docentes	29	[25][27][30][33][35][40][42][43][45][46][47][50][52][54] [55][65][69][72][76][79][81][84][87][88][89][92][94][97] [103]
Encuestas a estudiantes	44	[19][20][21][25][30][33][35][37][40][43][44][45][48][49] [51][52][53][56][59][60][61][63][65][66][67][71][72][73] [74][76][77][78][81][82][83][88][92][93][96][97][102] [103][104][105]
Encuestas a administrativos	14	[33][35][39][40][43][47][52][62][76][79][81][88][92][94]
Encuestas a graduados	10	[30] [33][57][61][64][68][81][91][92][103]
Encuestas a externos a la U	4	[33][42][43][68]
Seguimiento de procesos	2	[22][23]
Observaciones	5	[42] [76][92][104][105]
Varios/Otros/sin especificar	7	[24][31][34][75][80][85][86]
BD académicas nacionales	2	[36][95]

4.3 P3. ¿Qué tipos de publicaciones son las usadas para exponer resultados de calidad universitaria?

Tabla 3. Hallazgos por tipo de publicación.

Tipo de publicación	Cant.	Hallazgos
Artículo Journal	29	[11][13][24][26][27][31][38][39][41][42][43][45][49][53][64] [74][78][79][85][90][91][94][95][99][100][101][102][103][105]
Artículo Congreso	4	[29][57][61][68]
Capítulo de Libro	1	[23]
Libro	4	[73][75][80][97]
Proceeding Congreso	1	[86]
Tesis	50	[19][20][21][22][25][28][30][32][33][34][35][36][37][40][44] [46][47][48][50][51][52][54][55][56][58][59][60][62][63][65] [66][67][69][70][71][72][76][77][81][82][83][84][87][88][89] [92][93][96][98][104]

4.4 P4. ¿Qué países están abocados al estudio de la calidad universitaria?

Tabla 4. Hallazgos según origen de publicación.

País	Cantidad	Hallazgos
Angola	1	[39]
Argentina	11	[11][13][23][26][32][43][57][61][79][86][90]
Bolivia	1	[96]
Chile	3	[36][64][98]
Colombia	6	[58][68][69][70][73][83]
Latinoamérica	1	[75]

Cuba	6	[24][27][28][94][100][105]
Ecuador	11	[29][34][38][41][42][47][78][80][88][97][104]
España	3	[76][91][99]
México	6	[22][31][49][74][85][101]
Nicaragua	1	[95]
Perú (solo 4 no son tesis)	37	[19][20][21][25][30][33][35][37][40][44][46][48][50][51][52][53] [54][55][56][59][60][62][63][65][66][67][71][77][81][82][84][87] [89][92][93][102][103]
Portugal	1	[72]
Venezuela	1	[45]

5 Conclusiones y trabajo a futuro

Como conclusión de la P1 los objetos de estudio que mayor análisis recibieron fueron los relacionados directamente a Acreditación/Evaluación/Autoevaluación, Calidad del Servicio Académico / Educativo y Calidad de la gestión universitaria, correspondiendo a un 64% sobre el total de objetos de estudio encontrados.

Respecto a la P2, el instrumento de recolección más ampliamente usado es la encuesta (ya sea entrevista o cuestionario), con un 73% en relación a todos los instrumentos o métodos de recolección, siendo la perspectiva del alumno principalmente y luego la del docente, los más requeridos para la captura de datos. Le sigue el análisis documental o bibliográfico. Estos tres tipos de método de recolección de datos representan el 68% del total de los instrumentos identificados en los estudios. Es decir que generalmente se mide la calidad universitaria por la percepción de sus actores

Tanto para los hallazgos para P1 y P2, una misma contribución científica puede hacer referencia a más de un objeto de estudio y a más de un método o instrumento de recolección de datos. Cabe notar que todos los estudios relacionados a ranking usan el análisis documental como método de recolección de datos.

La P3 arroja resultados inesperados: la mayor cantidad de contribuciones científicas en relación a calidad universitaria se realizan a través de tesis (de grado y posgrado), un 56%, en tanto los artículos de revistas aportan con un 33%.

En cuanto a la P4, es Perú quien se lleva la bandera en aportes a la investigación de la calidad universitaria, casi un 46%, siguiendo en segundo lugar, con un 12% de aportes Argentina y Ecuador por partes iguales. Es de notar que de las contribuciones de Perú son, a excepción de 4, todas investigaciones de tesis, y, por ejemplo Cuba, sus 6 aportes corresponden a artículos de revistas.

Como trabajo a futuro, se debe analizar lo relacionado a la P3, porqué la mayoría de las contribuciones sobre esta temática se vuelcan en tesis y no en Journals indexados?, porque incluso muchas de las revistas de Ciencias y Educación en donde están publicados los artículos encontrados, no están indexadas?. Esto ya fue avizorado cuando en un principio se indagó por los buscadores especializados, como se comenta en la sección 3.

Respecto al instrumento o método de recolección de datos, se debería examinar más profundamente e identificar indicadores que permitan generar métricas en base a

datos de medición cuantitativos, que sumados a los cualitativos de las encuestas aporten un panorama más real de la calidad universitaria.

También sería interesante indagar sobre los distintos modelos educativos o de gestión de calidad educativa-universitaria, usados especialmente en Latinoamérica, donde la cultura y la idiosincrasia es bien diferente a Europa y EEUU, con el fin de encontrar un modelo que se adapte y respete las características propias de cada universidad y su contexto.

Los resultados obtenidos en este estudio exploratorio sistemático dan lugar para más conclusiones y generan más interrogantes para su análisis, que por limitaciones de espacio no se pueden plasmar.

Referencias

[1] Ministerio de Eduicación, *Ley Nº 24.521 de Educación Superior*. Miniserio de Educación de la Nación Argentina, 1995, p. 16.

[2] CONEAU, "Documentos Institucionales | CONEAU," *CONEAU. Calidad en la Educación Superior*, 2015. [Online]. Available: https://www.coneau.gob.ar/archivos/publicaciones/estudios/20anios_CONEAU.pdf. [Accessed: 21-Feb-2019].

[3] "Google Académico," 2004. [Online]. Available: https://es.wikipedia.org/wiki/Google_Académico. [Accessed: 21-Feb-2019].

[4] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic Mapping Studies in Software Engineering," 2008, pp. 68–77.

[5] B. Kitchenham, P. Brereton, and D. Budgen, "The educational value of mapping studies of software engineering literature," in *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering - ICSE '10*, 2010, vol. 1, p. 589.

[6] L. Korsunsky, D. Del Valle, E. Miranda, and C. Suasnábar, Eds., "Internacionalización y producción de conocimiento.El aporte de las redes académicas," in *Coloquio Regional: Balance de la declaración de Cartagena y aportes para la CRES 2018*, 2018, p. 206.

[7] F. J. García-Peñalvo, "Innovación educacional," Grupo GRIAL, Salamanca, May 2018.

[8] F. J. GARCÍA-PEÑALVO, "Mapa de tendencias en Innovación Educativa," *Educ. Knowl. Soc.*, vol. 16, no. 4, p. 6, Dec. 2015.

[9] S. M. Araujo, "La evaluación y la Universidad en Argentina: políticas, enfoques y prácticas," *Rev. la Educ. Super.*, vol. 43, no. 172, pp. 57–77, Oct. 2014.

[10] A. T. Requena, "La evaluación en la universidad española.," *Calid. en la Educ.*, vol. 0, no. 16, p. 22, May 2002.

[11] S. Emiliozzi, "El influjo de los rankings sobre la calidad en las universidades argentinas," *Rev. Iberoam. Ciencia, Tecnol. y Soc. - CTS*, vol. 13, no. 37, Feb. 2018.

[12] O. Barsky, *Los rankings internacionales de las universidades y el rol de las revistas científicas*. Buenos Aires: Teseo, 2014.

[13] O. Barsky, "La inconsistencia de los rankings internacionales de las universidades y la debilidad del debate académico en la Argentina," *Rev. Iberoam. Ciencia, Tecnol. y Soc. - CTS*, vol. 13, no. 37, Feb. 2018.

[14] A. Oppenheimer, *¡Basta de historias! : la obsesión latinoamericana con el pasado y las doce claves del futuro*. Vintage Español, 2010.

[15] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Inf. Softw. Technol.*, vol. 64, pp. 1–18, 2015.

[16] ACM, "ACM Digital Library," *http://dl.acm.org*. [Online]. Available: http://dl.acm.org.

[17] IEEE, "IEEE Xplore Digital Library," *http://ieeexplore.ieee.org*. [Online]. Available: http://ieeexplore.ieee.org.

[18] Scopus, "Scopus - bibliographic database (ELSEVIER)," *www.scopus.com*. [Online]. Available: www.scopus.com.

[19] E. V. Muñoz Javier, "Acreditación y el desempeño profesional de los docentes de la Facultad de Educación de la Universidad Nacional Daniel Alcides Carrión, 2016," Universidad Nacional de Educación Enrique Guzmán y Valle. Escuela de Posgrado., 2018.

[20] K. L. Sialer Pérez, "Acreditación programática SINEACE 2015 y satisfacción del estudiante de la carrera de Ingeniería Agroindustrial y Agronegocios de una Universidad Privada de Lima en el primer semestre 2018," Universidad San Ignacio de Loyola, 2018.

[21] M. Collazos Tunjar, "Acreditación y gestión de calidad educativa en la Institución Educativa Emblemática Antenor Orrego Espinoza - Zarate San Juan de Lurigancho 2017," Universidad Nacional de Educación Enrique Guzmán y Valle. Escuela de Posgrado., 2018.

[22] E. Muñoz-Peñuela and Elizabeth, "Actualizar el modelo de autoevaluación institucional y de los programas de la universidad piloto de Colombia para desarrollar la cultura de calidad," ITESO, 2018.

[23] A. Guaglianone, "Aportes para el análisis del impacto de las políticas de evaluación y acreditación en Argentina," in *La Ley de educación superior : impactos, desafíos e incertidumbres*, O. Barsky and Á. (compiladores) Corengia,

Eds. CABA, Argentina: Editorial Teseo, 2018, 2018, p. 414.
[24] E. González Cruz, F. A. Fernández, and Z. M. Ledesma Martínez, "Cualimetría en instituciones de educación superior cubanas: caso infraestructura y gestión de recursos," *Atenas. Rev. Científica Pedago.*, vol. 1, no. 41, pp. 65–82, Jan. 2018.
[25] I. A. León Castro, "El Modelo De Acreditación Para Programas De Estudios De Educación Superior Universitaria Y El Mejoramiento de la gestión de La calidad de La Facultad de Educación de La Universidad Nacional de Cajamarca, 2018," Universidad Nacional de Cajamarca, 2018.
[26] A. Stubrin, "Entre la uniformidad y la atomización. Un aporte sobre los impactos del aseguramiento de la calidad sobre instituciones y programas universitarios en América Latina," *Rev. Iberoam. Ciencia, Tecnol. y Soc. - CTS*, vol. 13, no. 37, pp. 85–94, Feb. 2018.
[27] Y. Ugarte Martínez, N. Blanco Balbeito, M. Amechazurra Oliva, Y. Betancourt Roque, I. García López, and T. Y. Fleites Did, "Estrategia metodológica para la acreditación de las especialidades de Nefrología, Urología y Terapia Intensiva," *Rev. Educ. Médica del Cent.*, vol. 10, no. 1, pp. 124–139, 2018.
[28] X. S. Guillén Vivas, "Estrategias para mejorar la gestión del proceso de autoevaluación institucional con fines de acreditación en la Universidad San Gregorio de Portoviejo. Ecuador," Editorial Ciencias Médicas, 2018.
[29] S. A. Acurio Maldonado, D. M. Acurio Maldonado, V. M. Pailiacho Mena, and D. J. Robayo Jácome, "ESTUDIO DE LA EVALUACIÓN DE LA CALIDAD DE LA EDUCACIÓN SUPERIOR ECUATORIANA," in *III Congreso: CIENCIA, SOCIEDAD E INVESTIGACIÓN UNIVERSITARIA*, 2018.
[30] C. R. Suarez Sánchez, "Evaluación De La Formación Integral Del Sociólogo En La Universidad Nacional De Cajamarca, Según El Modelo De Acreditación Para Programas De Estudios De Educación Superior Universitaria - 2016," Universidad Nacional de Cajamarca, 2018.
[31] J. Solis González, J. Rositas Martínez, and J. N. Barragán Codina, "Factores críticos para la implementación exitosa de sistemas de gestión de calidad ISO 9001:2015 en el Nivel Medio Superior de la UANL," *Daena Int. J. Good Conscienc. ISSN 1870-557X*, vol. 13, no. 1, pp. 151–177, 2018.
[32] T. D. Barrera Carhuancho, "Gestión del cambio en universidades públicas peruanas desde la percepción de los gestores, en el marco del licenciamiento institucional," Universidad Nacional de La Plata, 2018.
[33] F. Quispe Mamani, "Impacto del proceso de acreditación en la mejora continua en el Instituto Superior Pedagógico Público - Juli, 2017," Universidad Nacional del Altiplano, 2018.
[34] S. A. Acurio Maldonado, "Implementación de un sistema de evaluación de la calidad de la investigación y el postgrado. Caso: Pontificia Universidad Católica del Ecuador Sede Ambato," Pontificia Universidad Católica del Ecuador, 2018.
[35] R. M. Padilla Santos, "Indicadores de autoevaluación para la mejora del modelo de calidad institucional de la Universidad Peruana de Las Américas, Lima 2017," Universidad Peruana de las Américas, 2018.
[36] P. Ferreiro Gómez, "Indicadores de calidad y acreditación institucional," Universidad de Chile, 2018.
[37] L. D. Pacheco Rodríguez, "Influencia del sistema de evaluación implantado en la calidad educativa percibida por los estudiantes de la Universidad Nacional de San Agustín, Arequipa 2018," Universidad Nacional de San Agustín de Arequipa, 2018.
[38] J. M. Luperón Terry and G. G. López Bustamante, "LA AUTOEVALUACIÓN COMO FOMENTO DE LA CALIDAD EN LA FORMACIÓN DEL PROFESIONAL DE CULTURA FÍSICA," *Rev. Mag. las Ciencias. ISSN 2528-8091*, vol. 3, no. 2, pp. 103–116, Jun. 2018.
[39] M. Afonso Vindassi, M. Miranda Lopes, and R. Mário José da Costa, "LA AUTOEVALUACIÓN INSTITUCIONAL PARA MEDIR EL IMPACTO DE LA FORMACIÓN DEL PROFESIONAL EGRESADO DEL ISCED-HUAMBO," *Rev. Órbita Pedagógica*, vol. 5, no. 1, pp. 145–157, Sep. 2018.
[40] M. Mostajo Vargas, "La autoevaluación con fines de acreditación de la escuela de educación inicial de la Facultad de Educación de la Universidad Nacional Federico Villarreal 2011 -2012," Universidad Nacional Federico Villarreal, 2018.
[41] M. Ayala Bolaños, "La Gestión de la Autoevaluación como variable para el aseguramiento de la Calidad de la Educación Superior en el Ecuador.," *"Carácter" Rev. Cient. la Univ. Del Pacifico ISSN 2602-8476*, vol. 6, no. 1, Dec. 2018.
[42] L. A. Cadena Hurtado, A. A. Zambrano Rodríguez, A. M. León Castro, G. Real Zumba, and C. Marilud Filian, "Los procesos de autoevaluación en función de la calidad en la gestión de la educación superior.," *Dilemas Contemp. Educ. Política y Valores*, vol. IV, pp. 1–11, 2018.
[43] C. Marquis, "Reflexiones sobre el impacto de la evaluación universitaria en Argentina," *Rev. Latinoam. Educ. Comp. ISSN-e 1853-3744*, vol. 8, no. 12, pp. 59–69, 2018.
[44] M. Cruzalegui Delgado, "Calidad del servicio educativo y el desempeño docente en el ISPPEE 'María Madre' de la región Callao," Universidad Nacional de Educación Enrique Guzmán y Valle. Escuela de Posgrado., 2018.
[45] J. Gutierrez Zehr, L. Jauregui, and G. Cifuentes, "Diseño de un modelo para la evaluación integral del Profesor Universitario," *Rev. Pedagog.*, vol. 39, no. 105, pp. 83–109, 2018.
[46] M. Capcha Flores, "El clima laboral en el desempeño profesional académico de los docentes en la Universidad Nacional 'San Luis Gonzaga' de Ica en el año 2013 -2015," Universidad José Carlos Mariátegui, 2018.
[47] L. M. Moreira Moreira, "EL clima organizacional y su influencia en la productividad laboral en las Instituciones de Educación Superior (IES) públicas de la provincia de Manabí Ecuador," Universidad Nacional Mayor de San Marcos, 2018.
[48] L. F. Pari Aguilar, "El Desempeño docente y la satisfacción de los estudiantes del I.E.S.P.P. José Jiménez Borja de Tacna – 2017," Universidad Nacional de Educación Enrique Guzmán y Valle. Escuela de Posgrado., 2018.
[49] V. D. Melendres Carlos, "El grado de doctor como indicador de calidad académica en las instituciones de educación superior de México," *Trends challenges High. Educ. Lat. Am. Trends challenges High. Educ. Lat. Am.*,

pp. 11–19, 2018.
[50] M. M. T. Ishikawa Saldaña, "Influencia del clima organizacional en el desempeño de los docentes de la Facultad de Ciencias Económicas de la Universidad Nacional de Tumbes, año 2017," Universidad Nacional de Tumbes, 2018.
[51] R. Y. Velasquez Vargas, "Las estrategias didácticas y satisfacción académica de los estudiantes del primer ciclo de estudios de la Facultad de Pedagogía y Cultura Física de la Universidad Nacional de Educación Enrique Guzmán y Valle, 2017," Universidad Nacional de Educación Enrique Guzmán y Valle. Escuela de Posgrado., 2018.
[52] C. E. Martínez Campaña, "Modelo de evaluación para valorar el desempeño docente como factor de calidad en la Universidad Regional Autónoma de los Andes," Universidad Nacional de Trujillo, 2018.
[53] H. Leveau Bartra, J. Chávez Navarro, and L. Geng Olaechea, "Nivel educativo del docente y su influencia en la calidad de la enseñanza universitaria Universidad Alas Peruanas – 2017," *Rev médica panacea*, vol. 7, no. 1, pp. 28–33, 2018.
[54] B. R. Miguel Zisebuto, "Proceso de evaluación docente universitario y la calidad educativa en el Perú, año 2016," Universidad Peruana Los Andes, 2018.
[55] J. M. Miranda Tejada, "Satisfacción laboral en docentes de la Facultad de Ciencias Empresariales de una universidad privada del distrito de La Molina," Universidad San Ignacio de Loyola, 2018.
[56] A. Arce Cobeñas and V. Roque Lopez, "Análisis Comparativo de la percepción de la calidad de servicios académicos en los estudiantes de las carreras profesionales de la escuela profesional de administración de la Universidad Peruana Unión filial Tarapoto -2018," Universidad Peruana Unión, 2018.
[57] M. M. Medina *et al.*, "Análisis de la calidad de la formación académica de los egresados 2011-2015 de la Facultad de Odontología UNLP," *Rev. Fac. Odontol. Univ. Nac. (Cordoba).*, vol. año 2018, pp. 12–16, 2018.
[58] L. P. Uribe Rico, "Aprendizaje basado en problemas: un método que contribuye en la calidad de la educación superior en Colombia," Universidad Militar NUEVA GRANDADA, 2018.
[59] M. E. Chenet Zuta, "Calidad académica según los estudiantes del último semestre de la Universidad Privada de Huancayo Fanklin Roosevelt; en el año 2017," Universidad Peruana Los Andes, 2018.
[60] K. C. Florian Cardenas and A. K. Soto Dueñas, "Calidad de servicio y la satisfacción del estudiante en la Facultad de Ciencias Empresariales de la Universidad Peruana Unión, Lima-2018," Universidad Peruana Unión, 2018.
[61] M. Herrera, M. I. Lund, S. Ruiz, and M. G. Romagnano, "Calidad universitaria mediante técnicas del data mining," in *Workshop Argentino de Ciencias de la Computación*, 2018.
[62] E. M. Guerrero López, "El Kaizen como proceso de mejora continua, en el aseguramiento de la calidad de las instituciones educativas superiores del Ecuador, periodo 2015-2016," Universidad Nacional Mayor de San Marcos, 2018.
[63] K. X. Becerra Cabrejos and M. Raffo Ticlavilca, "Estudio de la calidad del servicio académico y el nivel de satisfacción en los estudiantes de la facultad de derecho de la Universidad Católica de Trujillo Benedicto XVI - 2018 (Tesis parcial)," Universidad Privada del Norte, 2018.
[64] Ó. G. Espinoza Díaz, L. E. González Fiegehen, J. I. Loyola Campos, Ó. G. Espinoza Díaz, L. E. González Fiegehen, and J. I. Loyola Campos, "Evaluación de la satisfacción de titulados de la carrera de psicología en Chile," *Innovación Educ. (México, DF) ISSN 1665-2673*, vol. 18, no. 76, pp. 171–192, 2018.
[65] B. D. Herrera Vilca, "Gestión Pedagógica Y La Calidad Académica En La Facultad De Ingeniería Económica De La Universidad Nacional Del Altiplano Puno - 2017," Universidad Cesar Vallejo, 2018.
[66] L. M. Del Águila Echevarría, "Grado de satisfacción que perciben los estudiantes de pregrado de la Escuela de Sicología de la Universidad Científica del Perú, en relación a la calidad educativa, Iquitos-2018," Universidad Científica del Perú, 2018.
[67] E. Muñoz Miranda and R. Leoncio Ccorpa, "La eficiencia en la gestión del servicio educativo y su influencia en la satisfacción del usuario en el instituto superior tecnológico sise, sede Miraflores, Lima, 2017," Universidad Inca Garcilaso de la Vega, 2018.
[68] L. Manning Bula, "La formación técnico laboral en Colombia, evaluación del impacto de sus egresados desde la empleabilidad. El caso de Montería-Córdoba (1996 – 2015)," in *Congreso Internacional de Educación Evaluación: 2018-2019-ISSN: 2448-6574*, 2018, pp. 765–775.
[69] L. M. Moreno Rodríguez, "Percepciones sobre el Sistema de Evaluación Docente de la Universidad Nacional Abierta y a Distancia UNAD," Bogotá : Universidad Externado de Colombia, 2018, 2018.
[70] J. A. Meneses Velasquez and L. D. Suarez Castro, "Propuesta de mejoramiento para los procesos organizacionales en el Colegio Agustiniano Suba enfocado a la calidad," Universitaria Agustiniana, 2018.
[71] A. Arce Cobeñas, "Relación entre la calidad de servicios académicos y la satisfacción en los estudiantes de la Universidad Peruana Unión filial Tarapoto," Universidad Peruana Unión, 2018.
[72] C. A. Suasti Lopez, "Satisfacción de los estudiantes de la enseñanza superior con las clases virtuales: un estudio en la Universidad Técnica de Manabí," un estudio en la Universidad Técnica de Manabí, 2018.
[73] J. Sánchez Quintero, "Satisfacción estudiantil en educación superior: validez de su medición," *reponame Repos. Inst. Sergio Arboleda*, Jan. 2018.
[74] E. G. Surdez Pérez, M. del C. Sandoval Caraveo, and C. L. Lamoyi, "Satisfacción estudiantil en la valoración de la calidad educativa universitaria," *Educ. y Educ.*, vol. 21, no. 1, pp. 9–26, May 2018.
[75] J. C. (Editor compilador) Tovar-Galvez, Ed., *Trends and challenges in Higher Education in Latin America* -, First Edit. mexico: Adaya Press, 2018.
[76] M. Hornos Javierre, "Una universidad transformadora: análisis crítico de prácticas educativas de la Universidad Autónoma Latinoamericana en Medellín (Colombia).," Universidad Politécnica de Valencia, 2018.
[77] F. R. Polanco Chambi, "Percepciones, expectativas y actitudes hacia la mejora de la calidad en educación a distancia universitaria de pregrado OPE Arequipa 2016," 2018.
[78] D. C. Fernández Lorenzo, Angie Gómez, Jumbo Gómez, Edison Guido Reinoso Guacapiña, S. I. Calvopiña

Espinoza, and D. Regueira Martínez, "Análisis de satisfacción estudiantil en ingeniería comercial de la Universidad de las Fuerzas Armadas ESPE," *Educ. Física y Deport.*, vol. 22, no. 237, pp. 29–33, Feb. 2018.

[79] D. A. García Arango, G. Luján Rodríguez, M. G. E. Zemel, and D. E. Gallego-Quiceno, "Calidad en instituciones de educación superior: Análisis comparativo entre modelos," *Rev. Venez. Gerenc.*, vol. 23, no. 1, pp. 200–217, Jul. 2018.

[80] N. A. COMPILADOR González Rodríguez, Ed., "Desafios de la Educación Superior para la gestión de la calidad," in *Ill CONGRESO CIENTÍFICO INTERNACIONAL "Sociedad del conocimiento: Retos y Perspectivas" ISBN: 978-9942-960-39-9*, 2018, p. 130.

[81] W. T. Orellana Mendoza, "El modelo de excelencia en la EFQM y su aplicabilidad a la Facultad de Ingeniería Eléctrica y Electrónica de la Universidad Nacional del Centro del Perú – 2017," Universidad Nacional Federico Villarreal, 2018.

[82] E. M. Sánchez Durand, "El servicio educativo, la percepción de la calidad y el nivel de satisfacción en los estudiantes de la Universidad Nacional de Educación Enrique Guzmán y Valle en el periodo 2017," Universidad Nacional de Educación Enrique Guzmán y Valle. Escuela de Posgrado., 2018.

[83] J. C. Oliveros Vélez, "Ìndice de cumplimiento de los objetivos misionales establecidos por la Universidad Nacional de Colombia Sede Medellìn, con base en la percepción de los estudiantes," Universidad Nacional de Colombia - Sede Medellín, 2018.

[84] L. B. Lecca Cruzado, "Innovación y gestión de la calidad en la Universidad Nacional Tecnológica Lima Sur/Villa El Salvador - 2017," Universidad Nacional de Educación Enrique Guzmán y Valle. Escuela de Posgrado., 2018.

[85] J. S. González, J. L. Abreu, and M. de J. Araiza Vázquez, "ISO 9001: 2015 en universidades públicas en México; La calidad de información y el compromiso de la dirección," *Daena Int. J. Good Conscienc.*, vol. 13, no. 1, pp. 49–69, 2018.

[86] M. J. Lemaitre, R. Aguilera, A. Dibbern, C. Hayt, A. Muga, and J. Téllez, "La educación superior como parte del sistema educativo de América Latina y el Caribe. Calidad y aseguramiento de la calidad," in *CRES 2018 - Conferencia Regional de Educación Superior*, 2018, p. 200.

[87] S. Gonzales Bellota, "La Gestión Administrativa y El Desempeño Docente En La Universidad Nacional Diego Quispe Tito Filial Calca.," Universidad Cesar Vallejo, 2018.

[88] E. M. Santelices Rivera, "La gestión de la información y su incidencia en la calidad Integral de la comunidad universitaria," Universidad de Guayaquil, Facultad de Comunicación Social, 2018.

[89] D. Diaz Silva, "La gestión de la calidad universitaria y su relación con la mejora de la cultura organizacional de los docentes de la Facultad de Derecho de la Universidad Privada Telesup, 2018," Universidad Nacional de Educación Enrique Guzmán y Valle. Escuela de Posgrado., 2018.

[90] M. C. Nosiglia, "La reforma universitaria de 1918 en debate: vigencia de sus principales postulados," *Rev. Latinoam. Educ. Comp. ISSN 1853-3744*, vol. 9, no. 13, pp. 93–104, 2018.

[91] C. M. Casierra Párraga, A. Muñoz Rodríguez, G. S. Venegas Álvarez, U. G. Farfán Corrales, and F. Á. Lemoine Quintero, "NIVEL DE SATISFACCIÓN DE EGRESADOS DE LA CARRERA DE INGENIERÍA EN MARKETING DE LA UNIVERSIDAD 'ELOY ALFARO' DE MANABÍ CAMPUS BAHÍA DE CARÁQUEZ," *Rev. Didasc@lia Didáctica y Educ. ISSN 2224-2643*, vol. 9, no. 4, pp. 23–38, Oct. 2018.

[92] J. E. Valero Cajahuanca, "Propuesta de un modelo de gobierno de tecnologías de información y su influencia en la calidad educativa en la carrera de ingeniería de sistemas de la Universidad Nacional de Huancavelica," Universidad Nacional Federico Villarreal, 2018.

[93] F. C. Arcos Ariza, "Satisfacción estudiantil en las Instituciones Educativas del distrito de Sapallanga, Huancayo 2018," Universidad César Vallejo, 2018.

[94] R. M. Jaquinet-Espinosa, D. Nogueira-Rivera, R. A. Frías-Jiménez, R. M. Jaquinet-Espinosa, D. Nogueira-Rivera, and R. A. Frías-Jiménez, "Variables organizacionales en la Facultad de Ciencias Económicas e Informática, Universidad de Matanzas," *Ing. Ind.*, vol. 39, no. 2, pp. 184–195, 2018.

[95] L. M. Dicovskiy Riobóo and M. E. Pedroza, "Minería de datos, una innovación de los métodos cuantitativos de investigación, en la medición del rendimiento académico universitario," *Rev. Científica FAREM-Estelí*, vol. 0, no. 24, pp. 143–152, 2018.

[96] L. A. Jurado Viscarra and I. A. T. Irazoque Tobías, "Relación estadística de la ejecución curricular y uso de tecnologías de información y comunicación con el rendimiento académico de los estudiantes de la Carrera de Ingeniería Eléctrica de la UMSA," Universidad Mayor de San Andrés, Bolivia, 2018.

[97] E. Loyola Illescas *et al.*, *Repitencia estudiantil en la Universidad Politécnica Salesiana*, 1ra ed. E: Universidad Politécnica Salesiana, 2018.

[98] M. A. Muñoz Araya, "CALIDAD DE LOS RANKINGS DE UNIVERSIDADES CHILENAS, UNA NUEVA PROPUESTA METODOLÓGICA. Evaluación entre los años 2012 a 2016," Universidad Diego Portales, Santiago de Chile, 2018.

[99] T. Blanco Sánchez, "Las universidades españolas en los rankings internacionales. Logros, carencias y desafíos," *Innovación Educ.*, vol. 0, no. 28, p. 171, Dec. 2018.

[100] A. P. López and A. P. Martínez, "Rankings universitarios. Evitando perseguir fuegos fatuos," *Estrateg. y Gestión Univ.*, vol. 6, no. 2, pp. 112–125, Dec. 2018.

[101] E. M. Carbajal, "Rankings universitarios internacionales," *Rev. Identidad Univ.*, vol. 1, no. 3, pp. 23–25, Dec. 2018.

[102] T. Rosales Cifuentes, "Modelo de Tutoría y Mejora de la Calidad Educativa de los Estudiantes en la Universidad Santo Domingo De Guzmán," *Cátedra Univ. ISSN. 2663-8401* , vol. 2, no. 1, p. 30, 2018.

[103] R. Rodríguez, "Propuesta de modelo educativo universitario basado en competencias para la Universidad Nacional de Trujillo, 2016," *Rev. Cienc. Y Tecnol.*, vol. 14, no. 3, pp. 107–124, Oct. 2018.

[104] M. J. Cabrera Ulloa, "Propuesta de un modelo de gestión de calidad en los servicios académicos, del área administrativa en la Universidad Técnica Particular de Loja, modalidad presencial, año 2017.," Universidad Tecnica particular de La Loja, Ecuador, 2018.

[105] A. I. Jaya Escobar, E. A. Olivo Fernández, and R. M. Guerra Bretaña, "Propuesta de modelo de gestión de la calidad para la Universidad Estatal de Bolívar," *Rev. Caribeña Ciencias Soc. ISSN 2254-7630*, pp. 1–19, 2018.

API REST extensible para Robots no Industriales

César Omar Aranda
IISCOT/ FCSyA, Universidad del Aconcagua, Catamarca 147, Mendoza, Argentina.
cesar.aranda@uda.edu.ar

Resumen. Este trabajo es el resultado de un proyecto de tecnologías aplicadas, destinado a modelar la base de una plataforma capaz de ofrecer servicios para monitoreo y control remotos de uno o más robots independientes ubicados en un ámbito no industrial. Se realizan descripciones del robot antropomórfico elegido y se detallan los modelos, las definiciones y los servicios mínimos de la API REST. Se describe el prototipo del ecosistema donde coexisten los robots, los servidores, los usuarios y una aplicación cliente capaz de acceder a los recursos expuestos por la API. Se mencionan las principales alternativas de extensión que, con muy bajo esfuerzo, se pueden efectuar sobre la misma.

Palabras Clave: API REST, Robot antropomórfico, Plataforma, Monitoreo, Control, Sistemas embebidos.

1 Introducción

Dentro de las tecnologías, hasta hace poco denominadas emergentes, que empiezan a entrar la fase de madurez de sus ciclos de vida, relacionadas al uso domiciliario masivo, aparecen la robótica y la incorporación de API REST como tecnología para impulsar una idea o negocio.

Según se observa en la feria de tecnología CES 2018 [1], en Las Vegas, importantes empresas internacionales como Honda, LG y Ubtech, entre otras, se orientan a ofrecer robots de uso hogareño, siguiendo una tendencia iniciada en ediciones de años anteriores (CES 2017 [2], en Shangai o IFA 2016 [3], en Berlín).

Los robots ofrecidos, presentan un conjunto de servicios y funcionalidades, acompañados de una o más aplicaciones de configuración. En general, las posibilidades se encuentran limitadas a lo que cada fabricante incluya en sus programas.

Sin embargo, por propia definición, los robots podrían ser reprogramados y/o ampliados para que fueran capaces de ofrecer nuevas características.

La limitación principal está dada por la complejidad del robot y las dificultades del usuario de poder realizar la programación requerida.

La robótica es una más de estas tecnologías, donde la programación específica se encuentra todavía en un nivel complejo para la mayoría de los usuarios y en particular el desarrollo de aplicaciones web en este ámbito, aún es muy heterogéneo.

Por esta razón, se tratará de ocultar en la aplicación que provee la API o en el software de control, la complejidad inherente a la ingeniería del robot, como su cinemática directa o inversa, las consideraciones dinámicas propias de los movimientos y colisiones, las estrategias y mecanismos de control usados.

El enfoque principal será el de ofrecer, a los desarrolladores y eventuales usuarios, una capa abstracta que disminuya la distancia conceptual del trabajo con robots y facilite el desarrollo de aplicaciones Web que interactúen con robots de cierto tipo, usando estándares actuales.

Brevemente, se describen a continuación los elementos conceptuales fundamentales: robot, control y API REST.

1.1 Robot

Un robot es una máquina, "es un manipulador multifuncional reprogramable diseñado para mover materiales, partes, herramientas o dispositivos especializados a través de movimientos programados para la ejecución de una variedad de tareas" [4], que interacciona con su entorno, con un grado variable de autonomía en función de la tarea a realizar [5].

Existe una amplia variedad de diseños orientados a muy diversas tareas.

Los robots tienen diversas especificaciones y no pueden realizar más que el trabajo para el que fueron diseñados. Por ejemplo, un robot creado para ensamblar máquinas, no podrá ser adaptado para otras funciones.

La forma de organizar la inmensa cantidad de robots que existe puede ser utilizando su sistema de funcionamiento, como robots estacionarios (brazos mecánicos), robots cilíndricos, robots esféricos, robots paralelos, robots con ruedas, robots bípedos (de forma humanoide), robots nadadores, robots voladores, robots esféricos y móviles (como bolas robóticas) y enjambres de mini-robots. Otro criterio es por su arquitectura, de donde resulta posible encontrar robots de tipo androide, móvil, poliarticulado, zoomórfico o híbrido. Es habitual clasificarlos en función del área de aplicación, destacándose en este caso los robots de uso industrial, militar, médico, de servicio, de entretenimiento, espacial y doméstico.

1.2 Control

En todo sistema robótico es necesario cierto tipo de control, en general de lazo cerrado, para generar las actuaciones que permitan controlar las partes móviles.

En este proyecto existen 2 contextos de control: uno directo, ligado a la electrónica y a las características implícitas del robot usado, y otro asociado al control del usuario.

En los casos de robots con articulaciones, la forma más simple es controlar cada una de ellas de forma independiente o desacoplada de las restantes. Es habitual también emplear un controlador PID digital en cada articulación. Otras técnicas disponibles son el control basado en un modelo, el control adaptativo y el control con aprendizaje.

Es necesario controlar aspectos cinemáticos de posición, de trayectoria, También los efectos dinámicos producidos por aceleraciones o cargas, y el control de la fuerza [6].

En la necesidad de tener que manipular el robot de manera remota, deben definirse también los mecanismos por los cuales se realizará la comunicación, ya sea para recuperar información de estado como realizar la manipulación.

Si bien existen las posibilidades de comunicar directamente la computadora host a la electrónica de control, o de usar procesos que se ejecuten de manera concurrente con

el servidor Web y se comuniquen con él (por mecanismos RPC u otros), se procede a realizar el diseño asumiendo un escenario donde los dispositivos clientes poseen la capacidad de comunicarse con el robot, a través de internet usando una API REST.

1.3 API REST

REST es el acrónimo de REpresentational State Transfer.

"*REST es un estilo arquitectónico. Es un conjunto de restricciones, a respetar cuando se diseña la arquitectura de un servicio web*" [7].

Las restricciones propuestas por REST son las siguientes:

- REST está orientado a recursos lógicos. No se representan acciones, sino entidades de negocio.
- Cada recurso posee un identificador único universal (UUID o GUID) bajo el cual es posible referenciarlo de manera unívoca.
- La implementación, y la forma de representar internamente un recurso es privada y no accesible desde el exterior.
- Cada recurso tiene una interfaz, o conjunto de operaciones que admite.
- La interfaz es homogénea para todos los recursos. Y cada recurso debe escoger las operaciones que soporta de entre un conjunto cerrado de acciones.
- Las operaciones se realizan mediante la transferencia del estado de un recurso entre cliente y servidor.
- Las operaciones son stateless. Es decir, el resultado de una operación es independiente de la conversación C/S que haya existido anteriormente.
- Los recursos son multimedia. Esto es, flexibilidad para que el estado de un recurso pueda ser representado usando distintos formatos.

Cuando se define una API REST, se describe una capa lógica de aplicación con un conjunto de recursos que se desean exponer y un conjunto de métodos con los cuales acceder a esos recursos usando internet [8].

Podría afirmarse que REST es cualquier interfaz entre sistemas que use HTTP para obtener datos u operar sobre ellos, usando formatos estándar (XML, JSON, otro).

Es habitual definir los servicios con un enfoque orientado a datos, haciendo mención a operaciones CRUD[1]. En este tipo de enfoque existen 2 tipos de recursos: las entidades y las colecciones.

2 Infraestructura del prototipo

Para plantear el modelo básico de este sistema, se usa un robot prototipo, controlado mediante electrónica Arduino compatible y comunicada con un host basado en una SBC.

[1] CRUD es el acrónimo de las operaciones generales para datos: Create, Read, Update, Delete

2.1 Sistema Robot

Considerando el enfoque educativo y de bajo costo del ecosistema a controlar, la disponibilidad de la documentación, los diseños de las piezas 3D que conforman la estructura, el software de control entre otros factores, se selecciona como robot específico, uno de tipo poliarticulado.

Para este tipo, se evaluaron diferentes modelos de brazos robot, de 3 a 6 DoF. Los más significativos han sido Niryo1 (de Niryio Robotics), Thor (de AngelLM), RobotArm (de Florin Tobler), Zortrax (de Zortrax Technologies), LiteArm i2 (de Lance Pierson), Moveo (de BCN3D), V-Arm (de Nathan Barrow), AR2 i2 (de Annin) y EezybotArm MK2 (de Carlo Franciscone). Varios de ellos inspirados en el robot articulado de paletización industrial IRB460 [9] de ABB Robotics.

Luego de comparar diferentes parámetros, se adopta un robot bautizado robotArm, de tamaño y complejidad intermedios, de 4 DoF, con mecanismo de brazos paralelos, configuración RRR y efector final paralelo del tipo pinza.

Las piezas mecánicas del brazo robot se replican mediante impresión 3D, desde archivos de diseño provistos por Tobler [10], anexos realizados por Francis Ng [11] y variantes tanto en la estructura como en una nueva pinza, realizadas por el autor [12].

La electrónica de control se basa en las placas RAMPS y Arduino Mega 2560.

Ninguno de ellos ofrece aplicaciones que permitan controlar el robot usando internet.

2.2 Hardware del Servidor prototipo

Dadas las conclusiones arribadas en otro proyecto de este autor [13], se decide usar una placa SBC Raspbery Pi 3+ como dispositivo Servidor, acompañada de un Router TP-LINK AC750 y una cámara web Genius FaceCam 320X.

2.3 Software del Servidor prototipo

Como sistema operativo se adopta Debian GNU/Linux.
Se instalan y configuran en él las aplicaciones de servidor y de programación elegidas. Las razones de su elección se encuentran en las consideraciones del proyecto ya mencionado [13]:

- Lighttpd como servidor Web.
- Python como lenguaje para el desarrollo de la aplicación del lado servidor que ofrece la API REST.
- Javascript junto a HTML5 y CSS como lenguajes del lado cliente para el desarrollo de las interfaces de usuario.
- JSON como formato para intercambio de datos entre los clientes y el servidor.
- Archivos con formato XML para almacenar cada secuencia de trabajo.
- MySQL como servidor de base de datos.

Para la transferencia de los archivos requeridos en el servidor se usa sftp, provistos por el servidor openssh instalado y un cliente remoto.

Para manipular el servicio de emisión del video proporcionado por la cámara adicional, se usa MJPG_streamer, dados los mejores resultados obtenidos en las pruebas respecto de OpenCV y FSWebcam.

3 Prototipo de la Plataforma SaaS

3.1 Diseño lógico de la API REST

En el caso de esta API, varios servicios no se corresponden directamente con acciones sobre la base de datos, sino a operaciones de control sobre el robot.

En base a ello, los métodos disponibles en una petición HTTP, son usados para:

- GET: recuperar un elemento del sistema de archivos o datos provenientes de una sentencia SQL.
- POST: inserción de datos mediante una sentencia SQL, creación de un elemento en el servidor o ejecución de una operación en particular del robot.
- PUT: no usado.
- PATCH: modificar el estado de atributos de un recurso
- DELETE: eliminar un elemento (recurso) presente en el servidor.

Los recursos identificados del sistema que serán expuestos por la API son: el robot, los archivos de programa, los usuarios, las operaciones que el robot puede realizar y el estado del mismo durante la ejecución de un programa.

Cada URL de este trabajo se estructura de la siguiente forma:

{protocolo}://{hostname}[:puerto]/{v1/ruta_recurso}?{consulta de filtrado}

Los URIs han sido definidos en base a las entidades anteriores. Considerando que la API puede evolucionar, se ha agregado al URL el identificador de la versión: /v1

Para simplificar, en los siguientes ejemplos se indicará el URI del recurso, sin /v1.

Para los recursos expuestos se proponen los métodos detallados en la Tabla 1.

Para las operaciones de control se listan los códigos admitidos en la Tabla 2.

Dado que las operaciones de control del robot (y de otros dispositivos adicionales), producen un cambio de estado, sólo si la acción es viable y se pudo realizar, se adopta vincular el concepto de operación y selección directas al comando POST, en lugar de la recomendación [14][15] de usar PATCH o PUT, con una perspectiva puesta en datos.

Si bien se ha elegido para las pruebas del ecosistema un robot angular, en la API se ha previsto que pueda usarse un robot cuyo eslabón pueda desplazarse en lugar de rotar.

Tabla 1. Descripción de los recursos (URIs) y verbos asociados.

Recurso	GET	POST	DELETE	PATCH
/files3d	Lista los archivos de programa	Crea un archivo de programa	-	-
/files3d/4	Recupera los detalles del archivo 4	Selecciona para ejecutar el archivo con id 4	Elimina el archivo con id 4	-
/login3d	-	Crea (una vez validado) un	-	-

		nuevo token de acceso		
/process3d	Lista las tareas realizadas por el robot (historial)	Crea un nuevo proceso (registro de tarea)	-	-
/process3d /2	Recupera detalles del proceso con id 2	-	Elimina el proceso con id 2	Actualiza el proceso con id 2
/robots3d	Lista los robots disponibles	Agrega un nuevo robot	-	-
/robots3d /1	Recupera detalles del robot con id 1	-	Elimina el robot con id 1	-
/robots3d /1/operations	Lista las operaciones disponibles para el robot con id 1	-	-	-
/robots3d /1/operations/ 4	-	Ejecuta la operación 4 sobre el robot con id 1	-	
/robots3d /1/states	Recupera los detalles del proceso en curso	-	-	-
/users3d	Listado de los usuarios que pueden usar robots	Agrega un nuevo usuario de robots	-	-
/users3d/8	Recupera detalles del usuario con id 8	-	Elimina un usuario con id 8	Actualiza un usuario con id 8

Tabla 2. Códigos de operación del robot y la cámara

Cód.	Descripción de la Operación	Parámetros adicionales
1	Conectar robot	Denominación de la interfaz y velocidad de transmisión (en bps) del port
2	Desconectar robot	-
3	Encender/Activar el robot	-
4	Apagar/Desactivar el robot	-
10	Iniciar el programa/tarea	
11	Suspender el programa	-
12	Reasumir el programa	-
13	Abortar el programa	-
14	Activar el efector final	-
15	Liberar el efector final	-
20	Cambiar la velocidad de desplazamiento	Valor de velocidad (en mm/min) Identificador de la articulación
21	Cambiar la velocidad de rotación	Valor de velocidad (en º/min) Identificador de la articulación
22	Ir al origen de ejes X-Y-Z	-
23	Ir al origen del eje X	-
24	Ir al origen del eje Y	-
25	Ir al origen del eje Z	-

26	Ir al centro del plano X-Y prefijado para el espacio de trabajo	-
27	Mover la articulación lineal (movimiento relativo)	Valor de desplazamiento (en mm con signo) Identificador del eje (X | Y | Z)
28	Mover la articulación rotacional (movimiento relativo)	Valor de rotación (º sex. con signo) Identificador del eje de rotación (X | Y | Z)
29	Mover la articulación a una posición (movimiento absoluto)	Coordenadas X, Y, Z (en mm)
40	Conectar a servidor de stream de video	-
41	Desconectar de servidor de stream de video	-

En las respuestas dadas por un servidor, se han considerado 6 estados de error, cumpliendo con la recomendación de no definir más de 10 estados posibles en la API (códigos 200, 400, 401, 404, 405 y 500). Cada recurso tiene asociados un estado de respuesta satisfactoria (200) y una cantidad variables de estados de error.

Las respuestas a cualquier requerimiento HTTP presentan 2 partes: el header y el body. El header posee los campos: 'protocol_version', 'status_code' y 'status_message' (breve cadena descriptiva del resultado). El body posee 3 campos: 'error_code', 'error_message' y 'data'. El contenido de cada campo varía en función del requerimiento y del resultado del mismo.

3.2 Diseño de la API RESTful

Para el diseño de la plataforma, se ha aprovechado una propuesta anterior [16].

Esquema de capas del software

En la Figura 1 se observa que la aplicación proveedora de la API REST debe interactuar, por un lado con la aplicación del cliente web, y por el otro debe ser capaz de acceder al firmware de control del robot. La creación y exposición de recursos mediante la API le permite a una aplicación web interactuar con otras aplicaciones usando comunicaciones M2M.

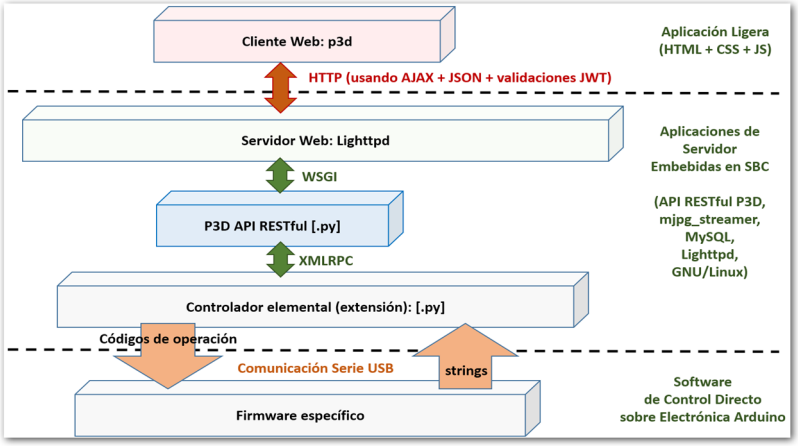


Figura 1: Capas y comunicaciones del prototipo (Fuente propia)

Modelos fundamentales

En las Figuras 2 y 3 se observan las entidades que conforman la API RESTful y una secuencia de trabajo ante un requerimiento HTTP de un recurso dado.

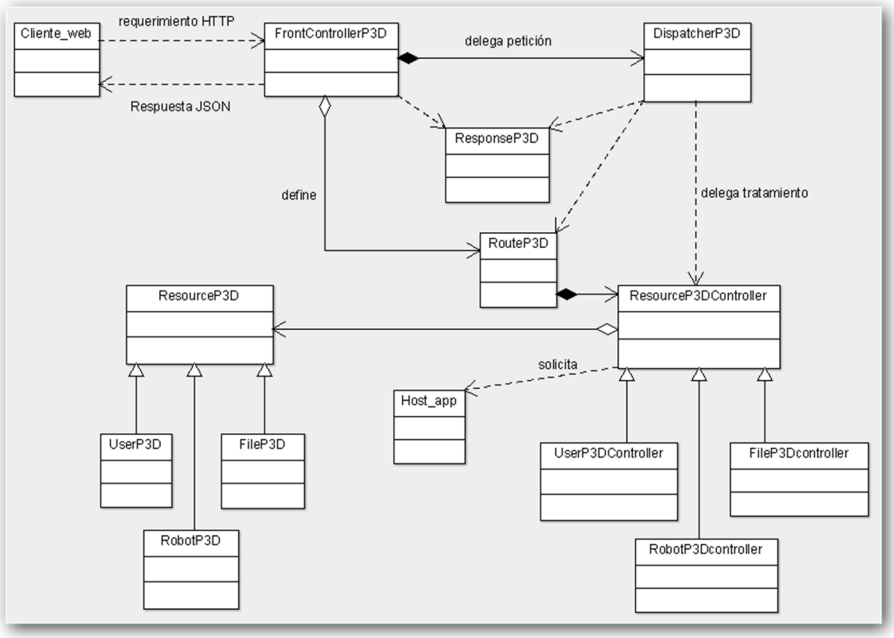


Figura 2: Diagrama de Clases de la API. (Fuente propia)

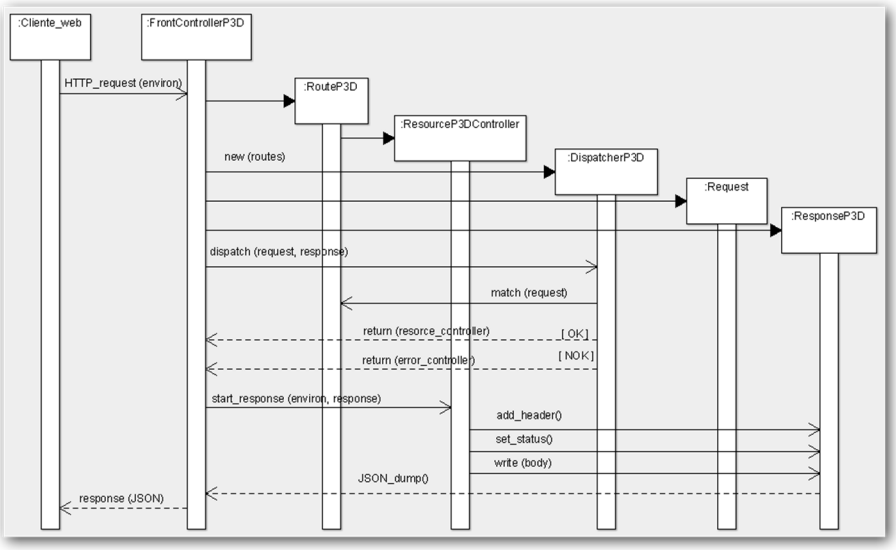


Figura 3: Diagrama de Secuencia para un requerimiento HTTP. (Fuente propia)

Diseño de la aplicación web cliente

Para las tareas de monitoreo y control remoto, se optó por una interfaz de usuario web para ser usada en un dispositivo móvil, desarrollada en lenguaje javascript con despliegue gráfico basado en HTML5 y CSS3.

Se trata de una estructura de página única, segmentada mediante elementos de capa en 3 bloques de interacción, denominados: Control, Programas e Información.

La interfaz de Control, permite seleccionar al robot remoto, establecer los parámetros del puerto de conexión del host al robot, controlar el robot y observar datos de estado (conexión, archivo y ejecución). Los controles permiten tareas de encendido y apagado, movimientos de cada articulación y operaciones del efector final entre otros. Selección del modo off-line y on-line que permite programar la tarea o ejecutar ordenes, respectivamente. También activar o desactivar la visualización remota del robot.

La interfaz de Programas, permite una gestión básica de los archivos de programa con las instrucciones para realizar cierta tarea, con acciones para ver detalles, listar o buscar bajo diferentes criterios, agregar o eliminar un archivo, o seleccionar un archivo para su carga en el robot.

La interfaz del menú Información, muestra detalles particulares acerca del aplicativo web descripto.

El acceso a los menús anteriores se realiza luego de una primera fase de validación usando usuario, clave y un mecanismo con JWT (basado en el estándar RFC 7519).

4 Conclusiones

Durante este trabajo, se instaló y configuró el software capaz de atender los requerimientos HTTP realizados por los clientes, se implementó la plataforma de control prototipo que ofrece los servicios a través de la API REST, y se elaboró un aplicativo cliente web básico para las pruebas, también se hizo la elección y ajuste de una aplicación de terceros ligada al control directo del robot conectado a la plataforma.

Se realizaron, en ámbito de laboratorio, pruebas primarias tanto de operación unitaria como integrada de los servicios, que muestran la capacidad del ecosistema para controlar y monitorear de manera remota la actividad de un robot, mediante el uso de redes de datos estándares.

Como resultados del diseño se han obtenido: el modelo de datos, las clases fundamentales de la API RESTfull, la definición de cada servicio de la API con ejemplos de cada petición/respuesta en los casos favorables.

Dado que el objetivo fundamental era el diseño de la API, algunos servicios de la misma fueron planteados como prototipos no funcionales.

Se han documentado los modelos y las consideraciones principales, a los efectos que el proceso sea replicable y las definiciones puedan extenderse según corresponda.

Si bien se han enfocado operaciones de cinemática directa y modo de trabajo desacoplada del robot, resulta conveniente avanzar sobre aspectos de control basados en la cinemática inversa.

Referencias

1. CES 2018, Las Vegas, https://www.clarin.com/sociedad/robots-uso-hogareno-protagonistas-mayor-feria-tecnologia-mundo_0_rkBv21l4G.html
2. CES 2017, Shangai, https://elpais.com/tecnologia/2017/06/07/actualidad/1496845192_806357.html
3. IFA 2016, Berlin, https://www.infobae.com/america/tecno/2017/09/01/el-rincon-de-la-locura-en-la-feria-ifa/
4. Reyes Cortés, F: Robótica, control de robots manipuladores. Alfaomega Grpo Editor, México (2011).
5. Ollero Baturone, A.: Robótica: manipuladores y robots móviles. Editorial Marcombo. Barcelona (2001).
6. Craig, J.: Robótica. 3º ed. Pearson Prentice-Hall. México (2006).
7. Amodeo, E.: Principios de diseño de APIs REST. Learn Publishing (2013).
8. Richardson, L. y Amundsen, M.: RESTful Web APIs. O'Reilly Media Publishing. California, USA (2013).
9. IRB 460, Paletizador robótico de alta velocidad. Obtenido en 2018 desde https://new.abb.com/products/robotics/es/robots-industriales/irb-460
10. RobotArm. Obtenido en 2018 desde https://www.thingiverse.com/thing:1718984
11. RobotArm by FrN. Obtenido en 2018 desde https://www.thingiverse.com/thing:2040526
12. RobotArm Extended. Obtenido en 2018 desde https://www.thingiverse.com/thing:3674358
13. Aranda, C.: Instalación y Configuración de Sistemas Embebidos en una SBC Intel Galileo 2 como Base de Plataforma IOT. Anales CICCSI 2018, 1ª edición, pp. 195--204. ISBN 978-987-45683-6-6. FUSMA ediciones. Universidad Champagnat. Mendoza (2019).
14. Microsoft Azure: Diseño de API Web. Obtenido en 2018 desde https://docs.microsoft.com/es-es/azure/architecture/best-practices/api-design
15. Internet Engineering Task Force: PATCH Method for HTTP (RFC 5789). Obtenido en 2018 desde https://tools.ietf.org/html/rfc5789
16. Aranda, C.: Ecosistema de IoT formado por dispositivos M2M de control y SaaS. Anales CICCSI 2018, 1ª edición, pp. 175--184. ISBN 978-987-45683-6-6. FUSMA ediciones. Universidad Champagnat. Mendoza (2019).

Performance evaluation of data base for big data in cloud

María Murazzo[1], Nelson Rodríguez[1], Pablo Gómez[2]

[1] Departamento e Instituto de Informática - F.C.E.F. y N. - U.N.S.J.
[2] Alumno Avanzado Licenciatura en Cs. de la Computación - F.C.E.F. y N. - U.N.S.J.

marite@unsj-cuim.edu.ar, nelson@iinfo.unsj.edu.ar, pablo.gomez.allende@gmail.com

Abstract. For very years, relational databases have been the leading model for data storage, retrieval and management. However, due to increasing needs for scalability and performance, alternative systems have emerged, namely NewSQL technology. NewSQL is a class of modern relational database management systems (RDBMS) that provide the same scalable performance of NoSQL systems for online transaction processing (OLTP) read-write workloads, while still maintaining the ACID guarantees of a traditional database system. In this research paper, the performance of a NewSQL database is evaluated, compared to a MySQL database, both running in the cloud, in order to measure the response time against different configurations of workloads.

Keywords: Big Data, Cloud Computing, Spanner, Google Cloud Platform

1 Introduction

At present databases are not only expected to be flexible enough to handle different variety of data formats, they are expected to deliver extreme performance as well as easily scale to handle big data [1].

Last two years 90 percent of the data in the world was generated. This growth of storage capacity, leading to the emergence of data management systems where data is stored in a distributed way, but accessed as well as analyzed as if it resides on a single machine. Different types of data as well as continuity in data availability has become more important than ever and expects data to be available 24×7 and from everywhere.

Structured Query Language (SQL) [2] became the standard of data processing because it contains such as data definition, data manipulation and data querying, all under one umbrella. RDBMS have always been distinguished by the ACID (Atomicity, Consistency, Isolation, Durability) principle set that ensures that data integrity is preserved at all costs. RDBMS (Relational database management systems) can guarantee performance on the order of thousands of transactions per second, but in this time online transaction processing (OLTP) [3] in stages such as games, advertising, fraud detection and risk analysis involves more than million transactions per second that traditional RDBMS cannot easily handle. High availability without any point of failure as well as durability such as challenges have created a new wave processing database solutions, that manages data in structured unstructured ways.

New type of data management solutions are emerging to handle distributed content on open platforms. Unstructured data, non-relational databases, distributed architectures and big data analysis can change the way in which data is stored and analyzed, with the aim of obtaining useful information for making decisions in real time.

The objective of this paper is to address the problem of storage, recovery and analysis of large volumes of data using non-traditional databases.

The rest of this paper is organized as follows. Section 2 explains the generalities of big data, as well as the challenges that must be faced; Section 3 explains the relationship between cloud computing and big data; Section 4 addresses the generalities of the database administrators and in particular NewSQL solutions. In the following section the working environment and framework used is detailed. Finally, the developed tests, conclusions and future works are explained.

2 Big data characterization

Under the explosive increase of global data, the term of big data is mainly used to describe enormous datasets [4]. Big data is a term utilized refer to the increase in the volume of data that are difficult to store, process, and analyze through traditional database technologies. Compared with traditional datasets, big data often includes masses of unstructured data that need more real-time analysis. In addition, big data also brings new opportunities for discovering new values, helps us to gain an in depth understanding of the hidden values, and also incurs new challenges, ex. how to organize and manage such datasets [5] .

Nowadays, big data related to the service of Internet companies grow rapidly. While the amount of large datasets is drastically rising, it also brings many challenging problems demanding prompt solutions:

- The latest advances of information technology (IT) make it more easily to generate data. Therefore, we are confronted with the main challenge of collecting and integrating massive data from widely distributed data sources.
- The rapid growth of cloud computing and the Internet of Things (IoT) further promote the sharp growth of data.
- In consideration of the heterogeneity, scalability, real-time, complexity, and privacy of big data, we shall the datasets at different levels during the analysis, modeling, visualization, and forecasting, so as to reveal its intrinsic property and improve the decisionmaking.

According to the analyzed, big data is an abstract concept. In general, big data shall mean the datasets that could not be perceived, acquired, managed, and processed by traditional IT and software/hardware tools within a tolerable time. Due to different concerns, scientific and technological enterprises, research scholars, data analysts, and technical practitioners have different definitions of big data. So, there have been considerable discussions between industry and academia about the definition of big data. In addition to developing a proper definition, the big data research should also focus on how to extract its value, how to use data, and how to transform “a bunch of data” into “big data” [6]. In that sense, NIST defines big data [7] as *“Big data shall mean the data of which the data volume, acquisition speed, or data representation limits the capacity*

of using traditional relational methods to conduct effective analysis or the data which may be effectively processed with important horizontal zoom technologies", which focuses on the technological aspect of big data. It indicates that efficient methods or technologies need to be developed and used to analyze and process big data.

3 Big data and cloud computing

In order to gain a deep understanding of big data, is necessary introduce a fundamental technology that are closely related to big data: cloud computing. There is no a consensual definition of Cloud Computing yet. One of the most cited definition is the NIST's [8], where is defined as being *"a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction".*

Cloud computing is a technology to perform massive-scale and complex computing. It eliminates the need to maintain expensive computing hardware, dedicated space, and software. Big data is the object of the intensive computation operation and stresses the storage capacity of a cloud system. The main objective of cloud computing is to use huge computing and storage resources under concentrated management, so as to provide big data applications with fine grained computing capacity [9]. The distributed storage technology based on cloud computing can effectively manage big data; the parallel computing capacity by virtue of cloud computing can improve the efficiency of acquisition and analyzing big data.

Cloud computing and big data are conjoined. Big data provides to users the ability to use commodity computing to process distributed queries across multiple datasets and return resultant sets in a timely manner. Cloud computing provides the underlying engine through the use of distributed data-processing platforms. Thanks to cloud, big data are stored in a distributed fault-tolerant database and processed through a programming model for large datasets with a parallel algorithm distributed in a cluster.

4 Big Data database

One of the challenges that confront organizations dealing with Big Data is how and where to store the tremendous amount of data. In this context, the most widespread data management technology is relational database management systems (RDBMS). With advent of Big Data however, the structured approach falls short to serve the needs of Big Data systems which are primarily unstructured in nature. Increasing capacity of SQL although allows huge amount of data to be managed, it does not really count as a solution to Big Data needs, which expects fast reply and quick scalability [10] [11].

This is where NewSQL [12] is a latest development in the world of database systems. NewSQL is a Relational Database with the scalability properties of NoSQL[13]. You can define NewSQL as a next generation scalable relational database management systems (RDBMS) for Online Transaction Processing (OLTP) that provide scalable performance of NoSQL systems for reading writing workloads, as well as maintaining the ACID guarantees of a traditional database system.

4.1 Architecture NewSQL

Traditional databases cannot deliver capacity on demand that application development might be hindered by all the work required to make the database scale. To overcome scalability challenges, developers add scaling like separating, sharding and clustering techniques. Another common approach is to add larger machines at more cost. An ideal DBMS should scale elastically, allowing new machines to be introduced to a running database and become effective immediately. Therefore, To adopt scale-out performance, DBMS that has been re-defined relational database technology and implement web-scale distributed database technology to tackle the multiple challenges associated with cloud computing and the rise of global application deployments [14].

Important Characteristics of NewSQL Solutions:

- NewSQL provides feature SQL as the primary mechanism for application interaction.
- NewSQL support ACID properties for transactions.
- NewSQL controls a non-locking concurrency control mechanism which is helpful for the real-time reads will not conflict with writes.
- NewSQL (dbShards) architecture providing much higher per-node performance than available fromtraditional RDBMS solutions.
- NewSQL support a scale-out, parallel, sharednothing architecture, capable of running on a large number of nodes without suffering bottlenecks.
- NewSQL systems are 50 times faster than traditional OLTP RDBMS.

NewSQL databases provide an SQL query interface, and clients (users and applications) interact with them the same way they interact with relational databases. They manage read/write conflicts using non-lock concurrency control.

4.2 Selected tool

There are many NewSQL databases being used in big data era, for real time web and big data applications. Every database is having a particular data format for storing its data. It has been decided to work with cloud computing as support architecture, more specifically Google Cloud Platform and Spanner as a tool.

Google Cloud Platform (GCP) [15] is a suite of cloud computing services that runs on the same infrastructure that Google contained in Google's data centers. The Storage services which provides, include: Cloud SQL (MySQL or PostgreSQL databases), Cloud Spanner and two options for NoSQL data storage (Cloud Datastore and Cloud Bigtable).

Cloud Spanner [16] [17] is a scalable, globally distributed database designed, built, and deployed at Google. At the highest level of abstraction, it is a database that shards data across many sets of Paxos state machines in datacenters spread all over the world. Replication is used for global availability and geographic locality; clients automatically failover between replicas. Spanner automatically reshards data across machines as the amount of data or the number of servers changes, and it automatically migrates data across machines (even across datacenters) to balance load and in response to failures.

5 Proposed work

In the present work, the performance comparison between a MySQL database and the Spanner database was performed, both working on the Google Cloud Platform.

Due to the need to work with large amounts of data, public data sets are used, in this case the data sets provided by the government of the Autonomous City of Buenos Aires are used. The first data set used contains data from the "Sistema Único de Atención Ciudadana" (SUACI), which is responsible for addressing the needs and claims of the residents of the CABA. This dataset is organized by year. For this work, only the year 2018 has been taken, which has 895,000 records with the following fields: contacto, periodo, categoria, subcategoria, concepto, tipo_prestacion, fecha_ingreso, hora_ingreso, domicilio_cgpc, domicilio_barrio, domicilio_calle, domicilio_altura.

The second dataset used shows a register of the streets of the CABA, where the name, meaning and code of the streets are included among other data.

To perform the performance evaluation, three queries were defined:

- **Query 1:** select count(*) from SUACI
- **Query 2:** select categoria, subcategoria, concepto, fecha_ingreso, fecha_cierre_contacto from SUACI
- **Query 3:** select RECLAMO.categoria, RECLAMO.concepto, RECLAMO.fecha_ingreso, RECLAMO.domicilio_calle as nombre_calle, CALLE.codigo as codigo_calle from SUACI RECLAMO, CALLEJERO CALLE WHERE RECLAMO.domicilio_calle = CALLE.nomoficial limit 11000;

The three queries were executed with the same syntax in both databases because the implementation of the SQL language in both databases is similar, however it is necessary to emphasize that the data types used by MySQL are different from those uses Spanner, so the migration of data between both databases is not direct and may require adjustments.

In order to start working with Spanner (as is the case with Cloud SQL) it is necessary to create an instance, which functions as a container for the databases that we are going to create.

The previously defined queries will be executed on three work scenarios: 2000, 10000 and 20000 records. This will allow to evaluate the behavior performance of MySQL and Spanner in front of the increase of workload.

To execute the queries to the MySQL database, the "MySQL Workbench" client was used, while to consult the Google Spanner database the console that provides the tool itself was used in the GCP.

In the case of the MySQL database, the way used to measure time is to connect the proxy "cloud_sql_proxy" provided by the GCP platform with the MySQL database through the console, then connect the MySQL Workbench client and subsequently execute the query and record the time. In the case of the Google Spanner database, the query is simply executed in the console provided by the tool and the time is recorded.

For both cases, five measurements are made for each query and the average is calculated, which will be taken as the final result.

5.1 Results analysis

The average times obtained for both databases are analyzed below. These times are expressed in seconds.

Execution with MySQL

Prior to the execution of the consultations it is necessary that the insertion of the records in the database is relied upon. Although these times are not of interest for the evaluation of the performance of the database in front of different workloads, if they are useful when evaluating which is the option that allows a data upload to the cloud faster. Table 1 shows the times incurred in the insertion of the corresponding records.

Table 1. Insertion times for MySQL

	2000 Records	10000 Records	20000 Records
MySQL	1856,77	9318,88	15482,95

In Table 2, the results of the executions of the three queries for 2000, 10000 and 20000 records are shown. Each value is the result of the average of five executions of each query. In Fig. 1, the results shown in Table 2 can be seen graphically.

Table 2. Execution times for MySQL

	2000 Records	10000 Records	20000 Records
Query 1	0,2494	0,2404	0,2666
Query 2	0,2500	0,2624	0,3266
Query 3	0,2624	0,2678	0,2714

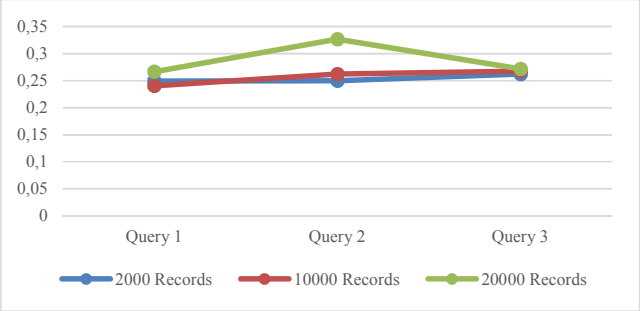


Fig. 1. Execution times for MySQL

Execution with Spanner

As in the case of the MySQL database, previous to the execution of the consultations it is necessary that the insertion of the records in the database is relied upon. Table 3 shows the times incurred in the insertion of the corresponding records.

Table 3. Insertion times for Spanner

	2000 Records	10000 Records	20000 Records
Spanner	34,21	39,72	41,75

In Table 4, the results of the executions of the three queries for 2000, 10000 and 20000 records are shown. Each value is the result of the average of five executions of each query. In Fig. 2, the results shown in Table 4 can be seen graphically.

Table 4. Execution times for Spanner

	2000 Records	10000 Records	20000 Records
Query 1	0,008878	0,03596	0,059554
Query 2	0,028214	0,054388	0,069064
Query 3	25,888	23,962	24,448

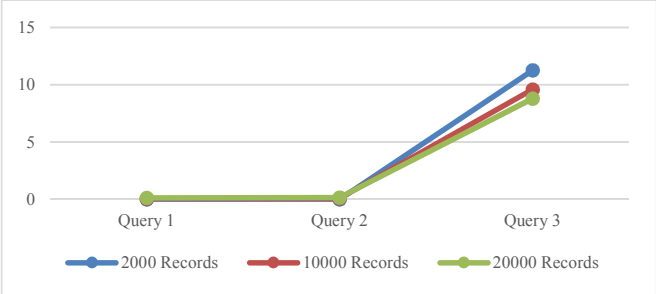


Fig. 2. Execution times for Spanner

Analyzing Tables 1 and 3 you can see that the insertion times of the registers in the database are significantly higher for MySQL. On the other hand, analyzing Tables 2 and 4, the best execution times for queries 1 and 2 are obtained with Google Spanner. However, the execution times of query 3 are significantly longer. It should be remembered that the executions were performed on an instance with a node.

To improve performance, the instance was configured with 4 nodes. In Table 5, the results of the executions of the three queries for 2000, 10000 and 20000 records are shown and in Fig. 3, the results shown in Table 5 can be seen graphically.

Table 5. Execution times for Spanner

	2000 Records	10000 Records	20000 Records
Query 1	0,01587	0,07524	0,1322
Query 2	0,02322	0,09514	0,1754
Query 3	11,256	9,577	8,788

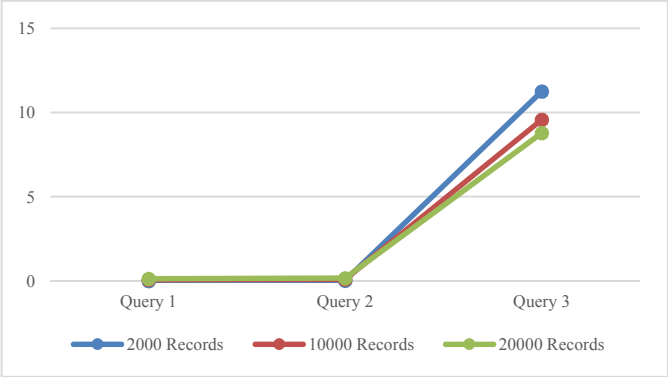


Fig. 3. Execution times for Spanner

As can be seen in Table 5, the execution times of query 3 improved with the addition of nodes for the instance. However, they are still significantly higher.

To improve performance, 2 instances are configured, with 4 nodes each. Table 6 shows the results of the executions of the three queries for records 2000, 10000 and 20000 and in Fig. 4, the results shown in Table 5 can be seen graphically.

Table 6. Execution times for Spanner

	2000 Records	10000 Records	20000 Records
Query 1	2,546	2,899	3,221
Query 2	2,665	2,985	3,198
Query 3	0,785	0,423	0,125

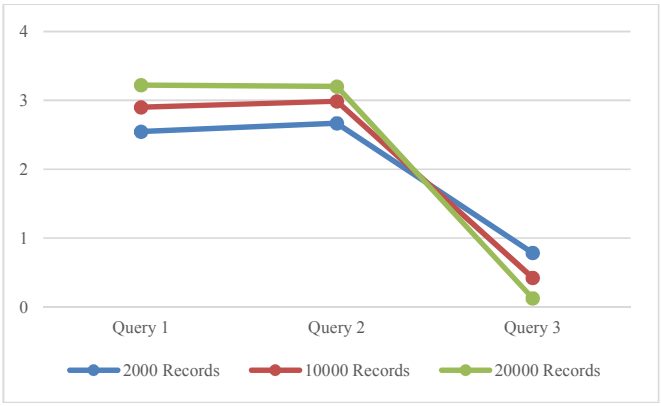


Fig. 4. Execution times for Spanner

Analyzing the results of Table 6, it can be seen that increasing Spanner resources yields a significant improvement. In the case of queries 1 and 2, using two instances of 4 nodes each, the performance in the execution of the queries declines. However, in the

case of query 3, the performance improves and the execution times are even lower than for the other queries.

6 Conclusion and Future work

NewSQL solutions aim to bring the relational data model into the world of NoSQL. Since there are a wide variety of NewSQL databases available now days, a new customer wishing to switch from the traditional physically hosted database server design to using NewSQL on the cloud, faces the problem of selecting the appropriate NewSQL database that can be used for his business needs without drastically changing his existing application. That is why this research offers a performance evaluation between a NewSQL database and a traditional database, MySQL, both hosted and running in the cloud; in order to measure the behavior against the increase of records and complexity of the queries. According to the experiments performed, the use of Spanner offers better results than MySQL in the face of workload and complexity of the queries.

It is proposed as future work, the use of greater workload which will be implemented by increasing records in the database and increasing complexity in queries.

References

[1] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Futur. Gener. Comput. Syst.*, vol. 29, no. 7, pp. 1645–1660, Sep. 2013.

[2] R. McFadyen and V. Kanabar, *An introduction to structured query language*. Wm. C. Brown, 1991.

[3] H. Plattner and Hasso, "A common database approach for OLTP and OLAP using an in-memory column database," in *Proceedings of the 35th SIGMOD international conference on Management of data - SIGMOD '09*, 2009, pp. 1–2.

[4] C. Kacfah Emani, N. Cullot, and C. Nicolle, "Understandable Big Data: A survey," *Comput. Sci. Rev.*, vol. 17, pp. 70–81, 2015.

[5] A. Oussous, F.-Z. Benjelloun, A. Ait Lahcen, and S. Belfkih, "Big Data technologies: A survey," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 30, no. 4, pp. 431–448, Oct. 2018.

[6] M. Barrionuevo *et al.*, "Estrategias y análisis orientados al manejo de datos masivos usando computación de alto desempeño," 2018.

[7] NIST, "NIST Big Data Working Group (NBD-WG)." [Online]. Available: https://bigdatawg.nist.gov/home.php. [Accessed: 18-Mar-2019].

[8] P. Mell and T. Grance, "The NIST Definition of Cloud Computing," 2011.

[9] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. Ullah Khan, "The rise of 'big data' on cloud computing: Review and open research issues," *Inf. Syst.*, vol. 47, pp. 98–115, Jan. 2015.

[10] J. Pokorný and Jaroslav, "Database technologies in the world of big data," in *Proceedings of the 16th International Conference on Computer Systems and Technologies - CompSysTech '15*, 2015, pp. 1–12.

[11] S. Madden, "From Databases to Big Data," *IEEE Internet Comput.*, vol. 16, no. 3, pp. 4–6, May 2012.

[12] R. Kumar, N. Gupta, H. Maharwal, S. Charu, and K. Yadav, "Critical Analysis of Database Management Using NewSQL," *Int. J. Comput. Sci. Mob. Comput.*, vol. 35, no. 5, pp. 434–438, 2014.

[13] Jing Han, Haihong E, Guan Le, and Jian Du, "Survey on NoSQL database," in *2011 6th*

International Conference on Pervasive Computing and Applications, 2011, pp. 363–366.

[14] A. B. M. Moniruzzaman, "NewSQL: Towards Next-Generation Scalable RDBMS for Online Transaction Processing (OLTP) for Big Data Management," Nov. 2014.

[15] Google, "Google Cloud Platform Overview | Overview | Google Cloud." [Online]. Available: https://cloud.google.com/docs/overview/. [Accessed: 01-Apr-2019].

[16] Google, "Cloud Spanner Documentation," 2017. .

[17] J. C. Corbett *et al.*, "Spanner: Google's Globally Distributed Database," *ACM Trans. Comput. Syst.*, vol. 31, no. 3, p. 8, 2013.

Orquestación de microservicios para realizar cómputo de grandes volúmenes de datos.

Miguel José Guevara Atencio[1], Maria Antonia Murazzo, Nelson Rodriguez.
miguelioseguevaratencio@gmail.com
nelson@iinfo.unsj.edu.ar
marite@unsj-cuim.edu.ar

Resumen. Los datos, aplicaciones y servicios informáticos son esenciales para el éxito de muchas empresas hoy en día. Para las pequeñas y medianas empresas que tienen sus propias bases de datos, el crecimiento exponencial de la información puede ser difícil de gestionar. Con el objetivo de lograr esto es que se considera que los contenedores pueden facilitar una infraestructura que permita soportar el tratamiento de grandes cantidades de datos generados por los dispositivos que hoy en día se conectan a Internet, además de proveer la capacidad de cómputo necesaria para procesar esos datos de manera dinámica y global. De esta manera los usuarios no deberán preocuparse por la escasez de recursos locales, sino centrarse en aprovechar la potencialidad de la aplicación. El presente trabajo tiene por objetivo identificar algunos de los aspectos a tener en cuenta en la orquestación de microservicios, la potencia de los contenedores, la arquitectura de Docker.

Palabras claves: Contenedores, Docker, Microservicio, Orquestación, Servicios.

1 Introducción

El Ministerio de Gobierno en conjunto con la Universidad Nacional de San Juan y la Policía de San Juan acordaron en la necesidad de un sistema informático que permita procesar toda la información existente en las distintas reparticiones y unidades policiales de la provincia de San Juan. Con este fin, se procede a iniciar el proyecto denominado; SIIS (Sistema Integral Informático de Seguridad), el cual debe ser, seguro, robusto, eficiente y que preste funciones las 24 horas del día, los 365 días del año. Esto último obliga a desarrollar una capa de abstracción de requerimientos de recursos, debido a que deben haber respuestas en tiempo real, con información integral y de alto impacto. Debe ejecutarse sobre una plataforma web que mediante redes de alta disponibilidad funcione ofreciendo tiempos de respuesta aceptables, en todo el territorio provincial. Durante el periodo de diseño y desarrollo surgieron necesidades en tiempos de respuesta, flexibilidad, facilidad de despliegue, dada la alta cantidad de transacciones realizadas a diario, y las prestaciones que el sistema debía entregar. Es por esto que se pretende estudiar y diseñar un esquema basado en

[1]

contenedores que tenga la capacidad de replicar recursos automáticamente y bajo demanda según sea necesario, todo esto de forma invisible para el usuario de la aplicación. A su vez también se busca que el despliegue de las versiones sea automatizado.

2 Conceptos Previos

El principal problema que se plantea cuando se necesita administrar recursos fijados como requerimiento de una aplicación de big data, es analizar y planificar la forma en la cual la aplicación debe escalar. Tradicionalmente, la construcción de aplicaciones ha seguido un paradigma monolítico como se menciona en [1].En este contexto, un desarrollo de software que se ajusta en una arquitectura monolítica, es una aplicación con una única base de código, que ofrece una variedad de servicios que utilizan diferentes interfaces tales como páginas HTML encapsulados a un mismo repositorio [2].

La principal desventaja que presentan las aplicaciones desarrolladas bajo este enfoque es que, si se necesita realizar una modificación sobre una funcionalidad, esta impactará en gran parte de las demás funcionalidades del sistema de modo tal que aumenta el costo de llevar a cabo dicha tarea. Otra desventaja es que, al tratarse de un esquema único, pueden generar un cuello de botella debido a que un error en la aplicación deja sin servicios todo el sistema.

Estas exigencias, se vuelven aún más críticas cuando se debe trabajar con grandes volúmenes de datos, dada las restricciones de volumen, velocidad y multiplicidad de formatos; y la necesidad de definir con exactitud los requisitos de capacidad de proceso y almacenamiento de datos. Trabajar con datos masivos obliga a contar con una plataforma escalable, capaz de crecer, adaptarse y responder a nuevos requerimientos por parte del usuario.

Es por ello que se presenta la necesidad de usar otro tipo de paradigma arquitectónico para desarrollar aplicaciones que gestionen grandes volúmenes de datos.

Una alternativa es implementar SOA, donde los servicios Web son componentes de software ligeramente acoplados entregados a través de tecnologías estándar de Internet. Es decir, los servicios Web son aplicaciones de negocio auto descriptivas y modulares que exponen la lógica empresarial como servicios a través de Internet, por medio de una interfaz programable y donde el protocolo de Internet (IP) puede ser utilizado para encontrar e invocar esos servicios. Un servicio web es el elemento clave en la integración de diferentes sistemas de información, ya que pueden basarse en diferentes plataformas, lenguajes de programación y tecnologías.

Una forma de implementar los servicios web es mediante microservicios. En [3] se define a los Microservicios como un nuevo método de diseño para la creación de servicios de bajo acoplamiento, pero autónomos, que permite a las empresas utilizar las tecnologías emergentes con el fin de crear y gestionar sistemas modulares a una escala que supera los demás enfoques, como la arquitectura orientada al servicio (SOA).

Los Contenedores son una capa de software que permiten correr múltiples instancias de Linux en un mismo host. Esto se logra aislando un conjunto de procesos y librerías de forma que cada uno tenga su propia vista privada del sistema operativo con un propio identificador de procesos, sistema de archivos e interfaces de red.

Cabe mencionar que dichas librerías permiten ejecutar una aplicación determinada en cualquier sistema operativo. Aportan ligereza, portabilidad, autosuficiencia, y una flexibilidad muy codiciada por los desarrolladores. Un contenedor es un cubo de software que comprende todo lo necesario para ejecutar el software de forma independiente.

En el caso de contendores Linux se utiliza LXC del kernel, y para contenedores Windows se utiliza el motor de virtualización Hyper-V. Para el proyecto que se utilizará de ejemplo se trabajará con Docker con sus contenedores, imágenes y repositorios lo cual permite lograr un despliegue exitoso, eficiente, ágil y portable. Son estas ventajas las que han traído el éxito, que se ha dejado notar tanto por la implicación de algunos grandes del sector como Google, Red Hat, IBM y Microsoft, como por su implantación y en grandes empresas y servicios cloud que todos usamos a diario como AWS, Digital Ocean, Google Cloud Platform.

Para comprender mejor, en la práctica un contenedor es un proceso que, internamente, contiene la aplicación que se quiere ejecutar y todas sus dependencias usando indirectamente el kernel del sistema operativo que se esté usando. Realizando un paralelismo entre un contenedor y una VM, ambos son sistemas auto contenidos que tienen como principal diferencia que, una VM necesita contener todo el sistema operativo mientras un contenedor aprovecha el sistema operativo en el que se ejecute como se puede ver en la figura 1.

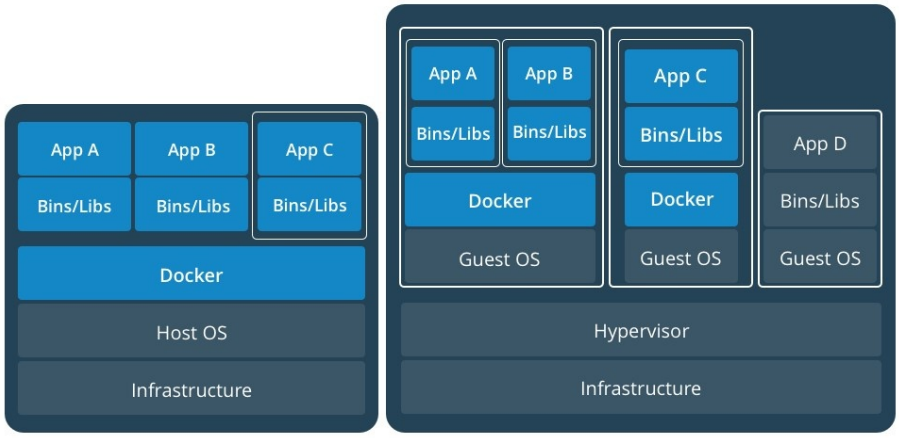


Figura 1. Contenedores vs Máquinas virtuales.

La principal ventaja de utilizar contenedores es que no depende de las instalaciones del sistema anfitrión. Otra ventaja es que es posible definir repositorios que

eventualmente alojarán a los contenedores con el objeto que otros usuarios los descarguen y consuman como servicio.

3 Características de Microservicios

Un modelo basado en microservicios trae aparejado la necesidad de administrar dichos servicios. Para esto surgen dos metodologías; la orquestación de microservicios donde cada proceso no tiene por que conocer la lógica de implementación de los demás, sino que existe un orquestador central encargado de dicha tarea. Y por otra parte, la coreografía de servicios donde cada microservicio debe conocer los tiempos y momentos de los procesos con los que coexiste.

Los beneficios de los microservicios, tales como su descubrimiento en tiempo de ejecución y su bajo acoplamiento, contribuyen a la orquestación de microservicios proporcionando un acercamiento al modelado y ejecución de procesos de negocio en tiempo real.

La implementación de orquestación en sistemas empresariales se vuelve fundamental si se trata de aplicaciones que utilizan múltiples microservicios. Donde cada contenedor representa un microservicio determinado que dicho sistema consume para llevar a cabo la tarea necesaria. Así, a medida que los sistemas crecen en demanda de recursos, resulta imprescindible la implementación de orquestadores de servicios.

Cuando la plataforma de despliegue son contenedores, esto habilita que los sistemas puedan pensarse y construirse de forma distribuida; y de esta forma contar con mecanismos de escalabilidad horizontal, que permita distribuir la carga de trabajo dinámicamente y paralelizar las tareas, más allá del nodo local. La escalabilidad horizontal implica balanceo de carga, replicación de recursos, reinstanciación de recursos en tiempo real y optimización en su uso, de forma que el usuario no perciba degradación de la experiencia de usuario.

El diseño de aplicaciones mediante una arquitectura basadas en microservicios, permite el escalado de aplicaciones a varios hosts mediante tecnología basada en contenedores.

Es por esto que utilizando los conceptos de contenedores y orquestación de recursos, será posible realizar el procesamiento sobre grandes volúmenes de datos independientemente de los recursos existentes.

4 Despliegue de microservicios con Docker

Los avances tecnológicos en telecomunicaciones, como dispositivos inalámbricos, entre otros, facilitan el acceso a la información. Por lo que el desarrollo de sistemas y aplicaciones distribuidas es altamente requerido”[4].

En la actualidad los sistemas basados en servicios corriendo en servidores dedicados ya sean físicos y/o virtuales, conllevan consigo mismo una carga extra física y logística para distribuir adecuadamente los recursos de los servidores.

Este nuevo enfoque reduce dichas tareas de gestión al punto tal de automatizar casi la mayoría de las operaciones de administración y mantenimiento posterior de servidores, dado que en fin es una filosofía que apunta a la gestión de servicios corriendo sobre contenedores, los cuales son económicos de implementar y se abstraen de las capas inferiores, dicho esquema brinda la libertad a los departamentos de IT de implementar la infraestructura deseada.

Surgen de esta forma, distintos esquemas y soluciones, que permiten reducir los costos de desarrollo, entrega y mantenimiento de sistemas de distinta índole que ofrecen servicios específicos. Entre las soluciones más conocidas se encuentran la arquitectura basada en micro servicios, la cual, está de auge dado el devenir de tecnologías altamente robustas y soportadas por extensas comunidades de desarrollo Open Source.

La arquitectura de micro servicios es un distintivo sistema de desarrollo de software que ha crecido en popularidad en los últimos años [5].

A continuación se detallan las características que una plataforma debe contemplar para llevar a cabo la implementación de un esquema basado en micro servicios.

4.1 Compose

Implementada en el lenguaje de scripts Python, su componente fundamental es un archivo de control central basado en el lenguaje de marcado YAML. En el archivo docker-compose.yml se pueden definir tantos contenedores de software como se desee incluyendo todas las dependencias, así como sus interrelaciones.

4.2 Dockerfile

Toda la configuración específica de cada uno de los contenedores se plasma en un archivo denominado Dockerfile como se puede ver en la figura 2. para poder dar soporte a los requisitos del servicio al que atenderán.

```
FROM webdevops/php-apache

ENV TZ=America/Argentina/Buenos_Aires

RUN apt-get update \
    && apt-get install php-pgsql -y \
    && echo display_errors=Off >> /opt/docker/etc/php/php.ini \
    && echo short_open_tag=On >> /opt/docker/etc/php/php.ini \
    && echo extension=php_pgsql.so >> /opt/docker/etc/php/php.ini

EXPOSE 80 443
```

Figura 2. Dockerfile

Docker gestiona un repositorio de imágenes denominado Docker Hub, el cual permite automatizar el despliegue y entrega continua de servicios creados a través de imágenes.

Con todo lo mencionado anteriormente se obtuvieron los contenedores listos para comenzar a configurar los respectivos microservicios.

Luego se debe ejecutar el script con el cual Compose brindará esta información a Docker. Éste levanta inicialmente una subred dentro de la cual van a coexistir todos los contenedores. Así Docker establece los puertos para acceder a cada microservicio. Una vez culminado el proceso de registración en el sistema subyacente, con la aplicación ya espejada entre el sistema operativo y el contenedor, el microservicio de php ejecuta la aplicación, internamente la aplicación accede a cada microservicio indicando su nombre por pasaje de variables de entorno que Docker ya se encargó de establecer cuando inició el Docker Compose. El pasaje de variables de entorno es invisible al desarrollador y automatizado por Docker.

Al final del proceso tenemos una aplicación consumiendo un pool de microservicios que Docker logra mantener disponibles, gestionando los contenedores por debajo.

```
version: '3'
services:
  phpservice:
    container_name: apache
    build:
      context: .
      dockerfile: DockerfilePhpApache
    volumes:
      - .:/app/siistesting/
    ports:
      - ${HTTP_PORT}:80
  mysqlservice:
    container_name: mysqldb
    build:
      context: .
      dockerfile: DockerfileDb
    depends_on: [phpservice]
    volumes:
```

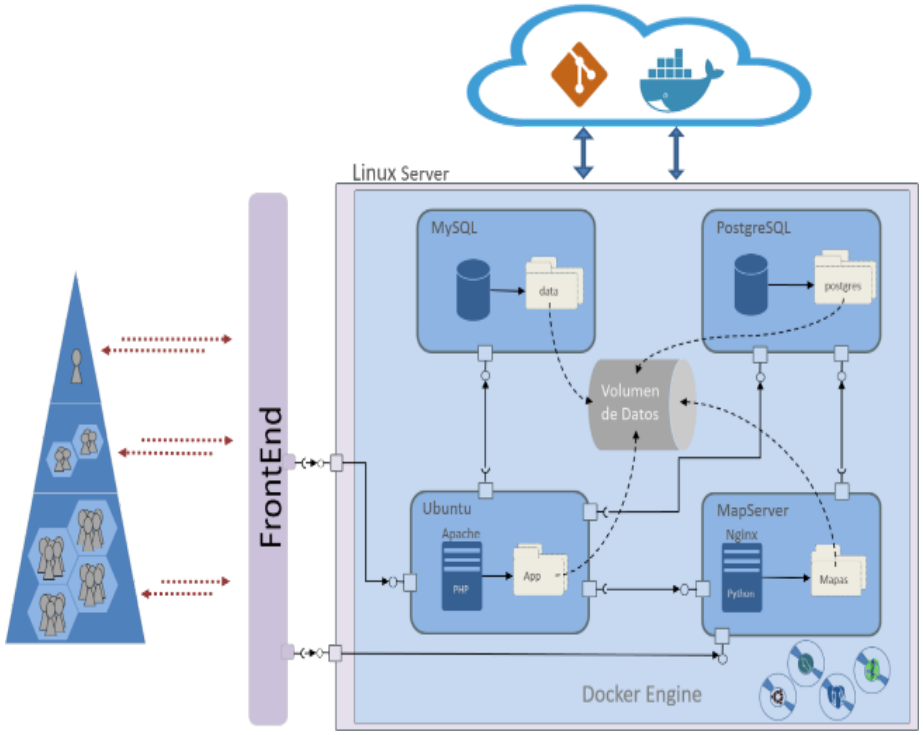
```
      - ./data:/var/lib/mysql
    ports:
      - ${MYSQL_PORT}:3306
  postgreservice:
    container_name: postgres
    build:
      context: .
      dockerfile: DockerfilePostgis
    depends_on: [mysqlservice]
    ports:
      - ${POSTGRE_PORT}:5432
  mapserverservice:
    container_name: mapsvr
    image: geodata/mapserver
    depends_on: [postgreservice]
    volumes:
      - ./mapas/:/maps:ro
    ports:
      - ${MAPSERVER_PORT}:80
```

5 Caso Practico

El Gobierno de la Provincia de San Juan a través del Ministerio de Gobierno y la Universidad Nacional de San Juan a través del Departamento de Informática se proponen la Vinculación de todos los sistemas existentes con la creación del SISTEMA INTEGRAL INFORMÁTICO DE SEGURIDAD – S.I.I.S. que permita la gestión automática de todas las tareas que se llevan a cabo en la Fuerza.

Desde el FrontEnd, el usuario realiza una consulta al sistema corriendo en el contendor con un servicio web, luego este consumirá un servicio de otro corriendo MySQL, para determinar los permisos del usuario, luego se comunicará con postgres para obtener los datos de geolocalización de la dependencia. Posteriormente, el usuario a través del FrontEnd consumirá los mapas del servicio de MapServer.

Luego, como se puede ver en la figura 3, los componentes del sistema diseñado bajo una arquitectura basada en micro servicios se vincula de la siguiente manera



.

Figura 3. Componentes.

6 Resultados Obtenidos

Para realizar las métricas se decidió utilizar la herramienta de desarrollo de Google Chrome. Se realizó una comparación de tiempos de respuesta entre una versión del sistema implementado bajo un esquema monolítico con el stack LAMP, y la versión del sistema implementado bajo el esquema basado en orquestando de micro servicios.

En la vista inicial del sistema, la versión monolítica tiene un tiempo de casi 3500 milisegundos como se puede ver en la figura 4.

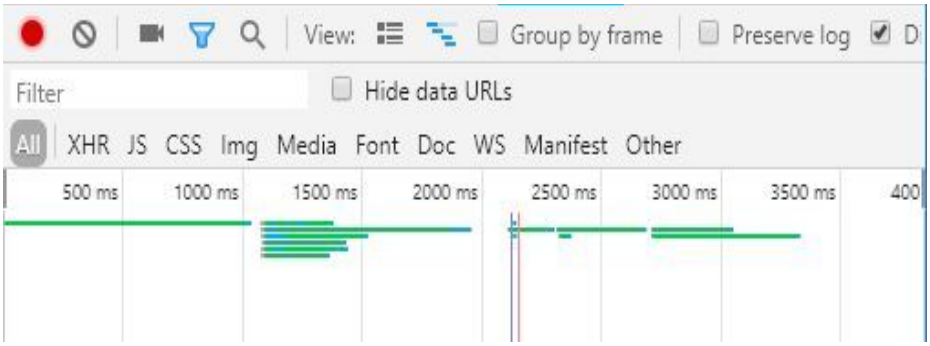


Figura 4. Esquema monolítico.

En la misma vista, como se puede observar en la figura 5, en el esquema basado en orquestación de micro servicios se puede observar un tiempo de respuesta de poco más de 600 milisegundos.

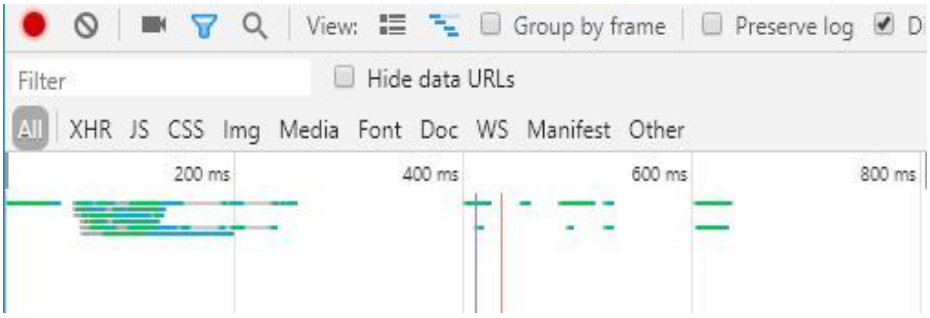


Figura 5. Esquema basado en orquestación de microservicios.

Con lo cual se logra reducir en un 80.79% el tiempo en comparación con la versión monolítica.

7 Conclusiones y futuros trabajos

Dados los resultados obtenidos mediante la implementación de Orquestación de microservicios se puede inferir que se logró disminuir el tiempo involucrado en el despliegue y puesta en producción del Sistema Integral Informático de Seguridad. Es decir, la implementación de contenedores como soporte para el despliegue de aplicaciones de procesamientos de grandes volúmenes de datos, facilitan las tareas de entrega continua como así también de integración continua.

Además, se logró disminuir los tiempos de respuesta de las consultas a cada microservicio debido a el balanceo de carga provisto por la implementación de múltiples contenedores.

Como resultado del uso de la tecnología de contenedores fue posible lograr interoperabilidad entre todos los microservicios; así como la independencia entre las tecnologías internas de cada microservicio y de tecnologías externas de desarrollo.

Como desventaja de este esquema, para trabajar con contenedores de software es necesario diseñar o migrar la arquitectura de las aplicaciones a una basada en microservicios, esto implica tareas como, reconfiguración de interfaces de aplicación que antes no existían, dividir cada capa o proceso en microservicios, configurar esquemas de red y unidades de almacenamiento, etc. Lo cual requiere un análisis del contexto cuando se decide implementar orquestación de microservicios ya que puede significar un costo de reingeniería excesivo para implementar aplicaciones legacy.

Por otra parte, una mala configuración de los microservicios puede llevar a un estado de entropía tal que el sistema pierda su disponibilidad, y por ende se degrada la performance del mismo. Por tanto es importante seguir adecuadamente el diseño de cada microservicio.

Con respecto a futuros trabajos se pretende implementar un motor como kubernetes. Este permite orquestar microservicios en un cluster de contenedores sobre infraestructura privada, por ejemplo.

Una de las características de Kubernetes es que permite aumentar la capacidad de procesamiento transfiriendo el procesos a plataformas públicas cuando se vea sobrepasado, de este modo a la hora de desplegar aplicaciones de procesamiento de grandes volúmenes de datos se pueda derivar tareas a infraestructuras adecuadas según sea necesario.

Referencias

[1] D. López and E. Maya, “Arquitectura de Software basada en Microservicios para Desarrollo de Aplicaciones Web.”

[2] T. DE Titulación Previo A La Obtención Del Título De and F. Naranjo Pamela Caterine, “ESCUELA POLITÉCNICA NACIONAL FACULTAD DE INGENIERÍA DE SISTEMAS PROPUESTA METODOLÓGICA PARA MIGRACIÓN DE SISTEMAS WEB CON ARQUITECTURA MONOLÍTICA HACIA UNA ARQUITECTURA BASADA EN MICROSERVICIOS CO-DIRECTOR: MSc. ANCHUNDIA VALENCIA CARLOS EDUARDO,” 2017.

[3] SOLUCIÓN INNOVADORA PARA UNA TI FLEXIBLE Y RECEPTIVA IMPLEMENTACIÓN MÁS RÁPIDA MÁS OPCIONES.”

[4] I. K. Farías Alejandro, “Definición de un ambiente de construcción de aplicaciones empresariales a través de Devops, Microservicios y Contenedores,” 2017.

[5] F. Humberto, V.-R. Universidad, F. De, and P. Santander, “Método de automatización del despliegue continuo en la nube para la implementación de microservicios,” 2018.

Integración de una BCI mediante la implementación de VRPN en escenarios de Serious Game para rehabilitación de problemas neurológicos

Javier J. Rosenstein, Rodrigo Gonzalez, Nicolás Ignacio Zárate Alvarez, Matías Campos

Instituto de Investigaciones, Facultad de Informática y Diseño, Universidad Champagnat, Belgrano 721, 5501 Godoy Cruz, Mendoza, Mendoza, Argentina.

rosensteinjavier@uch.edu.ar, gonzalezrodrigo@uch.edu.ar, nicolaszarate23@gmail.com, mgcampos9785@gmail.com

RESUMEN

En el desarrollo de sistemas de realidad virtual uno de los inconvenientes que se encuentran es la comunicación entre las aplicaciones y los dispositivos de adquisición de ondas cerebrales. Ya sea por no disponer de un método de acceso en forma directa o por necesitar independencia entre ambos, es decir que las aplicaciones corran en una plataforma y los dispositivos en otras. Para lograr esta independencia y a su vez permitir la integración de todo el sistema de realidad virtual, es necesario la implementación de algún protocolo de comunicaciones que permita esta vinculación heterogénea en tiempo real. Los dispositivos generalmente están asociados a funciones o características de los individuos que los utilizan y se necesita integrar los movimientos que estos representan a la aplicación de realidad virtual correspondiente. El presente trabajo trata del análisis e implementación del protocolo de comunicaciones VRPN (Virtual Reality Protocol Network) entre las partes de un entorno multimedia donde interactúan la adquisición de movimientos del usuario y la representación visual en un escenario virtual que permita la retroalimentación al usuario en tiempo real logrando una experiencia interactiva e inmersiva. Esto tiene aplicación directa en los tratamientos de rehabilitación en pacientes de patologías neurológicas y cognitivas.

Palabras clave: VRPN, EOG, EEG, BCI, Serious Games, Mirror Neurons, Neurofeedback

CONTEXTO

El presente proyecto forma parte del trabajo de tesis de Maestría en Teleinformática de la Universidad de Mendoza (Ciudad, Mendoza), correspondiente al tesista Javier Rosenstein, el mismo se desarrolla en el Instituto de Investigaciones de la Facultad de Informática y Diseño de la Universidad Champagnat (Godoy Cruz, Mendoza), en el marco de la Licenciatura en Sistemas de Información; en cooperación con el Laboratorio de I+D+i en Neurotecnologías de la empresa Neuromed Argentina S.A. (Godoy Cruz, Mendoza).
Este trabajo es parte del proyecto de investigación que dio inicio en Diciembre de 2017 denominado “Diseño y desarrollo de un prototipo de Serious Game destinado a la rehabilitación de problemas neurológicos implementando VRPN para la comunicación de la BCI”.

1. INTRODUCCIÓN

El objetivo principal de la presente línea de investigación consiste en el diseño y desarrollo de una BCI (Brain Computer Interface) [1] [2], que permita interactuar entre las señales generadas por un paciente neurológico [3] y una interfaz de realidad virtual. Este sistema debe permitir lograr el principio de neurofeedback [4], o retroalimentación hacia el paciente. De este modo se podrán mejorar sus capacidades cognitivas correspondientes. Incluso, estudios indican la posibilidad de tratar patologías psiquiátricas como la depresión [5] [6].

La implementación del trabajo se organiza de acuerdo a las siguientes etapas:

a) Adquisición de señales mediante técnicas de Electroencefalografía (EEG) [7] y adquisición de movimientos oculares mediante las técnicas de electrooculografía (EOG) [8] [9] [10] [11] [12].
b) Análisis de estas señales en tiempo real para poder identificar la voluntad de movimiento del individuo, así como la dirección del movimiento.
c) Transformar la voluntad de movimiento en comandos del protocolo VRPN (Virtual Reality Protocol Network) que permitan transmitir la información al componente software / hardware que la requiera.
d) Como continuación al resultado obtenido en el punto anterior, poder comandar una interfaz gráfica de aprendizaje o interfaz Cerebro / Computadora (BCI – Brain Computer Interface).

Finalmente se presenta la BCI como un sistema de adquisición de datos, procesamiento del protocolo serie a VRPN y luego la representación en nuestro modelo de prototipo de Serious Game [14] [15]. Una vez adquiridas las señales EEG/EOG, estas se analizan y codifican con las bibliotecas desarrolladas como parte de este proyecto. Esto permite su comunicación mediante VRPN con las interfaces virtuales que interpretan este protocolo.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

Este trabajo está compuesto por tres etapas o fases de trabajo. La primera de ellas corresponde a la implementación de VRPN entre un sistema simulado de captura de datos provenientes de un paciente, y una interfaz virtual básica. La segunda etapa del proyecto pretende avanzar sobre la captura de datos reales para que, luego de procesados, se transfieran a través de VRPN hacia la interfaz virtual de neurofeedback. Una breve descripción de cada una de las etapas se describe a continuación:

1. La primer parte consiste en la implementación de un simulador de señales EEG y EOG necesarias para el análisis e interpretación de la voluntad del usuario de la BCI. Estas señales una vez procesadas se deben codificar en comandos de VRPN para poder ser transmitidas hacia una interfaz virtual. Esta debe poder interpretar las señales transmitidas y representar la voluntad inicial del usuario correspondiente. De este modo se cumplen los objetivos de captura, análisis, procesamiento, transmisión, recepción y representación, lo cual produce el efecto de neurofeedback deseado sobre una interfaz virtual de capacitación a nivel prototipo.
2. La segunda etapa consiste en el desarrollo y mejora del proyecto mediante la adquisición real de señales EEG y EOG por electrodos ubicados superficialmente sobre la cabeza del paciente. Esta modificación al sistema requiere de un diseño e implementación electrónica así como del desarrollo del firmware que permita adquirir, analizar y procesar estas señales, que luego serán transmitidas a través del protocolo VRPN hacia una interfaz virtual. Así, el sistema diseñado se incorporará en forma transparente al proyecto implementado en la primer etapa.

3. La etapa final de este proyecto consiste en el diseño y desarrollo del escenario virtual de rehabilitación cognitiva para pacientes neurológicos según las indicaciones concretas por parte del especialista en neurología. Se partirá de un relevamiento de las técnicas de aprendizaje que se requieren implementar y los resultados que se pretenden obtener, indicados por el neurólogo o experto afín.

3. RESULTADOS ESPERADOS

Los resultados esperados se pueden dividir en dos grandes grupos:

1. BCI e interfaz virtual de neurofeedback: se espera obtener como producto final un sistema de retroalimentación a un supuesto usuario simulado por la generación de señales que comandan el escenario de realidad virtual via comandos codificados en el protocolo VRPN que se transmiten a través de una red Ethernet.

2. Adquisición de datos reales de EEG y EOG: análisis y preprocesamiento; se identifican los comandos necesarios para su transmisión hacia el equipo generador de tramas VRPN para que finalmente se tenga el proyecto totalmente terminado y funcional que implemente el efecto de neurofeedback propuesto.

4. FORMACIÓN DE RECURSOS HUMANOS

La línea de I+D presentada está vinculada con el desarrollo de una tesis de postgrado, por parte del estudiante de maestría en teleinformática de la Universidad de Mendoza, Javier J. Rosenstein. Dicha tesis se centra en la implementación del protocolo VRPN demostrando su uso en un sistema BCI.

Además cuenta con la dirección del Lic. Javier J. Rosenstein (UCH) y como investigador principal el Dr. Rodrigo Gonzalez (UCH). En lo que respecta a la formación de estudiantes de la licenciatura, esta línea de investigación cuenta con dos tesis de grado en curso, pertenecientes a los estudiantes Nicolás Ignacio Zárate Alvarez y Matías Campos, cuyos planes de tesis se encuentran específicamente dentro del marco de este proyecto. Ambos cursan la licenciatura en Sistemas de Información en la Universidad Champagnat.

Una vez que el proyecto se encuentre implementado desde su primer etapa será utilizado como recurso para el dictado de talleres de comunicaciones, redes, programación de microcontroladores, Programación en C/C++ y Diseño y programación de interfaces virtuales de capacitación/rehabilitación en general, tanto para estudiantes de la universidad, como así también para aquellos interesados externos a la UCH.

5. BIBLIOGRAFÍA

1. Y. Wang, X. Gao, B. Hong, and S. Gao, “Practical designs of brain-computer interfaces based on the modulation of EEG rhythms,” in Brain-Computer Interfaces. Springer, 2009, pp. 137–154.
2. J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, “Brain-computer interfaces for communication and control,” Clinical neurophysiology, vol. 113, no. 6, pp. 767–791, 2002.
3. J. A. Pineda, “The functional significance of mu rhythms: translating “seeing” and “hearing” into “doing”,” Brain Research Reviews, vol. 50, no. 1, pp. 57–68, 2005.
4. S. Enriquez-Geppert, R. J. Huster, and C. S. Herrmann, “Boosting brain functions: Improving executive functions with behavioral training, neurostimulation, and neurofeedback,” International Journal of Psychophysiology, vol. 88, no. 1, pp. 1–16, 2013.

5. R. Ramirez, M. Palencia-Lefler, S. Giraldo, and Z. Vamvakousis, "Musical neurofeedback for treating depression in elderly people." Frontiers in neuroscience, vol. 9, pp. 354–354, 2014.
6. W. Rief, "Getting started with neurofeedback," 2006.
7. J. D. Kropotov, Quantitative EEG, event-related potentials and neurotherapy. Academic Press, 2010.
8. A. Bulling, J. A. Ward, H. Gellersen, and G. Troster, "Eye movement analysis for activity recognition using electrooculography," IEEE Transactions on pattern analysis and machine intelligence, vol. 33, no. 4, pp. 741–753, 2011.
9. H. Singh and J. Singh, "A review on electrooculography," International Journal of Advanced Engineering Technology, vol. 3, no. 4, pp. 115–122, 2012.
10. D. P. Bautista, I. A. Badillo, D. De la Rosa Mejía, and A. H. H. Jiménez, "Interfaz humano-computadora basada en señales de electrooculografía para personas con discapacidad motriz," ReCIBE, vol. 3, no. 2, 2016.
11. S. Yathunanthan, L. Chandrasena, A. Umakanthan, V. Vasuki, and S. Munasinghe, "Controlling a wheelchair by use of EOG signal," in 2008 4th International Conference on Information and Automation for Sustainability, IEEE, 2008, pp. 283–288.
12. V. C. C. Roza, "Interface para tecnologia assistiva baseada em eletrooculografia," 2014.
13. A. C. Gaviria, I. C. Miller, S. O. Medina, and D. R. Gonzales, "Implementación de una interfaz hombre-computador basada en registros EOG mediante circuitos de señal mixta PSoC," in V Latin American Congress on Biomedical Engineering CLAIB 2011 May 16-21, 2011, Habana, Cuba. Springer, 2013, pp. 1194–1197.
14. P. Rego, P. M. Moreira, and L. P. Reis, "Serious games for rehabilitation: A survey and a classification towards a taxonomy," in 5th Iberian Conference on Information Systems and Technologies. IEEE, 2010, pp. 1–6.
15. J. S. Breuer and G. Bente, "Why so serious? On the relation of serious games and learning," Eludamos. Journal for Computer Game Culture, vol. 4, no. 1, pp. 7–24, 2010.

DNS Tunneling: A Deep Learning based Lexicographical Detection Approach

Franco Palau[1,3], Carlos Catania[1], Jorge Guerra[1], Sebastian Garcia[2], and Maria Rigaki[2]

[1] LABSIN, Facultad de Ingenieria. UNCuyo. Mendoza. Argentina
franco.palau@ingenieria.uncuyo.edu.ar|harpo@ingenieria.uncuyo.edu.ar|jr@uncuyo.edu.ar
[2] CTU - Czech Technical University. Prague. Czech Republic
sebastian.garcia@agents.fel.cvut.cz
[3] UCH - Universidad Champagnat. Mendoza. Argentina

Abstract. Domain Name Service is a trusted protocol made for name resolution, but during past years some approaches have been developed to use it for data transfer. DNS Tunneling is a method where data is encoded inside DNS queries, allowing information exchange through the DNS. This characteristic is attractive to hackers who exploit DNS Tunneling method to establish bidirectional communication with machines infected with malware with the objective of exfiltrating data or sending instructions in an obfuscated way. To detect these threats fast and accurately, the present work proposes a detection approach based on a Convolutional Neural Network (CNN) with a minimal architecture complexity. Due to the lack of quality datasets for evaluating DNS Tunneling connections, we also present a detailed construction and description of a novel dataset that contains DNS Tunneling domains generated with five well-known DNS tools. Despite its simple architecture, the resulting CNN model correctly detected more than 92% of total Tunneling domains with a false positive rate close to 0.8%.

Keywords: Neural Networks · Network Security · DNS Tunneling.

1 Introduction

The Domain Name System (DNS) is a well-established and trusted protocol. It was originally made for name resolution and not for data transfer, so it is often not seen as a malicious communications and data exfiltration threat and therefore, organizations rarely analyze DNS packets for malicious activity. This lack of attention has made it attractive to attackers, who exploit the DNS protocol by using a technique called **DNS Tunneling**. With this method, data is encoded inside DNS queries and responses, allowing attackers to exchange information in an obfuscated way.

A requirement to use DNS Tunneling is that the attacker must control a domain and a DNS server, which receives the DNS requests for the domain and

act as an authoritative server for that domain in order to run the server-side tunneling and decoding programs. By doing so, the attacker can observe all incoming DNS queries and have control under the answers to the queries.

Consider a machine that has been comprised, i.e. infected with malware. DNS Tunneling can be used by the attacker in two ways [3]:

- to exfiltrate data from the compromised machine to an external server controlled by the attacker.
- to establish a tunnel and send communications and instructions from the external server to the compromised machine.

If the attacker registers the domain `evilsite.com`, then data from comprised machine can be transmitted as a DNS request to
<`base64_encoded_data`>`.evilsite.com`. Then the server controlled by the attacker could send any command/data by responding to the query with a CNAME, TXT or NULL record as response. Hence, the attacker can receive and send data, i.e. a bidirectional data transfer channel is achieved using DNS tunneling. The advantages of this method include that DNS is almost always accessible in the network , no direct connection is established between the comprised machine and attacker, and pure data exfiltration (upstream only) is difficult to detect.

Catching threats which exploit the DNS has become a central topic in network security and various types of malicious domains detection methods have been proposed. A common approach for detecting these DNS Threats are the so-called lexicographical approaches. Under such approaches domains are classified by studying the statistical properties of the characters conforming the domain name, such as the frequency distribution of properties, domain character length, Shannon entropy, the vowel/consonant ratio and dictionary-based similarity among others [1, 3, 11].

Several other DNS Threats detection approaches extended the idea of using the information provided by domain names properties to train a machine learning classifier such as Support Vector Machine (SVM), Naive Bayes (NB), Decision Tree (DT) [13] or Logistic Regression [3]. Recently, to avoid the need of designing the right set of features for training machine learning classifiers, some authors explored the application of Deep Learning (DL) techniques. In [16], an LSTM network was presented as a Domain Generation Algorithms (DGA) classifier reaching a 90% detection rate. With a similar focus, in [2] a DGA classifier was built, but in this case based on a simple 1D Convolutional neural network (1D-CNN) capable of detecting more than 97% of total DGA domains with a false positive rate close to 0.7%.

As state in [6], research on DGA detection can be used for DNS Tunneling detection since DGA generated domains are abnormal in a similar way to names from data encoding, so it is reasonable to apply the same detection approach. Following this idea, in the present work we focus on analyzing the DNS Tunneling detection performance of a 1D-CNN with an architecture similar to [2]. In addition and given the lack of quality datasets for evaluating DNS tunneling connections, we create a new dataset from five well-known DNS tunneling tools.

The dataset was created on Virtual Machine infrastructure using a time injection methodology [14] for executing and logging several DNS tunneling tools.

The main contributions of the present article are:

- A 1D convolutional neural network (1D-CNN) capable of detecting DNS Tunneling in Domain names.
- The evaluation of the 1D-CNN model on a new generated dataset following a well defined methodology for labeling tunneling connections.

The rest of this paper is organized as follows: Section 2 describes the network architecture, Section 3 explains the dataset construction, Section 4 illustrates the experimental design while Section 5 details the results, and finally in Section 6, we present the conclusions.

2 Deep Learning model Lexicographical Detection

The Neural Network Architecture model is a 1D Convolutional Neural Network (CNN). This CNN is composed of three main layers. The first one is an *Embedding* layer, then there is a *Conv1D* layer, and finally a Dense fully connected layer. The first two layers are the most relevant components of the architecture regarding the problem of detecting Tunneling domains. Both layers are responsible for learning the feature representation in order to feed the third Dense and fully connected layer. Beside the three layers previously described, the complete Neural Network Architecture includes some other layers for dealing with the dimensions output of the *Conv1D* layer as well as layers for representing the input domain and the output probability. A detail of the complete architecture together with the used activation functions is shown in Table 1, whereas the three main layers are described in the following subsections.

Table 1: The complete network architecture including the corresponding output dimensions and activation function used in each layer

Layer (type)	Activation Function
input (Input Layer)	-
embedding (Embedding)	-
conv1d (Conv1D)	relu
dense_1 (Dense)	relu
dense_2 (Dense)	sigmoid

2.1 Embedding Layer

A character embedding consists in projecting l-length sequences of input characters into a sequence of vectors R^{lxd}, where l has to be determined from the

information provided by the sequences in the training set and d is a free parameter indicating the dimension of the resulting matrix [16]. By using an *Embedding* layer in the architecture, the neural network learns in an efficient manner the optimal set of features that represent the input data.

2.2 1D Convolutional Layer

The *Conv1D* layer refers to a convolutional network layer over one dimension. For the Tunneling detection problem, such dimension consists of the length of the domain name sequence. The convolutional layer is composed of a set of convolutional filters that are applied to different portions of the domain name. A visual example of the feature extraction process for a 1D Conv layer is shown in Fig. 1. The figure depicts a 1D convolutional layer constructing 256 filters (features) ($nf = 256$), with a window (kernel) size of 4 ($ks = 4$) and a stride length value of 1 ($sl = 1$). The layer selects from groups (also referred as patches) of 4 characters to apply the convolutional filters, and continues shifting one character at a time (stride value) applying the same convolutions filter over the rest of the sequence. Consequently, the neural network generates 4-grams features. These features represent the discriminative power of these group of letters in the domain names.

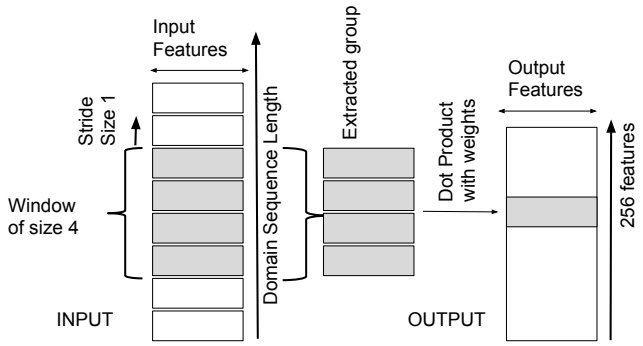


Fig. 1: Feature extraction process of the 1D convolutional layer

By applying the same filter all over the sequence the required computation time is considerable reduced when compared with traditional Multilayer Perceptron layers. Additionally, since a convolutional kernel independently operates on each 4-gram it is possible to go over the entire input layer concurrently. This paralellization and its consequent low computing time is one of the major benefits of using convolutional networks instead of other deep learning approaches usually used for text processing such as Long Short Term Memory (LSTM) [7, 15, 16].

2.3 Dense Layers

The features extracted by both previous layers are be used by traditional Multilayer Perceptron Network (MLP) to output the probability of a given domain belongs to Tunneling or Normal class. MLP is composed of two layers: A first fully connected layer of size hn (*Dense* layer) connected to a second *Dense* layer of size 1 used for actually giving the probability output about the considered domain.

3 Dataset

3.1 A Methodology for Collecting Tunneling DNS Data

A Virtual Machine Infrastructure (VMI) including all the required components for performing a DNS tunneling connection was deployed in order to proceed with the acquisition of DNS tunneling data. In the diagram from Fig. 2 it is possible to recognize the three main components involved in DNS tunneling: (a) the clients, which are a group a (possible) compromised machines inside the local network, (b) a local DNS server in charge of logging and resolving all local clients DNS requests. and (c) a computer outside the local network where DNS tunneling server side is running. In addition, two domain names were registered in order to provide authoritative answers: `harpozedcompute.com` and `securitytesting.online`. Both domains were properly configured for pointing to computer running the DNS tunneling servers.

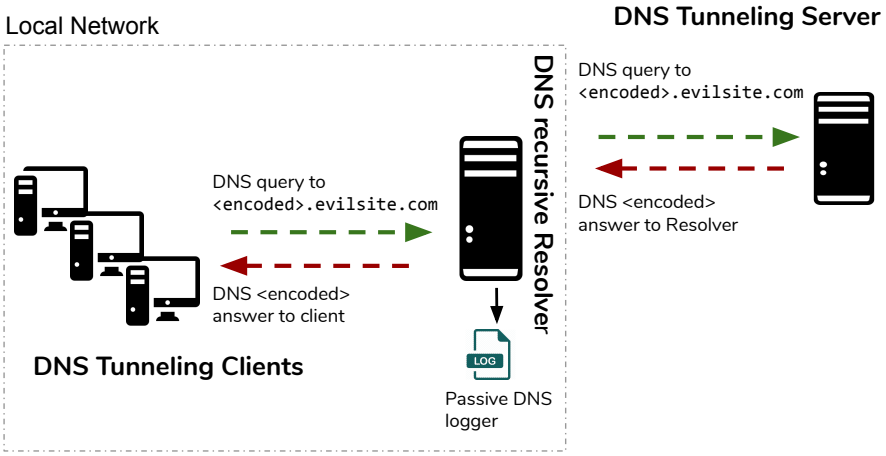


Fig. 2: Virtual Machine infrastructure (VMI) for acquiring DNS tunneling data

The use of VMI facilitates the execution of different DNS tunneling tools under a controlled environment. In this work, five different DNS tunneling tools

were executed inside the VMI. The name along with a summary description of each DNS tools is shown in Table 2. Four out the five tools are oriented to high throughput connections. The only exception is the `DNSexfiltrator` tool which performs sporadic data exchanges using DNS. Certainly, the detection of such low throughput data exchanges results in a considerable more difficult task.

Table 2: DNS tunneling tools executed inside the VMI

Tool	Description	Throughput
`tuns`	IpV4 over DNS tunneling	High
`dnscat2`	C&C oriented DNS tunneling	High
`dns2tcp`	TCP over DNS tunneling	High
`iodine`	IpV4 over DNS tunneling	High
`DNSexfiltrator`	simple data exchange over DNS	Low

A set of Unix shell scripts were implemented for automatizing different common actions performed through the tunnel. In the case of the high throughput tools, the script consisted of a number of ICMP ECHO REQUEST packets and the transfer of files of different sizes. In the case of `DNSExfiltrator`, only the transfer were implemented.

A simple approach for collecting all the tunneling conections inside the recursive DNS logs consists of searching for all requests coming from `harpozedcomput.com` and `securitytesting.online`. In addition, for precisely recognizing the request corresponding to each tool, a time injection strategy [14] was used. Under such strategy, the script execution timestamp is recorded for further timestamp matching with the recursive DNS server logs in the VMI.

Finally, since both `tuns` and `dns2tcp` failed at establishing the tunnel connections, neither file transfer nor ICMP packet were transmitted. However, the repeated failed attempts were still logged.

3.2 Dataset Description

The dataset used in this paper contains both DNS Tunneling and normal domain names.The resulting distribution for both classes can be observed in Table 3.

For DNS Tunneling, 8000 tunnel domains were generated as explained in Section 3.1. Three tunneling tools were discriminated,`dnscat2` [4], `iodine` [8] and `DNSExfiltrator` [5]. Several DNS requests coming from unspecified tunneling tools, as well as failed attempts to establish the tunneling connection, are refered as `Not Specified` in the table.

In order to balance the dataset for both domains classes (i.e. DNS Tunneling and normal), we collected 8000 normal domains that come from the Alexa top one million domains and the Bambenek Consulting feed. Finally, a small portion of the dataset contain commonly accessed domains inside Czech Republic domain

Table 3: DNS tunneling tools executed inside the VMI

Class	Source	Total
Tunneling	`dnscat2`	23
	`DNSexfiltrator`	78
	`iodine`	346
	`Not specified`	7553
	`cz`	511
Normal	`bambenek`	3000
	`alexa`	5000

name servers. The Czech language is particularly difficult to detect given its low vocal/consonant ratio.

The complete dataset detailed in this section was split into two new datasets: the first one, containing 80% of total entries, is used for the training and tuning of the network. The second one, containing the remaining 20% of total entries, is used for the evaluation of the proposed model on unseen domains. Both training and testing datasets can be found in [10].

4 Experiment Design

In this section, we describe the selected metrics and the hyper-parameters fine-tuning methodology for improving the model results.

4.1 Metrics

Several standard metrics are used for evaluating the network. These metrics are **Precision**, **Recall** or True Positive Rate (**TPR**), False Positive Rate (**FPR**) and **F1-Score**. Precision is computed as the ratio of items correctly identified as positive out of total items identified as positive while Recall is computed as the ratio of items correctly identified as positive out of total true positives. The F1-Score is the harmonic mean of Precision and Recall. Finally, False Positive Rate is computed as the ratio between items incorrectly identified as positive and the total number of actual negative predictions.

4.2 Hyper Parameters Tuning

The proposed model possesses many hyper-parameters that need to be tuned. In order to complete the fine tuning task, a traditional Grid Search was conducted. Among all the possible hyperparameters, we particularly focused on finding the optimal values related to the Embedding, Conv1D and Dense layers. For a robust estimation, the evaluation of each parameter combination was carried out using a k-fold cross validation with k = 5 folds. The 1D-CNN was trained using the back propagation algorithm [12] considering the Adaptive Moment Estimation

optimizer [9]. The 1D-CNN training was carried out during 10 epochs. Table 4 shows the parameter combinations with the best performance detection in terms of the F1-Score.

Table 4: Best Hyper-parameters subset. For space reason only the higher average F1-Score parameter combination is shown. This combination was chosen for all the remaining experiments.

avg. F1	**sd**	*nf*	*ks*	*sl*	*d*	*l*	*hn*	**parameters**
0.936820	**0.011219**	**1024**	**4**	**1**	**100**	**45**	**256**	11,425,685

5 Results

The hyper-parameters selected in the previous section were employed for training the network. Then, the model was tested using the explained testing dataset. The experiment results are shown in Table 5 where metrics illustrated in Section 4.1 are calculated with a decision boundary threshold set to 0.90. The support (size of test set) is given in the last column.

The resulting F1-Score around 96% and the low FPR achieved for the two considered classes (i.e Normal and DNS Tunneling), demonstrate that the 1D-CNN detection method was capable of extracting common patterns that are important to discriminating malicious domains (DNS Tunneling) from non-malicious domains.

Table 5: Resulting metrics of the proposed model evaluation on unseen domains

Domain Type	**Precision**	**Recall (TPR)**	**FPR**	**F1-Score**	**Support**
Normal	0.9342	0.9948	0.0762	0.9635	1726
Tunneling	0.9939	0.9237	0.0052	0.9575	1586

In Figure 3 we show the probability of being Tunneling given by the network to each domain in testing set. Dots located on the left correspond to the true Normal domains and dots located on the right are the true Tunneling domains. The decision boundary threshold of 0.90 is plotted as a red line which indicates that all dots above this line are classified as Tunneling domains whereas dots below are classified as Normal domains.

When Tunneling domains are discriminated by the tunneling tool used to generated them, it can be seen that we were able to identify all Tunneling domains generated with `dnscat2` (violet dots), while in the case of `DNSExfiltrator` (red

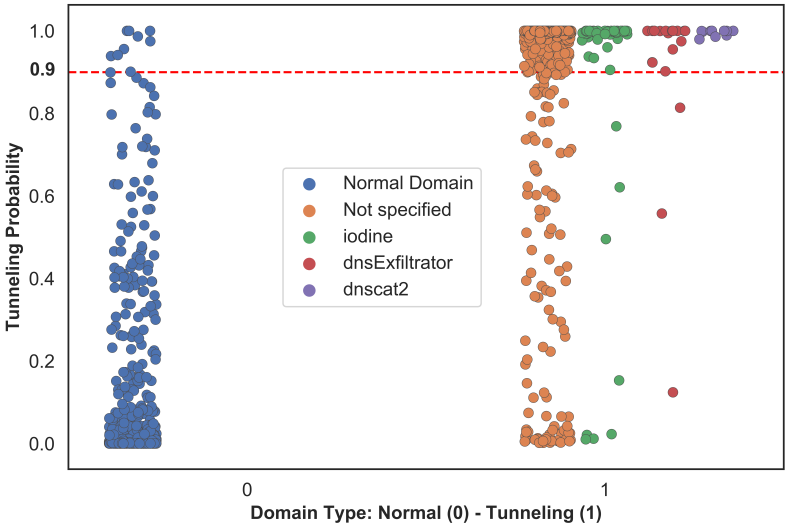


Fig. 3: Tunneling probability explicit for each domain in testing set. Dots corresponding to DNS Tunneling class have been colorized according to their tunneling tool

dots) and `iodine` (green dots), we detected correctly 85% and 87% of these Tunneling domains respectively. There is a group of domains (orange dots) that, as explained in Section 3, failed at establishing the tunnel connection, so they don't have a tunneling tool specified. We detected correctly 96% of these domains, being these results consistent with the overall performance of the network.

6 Concluding Remarks

In the present work, we explored the viability of 1D-CNN for lexicographical DNS Tunneling detection.

A new dataset containing 8000 DNS requests was specially created for evaluating the proposed model on domains generated with 5 common tunneling tools as well as normal domains.

The dataset was properly split into training and testing sets. A hyper-parameters grid search was conducted on training set and the best resulting model was then evaluated on the testing set.

The resulting 1D-CNN was able to detect 99% of total normal domains and 92% of total Tunneling domains, with a FPR of 0.07% and 0.005% respectively. Although its simple architecture, such results make it suitable for real-life networks.

Inspecting the Tunneling tool that were used to generate each malicious domain, we observed that we were able to identify all domains generated with `dnscat2` and around 86% for `DNSExfiltrator` and `iodine`. An analysis of the lexicographical similarities and differences between domains generated with these tools is left for future work. Moreover, since both `tuns` and `dns2tcp` failed at establishing the tunnel connections, domains considered for the analysis of detection rate discriminated by tunneling tool correspond to only 3 different tools. Therefore, it could be necessary an in-depth analysis of how including more tunneling tools affects the performance of the network.

Acknowledgments

The authors would like to thank the financial support received by Universidad Champagnat, CVUT and UNCuyo during this work. In particular the founding provided by the Czech TACR project no. TH02010990 and the PICT 2015-1435 granted by ANPCyT. The authors would also like to specially thank Whalebone s.r.o., whose technical support and help have been fundamental to the complete research process. In addition, we want to gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan V GPU used for this research.

References

1. Antonakakis, M., Perdisci, R.: From throw-away traffic to bots: detecting the rise of dga-based malware. Proceedings of the 21st USENIX Security Symposium p. 16 (2012)
2. Catania, C., Garcia, S., Torres, P.: An analysis of convolutional neural networksfor detecting dga (2018). https://doi.org/10.13140/RG.2.2.11771.16165
3. Das, A., Shen, M.Y., Shashanka, M., Wang, J.: Detection of exfiltration and tunneling over dns. 16th IEEE International Conference on Machine Learning and Applications (2017)

4. Dnscat2: Tool for creating an encrypted command-and-control channel over the dns protocol, https://github.com/iagox86/dnscat2
5. DnsExfiltrator: Data exfiltration over dns request covert channel, https://github.com/Arno0x/DNSExfiltrator
6. Farnham, G.: Detecting dns tunneling. SANS Institute Information Security Reading Room (2013)
7. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (Nov 1997), http://dx.doi.org/10.1162/neco.1997.9.8.1735
8. Iodine: Tool for tunneling ipv4 data through a dns server, https://code.kryo.se/iodine/
9. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. CoRR abs/1412.6980 (2014)
10. Palau, F., et al.: DNS Tunneling Dataset. https://datahub.io/palaufranco12 (2019)
11. Qi, C., Chen, X., Xu, C., Shi, J., Liu, P.: A bigram based real time dns tunnel detection approach. Procedia Computer Science (2013)
12. Rumelhart, D.E., Hinton, G., Williams, R.: Neurocomputing: Foundations of research, chap. Learning Representations by Back-propagating Errors, pp. 696–699. MIT Press (1988)
13. Sammour, M., Hussin, B., Iskandar Othman, M.F.: Comparative analysis for detecting dns tunneling using machine learning techniques. International Journal of Applied Engineering Research ISSN, Volume 12, Number 22 pp. 12762-12766 (2017)
14. Sharafaldin., I., Lashkari., A.H., Ghorbani., A.A.: Toward generating a new intrusion detection dataset and intrusion traffic characterization. In: Proceedings of the 4th International Conference on Information Systems Security and Privacy - Volume 1: ICISSP,. pp. 108–116. INSTICC, SciTePress (2018). https://doi.org/10.5220/0006639801080116
15. Torres, P., Catania, C., Garca, S., Garcia Garino, C.: An analysis of recurrent neural networks for botnet behavior detection (06 2016). https://doi.org/10.1109/ARGENCON.2016.7585247
16. Woodbridge, J., Anderson, H., Ahuja, A., Grant, D.: Predicting domain generation algorithms with long short-term memory networks (2016), http://arxiv.org/abs/1611.00791

Action Recognition in Videos: LSTM Network Parameter Setting

Carlos Ismael Orozco[1], María Elena Buemi[2], and Julio Jacobo Berlles[2]

[1] Departamento de Informática, FCE. Universidad Nacional de Salta, Argentina
[2] Departamento de Computación, FCEyN. Universidad de Buenos Aires, Argentina
ciorozco.unsa@gmail.com, {mebuemi,jacobo}@dc.uba.ar

Abstract. Action recognition in videos is currently a topic of interest in the area of computer vision, due to potential applications such as: multimedia indexing, surveillance in public spaces, among others. In this paper we propose (1) Implement a CNN–LSTM architecture. First, a pre-trained VGG16 convolutional neuronal network extracts the features of the input video. Then, an LSTM classifies the video in a particular class. (2) Study how the number of LSTM units affects the performance of the system. To carry out the training and test phases, we used the UCF-11. (3) Evaluate the performance of our system using accuracy as evaluation metric. We obtain 93% for the best model.

Keywords: Action recognition · Convolutional neural network · Long short-term memory · UCF-11.

1 Introduction

The action recognition problem in videos is of great interest in the area of pattern recognition and computer vision due to its potential applications such as: multimedia indexation, information recovery, patient monitoring and control, automated surveillance in public spaces, among others. The objective of the action recognition systems is to classify each video into the class that represents the action that happens in the video. To this end, the interactions between the subjects and/or objects within it, must be accounted for. This problem has been investigated by other works:

Liu et al. [10] propose a framework for detecting and recognizing human actions. To achieve a robust estimation of the region of interest, they use a combination of optical flow together with a Harris $3D$ edge detector to obtain space-time information from video. Then, with the calculation of the local features SIFT and STIP, they train a universal model background (UBM) for the task in question.

Wang et al. [16] propose a dense trajectory approach. They take dense points in each frame of the video and track based on the optical flow displacement information.

Sharma et al. [13] They propose a model based on attention mechanisms for the task of recognizing actions in videos. They use an LSTM neural network

contemplating the spatial and temporal space of the video.

The objective of this work is to implement a video action recognition system. For this we propose the use of a CNN–LSTM architecture. A convolutional neural network extracts the features of the video while an LSTM neural network classifies the video into a certain category. The work is organized as follows: in the section 2, the general structure of the system is described; in section 3, the database used, the evaluation method, the experiments carried out and the results obtained are described. Finally, in section 4 the conclusions and future work are presented.

2 Our Proposal

In this paper we propose the use of a CNN–LSTM architecture. Figure 1 shows a general scheme of the system in its different stages. Input: the video is normalized for a total of 40 frames. CNN: A pre-trained VGG16 extracts the characteristics of the video by obtaining features of size 40×25088. LSTM: takes each vector from the previous stage and processes it in n_u LSTM units. Finally, the output stage consists of a dense layer with n_c nodes, one for each class. Section 3 shows in detail the implementation proposed.

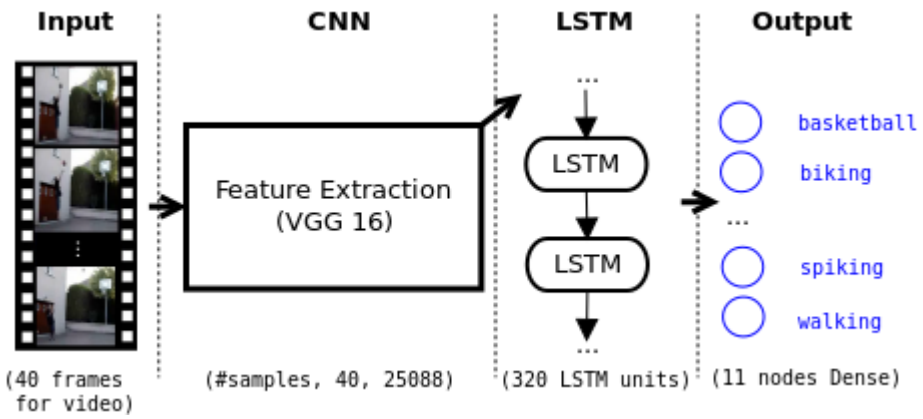


Fig. 1. System architecture. First a CNN VGG16 extracts the features. Then an LSTM classifies a certain class.

2.1 Convolutional Neural Network: VGG16

We use the convolutional architecture of VGG16 proposed by Zisserman et al. [14] because it obtained a very good result in the ImageNet Large-Scale

Visual Recognition Challenge (ILSVRC- 2014) Classification and location tasks. Figure 2 shows the layers that make up the architecture.

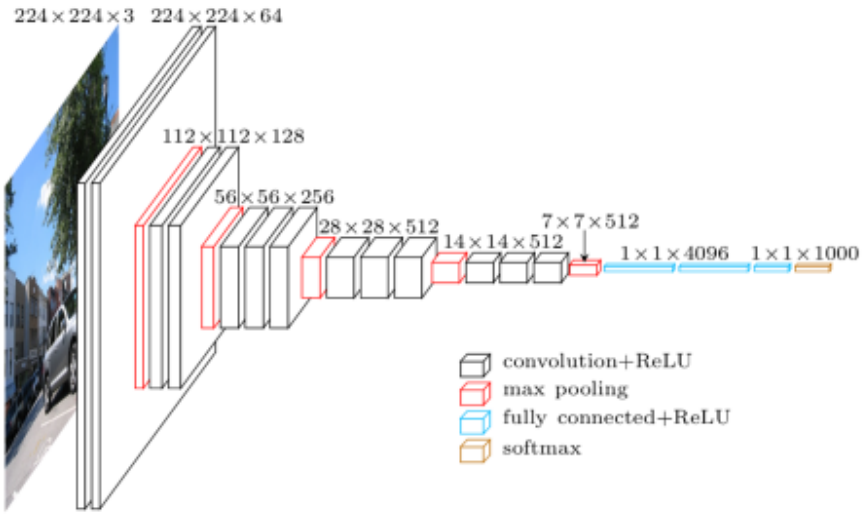


Fig. 2. VGG16 pre-trained implemented in the library Keras [5].

We realized a vectorization of the max pooling of layer 10 obtaining a vector of dimension $7 \times 7 \times 512 = 25088$ to feed the neural network LSTM.

2.2 Long Short-Term Memory

The neural networks LSTMs [7] are a special type of recurrent neural network (RNN) that are formulated in such a way that remembering information for long periods of time is their natural behavior.

We use the LSTM unit proposed by Donahue et al. [6]. Its general structure is shown in Figure 3. The main characteristic of an LSTM unit is a memory cell c which encodes, at each time step, the knowledge of the entries that have been observed up to that moment. The cell is modulated by three types of gates:

1. Input gate (i): Controls whether the current entry is considered (x_t).
2. Forget gate(f): Allows the LSTM to forget the previous memory (c_{t-1})
3. Output gate(o): Decides how much memory will be transferred to the hidden state (h_t)

They are calculated as:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1}) \tag{1}$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1}) \tag{2}$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1}) \tag{3}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \phi(W_{xc}x_t + W_{hc}h_{t-1}) \tag{4}$$

$$h_t = o_t \odot \phi(c_t) \tag{5}$$

Where: σ is the sigmoidal function, ϕ is the hyperbolic tangent, $\odot$ represents the product with the value of the gate and the weights of the matrix denoted by W_{ij}.

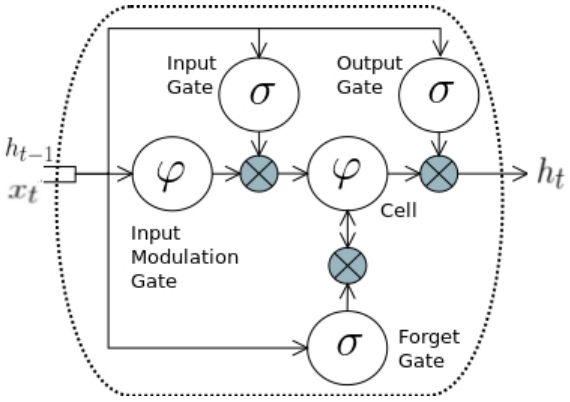


Fig. 3. LSTM unit

3 Experiments and Results

UCF-11 [11] has 1600 videos that belong to one of the following $n_c = 11$ classes: basketball, biking, diving, golf_swing, horse_riding, soccer_juggling, swing, tennis_swing, trampoline_jumping, volleyball_spiking and walking. Performance measure is calculated by average accuracy over all classes, using configuration as the original [11], and we use leave one out cross validation (LOOCV) for a pre-defined set of 25 folds. Figure 4 shows the relationship between number of units and final precision. For $n_u = 320$ LSTM units the obtained precision was 91.94% (best result). Table 1 summarizes the results obtained by our system compared to the systems mentioned in the bibliography. Our result is better than the one proposed by [11,10,16,13,4].

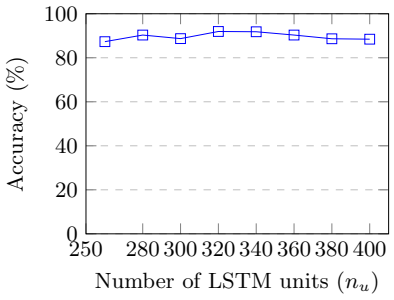


Fig. 4. ROC graphic for the system with varying number of LSTM units against Accuracy.

Approach	**Acc (%)**
Liu et al. [11]	71.2%
Liu et al. [10]	76.1%
Wang et al. [16]	84.2%
Sharma et al. [13]	85.0%
Cho et al. [4]	88.0%
CNN–LSTM (Our proposal)	**91.94%**

Table 1. Results of the video classification (with $n_u = 320$) using the database UCF-11 [11].

4 Conclusions and future work

In this work we implement a video action recognition system, using a CNN–LSTM neural network. First, a VGG16 extracts the characteristics of the video. Then an LSTM neural network classifies the class to which it belongs. It was implemented in Python using the library Theano [2,3], trained and tested using the database [11].

We evaluated the performance of our proposal using the precision evaluation metrics. We obtained $\sim 91\%$ for testing for $n_u = 320$.

As future work we consider the use of other databases, such as Hollywood2 [12], HMDB [8], UCF-101 [15] to make our system more robust. Another goal is to implement the attention mechanisms proposed by [1,9].

Acknowledgement

This work has been supported by Universidad Nacional de Salta (Proyecto CIUNSa A 2364)

References

1. Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *CoRR*, abs/1409.0473, 2014.
2. Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, James Bergstra, Ian J. Goodfellow, Arnaud Bergeron, Nicolas Bouchard, and Yoshua Bengio. Theano: new features and speed improvements. Deep Learning and Unsupervised Feature Learning NIPS 2012 Workshop, 2012.
3. James Bergstra, Olivier Breuleux, Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, Guillaume Desjardins, Joseph Turian, David Warde-Farley, and Yoshua Bengio. Theano: a CPU and GPU math expression compiler. In *Proceedings of the Python for Scientific Computing Conference (SciPy)*, June 2010. Oral Presentation.

4. Jungchan Cho, Minsik Lee, Hyung Jin Chang, and Songhwai Oh. Robust action recognition using local motion and group sparsity. *Pattern Recognition*, 47(5):1813 – 1825, 2014.
5. François Chollet et al. Keras. https://keras.io, 2015.
6. Jeff Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. Long-term recurrent convolutional networks for visual recognition and description. *CoRR*, abs/1411.4389, 2014.
7. Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, November 1997.
8. H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre. HMDB: a large video database for human motion recognition. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2011.
9. Natsuda Laokulrat, Sang Phan, Noriki Nishida, Raphael Shu, Yo Ehara, Naoaki Okazaki, Yusuke Miyao, and Hideki Nakayama. Generating video description using sequence-to-sequence model with temporal attention. In *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*, pages 44–52, 2016.
10. D. Liu, M. Shyu, and G. Zhao. Spatial-temporal motion information integration for action detection and recognition in non-static background. In *2013 IEEE 14th International Conference on Information Reuse Integration (IRI)*, pages 626–633, Aug 2013.
11. Jingen Liu, Jiebo Luo, and Mubarak Shah. Recognizing realistic actions from videos in the wild. pages 1996 – 2003, 07 2009.
12. M. Marszalek, I. Laptev, and C. Schmid. Actions in context. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2929–2936, June 2009.
13. Shikhar Sharma, Ryan Kiros, and Ruslan Salakhutdinov. Action recognition using visual attention. *CoRR*, abs/1511.04119, 2015.
14. K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.
15. Khurram Soomro, Amir Roshan Zamir, and Mubarak Shah. UCF101: A dataset of 101 human actions classes from videos in the wild. *CoRR*, abs/1212.0402, 2012.
16. H. Wang, A. Kläser, C. Schmid, and C. Liu. Action recognition by dense trajectories. In *CVPR 2011*, pages 3169–3176, June 2011.

Incorporar Análisis de Dominio (Joda) al Proceso de Desarrollo Uml Components (DSBC)

Ezequiel Moyano[1], Adriana Urciuolo[1], Matías Gel[1], Rodolfo Iturraspe[1], Emilio Izarra[1], Martín Villareal[1]

[1] Instituto de Desarrollo Económico e Innovación, UNTDF, Yrigoyen 879, (9410) Ushuaia, Argentina. www.untdf.edu.ar
{emoyano, aurciuolo, mgel, riturraspe, mvillarreal}@untdf.edu.ar

Resumen: El desarrollo de software basado en componentes es uno de los mecanismos más efectivos para la construcción de grandes sistemas, sobre todo de gran complejidad. UML Components presenta un gran potencial para construir sistemas en dominios complejos, y permite definir una manera de modelar.

El proceso de desarrollo UML Components debe adaptarse a los fines de proveer una mejor definición del dominio, para representar la complejidad del mismo, se ve la necesidad de completar la etapa de modelado de requerimientos propuesta, incorporando una etapa de Análisis de Dominio en forma previa.

El propósito del Análisis de Dominio es el de proporcionar la reutilización de la especificación de un dominio específico. Una de las técnicas de análisis de dominio que mejor se ajusta a sistemas complejos es JODA.

El presente trabajo incorpora la técnica JODA como primer etapa al proceso de desarrollo de software basado en componentes UML Components.

Palabras clave: Desarrollo de software basado en componentes, Análisis de dominio, Sistemas complejos, Reuso.

Contexto

La línea de investigación se desarrolla en el Instituto de Desarrollo Económico e Innovación (IDEI) de la UNTDF, por docentes-investigadores que llevan adelante proyectos en la temática de Sistemas de Información Ambiental – Hidroinformática (más de 10 años, acreditados por la UNPSJB,previamente, y la UNTDF).

Actualmente esta línea se desarrolló en el marco de un proyecto de investigación (Abr/2017-Mar/2019) – en Sistemas de Información Ambiental. El mismo ha obtenido aval UNTDF.

1 Introducción

Muchos dominios, como los sistemas de información ambiental, se caracterizan por su complejidad [07], con una gran cantidad de subdominios heterogéneos.[05]

Los trabajos de investigación actuales en esta problemática, no se han detenido en el estudio de técnicas de análisis de dominio apropiadas para comprender la complejidad y sus efectos, por cual no se cuenta con modelos del dominio que puedan reutilizarse para distintos sistemas.

La ingeniería de dominio tiene como objetivo lograr la reutilización del conocimiento [09], el Análisis del Dominio permite el estudio de un dominio o un conjunto de aplicaciones[03]. Las técnicas de AD están asociadas a la reutilización, su principal característica es capturar información relacionada con el dominio, y determinar qué y cómo este conocimiento se reutilizará.[04]

Por tanto se constituye en el punto de partida para obtener software reusable, sobre todo en sistemas de gran complejidad, facilitando y permitiendo el reuso de los productos obtenidos para futuros del desarrollo de software.

La necesidad de contar con nuevos sistemas que no demanden grandes periodos de tiempo y esfuerzos motivó la idea de contar con componentes desarrollados que permitieran su reutilización. Esto favoreció el avance de lo que se conoce como Desarrollo de Software Basado en Componentes (CSBD)[01].

El DSBC se ha convertido actualmente en uno de los mecanismos más efectivos para la construcción de grandes sistemas y aplicaciones de software[08].

Existen numerosas técnicas, herramientas y procesos, para la construcción de software basado en componentes. El DSBC se define en diferentes etapas, como las propuestas por RUP (Rational Unified Process) o las propuestas por el proceso de desarrollo “UML Components”, pero ninguna aborda análisis de dominio.

El presente trabajo centra su interés en el Análisis de Dominio y su aplicación en las etapas tempranas de los procesos de Desarrollo de Software Basados en Componentes, como punto de partida para el desarrollo de sistemas que presentan gran complejidad, que permita definir componentes verdaderamente reusables.

Para lo cual se utilizó la técnica JODA incorporada a UML Components como proceso de DSBC.

La contribución principal de JODA es la separación de la ingeniería del dominio de la ingeniería de aplicación y su interacción[03]. La meta de JODA es definir un modelo del dominio que se pueda utilizar para producir objetos reutilizables del software satisfaciendo sus requerimientos [10]. Se escogió JODA, entre otros, por su versatilidad y dado que trabaja con análisis orientado a objetos y notación UML, a los fines de obtener consistencia en el proceso de modelado DSBC.

UML Components, utiliza diagramas de UML y estereotipos para extender UML[02], se basa en un modelo conceptual de la información que existe en el dominio del problema, y adaptable a la técnica de análisis de domino[06], su objetivo principal es capturar los conceptos e identificar las relaciones.

El Proceso UML Component abarca desde la especificación de los requerimientos hasta el ensamblaje, donde cada proceso presenta los objetivos principales.

2 Líneas de Investigación y Desarrollo

La línea de investigación general se enfoca en el estudio y desarrollo para incorporar técnicas de análisis de dominio al proceso de desarrollo basado en componentes en dominios complejos. Para tal fin se utilizó la técnica de Análisis de dominio JODA al Proceso UML Components en su primer estapa.

JODA divide el proceso del análisis del dominio en tres fases [03]: Preparación del dominio, Definición del dominio y Modelado del dominio.

Uno de los principales problemas observados en los procesos de DSBC, es que las técnicas de especificación y modelado de requerimientos no contemplan métodos de análisis del dominio. Se ve la necesidad de completar la etapa de modelado de requerimientos de UML Components, incorporando técnicas de análisis de dominio en forma previa. Esto permitó una mejor definición del dominio y lograr comprender sus aspectos comunes, para representar el conocimiento y permitir su reutilización.

Primero se redefinió el proceso de Requerimientos de UML Components, incorporando el modelado del dominio como entrada, y redefinir el proceso de especificación de requerimientos, considerando las etapas definidas en la técnica JODA para el modelado del dominio.

Posteriormente se generó un modelo basado en UML para la etapa de modelado del dominio incluida al Proceso de definición de Requerimientos, según las fases que define JODA.

3 Resultados

Para validar el proceso se definió como dominio específico, que contenga las características de un dominio complejo, los Sistemas de Información Meteorológicos Los fenómenos físicos en la atmósfera ocurren en todas las escalas espaciales y temporales y sus impactos son relevantes para muchas actividades.[14].

Se trabajó con expertos del dominio para obtener el conocimiento necesario. Siguiendo el proceso de UML Components, se especificó el modelo de casos de uso de un sistema genérico del dominio y su respectiva descripción, y se definieron los componentes conceptuales del dominio, sus interfaces y la interacción entre los componentes.

Para la validación respecto a que los componentes obtenidos a través del análisis previo sean verdaderamente reutilizables se utilizaron dos aplicaciones del dominio específico. Un *Sistema de Alertas*: cuya finalidad es advertir fenómenos meteorológicos que se producen bajo ciertas circunstancias (inundaciones, temporales, sequías, granizo, etc). Y un *Sistema de Modelización de Caudales:* Su finalidad es conocer los caudales de un río para el usufructo del agua (agua potable, turismo, acuicultura, riego, etc.) Del análisis requerimientos de cada sistema, los mismos se pueden satisfacer a través de los componentes especificados.

Como conclusión a partir de aplicar Análisis de Dominio a UML Components permitió obtener un proceso de desarrollo (sobre todo en la etapa de especificación de requerimientos) más detallado, preciso y formalizado, extendiendo UML Components a sistemas de mayor complejidad[13].

Utilizar la notación UML en la técnica JODA, permitió definir un número de modelos y diagramas para la especificación del software con varios grados de

abstracción, con lo cual se obtuvo un modelado del dominio más semejante a la realidad. El presente trabajo permite concluir que los componentes que se obtienen del proceso son verdaderamente reutilizables, aplicando debidamente sus interfaces.

Los resultados mencionados constituyen aportes al conocimiento del dominio, sus características y avances en el desarrollo de una arquitectura referencial.

Como futuros trabajo en el marco de este estudio se espera poder validar otras aplicaciones dentro del dominio, y ampliar el mismo a otros subdominios dentro de las aplicaciones ambientales, a los efectos de encontrar más entidades y procesos comunes modelados en el dominio.

4 Formación de Recursos Humanos

El Equipo de Trabajo es multidisciplinario; conformado por docentes investigadores de la UNTDF y expertos del dominio (hidrólogos). En esta línea de investigación se encuentran en desarrollo tres tesis de (Magister en Ing. de Software, UNLP, Ezequiel Moyano y Martín Villarreal; Magister en Sistemas Inteligentes España, Matías Gel). Una tesis de Doctorado en Cs. Informáticas en UNLP de Adriana Urciuolo.

5 Bibliografía

1. Bertoa M.F., Troya J.M. y Vallecillo A.: Aspectos de Calidad en el Desarrollo de Software Basado en Componentes, 2002.
2. Cheesman J., Daniels J.: UML Components, 2000.
3. Holibaugh R.: Working Group (JIAWG) Object-Oriented Domain Analysis Method (JODA)
4. Harsu M.: A Survey On Domain Engineering, Institute Of Software Systems Tampere University Of Technology.
5. Koschel A., Kramer R., Nikolai R.: A Federation Architecture For An Environmental Information System Incorporating GIS, Germany, 2000.
6. Estévez E.: Desarrollo De Software Basado En Componentes, Neuquen 2003.
7. Günther O.: Environmental Information Systems, Berlín Heidelberg, Germany, 1998.
8. Iribarne Martínez L.F.: Tesis Doctoral Un Modelo De Mediación para el Desarrollo de Software Basado En Componentes COTS, Universidad De Málaga, 2003.
9. Hamar V.: Aspectos Metodológicos del Desarrollo y Reutilización de Componentes de Software, Mérida, 2003.
10. Regis E., Tondello G.: Análise de Domínio Pesquisado, Universidad e Federal De Santa Catarina,2000.
11. Urciuolo A., Iturraspe R., Moyano E. Perfil UML 2.0 para Aplicaciones de Monitoreo Ambiental, VI JIISIC'07, pp. 393--401, ISBN 978-9972-2885-1-7, Lima, Perú..
12. Urciuolo A., Iturraspe R., Conceptual Patterns for Water Resources Information Systems, Journal of Computer Science and Technology pp 20—26, ISSN: 1666-6038,2003.
13. Urciuolo A., Iturraspe R., Moyano E., Incorporar Análisis de Dominio (Joda) en Uml Components (DSBC) WIIC 2019 – San Juan, 2019.
14. Servicio Meteorológico Nacional https://www.smn.gob.ar

Uso del Estándar IEEE-830 en la Especificación de Requerimientos para el Manejo Sostenible de Cuencas en TDF

Ezequiel Moyano[1], Adriana Urciuolo[1], Matías Gel[1], Rodolfo Iturraspe[1], Emilio Izarra[1], Martín Villareal[1], Jhonatan Vieira[1]

[1] Instituto de Desarrollo Económico e Innovación, UNTDF, Yrigoyen 879, (9410) Ushuaia, Argentina. www.untdf.edu.ar
{emoyano, aurciuolo, mgel, riturraspe,eizarra,mvillarreal,}@untdf.edu.ar, jhonatanvieira@live.com.ar

Resumen: En el marco de un proyecto de investigación sobre el manejo de cuencas en Tierra del Fuego, con el propósito de generar un Sistema de Información socio-ambiental de las cuencas de estepa de la Provincia, se realizó un análisis de requerimientos de información y se obtuvo un documento como resultado de la Especificación de Requerimientos (ERS) de acuerdo a la última versión del estándar IEEE 830.
En el cual se presentan los principales requerimientos funcionales y no funcionales, en el marco de los avances de un proyecto de investigación, en base a las características propias que presentan los modelos ambientales, a los fines de obtener una descripción completa sobre el estudio de las cuencas de estepa de la Provincia de Tierra del Fuego.

Palabras clave: Especificación de Requerimientos, IEEE 830, Sistemas de Información Ambiental.

Contexto

La línea de investigación se desarrolla en el Instituto de Desarrollo Económico e Innovación de la UNTDF, por docentes-investigadores que lleva adelante proyectos en la temática de Sistemas de Información Ambiental desde hace más de 15 años.

Actualmente esta línea se desarrolla en el marco de un proyecto de investigación denominado: "Bases para el manejo sostenible de las cuencas de la estepa de la Provincia de Tierra del Fuego" avalado por la UNTDF según Resolución Rectoral N° 60/2017 en el período Mayo 2017 a Julio 2019.

El Proyecto se lleva adelante con un enfoque multidisciplinario por parte de organismos de investigación y gestión en la provincia de Tierra del Fuego, Dirección de Recursos Naturales, Docentes e Investigadores de IPCA (Instituto de Ciencias Polares y Ambientales de la UNTDF) y del IDEI.

1 Introducción

1.1 Ingeniería de Requerimientos y Especificación de Requerimientos

La ingeniería de requerimientos es la primera, y más importante, etapa de todo paradigma de ingeniería de software, según Brooks *"La parte más difícil de construir un sistema de software es decidir qué construir"*.

La especificación de requerimientos representa la comprensión de las necesidades, y expectativas de los interesados y/o expertos del dominio (stakeholders). Los requerimientos de un sistema de software determinan lo que el sistema debe hacer y define las restricciones de su operación e implementación.

Los beneficios de obtener buenos requerimientos incluyen acuerdos junto a los stakeholders de la tarea a ejecutar y del criterio de aceptación del sistema concluido.

Las dificultades que se plantean al obtener los requerimientos para un sistema de Información (SI), radican fundamentalmente en la adquisición del conocimiento del dominio y los problemas que se presentan en la captura de la información. Los requerimientos tienen solamente un sentido en el ámbito donde se elicitan y se ubican en el dominio de la aplicación donde está el problema.

Estudios realizados concluyen que los proyectos fallan porque sus requerimientos no tuvieron una correcta especificación o fueron inadecuados[2].

La ingeniería de requerimientos es un proceso sistemático de especificar requerimientos[1], en tres fases específicas: Elicitación (obtener el conocimiento del dominio o problema), Especificación (documento que refleja un acuerdo entre los stakehokders y analistas) y la Validación (Certifica la consistencia del modelo de requerimientos con las intensiones de clientes y usuarios).

Fig 1. Kotonya y Sommerville [4]

Fig 2.Loucopoulos [1]

La ingeniería de requerimientos es un área crítica en el desarrollo de software y particularmente lo es la especificación de requerimientos.[6]

1.2 Estándar IEEE-830

La IEEE define los requerimientos como:[3]

1. Condición o capacidad que necesita el usuario para resolver un problema o alcanzar un objetivo.
2. Condición o capacidad que debe satisfacer o poseer un sistema , un contrato, un estándar, una especificación u otro documento formalmente impuesto.
3. Representación documentada de una condición o capacidad como en 1 o 2.

El estándar IEEE 830-1998 para el SRS (en inglés) o ERS (Especificación de requerimientos de software) es un conjunto de recomendaciones para la especificación de los requerimientos de software el cual tiene como producto final la

documentación de los acuerdos entre los stakeholders y el grupo de desarrollo para así cumplir con la totalidad de exigencias estipuladas[3].

Según IEEE un buen Documento de especificación de Requerimientos, pese a no ser obligatorio el estándar 830, deberá incluir toda la información presentada en dicho estándar. Si bien el estándar de IEEE 830 no está libre de defectos u observaciones, es una descripción completa del comportamiento del sistema de información que se va a desarrollar; y también define los requerimientos no funcionales como las reglas del negocio o dominio[5]. Altamente aceptado en el ámbito. Considerando que los sistemas de información ambiental presentan características particulares, obtener una especificación de requerimientos posibilitará obtener el sistema de información adecuado.

1.3 Sistemas de Información Ambiental

Los sistemas de información ambiental se caracterizan por su complejidad y tiene como objetivo recoger, sistematizar y actualizar datos ambientales y entender su significado. Estos sistemas se relacionan con el manejo de los datos de distintos componentes del ambiente: el suelo, el agua, el aire y las especies existentes[7], y cuentan con cinco dimensiones: su localización en el espacio (latitud, longitud y altitud), el tiempo en que se obtiene y los fenómenos particulares que son analizados[8].

La complejidad de estos sistemas está dada por la necesidad de manejar grandes volúmenes de datos, objetos espacio / temporales, factores ambientales, políticas de gestión, normativas legales y técnicas[10].

Las cuencas de estepa de la zona Norte de la Provincia de Tierra del Fuego, presentan características biofísicas particulares que determinan una alta fragilidad ambiental y requieren de estudios específicos para su adecuado manejo[13]. Existen escasos estudios que permitan conocer el potencial hídrico de estas cuencas, y los efectos que las diferentes actividades realizadas provocan[14].

2 Líneas de Investigación y Desarrollo

La línea de investigación se enfoca en la estándar de la ERS de la IEEE-380, el presente trabajo muestra la especificación de requerimientos obtenida para generar un Sistema de Información socio-ambiental de las cuencas de estepa de la Provincia, que permita la toma de decisión a los responsables de su manejo utilizando el estándar IEEE – 830.

Si bien el tema central está en la especificación de los Requerimientos Funcionales y No funcionales, se completaron todos los apartados del estándar: el Propósito, Alcance, Resumen, Perspectiva del producto (el sistema de información será un producto diseñado para trabajar en entornos WEB, lo que permitirá su utilización de forma rápida y eficaz, además se integrara conjuntamente con otros productos software de diversa naturaleza como GIS, Google Earth, etc.). La funcionalidad se modelo a través de casos de uso y diagramas de actividad.

Se caracterizaron los diversos usuarios intervinientes, principales Restricciones: Plataforma Web, Lenguajes: HTML, PHP o Paython.

Los Requerimientos Funcionales y No Funcionales, se especificarón siguiendo los siguientes ejemplos.

Ident. Req.:	RF001
Nombre:	Identificar y almacenar cuencas
Descrip Req.:	El sistema debe identificar cada cuenca y almacenar la información.
Entrada	Identificación de la cuenca, nombre, zona geográfica y características.
Prioridad del requerimiento: Alta	

Ident. Req.:	RF002
Nombre:	Manejo de secciones de una cuenca.
Descrip Req.:	El sistema debe obtener y manejar diversas secciones de una cuenca.
Entrada	Ident. de la cuenca, ident. de la sección, nombre, zona, longitud, etc.
Prioridad del requerimiento: Alta	

Ident. Req.:	RNF002
Nombre:	Tiempos de respuestas.
Descrip Req.:	El sistema debe realizar el 90% de las búsquedas en no más de 10 seg. y el resto no más de 15 seg.
Entrada	Infraestructura.
Prioridad del requerimiento: Media	

Ident. Req.:	RNF003
Nombre:	Patrones de navegación.
Descrip Req.:	El sistema usará patrones de navegación (entorno WEB) multi-plataforma.
Entrada	Infraestructura.
Prioridad del requerimiento: Alta	

3 Conclusiones

El uso del formato de la IEEE-830 permitió obtener una Especificación de Requerimientos organizada, completa y detallada del Sistema de Información Ambiental en estudio (cuencas de TDF); resultando sencilla su organización tanto para los expertos del dominio como los profesionales de sistemas e investigadores.

A partir de los Requerimientos obtenidos se espera poder realizar el desarrollo de aplicaciones de visualización Web del impacto ambiental de las cuencas como resultado final del proyecto de investigación en que se encuentra incluido.

Los resultados mencionados constituyen aportes al conocimiento del dominio, sus características y avances en el desarrollo de una arquitectura referencial.

5 Bibliografía

1. Loucopoulos P., Karakostas V.," System Requirements Engineering", 1995, London.
2. Pesado P, Feierherd G, Pasini A. Especificación de Req. para Sistemas de Voto Electrónico
3 IEEE https://standards.ieee.org/standard/830-1998.html.
4 Katonya, G., Sommerville, "Requirements Engineering Processes and Techniques", 1998.
5 Raymond Turner. "The Foundations of Specification".Vol. 15, No. 5, 2005, pp. 623–663.
6. Rossi G, Antonelli L, Sampaio JC, Oliveros A, Buenas prácticas en la especificación del dominio de una aplicación, WER 2013.
7. Günther, O., Environmental Information Systems, Berlín Heidelberg, Germany, 1998.
8. Moyano E, Urciolo A, Iturraspe R, Una Metodologia para Requerimientos Referenciales en dominios complejos. WICC 2019, San Juan.
9. Iribarne Martínez L, Un Modelo de Mediación para el Desarrollo de Software Basado en Componentes COTS, Universidad De Málaga, 14 De Julio De 2003
10. Urciuolo A. Sistemas De Hidroinformática Oo + Soluciones Evolutivas 2005.
11. Rol Of Information Technology In Environmental Information System, 2003
13. Urciuolo A., Iturraspe R., Iturraspe R.(h), Noir G.,2015. Modelación de servicios ecosistémicos de la cuenca del Río Grande. XXV Congreso Nacional del Agua, Paraná.
14. Lofiego R., Urciuolo A., Iturraspe R.,. Evaluación de la situación ambiental del estuario del Río Grande de Tierra del Fuego. Actas del XXII Cong. Nac. del Agua, 2009 Trelew.

CRITERIOS PARA EL ANÁLISIS PRELIMINAR DE DATOS EXTRAÍDOS MEDIANTE WEB SCRAPING

Roxana Martínez [1], Christian Parkinson [1], Rocío Rodríguez [1], Pablo Vera [1]

[1] Centro de Altos Estudios en Tecnología Informática (CAETI).
Universidad Abierta Interamericana (UAI), Ciudad Autónoma de Buenos Aires, Argentina
{Roxana.Martinez; Christian.Parkinson; Rocioandrea.Rodriguez;
Pablomartin.Vera;}@uai.edu.ar

Abstract. Al extraer datos por medio de técnicas de raspado de datos desde los sitios web, es necesario el control de los mismos para cotejar los datos originales versus los obtenidos como resultado final. El tratamiento de extracción de los datos exige pensar en la calidad de la información, por lo que es importante el enfoque de técnicas para la verificación y la validación de éstos. Algunos tipos de validaciones de datos pueden ser más eficaces que otros. Una vez que son validados los datos, se puede dar inicio al proceso de ingesta de datos, el cual introduce diversos datos de diferentes fuentes, estructuras o bien características en un procesamiento de datos para su reutilización. Este trabajo presenta una propuesta de las mejoras prácticas de validación de datos en raspado web.

Keywords: validación de datos, scraping, minería de datos

1 Introducción

El método de web scraping es una técnica utilizada mediante programas de software para extraer información de sitios web. "Usualmente, estos programas simulan la navegación de un humano en la World Wide Web ya sea utilizando el protocolo HTTP manualmente, o incrustando un navegador en una aplicación" [1]. Esta técnica permite la extracción de datos de cualquier sitio web, armando previamente un mapa de los datos que se desean extraer. Los datos extraídos son exportados en un formato conveniente para su análisis permitiendo luego cruzar información de diversas fuentes.

"Para hacer web scraping es necesario analizar aspectos como: accesibilidad de los datos de origen, análisis de patrones de los datos, frecuencia de extracción de los datos con el objetivo de buscar la vía más óptima para obtener los datos" [2].

El análisis preliminar de los datos es muy importante, sin ir más lejos, se debe tener en cuenta el concepto de ingesta de datos, que es el proceso mediante el cual se introducen datos, de diferentes fuentes, estructura y/o características dentro de otro sistema de almacenamiento o procesamiento de datos. "La administración de la ingesta permite el control de los datos ingresados, de dónde provienen, cuándo llegan y dónde se almacena" [3].

Web scraping tomará los datos de entrada desde la web, en donde ya han sido publicados, es por ello que el proceso deberá enfocándose en los datos que serán recolectados de la web, para lo cual resulta necesario que las fuentes a las que se recurre para extraer la información

sean confiables. Es por ello que los datos deben ser validados previamente a su publicación. La validación es una verificación automática para garantizar que los datos sean confiables, completos y fiables, si bien la validación no puede garantizar la precisión de los datos, puede lograr reducir la probabilidad de tener datos sucios. "Un dato sucio se considera, por ejemplo, cuando existe ausencia de un valor, que puede ser la cantidad o el precio de un producto. Otro caso de datos sucios se puede presentar en el campo dirección, haciendo un uso inapropiado [4]".

"Después de que los datos se bajan de la web el proceso de extracción continua para parsear, reformatear y reorganizar los datos en una forma estructural"[5].

1.1 Escenarios de Web Scraping

Existen distintos escenarios en los cuales se puede extraer datos que sean relevantes y monitorear sus cambios, entre la información a monitorear en [5] se enumera: Contactos, Monitoreo y Comparación en el cambio de precios, Recolección de revisión de productos, Listado de Propiedades en Venta, Monitoreo del Clima, Detectar cuando cambia un sitio web. En un nivel macro los sitios son constantemente raspados para obtener información los buscadores.

En este artículo se extraerá información de sitios web gubernamentales lo cual permitirá ofrecer al ciudadano un repositorio centralizado con información estructurada y organizada que ponga a disposición los datos abiertos. El enfoque del presente artículo está puesto en ofrecer criterios para validar la información obtenida con web scraping y ponerla a disposición de los ciudadanos.

1.2 Validación de Datos

La etapa de la validación de datos es un proceso que se centra en entregar finalmente, los datos ingresados como limpios, consistentes y claros a los diversos softwares y servicios que van a consumirlos. Parte de este paso, implica la comprobación de la integridad y validez de los datos.

Dicha validación permite garantizar que se cumplan los requerimientos a nivel objetivos y estándares básicos de calidad del dato. Este paso también es conocido como validación de entrada y es una parte fundamental del procesamiento de datos.

La gestión de la calidad de datos implica un conocimiento de los componentes de la calidad de datos. Se pueden utilizar para asegurar la limpieza y la exactitud de la información.

La calidad de los datos normalmente implican los siguientes servicios [6]:

a) Análisis de datos: es la separación de datos y su análisis en un formato determinado;
b) Estandarización: determinar cuáles datos se sitúan en cada campo y aseguran que se almacene en un formato estándar.
c) Validación: asegurarse que los datos sean consistentes. Se pueden agregar validaciones de campos cruzados como verificaciones de códigos y demás.
d) Verificación y Comprobación: se comprueba la fuente de los datos, con el fin de asegurar que los datos son válidos.

e) Coincidencias (matching): se identifican los registros duplicados y mezclas de registros en forma correcta.

La validación de datos permite garantizar que los datos que alimentan a las aplicaciones puedan ser precisos, completos y consistentes. Este paso implica que se apliquen correctamente los controles de validación y reglas de datos. Por lo general, las reglas de datos con su tipo de dato, descripción, dominio y demás, se definen en un diccionario de datos o bien en la propia validación que incluye un software.

2 Técnicas de Validación de Datos

La importancia en hacer foco en la calidad de datos es un tema fundamental a la hora de extraerlos desde un sitio web. Éstos por sí solos no aportan una información puntual, sino que se deben realizar tareas de análisis y validación para llegar al objetivo de la obtención de información válida y rica en decisiones. "La posesión de grandes volúmenes de datos no es siempre valiosa en sí misma, los datos adquieren un valor real cuando son analizados y explotados con el objetivo de extraer conocimiento de los mismos" [7].

Uno de los puntos a tener en cuenta es la validación de los datos que se extraen, debido a que incorporan técnicas que permiten que los datos sean seguros y confiables. "La seguridad, veracidad de los datos, ofrece confianza de los mismos. La confianza se establece por comparación de los datos recogidos con repetidas medidas obtenidas por distintas personas o en distintos tiempos o lugares. La utilidad de los registros depende básicamente de la calidad de los datos recogidos. Si los datos son de pobre calidad los resultados obtenidos con ellos pueden estar sesgados e inciertos" [8].

Los datos extraídos se pueden validar mediante técnicas de depuración. Si se presenta una interfaz de depuración, se puede visualizar los datos extraídos en una aplicación desarrollada y mostrar las diferencias entre los datos de la fuente original y los datos extraídos desde el sitio web, se propone que este contenido puede ser validado y verificado de dos formas posibles:

- Método Sincrónico: Se realiza una validación del tipo sincrónico, esto es, a través del desarrollo de un software, se puede mostrar toda la extracción de los datos en un sitio web, y a través de la aplicación web, se puede validar cada dato. A medida que se va confirmando un campo como validado, se puede avanzar con otro; caso contrario no se podrá proseguir. Es decir, este método, permite que la validación sea anidada, por ejemplo: un campo "Subtítulo", debe ser validado, antes de analizar y confirmar la validación de un campo "Descripción del Subtítulo". De esta manera, se van depurando los contenidos en forma cuasi conjunta de datos extraídos o raspados de la web.
- Método Asincrónico: Se realiza una validación del tipo asincrónica a través de un software en el cual, se muestra toda la extracción de los datos desde un sitio web, y a través de una interfaz de usuario, se va depurando campo por campo de manera independiente, los contenidos de datos extraídos o raspados de la web. Por ejemplo, los usuarios de la interfaz tienen que corregir los errores de los campos visualizados hasta finalizar con la depuración, sin importar el orden en que se muestran ya que se identifican por registros extraídos. Este método, permite resolver los errores en cualquier orden, antes de guardar y confirmar el contenido final obtenido.

3 Trabajos Relacionados

En la Tabla 1 se muestra el análisis de la comparación de diferentes trabajos relevados. Para ello se toma en cuenta el tema del trabajo, el objetivo, el raspado realizado y las validaciones de datos realizadas de los datos extraídos.

Tabla 1. Trabajos relacionados

Trabajo	***Objetivo***	***Raspado***	***Validación de datos***
Metodología para el análisis de la demanda laboral [9]	Solución a la falta de información sobre los requerimientos cuantitativos y cualitativos	Recolección de información es el web-scraping, utilizando un algoritmo de recolección que contiene, además, un proceso que valida la información capturada, y que hace énfasis en el tipo de dato recogido, en la longitud y otros parámetros con los que se pueden identificar errores en los datos recolectados, para posteriormente exportarlo en formato XML y almacenarlo en una base de dato	El mecanismo de validación consiste en eliminar datos duplicados, dado que los mismos están suscriptos a una URL específica. Una vez almacenados los datos pasan por un proceso de homogeneización de los mismos, que consiste en la ordenamiento y organización de la información referida a un anuncio laboral provista por un empleador, para facilitar optimizar el proceso selección y búsqueda en función del hallazgo de patrones de búsqueda (habilidades, competencias, remuneraciones pretendidas, etc.)
Validación de Propuesta para medir la Interactividad en un ambiente de aprendizaje [10]	Creación de un modelo que permita medir la interactividad con los contenidos provistos en un ambiente de aprendizaje	Para la medición se utilizó el método Scraping o raspado, el cual consiste en utilizar la estructura de la aplicación web, en su conjunto llamado DOM. Esta extracción se realizó sobre algunas de las	La limpieza de los datos se realizaron de forma manual comparando la información resultante del raspado con el seguimiento de los participantes.

		aplicaciones de Erudito[11], de allí se guarda el código en estructuras de almacenamiento como variables, arreglos, etc.	
Técnicas de minería de datos para la detección de la visibilidad web de empresas españolas [12]	Desarrollo de un modelo de análisis cibermétrico (análisis basado en métricas de sitios web) orientado al análisis sectorial del mercado online español para su internacionalización	Utilización de un programa de raspado determinando una columna url contiene los datos de la url1 y de la url2, pero con el siguiente formato: <página web internacional> site: "<página web nacional>", e integración en la base de datos de MySQL, además, otra columna hit_number contiene el número de nombramientos de la url1 en la url2 y no precisa tratamiento previo	Realización de análisis estadísticos de datos, mediante la creación de scripts en el lenguaje de programación R
Estudio del estado del arte de las herramientas informáticas para la gestión de conocimiento externo a una empresa [13]	Estudio de las diferentes fases que atraviesan las organizaciones para la manipulación de la información	Análisis y descripción de diferentes herramientas informáticas para la recolección de datos, donde se desarrollan las diferencias entre las herramientas de raspado web.	Se presentan y detallan diferentes herramientas para almacenar, clasificar y realizar búsquedas de información o conocimiento a partir de los macrodatos ya validados y depurados
Uso ético de la información [14]	El objeto de la investigación es la experimentación de la técnica de web scraping en buscadores. En concreto se ha elegido el buscador Google por ser el más relevante y popular, así como su versión académica Google Scholar	Descarga del código fuente de la página web para crear una plantilla DOM. Esto permite almacenar temporalmente el código HTML de la página y estructurar su contenido mediante nodos, identificados como etiquetas embebidas y jerarquizadas.	Creación scripts que permitieran rastrear los contenidos de los buscadores y generar bases de conocimiento especializadas. Inspección de elementos y atributos mediante xpath. Comprobación de la integridad del código fuente del sitio descargado.

4 Propuesta de Variables para Validación

La verificación de los datos puede realizarse sobre datos de una base de datos en forma directa, es decir, no es necesario recurrir a ningún otro elemento. Sin embargo, el procedimiento de validación requiere necesariamente aprobación del usuario, por lo que se presenta un nivel de complejidad que lleva a un proceso, en la que hay que conciliar la información que se obtiene de fuentes externas, y así, lograr evaluar los datos en base a determinados aspectos como ser:

a) Criterios objetivos: son los que responden a la pregunta sobre si un campo de datos, realmente representa ese aspecto del mundo real que se destina al modelo;

b) Criterios subjetivos: son los que se utilizan para la evaluación de la idoneidad de los datos para un propósito determinado.

Es necesario presentar una herramienta correcta para la realización del proceso de validación de datos, y además, es importante tener personas dedicadas a la tarea de obtención válida del resultado final, es decir, parte de estas personas pueden contribuir a un mejor análisis de la información con el fin que sea confiable.

Algunos ejemplos de este tipo de personas son los usuarios que van a utilizar la información, ya que su aporte es en forma directa a la validación de los datos en todos los aspectos de forma subjetiva. Por otra parte, se tiene a los dueños de los datos, quienes pueden aportar criterios de valoración objetiva específica.

4.1 Buenas Prácticas

Algunas buenas prácticas a tener en cuenta para este proceso son:

- Desarrollar un proceso comparativo de medición entre la fuente original y la migrada.
- Juicio de expertos: es la opinión de la persona que está validando el dato, se debe tener en cuenta como una buena técnica de validación.
- Valoración en la validación: La validación debe pensarse como algo más elaborado que una mera comparación de los valores de datos. Parte de las transformaciones de los datos mismos, deben ser validadas para garantizar que el resultado tiene la calidad esperada en los datos de resultado final.
- Definir calidad a medir de los datos.
- Diseñar casos de pruebas especiales. Por ejemplo: Extraer un conjunto determinado de datos y validar el resultado final.
- Complementar con la utilización de herramientas de validación de datos, como ser: Microsoft Excel, el cual posee opciones de validación en sus celdas; Herramientas de base de datos con procesos internos de ETL. “La integración de las fuentes de datos se logra mediante el uso de procesos ETL (Extraer, Transformar y Cargar). Por lo tanto, se reconoce ampliamente que el diseño apropiado de los procesos ETL son factores clave” [15].

4.2 Criterios para la validación de datos

Se plantea como propuesta del presente artículo, en base a la experiencia adquirida en el área y a los trabajos académicos relevados en el apartado anterior, 9 criterios que serán los pilares para la validación de los datos. Estos criterios se describen en la tabla 2.

Tabla 2. Criterios para la validación de datos.

Criterio	***Descripción***
Integridad	Validar que la información esté correctamente relacionada. Por ejemplo: El dato de Subtítulo, que se encuentre correctamente relacionado con la descripción de éste.
Existencia	Validar que el dato extraído contenga un valor.
Formato	Validar que el formato extraído sea un formato válido para su representación como objeto del mundo real. Por ejemplo: que las extracciones de los datos DNI, sean sólo números, ya que ese tipo de campo, no debe permitir caracteres alfanuméricos.
Consistencia	Verificar que los datos sean coherentes entre sí y tengan un sentido lógico.
Presición	Validar el dominio que es el conjunto de valores posibles que puede adoptar el campo. De esta forma, se asegura que el contenido del campo es adecuado.
Duplicidad	Validar que no exista redundancia de los datos para ese registro extraído.
Relevamiento	Evitar las anomalías en los datos, es decir, identificar valores por defecto e inválidos, para ello es necesario relevar reglas de negocio para tener veracidad de los datos en los contenidos de los campos trabajados, ya que éstos se tomarán como criterios de análisis.
Estandarización	Entender el conocimiento del dato desde la óptica lógica, es decir, identificar reglas específicas para algunos datos, como ser: nombres y apellidos, direcciones o bien formatos de fechas. Básicamente se apunta a la normalización de la forma de escribir los datos.
Data maching	Puede usarse para comparar los datos y luego definir un porcentaje de aceptación que se considere válido. Es decir, se puede indicar una política de unificación que marque las instrucciones a seguir en este proceso.

5 Casos de prueba en base a parámetros definidos

Existen diversas herramientas para la extracción de datos, las cuales han sido analizadas también por los autores del presente artículo, dando como resultado una publicación previa [16], en la que se elige "*Import.io*" [17] por sobre las otras herramientas analizadas. Esta herramienta ha resultado ser la más adecuada, para llevar a cabo un trabajo de limpieza de datos de un nivel medio, y que posee una buena identificación de los registros extraídos para su reutilización y reconocimiento como datos abiertos a la reutilización de otras plataformas. Es por ello que se decide realizar utilizando esta herramienta pruebas de raspado de datos para 3 casos de estudio a modo de ejemplo. Una vez extraídos los datos se analizarán los diferentes criterios de la Tabla 2, descritos anteriormente.

Los casos de estudio seleccionados son sitios de gobierno con temáticas y características disímiles:

Tabla 3. Casos de estudio.

	Caso de Estudio	***Se extrajeron los datos***	***Criterio de Normalizaión***	***URL***
A	Portal del Gobierno Argentino ARGENTINA.GOB.AR	- Subtítulo del tema - Descripción/Contenido - Link de cada Subtítulo - Link del sitio	Se extraen los contenidos de los datos referidos a la salud, todo lo que no cumpla con dicho criterio, no se toma en cuenta	https://www.argentina.gob.ar/tema/cuidarlasalud
B	Municipalidad de Lanús	- Subtítulo del tema - Link de cada Subtítulo - Link del sitio	Se extraen los temas referidos a la seguridad de la municipalidad de Lanús	http://www.lanus.gob.ar/fichas/63-municipio-de-lanus
C	Trámites a distancia - Organismos	temas relacionados con los organismos - Organismo - Link de cada organismo - Link del sitio	Se extraen los organismos relacionados a los trámites a distancias	https://tramitesadistancia.gob.ar/tramitesadistancia/tad-publico

5.1 Comparativa de Casos de Estudio

Para todos los casos de estudio, se extraen los datos en formato de archivo .xls. En la Tabla 3 se muestra la comparativa de los casos de estudio con las distintas variables para la validación de datos.

Tabla 3. Comparativa de los casos de estudio con respecto a los criterios.

Criterio	***Caso A***	***Caso B***	***Caso C***
Integridad	CUMPLE	CUMPLE	CUMPLE
Existencia	CUMPLE	CUMPLE	CUMPLE
Formato	NO CUMPLE, se muestran los datos de Subtítulo y Descripción juntos	CUMPLE	NO CUMPLE, se muestran los datos de Subtítulo y Cantidad juntos
Consistencia	CUMPLE	CUMPLE	CUMPLE
Precisión	NO CUMPLE, ya que hay dos datos en la misma columna y no posee un carácter de espacio o diferenciación entre uno y otro.	CUMPLE	NO CUMPLE, ya que hay dos datos en la misma columna y no posee un carácter de espacio o diferenciación entre uno y otro.
Duplicidad	El dato de link de URL es el mismo y se	El dato de link de URL es el mismo y se	El dato de link de URL es el mismo y

	repite por cada uno de los registros extraídos	repite por cada uno de los registros extraídos	se repite por cada uno de los registros extraídos
Relevamiento: Anomalías de datos	No posee valores inválidos	No posee valores inválidos	No posee valores inválidos
Estandarización	No Cumple con el criterio de normalización de datos.	Cumple con el criterio de normalización de datos	No Cumple con el criterio de normalización de datos
Data Maching	Porcentaje de aceptación válido	Porcentaje de aceptación válido	Porcentaje de aceptación válido

A partir de la comparativa presentada en la tabla 3, se puede observar que los criterios de integridad y existencia se cumplen en todos los casos. Para el criterio de formato con el que fueron extraídos, es decir, para el caso A y C no se cumple, ya que los datos de subtítulo y la descripción de este, se encuentran juntos en la misma celda del archivo en .xls, esto va de la mano con que el dato no es preciso en su identificación de celda, debido a que representa dos datos a la vez. Para el criterio de precisión, no se cumplen en los casos A y C, ya que el dato no es exacto al de la fuente original del sitio del que se extrajo. Por otra parte, estos casos tampoco cumplen con los criterios finales de normalización ya que hay datos fusionados como ser el subtítulo y la descripción.

En el caso del criterio de duplicidad, sólo se observa que la herramienta extrae el mismo dato del link URL del sitio original, esto es, en cada registro, la URL del sitio se repite. Es necesario realizar foco a estas cuestiones, ya que la extracción puede traer datos redundantes los cuales ocupan mayor espacio de almacenamiento. Para el criterio de anomalías de datos, en todos los casos no existen inconvenientes. Cabe destacar que, prestar atención a los criterios en forma correcta, permite que los datos obtenidos sean más limpios y confiables, dando mayor seguridad en la calidad del dato a trabajar. Finalmente el porcentaje de aceptación es aceptable.

6 Conclusiones

La importancia de la calidad de los datos es necesaria para la toma de decisiones, es por ello que, se observó que la verificación y validación de datos, ayuda a aumentar el grado de confianza de la información brindada a los usuarios finales.

Este procedimiento de verificación de los datos extraídos en el raspado de datos, ayuda a la verificación e implementación de una correcta aplicación de reglas de negocio, calidad y análisis de la misma.

Cada uno de los criterios presenta los conceptos que se deben tener en cuenta al realizar la extracción de datos en un análisis preliminar. Como se muestran en los casos de estudio, los datos pueden ser variados, pero los conceptos a validar para cuidar la calidad del dato extraído, serán los mismos. Se debe prestar especial atención a la precisión y formato de los datos rapados como así al resultado obtenido. Este resultado será la fuente de origen de datos otras aplicaciones las cuales se utilizarán para diversos fines colaborativos.

Referencias

1. Khabsa, M., & Giles, C. L. (2014). The number of scholarly documents on the public web. PloS one, 9(5), e93949.
2. Murillo, D., & Saavedra, D. (2017, July). Web Scraping de los Perfiles y Publicaciones de una Afiliación en Google Scholar utilizando Aplicaciones Web e implementando un Algoritmo en R. In Memorias de Congresos UTP (pp. 8-15).
3. Aucancela Guamán, M. A., Naranjo Vaca, M. J., & Betún Yuquilema, J. F. (2018). Mapeo sistematico de literatura de un data lake.
4. Vega, J. J. C., Aguilar, L. J., & Marín, L. M. G. La inteligencia de negocios como una herramienta en la gestión académica. Revista Científica. ISSN, 124, 2253 (2016).
5. Zhao, B. (2017). Web scraping. Encyclopedia of big data, 1-3.
6. Aguilar, L. J. Big Data, Análisis de grandes volúmenes de datos en organizaciones. Alfaomega Grupo Editor (2016)
7. Albright, S. C., Winston, W., & Zappe, C. (2010). Data analysis and decision making. Cengage Learning.
8. Planas, M., Rodríguez, T., & Lecha, M. (2004). La importancia de los datos. Nutrición hospitalaria, 19(1), 11-13.
9. Cardenas Rubio, J., Guataqui Roa, J., & Montaña Doncel, J. (2015). Metodología para el análisis de demanda laboral mediante datos de Internet: el caso colombiano. Revista de Economía del Rosario, 18(01), 93-126. doi:http://dx.doi.org/10.12804/rev.econ.rosario.18.01.2015.03
10. González, R. U., Méndez, N. D., & Moreno, J. (2018, October). Validación propuesta para medir la interactividad en un ambiente de aprendizaje caso con plataforma erudito. In Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE) (Vol. 29, No. 1, p. 1756).
11. Erudito, "Erudito Beta", Disponible en: http://erudito.medellin.unal.edu.co/
12. Edo, M., & Amparo, G. (2016). Técnicas de minería de datos para la detección de la visibilidad web de empresas españolas (Doctoral dissertation).
13. Jibaja Ayala, P. D., & Lascano Ramirez, E. S. (2018). Estudio del estado del arte de las herramientas informáticas para la gestión de conocimiento externo a una empresa (Bachelor's thesis).
14. OchandO, M. B. Google scraping. Uso ético de la información: implicaciones y desafíos, 167.
15. Munoz, L., Mazon, J. N., & Trujillo, J. (2011). ETL process modeling conceptual for data warehouses: a systematic mapping study. IEEE Latin America Transactions, 9(3), 358-363.
16. Martínez, R., Rodríguez, R., Vera, P. & Parkinson, C. (2019). Análisis de técnicas de raspado de datos en la web aplicado al portal del Estado Nacional Argentino. Workshop de Investigadores en Ciencias de la Computación (WICC 2019, Universidad Nacional de Río Cuarto).
17. Import.io, "Proporcionar los datos web que informan las decisiones comerciales cotidianas", Disponible en: https://www.import.io

Mantenimiento del software, aspectos aplicados en sitios web accesibles

Mariño Sonia, Juan Gallardo, Pedro Alfonzo

Universidad Nacional del Nordeste Fac. de Cs. Exactas y Naturales y Agrimensura

Resumen. El ciclo de producción del software puede presentarse en distintos modelos de procesos y en todos ellos se incorpora el mantenimiento como un aspecto fundamental. En este trabajo se presenta una experiencia de mantenimiento Correctivo, aspectos tendientes a asegurar la calidad del software, en particular la Accesibilidad Web y de esta manera generar productos tecnológicos de calidad y permitir la inclusión digital de los individuos independientemente de sus capacidades. Se describe la metodología utilizada, las etapas abordadas para la elaboración del presente trabajo y se detallan las iteraciones realizadas, con miras a incluir en los sucesivos incrementos mejoras en torno a la Accesibilidad Web.

Palabras claves: Mantenimiento del software, accesibilidad web, W3C, WCAG 2.0.

1 Introduccion

El ciclo de producción del software puede presentarse en distintos modelos de procesos y en todos ellos se incorpora el mantenimiento como un aspecto fundamental. En este sentido en [1] se afirma que existen cuatro actividades fundamentales de proceso: Especificación del software, Desarrollo del software, Validación del software y Evolución del software. Esta última incluye el proceso de mantenimiento.

1.1 Mantenimiento del software

El Instituto de Ingeniería Eléctrica y Electrónica (IEEE, 1993 [2]) define al mantenimiento del software (MS) como la modificación de un producto software, después de su entrega para corregir fallos, mejorar el rendimiento u otros atributos, o para adaptar el producto a otro entorno.

Respecto al mantenimiento del software (MS), en el IEEE [2] se establecen los siguientes tipos: i) Mantenimiento adaptativo: modificación de un producto de software que se realiza después de su entrega para mantener un programa operativo mientras se cambia el entorno de producción; ii) Mantenimiento correctivo: modificación de un producto software realizado después de su entrega para corregir fallas encontradas; iii) Mantenimiento de emergencia: implica un mantenimiento

correctivo no programado, ejecutado para mantener un sistema operacional; iv) Mantenimiento perfectivo: modificación de un producto software después de su entrega para mejorar su rendimiento o mantenibilidad.

Actualmente, existe un gran interés en la incorporación de pautas de calidad en el desarrollo de software, que impulsen la interacción y fomenten la inclusión digital de los individuos independientemente de sus capacidades. De esto se trata la accesibilidad web (AW), la cual se ha convertido en varios países del mundo en una preocupación [3].

1.2 Accesibilidad a los contenidos web

La Accesibilidad Web (AW) refiere al acceso universal a la Web, independientemente del tipo de hardware, software, infraestructura de red, idioma, cultura, localización geográfica y capacidades de los usuarios [4]. Numerosos sitios web presentan barreras de accesibilidad, lo que dificulta la utilización de la Web para muchas personas [5] [6], aun cuando es vasta la legislación que intenta mejorar y asegurar el acceso a contenidos.

Existen organizaciones han propuestos estándares en torno a la AW, una es el W3C (World Wide Web Consortium que promueve las Pautas de Accesibilidad para el Contenido Web (WCAG) 2.0.

Los individuos y organizaciones que emplean las WCAG son un grupo amplio y variado que incluye diseñadores y desarrolladores web, reguladores, agentes de compra, profesores y estudiantes. Para satisfacer las necesidades tan variadas de esta audiencia, se proporcionan varios niveles de orientación: principios generales, pautas generales, criterios de conformidad verificables y una amplia colección de técnicas suficientes, técnicas recomendables y fallos comunes documentados con ejemplos, enlaces a recursos adicionales y código [7].

La WCAG 2.0 se compone de 4 principios, 12 pautas y 61 criterios de conformidad (éxito), más un número no determinado de técnicas suficientes y técnicas de asesoramiento [8]:

Principio 1: Perceptibilidad - La información y los componentes de la interfaz de usuario deben presentarse a los usuarios de la manera en que puedan percibirlos.: 4 directrices y 22 criterios de cumplimiento.

Principio 2: Operabilidad - Los componentes de la interfaz de usuario y la navegación deben ser operables.: 4 directrices y 20 criterios de cumplimiento.

Principio 3: Comprensibilidad - La información y el manejo de la interfaz de usuario deben ser comprensibles.: 3 directrices y 17 criterios de cumplimiento.

Principio 4: Robustez - El contenido debe ser lo suficientemente robusto como para confiarse en su interpretación por parte de una amplia variedad de agentes de usuario, incluidas las tecnologías asistidas: 1 directriz y 2 criterios de cumplimiento.

Cada principio está compuesto por una cantidad determinada de pautas [7].
Principio 1: Perceptible - La información y los componentes de la interfaz de usuario deben ser presentados a los usuarios de modo que ellos puedan percibirlos.

- Pauta 1.1 Alternativas textuales: Proporcionar alternativas textuales para todo contenido no textual de modo que se pueda convertir a otros formatos

que las personas necesiten, tales como textos ampliados, braille, voz, símbolos o en un lenguaje más simple.

- Pauta 1.2 Medios tempodependientes: Proporcionar alternativas para los medios tempodependientes.
- Pauta 1.3 Adaptable: Crear contenido que pueda presentarse de diferentes formas (por ejemplo, con una disposición más simple) sin perder información o estructura.
- Pauta 1.4 Distinguible: Facilitar a los usuarios ver y oír el contenido, incluyendo la separación entre el primer plano y el fondo.

Principio 2: Operable - Los componentes de la interfaz de usuario y la navegación deben ser operables.

- Pauta 2.1 Accesible por teclado: Proporcionar acceso a toda la funcionalidad mediante el teclado.
- Pauta 2.2 Tiempo suficiente: Proporcionar a los usuarios el tiempo suficiente para leer y usar el contenido.
- Pauta 2.3 Convulsiones: No diseñar contenido de un modo que se sepa podría provocar ataques, espasmos o convulsiones.
- Pauta 2.4 Navegable: Proporcionar medios para ayudar a los usuarios a navegar, encontrar contenido y determinar dónde se encuentran.

Principio 3: Comprensible - La información y el manejo de la interfaz de usuario deben ser comprensibles.

- Pauta 3.1 Legible: Hacer que los contenidos textuales resulten legibles y comprensibles.
- Pauta 3.2 Predecible: Hacer que las páginas web aparezcan y operen de manera predecible.
- Pauta 3.3 Entrada de datos asistida: Ayudar a los usuarios a evitar y corregir los errores.

Principio 4: Robusto - El contenido debe ser suficientemente robusto como para ser interpretado de forma fiable por una amplia variedad de aplicaciones de usuario, incluyendo las ayudas técnicas.

- Pauta 4.1 Compatible: Maximizar la compatibilidad con las aplicaciones de usuario actuales y futuras, incluyendo las ayudas técnicas.
- Para que una página web sea conforme con las WCAG 2.0, deben satisfacerse todos los requisitos de conformidad siguientes [7]:
- Nivel A: la página web satisface todos los Criterios de Conformidad del Nivel A, o proporciona una versión alternativa conforme.
- Nivel AA: la página web satisface todos los Criterios de Conformidad de los Niveles A y AA, o se proporciona una versión alternativa conforme al Nivel AA.
- Nivel AAA: la página web satisface todos los Criterios de Conformidad de los Niveles A, AA y AAA, o proporciona una versión alternativa conforme al Nivel AAA.

Los métodos de evaluación de la accesibilidad se dividen en dos tipos: los analíticos y los empíricos. Los empíricos son utilizados para realizar las llamadas "evaluaciones payoff", que requieren una interacción entre los usuarios y el sitio web (Ej. Técnicas de Pantallas, test de usuarios y revisiones subjetivas) [9].

Los métodos analíticos se basan en la inspección de las páginas web que se realizan por medio de las llamadas "evaluaciones automáticas", las cuales validan el sitio de forma automática de acuerdo al cumplimiento de los criterios de la guía WCAG. Este método fue el utilizado para el proyecto.

En el trabajo se presenta una experiencia de mantenimiento Correctivo del software, aspectos tendientes a asegurar la calidad del software, en particular la Accesibilidad Web y de esta manera generar productos tecnológicos de calidad y permitir la inclusión digital de los individuos independientemente de sus capacidades.

2 Metodología

A continuación con fines de ilustrar el mantenimiento Correctivo, se describen las etapas abordadas para la elaboración del presente trabajo.

- Selección y estudio de herramientas de licencia libre para evaluar automáticamente y verificar el cumplimiento de las pautas WCAG 2.0. Se optó por eXaminator [10]. eXaminator es un servicio en línea para evaluar de modo automático la accesibilidad de una página web, usando como referencia algunas técnicas recomendadas por las Pautas de Accesibilidad para el Contenido Web 2.0 (WCAG 2.0). Adjudica una puntuación entre 1 y 10 como un indicador rápido de la accesibilidad de las páginas y proporciona un informe detallado de las pruebas realizadas.
- Selección y validación automática de dos páginas que componen un sitio e-commerce.
- Aplicación de procesos iterativos a la página principal y de catalogo de productos de un sitio e-commerce; y la posterior modificación del código necesario para mejorar la Accesibilidad Web de acuerdo a las WCAG 2.0 Nivel "A". Este proceso se iteró hasta lograr una página totalmente accesible.

3 Resultados

3.1 Proceso de mantenimiento Correctivos

Se aplicó iterativo a la página principal y de catalogo de productos de un sitio e-commerce. En este sentido, se realizaron las evaluaciones de la Accesibilidad de acuerdo a las WCAG 2.0 Nivel "A". En la Fig. 1 se detalla el análisis de las pautas de accesibilidad para el contenido web WCAG 2.0, donde se observa la nota (N)

asignada por cada prueba realizada, una ponderación (P), la puntuación obtenida (NP), un puntaje global un puntaje de 4.9 en la página principal del sitio evaluado y las puntuaciones obtenidas de acuerdo al tipo de discapacidad. Identificado el código se procedió al mantenimiento correctivo necesario.

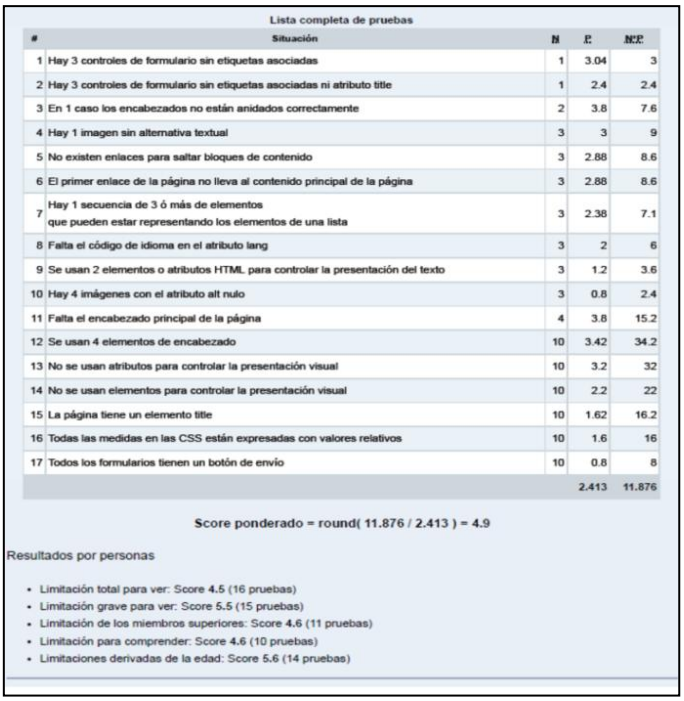
Las Tablas 1 y 2 sintetizan los errores detectados en la página principal y la página de catalogo de productos respectivamente, del sitio e-commerce analizado, en las sucesivas iteraciones realizadas.

Tabla 1. Análisis de errores detectados en la página inicial

Iteraciones	Número de pruebas	Excelente	Regular	Mal	Muy Mal	Puntaje
Iteración 1	17	6	1	8	2	4.8
Iteración 2	17	6	1	10	0	5.1
Iteración 3	12	9	1	2	0	7.9

Tabla 2. Análisis de errores detectados en la página de catalogo de productos

Iteraciones	Número de pruebas	Excelente	Regular	Mal	Muy Mal	Puntaje
Iteración 1	15	8	2	2	3	6.3
Iteración 2	14	9	2	2	1	7.3
Iteración 3	13	11	2	0	0	9.0



Lista completa de pruebas

#	Situación	N	P	NP
1	Hay 3 controles de formulario sin etiquetas asociadas	1	3.04	3
2	Hay 3 controles de formulario sin etiquetas asociadas ni atributo title	1	2.4	2.4
3	En 1 caso los encabezados no están anidados correctamente	2	3.8	7.6
4	Hay 1 imagen sin alternativa textual	3	3	9
5	No existen enlaces para saltar bloques de contenido	3	2.88	8.6
6	El primer enlace de la página no lleva al contenido principal de la página	3	2.88	8.6
7	Hay 1 secuencia de 3 ó más de elementos que pueden estar representando los elementos de una lista	3	2.38	7.1
8	Falta el código de idioma en el atributo lang	3	2	6
9	Se usan 2 elementos o atributos HTML para controlar la presentación del texto	3	1.2	3.6
10	Hay 4 imágenes con el atributo alt nulo	3	0.8	2.4
11	Falta el encabezado principal de la página	4	3.8	15.2
12	Se usan 4 elementos de encabezado	10	3.42	34.2
13	No se usan atributos para controlar la presentación visual	10	3.2	32
14	No se usan elementos para controlar la presentación visual	10	2.2	22
15	La página tiene un elemento title	10	1.62	16.2
16	Todas las medidas en las CSS están expresadas con valores relativos	10	1.6	16
17	Todos los formularios tienen un botón de envío	10	0.8	8
			2.413	11.876

Score ponderado = round(11.876 / 2.413) = 4.9

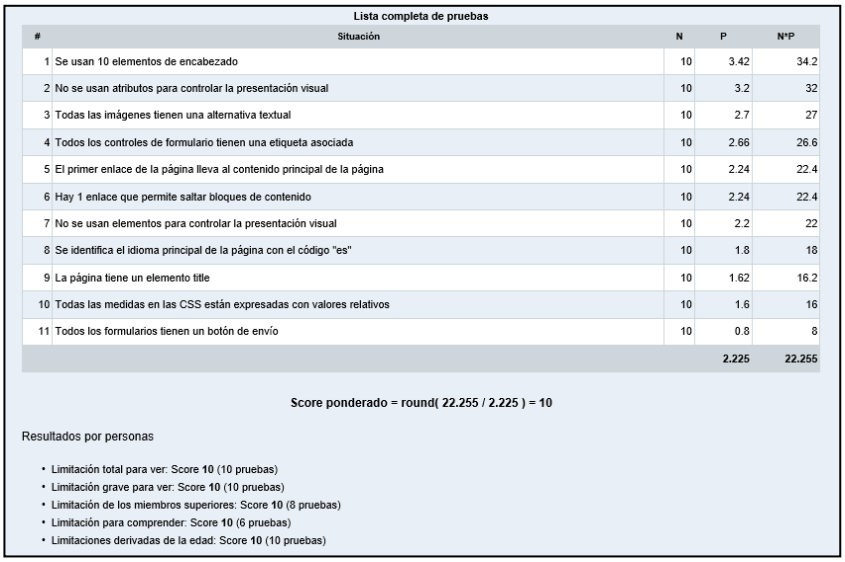
Resultados por personas

- Limitación total para ver: Score 4.5 (16 pruebas)
- Limitación grave para ver: Score 5.5 (15 pruebas)
- Limitación de los miembros superiores: Score 4.6 (11 pruebas)
- Limitación para comprender: Score 4.6 (10 pruebas)
- Limitaciones derivadas de la edad: Score 5.6 (14 pruebas)

Fig. 1. Análisis de accesibilidad de la página inicial

3.2 Evaluación de Accesibilidad

Una vez realizada todas las iteraciones (ver Tabla 1 y 2) y aplicado el mantenimiento Correctivo, se obtuvo como resultado que, tanto la página principal como la de catalogo de productos, lograron el "Nivel A" de los requisitos de conformidad establecidos por la WCAG 2.0. Lo expuesto se visualiza en las Figs. 2 y 3 donde se observan una puntuación de 10 puntos, de acuerdo a las métricas brindadas por el validador automático seleccionado.



Lista completa de pruebas

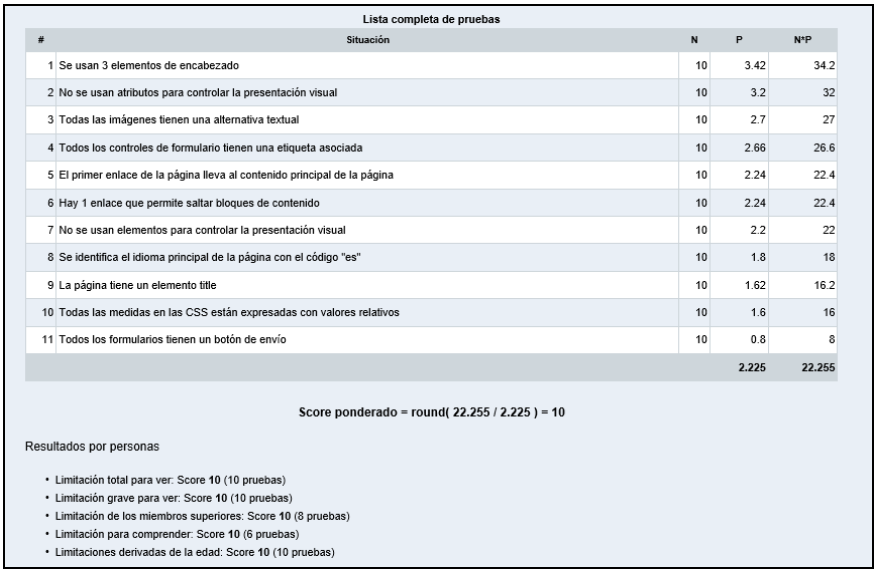
#	Situación	N	P	N*P
1	Se usan 10 elementos de encabezado	10	3.42	34.2
2	No se usan atributos para controlar la presentación visual	10	3.2	32
3	Todas las imágenes tienen una alternativa textual	10	2.7	27
4	Todos los controles de formulario tienen una etiqueta asociada	10	2.66	26.6
5	El primer enlace de la página lleva al contenido principal de la página	10	2.24	22.4
6	Hay 1 enlace que permite saltar bloques de contenido	10	2.24	22.4
7	No se usan elementos para controlar la presentación visual	10	2.2	22
8	Se identifica el idioma principal de la página con el código "es"	10	1.8	18
9	La página tiene un elemento title	10	1.62	16.2
10	Todas las medidas en las CSS están expresadas con valores relativos	10	1.6	16
11	Todos los formularios tienen un botón de envío	10	0.8	8
			2.225	**22.255**

Score ponderado = round(22.255 / 2.225) = 10

Resultados por personas

- Limitación total para ver: Score **10** (10 pruebas)
- Limitación grave para ver: Score **10** (10 pruebas)
- Limitación de los miembros superiores: Score **10** (8 pruebas)
- Limitación para comprender: Score **10** (6 pruebas)
- Limitaciones derivadas de la edad: Score **10** (10 pruebas)

Fig. 2. Informe general y puntuaciones globales de la página Principal



Lista completa de pruebas

#	Situación	N	P	N*P
1	Se usan 3 elementos de encabezado	10	3.42	34.2
2	No se usan atributos para controlar la presentación visual	10	3.2	32
3	Todas las imágenes tienen una alternativa textual	10	2.7	27
4	Todos los controles de formulario tienen una etiqueta asociada	10	2.66	26.6
5	El primer enlace de la página lleva al contenido principal de la página	10	2.24	22.4
6	Hay 1 enlace que permite saltar bloques de contenido	10	2.24	22.4
7	No se usan elementos para controlar la presentación visual	10	2.2	22
8	Se identifica el idioma principal de la página con el código "es"	10	1.8	18
9	La página tiene un elemento title	10	1.62	16.2
10	Todas las medidas en las CSS están expresadas con valores relativos	10	1.6	16
11	Todos los formularios tienen un botón de envío	10	0.8	8
			2.225	**22.255**

Score ponderado = round(22.255 / 2.225) = 10

Resultados por personas

- Limitación total para ver: Score **10** (10 pruebas)
- Limitación grave para ver: Score **10** (10 pruebas)
- Limitación de los miembros superiores: Score **10** (8 pruebas)
- Limitación para comprender: Score **10** (6 pruebas)
- Limitaciones derivadas de la edad: Score **10** (10 pruebas)

Fig. 3. Informe general y puntuaciónes globales de la página de Catálogo de Productos

4 Conclusiones

En este artículo se propone la introducción de conceptos de mantenimiento correctivo centrados en la AW. Se adaptó un proceso iterativo, con miras a incluir en los sucesivos incrementos mejoras en torno a la AW. Para ello se aplicó un evaluador automático que, además de indicar el puntaje proporciona un detalle de las dificultades referentes a AW localizadas. Es así como estas determinaciones permiten mejorar el código del desarrollo web y así a través de sucesivos incrementos aplicar detalladamente las pautas de accesibilidad para el contenido en la web (WCAG 2.0). Se prevé continuar la indagación empírica y desarrollar aquellos requerimientos que permitan lograr el cumplimiento del Nivel A en la totalidad del sitio web, dado que se expusieron resultados de aplicar evaluaciones sucesivas en sólo dos páginas concretas.

Agradecimientos. El trabajo se desarrolló en el marco del proyecto: TI en los sistemas de información: modelos, métodos y herramientas, acreditado por la Secretaria General de Ciencia y Técnica de la UNNE.

References

1. Sommerville, I.: Ingeniería del Software, Séptima Edición, Pearson Educación, 2005.
2. IEEE STD 1219:1993. Institute of Electrical and Electronics Engineers- ieee (1993). Standard for Software Maintenance. IEEE Computer Society Press.USA.
3. Gómez Codutti, A., Mariño, S. I. y Alfonzo, P. L.: "Una propuesta integradora de Mantenimiento Correctivo aplicada al Diseño Web Adaptativo y Accesibilidad Web", *Scientia et Technica* Año XXI, 21(1), 2016.
4. Luján Mora, S.: Accesibilidad Web, Definición de accesibilidad web, http://accesibilidadweb.dlsi.ua.es/?menu=definicion
5. World Wide Web Consortium (W3C), Introducción a la Accesibilidad Web, ¿Qué es la Accesibilidad Web?, https://www.w3c.es/Traducciones/es/WAI/intro/accessibility
6. ¿Qué hace el W3C?, https://www.w3c.es/ayuda/#activity
7. Pautas de Accesibilidad para el Contenido Web (WCAG) 2.0., http://www.sidar.org/traducciones/wcag20/es/
8. WCAG 2.0. Pautas de accesibilidad al contenido web 2.0., http://accesibilidadweb.dlsi.ua.es/?menu=pautas-2.0
9. Análisis de los Métodos de Evaluación de la Accesibilidad Web, Tipos de métodos de evaluación de la accesibilidad web, https://gplsi.dlsi.ua.es/almacenes/ver.php?pdf=102
10. eXaminator. Validador de Accesibilidad, http://examinator.ws

Integración de tecnologías de minería de datos y GIS. Generación de conocimiento en torno a la identificación y caracterización de clusters de robos y hurtos en una ciudad Argentina

Mariño Sonia, Lorena Flores, Sebastián Martins

Universidad Nacional del Nordeste Fac. de Cs. Exactas y Naturales y Agrimensura

Abstract. Se presenta la integración de tecnologías de minería de datos y GIS, orientadas a la generación de conocimiento para identificar y caracterizar clusters de robos y hurtos en una ciudad de Corrientes en el primer semestre de 2017. Se adaptó la metodología CRISP-DM, y se aplicó un conjunto de técnicas de minería de datos (SOM, TDIDT y Redes Bayesianas) para identificar y comprender los patrones delictivos. Además, se vincularon los patrones descubiertos con la tecnología GIS para comprender las zonas calientes de mayor ocurrencia de estos delitos. La finalidad es proponer innovadoras modalidades para apoyar procesos de decisión basados en TI.

Keywords: Minería de datos, GIS, toma de decisions, aplicaciones en gobierno.

1 Introducción

La inteligencia de negocios o Business Intelligence (BI, por sus siglas en inglés) ofrece un concepto clave desde una perspectiva empresarial, combina tecnologías y procesos que asisten al análisis de los datos y a la transformación de información empresarial, con el fin de obtener conocimiento válido que permita la toma de decisiones apropiadas orientadas a optimizar los recursos y mejorar los resultados.

La minería de datos (MD) es la subdisciplina de los sistemas de información que contribuye a la inteligencia de negocios las herramientas necesarias para explotar la información necesaria y transformarla en conocimiento útil. Para realizar este proceso de búsqueda de patrones se requiere el uso de técnicas de minería de datos (agrupación, árboles de decisión, redes neuronales artificiales, etc.) y la aplicación de algoritmos específicos dependiente de la problemática (agrupamiento, clasificación, predicción, entre otros) [1].

Existen diversos procesos de minería o explotación de información para obtener conocimiento a partir de los datos disponibles [2]:

Descubrimiento de reglas de comportamiento: El proceso de descubrimiento de reglas de comportamiento aplica cuando se requiere identificar cuáles son las condiciones para obtener determinado resultado en el dominio del problema.

Descubrimiento de grupos: Este proceso se aplica cuando se requiere identificar una partición en la masa de información disponible sobre el dominio de problema.

Ponderación de interdependencia de atributos: Este proceso se aplica cuando se requiere identificar cuáles son los factores con mayor incidencia (o frecuencia de ocurrencia) sobre un determinado resultado del problema.

Descubrimiento de reglas de pertenencia a grupos: Este proceso se aplica cuando se requiere identificar cuáles son las condiciones de pertenencia a cada una de las clases en una partición desconocida "a priori", pero presente en la masa de información disponible sobre el dominio de problema.

Ponderación de reglas de comportamiento o de la pertenencia a grupos: Este proceso se aplica cuando se requiere identificar cuáles son las condiciones con mayor incidencia (o frecuencia de ocurrencia) sobre la obtención de un determinado resultado en el dominio del problema, sean estas las que en mayor medida inciden sobre un comportamiento o las que mejor definen la pertenencia a un grupo.

En el análisis de delitos criminales, el descubrimiento de patrones significativos ha brindado la posibilidad de obtener datos de interés para interpretar y adecuar este conocimiento en la definición de los planes de prevención requeridos. La aplicación de diversas técnicas de minería de datos sobre el campo criminal se ha convertido en una herramienta con un gran potencial que permite diseñar estrategias específicas para esta área, resultando en un proceso automático de extracción de conocimiento útil [3].

Así mismo, el desarrollo de nuevas tecnología han logrado implementar nuevas herramientas como el GIS (Geographic Information System o Sistema de Información Geográfica, SIG su acrónimo en castellano), tal como lo define Chang [4], "…un sistema informático que permite capturar, almacenar, consultar, analizar y mostrar datos geoespaciales". Desde su lanzamiento, el GIS ha logrado traspasar las barreras tecnológicas presentando la información espacial a través de mapas georreferenciados. Estos mapas juegan un papel esencial en la vida cotidiana de las personas, su incorporación permite analizar, comunicar y compartir información con el fin de resolver y abordar problemáticas diarias, y ayudan a asistir a la toma de decisiones más inteligentes.

Desde esta perspectiva, el GIS puede tratarse como una base de datos que contiene información geográfica valiosa vinculada a un determinado territorio de interés [5]. Por ello, siguiendo a [6], resulta viable poder aplicar MD sobre base de datos georreferenciadas, para el hallazgo de patrones y regularidades significativas que resulten en conocimiento de interés sobre el territorio asociado. En el análisis delictivo, la MD aplicada sobre datos espaciales criminales permite la visualización de los hechos delictivos a través de mapas [7], con el fin de analizar, clasificar y predecir tendencias del delito de acuerdo a una determinada zona geográfica.

Este trabajo se validó en la detección y el análisis del comportamiento de hechos delictivos de robos y hurtos que afectan a la ciudad de Corrientes en el período comprendido por el primer semestre del año 2017, aplicando técnicas de minería de datos sobre una base de datos georreferenciada, y apoyados en tecnología GIS e Infraestructura de Datos Espaciales (IDE) para visualizar la información.

El conocimiento obtenido de la integración de la tecnología GIS y algunos algoritmos de la minería de datos se reflejó en la construcción del denominado Mapa del Delito o Mapa del Crimen de la ciudad de Corrientes. Esta es una herramienta cartográfica que permitió mapear y visualizar los patrones de delictualidad obtenidos

del análisis realizado, en el cual se identificaron las zonas de mayor riesgo de ocurrencia de hechos según distintos criterios.

Así, la implementación de esta solución tecnológica permitió identificar patrones para detectar y predecir la ocurrencia de estos tipos de delitos, relacionados con ubicaciones geográficas y otras variables de interés, entre las que se mencionan: i) características del hecho: lugar, día y horario de ocurrencia del delito, tipo de elemento sustraído (vehículo, domiciliario, otros), tipo de ataque (forcejeo, arrebato, etc.), tipo de arma u objeto utilizado en la escena y jurisdicción policial interviniente, ii) registro del autor del hecho: características del sospechoso (sexo del delincuente, edad aproximada, etc.), iii) registro del denunciante: sexo y edad de la víctima, entre otras.

Mediante este análisis es posible obtener información para la toma de decisiones, orientada el diseño de diferentes planes de prevención, mejora de la seguridad, alerta de situaciones de riesgo al ciudadano, reducción del impacto de la delincuencia, y mayor eficacia y eficiencia en el accionar de las fuerzas policiales ante estos actos delictivos, entre otros posibles impactos positivos.

2 Método

En la construcción de la solución propuesta, se contemplaron:

2.1 Adaptación de la metodología de CRISP-DM

En la literatura se mencionan diversas metodologías para modelar y procesar procesos de minería de datos. En este trabajo se optó por la metodología de CRISP-DM [4] para explotar los datos, que consta de las siguientes fases:

- Comprensión del negocio: se determinan los objetivos y requerimientos del proyecto desde una perspectiva del negocio, definiendo el problema de minería y el plan de trabajo.
- Comprensión de los datos: aborda la recolección de datos que se utilizarán en el proyecto y la familiarización con los mismos. Podrían surgir primeras hipótesis acerca de la información que podría estar oculta.
- Preparación de los datos: comprende actividades de tratamiento de los datos para construir el conjunto final sobre el cual se aplicarán las técnicas de minería.
- Modelado: se aplican las diversas técnicas y algoritmos de minería sobre el conjunto de datos para generar información y los patrones implícitos en ellos.
- Evaluación: se analizan los patrones obtenidos en función de los objetivos organizacionales. Se debería determinar si se ha omitido algún objetivo importante del negocio y si el nuevo conocimiento será implementado, es decir, si se pasará a la próxima etapa.
- Implementación: se comunica el nuevo conocimiento con fines de utilización. Éste debe ser representado de forma entendible para el usuario.

2.2 Algoritmos para modelar y explotar los datos

Se propone la aplicación de los procesos de explotación de información que comprende la combinación de las siguientes técnicas de minería de datos:

- SOM y TDIDT aplicados al descubrimiento de Reglas de Pertenencia a Grupos: Para el descubrimiento de reglas de pertenencia a grupos se propone utilizar mapas auto-organizados (SOM) y, una vez identificados los grupos, aplicar algoritmos de inducción (TDIDT) con el objeto de establecer las reglas de pertenencia a cada uno. Se optó por el algoritmo Kohonen-SOM para el descubrimiento de grupos y el algoritmo C4.5 para la caracterización o descubrimiento de reglas de cada clúster.
- Redes Bayesianas aplicadas a la Ponderación de Interdependencia entre Atributos: se opta por Redes Bayesianas Para determinar en qué medida la variación de los valores de un atributo incide sobre la variación del valor de un atributo clase. Se utilizó el algoritmo Naive Bayes para ponderar los atributos.

2.3 Software para explotar los datos

Se optó por Tanagra para ejecutar los procesos de explotación de información y generar información. Ofrece varios métodos de minería de datos de análisis exploratorio de datos, aprendizaje estadístico, aprendizaje automático y área de bases de datos. Implementa varios algoritmos de aprendizaje supervisado y no supervisado [5].

3 Resultados

Para la detección de los delitos de robo y hurto cometidos en la Capital de la Provincia de Corrientes en el periodo de tiempo de Enero–Junio de 2017, se utilizó una base de datos delictiva recibida del SAT (Sistema de Alerta Temprana), un sistema informático que almacena y concentra datos criminales referentes a los distintos tipos de delitos ocurridos en Argentina, y a partir del cual se realiza un análisis estadístico de la información. Actualmente se desconoce el uso de técnicas o herramientas de minería de datos aplicadas a estos datos.
Se estableció como objetivo de minería de datos, que sustenta el modelado y experimentación expuesta en el trabajo:

- Analizar una muestra de datos correspondientes al primer semestre del año 2017 registrados para la ciudad Capital de la Provincia de Corrientes y generar patrones de comportamiento relevantes a través de la agrupación de casos. Definidos los grupos, se deben determinar los factores que caracterizan la ocurrencia de delitos de robo y hurto en la ciudad, y detectar aquellos con mayor incidencia dentro de cada grupo.

Para lograr dicho objetivo se establecieron las siguientes necesidades:

- Identificar y caracterizar grupos que definan el comportamiento de delitos en base a las características temporales y del delito, para comprender con mayor detalle indicadores que definan a dichos grupos.
- A partir del comportamiento definido en el punto anterior, determinar cual/es de las características tiene un mayor nivel de incidencia en la ocurrencia de delitos en la ciudad.

Consideraciones:

- Los datos se obtuvieron de la base de datos gestionada/administrada por el SAT.
- La construcción del modelo, que responde al objetivo de minería de datos definido, implica: establecer los valores de los parámetros para ejecutar las técnicas "SOM y TDIDT aplicados al descubrimiento de reglas de pertenencia a grupos" y "Redes Bayesianas aplicadas a la ponderación de interdependencia entre atributos", ver Tabla 1.

Se presentan los resultados del descubrimiento de reglas de pertenencia a grupos que identifican y caracterizan el comportamiento del delito. Se contempla:

Formación de clusters

Para la formación de grupos se definen los atributos de entrada requeridos por el algoritmo Kohonen SOM: dia_m, mes_r, horas_r, delito_descrip, tipo_lugar, clase_arma, elemento_sustraido, tipo_ataque. Se aplicó el algoritmo y se obtuvieron cuatro grupos, de los cuales 79 registros forman parte del clúster c_som_1_1, 98 registros conforman el clúster c_som_1_2, 128 forman el clúster c_som_2_1 y 61 del clúster c_som_2_2.

El algoritmo Kohonen SOM agrupó los clusters basándose en la similitud de los valores de sus atributos. Los datos de un mismo grupo presentan características comunes, pero a su vez, los objetos entre los grupos deben ser diferentes.

- Descubrimiento del comportamiento del delito

Para descubrir el comportamiento del delito, se establecen los atributos para la ejecución del algoritmo C4.5. El atributo clase se define como la variable grupo generada por el algoritmo Kohonen SOM en la formación de clusters, y se seleccionaron como los atributos de entrada: dia_m, mes_r, horas_r, delito_descrip, tipo_lugar, clase_arma, elemento_sustraido y tipo_ataque

La aplicación del algoritmo C4.5 proporcionó un total de 17 reglas que caracterizan a los grupos identificados. A partir de las reglas generadas, se elaboró una primera interpretación de los clusters resultantes:

- Clúster c_som_1_1: Caracterizado por delitos mayoritariamente cometidos por medio del forcejeo o arrebato. Incluye casos en la vía pública o en un domicilio particular. Registra mayor actividad delictiva los días miércoles, jueves y viernes, durante los meses de marzo, abril, mayo y junio.
- Clúster c_som_1_2: En principio se trataría de delitos a través del forcejeo o arrebato y con arma blanca. Incluye casos de ocurrencia en la vía pública o en un domicilio particular. Además, se observa la característica que los delitos ocurrieron durante los días sábado, domingo, lunes o martes, en los meses de marzo, abril, mayo y junio.

- Clúster c_som_2_1: Es el grupo o clúster que más delitos agrupa. Se determinó que el delito aconteció en algunos casos en forma de ataque brutal y en otros se determinó que no existió ningún tipo de ataque. Predomina el uso de arma fuego y en otros casos, se determinó que no existió ningún elemento de ataque, e incluye lugares del hecho como en la vía pública, en un comercio, en el interior de un rodado, o en un domicilio particular. Este grupo está caracterizado por el robo objetos personales, motocicletas, de tipo domiciliario, vehículos y dinero. No obstante, en otros casos también se observa que no hubo elemento sustraído.
- Clúster c_som_2_2: Este grupo a diferencia del resto, posee mayor descripción de las características de los delitos. Particularmente, los delitos ocurrieron con ataque brutal a las víctimas y en otros casos, no existió ningún tipo de ataque. Se detectó como el arma de fuego y arma blanca, los elementos predominantes en este grupo, así mismo, también resultaron casos en donde no se registraron elementos de ataque. Este grupo incluye casos de ocurrencia en vía pública o en algún domicilio particular, en horarios de mañana o siesta. Asimismo, los meses que caracterizan a este grupo son enero y febrero. Los elementos sustraídos durante el delito que definen a este grupo son de tipo objetos personales, motocicletas, de tipo domiciliario, vehículos y dinero. En otros casos de igual forma se observa que no hubo elemento sustraído.

Ponderación de atributos con mayor incidencia en el comportamiento de delitos.

Se definen a continuación, los atributos para la ejecución del algoritmo Naive Bayes. El atributo clase se identifica como la variable de predicción generada por el algoritmo C4.5. en el descubrimiento del comportamiento del delito, y se establecen los atributos de entrada: dia_m, mes_r, horas_r, delito_descrip, tipo_lugar, clase_arma, elemento_sustraido y tipo_ataque (ver Anexo 1 donde se define el conjunto de datos final utilizado).

Los valores de incidencia de los atributos permiten analizar la información resultante Para ejemplificar la Fig. 1 muestra el comportamiento del atributo dia de la semana en los distintos clusters identificados.

	Sabado	Domingo	Lunes	Martes	Miercoles	Jueves	Viernes	Sum
c_som_1_1	0,0476	0,0357	0,0238	0,0952	0,2619	0,2381	0,2976	1,0000
c_som_1_2	0,5263	0,2421	0,1474	0,0842	0,0000	0,0000	0,0000	1,0000
c_som_2_1	0,2583	0,1917	0,0583	0,1250	0,1083	0,1417	0,1167	1,0000
c_som_2_2	0,3881	0,1940	0,1343	0,1194	0,0448	0,0448	0,0746	1,0000
Sum	0,3033	0,1694	0,0874	0,1066	0,1038	0,1093	0,1202	1,0000

Fig. 1. Ponderación de incidencia para el atributo día de la semana. Fuente: elaboración propia

A través del estudio de las características del delito con mayor incidencia en la ocurrencia del hecho, se presenta el análisis para cada clúster:

- Clúster c_som_1_1: La mayor cantidad de delitos ocurrieron por robo (94%). En la mayoría de los casos se trata de arrebatos (79,76%) y en la vía pública (90,48%). Existe un alto porcentaje de incidencia (84,52%) de la sustracción de algún objeto personal de la víctima. Se detectó al día viernes, como el día de la semana más frecuente (29,76%), y en cuanto a los valores de incidencia por mes, marzo y abril registraron mayor volumen de delitos (60,71%).
- Clúster c_som_1_2: Este clúster se caracteriza por la ocurrencia de robos en su totalidad por arrebato (79,76%) y de objetos personales. La mayoría de los casos ocurridos sucedieron en vía pública (90,48%) y con arma blanca (92,63%). Los hechos delictivos tienen mayor ocurrencia los días sábados y domingos, en el horario de siesta de 12:00 pm a 16:00 pm y de noche de 20:00 pm a 24:00 pm. Los robos tienen mayor ocurrencia en los meses de marzo y abril según los valores de incidencia.
- Clúster c_som_2_1: Este clúster está determinado por la distribución en cantidades iguales en la ocurrencia de ambos tipos de delitos (robo y hurto). El objeto sustraído con mayor incidencia con respecto al resto fue de tipo objeto personal (35%). De los lugares, el más frecuente fue en la vía pública (33,33%).
- Clúster c_som_2_2: Se analizó que los delitos de robo ocurrieron con mayor frecuencia en los meses de enero y febrero, los días sábados y en el horario de siesta de 12:00 pm a 16:00 pm. Este clúster se caracterizó por presentar mayores casos en la vía pública (95%), utilizando arma blanca (65,67%). Los robos ocurrieron en su totalidad por ataque brutal (52,24%). En cuanto a los elementos sustraídos los más frecuentes fueron motocicletas y objetos personales.

Finalizadas la ejecución de las técnicas de MD descriptas para: la formación de clusters, el descubrimiento del comportamiento de los delitos y en la ponderación de atributos con mayor incidencia en dicho comportamiento se obtuvo valiosa información. Ésta resume que: i] cuatro de los atributos más significativos de los clusters son: tipo de objeto sustraído (objeto personal), lugares y tipos de armas utilizadas durante el delito (vía pública y con arma blanca) y tipo de ataque más sufrido (arrebato).

Los cluster c_som_1_2 y c_som_2_2, permitieron identificar días de mayor porcentaje de ocurrencia de delitos (viernes, sábados y domingos) y horarios más frecuentes (horarios de siesta y de madrugada).

Para la representación espacial de los clusters generados se utilizó como complemento la generación de un mapa de calor, una herramienta capaz de detectar las zonas con mayor concentración de puntos sobre una determinada ubicación geográfica. A continuación se ejemplifica el tratamiento considerando el clúster c_som_1_1, y similares procesamiento de imágenes y análisis se aplicaron sobre los restantes clusters identificados

Identificadas las zonas, se analizó la modalidad del clúster c_som_1_1 presente en ellas, con el objetivo de caracterizar los delitos de acuerdo a los resultados derivados de la aplicación de los algoritmos de minería de datos. En la Fig. 2 se muestran distintos tonos de color rojo hacia colores más claros según la densidad de los puntos concentrados. Se aprecia que existe una mayor concentración de puntos sobre las zonas oeste, centro y norte de la ciudad de Corrientes. Los barrios que agrupan estos puntos: San Benito, Unión, Progreso, San Marcos, Primera Junta, Aldana y Sol de Mayo. La modalidad de delitos para este clúster se caracterizó por el uso de armas y mayoritariamente de tipo blancas, ocurrieron en la vía pública y a través del arrebato. Los días viernes se registraron mayor cantidad de hechos delictivos en el horario de siesta de 12:00 pm a 16:00 pm., detectándose los robos de objetos personales.

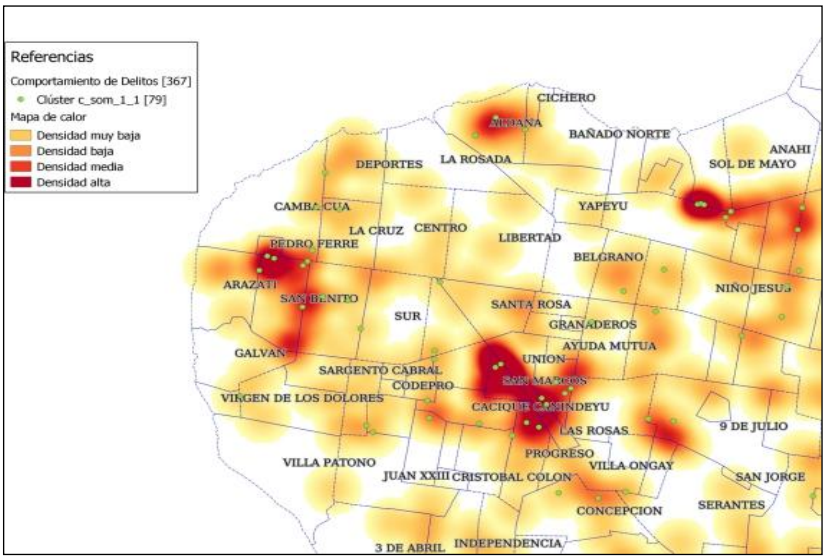


Fig. 2. Mapa de calor del clúster c_som_1_1 para caracterizar el comportamiento de delitos - Fuente: elaboración propia

3 Conclusiones

El trabajo describe una integración entre tecnología GIS y minería de datos: en particular se aplicaron los algoritmos Kohonem SOM, C4.5 Y Naive Bayes. Los resultados obtenidos muestran la buena calidad de la información generada que innovaría en el modo de apoyar procesos decisorios.

Los mapas delictivos creados corresponden a la relación entre el delito ocurrido y un espacio geográfico y tiempo determinado. Estas representaciones gráficas aportan a la identificación y localización del delito con mayor frecuencia de ocurrencia.

En este objetivo de minería de datos, el análisis caracteriza el delito de acuerdo con variables como: la modalidad, el empleo de armas, los objetos sustraídos, lugar de ocurrencia, el rango horario, los días de la semana y los meses con más delitos ocurridos.

De acuerdo con la modalidad del delito estudiada en cada uno de los clusters, el arrebato es la constante en los cuatro grupos, el mayor número se registra en la zona de la vía pública. Los objetos mayormente sustraídos son de tipo personal, y predomina el uso de las armas blancas utilizadas para cometer los delitos.

Con respecto a las ubicaciones, si se compara los mapas del delito (Figs. 10, 11, 12 y 13) a partir de la densidad de puntos concentrados se identifican regiones similares como ser, el centro y el este de la ciudad. Sin embargo, no todos los delitos ocurren en las zonas de mayor densidad, las avenidas se presentan como el principal escenario de ocurrencia de delitos. Por ejemplo, se observa la alta concentración de hechos a lo largo de la Av. Maipú.

Los delincuentes prefieren actuar los días viernes, sábados y domingos. En cuanto al rango horario, la tendencia de los hechos delictivos se registró entre las 12 p.m. y 16 p.m.

Respecto a los meses en los cuales se presentaron más delitos, los clusters coinciden en detectar a los meses de marzo y abril. Si bien este resultado no ofrece mayor información, como líneas futuras se podrían aplicar estas técnicas de minería de datos para los años siguientes y los resultados, y así conocer la influencia de delito por época o por estaciones del año.

Referencias

1. Anoopkumar, M., & Rahman, A. M. Z. (2016, March). A Review on Data Mining techniques and factors used in Educational Data Mining to predict student amelioration. In 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE) (pp. 122-133). IEEE.
2. García-Martínez, R., Britos, P., & Rodríguez, D. (2013, June). Information mining processes based on intelligent systems. In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (pp. 402-410). Springer, Berlin, Heidelberg.
3. McCue, C. (2014). Data mining and predictive analysis: Intelligence gathering and crime analysis. Butterworth-Heinemann.
4. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. SPSS inc, 16.

5. Rakotomalala, R. (2004,January). TANAGRA project. http://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.
6. Baghdadi, N., Mallet, C., & Zribi, M. (2018). QGIS and Generic Tools. John Wiley & Sons.

Importancia de la Innovación en los Sistemas de Salud Pública de San Juan

Alicia Aballay[1,2], Adriana Sarmiento[2], María R. Romagnano,[1,2,3] María del Carmen Becerra[2]

[1]Instituto de Informática, FCEFN, Universidad Nacional de San Juan
[2]Departamento de Informática, FCEFN, Universidad Nacional de San Juan
[3]Departamento de Geofísica y Astronomía, FCEFN, Universidad Nacional de San Juan

prof.alicia@gmail.com, adrivalle2005@gmail.com, maritaroma@iinfo.unsj.edu.ar, mcbecerra2008@gmail.com

Abstract. En este trabajo se abordan estudios organizacionales que pretenden servir de aporte a la innovación en los servicios prestados en materia de salud pública en la provincia de San Juan, los cuales ante la falta de interoperabilidad e integración de las bases de datos existentes no logran manejar la información sanitaria de los pacientes en tiempo real. Se identifica el conjunto de prácticas y recursos de gestión, resultante de la interoperabilidad semántica de estándares de salud, y se discuten brevemente diversos aspectos involucrados en la prestación de los servicios. A los efectos de lograr una gestión dinámica, eficiente e inteligente se describe la relación entre innovación y calidad, Desde el punto de vista de la seguridad, el modelo organizacional propuesto permite establecer las bases para que cualquier tipo de organización pública o privada, pueda realizar un manejo eficiente y seguro de los datos de salud de los pacientes.

Keywords: Calidad, estándares de salud, innovación, interoperabilidad semántica, recursos humanos en salud, seguridad.

1 Introducción

En este tercer milenio la innovación organizacional constituye un eje estratégico para la gestión de la calidad, enfocada en atender demandas y expectativas de los pacientes. La innovación es la introducción de un nuevo o significativamente mejorado servicio, aplicable en Salud Pública a través de un método organizativo, distinto en las prácticas internas de la gestión y tratamiento de los datos de Salud [1].

La arquitectura futura de los Sistemas de Historia Clínica Electrónica (HCE) divisa una tecnología más desvinculada de la capa de datos, su presentación y operaciones, de esta manera la salida de datos será genérica y podrá ser utilizada con diferentes modelos de prestación, más adaptables a cualquier entorno y no solo exclusivamente al de origen. Por lo tanto, se realiza la descripción de los factores de la atención donde la HCE es el centro de atención médica, atentos a que se debe comenzar con

información específica del paciente que ingresa al sistema de salud frente a un profesional médico, sea especialista o de un área o turno determinado, que culmina con el diagnóstico. Existen diferentes maneras de pensar y realizar la prestación de los servicios de salud pública, según la naturaleza, las opciones brindadas al paciente y la variedad de formas de realizar el proceso de prestación y el nivel de estandarización considerando además, el grado de contacto con el paciente.

En cada caso de implementación particular de estas técnicas, su integración y desarrollo es a medida, y en la actualidad no existe al momento, a nivel nacional, un sistema centralizado ni un estándar operativo jurídicamente exigible respecto de la modalidad a elegir para la implementación de la HCE, y como señala la autora, se puede arriesgar a decir que tal vez nunca exista [2].

En el presente trabajo se trata el tema de los estudios organizacionales como alternativa para evaluar el rol de los servicios de salud en el proceso de producción social de la salud y, en especial como un determinante de la salud pública. Se exponen aspecto relacionados con la innovación y calidad, y se presentan las bases conceptuales de los estudios organizacionales, justificando la necesidad de implementar este tipo de estudios en los Servicios de Salud Pública de la Provincia de San Juan, para explicar a nivel provincial cómo se produce la mencionada contribución a los sistemas de salud de la población, cuáles son sus factores de éxito y cuáles son las áreas susceptibles de mejora. Asimismo, se presenta un modelo de interoperabilidad semántica que ha surgido del proyecto de investigación "Aplicación de Modelos Semánticos a los Sistemas de Información para la Mejora y el Impacto de las Tecnologías de Información".

Como objetivos se proponen:

1. Describir las condiciones de calidad en los Sistemas Organizacionales de Salud Pública en la Provincia de San Juan.
2. Especificar las condiciones de Seguridad de la HCE.
3. Relevar los estándares para interoperabilidad semántica.

2 Metodología

Se realizó una investigación documental con el fin de conocer el estado del arte de los Sistemas de Salud. Es decir, se analizó la información disponible en organismos públicos. También se estudiaron experiencias realizadas en el campo público y privado, y por último se analizaron estándares del campo de las TIC (Tecnologías de la Información y de la Comunicación), para la interoperabilidad entre sistemas en el sector de la salud, reconocidos a nivel internacional.

3 Aspectos Clave para la Innovación en Sistemas de Salud

3.1 Vínculo entre Innovación y Calidad

En los sistemas organizacionales se debe llevar adelante un proceso para tomar conciencia y generar un marco normativo como el de la calidad con el fin de preparar el terreno para la creatividad y su producto, la innovación [3].

En los sistemas de información de las organizaciones, es indispensable generar un ambiente propicio para lograr la innovación y calidad en un esfuerzo coordinado de hombres y elementos materiales. La cultura de la organización debe reunir valores como la iniciativa y la responsabilidad y generar un clima laboral donde prime la cooperación. Como se puede ver, hoy en día los sistemas de HCE (Historia clínica Electrónica) son mayormente accedidos por personal sanitario (administrativo, médico y auxiliar, entre otros), y no de manera directa por el propio paciente, el que continúa solicitando el acceso a su HCE a la institución que efectúa su custodia [2].

El proceso creativo tiene éxito cuando da como resultado la innovación, es decir que pueden generarse múltiples ideas en el ámbito de la organización, aplicando técnicas y analizando su valor, pero al momento de la evaluación, se deben jerarquizar las ideas producidas y estudiar su factibilidad, terminando con la selección final de aquella que responde a los objetivos esperados.

La innovación se puede ver en el sistema organizacional y su medio una vez puesta en marcha la producción del servicio mediante un proceso tecnológico que pone en marcha las transformaciones, con la incorporación de la innovación y el posterior lanzamiento al mercado [4]. Un servicio representa un conjunto de operaciones que es diseñado y tiene un proceso, siguen una secuencia y se llevan a cabo con métodos y un sistema de prestación como en el caso de la HCE, siendo su característica distintiva que los productos resultantes son intangibles.

Un sistema de prestación de servicios de salud transforma insumos en productos, solo que, a diferencia de los bienes físicos, como se dijo antes, son intangibles. Así se puede afirmar que la productividad está expresada en la relación entre la producción de servicios y los recursos humanos utilizados para prestarlos, y la calidad percibida, es medida según los niveles de satisfacción de los usuarios y los tiempos de espera [5].

La innovación en las organizaciones que prestan servicios en los sistemas de salud se vuelve bastante compleja cuando se aborda desde la perspectiva organizacional, ya que ésta ocurre en diversos niveles de la organización, así como fuera de ella, donde el principal factor para su existencia es la necesidad de adaptabilidad del sistema organizacional para responder a los cambios del entorno. Es la primera etapa de vida del nuevo servicio o paquete de servicios. En este sentido, se debe comprender que Innovación y calidad están relacionados en el sentido que la mejora continua es un precepto que rige las políticas de calidad de las organizaciones. El sistema de calidad definido por la organización de servicios deberá contemplar variables de calidad específicas, que, en el caso de los servicios de salud, son enfocados en la atención mediada con la incorporación de tecnologías. Posteriormente, el proceso siguiente es el tratamiento y el seguimiento, finalizando con intervenciones médicas que adoptan variadas maneras, de acuerdo con la patología. Aquí se deben medir aspectos

referentes a la necesidad de cambio del sistema de atención basado en historia clínica en papel hacia la HCE, con sus variables de calidad específica. Se pueden identificar la calidad de la información, la calidad de la atención profesional y las cualidades de la historia clínica como factor orientador del proceso.

En el camino de la innovación la calidad aparece como eje fundamental de la gestión, representa un valor presente en todo el sistema que permite el logro de la misión y visión organizacional. La toma de conciencia de la calidad, el proceso de cambio, el aprendizaje y el compromiso de las personas, son factores interrelacionados para que mediante el liderazgo se genere el ambiente creativo y la inspiración para la innovación.

El enfoque principal de la gestión de la calidad es cumplir los requisitos del cliente y tratar de exceder sus expectativas. Entre las acciones posibles de la dirección para implementar una cultura innovadora hacia la calidad, se pueden mencionar las de Crear y mantener valores compartidos, fomentar el compromiso con la calidad y asegurarse que los líderes en todos los niveles sean modelos para toda la organización. En los servicios de salud, el papel de la HCE juega un rol fundamental dentro de los cambios requeridos para la satisfacción del usuario y la fidelidad del mismo al sistema prestador. Es dable mencionar la necesidad de conocer las demandas e intereses de los sectores partícipes (usuarios, directivos, trabajadores, proveedores, grupos sociales, gobierno, otras organizaciones).

La calidad total se aplica en todas las áreas y niveles de la organización y debe empezar en la cima de la jerarquía. El compromiso de la alta dirección es indispensable para garantizar un cambio a fondo en la cultura de la organización. Se busca el cambio paulatino y continuo para el logro de la excelencia en productos y procesos [6]. El proceso de mejora tiene las siguientes etapas (Fig. 1): elegir el área objeto de la mejora, Definir el equipo de trabajo, identificar los beneficios del benchmarking (estándar de excelencia, interno o externo); analizar el método actual; realizar un estudio piloto la mejora e implementar la mejora.

Fig. 1. Proceso de Mejora

Se pueden lograr beneficios clave al adoptar un estilo de gestión basado en Gestión de calidad, tales como el aumento de la eficacia y eficiencia en la

organización, la mejora en la coordinación de los proceso del valor para el cliente y en las comunicaciones entre las áreas y niveles, el desarrollo y la mejora de la capacidad de la organización y de sus personas para entregar los resultados deseados y, como coronación, el incremento de la satisfacción del cliente [7].

Es en la concepción del sistema de calidad donde sobresalen los valores señalados, con la estrategia y los objetivos enfocados en las personas, tanto quienes colaboran como los que son los prestadores, en cada área y función que intervienen en el proceso. Toda prestación de servicios posee características únicas, éstas varían en función de la naturaleza del servicio, pero también depende de la misión de la organización. Los directivos que deciden implementar la gestión de la calidad, toman una decisión que señala el rumbo organizacional cuando se llega a definir el nivel de prestación. Es indispensable involucrar a proveedores, miembros de la organización y usuarios. Los resultados de efectividad se conseguirán si se tiene contacto con el usuario (paciente) y cuando este manifiesta por escrito el grado de satisfacción. La disponibilidad de la información en la HCE refleja los valores de calidad, visibles al acompañar al paciente durante toda su vida, al servicio de la salud del mismo presente y futura.

En definitiva, contar con la HCE, es un medio para el logro de objetivos de calidad. La sostenibilidad se consigue cuando éste tiene la posibilidad de acceder a los datos e informes de su historia, lo que marca el nivel de organización interna. Se puede hacer llegar al mercado el brazo del servicio, ya que se trasciende las fronteras del sistema organizacional, va más allá de la atención en un momento o periodo determinado. Implica el logro de un servicio, medible, con valores individuales, grupales y organizacionales percibidos a nivel social.

3.2 Situación de los Servicios de Salud Pública en la Provincia de San Juan

El Sistema Sanitario Argentino ha dejado lugar a otras políticas dirigidas a la sociedad y vinculadas a la vida de las personas, tendientes a modificar los hábitos, los estilos de vida, la educación, las modalidades de trabajo y la distribución del ingreso en miras a la equidad, inclusión e igualdad [8]. La salud Pública "Es el esfuerzo organizado de la sociedad, principalmente a través de las instituciones de carácter público, para mejorar, promover, proteger y restaurar la salud de las poblaciones por medio de actuaciones de alcance colectivo" [9].

La salud se encuentra dentro del sector servicios. El diseño del servicio comprende todos los elementos que intervienen secuencialmente en la prestación del servicio desde la definición del negocio hasta su prestación, es decir, que se realiza desde la estrategia que define el negocio, siguiendo por el paquete que define al servicio y con el sistema que presta el servicio. El paquete de servicios está constituido por: servicios explícitos, servicios implícitos, bienes coadyuvantes e instalaciones de apoyo.

Los servicios de salud explícitos son los beneficios directamente relacionados con la esencia de la actividad, la calidad de las prestaciones de salud en una organización de naturaleza pública, privada o semiprivada. Todas las variables que intervienen en el servicio se vuelvan la HCD. Los servicios implícitos son los beneficios psicólogos que el paciente puede llegar a percibir y que normalmente hacen que valore más el

servicio de salud, aunque no varié la esencia de su prestación, la accesibilidad para usar el servicio, la privacidad de sus datos, etc.

Los bienes coadyuvantes son los artefactos tecnológicos comprados o consumidos por el paciente y que se requieren para que el servicio sea llevado a cabo. Las instalaciones de apoyo son los recursos físicos equipo tecnológico para procesar y actualizar informes para que el paciente tenga accesibilidad a su historia clínica, dado que para la prestación del servicio de Salud es importante tener el soporte tecnológico que permita tener la información completa del paciente.

Se puede hablar de la matriz de servicios que tienen como finalidad clasificar los servicios y además permite analizar las diferencias entre la administración de servicios encauzados por el proveedor, donde se requiere el seguimiento de la administración para automatizar e invertir capital, y los servicios encauzados por el paciente, los cuales demandan más atención en la gestión de recursos humanos y en el estudio de los aspectos de una tecnología flexible. Los servicios pueden reunir características de ambos, atendiendo requerimientos de los pacientes como el diseño de operaciones de servicio de la organización, denominados servicios coencausados entre clientes y proveedores. Los coencausados ofrecen una gama limitada de alternativas para los consumidores, con procesos moderadamente estandarizados, los servicios médicos entran en esta categoría.

Si bien la administración diseña la estrategia y define los objetivos del sistema de prestación del servicio, hay participación en el proceso por parte de los clientes, quiénes tienen influencia en la toma de decisiones, en cuanto a la forma de prestación del mismo. Las ideas acerca de nuevos productos provienen de diversas fuentes: Clientes, personal de marketing, de Investigación y desarrollo, de la alta dirección, de Producción, de Ingeniería Salud. Una vez lanzados al mercado todos los productos, incluyendo los productos buenos, tienen una duración limitada. Pueden permanecer tiempos muy variables dentro del mercado, pero la empresa debe anticiparse para efectuar los cambios necesarios, aunque sean pequeños, con el fin de lanzarlos cuando sea conveniente, como nuevos. Se debe tratar de poner en marcha un flujo de posibilidades de nuevos productos y servicios.

En la provincia de San Juan se firmó un convenio de colaboración entre el Ministerio de Salud Pública y la cartera homóloga de la provincia de Neuquén para la implementación de esta modalidad en el sistema público en un primer momento, según consta en los partes de prensa publicados oportunamente.El objetivo del convenio, es implementar en ambas provincias políticas de reforma y modernización de sus administraciones, orientadas a prestar más y mejores servicios a sus ciudadanos. Con este acuerdo de cooperación se busca, además, compartir experiencias para mejorar la salud de los habitantes, desarrollar un trabajo mancomunado de profesionales tecnológicos para lograr innovación permanente, aumentar la eficiencia y reducir los costos.

La provincia de San Juan eligió el proyecto neuquino por sobre otros de historias clínicas digitales de desarrollos públicos y privados. El interés se basa en los beneficios de esta herramienta en lo que hace a modernización del Estado y al valor agregado que le da el haber sido creado y generado dentro del sistema público neuquino, con recurso humano propio.

Neuquén y San Juan conformarán tres equipos –de Gestión, Implementación y Tecnológico-, que compartirán un proyecto de trabajo con encuentros virtuales y

presenciales de transferencia de conocimiento e implementación de la Historia Digital para la provincia de San Juan.

3.3 Condiciones de Seguridad de la HCE

Si analizamos los distintos modelos de implementación de las Historias Clínicas digitales en nuestro país y en el mundo, los esfuerzos realizados para brindarles integralidad, accesibilidad e interoperabilidad de los sistemas de información en salud, vemos que todos ellos se topan con su validez legal. La ley 26529 conceptualiza la HCE y en su Art.13 estipula que el contenido de la Historia Clínica Electrónica puede confeccionarse en soporte magnético siempre que se arbitren todos los medios que aseguren la preservación de su integridad, autenticidad, inalterabilidad, perdurabilidad y recuperabilidad de los datos contenidos en la misma en tiempo y forma Además, en la HCE se deberían especificar las propiedades de autenticación, no repudio e integridad. Sostenemos que sólo a través del diseño de un nuevo modelo de gestión organizacional [10], es posible particularizar y luego formalizar soluciones informáticas que permitan contar con un repositorio integral de documentos de salud de los pacientes.

Se debería construir un artefacto tecnológico firmado, contenido en un Sistema SI en sus funcionalidades y preferentemente externo a todos los sistemas productores de datos que hacen uso de este, que permite ser alimentado en la Organización Públicas, Privada y Semiprivadas en sus entradas por los datos a firmar y un conjunto de parámetros configuración que no afecte su funcionamiento dado en la dificultad de "acceder" de los distintos usuarios del sistema, de manera que una vez invocado el artefacto para el procesamiento de los datos, éste deberá gestionar el acceso a la información del paciente para producir la firma digital del Profesional médico y finalmente teniendo en cuenta la "seguridad" que debe brindarse al Paciente desde el punto de vista legal para resguardar la autoría, integridad y No repudio de los datos de salud del paciente, debiendo emitirse como resultado los datos de entrada junto a la firma digital correspondiente .La firma digital es una cadena o secuencia de caracteres que se adjunta al final del cuerpo del mensaje firmado digitalmente. La creación y verificación de firmas digitales se realiza mediante procedimientos técnicos, y supone la existencia de normas que respaldan su valor Legal. En nuestro país a partir de la sanción de la Ley 25.506/2001 y su normativa complementaria se cuenta con el marco regulatoria para su aplicación. Así la firma digital y la firma electrónica pueden ser aplicadas a la HCE, dependiendo de la autoridad emisora de certificados. En nuestra provincia existe la Ley N°8.128/2010, de Adhesión a la Ley 25.506 y se han firmado convenios de cooperación y Asistencia Técnica con la Secretaría de la Gestión Pública- Jefatura de Gabinete de Ministros, se ha impulsado el uso de esta nueva tecnología, creando los mecanismos y condiciones pertinentes que permiten su aplicación. La Dirección General de Recursos Humanos y Organización dependiente de la Secretaría de Gestión Pública, por convenio de Gobierno Provincial con la Oficina Nacional de tecnologías de Información (ONTI) ha sido designada como autoridad de Registro de Firma Digital del Poder Ejecutivo de la Provincia de San Juan. Se pretende brindar un aporte a la comunidad científica, al teorizar sobre los distintos componentes del modelo semántico para que las organizaciones tanto

públicas como privadas puedan gestionar la información de salud de los pacientes conformada en una HCE entendida como Artefacto TI [11] conceptualizado como las aplicaciones TI que habilitan y apoyan tareas incluidas en una estructura que a su vez está incluida en un contexto [12].

3.4 Estándares para el intercambio electrónico de datos del servicio de salud

Heath Level Seven HL7 es una iniciativa que comenzó en el año 1987, que se utiliza para normalizar las interfaces entre los sistemas heterogéneos de información y que se convirtió en el estándar de facto para el intercambio electrónico de datos clínicos de los pacientes y administrativos en los servicios de salud, es una organización que desarrolla estándares para minimizar las incompatibilidades entre sistemas de información en salud. En nuestro país no existe regulación para el intercambio estandarizado de información de salud, pero se cuenta con el apoyo de SNOMED que es una organización que aporta un estándar global para la representación de la información en salud. La terminología que define se una para encuestas sanitarias, registro de enfermedades y captura conceptos clínicos divididos en jerarquías como síntomas diagnósticos, antecedentes familiares, procedimientos quirúrgicos, estudios, diagnóstico. Etc. Entre los estándares médicos de interoperabilidad semántica se cita también a CIE que es la clasificación Universal de las enfermedades. El logical Observation identifiers Names and Codec (LOIC) es una base de datos y un estándar universal para identificar los resultados de los Laboratorios Médicos. La Sociedad Española de Radiología Médica (SERAM). La iniciativa I.H.E., acrónimo inglés de Integrating the Healthcare Enterprise (Integración de la Empresa/Organización de salud) nace como esfuerzo común de la R.S.N.A. y la H.I.M.S.S. (Healthcare Information and Management Systems Society). DICOM, IEEE 1073 y NCPDP SCRIPT en otros segmentos específicos del mercado de salud. EMR significa Electronic Medical Record que es el equivalente digital del expediente clínico que se lleva en papel y se guarda en archiveros enormes.

La ISO EN 13606 intenta dar una estructura para el contenido de EHR, se centra en las estructuras internas del mismo y a las que deben adherir todos los sistemas de comunicación. Respecto del formato a adoptar como estándar citamos el formato del documento portable ISO 32000-1-2008 y la firma de documentos digitales, y finalmente la norma ISO 32000-2-2017 nos ofrece un formulario digital para representar documentos electrónicos y permite a los usuarios intercambiar y ver documentos electrónicos independientemente del entorno en que se crearon, o del entorno en que se ven o imprimen.

A pesar que son muchos los estándares que desde la técnica se aportan para la innovación en la interoperabilidad de los sistemas de Salud, una gran cantidad de los mismos no logran manejar y gestionar la privacidad de la información sanitaria en tiempo real, de forma dinámica, eficiente e inteligente. EHD Establece las medidas básicas de seguridad que mantienen la confidencialidad de la información del paciente. Provee a las instituciones con mecanismos que consolidan eventos de auditoría de trazabilidad de la actividad de usuario.

3.5 Modelo de Interoperabilidad Propuesto

En el modelo de interoperabilidad propuesto (Fig. 2) se observa que la interoperabilidad se logra teniendo en cuenta la semántica (ontologías, tesauros, taxonomías), técnicas y herramientas, y estándares (SNOMED, CIE, HL7, CEN/ISO en 13606, etc.). Dicho modelo se encuentra en etapa de desarrollo y contrastación.

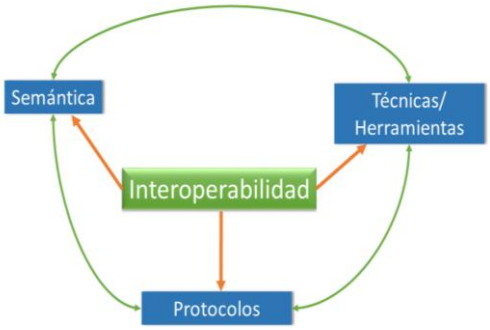


Fig. 2. Modelo de Interoperabilidad

Finalmente se remarca que el Modelo de Interoperabilidad básicamente representa la administración basada en Gestión de calidad, orientado a factores como el aumento de la eficacia y eficiencia en las organizaciones de salud, propiciando la mejora en la coordinación de los procesos de valor para el paciente y las comunicaciones entre las áreas y niveles.

4 Conclusiones

En el siglo XXI las organizaciones han tenido presiones para desarrollar estructuras y procesos duales que garanticen el rendimiento en un entorno complejo y expuesto a cambios rápidos. La doble búsqueda de estabilidad y cambio constituye una paradoja fundamental en todas las formas de organización y plantea un importante reto a las firmas que operan en el entorno empresarial actual [6]. En el sector de los servicios de salud, los directivos deberán conducir las organizaciones en el marco de las tendencias globales con sus desafíos tecnológicos, donde el aprendizaje será el fruto de los procesos de creatividad logrando innovación, cuyos cambios deberán monitorearse en forma continua. Las organizaciones innovadoras se caracterizan por generar un cambio cultural, propiciando actitudes de apertura y compromiso en las personas, quienes son los encargados de aplicarlos en el diseño y producción de servicios en todos los niveles y áreas.

Se deben combinar estructuras, procesos y capacidades actuando con flexibilidad para favorecer la convivencia de dos modelos de gestión, donde por una parte se busque la prevalencia de la estabilidad y el control con el fin de garantizar la eficiencia a corto plazo y, por otra parte, se promueva la asunción de riesgos,

explorando posibilidades nuevas. Las personas de la organización juegan un rol fundamental en la implementación de cambios, siempre de la mano de un liderazgo transformador que acompañe la transición en procesos como el que demanda la utilización de la Historia Clínica Electrónica (HCE).

Los resultados exitosos se pueden lograr con el impulso del espíritu de equipo, con el desarrollo de valores de calidad. Se deben explicitar las contribuciones de todos los miembros a la organización y a la sociedad en su conjunto. Así se aumenta la intensidad del esfuerzo que exige la motivación, logrando que se involucren los profesionales y el personal de apoyo, dando sentido de realización a las personas, al momento de observar la realización de las mejoras y los resultados del esfuerzo sistémico. Se alcanza la visión compartida de todos los miembros de que es posible satisfacer a todos los partícipes, trabajando en lo personal y organizacional se oriente a incorporar los cambios.

Los nuevos conocimientos e ideas se deben combinar con el conocimiento existente para el logro de la efectividad organizacional, cualquiera sea el modelo de organización asumiendo los riesgos para generar innovación a largo plazo. La clave para obtener y mantener la ventaja competitiva está en evaluar los resultados obtenidos y realizar los ajustes que garanticen la mejora continua. En conclusión, quienes asumen la toma de decisiones deben asumir riesgos, transitar la vida organizacional en busca de estrategias mediadas por políticas a largo plazo, que logren un balance entre calidad e innovación.

Referencias:

1. OCDE-EUROSTAT, Manual de Oslo, “Guía para la Recogida e interpretación de datos sobre
Innovación”, grupo TRAGSA (Traducción Española), 3° edición, 2005.
2. Faliero, Johanna C. Historia Clínica Electrónica: El futuro de la gestión documental sanitaria en la era de la e-salud. Editorial Ad-Hoc.: 2018. Año de edición ISBN: 978-987-745-108-5.
3.-Schroeder Roger, Susan Meyer y Johnny Rungtusanatham, “Administración de Operaciones”, 5°,Edición, Ed. Mc. Graw-Hill, Buenos Aires, 2011.
4.Adler, Martín Oscar y otros, “Producción y Operaciones”, Ed. Macchi, Buenos Aires, 2004.
5.Farjoun, M. Beyond dualism: stability and change as a duality; Academy of Management Review,35 pp. 202-225.
6.Chiavenato, Idalberto” Introducción a la Teoría General de la administración,” Mc. Graw-Hill.Educación, 8° edición, 2014.
7.Norma Iso 9000-2015. “Sistema de Gestión de la Calidad, traducción oficial en español, Ginebra, 2015”.
8.Aisenberg y otro. El nuevo régimen de titularidad y guarda da la HCE.http:///observatoriodelasalud.blogspot.com.ar
9.Alvaro Franco E. Tendencias y teorías en salud pública. http://scielo.sld.cu/pdf/hie/v47n1/hie06109.pdf
10.Becerra, María, y otros. Gestión de la Historia Clínica Digital.48 JAIIO.SID 2019
11.García, Sebastián José, Gómez, María Claudia y Orellana, Alejandra. Un Modelo de Artefacto Sistema de Información. http://bibliotecas,ucasal.edu.ar/opaccss/doc num.php?esplnum ide=1007.
12.Weber, Ron, 2003, Editor’s Comments: Still Desperately Seeking the IT Artifact, MIS Quarterly, Vol. 27 No 2, pp. iii – xi/June 2003.

Detección de enfermedades o plagas en cultivos de papas andinas del norte argentino utilizando procesamiento de imágenes digitales

Fabiola Patricia Paz[1], C. Javier Izetta[1],

[1] Grupo de Investigación y Desarrollo en Informática Aplicada, Facultad de Ingeniería,
Universidad Nacional de Jujuy, Ítalo Palanca 10
4600 S. S. de Jujuy, Argentina
{fabyppaz, javierizetta}@gmail.com

Abstract. En este trabajo se propone la detección de enfermedades o plagas en fotos de hojas de plantas de cultivos de papas andinas mediante la aplicación de técnicas bien conocidas de procesamiento de imágenes digitales. Una vez capturada la imagen el algoritmo propuesto determinará la existencia o no de alguna enfermedad o plaga mediante un análisis del color del área afectada. En el caso que exista alguna anomalía en la hoja, se podrá cuantificar y mostrar el límite del área dañada por la enfermedad o plaga. Los experimentos de este trabajo, se realizaron con varias muestras de hojas de plantas de cultivos dañadas y sanas extraídas de la web. La utilización de este algoritmo permite conocer con mayor precisión y rapidez el área afectada, permitiendo obtener resultados más confiables que puedan servir para tomar medidas oportunas para una mejor conservación y productividad del cultivo.

Keywords: Procesamiento de imágenes digitales, segmentación de imágenes digitales, hojas de plantas de cultivos, enfermedades o plagas del cultivo de papa.

1 Introducción

Los cultivos andinos históricamente formaron parte de la dieta de las poblaciones originarias, siendo cultivos de alto valor alimentario. Son considerados cultivos andinos por su resistencia y adaptación a sequias, heladas y suelos con características salinas [1]. Especialmente la papa, que poseen excelentes características nutricionales para su consumo ya que presentan altos porcentajes de materia seca, almidón, proteínas y minerales, así como cualidades culinarias destacables [2]. Sin embargo, cuando este cultivo se ve afectado por plagas y enfermedades, provoca pérdidas en los rendimientos y la calidad de los productos antes y después de la cosecha. Por lo tanto, tomar decisiones oportunas es vital para mantener y conservar sus propiedades como así también, una producción más eficiente [3].

En el área de la computación, el procesamiento de imágenes digitales ha despertado mucho interés por lo que ha evolucionado continuamente en estos últimos tiempos. Hoy en día, se puede encontrar numerosos trabajos de reconocimiento de

imágenes en diversas áreas, una de ellas es la agrícola en donde se han dado buenos resultados [4], [5], [6], [7], [8], [9]. En consecuencia y considerando además la importancia de conservar el valor alimenticio y cultural del cultivo de papa andina, es que en este trabajo se desarrolla un algoritmo para la identificación de plagas o enfermedades mediante el empleo de técnicas de procesamiento de imágenes. El reconocimiento de plagas y enfermedades es fundamental para determinar métodos de prevención y o control que pueden adoptarse en el momento oportuno.

Este trabajo se organiza de la siguiente manera: Sección 2 se define el concepto de procesamiento de imágenes digitales. Sección 3, se describe la metodología. Sección 4 se detalla algoritmo propuesto. Sección 5 se muestran los experimentos y resultados. Por último, en la Sección 6 se concluye.

2 Procesamiento de imágenes digitales

La visión computacional se define como un campo de la inteligencia artificial que, mediante la utilización de las técnicas adecuadas, permite la obtención, procesamiento y análisis de cualquier tipo de información especial obtenida a través de imágenes digitales [10]. Tradicionalmente se define en 4 pasos fundamentales [11]:

Fase 1, consiste en la captura o adquisición de las imágenes digitales mediante algún tipo de sensor.
Fase 2, consiste en el tratamiento digital de las imágenes, con objeto de facilitar las etapas posteriores. En esta etapa de procesamiento de imágenes es donde, mediante filtros y transformaciones geométricas, se eliminan partes indeseables de la imagen o se realzan partes interesantes de la misma.
Fase 3, se conoce como segmentación, y consiste en aislar los elementos que interesan de una escena para su posterior tratamiento.
Fase 4, es la etapa de reconocimiento o clasificación. En ella se pretende distinguir los objetos segmentados, gracias al análisis de ciertas características que se establecen previamente para diferenciarlos.

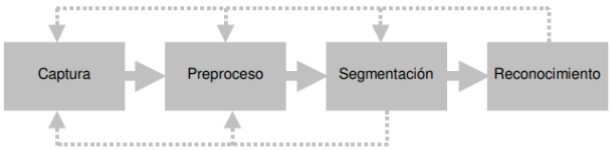


Fig. 1. Fases típicas de la visión computacional [11].

Estas cuatro fases, también representadas en la Fig. 1, no se siguen siempre de manera secuencial, sino que en ocasiones deben realimentarse hacia atrás [11] [12].

Este trabajo se centra en el procesamiento de imágenes digitales, abarcando las Fase

2, 3 y 4. En el procesamiento de imágenes digitales se encuentran todas las metodologías o algoritmos de edición de imágenes que tienen, entre otras, la finalidad de destacar o realzar elementos de importancia y ocultar o despreciar elementos de la

escena que no tengan un valor en el objetivo del procesamiento [11]. Con el fin de obtener el reconocimiento de anomalías en plantas de papas andinas, en la siguiente sección se describe los pasos realizados para la construcción del algoritmo y las herramientas del procesamiento de imágenes utilizadas en este trabajo.

3 Metodología

El reconocimiento de enfermedades o plagas en plantas de papas andinas se realiza tradicionalmente por inspección visual, lo cual posibilita la utilización de técnicas de procesamiento de imágenes digitales para desarrollar una herramienta que aportará no solo datos más rápidos y precisos sobre el área afectada de las plantas, sino también, hacerlo con bajo costo ya que se podría trabajar por ejemplo con cámaras de teléfonos móviles. A continuación, se describen las actividades realizadas para la construcción del algoritmo propuesto.

3.1 Adquisición de imagen

Se toma como base experimental el cultivo de papas andinas del norte argentino. Las imágenes son tomadas del trabajo de Pérez et. al. [3] y otras obtenidas de la web. Estas imágenes están capturadas directamente del cultivo con alta resolución en formato jpg. Pero para mejorar el reconocimiento de las áreas afectadas se modifican las mismas para obtener una imagen final con un fondo blanco y así reducir el error del reconocimiento tanto del área total de la hoja como el área de la zona dañada.
En la Fig. 2 se muestran algunas de las imágenes seleccionadas para este trabajo que presentan varias de las enfermedades y plagas a la que se encuentra expuesto este tipo de cultivo.

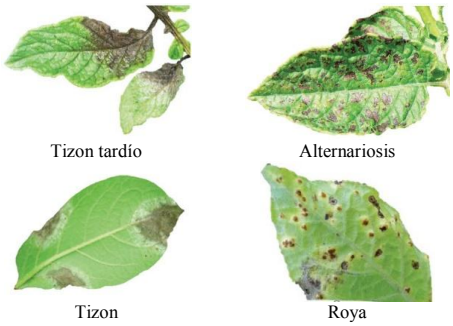


Fig. 2. Hojas de cultivos de papas con enfermedades y plagas.

3.2 Diseño del algoritmo de reconocimiento de enfermedades de hojas de cultivo

Una vez cargada la imagen de entrada en formato “jpg”, el algoritmo propuesto comienza seleccionando los colores del área dañada en las hojas ya que los mismos

no varían demasiado entre las hojas afectadas. Es decir, como la imagen de entrada está en RGB, crea una máscara para segmentar el área dañada dejando pasar el color de interés de las hojas. Para esto, se podría optar en trabajar con escalas de grises, pero los resultados no serían los adecuados ya que utilizar la información que brinda el color nos posibilita obtener una mejor área de la zona de interés. Por lo tanto, se decide trabajar directamente con los colores RGB para segmentar las zonas dañadas y obtener muestras de las mismas de manera exitosa, los valores de los pixeles de estas muestras se almacenan en un vector. Cabe aclarar que las capturas almacenadas de estos valores, correspondientes a cada nivel RGB, son únicamente para obtener un intervalo de colores el cual el algoritmo deberá aislar cuando lea una nueva imagen de entrada. El siguiente paso es utilizar operadores que permitan capturas las regiones de las hojas de nuevas imágenes que estén dentro de los valores de colores correspondientes a las muestras almacenadas. Esto se realizó de dos maneras diferentes. Primero se generó una máscara utilizando los valores máximos y mínimos correspondientes a las muestras almacenadas. Es decir, el algoritmo toma los pixeles de la imagen de entrada que estuvieran comprendidas entre esos valores máximos y mínimos. Este primer intento arrojó resultados que no fueron los esperados. El área obtenida estaba muy alejada del área real. Durante las pruebas se observó que en algunas imágenes funcionaba mejor que en otras, por lo que este método parece ser muy sensible a los ruidos presentes en las imágenes de entrada. Lo que llevó a buscar otra manera de capturar la zona de interés de las imágenes. Esta vez el algoritmo calculó la media de las muestras extraídas de las áreas dañadas obteniendo un vector con los valores medios para cada color RGB. Luego, para medir la similitud entre los colores utilizó la distancia euclidiana de los valores rojo (R), verde (G) y azul (B) de la imagen de entrada y el valor medio de los colores rojo (R), verde (G) y azul (B) de las muestras. La distancia euclidiana está definida por:

$$D(z,a)=((zR-aR)^2+(zG-aG)^2+(zB-aB)^2)^{1/2} \quad (\mathbf{1})$$

Donde z es similar a a si la distancia entre ellos es menor a un umbral especificado D. La media fué calculada en base a 10 muestras y un umbral de 0.11. Si los píxeles superan el umbral pasan a 0 y si están por debajo del umbral toman el valor 255, de esta manera el algoritmo obtiene la imagen segmentada. Luego para perfeccionar las zonas de interés de la imagen segmentada, el algoritmo utiliza operadores morfológicos como la dilatación y cierre para expandir un poco las zonas capturadas y eliminar algunos ruidos. En la dilatación se utilizó un elemento estructural de tipo 'esfera' de 3x3 y para el cierre un elemento estructural de tipo 'disco' de 3x3 ya que fueron los que obtuvieron los mejores resultados en los experimentos realizados.

Para obtener el porcentaje de daño de las imágenes de entrada primero se debe obtener el área total de la imagen original y luego de la zona dañada. Para esto, convierte la imagen original a escala de grises y la binariza utilizando un umbral alto entre 0.7-0.9 (siendo 0.9 con el que obtuvo un mejor resultado). Para eliminar el ruido utiliza nuevamente el operador cierre con un elemento estructural de "disco" de 5x5 varias veces (4 veces) y saca el complemento de la imagen resultante para tener los 1 en el área la\s hoja\s.

Luego, para obtener el área total de la imagen de entrada el algoritmo utiliza el operador suma para obtener el total de pixeles de la imagen (área 1) De la misma manera, con el operador suma obtiene el área de la zona detectada (área 2). Luego, el

área dañada detectada es multiplicada por 100 y se divide por el área de la imagen de entrada, esto se interpreta como el porcentaje de daño que tiene la planta.

Por último, el algoritmo dará como resultado la identificación o no de plagas y/o enfermedades de la hoja de cultivo. Cuando el porcentaje sea mayor al 1% el algoritmo determinará que la planta de cultivo está afectada por alguna enfermedad o plaga y además mostrará el porcentaje de daño de dicha planta. En caso contrario, cuando el porcentaje de detección sea menor al 1% el algoritmo no detectará algún tipo de anomalía, por lo que indicará que la hoja de planta de cultivo está sana. En la Fig. 3 se muestra el pseudocódigo del algoritmo.

```
# Algoritmo de reconocimiento de enfermedades de hojas de
cultivo de papas andinas

#Se lee la imagen en RGB
Leer imagen I
Mostrar la imagen I
Copiar I en I2
Recorrer cada pixel de la imagen I2 separando las capas (R G y
B)
# Capturar el área daña
Calcular con los valores anteriores la media total de la
distancia euclideana
Si distancia>0.11 # el radio se determinada por pruebas (0.1 a
0.2)
     Asignar al pixel los valores [0 0 0]
Sino
     Asignar al pixel los valores [255 255 255];
Fin
Definir el elemento estructural E de tipo esfera de 3x3
Aplicar 2 veces el operador de dilatación para imagen I2 con E
Definir elemento estructural E2 de tipo disco de 3x3
Aplicar 2 veces el operador cierre para I2 con E2
Binarizar la imagen I2
Mostrar la imagen I2 con la zona dañada detectada
# Cálculo del área total de la hoja
Convertir la imagen original I de color (rgb) a escala de grises
obteniendo I3
Luego binarizar I3 para trabajar con 0 y 1
Definir elemento estructural E3 de tipo disco de 5x5
Aplicar 4 veces el operador cierre a I3 con E3
Calcular el complemento de I3
# Cálculo el área completa de la hoja sin el área dañada
Restar a I3 el área de la hoja dañada I2
Mostrar la imagen I3 que contendrá la imagen de la hoja
analizada
#Cálculo del porcentaje de área dañada con respecto al área de
la imagen de entrada
area1=Calcular el total de pixeles de la hoja completa con el
operador suma.
area2=Calcular la suma de los pixeles que pertenece solo al área
dañada de la hoja
porcentaje=((area2*100)/area1);
```

```
# Resultado
Si porcentaje>1 # si la hoja analizada tiene un área menor
   Mostrar el mensaje:
   "La\s hoja\s de cultivo presenta una enfermedad o plaga'"
   "Con un porcentaje de daño del (porcentaje) del total de la
hoja"
sino
    Mostrar el mensaje:
    "Es un cultivo sano"
fin
```

Fig. 3. Pseudocódigo del algoritmo.

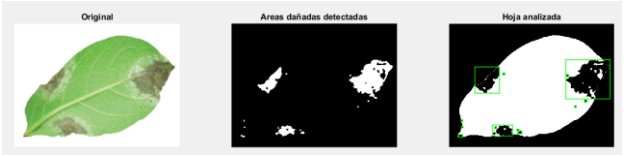
4 Análisis de los resultados

Para analizar los resultados mediante el algoritmo propuesto, se realizan ajustes individuales para cada imagen de entrada. En la Fig. 4 se muestran los resultados con las mejores respuestas obtenidas del algoritmo.

Imagen 1

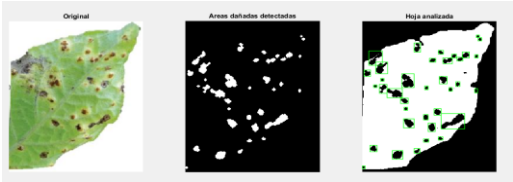
Las hojas presentan una enfermedad o plaga con un daño aproximado del 33.14 % del total.

Imagen 2



La hoja presenta una enfermedad o plaga con un daño aproximado del 14.13 % del total.

Imagen 3



La hoja presenta una enfermedad o plaga con un daño aproximado del 7.29 % del total.

Imagen 4

La hoja presenta una enfermedad o plaga con un daño aproximado del 18.53 % del total.

Imagen 5

Las hojas presentan una enfermedad o plaga con un daño aproximado del 21.37 % del total.

Fig. 4. Resultados óptimos de cada imagen de entrada.

Sin embargo, para lograr resultados aceptables con un algoritmo robusto, se decide calcular un valor medio de cada uno de los colores RGB que mejor se acerque a los resultados óptimos vistos en la Fig. 2. Para ello se calculó una media entre los valores medios de las muestras para cada una de las imágenes de prueba, y así obtener un vector que permita detectar enfermedades y plagas de cualquier imagen de entrada, además de calcular un porcentaje aproximado y aceptable del área de la zona dañada.

Este proceso se observa en la Tabla 1.

Tabla 1. Valores de la media obtenidos de 10 muestras con resultados vistos en Fig.2.

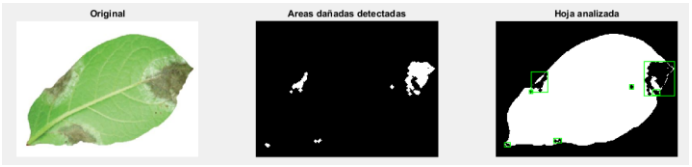
Imagen de entrada	RGB media R(rojo)	G(verde)	B(azul)
Imagen 1	124.37	128.94	84.80
Imagen 2	108.27	101.00	76.67
Imagen 3	96.15	84.25	54.70
Imagen 4	92.40	72.20	37.40
Imagen 5	89.37	79.60	69.70
Media total	102.11	93.20	64.65

A continuación, en la Fig. 3 se muestran los resultados de las imágenes de entradas con el nuevo vector de la media total calculada en la Tabla 1. Y en la Fig. 4 se muestra las diferencias absolutas de los porcentajes de las áreas calculadas entre los resultados obtenidos en la Fig. 2 y los resultados obtenidos en la Fig.3. Esta diferencia podría interpretarse como el error en el cálculo de las áreas dañadas.

Imagen 1

Las hojas presentan una enfermedad o plaga con un daño aproximado del 31.10 % del total.

Imagen 2

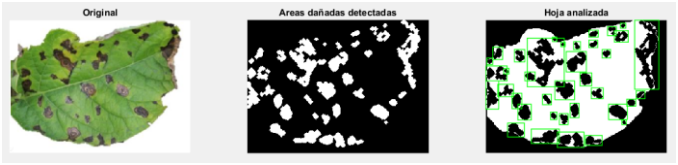


La hoja presenta una enfermedad o plaga con un daño aproximado del 5.36 % del total.

Imagen 3

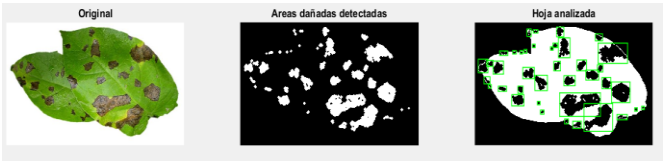
La hoja presenta una enfermedad o plaga con un daño aproximado del 10.87 % del total

Imagen 4



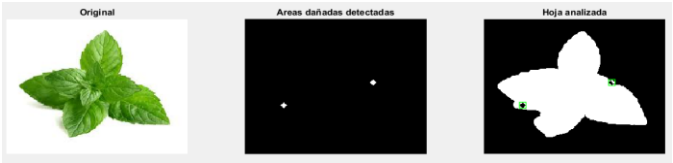
La hoja presenta una enfermedad o plaga con un daño aproximado del 24.99 % del total.

Imagen 5



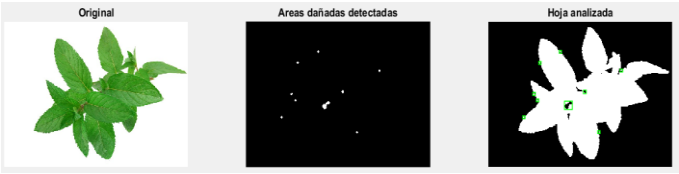
Las hojas presentan una enfermedad o plaga con un daño aproximado del 21.46 % del total.

Imagen 6



Es un cultivo sano.

Imagen 7



Es un cultivo sano.

Fig. 5. Resultados con la media total de cada imagen de entrada.

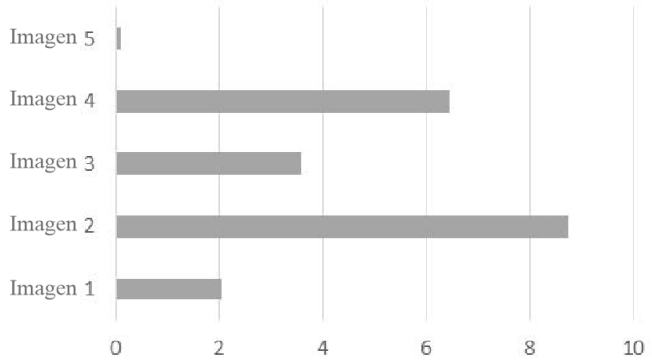


Fig. 6. Diferencias de los porcentajes entre las medias calculadas en las Fig 4 y 5.

5 Conclusiones

En este trabajo se presentó un algoritmo para el procesamiento de imágenes de plantas de cultivos de papas andinas. El objetivo de algoritmo propuesto fue detectar, algunas de las enfermedades y plagas, indicando además el grado de daño representado en dicha planta. En la implementación del algoritmo se utilizó varias de las técnicas propias del procesamiento de imágenes, para extraer las áreas dañadas de las plantas enfermas, separarlas de las zonas sanas, visualizarlas y calcular el grado de daño ocasionado por la enfermedad. El algoritmo propuesto pudo cumplir con el objetivo de detectar o no enfermedades en las plantas de cultivo de papas y determinar un porcentaje de daño con una exactitud aceptable. Sin embargo, observando las Fig. 4 y Fig.5 la imagen 4 es la que más sufre variabilidad en la detección del grado de la zona dañada, de acuerdo a la imagen original. Luego le sigue la imagen 2 y el resto de las imágenes con un grado de variabilidad más reducido. En el caso de la imagen 5 esta fue la más favorecida con el método propuesto de los valores medios. Por lo tanto, los resultados aún se deben mejorar, no se logró eliminar del todo el ruido, por lo que se podría utilizar otras técnicas para determinar más precisamente la zona del color de interés y así mejorar los resultados.

References

1. Alvarez, S., Cáceres, M., Calle, M., Córdoba, P., Giménez, L., López, A., Méndez, D., Quispe, J., Rodríguez, J., Singh, G. Conociendo los cultivos andinos de mi provincia. Calidad y educación alimentaria. Comisión Nacional de Bibliotecas Populares (CONABIP). ISBN 978-987-3926-00-6. Universidad Nacional de Jujuy (2015).
2. INTA informa. Las papas nativas son de Interés Nacional. Recuperado en junio 2019 de intainforma.inta.gov.ar/?p=14123 (2012).
3. Pérez, W., & Forbes, G. Guía de identificación de plagas que afectan a la papa en la zona andina. International Potato Center (2011).

4. Garnique, L. A. V., Cabrera, H. I. M., Chavarry, K. L. V., & Barco, M. J. V. Visión artificial: Aplicación de filtros y segmentación en imágenes de hojas de café. Ingeniería: Ciencia, Tecnología e Innovación, 1(2), 71-71 (2014).
5. Najafzadeh, Mohsen & Tavakoli, Maryam. Application of the Image Processing Technique for Separating Sprouted Potatoes in the Sorting Line. 10.13140/2.1.2093.0887. (2014).
6. Chahar, R., & Soni, P. A study of image processing in agriculture for detect the plant diseases. International Journal of Computer Science and Mobile Computing., 4, 581-587. (2015).
7. Janwale, Asaram. Digital Image Processing Applications in Agriculture: A Survey. International Journal of Advanced Research in Computer Science and Software Engineering. 5. 622. (2015).
8. Prakash, K., Saravanamoorthi, P., Sathishkumar, R., & Parimala, M. A study of image processing in agriculture. International Journal of Advanced Networking and Applications, 9(1), 3311. (2017).
9. Latha, M., Poojith, A., Reddy, B. A., & Kumar, G. V. Image processing in agriculture. International journal of innovative research in electrical, electronics, instrumentation and control engineering, 2(6). (2014).
10. Ordieres, Joaquín, Castejón Limas, Manuel, Martinez de Pisón Ascaibar, Francisco Javier, & lb Elías, Fernando. Técnicas y algoritmos básicos de visión artificial. ISBN 84-689-9345-X. Universidad de la Rioja. (2006).
11. Vélez, J., Moreno, A., Sánchez, A., Sánchez-Marín, J. Vision por Computador. 2da Edición. Editorial Dykinson. (2003).
12. Gonzalez, R. C., and R. E. Woods. "Digital image processing: Pearson prentice hall." Upper Saddle River, NJ 1 (2008).

De los autómatas finitos a las expresiones regulares: Un análisis áulico de identificación de expresiones equivalentes

Lic. Mathias G. Diaz Ogás[1], MSc. Prof. Elisa S.Oliva[1], Lic. Ana L. Molina[1], MSc. Prof. Nancy Alonso[1].

mathiasdiaz@conicet.gov.ar, eoliva@iinfo.unsj.edu.ar, lauramolina@outlook.com, nalonso@unsj-cuim.edu.ar.

[1] Departamento de Informática. Facultad de Ciencias Exactas, Físicas y Naturales. Universidad Nacional de San Juan, Argentina.

Abstract. En el presente trabajo se propone una discusión acerca del análisis de expresiones regulares (ER), de estructuras sintácticas diferentes, que surgen a partir de la aplicación de una técnica de conversión de un autómata finito (AF) en dichas ER. Cada ER tiene una forma diferente, y por lo tanto parecen representar a lenguajes diferentes. La propuesta del trabajo es verificar esta diferencia, obteniendo como resultado que estas ER resultan ser equivalentes luego de implementar la minimización de los AF derivados a partir de cada una de ellas. Este análisis y validación surge desde la identificación en el trabajo áulico de la cátedra de Matemática Discreta con alumnos de las carreras de Licenciatura en Ciencias de la Computación y en Licenciatura de Sistemas de Información. Se realizó una experimentación utilizando el software de código abierto JFLAP, que permite sintetizar y relacionar los conceptos teóricos con la identificación práctica de la equivalencia entre las ER trabajadas. Además, se validó dicha equivalencia mediante la aplicación de propiedades algebraicas de las ER, permitiendo obtener el mismo resultado, de que las ER trabajadas son equivalentes a pesar de ser totalmente distintas.

Keywords: Expresiones Regulares; Equivalencia entre Expresiones Regulares; Autómatas Finitos; Lenguajes Regulares; JFLAP.

1 Introducción

En la segunda mitad del siglo XX el lingüista estadounidense Noam Chomsky crea la corriente conocida como generativismo o teoría de lenguajes formales, que convirtió la lingüística en una ciencia. Según Chomsky, la persona que ha aprendido una lengua es capaz de formular enunciados que nunca antes ha escuchado, porque conoce las reglas según las cuales los enunciados deben formarse. Este conocimiento no es adquirido mediante el hábito (sería imposible) sino que es una capacidad innata. Todo ser humano que nace ya lleva consigo esta capacidad, que es la Gramática Universal, reglas gramaticales que rigen a todas las lenguas por igual [1]. Esta teoría creada no sólo podía aplicarse a los lenguajes naturales, sino que además facilitaba el estudio y

formalización de los lenguajes para programación de computadoras. La jerarquía de Chomsky es una clasificación de lenguajes formales de gran importancia en teoría de la computación, tal como se visualiza en la Tabla 1.

Tabla 1. Jerarquía de Chomsky

Gramáticas	Lenguajes	Máquinas de Estados
Tipo 0 de Chomsky	Lenguajes sin restricciones o recursivamente enumerables	Máquina de Turing
Tipo 1 de Chomsky	Lenguajes sensibles al contexto	Autómatas lineales acotados
Tipo 2 de Chomsky	Lenguajes independientes del contexto	Autómatas a pila
Tipo 3 de Chomsky	Lenguajes Regulares	Autómatas finitos

Las gramáticas de Tipo 3, según la jerarquía de Chomsky, permiten generar lenguajes que pueden describirse con herramientas matemáticas, usando el álgebra de expresiones regulares (ER) y con autómatas finitos (AF) [2]. Estas construcciones teóricas son objeto de este trabajo.

Los AF se pueden definir de formas diversas, incluyendo Autómatas Finitos Determinísticos (AFD), No Determinísticos (AFND), y No Determinísticos con transiciones vacías (AFND-Ɛ). Existen diferentes teoremas que prueban las equivalencias entre estos autómatas y las ER que representan a los Lenguajes Regulares [2]. La Figura 1 muestra algunas de dichas equivalencias.

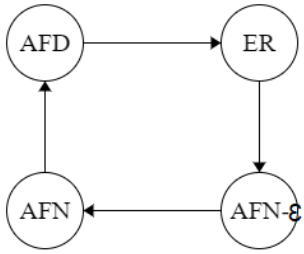


Fig. 1. Equivalencias entre los AF y ER.

En las carreras de Lic. en Ciencias de la Computación y Lic. en Sistemas de Información, a través de la asignatura Matemática Discreta, se busca que los alumnos logren formalizar nociones relacionadas con la computación, aplicadas en el estudio de los lenguajes formales y autómatas.

Como los alumnos no son matemáticos expertos, la intención es motivarlos a conocer los cimientos fundacionales de la informática (lenguajes formales, gramáticas, autómatas, etc.), subrayando los aspectos esenciales de la disciplina que permanecen inalterables ante el cambio tecnológico. Por ello, la metodología aplicada, está enfocada en que los alumnos aprendan a desarrollar su capacidad lógico-matemática, lograr una participación activa en cada uno de los procesos de enseñanza-aprendizaje y, con el uso de nuevas tecnologías, lograr una síntesis entre la teoría y la práctica.

Meirieu [3] sugiere entusiasmar a los alumnos, para lo cual deben emplearse estrategias "accesibles y difíciles al mismo tiempo". Estas estrategias son tales cuando

el alumno siente que es capaz de lograr el objetivo propuesto y percibe la existencia de una hipótesis que todavía no le es propia.

La práctica docente de la cátedra permite señalar que hay momentos en el proceso evolutivo de la asignatura que requiere sólo el fundamento matemático y otros en los que las acciones de las herramientas tecnológicas mejoran los procesos de enseñanza–aprendizaje.

En la asignatura se hace uso del software libre JFLAP[1], con el objeto de que el alumno pueda comprender el funcionamiento de los modelos teóricos incluidos en la asignatura. Además, permite visualizar simbólica y gráficamente la relación entre las ER y las máquinas de estados finitos. Esta herramienta se trabajó en otra investigación, donde se compara la construcción de un AFD que responde a un cierto lenguaje, realizado con las nociones teórico-prácticas adquiridas y la misma elaboración con el software JFLAP. Y se discute el uso de procesos no explícitos en el software que llevaron a obtener distintos AFD que, con intervenciones teóricas, se pudieron concluir en equivalencias entre los autómatas construidos [4].

La competencia matemática que se establece en el presente trabajo mediante una experiencia de cátedra se centra en desarrollar habilidades en razonamiento matemático y la utilización del lenguaje algebraico. Sobre esta línea se estable el objetivo del trabajo como el análisis de equivalencia entre ER a través del trabajo con propiedades algebraicas y equivalencias teóricas de las operaciones involucradas.

2 Problemática

Aunque las ER describen los lenguajes de manera completamente diferente a como lo hacen los AF, ambos conceptos representan exactamente el mismo conjunto de lenguajes, que se denominan Lenguajes Regulares. Todo lenguaje definido mediante uno de estos autómatas puede ser representado mediante una ER. Para demostrar esto, se puede suponer que el lenguaje es aceptado por algún AF [2].

La construcción de una ER para definir el lenguaje de cualquier AF consiste básicamente en construir expresiones que describan conjuntos de cadenas que etiqueten ciertos caminos del diagrama de transiciones de un AF. El método que se va a utilizar para construir las ER implica a un "algoritmo de eliminación de estados" [5]. Dado un AF como dato si se elimina un estado *S*, todos los caminos que pasen a través de *S* ya no existirán en el autómata. Para evitar perder el lenguaje del autómata, se debe incluir, sobre un arco que vaya directamente desde un estado *Q* hacia un estado *P*, las etiquetas de los elementos de los caminos que vayan desde *Q* a *P*, pasando por el estado eliminado *S*. Dado que la etiqueta de este arco ahora puede implicar cadenas en lugar de sólo símbolos, no se puede simplemente enumerar las cadenas como una etiqueta. Para ello la forma de representar dichas cadenas son las ER.

La Figura 2 muestra cómo queda un AF al aplicar el algoritmo y eliminar un estado *S*. Se puede ver claramente que las transiciones que pasan por *S* son reemplazadas por

[1] JFLAP es un software libre para experimentación con lenguajes formales y máquinas de estado como AFD, AFN, Autómatas Push Down, Máquinas Turing de cintas múltiples, expresiones regulares y diferentes tipos de gramáticas. Disponible en: http://www.jflap.org/

una transición con etiqueta con una ER que representa un camino equivalente, para el lenguaje regular subyacente, pero sin el estado *S*.

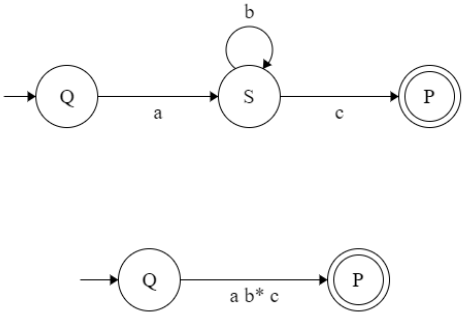


Fig. 2. Eliminación del estado S del AF.

Entonces, se pueden considerar autómatas que tienen expresiones regulares como etiquetas. Es posible de esta manera, identificar que el lenguaje de un AF es la unión, para todos los caminos desde el estado inicial hasta un estado aceptación, del lenguaje formado mediante la concatenación de los lenguajes de las expresiones regulares a lo largo de cada camino [2]. Los elementos de arcos adyacentes se etiquetan como concatenaciones de estos elementos, mientras que los elementos de arcos paralelos son uniones de "expresiones", que en la sintaxis de las ER se describen con el símbolo *"+"*. Finalmente, los elementos de los bucles o "loops" se representan con la clausura de Kleene[2] de todos los elementos que se encuentren en el bucle.

Para poder obtener la ER que representa al Lenguaje Regular de un AF se utiliza un nuevo tipo de AF denominado Autómata Finito No Determinístico Generalizado (AFNG) [2]. Un AFNG puede obtenerse fácilmente desde un AF dado como dato (Determinístico o No Determinístico). La transformación consiste en agregar 2 estados nuevos que se denominaran S y F.

El estado S (Start) representará un nuevo estado inicial. Este no contendrá ninguna transición de entrada, y se conectará al estado inicial del AF original, mediante una transición de cadena vacía (Ɛ). No contendrá ningún otro tipo de transición de salida. El estado F (Finish) representará un nuevo y único estado final. Este no contendrá ninguna transición de salida, y se conectaran a él todos los estados finales del AF original, mediante una transición de cadena vacía (Ɛ). No contendrá ningún otro tipo de transición de entrada.

Una vez obtenido el AFNG, se aplica el algoritmo de eliminación de estados para obtener las etiquetas que representan a la ER de cada transición en el autómata. Se elimina de un estado a la vez, hasta obtener una única transición que unirá los estados S y F, donde en la etiqueta de la misma podrá identificarse la ER del lenguaje regular representado por el AF original.

[2] La clausura de Kleene (estrella Kleene o cierre estrella) es una operación unaria que se aplica sobre un conjunto de cadenas de caracteres o un conjunto de símbolos o caracteres (alfabeto), y representa el conjunto de las cadenas que se pueden formar tomando cualquier número de cadenas del conjunto inicial, posiblemente con repeticiones, y concatenándolas entre sí.

El algoritmo de eliminación de estados no indica ni obliga a eliminar dichos estados en un determinado orden. Realizando pruebas en el aula, con diferentes AF, se pudo comprobar que, alterando el orden de eliminación de estos estados, se obtienen diferentes ER para un mismo autómata. De esta manera surge la siguiente pregunta que es el objeto de investigación de este trabajo, ¿Las ER resultantes de la aplicación del algoritmo de eliminación de estados son equivalentes? De ser así, estas ER representan al mismo Lenguaje Regular.

Para poder comprobar esto se utiliza el software JFLAP, que permite obtener el AFN- Ɛ equivalente a las ER, y mediante las equivalencias entre autómatas (Ver Fig. 1) lograr el AFD de cada expresión. Una vez obtenido el AFD de cada expresión se aplica el método de minimización de autómatas, para obtener así el AFD mínimo equivalente y poder comprobar fehacientemente si estos representan al mismo lenguaje regular. La prueba de equivalencia es eficiente y decidible, ya que la forma mínima de un AFD es única hasta el isomorfismo [5]. Además, también se valida la equivalencia mediante la aplicación de propiedades de las ER de forma algebraica.

3 Resultados

Establecida la problemática descrita en la Sección 2, se seleccionó un AF del conjunto ejercicios realizados en el aula, para poder comprobar que las expresiones resultantes luego de la aplicación del algoritmo de eliminación de estados, son equivalentes.

Se define entonces el siguiente AFN M = <Q, ∑, δ, q0, F>. Con conjunto de estados Q = {q0, q1}; conjunto de símbolos de entrada (alfabeto) ∑ = {a, b}; estado inicial q0; conjunto de estados finales o de aceptación F = {q1}; función de transición δ definida a través del diagrama de transición de estados de M, que se describe en la Figura 3.

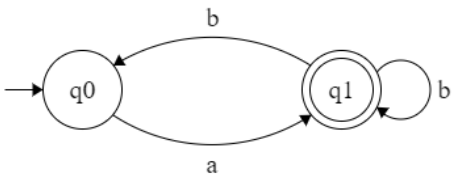


Fig. 3. Diagrama de transición de estados del AFD M.

Definido el AFN M, se procede a aplicar el algoritmo de eliminación de estados. Para ello, como fue descrito en la Sección 2, es necesario transformar el AFN M en un AFNG. Se define entonces el AFNG M' = <Q, ∑, δ, q0, F>. Con conjunto de estados Q = {S, F, q0, q1}; conjunto de símbolos de entrada (alfabeto) ∑ = {a, b}; S como estado inicial; conjunto de estados finales o de aceptación F = {F}; función de transición δ definida a través del diagrama de transición de estados de M', que se describe en la Figura 4.

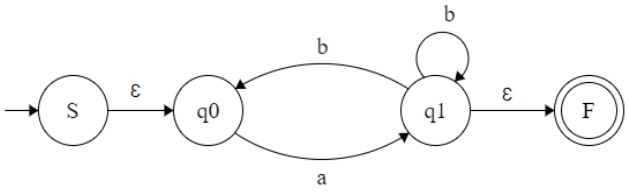


Fig. 4. Diagrama de transición de estados del AFNG M'.

Obtenido el AFNG M', se aplica sobre éste, el algoritmo de eliminación de estados. Este algoritmo se aplicará dos veces, en una primera instancia se eliminarán los estados en orden q0, q1. En la segunda instancia se eliminarán en orden q1, q0.

La Figura 5 describe el algoritmo de eliminación de estados para la primera instancia. En Fig. 5 (a) se puede ver el AFNG M' obtenido. En Fig. 5 (b) se elimina el estado q0. Es posible ver como las entradas de q0 (desde S y desde q1), se conectan con cada salida de q0 (en este caso una sola, hacia q1). De esta manera S queda conectado con q1 con etiqueta "*Ɛ a"* (Ɛ no se coloca en la transición), y en q1 se genera un bucle (debido a que q1 es entrada y salida de q0) con etiqueta "*b a"*. De esta manera se conserva el valor de los símbolos para cada camino que involucraba al estado eliminado q0. En Fig. 5 (c) puede verse como los dos bucles de q1 que unen en una misma transición etiquetada como la unión de las expresiones de cada transición *"b a + b"*. Esto es debido a que, dos caminos que inician en un mismo estado y terminan en un mismo estado (son paralelos), pueden ser representados como la unión de las expresiones de cada camino. En Fig. 5 (d) se elimina q1, la entrada a este estado desde S queda conectada a la única salida hacia F. La expresión del bucle en q1 representa a una clausura de Kleene para toda la expresión. La ER resultante para esta primera instancia de implementación del algoritmo corresponde a la etiqueta resultante en la transición entre los estados S y F, se denomina a dicha expresión, ER1 = *a (b a + b)* *.

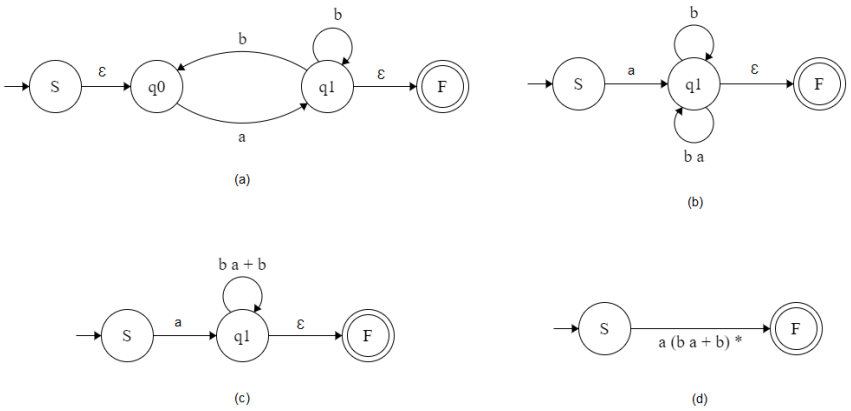


Fig. 5. Aplicación del algoritmo de eliminación de estados sobre M'. Eliminando los estados en orden q0, q1.

La segunda instancia de aplicación del algoritmo, se realiza eliminando los estados en orden q1, q0. La Figura 6 describe el algoritmo para esta segunda instancia. En Fig. 6 (a) se visualiza el AFNG M' obtenido a partir del AFN original. En Fig. 6 (b) se elimina el estado q1. Se puede ver como las entradas de q1 (en este caso una sola desde q0), se conectan con cada salida de q1 (hacia q0 y F). De esta manera q0 queda conectado con F con etiqueta "*a b * Ɛ"* (Ɛ no se coloca en la transición), y en q0 se genera un bucle (debido a que q0 es entrada y salida de q1) con etiqueta "*a b* b"*. De esta manera se conserva el valor de los símbolos para cada camino que involucraba al estado eliminado q1. En Fig. 6 (c) se elimina q0, la entrada a este estado desde S queda conectada a la única salida hacia F. La expresión del bucle en q0 representa a una clausura de Kleene para toda la expresión. La ER resultante para esta segunda instancia de implementación del algoritmo corresponde a la etiqueta resultante en la transición entre los estados S y F, se denomina a dicha expresión, ER2 = *(a b* b) * a b**.

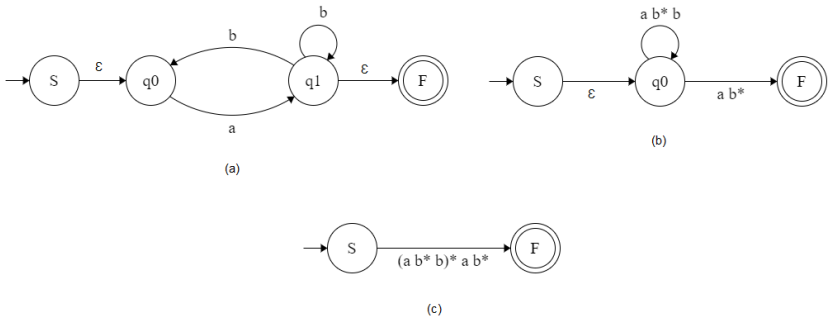


Fig. 6. Aplicación del algoritmo de eliminación de estados sobre M'. Eliminando los estados en orden q1, q0.

Se puede ver que las expresiones ER1 = *a (b a + b) **, y ER2 = *(a b* b) * a b**, son aparentemente, totalmente diferentes. Sin un análisis profundo no se podría establecer si estas expresiones representan el mismo Lenguaje Regular. Esto propuso la discusión en el aula de que si el resultado del algoritmo de eliminación de estados, para obtener una ER a partir del AFN M, es realmente correcto. Los resultados de las ER variaron al aplicar un orden distinto de eliminación, por lo tanto, se puede inducir que estas expresiones son equivalentes. Es decir que representan al mismo lenguaje. Pero es necesario probar que esto es verdadero.

Para probar si estas ER son equivalentes se hizo uso de la tecnología, mediante JFLAP. En este software se cargan las expresiones ER1 y ER2 para obtener los AFD equivalentes a dichas expresiones. La herramienta realiza esta transformación en el orden de equivalencias mostradas en la Fig. 1. De cada ER obtiene un AFN-*Ɛ*, luego sobre este último se obtiene el AFN, y finalmente el AFD equivalente. De esta manera JFLAP permite acortar un largo proceso de cálculo de equivalencias entre autómatas, que, si bien puede realizarse de forma manual, si la longitud de la ER es extensa puede ser un proceso complejo y largo de resolver.

En la Figura 7 (a) se visualiza el AFD resultante para la ER1, y en la Figura 7 (b) el AFD correspondiente a la ER2. Ambos AFD obtenidos desde JFLAP. Como se puede ver ambos AFD son diferentes, tienen diferente cantidad de estados y transiciones.

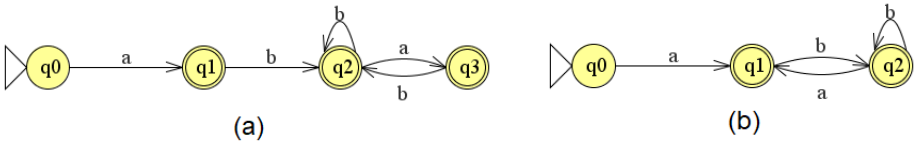


Fig. 7. (a) AFD equivalente a la ER1 = *a (b a + b) **. **Fig. 7.** (b) AFD equivalente a la ER2 = *(a b* b) * a b**.

Aplicando la técnica de minimización de AFD, es posible particionar el conjunto de estados de cualquier autómata en grupos de estados mutuamente indistinguibles. Si se reemplaza cada partición por un solo estado que la represente, se obtiene un AFD equivalente que tiene menos estados que cualquier AFD que reconoce el mismo lenguaje. Al aplicar la minimización a los AFD correspondientes a ER1 y ER2, se puede comprobar finalmente si estas expresiones son equivalentes.

Nuevamente, haciendo uso de la tecnología mediante JFLAP, se pueden obtener los AFD mínimos equivalentes para los AFD de las Fig. 7 (a) y Fig. 7 (b). Aplicada esta acción sobre ambos autómatas se obtuvo el resultado de que ambos minimizan a un mismo AFD equivalente. En la Figura 8 se puede ver el resultado. Para no confundir los autómatas, en el AFD mínimo se colocaron nombres de estados diferentes a los anteriores.

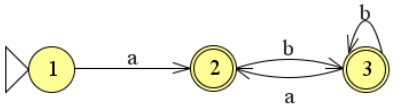


Fig. 8. AFD mínimo equivalente a los AFD de las ER1 y ER2.

De esta manera, es posible afirmar que es verdadera la pregunta que fue planteada en el presente trabajo, surgida de la discusión áulica, donde se inducía que ambas expresiones eran equivalentes. Al determinar el mismo AFD mínimo equivalente, desde cada ER, se concluye que las expresiones ER1 y ER2 representan el mismo lenguaje. Por otro lado, se puede ver como el AFD mínimo, coincide con el AFD que representaba a la ER2 (Fig. 7 b). Es decir que el AFD de la ER2 realmente no fue minimizado, ya que justamente es el AFD mínimo posible de obtener para el lenguaje de las ER analizadas.

Otra forma de verificación de que efectivamente ambas expresiones ER1 y ER2 son equivalentes, es calculando la gramática regular equivalente a los AFD obtenidos para cada ER. Luego, mediante JFLAP obtener el lenguaje que genera cada gramática, permitiendo comparar si ambos lenguajes son iguales.

Para ello se modificaron los nombres de los estados debido a las restricciones del software. En la Tabla 2 se muestran las equivalencias entre los nombres de los estados de los AFD de las expresiones ER1 y ER2, y los nombres de los símbolos no terminales de las gramáticas regulares equivalentes.

Tabla 2. Equivalencia entre los estados de los AFD de ER1 y ER2, y los símbolos No Terminales de las gramáticas regulares equivalentes.

Estados de los AFD	Símbolos No Terminales
q0	S
q1	A
q2	B
q3	C

En la Figura 9 (a) se puede observar la gramática regular equivalente a ER1, definida como GER1 = <N1, T1, S, P1>. Con símbolos no terminales N1 = {S, A, B, C}; símbolos terminales T1 = {a, b}; símbolo inicial o axioma S; y reglas de producción P1 definidas en la Fig. 9 (a). En la Figura 9 (b) se puede observar la gramática regular equivalente a ER2, definida como GER2 = <N2, T2, S, P2>. Con símbolos no terminales N2 = {S, A, B}; símbolos terminales T2 = {a, b}; símbolo inicial o axioma S; y reglas de producción P2 definidas en la Fig. 9 (b).

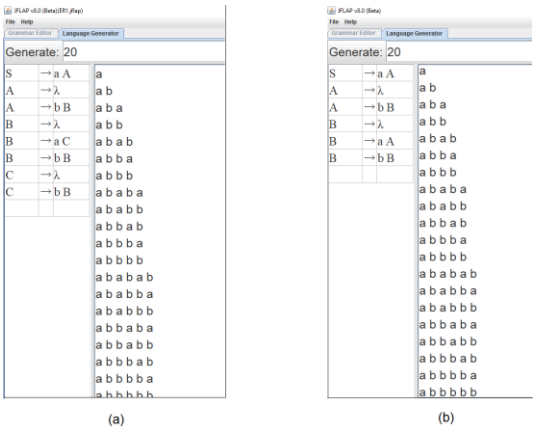


Fig. 9. Reglas de producción de GER1 y GER2, y las primeras 20 palabras del lenguaje generado por cada gramática.

Como podemos ver, ambas gramáticas GER1 y GER2 son diferentes, poseen diferentes elementos no terminales y sus reglas de producción son distintas. Desde JFLAP se listaron los lenguajes que generan las gramáticas GER1 y GER2. En las Fig. 9 (a) y (b) se muestran las primeras 20 palabras de cada lenguaje, por ser un lenguaje infinito no es posible mostrar el lenguaje completo en una pantalla. Como puede apreciarse en la figura, los lenguajes generados por ambas gramáticas son iguales. De esta manera se refuerza la conclusión obtenida con la minimización de los AFD de ER1 y ER2, de que ambas expresiones son equivalentes, y como se definió, representan el mismo lenguaje.

Para finalizar, y dar un respaldo teórico adicional al resultado obtenido, se aplicaron propiedades de las ER sobre ER1 y ER2 a fin de determinar que ambas expresiones son equivalentes, a pesar de ser totalmente distintas. Las siguientes propiedades sobre ER son las utilizadas para probar dicha equivalencia:

$$(1)\ \ (r+s)^* = (r^* s^*)^* = (r^* + s^*)^* = (r^* s)^* r^* = r^*(s\, r^*)^* \qquad (2)\ \ (r\, s)^* r \equiv r(s\, r)^*$$

Dadas las propiedades anteriores, se aplican sobre ER2 = *(a b* b) * a b** como sigue:

$$(a\, b^* b)^* a\, b^* \equiv_{(2)} a\, b^* (b\, a\, b^*)^* \equiv a[\, b^* (b\, a\, b^*)^*\,] \equiv_{(1)} a[\, (b + b\, a)^*\,] \equiv a\, (b + b\, a)^*$$

$$\therefore\ (\boldsymbol{a\, b^* b})^* \boldsymbol{a\, b^*} \equiv \boldsymbol{a\, (b + b\, a)^*}$$

El resultado de aplicar las propiedades (1) y (2) sobre ER2 da como resultado exactamente la misma expresión que ER1 = *a (b a + b) **. Verificando entonces que ER1 y ER2 son equivalentes.

4 Conclusiones

El uso de TICs en el ámbito de enseñanza universitaria aporta en la gestión educativa, espacios para apropiación de contenidos y de autoevaluación de los estudiantes. Su utilización en la asignatura permite incursionar en nuevas metodologías de trabajo que no limiten las prácticas áulicas con sólo el uso de recursos tradicionales. Los argumentos teóricos son la base en este trabajo de investigación, sin cuya fundamentación no se podría haber discernido si las ER, que se obtuvieron desde un AFD, eran consistentes con la discusión surgida en el aula acerca de si estas expresiones ER1 y ER2 eran equivalentes. Los resultados obtenidos muestran que efectivamente ambas expresiones son equivalentes, es decir que representan el mismo lenguaje.

Este resultado permite valorar el uso de JFLAP como verificador de ejercicios para la realización de un trabajo autónomo. Se puede ponderar de esta manera que el uso de la tecnología puede, y es razonable que, sea acompañado por los procesos teóricos. Y de la misma manera, la teoría aprendida en las aulas tiene un alto beneficio en el aprendizaje si es acompañada con herramientas de uso tecnológico que permitan validar los procesos estudiados.

Como trabajos futuros se plantea ampliar la simulación propuesta con un análisis de otras situaciones prácticas más complejas que requieren una formalización teórica para poder verificar los resultados obtenidos con JFLAP. Como así también establecer una estadística del aporte que brinda el uso de la herramienta respecto al rendimiento de los alumnos para la validación de conceptos teóricos en el aula.

5 Referencias

[1] H. W. Peter and N. Chomsky, “Aspects of the Theory of Syntax,” *Mod. Lang. Rev.*, 1968.
[2] J. Hopcroft, R. Motwani, and J. Ullman, *Teoría de autómatas, lenguajes y computación*. 2007.
[3] P. Meirieu, *Aprender, sí, pero ¿cómo?* 2002.
[4] E. Oliva, M. Diaz, A. Molina, N. Alonso. “Lenguajes regulares y autómatas finitos determinísticos asociados. discusión de implementación con JFLAP” *Congreso Internacional de Ciencias de la Computación y Sistemas de Información*. Mendoza, Argentina. 2018.
[5] M. Sipser, *Introduction to the Theory of Computation*. 2013.

Una aproximación para construir sistemas de programación de la producción utilizando servicios en la nube

Daniel Díaz[1,2], Sandra Oviedo[1,2], Fernando Guardia[1,2], Francisco Ibañez[2], Juan Aranda[2], Nancy Alonso[2], Juan Cuneo[2]

[1] Laboratorio de Informática Aplicada a la Innovación
Instituto de Informática, [2]Departamento de Informática
FCEFN- UNSJ

ddiaz, soviedo, fguardia, fibanez, jaranda, nalonso,
jcuneo@iinfo.unsj.edu.ar
San Juan, Argentina

Resumen.
La computación en la nube y la industria 4.0 están produciendo una revolución en el dominio de los sistemas de planificación y programación de la producción. Por un lado computación en la nube ha democratizado el acceso a las tecnologías de optimización, logrando que muchas medianas y pequeñas empresas aprovechen lo mejor de la tecnología existente, sin necesidad de incurrir en los grandes costos de licenciamiento de estas tecnologías y sin realizar la cuantiosa inversión en hardware que requieren dichas tecnologías. Por otro lado, en su definición de Industria 4.0 coloca a los sistemas ciber-físico como uno de los elementos básicos del paradigma. En este paradigma el proceso de planificación y programación de la producción se da en el contexto de un sistema ciber-físico de producción y por lo tanto el sistema de programación de la producción debe adecuarse a esta nueva forma de producción y los nuevos requerimientos que esta demanda. Este trabajo se presenta una aproximación para desarrollar sistemas de programación de la producción utilizando servicios en la nube.

Palabras clave: scheduling systems, optimización, cloud services

1 Introducción

Los beneficios de la computación en la nube ha llegado al dominio de los sistemas de optimización, dicho de otra forma existen servicios en la nube que nos permiten acceder a complejos motores de optimización (optimization solvers) que antes eran prohibitivos por sus costos de licenciamiento a pequeñas y medianas empresas. Estas herramientas son plataformas que permiten modelar, diseñar e implementar soluciones para problemas de optimización. Estas herramientas brindan diferentes interfaces o puertas de entrada, permitiendo de esta manera conectarse a diversos ecosistemas informáticos tales

como Java, NodeJs, .Net, entre otros. Esta característica permite una integración tecnológica con diversas herramientas y servicios que están disponibles en la nube. Los problemas de programación de la producción que de aquí en adelante denominaremos problemas de scheduling, son un caso particular de los problemas de optimización. En este trabajo se presenta un marco de trabajo que intenta servir de referencia para construir sistemas de scheduling. El marco está basado en una arquitectura monolítica y está pensado para construir sistemas de pequeño o mediano porte donde no ser requiere una administración compleja de la escalabilidad. El trabajo comienza haciendo una introducción a los sistemas de scheduling en dominios industriales, continua con una revisión de los diferentes modelos de servicios de la computación en la nube, despliegue de aplicaciones en la nube y arquitectura monolítica y las herramientas de desarrollo más importantes utilizadas en este trabajo. Luego, se revisan los conceptos referidos a optimizadores y servicios de optimización en la nube. Seguidamente se presenta el marco de trabajo para construir sistemas de programación de la producción utilizando servicios en la nube y un prototipo de un sistema de scheduling. Por último, se muestran las conclusiones y trabajos futuros.

2 Sistemas de scheduling en dominios industriales

Los problemas de Scheduling aparecen en diferentes dominios, este proyecto trata con problemas de scheduling que surgen en el dominio industrial y que en la industria que se denominan production scheduling o manufacturing scheduling.

La definición más clásica de la palabra scheduling dice : "Scheduling es el problema de asignar recursos limitados a tareas en el tiempo con el objeto de optimizar uno o más objetivos" [1]. En la industria, un sistema de scheduling es el corazón del sistema de planificación y control de la producción. Según la arquitectura jerárquica ANSI/ISA 95 que guía la organización de una planta industrial el sistema de scheduling interactúa con otros sistemas de fabricación, tales como ERP (Entreprise Resource Planning) y MES (Manufacturing Execution System). En los últimos años, con el surgimientos de la industria 4.0 el proceso de planificación y programación de la producción está en constante cambio debido al que las funciones del MES han sido absorbidas por el CPPS (Cyber-physical production systems) [2], [3]. Existe una literatura muy extensa sobre scheduling y se ha escrito mucho sobre modelos y algoritmos que resuelven problemas de scheduling. Sin embargo, poco son los artículos que tratan de cómo traer estos modelos y algoritmos a implementaciones reales. Esto se conoce como el "hueco o gap" entre la teoría y la práctica de scheduling [4]. En [5] se presenta una perspectiva más actual sobre este tema, en dicho trabajo se plantea como camino para cerrar este hueco entre la teoría y la práctica ese la implementación de una pieza de software que satisfaga las necesidades que tiene una empresa en el área de scheduling. Esto implica llevar a cabo un proceso de desarrollo de software para obtener un producto final, es decir un sistema de scheduling y que ayude a la toma de decisión. Según Yen y Pinedo [6] un sistema de scheduling se compone tres módulos: (1) módulos de base datos y base de conocimiento, (2) módulos del motor de scheduling y (3) módulo de interface de usua-

rio. La parte esencial de todo sistema de scheduling es el desarrollo del motor de scheduling, es donde se vinculan las necesidades que tiene la empresa con los modelos y algoritmos de scheduling.

En [7] se describen un conjunto de necesidades o requisitos que debe reunir un sistema de scheduling, algunos de los cuales se refieren a los algoritmos de secuenciación y al uso del sistema. Con respecto a los algoritmos de secuenciación, éstos deben ser flexibles y configurables. Deben ser capaces de manejar diversos criterios (scheduling multicriterio). A estos requisitos se les debe agregar otros tales como facilitar el mantenimiento y la reducción de costos de desarrollo.

Por último , [8] intenta integrar esta definiciones definiendo a un manufacturing scheduling como el proceso de toma de decisión que consiste de asignar un conjunto de operaciones/tareas requeridas para manufactura un conjunto de productos con los recursos existentes en una planta, como así también los tiempos necesario para iniciar estas operaciones o tareas. Un Schedule o plan se define como un conjunto específico de asignaciones de operaciones o tareas a los recursos sobre una escala temporal.

3 Computación en la nube.

Según [9] Cloud Computing es un modelo para habilitar acceso conveniente por demanda a un conjunto compartido de recursos computacionales configurables, por ejemplo, redes, servidores, almacenamiento, aplicaciones y servicios, que pueden ser rápidamente aprovisionados y liberados con un esfuerzo mínimo de administración o de interacción con el proveedor de servicios. Este modelo de nube promueve la disponibilidad y está compuesto por cinco características esenciales, tres modelos de servicio y cuatro modelos de despliegue.” Esta definición destaca que cloud es un modelo de prestación de servicios cuya principal orientación es la escalabilidad. Desde una perspectiva económica los servicios son elásticos, donde el usuario paga por lo que consume. Este modelo está compuesto por cinco características esenciales, tres modelos de servicio y cuatro modelos de despliegue. Las características esenciales son las que describen en esencia al modelo de cloud computing, en [10] se presentan como Auto-servicio por demanda, Acceso amplio desde la red, Conjunto de recursos, Rápida elasticidad, Servicio medido. Los Modelos de Servicios se refieren a la forma en la cual están organizados los servicios de cloud computing, son tres a saber *IaaS* (Infrastructure as a Service), *PaaS* (Platform as a Service), *SaaS* (Software as a Service), una descripción detallada se puede ver en [11]. Los Modelos de Despliegue hacen referencia a cómo se monta y pone en funcionamiento el sistema. Estos se presentan en [10] son *Nube privada*, *Nube comunitaria*, *Nube pública*, *Nube híbrida*.

3.1 Despliegue de aplicaciones en la nube – Arquitectura Monolitica

El desarrollo de aplicaciones basadas en los modelos IaaS o PaaS al momento del despliegue de las mismas se debe considerar los requerimientos sobre las capacidades de la computación en la nube que necesitan de las aplicaciones, tales como escalado au-

tomático, performance, despliegue en caliente, disponibilidad, monitoreo dinámico entre otros. Entre otras cosas una arquitectura de una aplicación en la nube está influenciada por las capacidades de la computación en la nube que se desean aprovechar. Generalmente, el desarrollo de aplicaciones en la nube es llevado a cabo siguiendo el estilo arquitectónico monolítico. Una aplicación para la nube basada en el estilo monolítico normalmente se compone de tres partes [12], en el lado del cliente está la interfaz de usuario, la cual normalmente se compone por archivo HTML, un archivo javascript y una hoja de estilo. Estos archivos se ejecutan sobre un browser instalado en la máquina del usuario. El segundo componente es una base de datos, que contiene todos los datos con los que trabaja la aplicación. Por último, en el lado del servidor existe una aplicación que gestiona las solicitudes HTTP, ejecuta la lógica del dominio, lee, escribe y modifica datos de la base de datos, y construye los archivos HTML que son enviados al cliente. Esta aplicación es monolítica, todo está empaquetado como un único ejecutable. En la nube es sencillo su despliegue y mantenimiento, es importante mencionar que cualquier cambio en el sistema implica construir y desplegar una nueva versión de la aplicación en el servidor. En desarrollo la aplicación completa está contenida en un solo proyecto y se utiliza la tecnología de contenedores no hay se necesitan cambios para pasar del servidor de desarrollo al servidor de producción.

3.2 Meteor.js

Meteor [13] es un framework de código abierto para el desarrollo de aplicaciones web, utiliza Nodejs. Meteor permite la creación rápida de prototipos y produce multiplataforma de código (web, Android, iOS). Se integra con MongoDB, utiliza el Protocolo de Datos Distribuidos y un patrón de publicación-suscripción para propagar automáticamente los cambios de datos a los clientes sin que el desarrollador tenga que escribir ningún código de sincronización. En el cliente, Meteor depende de jQuery y se puede utilizar con cualquier biblioteca del widget de JavaScript UI. Meteor es un entorno ultra-simple para la construcción de sitios web modernos. Meteor proporciona una forma fácil de hacer todo esto: la reactividad, que, en esencia es un enfoque declarativo. De esta forma, podemos definir la relación entre los objetos una sola vez y podemos estar seguros de que se mantendrán siempre sincronizados, en vez de tener que especificar el comportamiento de cada uno de los posibles cambios. Este es un concepto realmente poderoso, ya que, en un sistema de tiempo real puede haber muchas entradas y todas pueden cambiar de forma impredecible. Con “declarativo”, se quiere decir que, cuando se renderiza HTML basado en una o varias fuentes de datos reactivos, Meteor se hace cargo de la tarea de sincronizar las fuentes y, de forma transparente, asumir el trabajo sucio de mantener la interfaz de usuario actualizada.

3.3 Mongo DB

Mongo DB es una base de datos NoSQL. Es una base de datos ágil que permite cambiar los esquemas rápidamente cuando las aplicaciones evolucionan, proporcionando siempre la funcionalidad que los desarrolladores esperan de las bases de datos tradicionales,

tales como índices secundarios, un lenguaje completo de búsquedas y consistencia estricta.

MongoDB fue creado para brindar escalabilidad, rendimiento y gran disponibilidad, permite escalar de una implantación de servidor único a grandes arquitecturas complejas de centros multidatos. MongoDB brinda un elevado rendimiento, tanto para lectura como para escritura. Según [14] MongoDB guarda la estructura de los datos en documentos tipo JSON con un esquema dinámico llamado BSON, no existe un esquema predefinido. Los elementos de los datos son llamados documentos y se guardan en colecciones, las cuales pueden tener un número indeterminado de documentos. Comparando con una base de datos relacional, se puede decir que las colecciones son como tablas y los documentos son registros. La diferencia es que en una base de datos relacional cada registro en una tabla tiene la misma cantidad de campos, mientras que en MongoDB cada documento en una colección puede tener diferentes campos. En un documento, se pueden agregar, eliminar, modificar o renombrar nuevos campos en cualquier momento, ya que no hay un esquema predefinido. MongoDB soporta la búsqueda por campos, consultas de rangos y expresiones regulares. El sistema permite al usuario personalizar una consulta en tiempo real y pueden devolver un campo específico del documento pero también puede ser una función definida por el usuario.

4 Software para solucionar problemas de optimización (Optimization Solvers)

Los solucionares para optimización, del término en inglés “optimization solvers”, son plataformas que permiten modelar, diseñar e implementar soluciones para problemas de optimización. Los problemas de optimización se pueden clasificar en optimización si restricciones y con restricciones. Los problemas de scheduling son un caso particular de los problemas de optimización con restricciones. Los solucionadores para optimización implementan diferentes algoritmos y técnicas para solucionar problemas de optimización con restricciones. Las diferentes técnicas que utilizan los solucionadores para resolver los problemas de optimización se pueden clasificar en técnicas provenientes de la investigación de operaciones y técnicas provenientes de la inteligencia artificial o también llamadas metaheurísticas. Entre las técnicas provenientes de la investigación de operaciones se tienen las siguientes: Programación lineal (LP), Programación lineal mixta entera (MILP), Programación cuadrática (QP), Programación cuadrática entera mixta (MIQP), Programación restringida cuadráticamente (QCP), Programación entera mixta restringida cuadráticamente (MIQCP),

Las técnicas provenientes de la Inteligencia Artificial más implementadas por los solucionadores son: Programación por restricciones (Constraint Programming), Algoritmos genéticos, Colonia de Hormigas, Búsqueda Tabú.

4.1 IBM Decision Optimization on Cloud

Es la solución en la nube de IBM que utiliza como motor de soluciones CPLEX y CP Optimizer [15], con IBM Decision Optimization on Cloud (**DOcplexcloud**) [16] se pueden solucionar problemas de optimización en la nube sin necesidad de instalar o configurar ningún software de optimización, simplemente accediendo a sus servicios en la nube. Brinda el poder de los optimizadores CPLEX en un entorno de nube de autoservicio. Este servicio puede llamar a los optimizadores CPLEX basados en la nube desde diferentes formas.

- Se puede llamar a un solucionador desde una aplicación usando una API DOcplexcloud
- Se puede cargar un problema de optimización en el sitio web de DropSolve.
- Se puede configurar el acceso al servicio desde una herramienta de modelado propietaria

El problema de optimización necesita ser descripto en OPL, Optimization Programming Language, este es un lenguaje de programación que proporciona una descripción matemática natural de los modelos de optimización [17]. En OPL, un modelo de optimización queda descripto mediante dos archivos, el archivo del modelo y el archivo de datos. La primera manera de usar este servicio en la nube de IBM, la API de Decision Optimization on Cloud, también llamada DOcplexcloud APIs, hace posible invocar a un motor de optimización desde su aplicación para solucionar un problema de optimización. Solucionar un problema involucra tres pasos básicos tal como se explica en la **¡Error! No se encuentra el origen de la referencia.** que se presenta a continuación.

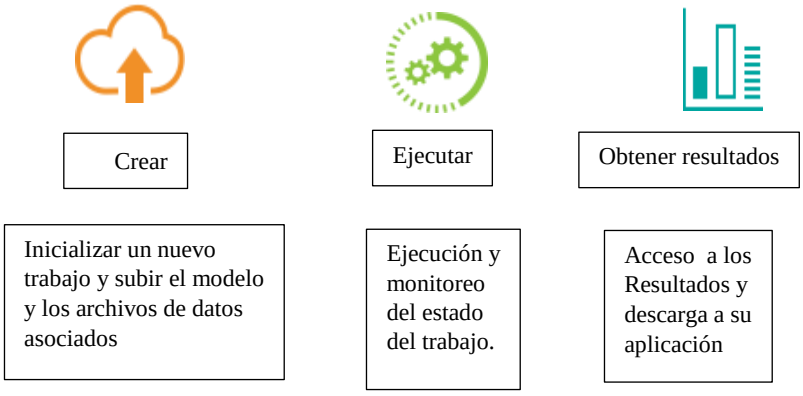


Fig. 1. Pasos para ejecutar un modelo en la DOcplexcloud API

Esta herramienta presenta todos los beneficios de computación en la nube para resolver problema de optimización, entre los cuales están los problemas de scheduling. No se necesita invertir en costosas licencias, el servicio es tasado por horas de uso, sin embargo, si una solución ocupa mucho tiempo por cada ejecución será necesario hacer un análisis del licenciamiento versus la solución en la nube. Si el uso está destinado a

un solo cliente cuyo algoritmo ocupa solo segundos y son pocas las ejecuciones por día, sin duda es una buena solución este servicio. Esto permite a las pequeñas empresas, cuyos problemas requieren muy poco procesamiento, acceder a motores de optimización cuyo licenciamiento es privativo por sus costes.

5 Marco para desarrollar sistemas de programación de la producción

En la **¡Error! No se encuentra el origen de la referencia.** se presenta la arquitectura de un marco de trabajo para desarrollar sistemas de scheduling utilizando servicios en la nube. El marco está constituido por un conjunto de módulos integrados que se describen a continuación. Hay tres partes involucradas: Cliente, Servidor y Servicios en la Nube. En el cliente están los módulos de la interfaz de usuario, y el módulo cliente para la colaboración en tiempo real que permiten crear interfaces reactivas. La reactividad responde en tiempo real web y permite crear sistemas colaborativos. De esta manera es posible crear sistemas donde los cambios que realiza un cliente en cualquiera de las interfaces de usuario son inmediatamente propagados a todos los clientes del sistema. Un usuario remotamente puede colaborar con otro en una determinada interfaz, por ejemplo, supongamos una interfaz que contiene a un gráfico interactivo para el análisis de los datos de la salida de un problema de optimización, el usuario 1, puede mostrar su opinión acerca de los datos de la salida a los demás usuarios modificando ciertos parámetros. También es posible la sensorización, es decir monitorear sensores en tiempo real que permiten la implementación de sistemas cibernéticos y por lo tanto construir sistemas de scheduling para la industria 4.0 [18].. En el servidor el módulo Gestión de la Colaboración en Tiempo Real que es el encargado de gestionar este tipo de colaboración.

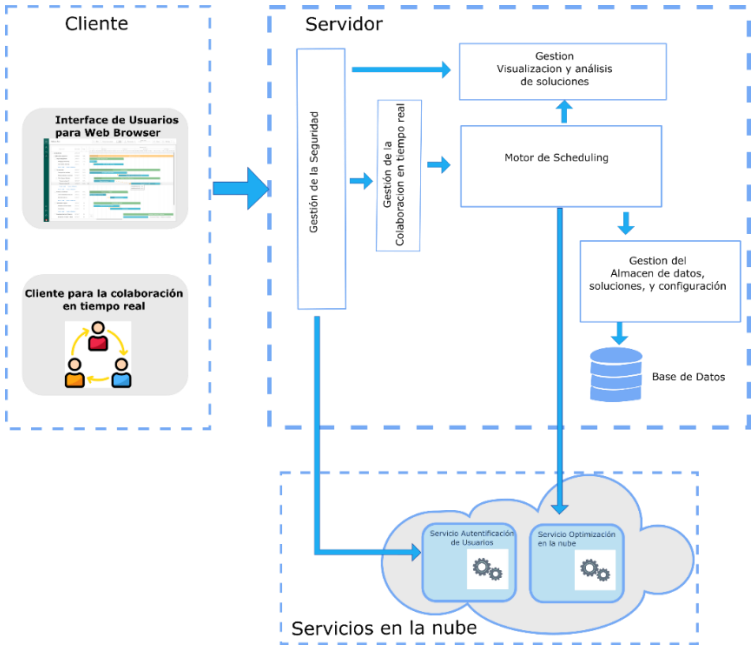


Fig. 2. Arquitectura del marco para desarrollar sistemas de programación de la producción

El usuario inicia una sesión en el sistema a través de la interfaz de autentificación de usuarios, esta solicita al servidor que valide al usuario, el servidor mediante el módulo de Gestión de la Seguridad autentifica al usuario y autoriza a los demás módulos a interactuar con el cliente. El módulo Motor de Scheduling es optimizador basado en modelos, su tarea es gestionar el modelo y los datos que representan un determinado problema de scheduling. Cuando un usuario interactúa con el problema a través de la interfaz de usuario se generan nuevos datos que se asocian a un archivo de datos, pueden existir opciones en la interfaz que permitan afinar el motor de optimización, estas opciones se asocian a un archivo de configuración. Una vez que el usuario ha terminado de configurar el problema de scheduling lo remite al cliente. En el servidor, el Motor de Scheduling, recibe el problema, y realiza un análisis del mismo llamando al módulo de Gestión del Almacén de datos, soluciones y configuración para determinar si el problema ha sido resuelto anteriormente, si es así, el modulo encargado de la gestión del almacén remite la solución correspondiente. Si el gestor del Almacén responde que no existe solución almacenada para el problema el Motor de Scheduling llama al servicio de optimización en la nube para que resuelva el problema. El servicio de optimizacion web remite la solución al Motor de Scheduling. Por ultimo una vez que el Motor de Scheduling ha obtenido una solución para el problema lo envía al módulo de visualización y análisis, el cual formatea la salida y lo remite a la interfaz de usuario para que presente la solución.

6 Prototipando un sistema de scheduling mediante el marco

Con el objeto de ilustrar el uso del marco propuesto se describe a modo general, y sin entrar en detalle, la construcción de un muy simple sistema de scheduling utilizando el marco. La problemática se desarrolla en un contexto industrial donde se desarrolla la problemática es una fábrica de sillas. La entrada al sistema de scheduling está compuesta por un conjunto de actividades, sus duraciones y sus precedencias. Por otra lado para ejecutar esta actividades son necesarios recursos, estos pueden ser los operarios. Existe un conjunto de restricciones tales como que una actividad requiere de determinado recurso para ejecutarse. Por ultimo en los objetivos del scheduling pueden ser de diversos, desde optimizar el tiempo de entrega hasta optimizar el uso de los recursos. Para desarrollar el sistema siguiendo el marco propuesto se seleccionan las siguientes tecnologías: Meteor.js [13] para implementar tanto en el cliente como en elservidor. El servicio de optimización en la nube es IBM Decision Optimization on Cloud [16]. El servicio de Autentificación se implementa mediante Meteor Accounts system, el que provee una serie de paquetes que permiten autentificación mediante password, integración con terceros para iniciar sesión (Google, Facebook, entre otros) y una interface de usuario preconstruida y customizable. MongoDB como motor de base de datos. La gestión de la colaboración en tiempo real se realiza con los mecanismos de publicar / suscribir y variables reactivas que provee Meteor. Para la visualización se selecciona el paquete anyChart de anychart.com. La plataforma cloud IaaS para despliegue de todo el sistema será digitalOcean. El despliegue se realizara utilizando una máquina virtual con contenedores Docker. Se utiliza un contenedor para la gestión de la base de datos MongoDB, un contenedor para la aplicación, un contenedor nginx que se utiliza como proxy y servidor web.

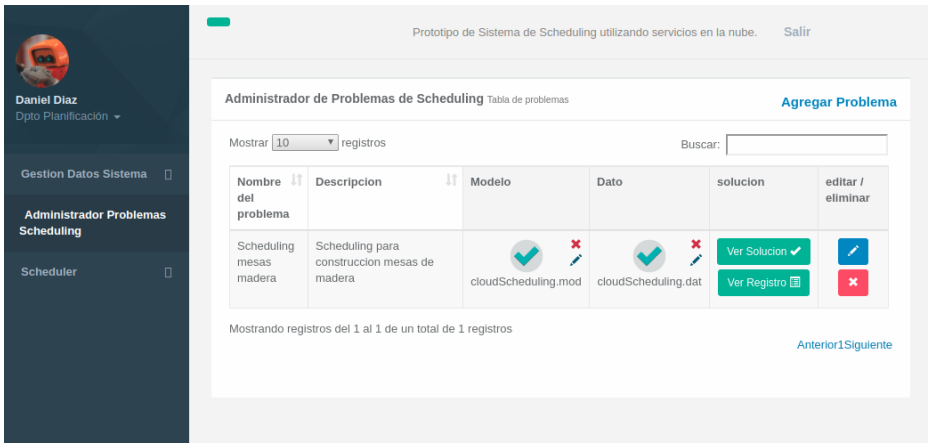


Fig. 3. Interfaz del gestor del Almacén de Datos

La figura 3 muestra parte de la interfaz del prototipo, más específicamente se puede observar la interfaz del "Administrador de Problemas de Scheduling", mediante el cual

se pueden cargar diferentes problemas de scheduling y solicitar su solución. A la derecha arriba, se encuentra el link *Agregar Problema*, el cual permite ingresar un nuevo problema. En la tabla de la interfaz se visualizan todos metadatos de un problema, las columnas de la izquierda Nombre, Descripción, Modelo y Datos, conforman la entrada, mientras que la columna Solucion es la salida, la última columna permite editar o eliminar un problema. De esta forma en la pantalla se muestra la existencia de un problema del sistema de scheduling de mesas de madera, que tiene solución asignada, es decir una acción previa solicito una solución para el mismo.

7 Conclusiones y trabajos futuros

En este trabajo se ha presentado una aproximación basada en un marco de trabajo para desarrollar sistemas de scheduling utilizando servicios en la nube. En el apartado 6 se presentó la arquitectura del marco de trabajo, el cual hace uso de servicios de optimización en la nube. Con la arquitectura presentada fue posible construir un prototipo básico de una aplicación que tiene por objeto simular un sistema de scheduling. El prototipo utiliza el motor de soluciones de *IBM Decision Optimization*, un servicio en la nube del tipo SaaS. Si bien en el prototipo no se desarrolló ningún tipo de capacidad de colaboración, la arquitectura propuesta propone los mecanismos necesarios para implementar varias formas de colaboración tal como interfaces reactivas con interacción en tiempo real.

Como trabajos futuros se propone mejorar el prototipo incorporando nuevas funcionalidades, trabajando preferentemente en las necesidades que impone la industria 4.0, es decir hacer uso de las capacidades de colaboración en tiempo real que propone la arquitectura del marco de trabajo presentado.

Referencias

[1] K. R. Barker, *Elements of sequencing and scheduling*. New York: John Wiley and Sons, 1974.

[2] G. Guizzi, S. Vespoli, and S. Santini, "On The Architecture Scheduling Problem Of Industry 4.0," in *CIISE*, 2017, pp. 94-100.

[3] D. Rossit and F. Tohmé, "Scheduling research contributions to Smart manufacturing," *Manufacturing Letters*, vol. 15, pp. 111-114, 2018.

[4] B. L. MacCarthy and J. Liu, "Addressing the gap in scheduling research: A review of optimization and heuristic methods in production scheduling," *International Journal of Production Research*, vol. 31, pp. 59-79, 1993.

[5] G. P. Henning, "Towards a software engineering approach to the deployment and implementation of scheduling systems," in *Computer Aided Chemical Engineering*. vol. 40, ed: Elsevier, 2017, pp. 1405-1410.

[6] B. P.-C. Yen and M. Pinedo, "On the design and development of scheduling systems," in *Fourth International Conference on Computer Integrated Manufacturing and Automation Technology*, 1994, pp. 197 - 204.

[7] K. McKay, M. L. Pinedo, and S. Webster, "Practice-Focused Research issues for scheduling systems," *Production and Operations Management,*, vol. 11, pp. 249-258, 2002.

[8] J. M. Framinan, R. Leisten, and R. R. García, "Manufacturing scheduling systems," *An integrated view on Models, Methods and Tools*, pp. 51-63, 2014.

[9] N. Antonopoulos and L. Gillam, *Cloud computing*: Springer, 2010.

[10] G. Taylor, "Definición de Cloud Computing por el NIST," *https://blogs.technet.microsoft.com/guillermotaylor/2010/08/25/definicin-de-cloud-computing-por-el-nist/*, 2010.

[11] F. C. Castiglione, "Virtualización en Entornos Cloud," *Tesis de grado* vol. FCEFyN - UNSJ, 2017.

[12] M. Fowler and J. Lewis, "Microservices a definition of this new architectural term," *URL: http://martinfowler. com/articles/microservices. html*, p. 22, 2014.

[13] Meteor, "Meteor documentation," *www.meteor.com*, 2019.

[14] A. Martín, S. Chávez, M. A. Murazzo, N. R. Rodríguez, and A. Valenzuela, "MongoDB en ambiente Cloud Híbrido con OpenStack," in *XVII Workshop de Investigadores en Ciencias de la Computación (Salta, 2015)*, 2015.

[15] IBM-Cplex, "IBM ILOG CPLEX Optimization Studio," *https://www.ibm.com/support/knowledgecenter/en/SSSA5P_12.7.1/ilog.odms.studio.help/Optimization_Studio/topics/COS_intro_features.html*, 2017.

[16] DOcplexcloud, "Decision Optimization on Cloud," *https://developer.ibm.com/docloud/documentation/decision-optimization-on-cloud/*, 2019.

[17] OPL, "Optimization program language," *https://www.ibm.com/analytics/data-science/prescriptive-analytics/optimization-modeling*, 2017.

[18] D. A. Rossit, F. Tohmé, and M. Frutos, "Industry 4.0: Smart Scheduling," *International Journal of Production Research*, pp. 1-12, 2018.

Criterios de calidad del software. Un enfoque de Accesibilidad Web vinculado a la gestión de recursos físicos

José M. Bordón[1], Pedro L. Alfonzo[1], Sonia I. Mariño[1]

[1]Facultad de Ciencias Exactas y Naturales y Agrimensura, Universidad Nacional del Nordeste, Argentina
josebordon1994@gmail.com, plalfonzo@hotmail.com, simarinio@yahoo.com

Resumen. La Accesibilidad Web es uno de los criterios de calidad del software, que aseguran acceso a la información sin restricciones espacio-temporales. Un sitio web accesible significa que el diseño de la interfaz del usuario debe facilitar el acceso universal a su información en condiciones de igualdad. Este trabajo presenta un aporte al ámbito académico, a través del desarrollo de una herramienta software parametrizable orientada a la gestión de recursos físicos, como ser las aulas y laboratorios de informática, considerando en etapas temprana de su construcción criterios relacionados con la Accesibilidad Web, en particular con accesibilidad visual y de esta manera contribuir a la inclusión social de los e-ciudadanos. Además, se aplicó un proceso iterativo que al finalizar cada incremento verifica el cumplimiento de los criterios de calidad considerados.

Palabras clave: Accesibilidad web, gestión de aulas, W3C, ingeniería del software, calidad del software.

1 Introducción

En la sociedad del conocimiento un tema de relevancia abordado por numerosos organismos internacionales es la Accesibilidad Web (AW). Tal es su importancia, que se encuentra en la agenda de la Fundación Sidar [1], el Consorcio World Wide Web (W3C) [2], la Organización Internacional para la Estandarización (ISO) [3], entre otros.

El principal objetivo de la AW es aportar en la definición de estándares, métodos y herramientas para lograr que las tecnologías y las Tecnologías de la Información y Comunicación (TIC) mejoren la calidad de vida de los sujetos inmersos en una sociedad digital.

La Red UNCI [4] establece que la Informática se compone de nueve disciplinas, y una de ellas la Ingeniería del Software (IS). La Ingeniería del Software (IS), como disciplina de Ingeniería se ocupa de todos los aspectos de la producción de software - desde las primeras etapas de especificación del sistema hasta su mantenimiento después de que entra en uso. Menciona como elementos clave: los métodos, las

herramientas y los procedimientos que facilitan el control del proceso de desarrollo de software y brinda a los desarrolladores las bases de calidad de una forma productiva [5]. Una de sus principales áreas de estudio de investigación es la calidad del software.

En [5] se establece que la calidad del software versa sobre el cumplimiento de los requisitos de funcionalidad y desempeño explícitamente establecidos, de los estándares de desarrollo explícitamente documentados, y de las características implícitas que se espera de todo software desarrollado profesionalmente.

Esta definición subraya la importancia de los estándares, que definen un conjunto de criterios de desarrollo que guían la forma en que se aplica la IS, para lograr productos y servicios de calidad. La AW es uno de los criterios de calidad del software, que aseguran acceso a la información sin restricciones espacio-temporales.

Un sitio web accesible significa que el diseño de la interfaz de usuario debe facilitar el acceso universal a su información en condiciones de igualdad. La mejora de la AW impacta sobre numerosos tipos de discapacidades, tanto físicas (visuales, auditivos, cognitivos, neurológicos, entre otros) como aquellas no oficiales derivadas de las tecnologías (falta de acceso o equipamiento adecuado). A través de herramientas se evalúa y repara la accesibilidad y por consiguiente el diseño de dichos entornos.

Por otra parte, en los últimos años se ha observado un incremento en el uso de las TIC y de la multiplicidad de soluciones orientadas para facilitar el manejo, el acceso y la difusión de la información. Cada vez es mayor el número de personas e instituciones que integran la tecnología al trabajo y al quehacer cotidiano [6]. El uso de TIC y su implementación en las instituciones contribuye a incrementar su inteligencia organizacional.

Las asignaciones de recursos en cualquier organización dejan de ser un problema trivial cuando los recursos son limitados y las necesidades cambian. Una problemática de las instituciones educativas es la asignación de espacios físicos. Las asignaciones, se logran mediante la toma de decisiones, que surgen de comparar las disponibilidades edilicias, las capacidades áulicas, los horarios solicitados, entre otros, siendo estas decisiones las más tolerables para que la institución pueda llegar a su correcto funcionamiento [7].

Se coincide con [8], en que el problema de la concesión óptima de aulas ha sido tratado con diversas técnicas, herramientas y tecnologías. Además, siendo escasos los productos software, que determinen la mejor distribución de espacios en función de las características edilicias y de las franjas horarias solicitadas por los docentes.

En el desarrollo Web, los requerimientos puede complementarse con la integración de estándares de accesibilidad a los efectos de generar productos de calidad, por lo tanto la inclusión de éstos estándares en etapas tempranas del proceso de elaboración de sitios web, constituye una manera de construir software de calidad.

Este trabajo aportará al ámbito universitario, mediante el diseño y desarrollo de una herramienta software parametrizable orientada a la gestión de recursos físicos, como ser las aulas y laboratorios de informática, considerando en etapas temprana de su construcción criterios relacionados con la AW, en particular con accesibilidad visual y de esta manera contribuir a la inclusión social de los e-ciudadanos.

2 Metodología

La metodología utilizada constó de las siguientes actividades:

- Revisión bibliográfica. Se recopiló, analizó y seleccionó información referente a:
    - Los métodos y herramientas de la Ingeniería del Software [5], [9].
    - Los Frameworks orientados al desarrollo del software [10], [11].
    - Selección y estudio de herramientas disponibles en la web, que permiten comprobar si la combinación de color es adecuada para su percepción visual en base a los requerimientos de las Pautas de Accesibilidad a los Contenidos Web (WCAG 2.0) [12], se optó por WCAG Contrast Ckecker [13].

- Adaptación de un proceso iterativo, cuyo esquema se visualiza en la Fig. 1. Este proceso se propuso considerando que permite analizar la AW, en particular la percepción visual en base a los requerimientos de las WCAG 2.0 [12]: cumplimiento del criterio 1.4.1 Uso del color (Nivel A) y 1.4.3, que especifica un contraste mínimo en el texto (Nivel AA).

- Diseño y construcción del software que automatice la gestión de recursos físicos en un contexto universitario aplicando criterios relacionados con la accesibilidad visual para lo cual se utilizó un proceso iterativo e incremental compuesto por las siguiente etapas:
  Etapa 1: Análisis y diseño de la herramienta software.
    - Análisis de factibilidad, de requerimientos funcionales y no funcionales del sistema, definición de limitaciones tecnológicas.
    - Definición de seguridad en el acceso a la información. Se definieron a partir de los distintos perfiles de usuarios que accedan al sistema.
    - Especificación de requerimientos. Se especificaron las funcionalidades solicitadas.
    - Diseño de la base de datos. Se modeló y diseñó una base de datos relacional.
    - Selección de herramientas. Se utilizaron herramientas de libre distribución [10], [11].
    - Documentación. Se elaboraron documentaciones de soporte al análisis y diseño de las soluciones propuestas.

  Etapa 2: Desarrollo de la herramienta software.
    - Construcción de las interfaces. Se desarrolló de acuerdo a los diversos perfiles de usuarios a los cuales está dirigido el producto de software.

- Desarrollo. Se procedió a la codificación iterativa e incremental de las versiones del software.
- Pruebas. Realizada la codificación del sistema en cada incremento, se procedió a realizar pruebas para verificar el correcto funcionamiento de cada funcionalidad desarrollada con miras a realimentación.
- Pruebas de integración. Se realizaron pruebas sobre la herramienta software completamente integrado.

- Aplicación del proceso propuesto en la Fig. 1 al finalizar cada incremento y la posterior modificación del código necesario para mejorar los criterios relacionados con la accesibilidad visual.

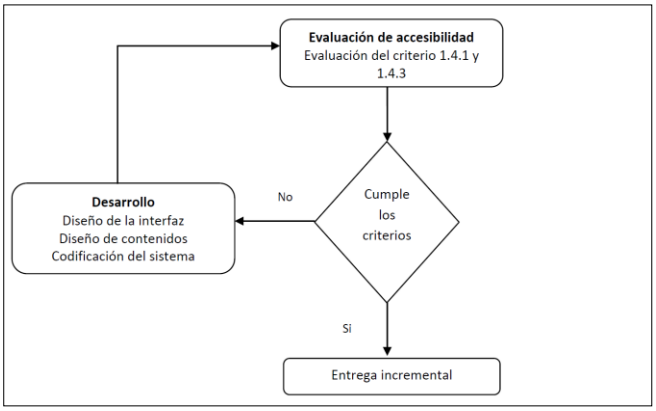


Fig. 1. Proceso de iterativo de evaluación de la accesibilidad

3 Resultados

En este trabajo se presentan las páginas desarrolladas de acuerdo al usuario con perfil "público" y perfil "Bedel" - usuario registrado encargado de realizar las reservas - consideradas en proporcionar una solución acorde con las necesidades y requerimientos edilicios demandados por las distintas carreras dictadas en una determinada unidad académica.

El sitio web desarrollado debe cumplir una serie de características para satisfacer los requerimientos de accesibilidad visual, en particular la percepción visual en cuanto al color, es decir, dar cumplimiento del criterio 1.4.1 Uso del color (Nivel A) y 1.4.3, que especifica un contraste mínimo en el texto (Nivel AA), de acuerdo a las WCAG 2.0 [12]. En la Tabla 1 se mencionan algunos beneficios de implementar los mencionados criterios [14], [15].

Tabla 1. Beneficios de cumplimentar los criterios 1.4.1 y 1.4.3

Criterio 1.4.1 - Uso del color [14]	**Criterio 1.4.3 - contraste mínimo [15]**
- Usuarios con visión parcial a menudo experimentan una visión limitada del color. - Algunos usuarios mayores no puedan ver bien el color. - Los usuarios con daltonismo se benefician cuando la información transmitida por color está disponible de otras formas visuales. - Es posible que las personas que utilizan pantallas de solo texto, color limitado o monocromo no puedan acceder a información dependiente del color. - Los usuarios que tienen problemas para distinguir entre colores pueden buscar o escuchar señales de texto.	- Las personas con baja visión pueden tener dificultades para leer textos que no contrastan con su fondo. Proporcionar una relación de contraste de luminancia mínima entre el texto y su fondo puede hacer que el texto sea más legible incluso si la persona no ve la gama completa de colores.

Por lo expuesto, se utilizó el proceso propuesto en la Fig. 1 a las páginas mencionadas anteriormente al finalizar cada incremento; y se procedió a la posterior modificación del código necesario para cumplimentar con los requerimientos de accesibilidad visual establecidos. Este proceso se iteró hasta lograr una página totalmente accesible. La Tabla 2 presenta los incrementos realizados.

Tabla 2. Incrementos realizados y evaluación de criterios relacionados con la accesibilidad visual

Incrementos	**Desarrollo**	**Criterio 1.4.1 - Uso del color (Revisión manual)**	**Criterio 1.4.3 - contraste mínimo (Revisión automática)**
Incremento 1	Página de disponibilidad de aulas	Cumple el criterio.	Ver Fig. 2
Incremento 2	Página de filtro y búsqueda de aulas	Cumple el criterio	Ver Fig. 3
Incremento 3	Página de un aula específica	Cumple el criterio	Ver Fig. 4
Incremento 4	Página de reserva de aulas	Cumple el criterio	Ver Fig. 5

En las Figs. 2, 3, 4 y 5 se observan el resultado de la verificación utilizando las herramientas WCAG Contrast Checker [13] antes y después de realizar las correcciones necesarias en cada incremento.

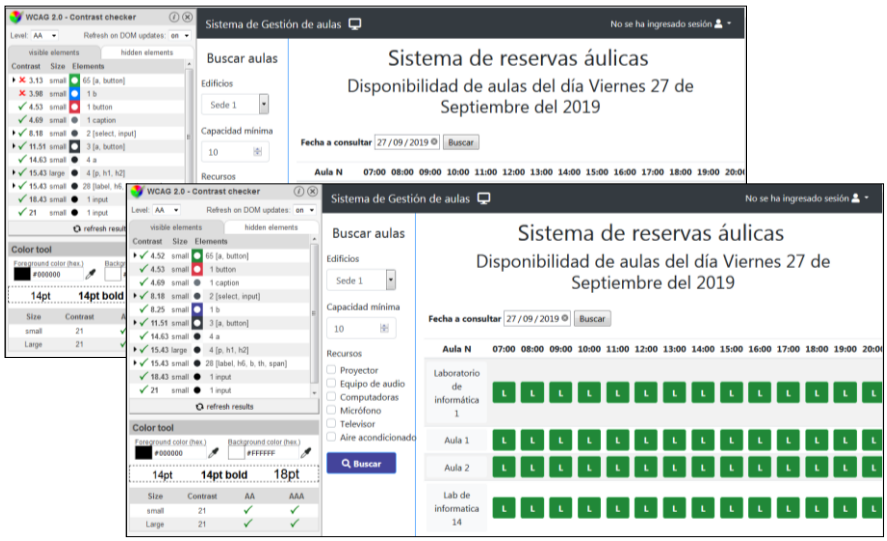


Fig. 2. Página de disponibilidad de aulas – Perfil público

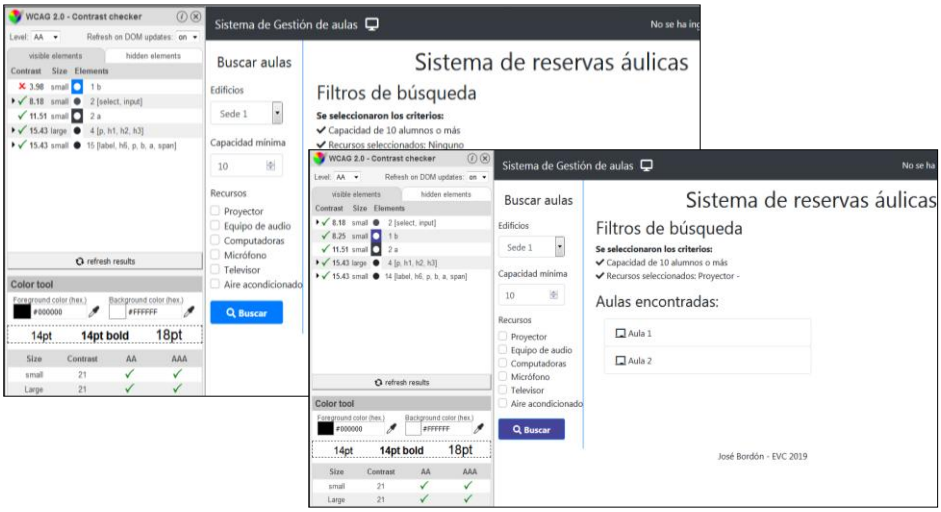


Fig. 3. Página de filtro y búsqueda de aulas - Perfil público

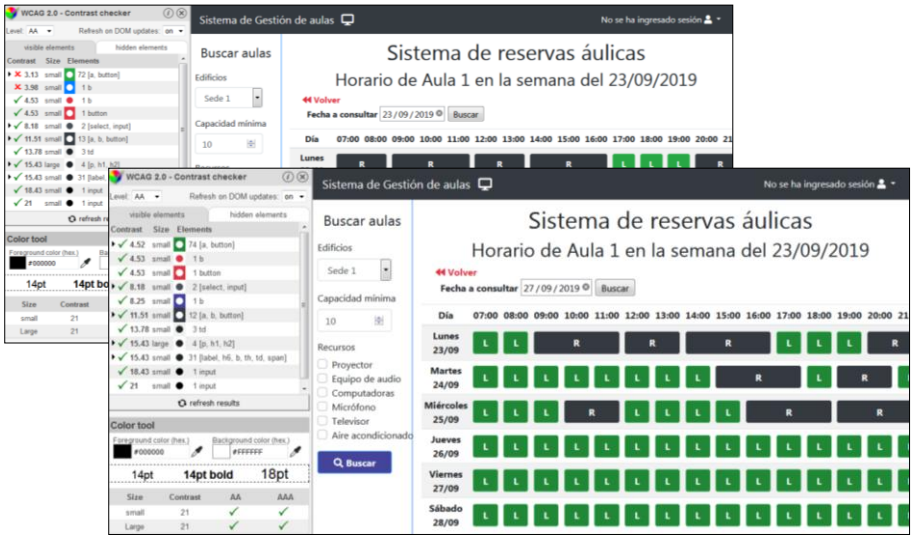


Fig. 4. Página de un aula específica – Perfil público

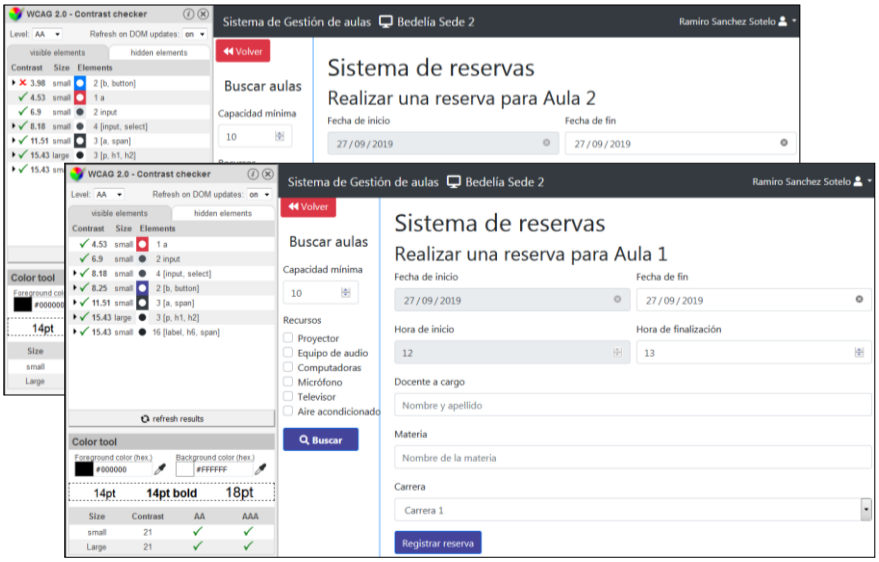


Fig. 5. Página de reserva de aulas – Perfil Bedel

4 Conclusiones

Se presentó el diseño y desarrollo de una herramienta software parametrizable orientada a la gestión de recursos físicos y su posible implementación en una unidad académica. En la misma, se consideraron criterios relacionados con la accesibilidad

visual en etapas temprana de su desarrollo para de esta manera lograr un producto de software calidad y contribuir a la inclusión social de los e-ciudadanos. Las implementaciones realizadas permiten expresar que se dispone de un producto software útil, fiable y accesible de acuerdo a los criterios considerados.

Por lo expuesto es menester que los equipos desarrolladores validen y apliquen los cambios necesarios a fin de asegurar la igual de acceso a la información

Por otra parte, se indica la relevancia de continuar tratando procedimientos metodológicos y tecnológicos para potenciar esta línea de investigación: la Accesibilidad Web, focalizada en la accesibilidad visual. Avanzar en el desarrollo, implementación y difusión de estos temas, permite disponer de propuestas teóricas, reflejadas en la utilización práctica y efectiva de los conceptos reflejados en artefactos software.

Agradecimientos. El trabajo se enmarca en una Beca de Estímulo a las Vocaciones Científicas, otorgada por el Consejo Interuniversitario Nacional convocatoria 2018 (Resolución P. N° 403/18) y en un Proyecto de I+D "TI en los Sistemas de Información: modelos, métodos y herramientas", Acreditado por la Secretaría General de Ciencia y Técnica de la Universidad Nacional del Nordeste.

Referencias

1. Fundación Sidar, http://www.sidar.org/
2. W3C. About W3C, https://www.w3.org/Consortium/
3. ISO. Organización Internacional para la Estandarización, http://www.iso.org/iso/home.html
4. Red UNCI. Red de Universidades Nacionales con Carreras en Informática, http://redunci.info.unlp.edu.ar/
5. Pressman, R.: Ingeniería de Software: Un Enfoque Práctico. Pearson Education, 2010.
6. Hidalgo M., Fernández M., Godoy M. V. y Mariño S. I.: Prototipo de Sistema Informático destinado a la Gestión de un Seguro Público de Salud, XXIV Congreso Argentino de Informática y Salud, CAIS 2013.
7. Badaracco, N., Alfonzo, P. L., Mariño, S. I.: Modelización de la asignación de aulas con técnicas simbólicas de la IA como ayuda a la toma de decisiones. Revista Electrónica Científica de Estudios Telemáticos, 13(2), pp. 16-35, 2014.
8. Cancelo, F., Cababie, N, Barrera, G. y López De Luise, D.: Un nuevo enfoque para asignación óptima de múltiples recursos. Universidad de Palermo, ITLab, AIGROUP, 2009.
9. Sommerville, I.: Ingeniería del Software. Ed. Pearson Educación, México, 2011.
10. CodeIgniter. Framework de PHP, https://codeigniter.com/
11. Bootstrap. Framework de desarrollo web, https://getbootstrap.com/
12. Web Content Accessibility Guidelines (WCAG) 2.0, https://www.w3.org/TR/
13. Firefox add-ons-Mozilla, WCAG Contrast checker, https://addons.mozilla.org/es/firefox/addon/wcag-contrast-checker/.
14. Comprender el Criterio de Conformidad 1.4.1: Uso del color, https://www.w3.org/WAI/WCAG21/Understanding/use-of-color
15. Comprender el Criterio de Conformidad 1.4.3: Contraste (Mínimo), https://www.w3.org/WAI/WCAG21/Understanding/contrast-minimum

Técnicas de Análisis y Visualización de Minería de Datos para la Reducción de Dimensiones en Tablas de Datos

Laura Viviana Gutiérrez , Manuel Oscar Ortega, Maria Isabel Masanet,Facundo De La Jara

Departamento de Informática
Facultad de Ciencias Exactas, Físicas y Naturales
Universidad Nacional de San Juan
{gutierrez.laura, manuel.ortega, mimasanet, facundodelajara}@gmail.com

RESUMEN

Según un informe de Data NeverSleeps de Domo prevé que para el año 2020, cada persona del planeta creará 1,7MB de datos cada segundo. En este contexto uno de los problemas a los que se enfrenta el Científico de Datos lo constituye el volumen de datos a procesar. A veces esta cantidad de datos provoca que a simple vista no se puedan observar relaciones entre ellos. Una de las etapas previas al análisis de datos al que se enfrenta el Científico de Datos, es la limpieza de los datos. Cuando se habla de limpieza de los datos, debe entenderse que se refiere, entre otros, a reducir las dimensiones de los datos a efectos de poder tratarlos. Para llevar a cabo esta tarea se propone en este trabajo, utilizar técnicas de análisis de datos y visualización para reducir dimensiones en los datos a tratar.

Frases claves: Ciencia de Datos, Visualización, Correlación, Reducción de dimensiones.

1. Introducción

Gracias a los nuevos desarrollos tecnológicos, como Internet de las cosas, se estima que la cantidad de información digital generada en 2035 ascienda a los 2.142 zettabytes [1].
Estos grandes volúmenes de datos, dan origen a lo que se denomina Big Data y a los problemas asociados al tratamiento de grandes volúmenes de datos.
Big Data hace referencia a una colección de datos de considerable dimensión que hacen imposible el procesamiento con aplicaciones tradicionales de base de datos.
En [2], Furth, B. y Villanustre, F. definen Big Data usando las tres V's: Volumen, Velocidad y Variedad. El volumen se refiere a la cantidad de datos, desde Terabytes

(TB) a Petabytes (PB), relacionado con la estructura de estos datos incluyendo registros, transacciones, archivos, y tablas. El volumen de estos datos, se prevé que crecerá 50 veces para el 2020. La velocidad se refiere a la forma de transferir los grandes volúmenes de datos incluyendo transmisión en batch, tiempo real y flujos. La velocidad, incluye tiempo y latencia, características propias del manejo de datos. Los datos pueden ser analizados, procesados, almacenados y manipulados en forma rápida, o con un retardo entre eventos. La variedad de los grandes volúmenes de datos, se refiere a los diferentes formatos que pueden adoptar los datos, incluyendo estructurados, semi estructurados, no estructurados y todas las combinaciones de estos tres. El formato de los datos incluye: documentos, mails, mensajes de texto, audio, imágenes, video, gráficos, entre otros.

La Minería de Datos (MD) involucra e integra técnicas de diferentes disciplinas tales como tecnologías de bases de datos y data warehouse, estadística, aprendizaje de máquina, computación de alta performance, computación evolutiva, reconocimiento de patrones, redes neuronales, visualización de datos, recuperación de información, procesamiento de imágenes y señales, y análisis de datos espaciales o temporales.

La comprensión e interpretación apropiada de los resultados de un modelo es el mayor problema o desafío en un proceso de MD. Algunas investigaciones sostienen que una de las formas de apoyar esta problemática es utilizando visualización en la etapa de construcción del modelo, y no solamente en el análisis exploratorio de los datos (entrada), y los patrones o reglas obtenidas en un proceso de MD (salida) [3][4].

En [5], Blitzstein J. propone un modelo de Procesos de Ciencias de Datos, basado en 5 etapas.

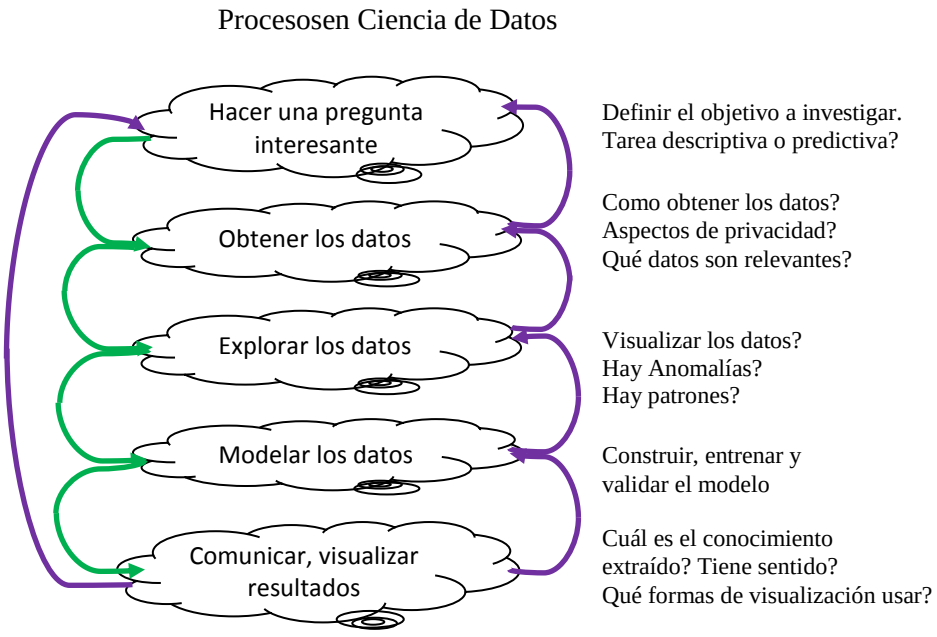


Figura 1 – Procesos en Ciencia de Datos (Blitzstein)

Como puede observarse en la Figura 1, el análisis de los datos se inicia con una pregunta, que plantea o servirá para plantear el objetivo del análisis de datos, especifica la meta que se persigue, lo que se pretende predecir o estimar. Luego sigue en el modelo, la obtención de los datos.La obtención de los datos, es una tarea que a veces se complica, el propietario de los datos puede tener impedimentos legales para entregarlos, o simplemente negarse a entregar los datos para su procesamiento, por razones de negocio, patentes, etc. Ante la imposibilidad de acceder datos privados, el Científico de Datos puede trabajar con datos abiertos. Una vez obtenidos los datos, comienza la etapa de exploración de los mismos.
A este modelo se lo ha ampliado agregando una etapa previa, denominada limpieza de los datos, presentada por O'Neil, C. y Schutt, R. en [6]. Los autores reconocen otras etapas antes de la exploración de los datos, como puede observarse en la Figura 2.

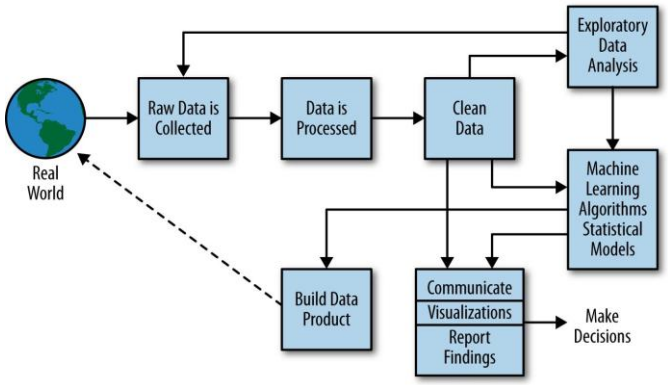


Figura 2 – Procesos en Ciencia de Datos (O'Neil y Schutt)

En esta etapa el Científico de Datos deberá encontrar qué datos son relevantes y cuales no lo son, con la finalidad de reducir la cantidad de dimensiones para poder tratar los mismos.
Los mismos autores, proponen el rol del Científico de Datos, como puede apreciarse en la Figura 3.

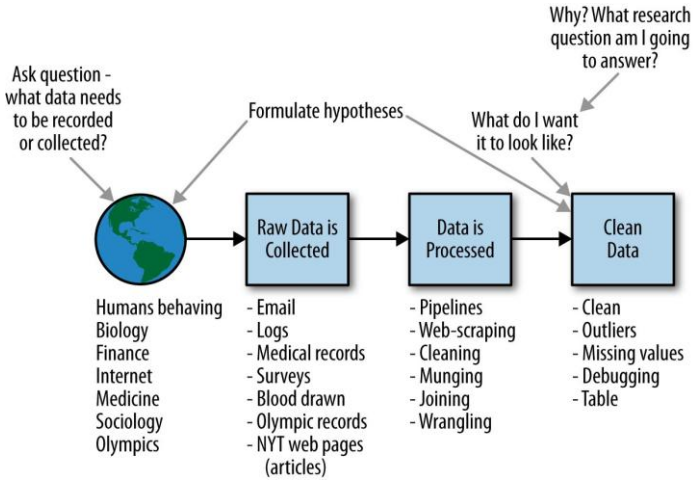


Figura 3 – Rol del Científico de Datos (O'Neil y Schutt)

El Cientfíco de Datos, debe responder a la pregunta ¿qué datos necesita?, estos pueden provenir de diversas fuentes, por ejemplo, comportamiento humano, biología, economía, finanzas, medicina, juegos olímplicos, deportes, etc. Una vez idenficada la fuente de datos, la tarea siguiente, es obtener los mismos. Posteriormente, el Científico de Datos, comienza por conocer el tipo de datos, los clasifica, los agrupa, encuentra valores máximos, mínimos y medios, cuartiles. Puede estudiar la distribución de los datos visualizando en gráficos las distribuciones a las que tienden. Esta etapa, denominada preprocesamiento de los datos involucra tareas como: *limpiar los datos*, eliminando datos anómalos, duplicados, reduciendo cantidad de dimensiones, eliminando columnas irrelevantes a la problemática, etc., *transformar datos*, cambiando el tipo de datos para poder procesarlos, *calcular nuevos datos* según la problemática y el modelo de aprendizaje a aplicar.

Para visualizar, la visualización matricial tiene una propiedad interesante: es capaz de mostrar patrones de datos a nivel local y global en un solo gráfico. Esto se ve facilitado por el algoritmo de diseño, que es simple y efectivo. Muestra las entidades de datos en filas y columnas, asignando a la celda correspondiente a la intersección los atributos de los datos [7].

Este trabajo presenta el uso de técnicas de análisis y visualización para reducir dimensiones en los datos aplicado a un caso de estudio. Se enmarca en el proyecto bianual 2018-2019 "Visualización y DeepLearning en Ciencia de Datos", código 21/E1071, que se desarrolla en el ámbito de Instituto y Departamento de Informática de la FCEFN-UNSJ, aprobado por el Consejo de Investigaciones Científicas Técnicas y de Creación Artística (CICITCA), financiado en su totalidad por la propia Universidad.

2. Definición del Problema

El trabajo se centra en hacer uso de análisis de datos y una estrategia de visualización, denominada matriz de correlaciones o mapa de calor, con el objetivo de tratar de reducir la cantidad de dimensiones de los datos a procesar en un problema. Esta tarea está enmarcada en la fase del proceso de Ciencia de Datos, denominada limpieza de los datos. La alta dimensión de los datos puede ser un obstáculo a la hora del procesamiento de los mismos.
Para el análisis de datos a través de la matriz de correlaciones, se ha seleccionado el conjunto de datos de un juego denominado FIFA 2018, que tiene publicado en formato abierto los datos de los jugadores de fútbol, con un conjunto de características de los mismos. La pregunta para el análisis, tal como la propone Blitzstein J. en [5], será: ¿qué variables son más relevantes o brindan mayor información sobre el valor de transferencia de un jugador?. En función de esta pregunta, se plantea como propósito encontrar las variables correlacionadas que determinan el valor de transferencia de jugador de fútbol del FIFA 2018[1], de modo que se reduzca la cantidad de dimensiones del conjunto de datos bajo estudio.

3. Desarrollo y Resultados

El conjunto de datos de FIFA 2018, tiene 17.981 filas, se corresponden con los jugadores de fútbol que tiene el juego en esta versión, y por cada jugador 75 columnas, que corresponden a los atributos de los jugadores.
En el proceso se utilizó el lenguaje de programación Python. Se eligió dicho lenguaje por poseer librerías para el procesamiento, análisis y visualización de datos. Las librerías que se utilizaron fueron: Numpy, Pandas, Matplotlib y Seaborn.
Numpy es una librería especializada para la computación científica en Python. Numpy incluye cosas como: arreglos n-dimensionales, operaciones sobre arreglos n-dimensionales altamente eficientes, operaciones del álgebra lineal, transformada de Fourier.
La librería Pandas, posee instrucciones específicas para la manipulación de datos en formato de tablas, así como para el análisis de datos.
La librearía Matplotlib se especializa en la creación de gráficos de gran calidad, en dos dimensiones (2D). Posee instrucciones para la creación y visualización de diversos tipos de gráficos, de barras, de puntos, histogramas, circulares, entre otros.
La librería Seaborn, está basada en Matplotlib y se especializa en la visualización de gráficos estadísticos de gran calidad.

[1] https://www.kaggle.com/thec03u5/fifa-18-demo-player-dataset/version/5#CompleteDataset.csv

Desarrollo

Instrucciones para la carga del conjunto de datos bajo estudio:

```
importnumpy as np
import pandas as pd
df = pd.read_csv("/home/Documentos/Data/Dataset.csv")
print(df)
```

El resultado de la ejecución de dicho código se puede observar en la Figura 4

Unnamed: 0	Name	Age	...	RW	RWB	ST
0	Cristiano Ronaldo	32	...	91.0	66.0	92.0
1	L. Messi	30	...	91.0	62.0	88.0
2	Neymar	25	...	89.0	64.0	84.0
3	L. Suárez	30	...	87.0	68.0	88.0
4	M. Neuer	31	...	NaN	NaN	NaN
5	R. Lewandowski	28	...	84.0	61.0	88.0
6	De Gea	26	...	NaN	NaN	NaN
7	E. Hazard	26	...	88.0	64.0	82.0
8	T. Kroos	27	...	80.0	78.0	77.0
9	G. Higuaín	29	...	82.0	55.0	87.0
10	Sergio Ramos	31	...	69.0	81.0	72.0
11	K. De Bruyne	26	...	85.0	71.0	81.0
12	T. Courtois	25	...	NaN	NaN	NaN
13	A. Sánchez	28	...	86.0	66.0	83.0
14	L. Modrić	31	...	83.0	80.0	76.0
15	G. Bale	27	...	87.0	74.0	87.0
16	S. Agüero	29	...	86.0	57.0	86.0
17	G. Chiellini	32	...	56.0	75.0	59.0
18	G. Buffon	39	...	NaN	NaN	NaN
19	P. Dybala	23	...	87.0	60.0	83.0
20	J. Oblak	24	...	NaN	NaN	NaN

Figura 4 – Carga y visualización de datos

El paso siguiente que se llevó a cabo en esta investigación, consistió en comprender los datos con los que se trabajó.

Luego del análisis de las 75 columnas, se descubrió que los tipos de datos de las columnas se agrupan en dos grandes clases: tipo texto y tipo numérico ordinal. Entre las columnas de tipo texto se encontraron, por ejemplo el nombre del jugador (Name), la nacionalidad (Nationality), club (Club), etc. Algunas de las columnas de tipo texto, poseen links externos, que redirigen hacia la foto del jugador (Photo), la bandera del país de nacimiento del jugador (Flag), el logo del club (Club logo), entre otros. Las columnas numéricas ordinales encontradas corresponden, a la edad del jugador (Age), calificación general (Overall), potencial (Potential), valor (Value), salario (Wage), aceleración (Acceleration), agilidad (Agility), etc.

Los distintos tipos de datos pueden analizarse usando la función *describe()* de Pandas. Si se ejecuta sobre toda la tabla, sólo muestra la estadística de los datos numéricos.

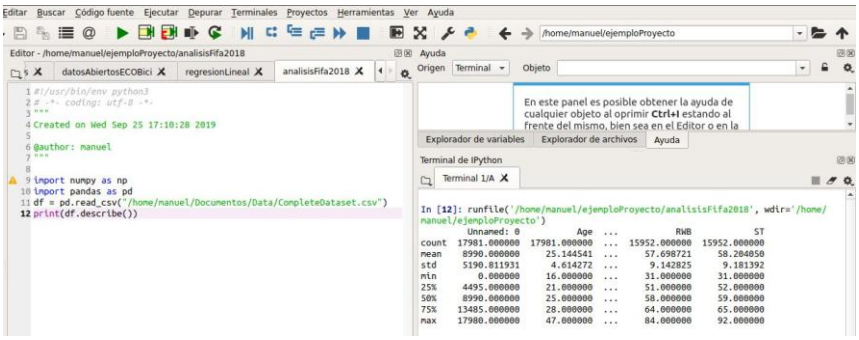


Figura 5 – Ejecución de la función *describe()*

La Figura 5, muestra que luego de cargado el conjunto de datos en la variable denominada *df*, será *print(df.describe())* el encargado devisualizar la cantidad de valores, la media aritmética, la desviación estándar, el valor mínimo el valor máximo, y los cuartiles.

Para el caso de la edad, se observa que la media aritmética es de 25 años de edad, el valor mínimo es de 16 años, el máximo es de 47 años. De este mismo modo se podrían analizar todas las columnas con valores de tipo numérico.

Una vez familiarizado con los datos, el Científico de Datos, analizará de manera más profunda el contenido de las columnas. Uno de los análisis que se hace es el tratamiento de los valores nulos o ausentes.

La Figura 4, muestra que hay columnas que poseen valores, NaN, Pandas identifica con este valor, las columnas en las que hay ausencia de valor, valores nulos, que en la literatura de Ciencia de Datos, son los denominados datos perdidos o ausentes. Luego de un análisis exhaustivo, se detectó que la mayor parte de valores ausentes, corresponden a los arqueros. Se tomó la decisión de trabajar con un subconjunto de jugadores, ya que el estudio en proceso, tiende a buscar la relación que existe entre el valor de venta de los jugadores, los jugadores de campo, en especial los delanteros y medio campistas, que en la vida real son los de mayor valor de mercado.

El tratamiento de los valores nulos tiene varias acciones posibles. Una es eliminar las filas que poseen valores nulos, para ello Pandas provee de la función *dropna.* Otra opción es reemplazar los valores nulos por el valor medio de la columna, sin indicar que se hizo dicho reemplazo, y por último es agregar una columna al set de datos, en dicha columna se identifica si la columna posee o no valores nulos, y éstos se reemplazan por los valores medios.

Se tomó la decisión de optar por la eliminación de las filas con valores nulos, ya que como se expresó anteriormente, estos corresponden a arqueros, que merecen un tratamiento distinto, para el análisis de ellos interesan características distintas a las de los jugadores de campo.

La sentencia que se ejecutó para la eliminación es: *df.dropna().*

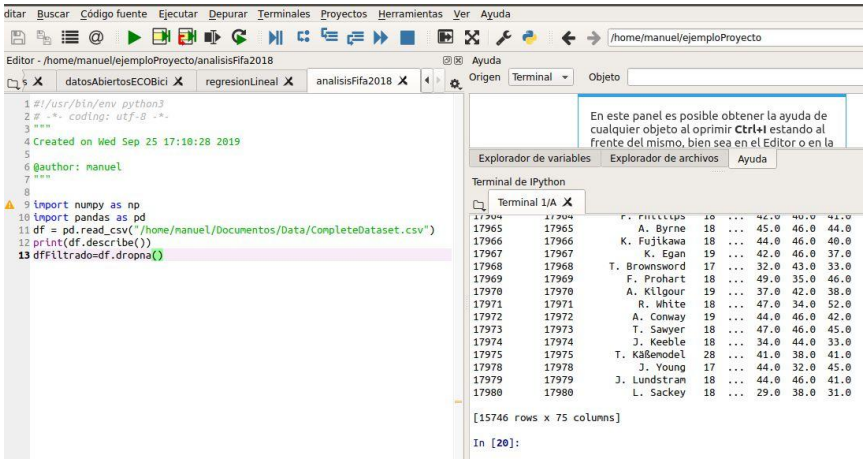


Figura 6 – Eliminación de filas con valores nulos

Luego de todo el análisis de datos, se procedió a analizar gráficamente la correlación de las columnas en el valor de venta del jugador. Éste análisis se hizo gráficamente, a través de la matriz de correlación o mapa de zonas calientes (heatmap). La Figura 7 muestra dicha matriz.

La matriz de correlación muestra en una gama de colores, las variables poco relacionadas valor -1.00 de la gama, a las muy relacionadas, valor 1.00 de la gama.

La matriz de correlación permite realizar observaciones como: el valor de venta del jugador es una variable que se correlaciona con su salario, a más valor de venta, más salario cobra. Las variables como compostura, control del balón, también tienen que ver con el valor de venta del jugador, en menor medida que el sueldo. Una sorpresa fue encontrar que las reacciones del jugador en la cancha, también tienen que ver con el valor de venta del mismo. Si bien no estaba bajo estudio, se puede ver que la calificación general (Overall) tiene una relación muy fuerte con el salario.

La figura 7 muestra las correlaciones de 34 atributos, que abarcan desde la edad del jugador hasta las boleas que puede realizar. Se utilizó esta matriz para encontrar las variables más correlacionadas con el objeto de reducir las dimensiones de las tablas.

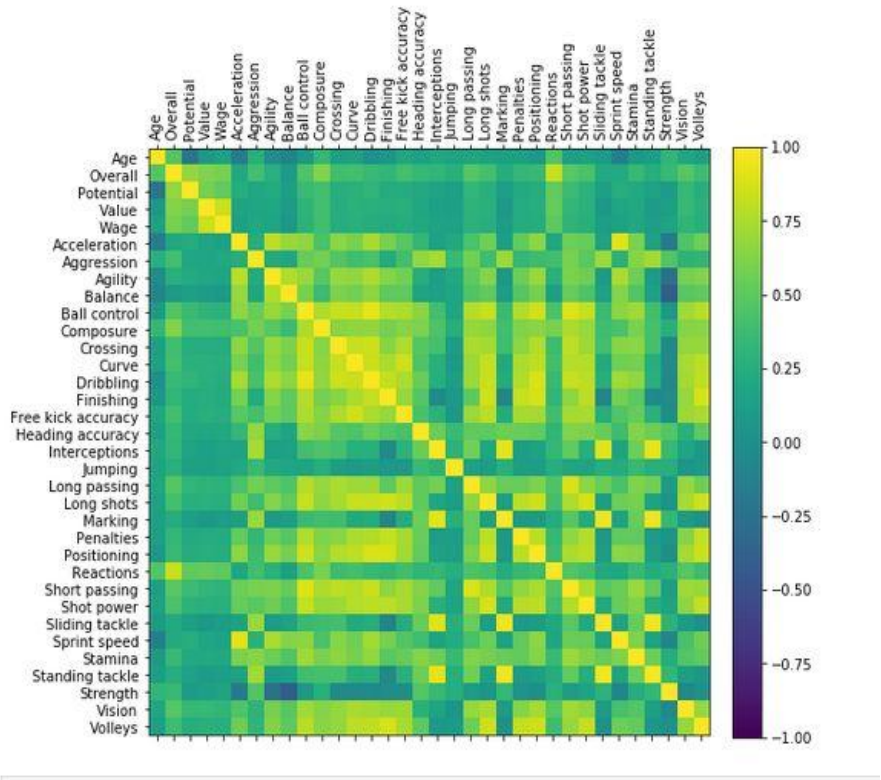


Figura 7 – Matriz de Correlación (HeatMap)

4. Conclusiones

La investigación se centró en un caso de estudio del conjunto de datos abiertos del juego FIFA 2018, buscando reducir las dimensiones de la tabla de datos, que originalmente tenía 17.981 filas con 75 columnas cada una. Luego del análisis exploratorio que se realizó sobre los datos, se logró reducir al 87,57% la cantidad de filas, quedando 15.746 filas de 75 columnas cada una. Con el análisis visual de la matriz de correlación, y atendiendo a la pregunta de investigación que se planteara, se llegó a reducir el conjunto de datos de a 15.746 filas de 9 columnas cada una, 6 de las cuales corresponden a los valores relacionados con el valor de venta del jugador, y las 3 restantes al nombre del jugador, club y nacionalidad. El volumen de datos, cantidad de celdas, a procesar se redujo de 1.348.575 a 141.714, lo que representa el 10,50% de los datos originales. El objetivo de reducir la cantidad de dimensiones se logró, ya que el conjunto de datos se redujo en un 89.50%, encontrando efectivamente las variables correlacionadas. Además, y si bien no estaba bajo estudio, apareció una relación llamativa entre Salario y Calificación General, a mayor calificación mayor salario. Para trabajos futuros, se pretende estudiar la relación de las habilidades de los jugadores, con la nacionalidad, el sueldo con el lugar donde juega, hay varios

supuestos al respecto. Los supuestos para dichas investigaciones tienen que ver con que los jugadores latinos, brasileros, argentinos y uruguayos son los que mejor ganan y son los más hábiles. El otro supuesto que se estudiará, tendrá que ver con que los jugadores que juegan en las ligas española, italiana, inglesa y alemana son los que poseen mejores salarios.

5. Referencias Bibliográficas

[1] G. Moreno, «Infografía: A la espera de un Big Bang de datos», *Statista Infografías*, 2019. [En línea]. Disponible en: https://es.statista.com/grafico/17734/cantidad-real-y-prevista-de-datos-generados-en-todo-el-mundo/. [Accedido: 18-sep-2019].

[2] B. Furht y F. Villanustre, *Big Data Technologies and Applications*. Suiza: Springer, 2016.

[3] I. H. Witten y E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques, 2nd. Edition.*, 2nd. Morgan Kaufmann series in data Management Systems, 2005.

[4] C. Meneses y G. Grinstein, «Visualization for Enhancing the Data Mining Process», en *In Proceedings of the Data Mining and Knowledge Discovery: Theory, Tools, and Technology*, Orlando, FL, 2001.

[5] J. Blitzstein, «Joe Blitzstein's answer to What is it like to design a data science class? In particular, what was it like to design Harvard's new data science class, taught by professors Joe Blitzstein and Hanspeter Pfister? - Quora», 2013. [En línea]. Disponible en: https://www.quora.com/What-is-it-like-to-design-a-data-science-class-In-particular-what-was-it-like-to-design-Harvards-new-data-science-class-taught-by-professors-Joe-Blitzstein-and-Hanspeter-Pfister/answer/Joe-Blitzstein. [Accedido: 19-sep-2019].

[6] C. O'Neil y R. Schutt, *Doing Data Science*. O´Reilly, 2015.

[7] M. Behrisch *et al.*, «Quality Metrics for Information Visualization», en *EuroVis 2018*, República Checa, 2018, vol. 37.

Una Estrategia para la Evaluación y Selección de Ideas en la Plataforma Ideas en Tiempo Real

Sandra Oviedo[1], Sandra Rodriguez[3], Daniel Díaz[1], Alejandra Otazú[2], Juan M. Cúneo[2], Marta Cabrera Villafañe[3]

[1]Laboratorio de Informática Aplicada a la Innovación, Instituto de Informática-
[1,2]Departamento de Informática - Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de San Juan, CUIM, San Juan, Argentina.
[3]Centro de Investigación y Desarrollo Informático (CIDI) -Sede Universitaria Chamical -Universidad Nacional de la Rioja. Chamical, La Rioja. Argentina
soviedo@iinfo.unsj.edu.ar; sandraorona27@yahoo.com.ar; ddiaz@iinfo.unsj.edu.ar; aleotazu@hotmail.com, jmacuneo@gmail.com, marta.cabreravilla@gmail.com

Resumen. La Plataforma Ideas en Tiempo Real es un asistente para el proceso innovador. Este trabajo presenta una estrategia para evaluar ideas y su aplicación en un estudio de caso. La estrategia se basa en métricas y se propone a fin de evitar la subjetividad en la evaluación de las ideas. Se prueba en la evaluación de las ideas de un grupo de jóvenes emprendedores, las ideas son para desarrollar una aplicación móvil para apoyo del turismo en un departamento cordillerano. Se presenta un conjunto de criterios de evaluación resultante de un trabajo colaborativo obtenido a partir de una encuesta. Los criterios, con una escala y un método asociado, se convierten en indicadores que permiten evaluar cada idea, obteniéndose una evaluación elemental. Mediante una función de agregación se obtiene una evaluación global, entregando como resultado un ranking de preferencias de ideas.

Palabras claves: Evaluación y Selección de Ideas. Ideas en Tiempo Real. Métricas. Innovación.

1 Introducción

Una idea innovadora se refiere específicamente a un producto de tecnología potencialmente convertible en un negocio sustentable; esto es, económicamente rentable, social y ambientalmente beneficioso, así como sostenible en el tiempo. Pues bien, ante esta necesidad de evaluar y seleccionar una idea innovadora de un conjunto de ideas resultantes de una sesión de creatividad es una tarea compleja. Se destaca el valor de las ideas, donde una mala evaluación puede llevar a descartar ideas con mucho potencial, provocando grandes pérdidas. La función de seleccionar una idea para su desarrollo está afectada por múltiples intereses y de distinta naturaleza u origen, muchas veces antagónicos. Con vistas a evitar la subjetividad en la evaluación de las ideas de un portfolio, se propone una estrategia basada en métricas que

permitirán cuantificar los atributos (preferencias) de cada idea. Tales atributos estarán en función de los objetivos de la organización, sus clientes o mercado objetivo y capacidades de producción, es decir que la evaluación que se propone responde a un contexto de análisis. Como resultado se espera obtener un ranking de ideas ordenado por preferencias.

Este trabajo se propone en el ámbito de la Plataforma Ideas en Tiempo Real (IdeTR), es una Plataforma Tecnológica para la Generación de Ideas de Productos Tecnológicos, se desarrolla en el Laboratorio de Informática Aplicada a la Innovación del Instituto de informática de la Facultad de Ciencias Exactas Físicas y Naturales de la Universidad Nacional de San Juan. Se trata de una aplicación web de tiempo real que asiste al proceso innovador. En su etapa inicial el sistema tiene dos grandes componentes, la gestión de usuarios y grupos y, un componente para asistir en la generación de ideas, basado en la técnica de Brainstorming. Permite planificar y ejecutar sesiones de creatividad online, de manera que los participantes pueden hacerlo de manera remota o localizada. En [1] se describe con más detalle la plataforma. Como resultado de una sesión de creatividad se obtiene un conjunto o portfolio de ideas. Actualmente se está trabajando en el diseño del componente para asistir a la evaluación de las ideas (proyecto denominado IdeTRES, Evaluación y Selección de Ideas en la Plataforma Ideas en Tiempo Real). Se espera ir agregando componentes a IdeTR a fin de lograr integrar herramientas para las diferentes etapas del proceso innovador.

2. Evaluación de ideas

Centrados en las empresas, la principal preocupación no es solo la selección de ideas, sino también, si las ideas seleccionadas podrían contribuir a la competitividad sostenida y crear un nivel de beneficio suficiente [2]. Por lo tanto, se debe tener en cuenta que la creación de ideas y su evaluación es parte de un entorno más amplio que incluye el establecimiento de objetivos, el mercado objetivo y la evaluación de capacidades que la empresa tiene a su disposición, es lo que se denomina Estructura del porfolio de ideas.

Existe muy poca literatura acerca de la evaluación de ideas [3], entre lo encontrado se destacan algunos trabajos como:

- En [2] se presenta una revisión de técnicas y métodos que podrían emplearse para evaluar ideas.
- En el trabajo de [4] se presenta un experimento realizado con 120 personas, la mitad evaluó ideas siguiendo una escala, asistidos por una herramienta, y la otra mitad evaluó ideas por intuición, los autores llaman a esta estrategia como preferencia del mercado. Como conclusión, la evaluación para los que usaban una escala ayudados por una herramienta resultó muy fácil, mientras que para los evaluadores por preferencia del mercado se tornó en una tarea dificultosa, lo cual afectó la calidad de la decisión.
- En [3] se plantea que la evaluación y selección de ideas es un problema que tiene lugar en el presente. Implementar todas las ideas para determinar cuál es la mejor no es práctico y sigue siendo imposible seleccionar ideas del futuro. Proponen

intentar predecir el futuro (definiendo mundos) y evaluar qué tan bien funcionará una idea determinada dentro de él. Se presenta como un trabajo en curso.

- En [5] se presenta un experimento para evaluación de ideas de puntuación (calificación) vs. clasificación. En el proceso de puntuación, se les pide a los evaluadores que califiquen la calidad de cada idea asignándole una puntuación, de 0 a 10, mientras que, en el proceso de clasificación, el evaluador simplemente ordena todas las ideas según sus cualidades percibidas.

En [6] se presenta un estudio acerca de la relación entre los creadores de las ideas y los evaluadores, se analizan diferentes hipótesis acerca de la distancia y nivel jerarquico que conviene que tengan unos sobre otros en una organización.

2.1 Métricas para reducir la subjetividad en la evaluación

Cuando se trata de evaluar una idea ya sea por el atributo que sea, por ejemplo, por novedad, por precio, por usabilidad, por tecnología que emplea, etc., es muy fácil ir al plano de lo subjetivo. Para vencer la subjetividad y lograr una evaluación de lo más objetiva, es necesario recurrir a una estrategia que garantice, en lo posible, la objetividad. Es por ello que se propone el uso de métricas para evaluar las ideas.

Una métrica consiste básicamente en la medición numérica de un atributo. Las métricas existen debido a la necesidad de tener información cuantitativa para la toma de decisiones [7]. Las métricas son específicas, en muchos casos, para un contexto determinado. Son válidas solamente dentro de cierto dominio y no pueden ser directamente comparadas o interpretadas fuera de él. Los siguientes elementos deben conocerse para que una métrica quede completamente definida en tabla 1.

Tabla 1. Elementos de una métrica

Elementos	Descripción
Métrica	Nombre de la métrica
Descripción de la métrica	Descripción de que se mide
Frecuencia de las mediciones	Cuantas veces es la medida tomada
Procedimiento de medición	Como es la métrica medida
Valoración de los umbrales	Como son los umbrales calculados
Umbrales actuales	Rango actual de los umbrales que se consideran normales
Valor Objetivo	El mejor valor posible para la métrica
Unidad de medida	Unidades de medida

La selección de las métricas a aplicar es equivalente a describir las características de las ideas requeridas por quien encara una evaluación. En general se evaluará considerando los objetivos de la organización, su mercado objetivo, clientes y capacidades. Generalmente, no puede decirse que una idea sea de mejor calidad que otra, en cambio, si puede decirse que una idea tiene preferencias sobre otra, entendiéndose como preferencia el peso de un atributo, o la satisfacción de requerimientos planteados. Todo dependerá del contexto en que se realiza la evaluación.

Realizada la determinación de las métricas a emplear, se define el procedimiento para su medición. Esto dependerá de la naturaleza de la métrica, en muchos casos, tratándose de atributos de las ideas, el procedimiento de medición podrá ser un cuestionario. Asociada al procedimiento de medición habrá una función, en muchos casos, una escala, que permita llevar la medición del atributo a un valor, obteniéndose así un indicador.

Se puede realizar una evaluación elemental obteniendo todos los indicadores de los atributos evaluados para una idea.

2.2 Selección de ideas

Una vez que se haya hecho la evaluación elemental donde se valoró cada atributo de cada idea, se puede realizar una evaluación global. Mediante un modelo de agregación, que puede incluir la aplicación de un método de decisión multicriterio [8], se puede obtener un valor para cada idea. Según [3] hasta una idea puede ser medida y expresada numéricamente.

El modelo de agregación podría, además, quedar definido por reglas de exclusión y preferencias.

Luego de aplicado el modelo de agregación se podrá obtener un valor para cada idea, con lo que se podrá finalmente hacer las comparaciones necesarias, y establecer un orden o ranking de ideas, dejando en los primeros puestos las consideradas "mejores". Posibilitando de esta manera la selección de la o las ideas, según el contexto en el que fueron evaluadas.

3. Estrategia para evaluación y selección de ideas

A continuación, se propone una estrategia para evaluar las ideas de un portfolio de ideas, donde se integran los conceptos del apartado precedente.

La estrategia propuesta está basada en un modelo jerárquico de requerimientos de evaluación, partiendo de la estructura del portfolio. A partir de esas características, se derivan subcaracterísticas y, a partir de éstas, siguiendo un proceso de descomposición jerárquico, se especifican (asocian) atributos, el resultado de este proceso se denomina árbol de requerimientos.

El árbol de requerimientos permite capturar la estructura organizacional y además con una correcta segmentación se llega fácilmente a la agregación, la cual permite hacer una evaluación en todos los aspectos.

El principal objetivo de esta estrategia consiste en evaluar y determinar un ranking de ideas que conforman un portfolio de ideas. Para esto, se analizan los indicadores (también llamados variables de preferencia o de performance) obtenidos para cada atributo, luego su integración en un valor total asignado a cada idea y finalmente estableciendo un orden de preferencias de las ideas evaluadas. El resultado del proceso de evaluación (y eventualmente de comparación y selección) puede ser interpretado como el ordenamiento de las ideas que pueden ser de mayor interés para determinado momento o circunstancia de la organización y del mercado.

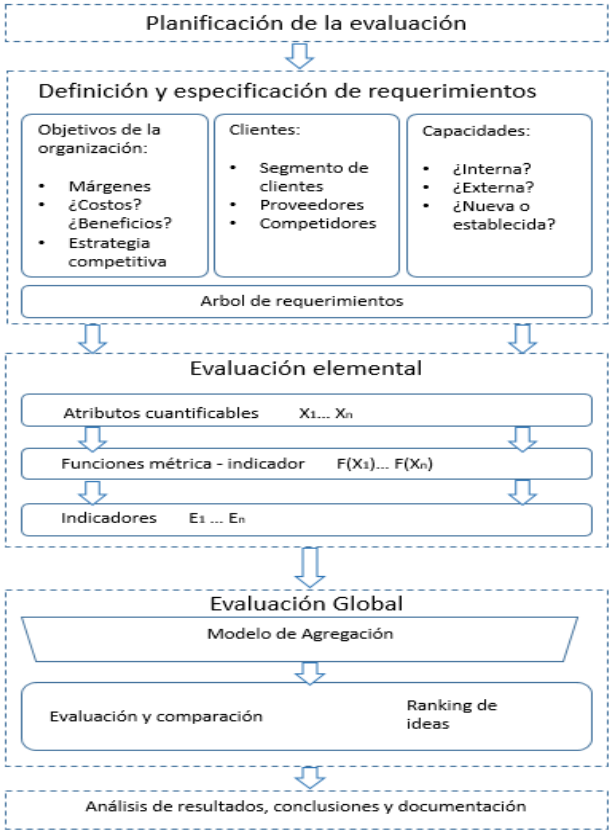


Fig. 1. Estrategia para evaluación de ideas

A continuación, se realiza una breve explicación de cada una de las fases, según se muestra en la figura 1.

1. Planificación de la evaluación. Contiene actividades y procedimientos de soporte, con el fin de determinar objetivos estratégicos, tácticos y operativos. Permite establecer las principales estrategias y metas del proceso en un contexto organizacional, seleccionar un proceso de evaluación, programar y asignar métodos, agentes y recursos a las actividades de evaluación. Es importante en esta etapa establecer quienes serán los evaluadores de las ideas, según [6] es inevitable la parcialidad si las mismas personas que generan las ideas, luego las evalúan.
2. Definir y especificar los requerimientos, entendiéndose los requerimientos como atributos deseables en las ideas. Lo constituyen un conjunto de actividades y modelos para la identificación, determinación, análisis y especificación de los requerimientos. A partir de un proceso de medición y evaluación orientado a objetivos comunes de la organización, y con el fin de analizar, evaluar, comparar, y mejorar la selección de ideas, se establece los requerimientos que deben responder a necesidades explícitas e implícitas de un perfil para un dominio dado. El producto de esta fase es

un documento que jerárquicamente especifica todas las características y atributos cuantificables que se deberían tener en cuenta en la evaluación de las ideas.

3. Definir e implementar la evaluación elemental. Tiene relación con actividades, modelos, técnicas y herramientas para determinar métricas y criterios de evaluación para cada atributo cuantificable. Se consideran funciones para determinar indicadores elementales a partir de métricas, rangos de aceptabilidad, entre otros asuntos. Una vez definidos y relacionados los criterios para medir cada atributo, se debe analizar cada idea, computar las métricas e indicadores elementales, y por último documentar los resultados encontrados.
4. Definir e implementar la evaluación global. Comprende las actividades, modelos, y herramientas para determinar los criterios de agregación de los indicadores (preferencias) elementales para producir la preferencia global, para cada idea analizada. Como función de agregación se podría utilizar la suma de indicadores pesados, el peso del indicador representa la importancia de cada indicador.
5. Analizar los resultados, concluir y documentar. Trata de actividades de análisis, comparación de las preferencias de indicadores elementales y globales, y la justificación de los resultados. Se utilizan herramientas y mecanismos de documentación para facilitar la interpretación de los datos y su seguimiento.

4. Un componente para asistir en la evaluación de ideas en IdeTR

La estrategia de evaluación de ideas que se presenta es una primera aproximación para diseñar un componente funcional para la plataforma IdeTR. En el workflow de la plataforma, las actividades 1 y 2 de la estrategia se realizarían en el momento de planificar la sesión de creatividad, esto es cuando se establece la Temática sobre la que se generarán las ideas, deberían establecerse los criterios, métodos y escalas de medición, así como designarse los evaluadores. Una vez obtenidas las ideas, se podrá avanzar con las demás actividades para la evaluación.

5. Aplicación en un estudio de caso

Se pretende poner a prueba la estrategia definida en el contexto que a continuación se describe.

5.1 Contexto

Un grupo de estudiantes del último curso de la Licenciatura en Ciencias de la Computación de la Facultad de Ciencias Exactas físicas y Naturales de la Universidad Nacional de San Juan, se propone como objetivo iniciar un emprendimiento mediante el desarrollo de una aplicación para celulares a solicitud de un municipio cordillerano de San Juan, Argentina. Se trata de un departamento turístico, donde se destaca la práctica del turismo aventura, así como el turismo rural, con cabalgatas y senderismo de montaña. El departamento dispone de acceso a tecnología móvil 4G.

El equipo interesado tiene por objetivo desarrollar una aplicación para dispositivos móviles en una primera versión que tenga la mayor cantidad de usuarios posibles.

Las tecnologías de desarrollo que los estudiantes dominan son: JavaScript, HTML5, .Net, Android, entre las más conocidas.

Mediante una sesión de brainstorming los estudiantes obtuvieron un conjunto de ideas, entre las que deberán elegir una para desarrollar. Por lo tanto, se pretende evaluar las ideas de un conjunto de ideas de aplicaciones para móviles aplicadas al turismo de la región cordillerana de San Juan.

En el contexto planteado, las ideas propuestas están en tabla 2.

Tabla 2 Porfolio de ideas a evaluar.

Id	Idea descripta
1	Juegos de búsqueda de tesoro (orientados a alguna localidad específica) considerando el uso de GPS.
2	A partir de una imagen propuesta que los usuarios determinen las coordenadas del punto desde donde se tomó la fotografía.
3	Un navegador localizado que permita mediante la aplicación ir haciendo un recorrido turístico.
4	Un sistema de alarmas que avise cuando en una excursión haya algún punto de interés.
5	Rastreador de mascotas perdidas.
6	Rastreador de cámaras fotográficas u otros objetos de valor perdidos.
7	Una aplicación que grabe lo que una persona cante siguiendo un karaoke para que lo pueda escuchar con el celular.
8	Una aplicación que cuente los km recorridos, calcule la velocidad y la cantidad de calorías quemadas por el usuario.
9	Una aplicación que dé la alarma de cuándo es menos nocivo exponerse al sol, es decir que avise cuándo es recomendable o no tomar sol.

5.2 Evaluación

Para aplicar la estrategia de evaluación se procede a aplicar cada una de las actividades previstas:

1. Planificación de la evaluación. Se determinó que los criterios de evaluación y sus pesos (o importancia) serían definidos mediante una encuesta realizada a personas de opinión autorizada, en este caso, profesionales informáticos, de marketing, y experimentados en formulación y evaluación de proyectos tecnológico. La encuesta se realizará de manera online mediante el uso de un formulario de Google. También se determinó quienes formaran parte del equipo evaluador, recayendo la tarea en 3 profesores de las carreras de informática, de diferentes áreas curriculares.
2. Definir y especificar los requerimientos, para obtener un árbol de requerimientos, se propone la pregunta ¿qué se espera de la idea elegida? Para ello se describió un contexto que destaca características de potenciales usuarios de las aplicaciones: personas que realizan turismo aventura, turistas que disfrutan del turismo rural. Se describió algunas aspiraciones de los interesados en desarrollar la idea elegida, consistente en una aplicación: que sea para la mayor cantidad de usuarios posibles, (tratándose de un grupo de estudiantes que quiere insertarse en un

mercado, su interés está en lograr la mayor visibilidad posible, en su primera incursión). También se destacó las capacidades que tienen en cuanto a posibilidades técnicas los miembros del equipo ejecutor de la idea y las posibilidades tecnológicas que ofrece el lugar para su despliegue. De esta manera los requerimientos deberán cubrir los 3 aspectos, como se describió en apartados precedentes: Clientes, Objetivos y Capacidades de la organización (en este caso del equipo de emprendedores). La interpretación de estas características se presenta en los criterios de evaluación que se proponen para evaluar las ideas. Dichos criterios son obtenidos a partir de una encuesta realizada al grupo de expertos descripto en el paso 1. También en la misma encuesta se asignan pesos o nivel de importancia a tales criterios.

Se analizaron los resultados de la encuesta, los criterios propuestos por los encuestados son los que se describen a continuación. En la actividad 3 Definir e implementar la evaluación elemental, se definieron métodos para evaluación y escala de valores que pueden tomar los criterios, que también se describen a continuación:

C1. Originalidad: Da una noción del grado de novedad de la idea propuesta. Se puede medir mediante las siguientes preguntas ¿Existen aplicaciones con funcionalidades similares? Si la respuesta es NO se asigna un valor igual a 1, si la respuesta es SI, se podría hacer otra pregunta, ¿Cuán parecida son las aplicaciones que conoce con esta propuesta? 0 para la respuesta "muy parecida", 0,5 para una respuesta "poco parecida".

C2. Factibilidad Técnica: El evaluador podrá asignar un valor a este criterio teniendo en cuenta: Los requerimientos tecnológicos para desarrollar la idea vs. la capacidad del equipo ejecutor. Los requerimientos de uso vs. las posibilidades que ofrece el lugar y sus turistas al respecto. La respuesta a cada uno de estos ítems será en una escala de 0 a 1, y el criterio tomará el valor promedio entre ambos.

C3. Factibilidad Económica: Con este criterio se evaluará la posibilidad económico financiera para desarrollar y poner en marcha la aplicación de la idea propuesta. Siempre teniendo en cuenta las capacidades del equipo ejecutor, se evaluará la capacidad para afrontar la inversión que implica el desarrollo de la idea. Según el grado de dificultad que esto represente podrá asignarse un valor entre 0 y 1.

C4. Potencial cantidad de usuarios, teniendo en cuenta que el interés está en lograr la máxima visibilidad, es importante que la aplicación atraiga a un público variado, eso es lo que se evaluará en este criterio. La evaluación podrá enfocarse en analizar qué público podría atraer, si es para uso muy exclusivo se asignará un puntaje inferior a 0,5 si es para un público masivo puede alcanzar un valor igual a 1.

C5. Propicia buenas prácticas ambientales, si la idea involucra alguna acción de cuidado del medio ambiente toma valor 1 este criterio, sino, valor 0. Se propuso con objetivo de estimular el cuidado ambiental ya que en la región donde se implementará se trabaja fuertemente en esta temática.

Otros criterios son: C6. Funcionalidad. C7. Escalabilidad. C8. Impacto social C9. Diseño, por cuestiones espacio no se explica el método de medición, lo cual no les resta importancia a tales criterios. Se aclara que se realizan siguiendo la misma idea de las evaluaciones explicadas en los otros criterios.

Continuando con la actividad 3 se asignan pesos a cada criterio a evaluar, tabla 3.

Tabla 3 Criterios a evaluar y pesos asignados a cada uno.

Criterio		Peso
C1	Originalidad	20
C2	Factibilidad técnica	22
C3	Factibilidad económica	21
C4	Potencialidad de cantidad de usuarios	8
C5	Propicia buenas prácticas ambientales	3
C6	Escalabilidad	4
C7	Funcionalidad	11
C8	Impacto social	7
C9	Diseño	4

Teniendo en cuenta la tabla 3, el documento que describe los criterios, métodos a aplicar para evaluación y escalas de valores asignadas, los evaluadores proceden a evaluar las ideas de tabla 2. Se obtienen los indicadores (valores que toman los atributos o criterios), tabla 4.

Tabla 4. Evaluación elemental

Id idea	C1	C2	C3	C4	C5	C6	C7	C8	C9
1	0	0,75	1	0,8	0	1	0,4	0,5	0,8
2	1	0,5	1	0,8	0	1	0,4	0,5	0,5
3	0	0,5	1	0,7	1	1	0,7	0,5	0,8
4	1	0,5	1	0,7	0	1	0,5	0,5	0,8
5	0,5	0,5	1	0,25	0	1	0,8	0	0,8
6	0,5	0,5	1	0,6	0	1	0,8	0	0,8
7	0,5	1	1	0,1	0	1	0,3	0	0,8
8	0	0,3	1	0,2	0	1	1	0,5	0,8
9	0	0,5	1	1	1	1	1	0,3	0,8

Luego se puede aplicar el siguiente paso en la evaluación. 4. Definir e implementar la evaluación global. Se aplica como modelo de agregación la suma combinada de los valores elementales por el peso asignado a cada criterio, se obtienen los siguientes valores para cada función: en la tabla 5 se presentan ordenados según el ranking de valores alcanzados.

Además, se completa la evaluación aplicando reglas de exclusión que se propusieron en la encuesta, entre las que más hincapié se hizo están las siguientes reglas “que propicie prácticas ilegales” y “que su uso atente contra la seguridad de las personas”. Dadas las características de las ideas evaluadas ninguna de ellas es excluida y el ranking de ideas se conserva.

6. Conclusiones y trabajos futuros

Se propuso una estrategia de evaluación de ideas basada en métricas, para evaluar y seleccionar ideas en la Plataforma ideas en Tiempo Real. Se hizo el seguimiento de la estrategia para evaluar las ideas de un contexto y dominio determinado, lográndose probar con buenos resultados la estrategia.

Como trabajo futuro, se propone avanzar en la implementación de la estrategia probada en la Plataforma IdeTR.

Desde el punto de vista metodológico el trabajo futuro que se propone a la brevedad es experimentar con el uso de Métodos de Análisis Multicriterio para mejorar el modelo de agregación, en la etapa de Evaluación Global.

Tabla 5. Evaluación global. Ranking de preferencias de las ideas evaluadas.

Id idea	Idea descripta	Evaluación Global
4	Un sistema de alarmas que avise cuando en una excursión haya algún punto de interés.	73,8
2	A partir de una imagen propuesta que los usuarios determinen las coordenadas del punto desde donde se tomó la fotografía.	72,3
7	Una aplicación que grabe lo que una persona cante siguiendo un karaoke para que lo pueda escuchar con el celular.	64,3
9	Una aplicación que dé la alarma de cuándo es menos nocivo exponerse al sol, es decir que avise cuándo es recomendable o no tomar sol.	63,3
6	Rastreador de cámaras fotográficas u otros objetos de valor perdidos.	62,8
5	Rastreador de mascotas perdida.	60
1	Juegos de búsqueda de tesoro (orientados a alguna localidad específica) considerando el uso de GPS.	59
3	Un navegador localizado que permita mediante la aplicación ir haciendo un recorrido turístico.	59
8	Una aplicación que cuente los km recorridos, calcule la velocidad y la cantidad de calorías quemadas por el usuario.	50,9

Referencias

1. Diaz Araya, D., et al., *Ideas en tiempo real, un asistente para la generación de ideas en tiempo real.* Actas de las Jornadas Argentinas de Tecnología, Innovación y Creatividad 2017. JATIC 2017. Mar del Plata. ISBN 978-987-46267-2-1, 2017.
2. Rebernik, M. and B.B. Hojnik, *Module 4: Idea Evaluation Methods and Techniques*. 2008: Ekonomska-poslovna fakulteta.
3. Forde, A.N. and M.S. Fox, *A Proposed Approach for Idea Selection in Front End of Innovation Activities* Technology Innovation Management Review, 2016. **6**(8).
4. Blohm, I., et al., *Rate or Trade? Identifying Winning Ideas in Open Idea Sourcing.* Information Systems Research, 2016. **27**(1): p. 27-48.
5. Cui, Z., S. Kumar, and D. Goncalves, *Scoring vs. Ranking: An Experimental Study of Idea Evaluation Processes.* Production and Operation Management Society, 2019. **28**(1): p. 176-188.
6. Schweisfurth, T., M. Zaggl, and C.P. Schöttl, *Does similarity between evaluator and creator affect the evaliation of ideas?* . Academy of Management Proceedings, 2017(1).
7. Screpnik, C.R., *Métricas Aplicables a la Evaluación de Sitios e-government y su Impacto Social.* Universidad de La Plata, 2013.
8. Rodriguez, S., E. Rearte, and M. Cabrera Villafañe, *Una propuesta de decisión multicriterio en IdeTR.* Workshop de investigadores en ciencias de la computación 2019. WICC 2019, 2019.

Aplicación de Internet de las Cosas a la administración y control de un robot autónomo móvil

Amaro, M. , Azcurra, D.

Dpto. Desarrollo Productivo y Tecnológico
Universidad Nacional de Lanús.
29 de Septiembre 3901 (1826) Remedios de Escalada, Lanús
Buenos Aires, Argentina.
marcos.n.amaro@gmail.com

Resumen. Internet de las Cosas (IoT, por sus siglas en inglés) es un concepto de computación en el que todas las cosas, incluyendo todo objeto físico, pueden ser interconectadas a través de Internet. Una plataforma de IoT, por su parte, es el soporte que permite lograr dicha interconexión y se genere un "ecosistema" propio. Está constituida –principalmente- por el software al que se conectan los dispositivos de hardware, brindando puntos de acceso y comunicación para el desarrollo de aplicaciones. En este trabajo se presenta la implementación de un modelo de administración y control de un robot autónomo móvil basado en una plataforma de IoT y hardware de bajo costo.

Palabras clave: Internet de las cosas, robot autónomo móvil, sistemas de control, plataforma de Iot, IBM Watson, Arduino, NodeMCU.

Introducción

El presente trabajo se desarrolló en el marco del proyecto “Robótica basada en Internet de las Cosas”[1] , el cual integra la línea de investigación en aplicaciones de sistemas industriales, robótica y telecomunicaciones en la carrera de la Carrera de Licenciatura en Sistemas de la Universidad Nacional de Lanús.
Durante el desarrollo del mismo se buscó llevar a adelante la implementación de un modelo de administración y control de un robot autónomo móvil utilizando una plataforma de IoT.

Por robot autónomo móvil se entiende un vehículo con características que le permiten acelerar, frenar, y manejar con una limitada o ninguna interacción por parte de un conductor y puede dividirse en dos tipos:

- Semiautónomo
- Completamente autónomo

Los semiautónomos son aquellos que pueden acelerar, frenar y manejarse, pero también requiere de un conductor el cual puede estar en completo control del

vehículo. En cambio, los completamente autónomos son aquellos que se pueden manejar sin la necesidad de interacción humana [2].
En otras palabras, el funcionamiento de un vehículo autónomo busca imitar las capacidades humanas de manejo [3]. Para alcanzar dicho propósito se utilizan diferentes sensores que van desde cámaras de video, GPS, sensores de proximidad, etc., para obtener datos del contexto que lo rodea, procesarlos y tomar una decisión como puede ser cambiar de carril, frenar, esquivar un peatón o recorrer una ruta determinada.

Plataformas de IoT

Antes de explicar que se entiende por plataforma de IoT, es conveniente ver los componentes que conforman un 'ecosistema' IoT:

1. **Hardware**: sensores, actuadores y demás dispositivos electrónicos que obtienen datos del medio ambiente (por ejemplo, un sensor de humedad) o ejecutan acciones en él (por ejemplo, un sistema de riego).
2. **Conectividad**: el hardware necesita transmitir los datos recolectados hacia la nube (como el caso de la medición de humedad) así como también puede necesitar recibir órdenes (iniciar el regado).
3. **Software**: el software alojado en la nube es responsable de analizar los datos enviados por los sensores y de tomar decisiones en respuesta al análisis realizado.
4. **Interface de usuario:** es el medio por el cual los usuarios pueden interactuar con el sistema de IoT, generalmente será mediante un panel de control basado en una aplicación web que permitirá al usuario ver la información recolectada por los sensores y enviar órdenes en forma manual hacia los dispositivos.

Entonces, una plataforma de IoT es el software de soporte que conecta todo en un "ecosistema" IoT. Facilita la comunicación, el flujo de datos, la administración de dispositivos y la funcionalidad de las aplicaciones. Se puede entender a una plataforma de IoT como una tecnología multicapas, que conecta sensores/dispositivos, los administra y brinda la posibilidad de automatizar la conexión de los dispositivos dentro del mundo de Internet de las Cosas.

Constituyen la clave para el desarrollo de aplicaciones IoT escalables y servicios que conectan el mundo real con el virtual, y presentan las siguientes ventajas:

1. **Escalabilidad**: facilita la adición de más conexiones a dispositivos.
2. **Personalización**: modificables a las necesidades del negocio y la adición de más características.
3. **Seguridad**: ofrecen encriptamiento y segurización de los datos.
4. **Facilidad de uso**: brindan al usuario un entorno gráfico y visualización sencilla de su ambiente IoT.

Componentes del proyecto

Para la implementación del proyecto se desarrolló un robot Rover de 4 ruedas con las siguientes características de hardware:

- 4 motores DC.
- Sensor de distancia de ultrasonido.
- Placa Arduino Uno R3 para la ejecución de comandos, control del sensor de distancia e integración con el motorshield.
- Placa NodeMCU V3 Amica, para la comunicación vía Wi-Fi con la plataforma de IoT, envío y recepción de estados y comandos.
- Adafruit Motorshield V1, para el funcionamiento y control de los motores.

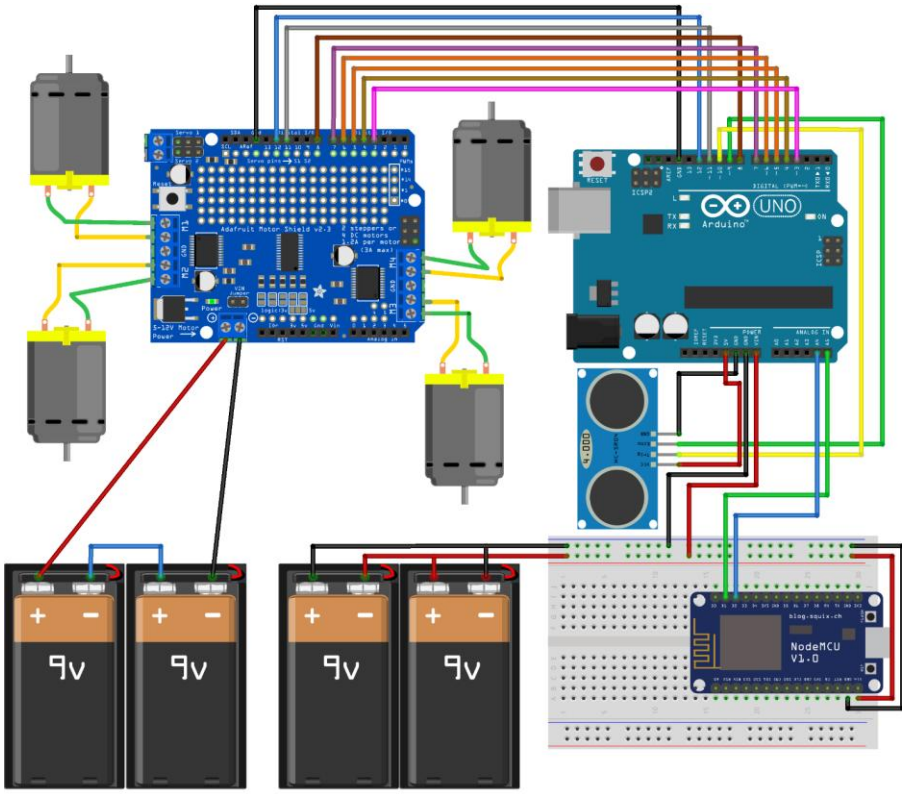


Fig. 1. Diagrama de conexión del robot autónomo móvil

Referido a la elección de la plataforma de IoT, se tuvieron en cuenta las plataformas estudiadas por Higa,H. en su trabajo “Estudio de las Plataformas de Internet de las

Cosas Utilizables en Robótica Autónoma" [4], las cuales son ThingsBoard [5], Thinger.io [6], Adafruit [7], MyDevices Cayenne [8], AWS IoT Core [9], Azure IoT hub [10], Google Cloud IoT Core [11], y por último IBM Watson IoT [12].

Entre las mencionadas, se eligió llevar a cabo el proyecto utilizando la plataforma de IBM Watson, por las siguientes razones:

- Soporta el hardware a utilizar en este proyecto sin necesidad de agregar un sistema operativo intermediario, en este aspecto, por ejemplo, Google Cloud requiere utilizar Mongoose OS para conectar un NodeMCU a la plataforma.
- Si bien se puede utilizar certificados para la seguridad con IBM Watson, no es un requisito necesario y ofrece conexión segura mediante usuario y token. Esto, en placas como la NodeMCU, es fundamental ya que la inclusión de certificados con la misma es muy compleja de realizar.
- Cuenta con la posibilidad de programar scripts en la plataforma utilizando Node-RED, esta herramienta, creada por IBM, brinda la posibilidad de escribir scripts del lado de la plataforma en una forma visual. Permite conectar los dispositivos, las APIs y los servicios IoT para trabajar en conjunto. Este aspecto fue determinante en la elección, ya que IBM permite conectar directamente el servicio Bluemix con Node-RED, manteniendo la conexión activa en todo momento y reduciendo los tiempos de respuesta al no tener que estar reconectándose continuamente.
- Cuenta con un plan gratuito y no requiere de datos personales (tales como la tarjeta de crédito) para su utilización.
- La documentación, tanto de la plataforma como de Node-RED es muy amplia, otra característica importante para reducir la curva de aprendizaje y solucionar rápidamente los problemas que se presentan en su uso.

En cuanto al software embebido en las placas, se programaron en C utilizando Arduino IDE, y del lado del servidor en la plataforma de IoT, se utilizó Node-RED programando las rutinas de control y procesamiento de datos recibidos en lenguaje Javascript.

Administración y control

El sistema de administración y control del robot puede representarse como un esquema de maestro y esclavo, donde el esclavo es el robot, quien expone las funciones o acciones que sabe hacer, y el maestro, situado en la nube y accesible mediante IBM Watson IoT/Node-RED, quien envía las órdenes al esclavo para que éste las ejecute. En la tabla 1 se especifican las funciones programadas para el sistema embebido, en la figura 2 se muestra el diagrama de flujo en Node-Red correspondiente al control y administración desde la plataforma de IoT mediante un

panel de control (o dashboard) mostrado en la figura 3, donde el selector "Habilitar" corresponde a la habilitación de la rutina de control autónomo del robot.

Tabla 1. Resumen de funciones embebidas a las placas del robot.

Función	NodeMCU	Arduino UNO R3
Cambiar Modo	Recibe esta orden de la plataforma de IoT y la envía a la Arduino UNO R3	Los valores posibles son 0 y 1: Manual (0): el robot ejecuta la orden por pasos, cada paso es cada orden recibida. Automático (1): ejecuta la orden recibida sin detener los motores, hasta que reciba una orden de detenerse o se cambie al modo Manual.
Distancia	Recibe esta orden de la plataforma de IoT y la envía a la Arduino UNO R3	Representa el valor por el cual el robot considera que está frente a un obstáculo y se detiene si está a la misma o menor distancia de la configurada por esta función.
Avanzar	Recibe esta orden de la plataforma de IoT y la envía a la Arduino UNO R3	Gira los motores en sentido horario hacia adelante junto a la velocidad a la que deben girar los motores.
Retroceder	Recibe esta orden de la plataforma de IoT y la envía a la Arduino UNO R3	Gira los motores en sentido antihorario, hacia atrás junto a la velocidad a la que deben girar los motores.
Girar hacia la izquierda	Recibe esta orden de la plataforma de IoT y la envía a la Arduino UNO R3	Gira el robot hacia el lado izquierdo a la velocidad indicada desde la plataforma de IoT.
Girar hacia la derecha	Recibe esta orden de la plataforma de IoT y la envía a la Arduino UNO R3	Gira el robot hacia el lado derecho a la velocidad indicada desde la plataforma de IoT.
Detener	Recibe esta orden de la plataforma de IoT y la envía a la Arduino UNO R3	Detiene el robot.
Delay	Desde la plataforma se puede especificar un delay sanitario entre órdenes que finalmente debe ejecutar la Arduino UNO. Esto se hizo necesario para dar un tiempo prudencial a la comunicación I2C.	No recibe esta orden.

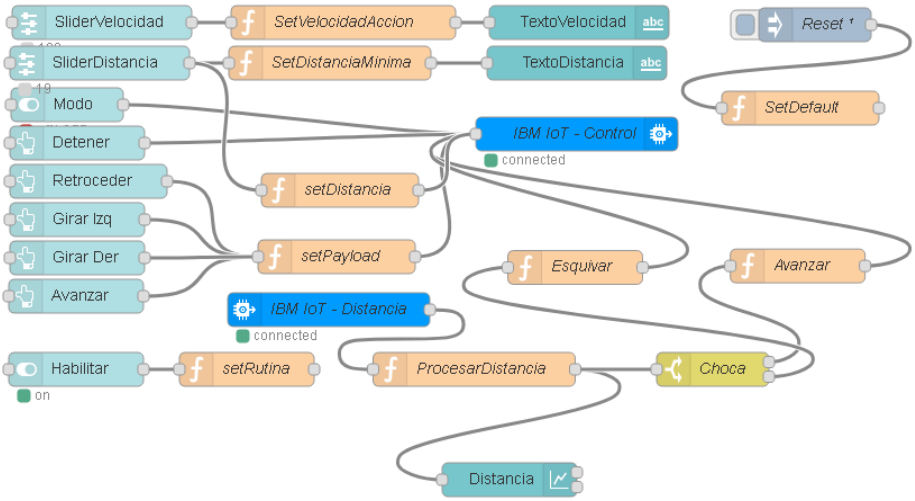


Fig. 2. Diagrama de flujo de Node-RED

Fig. 3. Panel de control (o dashboard) para administrar y controlar el robot en forma remota

Pruebas

Se realizaron las dos pruebas correspondientes tanto para el control manual del robot como para el control autónomo. Las características probadas fueron las siguientes:

1. **Avanzar**: el robot debe avanzar hacia adelante. La delantera del robot está determinada por la ubicación del sensor de distancia ultrasónico.
2. **Retroceder**: el robot debe retroceder, es decir, en sentido contrario donde está ubicado el sensor de distancia.
3. **Girar a la izquierda**: el robot debe girar a su izquierda.
4. **Girar a la derecha**: el robot debe girar a su derecha.
5. **Detener**: el robot se detiene.
6. **Cambiar a modo manual**: en este modo, la acción se ejecuta una sola vez, que es aproximadamente un solo giro del motor.
7. **Cambiar a modo automático**: este modo hace que la acción se ejecute indefinidamente, es decir, si se envió la orden de avanzar, lo seguirá haciendo hasta que se le indique una acción diferente.
8. **Cambio de velocidad**: el robot debe girar los motores según la velocidad en la que esté el control de velocidad en el dashboard.
9. **Control autónomo**: el robot debe avanzar de forma automática, al encontrarse frente a un obstáculo debe ejecutar la rutina programada en la plataforma IoT la cual consiste en dar unos pasos hacia atrás y hacer un giro de 45º aproximadamente y luego seguir avanzando. En caso de seguir frente al obstáculo, vuelve a repetir la rutina hasta poder avanzar sin obstáculos en frente.

Las pruebas realizadas concluyeron exitosamente. La demostración puede verse en los siguientes videos: https://youtu.be/wwJqtjVR7Do (control manual) y https://youtu.be/mVZugv5nK7w (control autónomo).

Además de las pruebas funcionales, se realizó un análisis cuantitativo mediante la medición de la latencia en la comunicación entre el dispositivo NodeMCU y la plataforma Watson IoT, en donde se ejecutaron 1000 llamados sobre la rutina de control autónomo, donde la placa NodeMCU envía el valor de distancia hacia la plataforma IoT y en Node-RED, a partir del valor recibido, se ejecuta el flujo correspondiente devolviendo el conjunto de instrucciones que debe ejecutar el robot. Los resultados fueron los siguientes:

Tabla 2. Parámetros sobre los que se realizó la prueba de latencia

Tipo de Conexón	Velocidad	Total de muestras tomadas
4G compartida desde celular por Access point Wi-Fi	13 Mbps, según medición en el sitio Fast.com	1000

Tabla 3. Resultados de la prueba de latencia

Unidad de tiempo	Máximo	Mínimo	Promedio
Milisegundos	625	469	504

Conclusiones

En la implementación realizada se exponen las funcionalidades que pueden ser tomadas como base para realizar desarrollos más avanzados, entre estas funcionalidades se pudieron implementar:

- Control remoto desde internet de un robot móvil.
- Control autónomo del robot basado en la programación de scripts que envían órdenes al dispositivo según los estados y valores que éste informa a la plataforma de IoT.
- Configuración de parámetros del robot desde la nube, como ser la distancia a la cual debe detenerse y esperar órdenes, la velocidad a la cual deben girar los motores, la forma en la que debe moverse.

En referencia a las plataformas de IoT, las interfaces ofrecidas al usuario, actualmente son bastantes intuititvas y amigables haciendo que la configuración de los dispositivos sea sencilla y con herramientas como Node-RED el desarrollo de un proyecto IoT se facilita aún más. Por otro lado, al ser una herramienta de código abierto también da la posibilidad de adaptarla, crear nuevos componentes y funcionalidades o simplemente añadirle los nodos que la comunidad de Node-RED publique para el resto de los desarrolladores.

En resumen y como conclusión final, en este trabajo se muestra que el desarrollo de un proyecto de IoT no requiere de una gran inversión inicial, ya que puede realizarse con materiales de bajo costo, el software requerido en su mayoría es libre y cuenta con una gran comunidad en la cual apoyarse y que existen diversas plataformas de IoT (muchas de ellas gratuitas o con períodos gratuitos) que se pueden utilizar según las necesidades que requiera la solución a implementar.

Referencias

1. Azurra, D. Santos D., Fernández, G., Fernández S.: Robótica basada en Internet de las Cosas. Proceedings del XX Workshop de Investigadores en Ciencias de la Computación. Pág. 835 – 837. ISBN 978-987-3619-27-4. (2018)

2. Krasniqi, X., Hajrizi, E.: Use of IoT Technology to Drive the Automotive Industry from Connected to Full Autonomous Vehicles. University of Business and Technology, Kosovo. (2016)
3. Prince, A. Jolías, K.: Tendencias tecnológicas (1ª ed.). Buenos Aires, Argentina: Ed. Autores Argentinos. ISBN 978-987-1791-10-1 (2016)
4. Higa, H.: Estudio de plataformas de Internet de las Cosas utilizables en robótica autónoma. Universidad Nacional de Lanús, Argentina. (2019)
5. ThingsBoard, https://thingsboard.io/docs/getting-started-guides/what-is-thingsboard/
6. Thinger.Io, https://thinger.io/
7. Adafruit, https://io.adafruit.com/
8. MyDevices Cayenne, https://mydevices.com/
9. AWS IoT-Core, https://aws.amazon.com/es/iot-core/
10. Azure IoT-Hub, https://azure.microsoft.com/en-us/overview/iot
11. Google Cloud, https://cloud.google.com/solutions/iot/?hl=es
12. IBM Watson, https://www.ibm.com/internet-of-things

Privacidad en Datos de Salud con Tecnología Blockchain

María del Carmen Becerra[1], Pedro Zarate[2]
[1,2] Instituto de Informática – Laboratorio de Análisis Forense e Informática Jurídica.
[1]Departamento de Informática – FCEFN-UNSJ
[1]mcbecerra2008@gmail.com, [2]pzarate@iinfo.unsj.edu.ar

Abstract. El trabajo propone una gestión eficiente de los datos de salud basado en las oportunidades que brinda la tecnología blockchain para la prestación eficiente de los servicios de salud insertos en sistemas distribuidos particularmente orientados a la descentralización. Se valora el aporte de blockchain para lograr una estructura más sólida y viable en cuanto a la organización de atención a la salud y el cumplimiento de reglas estrictas designadas en la legislación nacional para proteger la confidencialidad e integridad de los datos de salud. Se busca reforzar las relaciones de confianza a través de la gestión segura de la historia clínica electrónica Se analiza la estructura administrativa, técnica y física necesaria para la privacidad de los datos de salud del paciente. Se evalúa la res-ponsabilidad legal que provocan estas tecnologías en la privacidad de los datos de salud en relación al profesional, al paciente, a los establecimientos hospitalarios privados y semiprivados y al sistema de salud en general. Se concluye en que es necesario utilizar blockchain y redactar contratos inteligentes porque mediante su sofisticada codificación brindan una solución definitiva para mantener a salvo la privacidad del historial médico.

Keywords: Blockchain en Salud. Identidad Digital. Creación de contratos. Privacidad en blockchain.

1 Introducción

Al igual que la PC o Internet, la tecnología blockchain representa el poder transformador, lo que les permite lograr mejoras en la gestión de consentimiento de los pacientes y en la trazabilidad del dato médico. Sin embargo, en lugar de revolucionar la logística o la producción física, blockchain puede redefinir los procesos transaccionales y las relaciones que los acompañan, que existen entre el profesional, el paciente, los establecimientos hospitalarios públicos y privados y el sistema de salud en general.
Respecto a confidencialidad y disponibilidad, blockchain no elimina la confianza en los procesos negociales, sino que minimizan la cantidad de seguridad que se requiere de cada participante. Lo hacen mediante la distribución y el aporte de seguridad entre los diferentes actores del sistema de salud, a través del sistema económico que incentiva a los actores a cooperar siguiendo rigurosamente las reglas definidas por el protocolo.

La irrupción de la tecnología blockchain abre las puertas al nuevo modelo para una gestión eficiente y eficaz de los datos de salud del paciente logrando la inmutabilidad y trazabilidad de sus datos médicos en el sector de la salud, ello tendrá repercusión con implicaciones para todos los agentes de la cadena, tanto pacientes, industria y gestores. Los pacientes podrán acceder a sus datos de salud en cualquier lugar, en cualquier momento, de cualquier dispositivo, y podrán decidir según sea necesario, compartir sus atributos de identidad con el ecosistema de salud en general. En la cadena de producción de los medicamentos se tendría la total seguridad de cómo los medicamentos fueron creados, se han distribuido y han llegado a su punto final. Y las administraciones podrían lograr eficiencias hasta ahora difíciles de conseguir.

Sin embargo, los métodos para garantizar la integridad de los archivos médicos tienen como problemática el respaldo de dicha característica apoyado en un servidor central con lo cual, es relativamente fácil que alguien con malas intenciones pueda realizar alteraciones de estos datos, sin dejar rastros. Las políticas públicas se basan en la accesibilidad y transparencia. Desde el punto de vista de la técnica los sistemas y algoritmos deben ser transparentes. Consideraos que, la base de todas las soluciones de tecnología para habilitar Sistemas Seguros en salud debe ser de código abierto.

Se aborda una línea de investigación propia y sumamente fértil, con un vasto potencial para la creación o adaptación de contratos inteligentes a partir de Solidity y otras librerías que permiten la gestión del consentimiento del paciente, con pleno resguardo de su identidad digital y de sus datos personales.

Todo ello es indispensable para la seguridad jurídica, condición necesaria para el futuro desarrollo de una blockchain de tercera generación que se adapte a los servicios públicos. La Metodología utilizada busca describir las condiciones del blockchain en los sistemas de salud, analizar las capacidades de administración de identidad del paciente en escenarios de blockchain y seleccionar mecanismos de creación automatizada de contratos para la salud a partir de patrones disponibles.

2 Blokchain en salud

Blockchain en el sector salud, significa una tecnología de confiabilidad distribuida de diferentes tipos. Mediante esta tecnología, se registra y protege los archivos mediante el uso de criptografía, resultando así extremadamente difícil que alguien perturbe o cambie los datos si no dispone de la aprobación de todos los participantes de la Red. Por lo tanto, la inmutabilidad es una de las características que permiten la creación de bases de datos incorruptibles para registros médicos.

Si bien el término se utiliza a menudo como una frase de alcance general, diferenciamos tecnologías de blockchain a lo largo de dos características principales del acceso de lectura público y privado y el permiso de escritura con permiso y sin permiso, resultando al menos cuatro infraestructuras de blockchain con diferentes implicaciones, legales, comerciales y dominios técnicos. Esta es una simplificación excesiva en mira a la eficiencia ya que hay otras permutaciones de permisos que también pueden lograr fines similares.

Los procesos negociales que tienen por objeto la prestación de bienes y servicios vinculados a la información automatizada del dato médico, basados en tecnología blockchain [1] es un concepto reciente y en constante crecimiento posee características

que para el usuario común son difíciles de comprender. Una solución blokchain podría abordar las reglas de privacidad que fija la legislación argentina, Ley 25.506 (Ley de Firma Digital) que autoriza la firma electrónica para separar y encriptar la identidad de los pacientes. Los pacientes también pueden compartir sus datos sin revelar su identidad.

El país europeo ejemplo en el desarrollo de las nuevas tecnologías es sin duda Estonia, el sistema sanitario mientras los registros de eHealth se encuentran en las tradicionales bases de datos asistenciales a las que se puede acceder desde el portal del paciente, el uso de blockchain realiza el servicio de firma verificable. De esta forma la Estonian eHealth Foundation está asegurando los registros médicos de más de 1 millón de pacientes. El sistema es totalmente transparente y cualquier paciente puede iniciar sesión y ver exactamente quién ha estado viendo sus registros e incluso puede restringir el acceso a grupos de usuarios. De igual forma el gobierno de Dubái también ha lanzado un proyecto para establecer una plataforma blockchain con tres finalidades; el ahorro en transacciones administrativas, el fortalecimiento de una industria multisectorial de negocios que usen la plataforma y el liderazgo internacional. En el ámbito sanitario creará una prueba concepto para comprobar las posibilidades de la integridad de datos, seguridad y confianza el intercambio de registros digitales entre los diversos proveedores de salud [2].

En el sector sanitario público de España esto tiene aplicaciones en la protección de los registros de acceso y modificación de datos sensibles tales como la información de historias clínicas o los datos de gestión de listas de espera quirúrgica, de técnicas diagnósticos y de consultas. Las aplicaciones más interesantes empiezan a surgir con la posibilidad de introducción de lógica de negocio sobre las redes blockchain (smart contract o contrato inteligente). Al ser esta lógica de negocio indeleble, auditable e independiente de una entidad central, esta puede complementar o sustituir la labor de un tercero de confianza [3]. En España aún están en fase de observación, la creación de la Red Lyra, como un consorcio multisectorial encabezado por grandes empresas, y varias iniciativas profesionales como Blockchain España o LibroBlockchain, nos acercan a las realidades del blockchain en el mundo de la salud, lógicamente nos encontramos con los retos, dificultades y barreras que ralentizan su implantación. En algunos lugares, como Reino Unido o Canadá, hay sistemas nacionales para intercambiar historias médicas, pero son vulnerables a los hacker.

En Estados Unidos y varios otros países la salud combina salud pública y privada, lo que significa que sus datos están aún más fragmentados, por lo que el sistema público puede tener data que tu médico privado desconoce, y a la inversa. Además, están tratando de combatir la corrupción mediante esta tecnología [4]. Si tomamos los EE.UU. como ejemplo, vemos que las compañías del sector de la salud interesadas en la adopción de la tecnología blockchain deben cumplir con normativas de datos vigentes, como por ejemplo la Health Insurance Portability and Accountability Act de 1996 (HIPAA). A grandes rasgos, la HIPAA describe estándares para el almacenaje, compartición y protección de datos en el sector de la salud. Así que para cumplir plenamente con la normativa, las compañías radicadas en los EE.UU, deberían desplegar sistemas de registros blockchain personalizados, con funciones que otorguen una mayor privacidad, así como una accesibilidad limitada.

En algunas clínicas y hospitales de Argentina se está utilizando la Historia Clínica Electrónica (HCE), para crear registro con datos médicos y el historial médico de los pacientes. La prestación de servicios en el ámbito de la salud pública implica procesos con actividades diferenciadas según la especialidad y procedimientos. Tiene componentes, entre los que se puede citar el profesional médico, el paciente (usuario), el servicio (atención en un periodo de tiempo), las instalaciones (consultorio o sala, equipamiento e instrumentos), las plataformas de comunicación (web institucional) y las infraestructuras de los sistemas de información sanitaria global.

Al igual que la trazabilidad es indispensable para la trasparencia de los sistemas de salud pública, la integridad de los datos generados por los distintos sistemas de salud llegan a ser inaccesibles para algunos agentes de la cadena de producción como son los responsables de la Salud Pública e incluso para el mismo paciente como dueño de la HCE [5]. La Ley 26.529 que regula la HCE en Argentina la conceptualiza en su Artículo 12, Como *"..., entiéndase por historia clínica, el documento obligatorio cronológico, foliado y completo en el que conste toda actuación realizada al paciente por profesionales y auxiliares de la salud".* A su vez, el Artículo 13 conceptualiza la Historia Clínica Electrónica, estableciendo que: "*El contenido de la historia clínica, puede confeccionarse en soporte magnético siempre que se arbitren todos los medios que aseguren la preservación de su integridad, autenticidad, inalterabilidad, perdurabilidad y recuperabilidad de los datos contenidos en la misma en tiempo y forma. A tal fin, debe adoptarse el uso de accesos restringidos con claves de identificación, medios no re-escribibles de almacenamiento, control de modificación de campos o cualquier otra técnica idónea para asegurar su integridad. La reglamentación establece la documentación respaldatoria que deberá conservarse y designa a los responsables que tendrán a su cargo la guarda de la misma".*

Si analizamos los distintos modelos de implementación de las Historias Clínicas digitales en nuestro país y en el mundo, los esfuerzos realizados para brindarles integralidad, accesibilidad e interoperabilidad de los sistemas de información en salud, vemos que todos ellos se topan con su validez legal. La Ley Nº 26.529 que rige los "Derechos del Paciente, Historia Clínica y Consentimiento Informado" y, en algunos aspectos, en el Código Civil y Comercial de la Nación en su articulado referido a los derechos Personalísimos.

En líneas generales hay que asumir que es una tecnología inmadura, que en muchos aspectos todavía está en modo experimental y que tiene que demostrar la robustez de sus mecanismos. En la XVI Reunión de Salud, se afirma qe la gestión de los datos es clave para la transformación digital de la Sanidad [6].

Blockchain se puede clasificar en dos grandes tipos [7]: público y cerrado (permissioned). Como podrán imaginar, uno público está abierto a todo el mundo y los participantes son anónimos. Cualquiera puede descargar blockchain público, acceder todo el histórico de transacciones y agregar una nueva transacción. Así funcionan las criptomonedas. En comparación, en el blockchain cerrado solo pueden participar un número limitado de participantes cuyas identidades son conocidas. A su vez, se puede distinguir al cerrado cómo privado y semiprivados; los privados son utilizados internamente por una sola organización, mientras que los semiprivados —también llamados federados— son usados por un grupo de organizaciones, como por ejemplo un consorcio o una red de proveedores en una cadena de suministro. La figura del blockchain federado es la que es de principal interés en el contexto empresarial, ya que es donde están los

casos de uso en los que mejor se puede aprovechar las características del blockchain. Actualmente existen en España por lo menos 40 consorcios distintos de blockchain federado, aunque todavía se encuentran en su inicio y hay bastante volatilidad entre ellos. BAES Blockchain Lab Interuniversitario, investiga y desarrolla soluciones tecnológicas eficientes, en distintos sectores públicos y privados, dando contenido a la capacidad de Blockchain para eliminar intermediarios, ideal para la eficiencia y sostenibilidad de los servicios públicos, al agilizar varios procesos técnicos, y la reducción en gran medida de los costos. Para ellos, blockchain es una herramienta estupenda para poner en valor el sistema sanitario, es decir, que permite aflorar valor que de otra forma quedaría en manos de intermediarios. En cuanto a los clínicos dice que es importante tener toda la información de sus pacientes en el mismo repositorio. Los pacientes también tienen sus demandas cada vez que se pide un análisis de sangre, médico y paciente tienen su identificador único. La orden del análisis, los participantes y la fecha y hora serían guardados como un bloque más en la cadena de ese paciente en particular. Al tener dependencias con los registros pasados y futuros, alterar una pieza de información sin ser detectado se vuelve imposible sin alterar todos los registros de la cadena y de los nodos de la red [8].

Investigadores del MIT Media Lab [9] están trabajando en una solución de este estilo y han hecho un estudio piloto usando información de las historias clínicas del Hospital Beth Israel. Los resultados son interesantes y se está planteando expandir las pruebas a una red mayor de hospitales. Blockchain propone redefinir cómo se organiza la información y nuestra idea sobre la confianza entre las partes. Sin embargo, su uso en el sistema sanitario depende de la creación de una infraestructura técnica adecuada. Hospitales, clínicas, organizaciones gubernamentales y demás organizaciones del sistema deben estar dispuestos a colaborar para crear, prototipar y probar los conceptos fundamentales que harán las bases para las historias clínicas del futuro. MedRec [10] proporciona a los pacientes un registro completo de sus registros de atención médica que se centra en: acceso fácil e intuitivo; compartir datos; y contenido verificable. A través de la gestión de permisos en una cadena de bloques Ethereum, habilitamos el intercambio de datos iniciado por el paciente entre jurisdicciones médicas. Para respetar la necesidad de confidencialidad a gran escala, MedRec permite autorizaciones a nivel de registros individuales e incluso campos de metadatos específicos.

Los contratos inteligentes de blockchain ofrecen una flexibilidad convincente, que permite la caducidad de los derechos de audiencia, actualizaciones de revocación, procedimientos de registro de identidad personalizados y más. El libro mayor de cadena de bloques de MedRec mantiene un historial inmutable y auditable de interacciones médicas para pacientes y proveedores.

3 Identidad digital

La identidad es el núcleo de todas las transacciones e interacciones entre personas físicas (por ejemplo, ciudadanos y consumidores), entidades legales (por ejemplo, empresas, organizaciones y gobiernos), y otros componentes (por ejemplo, dispositivos inteligentes o con inteligencia artificial). La gente tiene un rico tapiz de identidades formado por sus propias identidades personales (por ejemplo, padre, marido), sus identidades sociales (por ejemplo, empleado, jugador de fútbol) y sus identidades regidas por las

normas (por ejemplo, pasaporte, permiso de conducir). Típicamente, estas identidades no son estáticas sino dinámicas, evolucionan durante la vida de una persona y en diferentes contextos. Las Identidades emitidas por el estado a menudo son más estables y tienen la capacidad de identificar de manera única individual, lo que la convierte en la base de nuestras interacciones con el mundo exterior.

Las identidades digitales en forma de "cuentas en línea" eran la forma de acceder a los servicios en los años noventa. Uno podría crear múltiples personajes y construir una capa protectora de privacidad alrededor de uno mismo, siempre que la identidad emitida por el estado no haya sido divulgada o descubierto. En la década de 2010, los servicios escalables se hicieron cargo de la Internet con "inicio de sesión único" (SSO) o identidades federadas (lo que permite a los usuarios iniciar sesión en un sitio web que utiliza una cuenta existente de proveedores como Facebook, Google o Twitter). Este enfoque alivia la carga de usabilidad en los sitios web y mejora la conversión de visitantes en usuarios activos. Sin embargo, estas soluciones crean puntos únicos de falla y correlacionan la actividad del usuario a lo largo del tiempo y en diferentes contextos, lo que requiere que el usuario negocie su privacidad por conveniencia.

Al juntar todas estas piezas, obtenemos una imagen en la que la identidad de los usuarios de la empresa se distribuye a través de muchos sistemas, pero los proveedores de identidad centrales la imponen. Cuando se trata de identidad, hay dos fricciones fundamentales que deben abordarse entre la arquitectura actual de los sistemas empresariales y el universo de las tecnologías de cadena de bloques. Habilitar las capacidades de administración de identidad en escenarios de blockchain con permiso va más allá de un desafío técnico, crea una fricción con uno de los principios fundamentales de las capas descentralizadas.

La mayor contribución de las tecnologías de blockchain fue que, por primera vez en la historia de la informática, tenemos un modelo en el que podemos confiar en las matemáticas y la criptografía en lugar de las partes centralizadas. Al utilizar ese principio como base, las arquitecturas de blockchain evolucionaron en base a protocolos de consenso como prueba de trabajo (PoW) o prueba de estaca (PoS) que se basan en cálculos para tomar decisiones. En ése mundo descentralizado, la identidad no ha sido un componente fundamental, ya que la dinámica de la red debería llegar a un proceso óptimo de toma de decisiones. Los modelos de consenso basados en cómputo de las pilas de blockchain contrastan fundamentalmente con las soluciones empresariales en las que las identidades de los participantes son bien conocidas. En ése sentido, puede argumentar que los protocolos de consenso ofrecen muy poco en un mundo de identidades conocidas. La tecnología blockchain puede servir cómo una solución de gestión de identidad, es importante comprender primero, a un nivel básico, cómo se han manifestado las debilidades en el sistema actual.

Algunos ven una solución futura prometedora para éste problema en el auge de la tecnología blockchain. Como lo explica un artículo reciente de Toptal Insights [7], la tecnología blockchain se refiere a un libro de información inmutable de igual a igual, distribuido de igual a igual. Cada "bloque" de información contiene un registro completo y preciso de cada transacción, que no se puede modificar una vez verificada y se asegura de forma criptográfica. Una característica definitoria de esta tecnología es su estructura distribuida de igual a igual, que teóricamente evita la necesidad de intermediarios cómo Yahoo o Equifax para almacenar datos.

Blockstack y Civic ofrecen soluciones de administración de identidad basadas en el uso de la tecnología blockchain para descentralizar y proteger los datos personales. Ambos también usan tokens (sus propias criptomonedas) para alimentar e incentivar la utilización de sus plataformas. En el caso de Blockstack, los desarrolladores consumen tokens al crear aplicaciones en la red de Blockstack, y los usuarios consumen tokens al registrar los nombres de usuario de Blockstack. Para Civic, los usuarios pueden recibir tokens para utilizar la aplicación de Civic.

La identidad soberana es el siguiente paso más allá de la identidad centrada en el usuario y eso significa que comienza en el mismo lugar: el usuario debe ser central para la administración de la identidad. Eso no requiere solo la interoperabilidad de la identidad de un usuario en múltiples ubicaciones, con el consentimiento del usuario, pero también el verdadero control del usuario de esa identidad digital, creando la autonomía del usuario. Para lograr esto, una la identidad soberana debe ser transportable; no se puede bloquear en un sitio o ubicación. La identidad es uno de los bloques de construcción fundamentales de las aplicaciones de blockchain permitidas y debe abordarse para permitir la adopción generalizada de estas tecnologías. Esfuerzos cómo la fundación de identidad descentralizada (DIF) [11] están liderando el camino para cerrar la brecha entre los sistemas de identidad tradicionales y el nuevo mundo de blockchains. Aunque hay algunos protocolos y herramientas en el espacio, habilitar las capacidades de identidad en soluciones de cadena de bloques empresarial sigue siendo un esfuerzo bastante complejo [12].

Con respecto a la protección de identidad, se ha argumentado que la tecnología de cadena de bloques puede eliminar la necesidad de intermediarios y permitir a los individuos el control total sobre sus identidades digitales. Además de simplificar el proceso de compartición de registros, los sistemas blockchain pueden también conferir a los pacientes un mayor nivel de accesibilidad y transparencia respecto a su propia información sanitaria. En algunas circunstancias, requerir la validación de las modificaciones efectuadas sobre documentos de pacientes, puede asegurar la exactitud de los registros; y si se realiza de forma adecuada, este tipo de verificación puede proporcionar una capa de seguridad extra que proteja contra errores humanos o falsificaciones deliberadas.

Nuestra Ley de protección de datos personales (Ley 25.326) establece distintos niveles de seguridad para el tratamiento de datos, las empresas de salud pueden procesar datos personales, usando la tecnología de cadena de bloques para acceder y verificar estos datos, mediante la designación de un responsable del tratamiento de los datos de salud.

4 Gestión de la Privacidad de los datos del paciente

Los datos de los pacientes se pueden almacenar en la blockchain y solo se pueden mantener accesibles para el paciente y el médico que está a cargo del caso. El acceso puede ser revocado en cualquier momento, lo que garantiza que el paciente tenga plena autoridad en sus informes médicos. Eso no significa que todos los datos sean inaccesibles. Los datos insensibles están disponibles públicamente, a los que puede acceder la organización de atención médica o las partes interesadas a través de la API. Esto asegura la colaboración adecuada a través de diferentes sistemas.

En su Artículo 14, nuestro CCC, dispone que "*El paciente es el titular de la historia clínica. A su simple requerimiento debe suministrársele copia de la misma, autenticada*

por autoridad competente de la institución asistencial". En el mismo artículo establece el plazo máximo de entrega de la misma, *"La entrega se realizará dentro de las cuarenta y ocho (48) horas de solicitada, salvo caso de emergencia".* Ahora bien, el Artículo 15, especifica qué debe contener en calidad de información, Además establece responsabilidades, a saber: "*Inviolabilidad. Depositarios. La historia clínica es inviolable. Los establecimientos asistenciales públicos o privados y los profesionales de la salud, en su calidad de titulares de consultorios privados, tienen a su cargo su guarda y custodia, asumiendo el carácter de depositarios de aquélla, y debiendo instrumentar los medios y recursos necesarios a fin de evitar el acceso a la información contenida en ella por personas no autorizadas. A los depositarios les son extensivas y aplicables las disposiciones que en materia contractual se establecen en el Libro II, Sección III, del Título XV del Código Civil, "Del depósito", y normas concordantes*.

En su versión más sencilla, la tecnología blockchain permite la grabación de un registro de transacciones con máximas garantías de integridad de tal manera que ni siquiera el administrador del sistema puede alterarlo. La tecnología blockchain aporta los estándares más altos hasta ahora conocidos en materia de seguridad respecto a la inmodificabilidad de su contenido. La criptografía asimétrica resulta fundamental para evitar la manipulación, hurto o introducción errónea de infracciones en la cadena de bloques. La clave privada se utiliza para desencriptar los datos y posibilita la firma para la realización de transacciones en la red. Blockchain permite garantizar la integridad e inmutabilidad de la información que registra, así como incorporar aquella que proviene de operadores de confianza y que es validada por los miembros de la red. Sin embargo, la utilización de esta tecnología no permite asegurar la veracidad de la información así adquirida, lo que abre las puertas a que pudieran existir errores de registro de información en la cadena de bloques. .

5 Creación de contratos con validez legal

Nuestro código Civil y Comercial de la Nación (CCyCN) en su art. 971 establece: *"Los contra-tos se concluyen con la recepción de la aceptación de una oferta o por una conducta de las partes que sea suficiente para demostrar la existencia de un acuerdo"*.
Respecto de los servicios de Salud la misma norma establece que respecto de los contratos celebra-dos a distancia que serán aquellos concluidos entre un proveedor y un consumidor con el uso exclusivo de medios de comunicación a distancia, entendiéndose por tales los que pueden ser utilizados sin la presencia física simultánea. En especial, se considerarán los medios electrónicos de las partes contratantes.
El deber de seguridad que se funda en el deber de buena fé art. 9 y 961 del CCy CN. Se encuentra regulado en el Art.5 de la Ley de defensa del consumidor que dice*: "Las cosas y servicios deben ser suministrados o prestados en forma tál que, utilizados en condiciones previsibles o normales de uso, no presenten peligro alguno para la salud o integridad física de los consumidores o usuarios".* El uso de tecnología blockchain y los contratos inteligentes que esta tecnología permite utilizar, tiene el potencial de generar: desarrollos, aplicaciones y soluciones en diversas áreas y servicios para la atención sanitaria. Ello es fundamental, porque además de la descentralización, el papel que jugará el contrato inteligente (smart contract), que básicamente es un compromiso entre distintas partes que obliga a los contrayentes y cuyo cumplimiento es exigible. Con las

salvedades de que éste contrato se auto ejecuta, es digital y no requiere de un tercero para validarse. El hecho de que el protocolo de blockchain utilice clave pública y privada consigue que la seguridad en las transacciones y la confidencialidad resulten máximas. También se pueden establecer cadenas públicas abiertas, compartidas o privadas en las que la participación se limita solo para entidades previamente aceptadas. Como registro para los datos almacenados externamente, como un índice que indica dónde está la información. Pero también es utilizado cómo huella digital o marca de tiempo de cada acción pudiendo saber quién hizo qué y cuándo.
En el caso del sector sanitario el contrato inteligente podría enviar facturas a pacientes o solicitar suministros médicos, evitando intermediarios. Los smart contract deben ser neutrales y conocidos, transparentes y nunca dar lugar a diferentes interpretaciones. Solidity ofrece un mecanismo de creación automatizada de contratos para la salud a partir de patrones disponibles. Solidity cómo lenguaje de programación [13], es el lenguaje más popular de contractos inteligentes ("smart contracts") en Ethereum. Lenguaje tipado (cada variable, argumento, función tiene un tipo asociado), con una sintaxis que recuerda a JavaScript y a otros lenguajes. Tiene el concepto de contrato incorporado, similar a las clases en otros lenguajes cómo Java y C#. Existe un compilador de línea de comando, y también es popular un módulo JavaScript que se puede usar desde NodeJS. El compilador produce bytecodes de la máquina virtual Ethereum. Solidity está tipado de manera estática y acepta, entre otras cosas, herencias, librerías y tipos complejos definidos por el usuario [14]. De esta manera también con el uso de esta tecnología se contribuirá a la identificación de soluciones coste efectivas, factibles y sostenibles para lograr una atención más segura del paciente [15].

6 Conclusiones

Este trabajo aborda los fundamentos propios que caracterizan al blockchain: la trasparencia, la capacidad de descentralizar el consenso, su registro inmutable y la confianza entre los actores desconocidos dentro de la red. Lo consideramos cómo una opción óptima para proporcionar una la gestión eficiente de los datos, de los pacientes que sería accesible a todos los profesionales intervinientes con total trasparencia no solo a nivel local sino también a nivel mundial.
Permite un nuevo modelo para manejar datos de pacientes, en un registro distribuido que se distribuye a través de la red, manteniendo la integridad a través de la criptografía. Ello posibilita la inmutabilidad del dato que puede ser leído y compartido más de una vez manteniendo la inmutabilidad, para lograr su trazabilidad conociendo en todo momento que ha ocurrido con cada agente de la cadena. Generando la trazabilidad de la información generada por las organizaciones públicas y privadas que podrán integrar la información generada en una HCE, que brindará total confianza al paciente cómo verdadero dueño de la información.
Con el aporte de la tecnológica se pueden acoplar nuevas posibilidades para la seguridad del dato cómo son las claves privadas, las claves inteligentes, los contratos inteligentes permiten desarrollar aplicaciones seguras. A pesar de que existen ciertos desafíos técnicos, logísticos y regulatorios, la implementación de estos sistemas jugará, con toda probabilidad, un importante rol en el futuro del almacenamiento y transmisión de los datos médicos.

7 Referencias

1. https://www.sadamweb.com.ar/news/2018_05Mayo/Guia-Blockchain-para-el-sector-de-la-salud-Curaesalud.pdf
2. https://2019.egovconference.ee/wp-content/uploads/sites/2/2019/02/Peeter-Ross-eHealth-in-Estonia.pdf
3. https://blockchainhub.one/en/http://www.intedya.com/internacional/1765/noticia-varios-paises-estan-estudiando-minuciosamente-la-tecnologia-de-blockchain-para-combatir-la-cor-rupcion.html
4. https://www.interactivity.la/2018/09/24/blockchain-y-salud/
5. www.perfil.com/noticias/ciencia/la-tecnologia-blockchain-dominara-2018-y-promete-una-revolucion-como-la-de-internet.phtml?fb_comment_id=1407690476024966_2005651092895565
6. https://issuu.com/editorialmic/docs/i_mas_s_133_febrero_2019
7. https://www.toptal.com/insights/innovation/blockchain-identity-management
8. https://ehealthreporter.com/es/noticia/blockchain-una-tecnologia-que-revolucionara-la-salud/
9. https://viral.media.mit.edu/pub/medrec
10. https://www.media.mit.edu/about/mission-history/
11. https://identity.foundation/
12. https://hackernoon.com/identity-the-elephant-in-the-enterprise-blockchain-room-6f31ed8d4132
13. https://github.com/ajlopez/AprendiendoSolidity
14. https://solidity-es.readthedocs.io/es/latest/
15. https://www.seguridaddelpaciente.es/resources/documentos/2015/Estrategia%20Seguridad%20del%20Paciente%202015-2020.pdf

Sistema de información Farmacéutica para la dispensación de Medicamentos e Insumos médicos en el Hospital Escuela y de Clínicas Virgen María de Fátima

Leonardo Garay[1], Enrique Mamani[1], Gabriel Pacheco[1], Ariel Rivadulla[1], Soledad Ciccarelli[1]

[1]Universidad Nacional de la Rioja. Capital. Argentina.
leo.garay9@gmail.com, mamanienrique77@gmail.com, alejandrocabj94@gmail.com, airguana@gmail.com, soledadcicarelli@yahoo.com.ar

Resumen. La tarea de dispensación de medicamentos es tradicionalmente realizada por el farmacéutico. El proceso de dispensación se define como el suministro o la entrega de uno o Más medicamentos e insumos médicos a un paciente y la respectiva información sobre su uso adecuado [1]. En el presente trabajo se desarrolla una alternativa de automatización para la Farmacia del Hospital Escuela y de Clínicas de la Universidad Nacional de la Rioja. El nuevo sistema (G-Farm) implementado permite al Servicio de farmacia tener el control de la distribución de medicamentos e Insumos médicos en forma eficiente, reduciendo los tiempos empleados de entrega por el personal de salud del nosocomio y llevando un seguimiento de los diversos productos.

Palabras claves: Servicio de farmacia hospitalaria. Dispensación. Medicamentos e Insumos médicos.

1- Introducción

A lo largo de la historia se ha demostrado que una de los principales recursos para combatir las enfermedades es el empleo de medicamentos, a través del proceso denominado farmacoterapia. De forma general, la utilización de fármacos comienza con el diagnóstico de la enfermedad del paciente y en base a esta, se establece un tratamiento farmacológico [2].

Los medicamentos son una parte importante de la medicina moderna y un producto esencial para mantener la vida, evitar el sufrimiento y prevenir enfermedades. El medicamento, como bien de consumo, tiene una característica única que lo distingue de los demás bienes: lo toma un paciente, lo prescribe un médico, lo dispensa un farmacéutico, lo administra un enfermero/a, en algunas ocasiones lo paga el sistema público de salud y en otras el usuario, pero no quien decide la prescripción.[3]

En farmacia: La dispensación (farmacia) es el acto en que el farmacéutico entrega la medicación prescrita por el médico al paciente, junto a la información necesaria para su uso racional. [4]

A medida que el Hospital Escuela y de Clínicas Virgen María de Fátima, perteneciente a la Universidad Nacional de la Rioja fue creciendo, el sistema de gestión de medicamentos implementado en el área farmacéutica fue quedando obsoleto y cada vez era más difícil; organizar las tareas de distribución que se realiza

en la farmacia, debido a que poseían un sistema enlatado. Luego de varios intentos fallidos, y por la complejidad requerida para la correcta dispensación se optó por el desarrollo propio por parte del área de Informática de la UNLaR.
La dispensación de medicamentos en las farmacias se realiza con el fin de racionalizar la distribución de medicamentos, disminuir los errores de medicación, garantizar el cumplimiento de la orden médica, dar un seguimiento a los tratamientos farmacológicos, brindar información sobre la correcta administración de los fármacos, así como también para fortalecer el papel del farmacéutico en el equipo asistencial y reducir el costo de medicación. [5,6]
El desarrollo de esta aplicación (G-Farma) está orientada a facilitar la dispensación y control de los medicamentos en el servicio farmacéutico del Hospital Escuela y de Clínicas "Virgen María de Fátima" normalizando los procedimientos y asegurando la calidad de la información registrada disponiendo de ella en cualquier momento, especialmente en momentos críticos donde la información y los datos de un medicamento facilitando la labor de las personas involucradas en el proceso.
La calidad de la dispensación de los medicamentos en las farmacias públicas, es un tema que debe ser evaluado de forma continua y detallada por el personal farmacéutico para su propio beneficio y el bienestar del paciente. Proceso que en ocasiones no recibe la suficiente importancia, por parte del sector salud en general.[7]

2- Línea de Investigación

La línea de investigación para el presente proyecto se encuentra respaldada bajo la Resolución CICYT N°044/11 del Consejo de Ciencia y Técnica de la Universidad Nacional de La Rioja, derivando para su aplicación las áreas de Ingeniería de Software, Base de Datos y Minería de Datos.

3- Modalidad

De acuerdo con la modalidad establecida en el Reglamento de Trabajo Final de la Carrera, el desarrollo se considera como un **Trabajo de extensión a la comunidad hospitalaria.**

4- Metodología

Para el desarrollo del presente trabajo se optó por la Extreme Programming (Programación Extrema). Esta contiene en sí planificación, diseño, desarrollo y pruebas como fases. El principal beneficio de esta metodología es la menor tasa de errores en el código. Permite tener un contacto directo con el usuario final del sistema en forma permanente, creando así las historias de usuarios que son de interés para la primera entrega del producto, favoreciendo el feedback con los desarrolladores y minimizando los tiempos de desarrollo del sistema.
El análisis funcional y el diseño de la aplicación ha sido realizada por el grupo de 3 tesistas en su etapa de finalización de carrera de grado para optar por el título de Licenciados en Sistemas de Información, con la participación del director del proyecto y el asesoramiento técnico de farmacéuticos con experiencia suficiente en investigación clínica en el hospital.

Luego de recopilar la información referente al sistema, se desarrolló una aplicación web que se pudiera integrar al entorno tecnológico actual, aprovechando así la actual infraestructura de servidores.
Las características del entorno tecnológico, y las herramientas a utilizar para el desarrollo son las siguientes:

- Arquitectura: Servidor Web.
- Tecnologías de desarrollo: Framework SIU-TOBA 3.2, Docker Compose, Git (Bitbucket)
- Servidor: Debían 9
- DBMS: PostgreSQL 9.6
- Diagramas, modelos y documentación: SQL Architect
- Gestión del Proyecto: Gantt Project. Trello
- Diseño de logotipo y edición fotográfica: GIMP 2
- Diseño de Manuales de Usuario: HelpNDoc
- Comunicación remota entre integrantes del proyecto: Facebook, Google+ Hangouts.

Los motivos que impulsaron esta decisión fueron:

- El aprovechamiento de los recursos existentes, evitando la necesidad de adquirir
- equipamiento nuevo o desarrollar infraestructura.
- Licencias Open Source del software que se utilizaron.

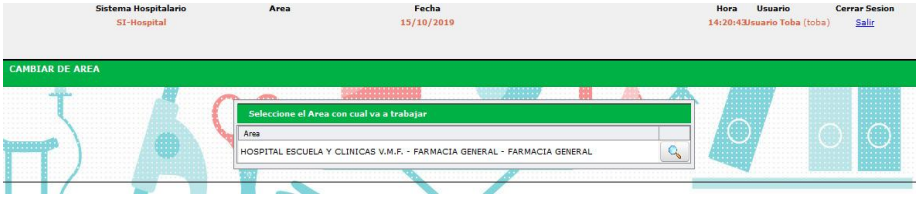


Fig. 1 Pantalla principal

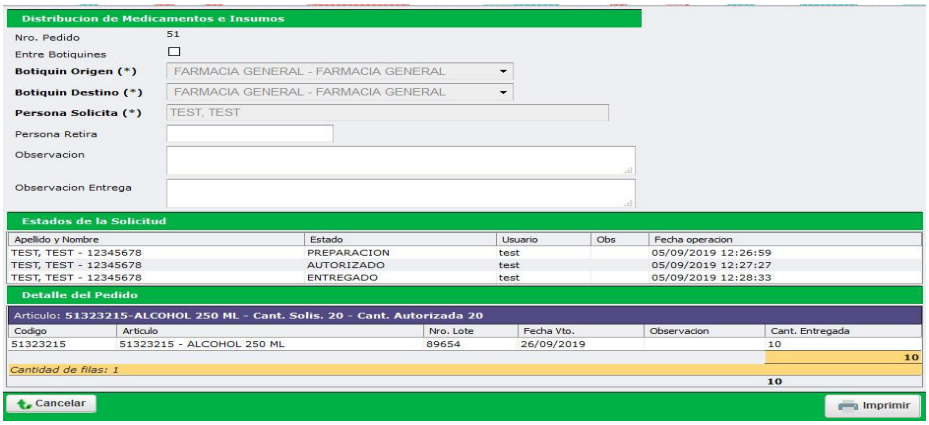


Fig 2:

Conclusiones y Trabajos a futuro

Trabajar con la metodología XP resultó ser un proceso de aprendizaje constante y un incremento a nuestro conocimiento que valoramos a la hora de desarrollar el proyecto. El producto concretado y puesto en marcha, es totalmente funcional responde a las expectativas de los usuarios de la Farmacia del Hospital Escuela, y nos señala que los objetivos propuestos en un principio se cumplieron. El sistema presenta simplicidad, flexibilidad e interfaces amigables, haciendo de la aplicación una herramienta útil en el proceso de dispensación..

Como conclusión queremos exponer que el hecho de ir haciendo entregas constantes y frecuentes a los usuarios, no sólo lo mantiene satisfecho, sino que también responde a la urgencia de este por obtener el producto esperado en periodos breves de tiempo. Además, se deja sentado con la documentación pertinente, la posibilidad de seguir ampliando el sistema con más módulos, a fin de desarrollar un Sistema de Información Hospitalaria que contribuya a las áreas administrativas y a la referida a la atención de los pacientes dispuestos por la instrucción para la atención hospitalaria.

Referencias

1. Sami Pérez Gomes. Buenas practicas de expendio. https://es.slideshare.net/samiperezgomes/buenas-practicas-de-expendio.
2. María José FAUS DADER, Francisco MARTÍNEZ ROMERO, La Atención Farmacéutica en farmacia comunitaria: evolución de conceptos, necesidades de formación, modalidades y estrategias para su puesta en marcha. 56 ◆Pharm Care Esp 1999; 1: 52-61.
3. Gustavo H. MARÍN *1,2, Martín CAÑÁS 2,3, Cecilia HOMAR 1,2 & Mariel PERROTTA- Utilización de Medicamentos del Programa REMEDIAR en la Provincia de Buenos Aires. Argentina - Latin American Journal of Pharmacy (formerly Acta Farmacéutica Bonaerense) Lat. Am. J. Pharm. 27 (4): 535-42 (2008).
4. https://es.m.wikipedia.org/wiki/Dispensaci%C3%B3n
5. Br. Millón Murillo Eduardo René, Br. Silva Hernández Wagner Javier, Br. Urbina Pedro Javier. Evaluación del proceso de dispensación de medicamentos en farmacia del Puesto de Salud El Calvarito, Ciudad de León, periodo Marzo-Junio 2018. Monografía para optar al título deLicenciado Químico Farmacéutico. **León, Nicaragua 2018**
6. Pérez Ruixo, J.J, Pastor Boix, E, Juan Colomer, J., Jiménez Torres, N.V, ANÁLISIS DE UN SISTEMA AUTOMATIZADO DE DISPENSACIÓN INDIVIDUALIZADA DE MEDICAMENTOS EN UNA UNIDAD DE CUIDADOS INTENSIVOS. Servicio de Farmacia. Hospital Universitario Dr. Peset. Valencia. Departamento de Farmacia y Tecnología Farmacéutica. Universidad de Valencia. Farm Hosp 1998; 22 (2): 81-87.
7. Idoate García, A.*, Giráldez Deiró, J.*, Idoipe Tomás, A.**, "TRIALS® : aplicación Informática para dispensación y control de medicamentos en fase de Investigación clínica" Año: Farm Hosp 1995; 19 (1): 24-30

PROPUESTA DE ANALISIS DESCRIPTIVO Y PRESCRIPTIVO EN ALUMNOS ASPIRANTES E INGRESANTES DE LA LICENCIATURA EN CIENCIAS DE LA COMPUTACIÓN UNIVERSIDAD NACIONAL DE CUYO

Marisa Haderne[1], María Alejandra Marín[2], Cristian Gamba[1], Raquel Zumel[1], Rosana Gimenez[2], Eleonora Luconi[1], Eleonora Valdivieso[1]

[1] Facultad de Ingeniería. Universidad Nacional de Cuyo.

[2] Facultad de Ciencias Económicas. Universidad Nacional de Cuyo

marisa.hadernea@ingenieria.uncuyo.edu.ar, alejandra.marin@fce.uncu.edu.ar, cristian.gamba@ingenieria.uncuyo.edu.ar, raquel.zumel@ingenieria.uncuyo.edu.ar, gimenezbula@hotmail.com, eloluconi@gmail.com, evaldivieso@itu.uncu.edu.ar

RESUMEN

En el sistema de educación superior, el ingreso y la deserción de los estudiantes son temas centrales en la definición de políticas universitarias. Dichas políticas promueven principios de equidad social y de oportunidades educativas, es decir, tratan que todas las personas logren iguales niveles de competencia, otorgando las mismas oportunidades de acceso y permanencia en las instituciones educativas. El problema planteado en el presente trabajo es relevante y de especial interés para la Facultad de Ingeniería de la Universidad Nacional de Cuyo. Por ello, resulta fundamental conocer al aspirante de la Licenciatura en Ciencias de la Computación, su problemática académica, trayectos educativos anteriores y otros componentes sociales que pueden afectar su rendimiento al momento del ingreso.

Se espera, mediante diferentes técnicas de minería de datos aplicadas a factores académicos, sociales y antecedentes de escolarización, caracterizar el comportamiento del aspirante a la Licenciatura en Ciencias de la Computación.

También se realizará un análisis de la articulación entre el colegio secundario y la universidad, su problemática y falencias y las expectativas de los docentes de primer año. Con los resultados obtenidos se espera rediseñar el curso de nivelación.

Palabras clave: 1) Ingreso 2) Minería de Datos Educacional 3) Rendimiento de los estudiantes

CONTEXTO

El presente trabajo es una introducción al proyecto bienal 2019-2020 aprobado por la Secretaría de Ciencia, Técnica y Posgrado de la Universidad Nacional de Cuyo. Se continuará con los avances del proyecto "SIPREU: Un sistema inteligente para caracterizar y predecir el rendimiento de estudiantes universitarios" [8] que comenzó en el año 2018.

Además de la relevancia general y local del problema corresponde analizar la importancia disciplinar del tema. La minería de datos educacional es una línea que ha cobrado mucha importancia mundial en los últimos años, como se observa en el portal dedicado al tema que lleva adelante la International Educational Data Minining Society [2].

INTRODUCCIÓN

Estamos ante una realidad en educación que impone cambios profundos. Según Norberto Fernández Lamarra, en su artículo "La Universidad en América Latina y Argentina: problemas y desafíos políticos y de gestión para la educación y la universidad" [10] explica que "el conocimiento ha dejado de ser lento, escaso y estable; la escuela y la universidad ya no son monopólicas en cuanto al conocimiento; el profesor y el texto han dejado de ser soportes exclusivos; las tecnologías de enseñanza tradicionales ya no son las únicas disponibles; las competencias, el aprendizaje y los tipos de inteligencia tradicionales están perdiendo pertinencia y validez". Así mismo presenta un disparador que genera la apertura al debate de repensar las acciones y compromiso en la Universidad al sostener que "la evolución de la educación superior en la última década ha intensificado su carácter elitista". Lamarra remarca que "el acceso a las instituciones de educación superior, públicas y privadas

de mejor calidad ha quedado reservado casi exclusivamente a los jóvenes de clase media y media alta, provenientes de escuelas medias públicas y privadas de buena calidad".

Uno de los fines de la política educativa en Argentina es asegurar igualdad de oportunidades en el acceso a la educación universitaria. En el plan estratégico de la Universidad Nacional de Cuyo 2021 [15] uno de sus objetivos incluye "Responder a la creciente demanda de educación superior en todos sus niveles, asegurando gratuidad e inclusión con calidad y pertinencia, y promoviendo una formación integral y de excelencia.". Para ello en su línea estratégica indica el "Desarrollo de mecanismos permanentes para eliminar las brechas sociales, culturales y educativas de los estudiantes preuniversitarios y universitarios" y "Fortalecimiento de las políticas de ingreso, permanencia y egreso de los estudiantes"

Asegurar la igualdad de oportunidades se puede traducir entre otros aspectos en garantizar el desarrollo de las mismas competencias en todas las personas. Esto implica oportunidad de acceso a las instituciones educativas y de permanencia en ellas.

La licenciatura en ciencias de la computación es reciente, comenzó en el año 2017 y ya cuenta con 3 cohortes y en este momento se está desarrollando la cuarta edición del curso de nivelación.

Como se puede observar en el cuadro 1, los porcentajes de ingreso por examen son bajos. A esta problemática se suma el rendimiento de primer año, donde más del 50% de los alumnos presenta problemas con las asignaturas relacionadas con las ciencias básicas (análisis matemático I, análisis matemático II, geometría analítica y álgebra).

Año de ingreso	Porcentaje de ingreso por examen
2017	13%
2018	29%
2019	23%

Cuadro 1: comparativa de ingreso

El problema planteado en este proyecto es relevante desde varios puntos de vista. En primer lugar, desde una valoración general, el acceso, la retención y el desgranamiento de los alumnos universitarios en nuestro país posee gran relevancia si se tiene en cuenta que de la Universidad egresa menos de un 30% de los alumnos ingresantes.

Entre las numerosas causas que generan el problema citado, este proyecto apunta a características y comportamientos propios en primer lugar del aspirante universitario. Entre otros factores que influyen en el rendimiento de éstos podrían citarse: situación socioeconómica, condición cultural, institución en la cual cursó estudios medios, rendimiento en los exámenes de ingreso, motivación personal, comportamiento actitudinal, entre otros. A continuación, se espera poder aplicar técnicas de minería de datos y otras herramientas de inteligencia artificial al seguimiento de los alumnos que ingresaron y cuál es la relación entre los resultados del ingreso y su rendimiento académico durante el cursado. La minería de datos educacional es un área de investigación interdisciplinar que trata con el desarrollo de métodos que exploran datos originados en el sector educativo y su contexto.

Este análisis es fundamental al plantear la necesidad de establecer objetivos y temas a incorporar en el ingreso y su impacto al momento de cursar las primeras asignaturas.

Esta propuesta se diferencia de otras dado que incluye a los aspirantes a ingresar a la facultad, por lo tanto, uno de sus objetivos sería relacionar el curso de nivelación dado por la facultad y su relación con los resultados académicos del primer año y la propuesta de modificación del actual sistema de ingreso.

Que ingrese a la licenciatura menos del 30% de los aspirantes a la carrera va en detrimento de la igualdad en oportunidad de acceso a la institución, y que menos de la mitad de alumnos logra cursar la totalidad de las asignaturas del segundo semestre muestra un problema difícil de ignorar.

Ante esta problemática el presente proyecto destaca la importancia del entendimiento del aspirante desde el momento que se inscribe en el curso de nivelación de la carrera, hasta que culmina el primer año de la misma y de la institución educativa que gestiona este proceso. Para analizar este trayecto se considera oportuno relevar los colegios donde el aspirante cursó el nivel secundario, identificar las capacidades que posee y detectar cuales son las expectativas de los docentes y la institución en cuanto a los conocimientos requeridos.

Con esta información cualitativa y cuantitativa se espera en primer lugar un replanteo curricular del curso de nivelación de la licenciatura y además el proyecto apunta a caracterizar el comportamiento de los aspirantes

mediante técnicas de minería de datos. Para ello entre las variables que se tendrán en cuenta están: su entorno social, económico, ambiental y cultural, registro de competencias, resultados del curso de nivelación, rendimiento académico de los ingresantes en primer año, entre otras.

La importancia de este proyecto es que analiza al sujeto desde el nivel secundario, pasando por el curso de nivelación hasta el cursado del primer año de la carrera, detectando las necesidades de todos los actores involucrados en este proceso: aspirantes, alumnos, docentes de primer año, dirección de ingreso, y de carrera y equipo de gestión entre otros.

HIPÓTESIS DE TRABAJO

Es posible realizar un replanteo del curso de nivelación a partir del análisis de los contenidos adquiridos en los colegios secundarios o experiencias anteriores de los aspirantes y la detección de expectativas de los profesores de primer año de los ingresantes.

Además, se puede caracterizar el rendimiento de los ingresantes mediante técnicas de Minería de Datos a partir del registro de diferentes variables de los aspirantes y su continuidad en el primer año. A partir de dicha caracterización se puede postular que los aspirantes que presenten fuertes similitudes con la población caracterizada tendrán rendimientos similares. Para ello, en base a los atributos considerados de interés se construirá un data set que se empleará como punto de partida para contar con datos de interés para el sistema propuesto.

RESULTADOS ESPERADOS

Al finalizar el trabajo se podrá obtener una caracterización del comportamiento de los aspirantes y estudiantes universitarios del primer año y utilizar la misma como herramienta predictiva de las trayectorias académicas a lo largo de las mismas, y poder colaborar con los actores principales de la Facultad que trabajan junto con los alumnos para poder disminuir la actual deserción.

Se espera poder completar un diagnóstico de la problemática de rendimiento académico de los alumnos de la carrera, así como el diseño de herramientas que faciliten la gestión de datos académicos y permitan la detección temprana de alumnos de alto riesgo.

Muchos de los análisis y estudios plasmados en las herramientas correspondientes podrían extenderse con relativa simplicidad a otras Unidades Académicas de la UNCuyo.

BIBLIOGRAFÍA

1. Araujo, S (Araujo 2017) “Entre el ingreso y la graduación: el problema de la democratización en la universidad”. Revista de Educación, núm. 27, junio 2017, pp. 35-61. Universidad Nacional del Centro de la Provincia de Buenos Aires, Buenos Aires, Argentina

2. Educational Data Mining Web Site: http://www.educationaldatamining.org/

3. A.M. Ezcurra (Ezcurra 2004). Diagnóstico preliminar de las dificultades de los alumnos de primer ingreso a la educación superior. Perfiles educativos, vol. XXVII, núm. 107, pp. 118-133

4. A.M. Ezcurra (2011) Igualdad en la educación superior. Un desafío mundial. IEC-UNGS, Los Polvorines, Argentina.

5. Feierherd, Guillermo & Depetris, Beatriz & Jerez, Marcela. (Feierherd 2019). Una evaluación sobre la incorporación temprana de algorítmica y programación en el ingreso a Informática.

6. D. Fernández (Fernández 2011): Proyecto 06B/246 Detección, identificación y análisis de los factores fundamentales, externos e internos, que influyen y definen el ingreso de los aspirantes a Ingeniería y Arquitectura de la Universidad Nacional de Cuyo. Proyecto bienal, Universidad Nacional de Cuyo, 2011-13

7. S. Formia, L. Lanzarini, y W. Hasperué (Formia 2013): Caracterización de la deserción universitaria en la UNRN utilizando Minería de Datos. Un caso de estudio, pp. 92-98, N°11, Diciembre de 2013, Revista Iberoamericana de Educación en Tecnología y Tecnología en Educación. ISSN 1850-9959

8. C.García Garino, C.Catania, M.Haderne, E.Pacini, L.Cortés, S.Raichman, L.Brottier, E.Valdivieso, Y.Garí Nuñez, J.Guerra Torres (Garino 2019) proyecto del SECTYP 2017-2019: SIPREU: Un sistema inteligente para caracterizar y predecir el rendimiento de estudiantes universitarios

9. S. Hernández, Montano A. Gillet, N. (Hernandez, 2016) Reflexiones acerca de la articulación entre la escuela secundaria y la universidad: la nivelación 2015 en la disciplina química. Mayo 2016. V jornadas Nacionales y I latinoamericanas de ingreso y permanencia en carreras científico-tecnológicas

10.Norberto Fernandez Lamarra (Lamarra 2007) La universidad en America Latina y Argentina: problemas y desafios políticos y de gestión

11.L. Lanzarini, E. Charnelli, G. Baldino y J. Díaz (Lanzarini 2015): Selección de atributos representativos del avance académico de los alumnos universitarios usando técnicas de visualización. Un caso de estudio. pp. 42-50, N°15, junio de 2015, Revista Iberoamericana de Educación en Tecnología y Tecnología en Educación. ISSN 1850-9959

12.C.Pérez (Perez 2013) El buen rendimiento escolar en los estudiantes que ingresan a la universidad a través del programa propedéutico: un análisis desde la motivación y el discurso de la UCSH. Calidad en la educación n. 38, julio 2013 pp. 147-179

13.M Pochulu (Pochulu 2009) análisis y categorización de errores en el aprendizaje de la matemática en alumnos que ingresan a la universidad. Revista Iberoamericana de Educación ISSN: 1681-5653

14.Segarra, C., Churio, S. y Bordino, N, (Segarra 2016) El desafío del ingreso y la permanencia a través de la Articulación entre cursos de nivelación Y asignaturas de primer año. Mayo 2016. V jornadas Nacionales y I latinoamericanas de ingreso y permanencia en carreras científico-tecnológicas

15.Universidad Nacional de Cuyo (UNCuyo 2014): Plan estratégico 2021. Disponible en: http://www.uncu.edu.ar/planificacion/upload/plan-estrategico-1.pdf

16.A.Vogliotti, (Vogliotti. 2016): Programa institucional para ingreso, continuidad y egreso de los estudiantes en la formación de grado en la Universidad Nacional de Rio Cuarto. Mayo 2016. V jornadas Nacionales y I latinoamericanas de ingreso y permanencia en carreras científico-tecnológicas

17.A.Zaldívar Colado, R.Estrada Lizárraga, L Nava Pérez, J.Peraza Garzón, E. Torregroza Fuentes, E. Sánchez Valdez (Zaldivar 2011) Comparativa entre los Métodos de Regresión Lineal y Minería de Datos para la Identificación de Variables Asociadas al Éxito Académico en Estudiantes de Educación Superior. Investigación y educación en un mundo en red. Madrid, 21-23 septiembre 2011

Análisis de procesos Maps en Hadoop

Rosana Gimenez[1], Miguel Méndez-Garabetti[1], Martín Eraso[1]

[1] Facultad de Ciencais Sociales y Administrativas, Universidad del Aconcagua, Catamarca 132, Mendoza, Argentina
{rgimenez, mmendez}@uda.edu.ar, mferaso@gmail.com

Resumen. Hadoop, emplea el paradigma de programación MapReduce, que ha sido ampliamente aceptado como el marco estándar para el análisis de grandes volúmenes de datos en entornos distribuidos. Desafortunadamente, este rico marco no estaba realmente explotado hacia el procesamiento a gran escala de datos espacio-temporales, sobre todo con la aparición y la popularidad de las aplicaciones que los crean en gran escala. Hadoop es una herramienta de programación basada en Java que soporta el almacenamiento y el procesamiento de grandes conjuntos de datos en un entorno informático distribuido y es muy apropiado para un gran volumen de datos. Utiliza HDFS para el almacenamiento y utiliza MapReduce para el procesamiento. El objetivo principal del modelo de programación MapReduce es poner en paralelo la ejecución del trabajo a través de múltiples nodos para su ejecución. Hoy en día, los investigadores y las empresas siguen sumando aportes mediante investigaciones para la mejora de Hadoop. Debido a esto, se han propuesto muchos algoritmos de planificación en las últimas décadas. Hay tres importantes problemas de programación en MapReduce como localidad, la sincronización y la equidad. El objetivo más común de algoritmos es reducir al mínimo el tiempo de finalización de una aplicación paralela. Parte de los problemas de rendimiento que se plantean para los programadores de Hadoop es el análisis de rendimiento comparativo entre diferentes casos de representados en varios puestos de trabajo.

Palabras Calve: Hadoop, Mapa Reduce, procesamiento en paralelo, algoritmos de planificación, clúster.

1 Contexto

Este trabajo forma parte de una investigación dentro del Laboratorio de Investigaciones de las carreras de Informática en la Universidad del Aconcagua, en forma paralela se está desarrollando una guía completa de Hadoop, que servirá de material didáctico, y quedará a disposición de los estudiantes. Es importante remarcar, que la presente temática cobra relevancia frente las necesidades que tienen las empresas para el procesamiento de grandes volúmenes de datos que cada vez les llegan en mayor cantidad, y las respuestas que deben generar exigen mayor inmediatez.

[1] Proyecto financiado por el ISCOTT, de la FCSA, UDA.

2 Introducción

Una revisión de la literatura nos indica que el problema de la planificación de trabajos que comparten un mismo conjunto de entradas, y el problema de la gestión de grandes volúmenes de datos intermedios, son tratados usando enfoques distintos y las soluciones para los problemas no están integrados en una misma política de planificación de trabajos.

Teniendo en cuenta que para el algoritmo FIFO -el clásico utilizado en Hadoop para la planificación de trabajos- el problema de la localidad no está bien resuelto, uno de los enfoques utilizados para resolverlo es aplicar técnicas destinadas a reducir el uso de recursos de disco y de red, tratando de mejorar la localidad de los datos y los accesos a disco, como propuso Zaharia, en el algoritmo de planificación de trabajos Delay Scheduling para aumentar localidad de datos en aplicaciones desarrolladas bajo el paradigma MapReduce. El algoritmo retrasa la planificación de tareas en unidades de tiempo cuando una tarea no encuentra los datos en el mismo nodo que esté libre. Sin embargo, su uso no es adecuado para tareas de tiempo de ejecución largos porque pueden causar degradación del rendimiento en los trabajos que tengan sus tareas retrasadas. Además, el valor de unidades de tiempo debe estar configurado para obtener buenos resultados. Establecer este valor depende del tipo de carga y el entorno de ejecución. Los valores bajos no pueden garantizar la localidad de datos y los demasiado alto pueden causar problemas de posposición indefinida.

Algunos trabajos proponen el almacenamiento en cache de datos para acelerar el acceso a ellos. Sin embargo, a menudo estas propuestas requieren modificaciones en la aplicación y/o en el flujo de trabajo o requieren un mecanismo sistemático de almacenamiento en cache, para la manipulación de archivos a través de múltiples dispositivos.

Las aplicaciones informáticas que manejan grandes volúmenes de datos pueden beneficiarse de la utilización de sistemas de computación paralelos y distribuidos tanto para mejorar el rendimiento como la calidad de la gestión y análisis de datos. Sin embargo, necesitan mejores diseños de sistemas y administradores de recursos porque introducen nuevas dimensiones, como la representación y almacenamiento de datos, el acceso descentralizado a los datos, múltiples estrategias de paralelismo y la heterogeneidad de entornos.

MapReduce está entre una de la más popular arquitectura distribuida comúnmente utilizada para el procesamiento y análisis de grandes volúmenes de datos analíticos. MapReduce es un sistema distribuido. Se ha adoptado ampliamente para diversas aplicaciones tanto de la industria y el mundo académico. La idea principal de la plataforma MapReduce es empujar el cálculo de los datos, a diferencia de otros paradigmas donde los datos son llevados a la plataforma de la memoria principal para sus cálculos.

2.1 MapReduce

Apache Hadoop es un marco de código abierto desarrollado en paradigma MapReduce. Se utiliza para el almacenamiento y procesamiento de datos masivos de forma distribuida. La plataforma Hadoop puede ser vista como un conjunto de máquinas que constan de un nodo maestro y varios nodos de trabajo. Ofrece diversas

herramientas, tales como Hadoop Distributed, File System (HDFS), planificación de tareas y funciones de gestión de recursos. El componente más crítico de Hadoop es su capa de almacenamiento HDFS, donde antes de procesar los datos en Hadoop, los datos tienen que estar en el HDFS. Inicialmente, Hadoop divide los datos en archivos de grandes bloques. Entonces, se los distribuye a través de los nodos de la agrupación. Después, se transfieren los paquetes de map-reduce, con las tareas en nodos en paralelo para el proceso. El objetivo de este método es llevar el cálculo a los datos donde residen, que ofrece procesamiento eficiente para la escala masiva de datos.

La *Fig.* 1 muestra como HDFS hace la distribución de los bloques en el cluster utilizando la política de distribución por defecto, que es aleatoria; y como mantiene el mapeo de los mismos en Nodo servidor.Remark 1.

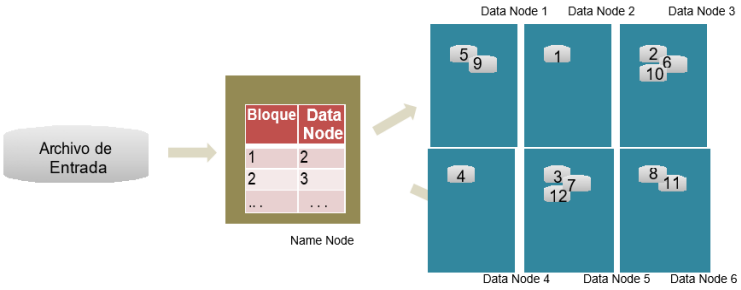


Fig. 1. - Mapeo y distribución de bloques de entrada en HDFS. – Fuente López 2014.

El creciente volumen de datos digitalizados y procesados por aplicaciones puede ser observado en simulación, sensores de datos, procesamiento de imágenes médicas y en prácticamente todas las áreas de conocimiento.

2.1 Gestión de trabajos en cluster Hadoop

La gestión de los trabajos en un cluster Hadoop se realiza por el planificador de trabajos Hadoop a través de dos procesos: jobtracker (en el nodo master) y rastreador de tareas (en los nodos workers). Los trabajos MapReduce presentados al cluster son inyectados al sistema en la cola de trabajos gestionada por el Jobtracker. El orden de ejecución de estos trabajos es definido por la política de planificación de trabajos de Hadoop. Por defecto esta política es basada en una cola del tipo FIFO. Los trabajos encolados son divididos por Hadoop en un conjunto de tareas Map y tareas Reduce. Para cada tarea Map se asocia un bloque de entrada. Y como Hadoop mantiene una tabla con la ubicación de cada bloque en el cluster, se define así donde preferiblemente se procesará la tarea Map. El mapeo es gestionado por el jobtracker.

Estas tareas son asignadas a cada uno de los rastreadores de tareas. Para cada tarea recibida, el servidor local crea una máquina virtual Java para ejecutarla. Cuando el nodo local finaliza su tarea, el rastreador de tareas finaliza la máquina virtual e informa de su estado al jobtracker a través de un mecanismo basado en heartbeat que

genera la asignación de una nueva tarea. La Figura 2 presenta la planificación de trabajos en Hadoop.

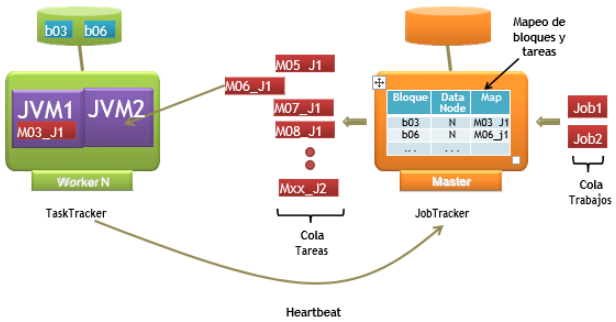


Fig. 2. - Presenta la planificación de trabajos en Hadoop. Fuente López 2014.

3 Resultados esperados

La investigación se encuentra en desarrollo. Actualmente se está evaluando, el trabajo en la nube hibrido, y cómo afectaría a nuestra propuesta de una nueva política de distribución en la etapa del MapReduce.

Referencias

1. Glushkova, Jovanovic, Abelló, "Mapreduce performance model for hadoop 2.x", El Sevier, 2017
2. Dean and Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," Commun. ACM, vol. 51, no. 1, pp. 1–13, 2008.
3. B. T. Rao and L. S. S. Reddy, "Survey on Improved Scheduling in Hadoop MapReduce in Cloud Environments," Int. J. Comput. Appl., vol. 34, pp. 29–33, 2012.
4. M. Zaharia, D. Borthakur, J. Sen Sarma, K. Elmeleegy, S. Shenker, and I. Stoica, "Delay scheduling: a simple technique for achieving locality and fairness in cluster scheduling," Proc. 5th Eur. Conf. Comput. Syst., pp. 265–278, 2010.
5. C. Li, J. Zhang, Y. Chen, Y. Luo, Data prefetching and file synchronizing for performance optimization in Hadoop-based hybrid cloud. El Sevier 2018
6. "FUSE Project." [Online]. Available: http://fuse.sourceforge.net.
7. S.-G. Kim, H. Han, H.-S. Jung, H.-S. Eom, and H. Y. Yeom, "Improving MapReduce Performance by Exploiting Input Redundancy.," J. Inf. Sci. Eng., vol. 27, no. 3, pp. 789– 804, 2011.
8. Y. Luo, S. Luo, J. Guan, and S. Zhou, "A RAMCloud Storage System based on HDFS: Architecture, implementation and evaluation," J. Syst. Softw., vol. 86, pp. 744–750, 2013.
9. J. Ousterhout, P. Agrawal, D. Erickson, C. Kozyrakis, J. Leverich, D. Mazières, S. Mitra, A. Narayanan, G. Parulkar, M. Rosenblum, S. M. Rumble, E. Stratmann, and R. Stutsman, "The case for RAMClouds: scalable high-performance storage entirely in DRAM," ACM SIGOPS Oper. Syst. Rev., vol. 43, pp. 92–105, 2010.

Transformation of Models as Key Element for Increasing the Productivity in Software Development: A Case Study.

Castro Rojas Luis Fernando, Esperanza Espitia Peña, Edwin Moreno Cuero

Universidad del Quindío

Abstract. The transformations between models are a key aspect that is carried out during the development of model-driven software. In this sense, many authors have generated diverse types of approaches, techniques and tools which are intended to sustain and validate the relevance of this field. Particularly, in the context of software development aimed to the generation of games, several challenges that prevent a fluid and efficient work must be faced. Some of these challenges are incompatibility between platforms and technologies, inconsistencies in code versioning, time spent in detecting and correcting bugs, and non-uniform programming standards, among others. This paper presents a case study in order to illustrate the application of MDA (model driven architecture) in the domain of the creation of computer games. So, we pretend to demonstrate that MDA supported by the development of a chain of transformation of models can improve the productivity and efficiency in software development.

Keywords: MDSD, MDA, model driven development, UML, model, metamodel, transformation, game.

Resumen. Las transformaciones entre modelos son un aspecto clave que se lleva a cabo durante el desarrollo de software dirigido por modelos. Con relación a este aspecto, muchos autores han generado una serie de planteamientos, técnicas y herramientas con las cuales se pretende sustentar y validar la pertinencia de dicha aproximación. Particularmente, en el contexto de desarrollo de software orientado a la generación de juegos, se deben enfrentar varios retos que dificultan un trabajo fluido y sin obstáculos. Algunos de dichos retos son incompatibilidad entre plataformas y tecnologías, inconsistencias en versionamiento de código, tiempo invertido en detectar y corregir errores, uso de modelos no homogéneos y estándares de programación poco uniformes. En este artículo se presenta un caso de estudio que ilustra la aplicación de Model Driven Architecture (MDA) en el dominio de la creación de juegos por computador. De esta formar se puede demostrar que MDA apoyado por el desarrollo de una cadena de transformación de modelos puede mejorar la productividad y eficiencia en el desarrollo de software.

Palabras clave: MDD, MDA, UML, modelo, metamodelo, transformación, juego.

1 Introduction

According to [1], a model is a simplification of a system built for a specific purpose, which must have the ability to answer questions instead of the current system. Under this premise and in order to introduce the development of this work, it is important to note that in late 2000, OMG members first reviewed the document entitled "Model Driven Architecture" and decided to form an architecture team to produce a more formal statement of the MDA. In mid-2014, OMG adopted the revised "MDA Guide Revision 2.0" as a more detailed definition of the architecture. Such proposal named MDA (model driven architecture) provides guidelines for structuring software specifications that are expressed as models and can be defined as the realization of the principles of MDE (model driven engineering) around a series of OMG standards such MOF, XMI, OCL, UML, CWM, SPEM, etc. [2]. Through these initiatives we can notice a special interest in demonstrating the advantages of moving from a code-centered approach to one based on models. Figure 1 illustrates a differentiation between the approaches based on object technology and based on models (MDA).

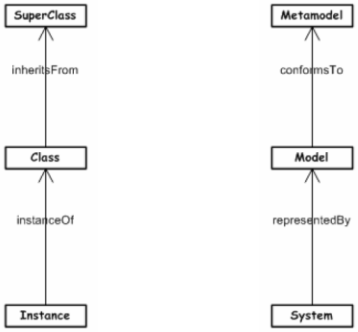


Figure 1. Differentiation between the object technology approach and MDA. Source: [3].

In this sense, the approach known as model-driven software development (MDSD) acquires a representative role in the context of software engineering, enhanced by various elements designed to support the foundation and realization of such an approach. One of these elements is the transformation of models, as documented in [4], the transformation *t* of a model *Ma* into another model *Mb* is defined by:

$t: Ma \rightarrow Mb$

Where the model *Ma* is supposed to be based on a metamodel *MMa* and the model *Mb* on a metamodel *MMb*.

Considering that transformation in itself is another model, then we have that:

Mt: Ma → Mb
In this sense the existence of a generic *MMT* transformation metamodel must be considered, which should be similar to any other MOF (Meta Object Facility) based on the MDA metamodel.

Based on the considerations previously described, this work presents the development of a case study whose intention is to become a reference through which it is possible to demonstrate the advantages and opportunities generated by the transformations of models in the scope of MDA. So, we are focused on a specific area of software development, such as the development of maze games.

2 Theoretical framework

2.1 Model Driven Architecture (MDA).

Model Driven Architecture® (MDA®) is an approach to software design, development and implementation spearheaded by the OMG. MDA provides guidelines for structuring software specifications that are expressed as models. This approach is based on several pillars namely: the separation of the domain from the business descriptions and the domain of the implementation dependent on the platform, the expression of specific aspects of a system in development with specialized domain-specific languages, the establishment of precise relationships between these different languages in a global framework and the ability to express transformations between them.

In [6] the author presents a description of the life cycle of MDA, which does not differ much from the traditional approach of the life cycle of software development, the difference lies in the importance of the role played by the models. The Figure 2 illustrates a proposal in that sense presented by [7].

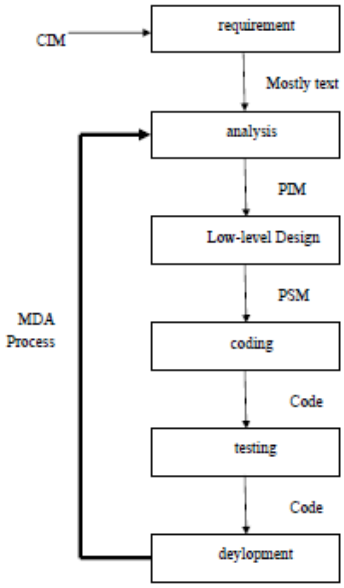


Figure 2. MDA life cycle Computation Independent Model (CIM). Source: [7]

2.2 Computation Independent Model (CIM).

It is the description of the system from the point of view of the expert of the domain which does not show the detailed structure of the system. It specifies only the characteristics of the system domain and is sometimes known as the domain model [12].

2.3 Platform Independent Model (PIM).

It is the first model defined in MDA as the highest level of abstraction that is independent of any implementation technology. Specifies a system to support business logic. This independence can guarantee that the model is used with several platforms of similar type to represent the abstraction of other platforms. A CIM only describes business concepts whereas a PIM may define a high-level systems architecture to meet business needs. For example, a PIM may define a Service Oriented Architecture (SOA) for an information sharing need defined in a CIM [12].

2.4 Platform Specific Model (PSM).

The PIM will be transformed into one more PSMs. This is a specific model for a platform or technology that allows you to describe the system in those same terms. The technology can be a framework like EJB or the relational database model [12].

2.5 Transformation of models.

According to [12] transformation deals with producing different models, viewpoints, or artifacts from a model based on a transformation pattern. In general, transformation can be used to produce one representation from another, or to cross levels of abstraction or architectural layers. When transformation is used to produce one representation from another, the conceptual content stays the same but the information is presented in another form. For example, you may create HTML documentation for an information model – but the information model content is the same – just available in HTML. Often, transformations combine multiple input models into an output model. For example, a technology-independent model may be transformed into a technology-specific model by combining it with a model of the technology to be used (e.g. J2EE).

In [7] the authors categorize the transformation of models into two dimensions: vertical and horizontal. The first occur when a source model is transformed into a target model at a different level of abstraction, when a model is refined. In the context of MDA, vertical transformations are useful when transforming in platform independent model (PIM) to a specific platform model (PSM). The horizontal transformations refer to transforming a source model into a destination model which has the same level of abstraction, are carried out to execute the evolution of models. [5]. Figure 3 illustrates a representation of the elements involved in the model transformation process:

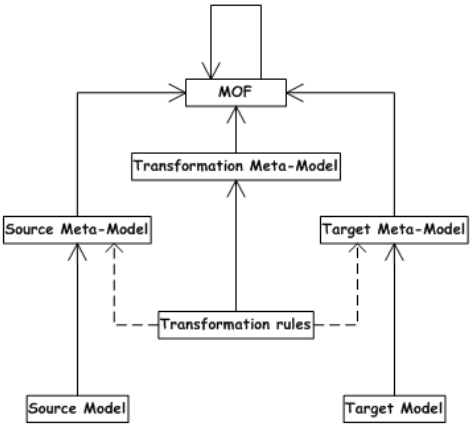


Figure 3. Diagram based on the transformation of models [4].

2.6 Languages and tools to support the transformation between models.

Eclipse ATL (ATL Transformation Language) [8].

Eclipse ATL (ATL Transformation Language) is a model transformation language and toolkit. In the field of Model-Driven Engineering (MDE), ATL provides ways to produce a set of target models from a set of source models. Developed on top of the Eclipse platform, the ATL Integrated Environment (IDE) provides a number of standard development tools (syntax highlighting, debugger, etc.) that aims to ease development of ATL transformations.

QVT (Query view transformation) [9]

The QVT specification has a hybrid declarative/imperative nature, with the declarative part being split into a two-level architecture. We start by explaining the two-level architecture of the declarative part, as it forms the framework for the execution semantics of the imperative part.
The declarative part of its specification is structured in two layers:

• A user-friendly Relations metamodel and language that supports complex object pattern matching and object template creation. Traces between model elements involved in a transformation are created implicitly.

• A Core metamodel and language defined using minimal extensions to EMOF and OCL. All trace classes are explicitly defined as MOF models, and trace instance creation and deletion are defined in the same way as the creation and deletion of any other object.

The imperative part is given by two mechanisms that invoke implementations of transformations from the Relations or the core: a standard language, Operational Mappings, as well as implementations of non-standard MOF operations, Black-box. Each relationship defines a class which will be instantiated to make the trace between the model elements that are being transformed and has a one-to-one mapping with a signature of the operation that implements the Operational Mapping or the Black-box. In Figure 4 these relationships are illustrated.

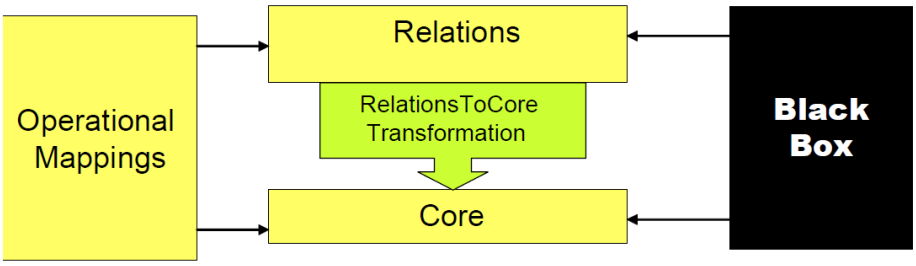


Figure. 4. Relationships between QVT metamodels [9].

ETL (Epsilon transformation language) [10, 12].

It is a model-to-model transformation language based on rules, built on Epsilon Object Language. It provides all the standard features of a transformation language and also expanded flexibility, being able to transform many input models to many output models, as well as consult, navigate and modify them. It is a hybrid language, which provides declarative rules with imperative bodies.

Acceleo [11].

Eclipse Acceleo is an open source code generator implementing the OMG's MOF Model to Text Language (MTL) standard that uses any EMF based models (UML, SysML, domain specific models...) to generate any kind of code (Java, C, PHP...). Acceleo is a template-based technology including authoring tools to create custom code generators. It allows you to automatically produce any kind of source code from any data source available in EMF format. It was created by the French company OBEO. Designed to complement the benefits of MDA technology, by increasing productivity during software development. The Figure 5 shows a representation of the Acceleo architecture.

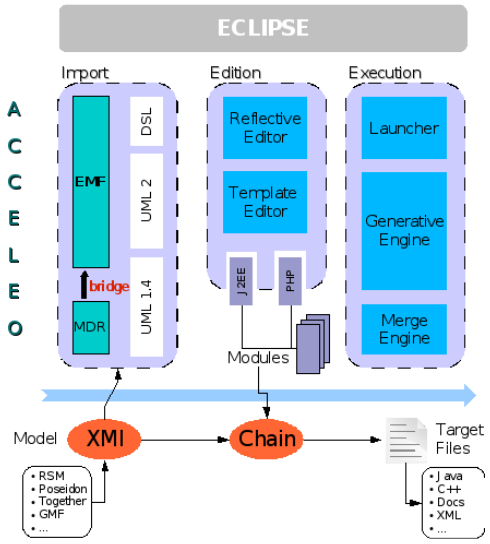


Figure. 5. Relations between Acceleo components [11].

3 Description of the Case Study.

The case study is framed within the domain of maze games. So, we can define one or several scenarios that share the same structural elements:

- A board on which the elements of the game are going to be displayed.
- Game elements of active type, are all those that have some type of behavior which is evidenced by the modification of one or more of its attributes.
- Game elements of passive type, are those that do not have any behavior, for example walls on the board.
- A player, who in turn is an active element and represents the main character of the game.
- The characterization of the game also involves a definition of challenge by means of which the player is intended to fulfill the objective of the game.

In all cases the main character must move through a maze and try to meet their goal to win the game or accumulate points, these objectives can be: collect other items, find a way out, uncover some keys, push other elements, etc. Well-known examples are Bomber Man and Pack Man.

Finally, we pretend to build a game in a semi-automatic way, with a development based on models and supported by a Model Transformation Chain (MTC). To do this, we use an iterative and incremental process and we execute the following activities:

For practical purposes, we use the following acronyms for naming the used artifacts:

MMN - Metamodel of the business.
mN - Business model compliant with MMN.
MMA - Metamodel of the architecture of the game.
mA - Architecture model compliant with MMA.

3.1 Definition of the business metamodel (MMN).

It is constituted by the concepts inferred from the specification of the game and from the point of view of domain experts. Figure 6 shows the metamodel obtained using EMF (Eclipse Modeling Framework SDK), used to model domain models, distinguishes between the metamodel and the current model. The metamodel describes the structure of the model and a model is the instance of this metamodel. EMF provides a framework to store the model information, by default it uses XMI format (XML Metadata Interchange).

Currently, EMF is based on two metamodels: Ecore and Genmodel. The Ecore metamodel contains the information about the defined classes and allows you to define

different elements (EClass, EAtribbute, EReference, EDataType). The Genmodel contains additional information for the generation of code.

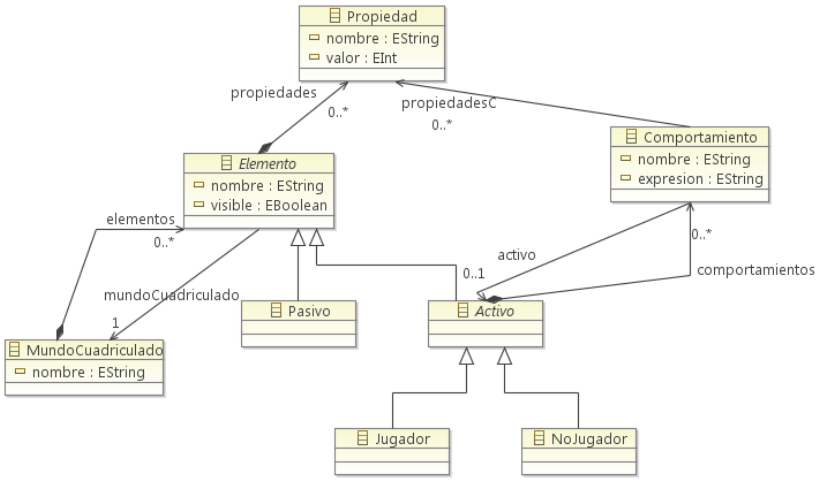


Figure 6. Business Metamodel. Source: the authors

3.2 Creation of a game model compliance to the business metamodel (mN).

EMF provides the easiness to create an instance of the previous metamodel in order to configure our particular game model. The content of the XMI file of a particular game example is illustrated below.

Example of content corresponding to the XMI file of the created model.

```
<?xml version="1.0" encoding="ASCII"?>
<JUEGO:MundoCuadriculado                    xmi:version="2.0"
xmlns:xmi="http://www.omg.org/XMI"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:JUEGO="/Juego2Arquitectura/Metamodelos/Juego.ecore"
xsi:schemaLocation="/Juego2Arquitectura/Metamode-
los/Juego.ecore  D:/DOCTORADO  SEMESTRE  2/DESARROLLO  Sw
MODELOS/workspaceTRAFINAL/Juego2Arquitectura/Metamodelos/
Juego.ecore" nombre="MundoPackman">
  <elementos  xsi:type="JUEGO:Jugador"  nombre="MiJugador"
visible="true" mundoCuadriculado="/">
    <propiedades nombre="posX"/>
    <propiedades nombre="posY"/>
```

10

```
    <propiedades nombre="numCookies"/>
    <comportamientos nombre="moveUp"
propiedadesC="//@elementos.0/@propiedades.1"expre-
sion="+1"/>
    <comportamientos nombre="moveDown"
propiedadesC="//@elementos.0/@propiedades.1"  expresion="-
1"/>
    <comportamientos nombre="moveRight"
propiedadesC="//@elementos.0/@propiedades.0"       expre-
sion="+1"/>
    <comportamientos nombre="moveLeft"
propiedadesC="//@elementos.0/@propiedades.0"  expresion="-
1"/>
    <comportamientos nombre="collectCookies"
propiedadesC="//@elementos.0/@propiedades.2"expre-
sion="+1"/>
  </elementos>
</JUEGO:MundoCuadriculado>
```

3.3 Definition of the architecture metamodel (MMA).

For this model, concepts of the implementation architecture of the game are added. In this case the adoption of the MVC (Model View Controller) pattern is added. This is a classic design pattern that allows the separation of the functionality of the business model from the presentation and control logic that uses this functionality.The proposed metamodel is illustrated in Figure 7.

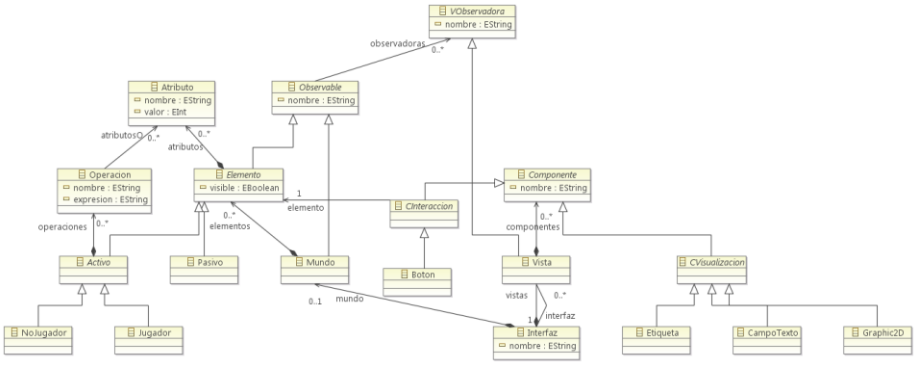
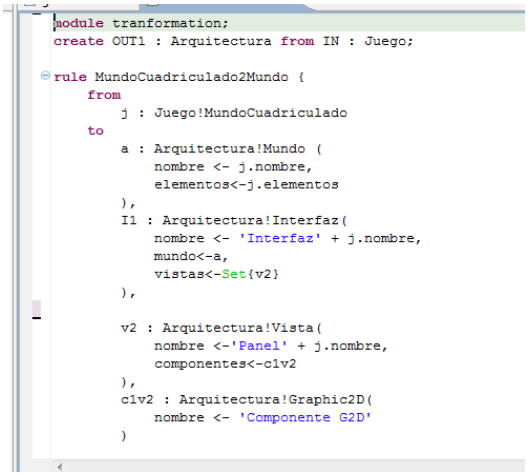


Figure 7. Metamodel of the architecture of the game. Source: the authors

3.4 Transformation of the MMN compliant model to the MMA compliant model.

We use ATL in order to perform the first model-to-model transformation, which allowed us to obtain a metamodel instance conforming to MMA, starting from a mN input model compliance to MMN. In Figure 8 we present an example of a rule applied during the transformation and the XMI file of the model generated as a result of the transformation.



```
module tranformation;
create OUT1 : Arquitectura from IN : Juego;

rule MundoCuadriculado2Mundo {
    from
        j : Juego!MundoCuadriculado
    to
        a : Arquitectura!Mundo (
            nombre <- j.nombre,
            elementos<-j.elementos
        ),
        I1 : Arquitectura!Interfaz(
            nombre <- 'Interfaz' + j.nombre,
            mundo<-a,
            vistas<-Set{v2}
        ),

        v2 : Arquitectura!Vista(
            nombre <-'Panel' + j.nombre,
            componentes<-c1v2
        ),
        c1v2 : Arquitectura!Graphic2D(
            nombre <- 'Componente G2D'
        )
```

Figure 8. Example of the generated transformation. Source: the authors.

3.5 Transformation of model to text.

In this step we use Acceleo, which is a powerful tool to generate the source code of the game. Among its main features we can find: it is Open Source type, has complete integration within Eclipse, native use of EMF, provides a friendly environment for editing and development, among others. In Figure 9 we show an example of the use of Acceleo as a code generator and a part of the source code generated by the tool.

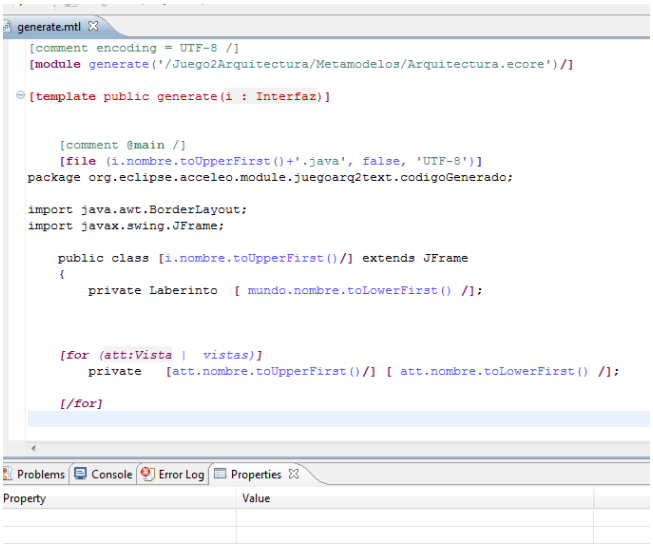


Figure 9. Code example generated by using acceleo. Source: the authors.

4 Evaluation and Analysis of results.

As a result of this case study, it was possible, in the first instance, to typify the modeling of this type of software development projects. From these models it is possible to standardize the design of the applications and, to ensure that the code finally obtained is consistent with the designs proposed in the early stages of the development process.

On the other hand, in table 1 we present a relation of the number of lines of code required in the implementation of a game taken as an example. For this test case, the functionality is defined by a player-type element that has the ability to move through a maze by responding to a directed control when pressing any of the buttons displayed on a button panel within the interface.

Table 1. Number of lines of code in Java of the game implemented. Source: the authors

Class Name	Number of lines of code
InterfazMundoPackman.java	39
PanelControles.java	59
PanelEstadsiticas.java	37
PanelMundoPackman.java	29
Jugador.java	33
Laberinto.java	36
TOTAL	233

Once the code generation template was developed, it was possible to automate the production of 183 lines of code, which independently of the input model that represents an instance of a particular game, will be updated with the corresponding information to generate a new code of the application. This represents an approximate gain of 78%. The remaining percentage is made up of elements that at the time were not contemplated within the scope of modeling for this work, but that in a later version it is possible to include them in such a way that the automation is close to 100%.

In addition to the obvious advantages with respect to savings in the manual generation of lines of code, an important element to consider is to evaluate the performance in terms of aspects inherent to the same programming process, which directly benefit several aspects. For example, quantity of defects, time of location and correction of defects, quality of the product generated among others, which results in the efficiency and productivity of the software development process.

5 Conclusions and future work.

It is possible to increase the performance, the efficiency and productivity of the development of applications using the transformation chain approach of models within the framework established by MDA. The case study that was used to perform the validation can easily be extended to other domains, even more complex, that allow us to consider and study other additional impact aspects to those that were discussed in this document.

As future work we pose that when the size of a project increases, likewise the benefits achieved should be increased. Therefore, an interesting work would be to apply this approach to more robust applications to evaluate the results and contrast them with those obtained through the use of traditional techniques for software development.

On the other hand, we pose to analyze other metrics, in addition to the number of generated lines of code. This will allow us to validate the approach of the chains of transformation of models from other points of view.

References

1. Bézivin, J., and Gerbé, O. (2001). Towards a Precise Definition of the OMG/MDA Framework. Presented at ASE, Automated Software Engineering.

2. Bezivin, J. (2005). On the unification power of models. Software and Systems Modeling. Presented at ASE, Automated Software Engineering.

3. Bézivin, J., Gérard, S., Muller, P. A., and Rioux, L. (2003). MDA components: Challenges and Opportunities". Metamodelling for MDA. First International Workshop Proceedings. York, UK, November.

4. Bézivin, J. (2004). In Search of a Basic Principle for Model Driven Engineering. Novática.Vol. 5 (2).

5. Judson, S.R., France, R.B., and Carver, D.L. (2003). Specifying Model Transformations at the Metamodel Level.

6. Wensheng W. (2005). Evaluation of UML Model Transformation Tools. Business Informatics Group.

7. Kleppe, A. (2003). MDA Explained- The Model-Driven Architecture: Practice and Promise. ISBN 0-321-19442-X.

8 Eclipse Foundation. (2018). Eclipse Project Documentation, User guide. ATL ver. 4.0.1. 2018-12-10.

9. Object Management Group.(2016). MOF Query/View/Transformation Specification Version 1.3. June 2016.

10. Eclipse Foundation. (2018). Eclipse Project Documentation. ETL, ver. 1.5. 2018-18-01.

11. Eclipse Foundation. (2019). Eclipse Project Documentation. Acceleo, ver. 3.7.8. 2019-03-20.

12. OMG. (2014). MDA Guide rev. 2.0 June 2014.

13. Kolovos, D., Rose, L., Garcia, A. and Paige, R.(2018).The Epsilon Book.

Representation of TROPOS by using Essence Kernel

Castro Rojas Luis Fernando, Esperanza Espitia Peña, Andrés Montilla

Universidad del Quindío

Abstract. The goal-oriented requirements engineering supports the requirements management by using an approach driven by goals. These goals can represent either strategic concerns or technical concerns. Several goal-oriented methods have been proposed, for example KAOS, GBRAM, TROPOS, and I*, among others. However, these approaches exhibit some shortcomings related to multiple interpretations. So, the integration and comparison of diverse GORE (Goal-Oriented Requirements Engineering) methods are difficult. In this paper, we propose and demonstrate that SEMAT can be used for achieving a better understanding of the GORE methods. Consequently, the relationships among these methods are easier to treat. This work demonstrates the benefits of using the Essence kernel for representing the TROPOS method.

Keywords: Essence Kernel, TROPOS, goal-oriented, method, requirement, AOSE.

1 Introduction

At present, the complexity of the information systems has forced the software engineers to have a deep knowledge of the organizational environment before starting the software system development. Martinez et al., (2005) highlight the considerable attention has been given in recent years to the early requirements phase in the software engineering community. Fuxman et al., (2001) emphasize about the importance of the early requirements analysis because is one of the most important and complex phases of the software development process, this phase focuses the understanding of a problem by studying the organizational context. The early phase is relevant because an error at this stage it can cause costly errors during late phase in the software development process.
Castro et al., (2014) establish a set of elements in the ESSENCE kernel that are defined in order to collect information from the software engineering process and control the activities performed during the software development process—e.g. stakeholder management, requirements elicitation, and software system development, among others.
Several methodologies with a goal-oriented requirements specification like I* (Yu, 1995), GBRAM (Anton, 1997), and TROPOS are suitable to be represented by using the ESSENCE kernel. Castro et al., (2014) proposes a representation model of GBRAM using ESSENCE kernel, the proposal is addressed by identifying the main concepts of the GBRAM methodology to represent into the concepts of ESSENCE (alpha, alpha-state, activity space, work products, patterns, among others.) (Jacobson et al., 2013). Another work of representation using ESSENCE is developed by Manrique, Gasca &

Gomez (2014), they propose to represent the best practices of the CMMI model using TSP (Team Software Process) framework into the ESSENCE kernel.
TROPOS is an agent-oriented software engineering (AOSE) and goal-oriented requirements engineering (GORE) methodology that based his efforts in the stage of early requirement, this methodology extend the I* methodology (Yu, 1995) into a formal specification language. TROPOS offers special features to be modeled in the ESSENCE kernel, the main feature is that base their efforts in the same context of ESSENCE, the early requirements stage. TROPOS exhibits a more detailed level than other similar approaches, thus the complexity of procedure. Some others agent-oriented software engineering methodologies are MASE (DeLoach & Wood, 2001) and Gaia (Wooldridge et al., 2000).
In this paper we propose a representation model of the main elements of AOSE methodology called TROPOS using the ESSENCE kernel, taking advantage of the main features of the kernel: drivable, extensible and practical (Jacobson et al., 2013).
The paper is organized as follows: In Section 2 we describe the main elements of the goal-oriented requirements specification and the agent-oriented software specification, primarily the TROPOS methodology and its main concepts. In Section 3 we describe the steps to representing TROPOS through the ESSENCE notation. Finally in Section 4 presents some concluding remarks and future work.

2 Theoretical framework

2.1 GORE (Goal – Oriented Requirements Engineering)

The conception of "Goals" is being using more in requirement engineering. Goals have long been recognized to be essential components involved in the requirements engineering (RE) process. GORE is concerned with the use of goals in various activities of requirement engineering like eliciting, elaborating, structuring, specifying, analyzing, negotiating, documenting, and modifying requirements. This area has received increasing attention. (Van Lamsweerde, 2001) (Anwer & Ikram, 2006)

2.2 I*

Yu (1995) proposed this Goal-Oriented methodology that include elements like actors, goals, tasks, belief and soft goals. Also have associations between actors, contribution links and strategic dependencies. In this methodology, the actors and the goals take an important role into of the notation.

2.3 GBRAM

GBRAM was designed for identifying, elaborating, refining, and organizing goals for requirements specification (Anton, 1997)**.** The GBRAM methodology include an

identification initial process of goals and final process that generate a specification requirement document. Among the main elements of the notation GBRAM are: Goal, requirement, agent, scenario, maintenance goal, goal obstacle and inputs/outputs.

2.4 **AOSE (Agent – Oriented Software Engineering)**

Jennings (1999) highlights the idea that many computer systems are being viewed in terms of autonomous agents, agents are being advocated as the next generation model for engineering complex. Agent-Oriented Software Engineering (AOSE), considered software engineering paradigm, bring a set a good practices in a new type of systems: the multi-agent systems, focused mainly on the use of agents. An agent can be a computer system or a person that have an interaction into an organization. "The interaction is probably the most important feature of complex software. Since the 80s systems software agents and multi-agent systems have grown in what is one of the most active areas of research and development in computing. One of the most important is the agents as an autonomous systems, capable of interacting with others agents to satisfy their design goals. The agent-oriented software engineering is at an early stage of its evolution. Preliminary methodologies and tools to support the development of agent systems are beginning to appear, but slowly" (Ciancarini & Wooldridge, 2001).

2.4 MaSE

DeLoach & Wood (2001) proposed this Agent-Oriented methodology that offers a patterns to guide to the developer in the analysis and design process. MaSE combines several established models in an integral methodology that allow derive new models from other existing models. The main concepts handled are: goals, roles and agents.

2.5 Gaia

Wooldridge *et al.,* (2000) argues that Gaia is based on the view of a multi-agent system as a computational organization consisting of various interacting roles. The methodology is applicable to a wide range of multi-agent systems, it covers the macro-level (organizational) and the micro-level (agent) aspects of systems. Gaia handles concepts such as: roles, responsibilities, agent types, services, acquaintances, among others.

2.6 TROPOS (GORE and AOSE)

Bresciani *et al.,* (2004) proposed an agent-oriented software development methodology, founded on the concepts of actor, goal, task, resource, dependency that are used

mainly to model and analyze early and late software requirements, architectural and detailed design. TROPOS is based in two key concepts: the notion of agent and the very early phases of requirements analysis, the second concept allow for a deeper understanding of the environment where the software must operate, and of the kind of interactions that should occur between software and human agents. TROPOS also provides a more refined analysis of system dependencies and a well-defined mechanism to represent the functional and nonfunctional requirements of the information system. The methodology consists of five phases of development but makes increased wear on the first phase (Early requirements analysis). The five phases are:

- **Early requirements analysis**: This stage consist to understanding of problem, there are identified and analyzed the stakeholders (actors that can affect and are affected) into an organizational context.
- **Late requirements analysis**: This analysis focuses in the system development, there is integrated the software system like an actor. The system to be obtained is described in its operating environment with their functions and qualities.
- **Architectural design:** In this phase the global architecture is defined based in subsystems (actors) interconnected by data and control flows (dependencies).
- **Detailed design:** In this phase actors are detailed according to their goals, capabilities and communication protocols.
- **Implementation:** Implementation of the system with an oriented language to agents, where has coherency with the previous designs.

Table 1 exhibits the basic concepts included in TROPOS, the Table 2 shows the elements of the methodology with a goal orientation and the Table 3 illustrates the elements with an agent orientation.

Table 1. Basic concepts include in TROPOS.
Source: (Martínez *et al.*, 2008)

Element	Description
Actor	Entity that have a purpose into the system or the organization, can be represent a person or a software system.
Agent	Actor with demonstrations or specific purposes, the term can refer to human agents or artificial agents.

Role	Abstract representation of a job or function that "plays" an actor into an organization.
Position	**Charge** that occupies an actor within an organization
Hard goal	Strategic interests of an actor that have criteria to corroborate if the goal are satisfied in time.
Soft goal	Strategic interests of an actor that don't have criteria to corroborate if the goal are satisfied in time.
Resource	Physical entity that refers to information.
Plan	Strategy used to achieve a hard goal or a soft goal

Table 2. Basic concepts include in TROPOS with a goal orientation.

Element	Description
Hard goal	Strategic interests of an actor that have criteria to corroborate if the goal are satisfied in time.
Soft goal	Strategic interests of an actor that don't have criteria to corroborate if the goal are satisfied in time.
Resource	Physical entity that refers to information.
Plan	Strategy used to achieve a hard goal or a soft goal

Table 3. Basic concepts include in TROPOS with an agent orientation.

Element	Description
Actor	Entity that have a purpose into the system or the organization, can be represent a person or a software system.
Agent	Actor with demonstrations or specific purposes, the term can refer to human agents or artificial agents.
Role	Abstract representation of a job or function that "plays" an actor into an organization.
Position	**Charge** that occupies an actor within an organization

3 TROPOS representation into the ESSENCE kernel.

In this section we propose a representation of the agent-oriented requirements specification and goal-oriented requirements specification (TROPOS) by using the ESSENCE kernel. We identify the main concepts include in TROPOS and we relate with them to the concepts in the ESSENCE kernel (mainly alphas, work products and patterns). TROPOS and ESSENCE give a special importance at the early requirements stage, this feature benefits TROPOS to be represented in the ESSENCE kernel. The representation of the main concepts of TROPOS with and agent orientation are shown in Figure 1, with goal orientation are show in Figure 2 and the subset of alphas needed for describing this concepts are shown in Figure 3. Figure 3 includes a set of universal alpha elements such as Opportunity, Stakeholders, and Requirements. These elements belong to the Customer and Solution areas of concern.

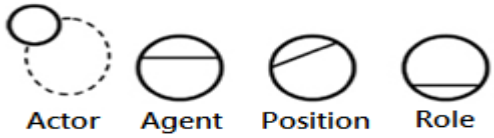


Figure 1. Representation of the concepts of TROPOS with an agent orientation.

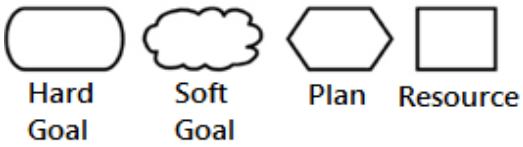


Figure 2. Representation of the concepts of TROPOS with a goal orientation.

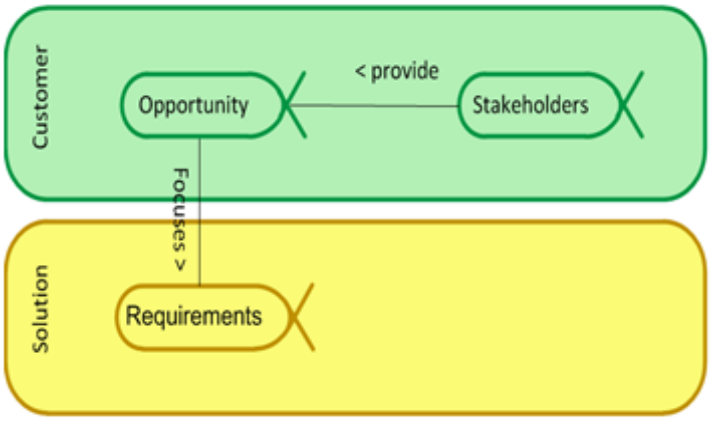


Figure 3. Subset of alphas in the ESSENCE kernel.

In Figure 4 we show the representation of TROPOS methodology by using ESSENCE kernel. In the representation is related elements with alphas (i.e., hard goals and soft goal into Opportunity, agents, actors and position into Stakeholders, plan into Requirements), with pattern (i.e., Role) and resource.

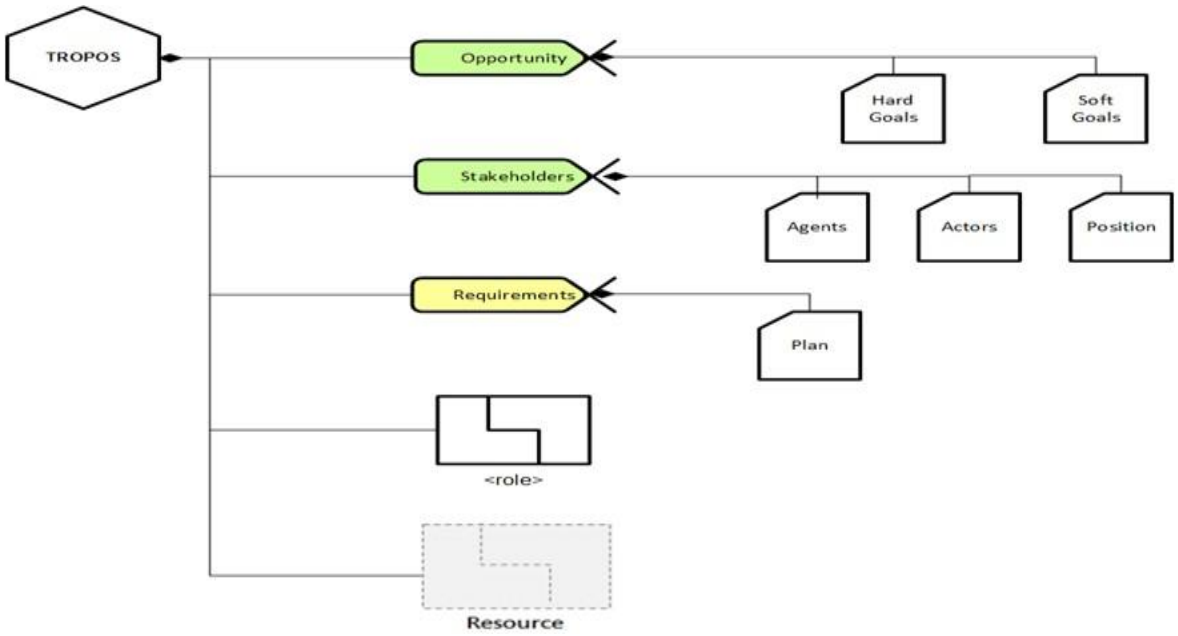


Figure 4. Representation of TROPOS by using the ESSENCE kernel. Source: the authors.

4 Conclusions and future work

In this paper was conceived a model of representation that was supported by ESSENCE kernel for a methodology with specific guidance, in this case, TROPOS. With this proposal of representation was contributed to the standardization an area of great importance as is the engineering requirements through a common framework drivable, extensible and practical like ESSENCE that is related to the different software engineering methodologies of requirements engineering. The main concluding remarks are:

• The important role of TROPOS into the development software process, mainly in the requirements analysis stage.

• TROPOS methodology is suitable to be represented in the ESSENCE kernel because it influences in two important areas of concern in software development process: the customer and the solution.

• GORE and AOSE methodologies have a strong interaction to the area of concern called customer, mainly in the requirements analysis stage.

• TROPOS offer a special interest with the organizational context to understanding the problem and customer needs.
• ESSENCE must allow integrate the good practices of a methodology into another (regardless of orientation) setting relationships between different software engineering methodologies.

Some future works that arises based in the proposed representation in this article are:

• Use the ESSENCE kernel to contribute to the teaching of the different methodologies of software engineering that have a complex structure and can create difficulties in their understanding.
• Integrate to ESSENCE in the representation of different methods that provide facilities and addressing the development at different stages of software development process as requirements analysis, design and architecture, testing, implementation, maintenance, among others.
• Research about the differences among the several goal-based requirements engineering and agent-based software engineering methods to identify and know the benefits of each one and apply them at the software development stage where provide higher performance and more advantages are obtained.

References

1. S. Anwer and N. Ikram, "Goal Oriented Requirement Engineering: A Critical Study of Techniques," pp. 121-130, 2006.
2. A. I. Anton, Goal identification and refinement in the specification of software-based information systems, 1997.
3. L. F. Castro, C. M. Zapata and F. Vargas, "GBRAM from a SEMAT Perspective," Software engineering: Methods, modeling and teaching, pp. 21-26, 2014.
4. B. Manrique, G. Gasca and M. Gomez, "Representation of TSP framework into the SEMAT kernel based on the best practices of Project management from CMMI-DEV," Software engineering: Methods, modeling and teaching, pp. 33-39, 2014.
5. A. Martínez Rebollar and O. Pastor López, "DE REQUISITOS TEMPRANOS A REQUISITOS TARDÍOS: UN ENFOQUE BASADO EN METAS," pp. 61-71, 2005.
6. A. Fuxman, M. Pistore, J. Mylopoulos and P. Traverso, "Model Checking Early Requirements Specifications in Tropos," pp. 174-181, 2001.
7. A. van Lamsweerde, "Goal-oriented requirements engineering: a guided tour," Requirements Engineering, 2001. Proceedings. Fifth IEEE International Symposium, pp. 249-262, 2001.
8. N. R. Jennings, Agent-Oriented Software Engineering, Cairo, Egypt: Springer Berlin Heidelberg, 1999.
9. M. Wooldridge, N. R. Jennings and D. Kinny, "The Gaia Methodology for Agent-Oriented Analysis and Design," Autonomous Agents and Multi-Agent Systems, pp. 285-312, 2000.
10. P. Ciancarini and M. Wooldridge, Agent-Oriented Software Engineering, Limerick, Ireland: Springer Berlin Heidelberg, 2001.

11. S. A. DeLoach and M. F. Wood, "An Overview of the Multiagent Systems Engineering Methodology," in Agent-Oriented Software Engineering, Limerick, Ireland, Springer Berlin Heidelberg, 2001, pp. 207-221.
12. P. Bresciani, A. Perini, P. Giorgini, F. Giunchiglia and J. Mylopoulos, "Tropos: An Agent-Oriented Software Development Methodology," Autonomous Agents and Multi-Agent Systems, pp. 203-236, 2004.
13. A. Martínez Rebollar, H. Estrada Esquivel and L. A. Gama Moreno, "UNA GUÍA RÁPIDA DE LA METODOLOGÍA TROPOS," Gerencia Tecnológica Informática, pp. 67-77, 2008.
14. I. Jacboson, P.-W. Ng, P. E. McMahon, I. Spence, S. Lidman and C. M. Zapata Jaramillo, "La Esencia de la Ingeniería de Software: El núcleo de Semat," Revista Latinoamericana de Ingeniería de Software, pp. 71-78, 2013.
15. E. Yu, Modelling Strategic Relationships for Process Reengineering. Ph.D. Thesis, University of Toronto, Toronto.

Repensando Software que requiere alto rendimiento Paradigma Reactivo

Susana B. Chavez[1], Nelson Rodriguez[1], Adriana Martin[1]

[1] Departamento e Instituto de Informática - F.C.E.F. y N. - U.N.S.J.
Complejo Islas Malvinas. Ignacio de la Roza y Meglioli. C.P. 5402. Rivadavia. San Juan, 0264 4234129
{schavez, schavez, schavez}@iinfo.unsj.edu.ar

Abstract. La aparición de Internet ha provocado un cambio radical en todo el mundo, sumado a esto el constante avance en la arquitectura de CPU multinúcleo y en los servicios de computación en la nube, sin duda han cambiado la forma en que las personas deben resolver los problemas y cómo deben percibir el futuro. El objetivo de este paper es comprender y evaluar estos cambios, y como el paradigma reactivo surge como una solución para abordar problemas relacionados con la disponibilidad, resiliencia, latencia y seguridad entre otros. Los tiempos cambian, los proyectos que se implementan cambian y las cajas de herramientas deben expandirse. Es necesario un nuevo conjunto de herramientas que ayuden a enfrentar eficientemente los desafíos que vienen con esta evolución. Es decir, debe ser posible crear aplicaciones reactivas que sean capaces de proporcionar un comportamiento en tiempo real, incluso bajo una carga elevada y en un entorno descentralizado.

Palabras claves: Paradigma Reactivo, Modelo de Actores, Programación Reactiva.

1 Introducción

La constante evolución tecnológica ha cambiado la forma en que las personas deben resolver los problemas y cómo deben percibir el futuro. Sin duda, la aparición de Internet ha cambiado drásticamente la forma en que las personas y las máquinas interactúan entre sí, permitiendo el acceso instantáneo a información y servicios en tiempo real. La mayoría de los dispositivos utilizan internet para compartir datos, realizar cálculos o proporcionar otros tipos de funcionalidades. Esta conectividad de dispositivos con Internet ha llevado a la creación de nuevos conceptos como Internet de las Cosas (IoT) que ha ganado popularidad en los últimos años. Varias aplicaciones dependen de la potencia informática y otros recursos proporcionados por la nube para su correcto funcionamiento. Los dispositivos como sensores y dispositivos portátiles utilizan los servicios en la nube para procesar los datos que generan. Este hecho pone de manifiesto la necesidad de nuevas formas de manejar la cantidad de datos que se originan en estos dispositivos.

Los desafíos no solo involucran elementos como el almacenamiento de datos o el poder de la computación, sino que también exigen soluciones de software que puedan administrar y procesar esta gran cantidad de información.

Para comprender el por qué y el cómo surge un nuevo paradigma (Paradigma Reactivo), es necesario pensar un poco en las computadoras. Ciertamente, han evolucionado mucho en las últimas décadas, especialmente en términos de velocidad de reloj de la CPU (MHz a GHz) y memoria (kilobytes a gigabytes). Sin embargo, el cambio más significativo que ha sucedido en los últimos años es la cantidad de núcleos que tiene cada CPU. La mayoría de las computadoras tienen al menos 4 núcleos de CPU, y ya hay proveedores que ofrecen CPU con 1024 núcleos. A pesar de esto, la arquitectura general de las computadoras y el mecanismo por el cual se ejecutan los programas no han experimentado una evolución significativa.

Si bien se han propuesto e implementado muchas soluciones posibles para lidiar con este escenario, la mayoría de ellas dependen de la nube como proveedor de servicios. Es justo que haya múltiples proveedores en la nube, como Amazon Web Services y Microsoft Azure, que ofrecen varias soluciones para diferentes problemas. Sin embargo, en los próximos años, una nueva ola entrante de dispositivos que generará una gran cantidad de datos, aumentará la necesidad de diferentes recursos y exigirá diferentes tipos de soluciones, abordando problemas relacionados con la disponibilidad, seguridad y latencia, entre otros.

Edge computing está destinado a ser uno de los principales focos de investigación en los próximos años, ya que se ocupa específicamente de los problemas mencionados [1]. La cantidad de datos generados desde múltiples dispositivos requerirá un cambio de enfoque, no solo en términos de conectividad de red, sino también en términos de soluciones de software.

Los tiempos cambian, los proyectos que se implementan cambian y las cajas de herramientas deben expandirse. Es necesario un nuevo conjunto de herramientas que ayuden a enfrentar eficientemente los desafíos que vienen con esta evolución. Es decir, debe ser posible crear aplicaciones reactivas que sean capaces de proporcionar un comportamiento en tiempo real a los usuarios, incluso bajo una carga elevada y en un entorno descentralizado.

1.1 Nuevos desafíos

En el mundo actual las aplicaciones corren desde dispositivos móviles hasta clústeres basados en la nube que ejecutan miles de procesadores multinúcleo. Los usuarios esperan una disponibilidad del 100% y un tiempo de respuesta de milisegundos. Estas demandas simplemente no son satisfechas por las arquitecturas de software de ayer.

Un enfoque coherente para la arquitectura de sistemas, necesita de características que se reconocen individualmente: sistemas que sean receptivos, resilientes, elásticos y basados en mensajes. Esto recibe el nombre de Sistema Reactivo.

El conjunto de aspectos centrales que caracterizan a los sistemas reactivos en general se publicó bajo el nombre de Manifiesto Reactivo [4]. El uso eficiente de los recursos y una oportunidad para que las aplicaciones escalen automáticamente es el impulsor de una gran parte de este manifiesto:

1. Responsive (Receptivo): Reacciona a los usuarios. Una aplicación receptiva debe satisfacer las expectativas del usuario en términos de disponibilidad y comportamiento en tiempo real. El intervalo de tiempo entre la solicitud y la respuesta se denomina latencia, y es una de las medidas clave a la hora de evaluar qué tan bien funciona un sistema.
2. Elastic (Escalable): Reacciona a la carga. Para poder interactuar continuamente con su entorno, las aplicaciones reactivas deben poder adaptarse a la carga a la que se enfrentan. Las ráfagas de tráfico repentinas pueden afectar una aplicación; por ejemplo, un tweet popular con un enlace a un artículo de noticias podría causar una carrera en un sitio web de noticias. Para este fin, una aplicación debe ser escalable: debe poder utilizar una mayor capacidad computacional cuando sea necesario. Esto significa que debe poder hacer un uso eficiente del hardware en una sola máquina (que puede tener uno o más núcleos de CPU), y también ser capaz de funcionar a través de varios nodos de cómputo a su disposición, dependiendo de la carga.
3. Message Driven (Event-Driven): Reacciona a un evento. Las aplicaciones basadas en la comunicación asíncrona reaccionan ante eventos discretos como las solicitudes HTTP sin monopolizar los recursos computacionales mientras espera que ocurra un evento. Este nivel natural de concurrencia produce mejor latencia que las llamadas a métodos síncronos tradicionales. Otro beneficio de la escritura de programas dirigidos por eventos es que los componentes se acoplan de forma flexible, lo que hace que el software sea mucho más fácil de mantener a largo plazo.
4. Resilient (Resistente): Reacciona a las fallas. Debido a que incluso los sistemas de software más sencillos son propensos a fallar (ya sea relacionados con el software o relacionados con el hardware), las aplicaciones reactivas deben ser resistentes a las fallas para satisfacer la demanda de disponibilidad continua. La capacidad de una aplicación para volver a ponerse de pie en caso de que se encuentre con un problema es posiblemente aún más importante cuando se trata de sistemas escalables, que son más complejos en su naturaleza y están distribuidos, debido a que aumenta la probabilidad de fallas en el hardware o la red.

1.2 Aspectos claves

Es fundamental detenerse en este punto para entender los conceptos básicos, como los modelos de concurrencia [9], utilizados para modelar y manipular operaciones asincrónicas sin perder de vista la tolerancia a fallos.
Debido a que los procesadores multinúcleo son cada vez más estándar, se han creado múltiples niveles de abstracción para simplificar la concurrencia y permitir un desarrollo más simple.
Cada uno de estos modelos tiene un enfoque diferente:

- Futures: son la base de la programación asíncrona y reactiva en Scala. Permiten manipular los resultados de una operación que aún no ha sucedido y lidiar efectivamente con la falla de dichas operaciones, permitiendo un uso más eficiente de los recursos computacionales.
- Memoria transaccional de software (STM): simplifica enormemente la comprensión conceptual de los programas multiproceso y ayuda a que los programas sean más fáciles de mantener al trabajar en armonía con las

abstracciones de alto nivel existentes, como los objetos y los módulos. Usado por Clojure.

- Modelo basado en Actores: es un modelo concurrente de cómputo para crear sistemas altamente concurrentes y paralelizables en un entorno distribuido. Popularizado por el lenguaje de programación Erlang y es implementado por Akka sobre JVM [10].
- Stream Reactivos: proporciona una abstracción para aplicaciones asíncronas altamente concurrentes con soporte para el procesamiento de flujo asíncrono con backpressure sin bloqueo. Ej.: RxJava, Reactor, etc.

No todos los modelos tienen un verdadero manejo de errores y recuperación de fallas. Algunos están diseñados para un alto rendimiento, pero no para escalar ya sea horizontal y/o verticalmente. Las ventajas que ofrecen cada uno de estos modelos son buenas ideas en algunos contextos, pero no en otros. Es preciso cuestionar todo y asegurarse de que realmente se comprenden las fortalezas y debilidades de todas sus opciones de diseño, ya sea que se esté construyendo o no un sistema reactivo.

1.3 Herramientas

Si bien, la creación de aplicaciones reactivas requiere más que simplemente usar tecnologías reactivas, deben cumplir con una serie de requisitos previos para habilitar el comportamiento reactivo, especialmente la capacidad para la ejecución de código asíncrono e impulsado por eventos.

Muchas prácticas comunes y modelos de programación aceptados no abordan desafíos importantes inherentes al diseño de sistemas para arquitecturas informáticas modernas. Para tener éxito, los sistemas distribuidos deben hacer frente a un entorno donde los componentes se bloquean sin responder, los mensajes se pierden sin dejar rastro y la latencia de la red fluctúa. Estos problemas ocurren regularmente en entornos de centros de datos cuidadosamente administrados, aún más en arquitecturas virtualizadas.

El toolkit Akka es un conjunto de bibliotecas de código abierto para diseñar sistemas altamente concurrentes, distribuidos, escalables y resistentes que abarcan núcleos y redes de procesadores [3]. Implementa el Modelo de Actor en la JVM. Esta herramienta permite que los desarrolladores se concentren en satisfacer las necesidades comerciales en lugar de escribir código de bajo nivel para proporcionar un comportamiento confiable, tolerante a fallas y de alto rendimiento.

Existen muchos de los llamados frameworks de aplicaciones web de pila completa, pero pocos de ellos también incluyen tecnologías reactivas, y muy pocos se construyen desde cero utilizando tecnologías reactivas, que incluyen principios reactivos en su núcleo. Los frameworks de pila completa se ocupan de todas las capas necesarias para crear e implementar una aplicación: tecnología de UI del lado del cliente (o un medio para integrarla), lógica empresarial del lado del servidor, autenticación, integración de acceso a la base de datos y varias bibliotecas para tareas más comunes (como llamadas a servicios web remotos).

El frameworks Play se ha diseñado desde cero para que sea posible crear aplicaciones web reactivas que sean capaces de proporcionar un comportamiento en tiempo real a los usuarios, incluso bajo una carga elevada y en un entorno descentralizado [6]. Está disponible en la máquina virtual java, y ha sido adoptado por grandes compañías como Morgan Stanley, LinkedIn y The Guardian.

Además, las distintas plataformas existentes han desarrollado extensiones reactivas como por ej.: Microsoft ofrece Rx (https://rx.codeplex.com/) que es una biblioteca para componer programas asíncronos y basados en eventos, disponibles en la plataforma .NET y otras plataformas como JavaScript. Node.js (http://nodejs.org) es una plataforma popular para crear aplicaciones controladas por eventos asíncronas en JavaScript. En la JVM, varias bibliotecas habilitan estas capacidades, como Apache MINA (https://mina.apache.org) y Netty (http://netty.io).

Todas esas tecnologías de bajo nivel ofrecen herramientas básicas para crear aplicaciones asíncronas y basadas en eventos, pero requiere un poco más de trabajo llegar al estado de una aplicación web completa que también tiene que lidiar con problemas como la organización de códigos, inclusión y organización de recursos del lado del cliente, como hojas de estilo y archivos JavaScript, conectividad de base de datos, seguridad, etc. Las aplicaciones web reactivas son una respuesta a los nuevos requisitos de alta disponibilidad y eficiencia de recursos que ofrece esta rápida evolución. La computación en la nube y la posterior aparición de servicios en la nube han cambiado el desarrollo de aplicaciones web de una actividad en la que una aplicación trata de resolver todo tipo de problemas, a un proceso de identificación y conexión a servicios en la nube adecuados que solo resuelve aquellos problemas que no se han resuelto de manera satisfactoria.

2 Adaptando modelos

Para cualquier aplicación que se desarrolle, el concepto central es el modelo de dominio, que es la representación de cómo funciona el negocio. Cuando se implementa un modelo de dominio, se traducen los procesos de negocio en software intentando hacer esta traducción de manera que resulte en un software lo más parecido a los procesos originales. Para lograr esto, se siguen ciertas técnicas y se adoptan paradigmas para el diseño, desarrollo e implementación. Para un sistema bancario, por ejemplo, las funcionalidades del sistema consisten en entidades como bancos, clientes, cuentas, divisas, transacciones e informes que trabajan juntos para brindar una buena experiencia bancaria al usuario. Es responsabilidad del modelo garantizar que los usuarios tengan una buena experiencia cuando usan el sistema.

Si bien, existen muchos enfoques para el modelado de dominios, es necesario tener siempre en mente que los tiempos cambian, los proyectos que se implementan cambian y las cajas de herramientas siguen expandiéndose. De esta manera es necesario repensar el modelo de manera que tenga una mayor perspectiva de futuro.

Durante la última década aproximadamente, las tecnologías orientadas a objetos (OO) han dominado por completo la creación de los modelos de dominio, incluso los más complejos. Sin embargo, para enfrentar los nuevos desafíos, la programación orientada a objetos (POO) se enfrenta a un problema muy interesante: la

implementación de concurrencia y paralelismo. Estos dos conceptos son el talón de Aquiles de la programación estructurada y orientada a objetos.

Otro aspecto que pone en desventaja a la POO es justamente su concepto central, el objeto, que encapsula el estado y el comportamiento. El estado de un objeto es *mutable*; y se prefiere sobre la inmutabilidad que requiere la construcción de nuevos objetos; la mayoría de las bibliotecas POO son muy ineficientes para hacer copias, lo que hace que la mutación de estado sea importante para el rendimiento. Por otro lado, resulta muy difícil de separarlos en servicios enfocados, livianos y rápidos.

Por ejemplo, si se necesita un objeto "cliente" grueso para muchas historias de usuarios (Fig. 1), la tendencia es forzar todas las rutas de código a través de "Cliente", en lugar de tener implementaciones separadas, con alguna reutilización de código cuando corresponda.

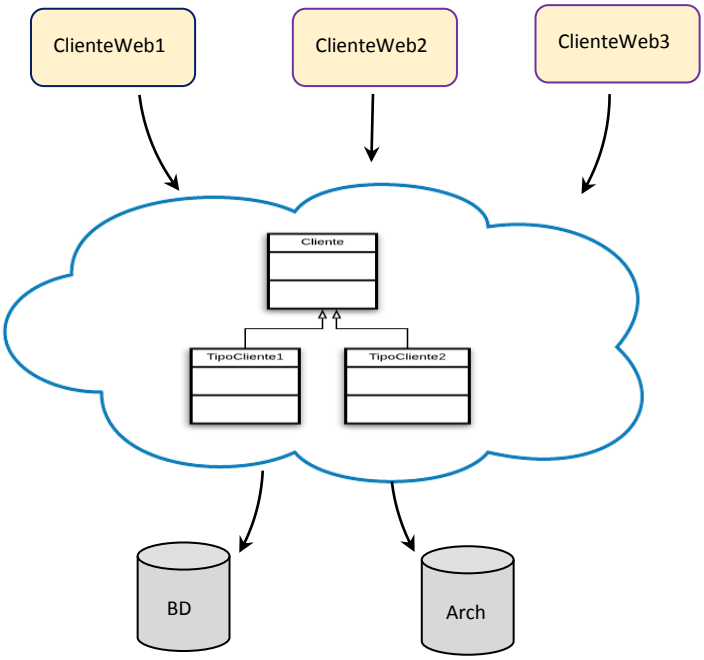


Fig. 1 – Sistema POO

Entonces, surgen cuestiones como: ¿Qué campos deberían estar en esta clase Cliente? ¿Qué pasa si un equipo necesita campos y métodos diferentes de otro equipo? ¿Se debería tener una clase Frankenstein, el superconjunto de miembros requeridos? ¿Se debería tener clases de Clientes separadas y abandonar un modelo uniforme para toda la organización? O bien, dado que cada equipo realmente obtiene los campos Cliente de un conjunto de resultados de base de datos, si cada equipo usa una tupla para los valores de campo devueltos (¡y no devuelve el registro completo!), entonces realice el trabajo requerido y luego genere nuevas tuplas o informe o lo que corresponda. ¡¡¡Parece muy sensata esta última opción!!!

Se difunde la lógica de cada historia de usuario (o caso de uso) a través de los límites de la clase, en lugar de ponerlo en un lugar, donde sea más fácil de leer, analizar y refactorizar. Se pone demasiada información en el código y esto lleva a aplicaciones abultadas que son difíciles de refactorizar para separar microservicios.

Entonces, ¿cómo se posiciona la POO como herramienta para la programación reactiva?

1. Receptivo: cualquier código inflado y lógica de implementación dispersa a través de los límites de la clase baja el rendimiento, posiblemente oscurece los errores y, por lo tanto, daña la capacidad de respuesta.
2. Escalabilidad: es mala, debido a la tendencia a sobre diseñar la implementación codificando más del modelo de dominio de lo absolutamente necesario. Esto hace que sea más difícil dividir el programa en microservicios, lo que limita la escalabilidad. Para sistemas de alto rendimiento, crear instancias de objetos para cada "registro" puede ser costoso.
   La concurrencia es un medio para lograr la escalabilidad, sin embargo, la concurrencia basada en threads no es fácil. Hacer más de una cosa con un programa se complica con el manejo de condiciones de carrera, semáforos, bloqueos, datos compartidos y todo lo relacionado con el multithreading. Esto incluye cuestiones básicas para determinar con precisión qué está haciendo un programa con varios threads, o cuándo se accede a una variable desde múltiples threads, o cuál es su valor en un momento dado, o cómo saber si hay dos threads en espera, y si esta condición los va a liberar (y cuándo) o si es un deadlock (punto muerto). Desafortunadamente, la concurrencia basada en threads presenta más desafíos que soluciones.
3. Event-driven: es bueno, los eventos crean objetos naturales.
4. Resistente: es regular, porque es muy difícil de materializar el manejo de errores, ya que es una preocupación transversal que atraviesa los límites de los objetos de dominio. La lógica dispersa (a través de los límites del objeto) y la mutación de estado hacen que los errores sean más probables.

En este contexto, la programación funcional (PF) experimentó un renacimiento con el lanzamiento de Scala en 2003, F # en 2005, Clojure en 2007 y Elixir en 2012, el enfoque del Modelo de Actor fue declarado ganador en la resolución de problemas de concurrencia.

Al parecer, la FP, promete ser el enfoque más natural para la programación reactiva. Proporciona los conceptos básicos que son mejores para abordar los cuatro rasgos. Se basa en un modelo Anémico, es decir, las funciones están separadas de los valores que representan el estado. Las funciones se aplican a los datos. Las mismas operaciones se pueden usar con cualquier colección, por ejemplo. Reduce en gran medida la hinchazón del código a través de una mejor reutilización. Mayor flexibilidad para componer comportamientos (Fig. 2). Facilita la construcción de microservicios que se pueden fragmentar (escalar) y replicar (resistente).

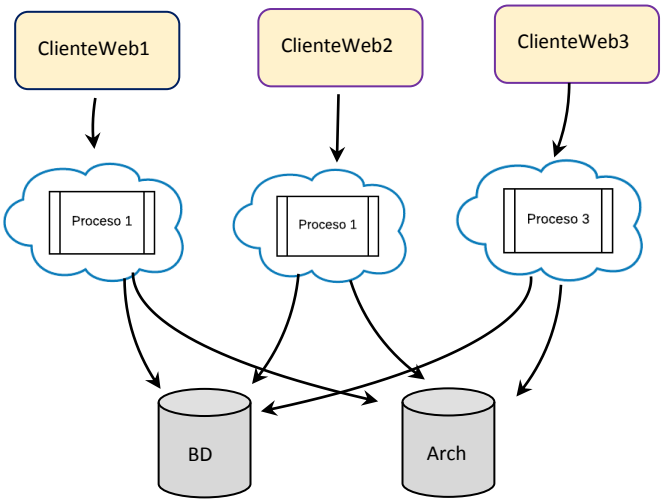


Fig. 2 – Sistema PF

La definición de Robert C. Martin [11], determina una diferencia fundamental con la programación imperativa:

"Functional programming is programming without assignment statements."

Decir que no hay sentencia de asignación equivale a decir que no se puede cambiar el valor de una variable una vez que se ha declarado. En Java, esto sería similar a usar la palabra clave final en todas partes, y solo trabajar con colecciones que no se pueden cambiar después de que se hayan inicializado con un conjunto dado de elementos, esto hace que tengan un mejor rendimiento.

Un valor inmutable, como su nombre lo indica, no cambia. En un programa, los valores inmutables se declaran una vez y sus valores son siempre los mismos durante toda la ejecución del programa, o hasta que ya no sean necesarios y se descarten. Las bibliotecas FP, que prefieren valores inmutables, han trabajado duro para implementar algoritmos eficientes para hacer copias.

Entonces, ¿cómo se posiciona la PF como herramienta para la programación reactiva?

1. Receptivo: manejo uniforme de errores y lógica normal. En general, promueve la separación natural que conducen a una mejor cohesión y un bajo acoplamiento.
2. Escalabilidad: el rigor de las matemáticas puede ayudar a evitar la lógica innecesaria y dividir el conjunto completo de comportamiento en conjuntos disjuntos. Las expresiones pequeñas y concisas minimizan la sobrecarga de recursos.

   Además, las celdas de datos son inmutables. Es importante aclarar en este punto que la inmutabilidad es más lenta, en el nivel micro, que la modificación

de un valor debido a la sobrecarga de la copia (a pesar de las API de copia muy eficientes) y probablemente a una mayor incoherencia de caché (porque tiene dos instancias, no una). En la escala macro, los valores mutables fuerzan una complejidad adicional para coordinar el acceso. Entonces, excepto cuando tiene un módulo muy enfocado que controla el acceso, la inmutabilidad es probablemente más rápida o al menos más confiable al evitar errores causados por mutaciones.

3. Event-driven: controladores fáciles de componer. La transmisión con eventos no se modifica, sino que se reemplaza, filtra, etc. Es natural en FP.
4. Resistente: las funciones sin efectos secundarios son mucho más fáciles de analizar, incluso reemplazar, por lo que es menos probable que tengan errores. Comportamientos complejos como la composición de funciones, solo es posible con funciones libres de efectos secundarios y valores inmutables.

3 Conclusiones

En el modelo conceptual de la POO se mezclan dos cosas: por un lado, se representan los procesos que ocurren en la vida real al describir entidades, y por otro se trata de representar cómo estas entidades evolucionan con el tiempo al mutarlas. Cuando en realidad se crea la ilusión de haber codificado el paso del tiempo en un objeto mutable, pero esta ilusión se desmorona tan pronto cuando alguien más mira este objeto desde un punto de referencia diferente.

Las cosas ya no funcionan de esa manera. Ahora los componentes interoperan mediante el intercambio de estructuras de datos inmutables (listas, mapas y conjuntos), que se parece más a la programación funcional.

Las estructuras de datos mutables conducen a una base de código frágil y dificultan la comprensión del código frente a operaciones concurrentes. Es preferible comenzar con estructuras de datos inmutables: si se necesita hacer que alguna parte del código que tenga mejor rendimiento que con la inmutabilidad, entonces buscar la mutación. Pero la API del cliente no debe ver la mutación; se debe encapsular detrás de una función para mantener la transparencia referencial.

Muchas ideas son buenas en algunos contextos, pero no en otros. A menudo, se piensa en introducir una abstracción como una buena idea, pero, de hecho, las abstracciones tienen sus propios costos que se deben medir, por ejemplo: rendimiento, oscuridad del comportamiento real, efectos secundarios, etc.

Al final de cuentas, lo que importa es asegurarse de que realmente se comprenden las fortalezas y debilidades de todas las opciones de diseño, ya sea que esté construyendo o no un Sistema Reactivo.

Referencias

1. Jim Waldo, Geoff Wyant, Ann Wollrath, and Sam Kendall, Sun Microsystems, Inc., (1994).

2. Manuel Bernhardt: Reactive Web Applications_ Covers Play, Akka, and Reactive Streams, Manning Publications (2016).
3. Raymond Roestenburg, Rob Bakker, Rob Williams: Akka in Action-Manning Publications (2016)
4. Manifiesto Reactivo, https://www.reactivemanifesto.org/
5. Nash, M._Waldron,W.: Applied Akka Patterns_ A Hands-On Guide to Designing Distributed Applications, O'Reilly Media (2016)
6. Framework Play: http://www.playframework.com/documentation
7. Debasish Ghosh: Functional and Reactive Domain Modeling, Manning Publ. (2016)
8. Paul Chiusano, Runar Bjarnason: Functional Programming in Scala, Manning Publications (2014)
9. Adam L. Davis: Reactive Streams in Java_ Concurrency with RxJava, Reactor, and Akka Streams-Apress (2018)
10. Modelo de Actores: https://doc.akka.io/docs/akka/current/guide/actors-motivation.html
11. Robert C. Martin, Functional Programming Basics, https://pragprog.com/magazines/2013-01/functionalprogramming-basics.

Diseño web responsive para el registro, control y difusión de donantes voluntarios de sangre del proyecto "Vida por vidas" en la provincia de La Rioja

Priscila Orona[1], Sandra Rodriguez[2], Emilio Rearte[3] , Diego Candelero[4]

Universidad Nacional de la Rioja. Capital , Chamical, Argentina
priscilaorona94@gmail.com[1], sandraorona27@yahoo.com.ar[2] , emiliorearte@yahoo.com.ar[3], diegocandelero35@gmail.com[4]

Resumen: La donación voluntaria de sangre es importante porque sus componentes y derivados son elementos insustituibles dado que no pueden ser reproducidos de manera artificial. La sangre no se puede "fabricar" en un laboratorio: los seres humanos dependemos de nosotros mismos para ayudarnos. Si bien es probable que hayamos escuchado acerca de los avances de la ciencia en relación con la producción de sangre artificial, esto está muy lejos de la realidad [1]. La propuesta de trabajo es desarrollar una herramienta web responsive que permita la administración y control de los donantes de sangre e interesados registrados en el proyecto de la Iglesia Adventista del Séptimo Dia "Vida por vidas", en la provincia de La Rioja. La aplicación permitirá la difusión y comunicación y la prestación de servicio a los centros de salud de la zona registrados para tal fin.

Palabras claves: donación de sangre, web, herramienta, vida por vidas, donantes voluntarios.

1. Introducción

Las enfermedades y los accidentes son una triste realidad con la que coexistimos día a día y nadie está exento de padecerlos. En estas situaciones y ante la necesidad de contar con suficientes suministros de pintas de sangre en Centros de Acopio y Hospitales esta problemática se torna muy marcada y, el déficit escenario que predomina en Sudamérica es cada vez mayor.

Por esta razón nace el proyecto sudamericano **"Vida por Vidas",** que es un plan fundamentado en el principio de la acción filantrópica e ideado por los jóvenes adventistas pertenecientes a la Iglesia Adventista del 7mo Dia. Este consiste en promover campañas de donación de sangre y médula ósea, e incentiva a las personas a ser parte de este propósito. Su objetivo es alcanzar e involucrar a los jóvenes para que participen como voluntarios en la donación de sangre y hemoderivados, en hospitales y bancos de sangre de ocho países de Sur América: Argentina, Bolivia, Brasil, Chile, Ecuador, Paraguay, Perú, Uruguay [2].

Este proyecto humanitario se viene desarrollando año a año en la ciudad capital de la provincia de La Rioja conjuntamente con la colaboración del servicio de hemoterapia del Hospital Enrique Vera Barros, actualmente en la institución se encuentra instalado un sistema informático llamado "Emot", en donde les permite a ciertas personas dentro del banco de datos poder manejar información, realizar registros de donantes y tener una mejor visualización en cuanto a la cantidad de donantes voluntarios que se acercan con las intenciones de prestar una ayuda a la persona que lo necesite. Este sistema tiene diferentes permisos, en donde solo los encargados autorizados de hemoterapia pueden ingresar y realizar modificaciones al sistema. "No está permitido que cualquier persona ajena al banco de sangre pueda extraer información de esta gran base de datos". [2]

La asesora regional de Servicios de Sangre de la OPS/OMS afirmó que "Los donantes voluntarios y regulares de sangre son los únicos que pueden prevenir la escasez de

sangre y de componentes sanguíneos seguros, deteniendo la necesidad de recurrir a donantes familiares o de reposición".
La OMS recomienda que al menos el 2% de la población de cada país debería donar sangre para tener centros bien dotados, en Argentina. [3].
Para este trabajo han sido seleccionados tres sitios web como puntos de referencias para este proyecto de trabajo final de carrera: 1- Bank India (BBI) en [4], 2- Banco de sangre de la Rioja Española, en [5] y, 3- El Centro Mutual RH (-) de la provincia de San Juan – Argentina, en [6].
Para concientizar acerca de la necesidad de donar sangre, es importante contar con programas de educación, que los bancos de sangre y los servicios de salud lo comuniquen bien y, al mismo tiempo, se cuente con el respaldo de todos los medios de comunicación.
El proyecto de concientización "Vida por vidas", perteneciente a la Iglesia Adventista del Séptimo Dia también carece de un sistema de información y de comunicación que gestione los datos clínicos de los donantes y/o potenciales donantes y del público en general, que registre la cantidad y tipo de sangre que se necesita en los distintos bancos de sangre de la provincia. También, que ofrezca alertas e informes sobre las inscripciones, promociones de sus campañas, como así dónde y cuándo será el próximo evento
Por estas razones, sería beneficioso para el proyecto contar con el sistema en línea que pueda resolver la problemática planteada, ofreciendo a la comunidad un servicio de calidad en los momentos de situaciones de riesgo.

2. Modalidad

De acuerdo con la modalidad establecida en el Reglamento de Trabajo Final de la Carrera de Licenciatura en Sistemas de Información, el desarrollo se considera como un **Trabajo de extensión a la comunidad sanitaria**.

3. Contexto

El proyecto se enmarca en el desarrollo del Trabajo Final de grado, requisito necesario para acceder a la titulación final de la carrera Licenciatura en Sistemas de Información en la Universidad Nacional de La Rioja, cuya línea de investigación se encuentra respaldada bajo la Resolución CICYT N° 044/11 del Consejo de Ciencia y Técnica de la UNLaR, a aplicar es Ingeniería de software y Base de Datos, debido a que ambas son básicas e importantes en el momento de la realización de la aplicación y del buen funcionamiento del sistema.

4. Beneficios de la aplicación

Se opto por la técnica web responsive. Esta técnica ha modificado totalmente la forma en la que se diseña. Ha modificado el patrón, los usuarios y sus hábitos de navegación y consumo han cambiado las reglas.

Fig. 1- Características del diseño responsive

Entre las ventajas que hemos detectado para la implementación del sistema responsive son:

- Mejora la vivencia del usuario ya que se adapta a la web y a las características de cualquier dispositivo, el proyecto en su gran mayoría se enfoca a la población de entre 18 a 40 años de edad.
- Se optimizan los contenidos para minimizar los tiempos de espera de los usuarios, debido al mayor ancho de banda con el que cuentan los distintos dispositivos móviles.
- Los costos de mantenimiento, cuando incluyen contenidos nuevos se realiza una sola actualización para todas las versiones de la web y los costos de desarrollo son menores, que el costo de desarrollo de una versión web para cada dispositivo.

5. Metodología

El enfoque de este proyecto está orientado a la "Investigación aplicada", se optó por elegir la metodología RUP (Proceso Racional Unificado) como herramienta metodológica en la realización de este proyecto debido a que su meta principal es asegurar la producción de software de alta calidad que cumpla con las necesidades de los usuarios, con una planeación y con un presupuesto previsible.
La metodología RUP cuenta con las siguientes fases:

1. **Inicio:**

Se definió y se acordó el alcance del proyecto con los patrocinadores, se realizaron encuestas a través de un cuestionario https://docs.google.com/forms/d/e/1FAIpQLScI5aRPQc4ZxGFe38wEd_P_FqFJ6SNBPukxURd5N6yc1UYoig/viewform, identificando estado de situación y los requerimientos asociados al proyecto, de esta manera se lo logro una visión muy general de la arquitectura de software

2. **Elaboración:**

Se selecciono el caso de uso que permita definir con más claridad el funcionamiento del sistema, se realizó la especificación de los casos de uso seleccionados y el primer análisis del dominio del problema.

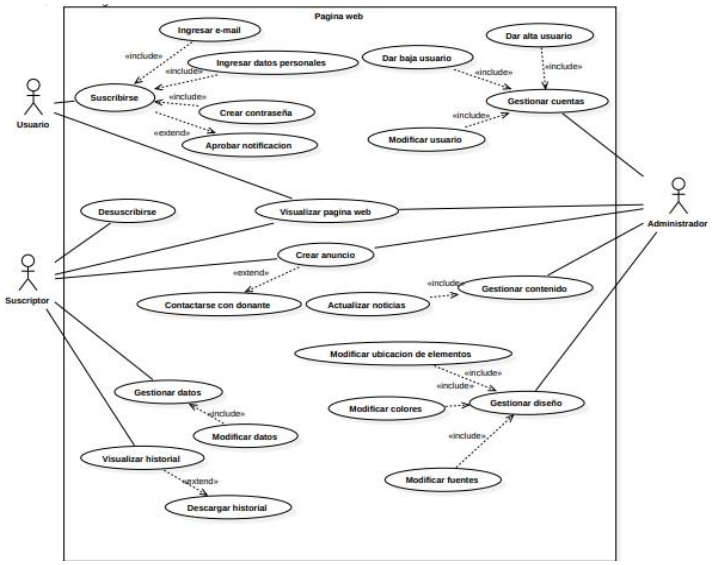


Fig. 2- Caso de Uso- Funcionamiento del sitio Web

3. **Construcción**:

Para el desarrollo de la web se utilizó el diseño web responsive o adaptativo, es una técnica de diseño web que busca la correcta visualización de una misma página en distintos dispositivos. Desde ordenadores de escritorio a tablets y móviles ya sean

ordenadores de sobremesa, portátiles, iPads, tablets Android o smartphones. Permitiendo que la estructura de página sea flexible a la hora de acceder a la web mediante multidispositivos.

4. **Transición:**

- Tecnologías de desarrollo: Framework de Frontend: Materialize, Git
- DBMS: MariaDB
- Servidor Web: WAMP
- Diagramas, modelos y documentación: StarUML
- Gestión del Proyecto: Gantt Project.
- Diseño de logotipo y edición fotográfica: Photoshop 6
- Diseño de Manuales de Usuario: Word.
- Comunicación remota con cliente: E-mail.

Fig 3. Página de Inicio en preparación

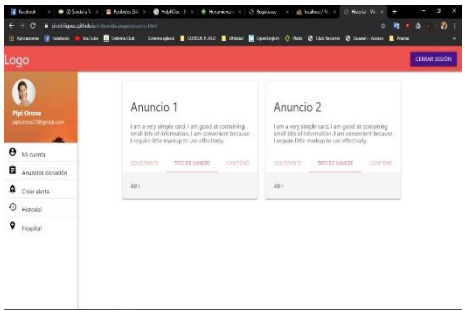


Fig. 4 Login del sitio

6. Conclusiones y trabajos futuros

Se pretende contribuir con este proyecto en brindar información y ser un medio de comunicación eficaz para la sociedad de referencia; en la necesidad que hoy tenemos de convertir el tipo de donantes que existe en nuestra provincia (de reposición) en donantes voluntarios, altruistas y repetidos. Su resultado permitirá gestionar datos clínicos de los donantes más relevantes y datos de grupo de sangre y factor, permitiendo hacer búsquedas por distintos ítems. El acceso para la gestión del sistema dará a usuarios específicos y que a su vez cumplan con de la Ley Nacional N°25.392 que contempla todo lo referente a la protección de datos personales [7].
Como trabajos a futuros y con la ayuda de aplicaciones como NetBean se confeccionarán las interfaces y relaciones en el sistema del banco de sangre para el Hospital de Chamical (nuevo proyecto de Trabajo Final) en la UNCHA, UNLaR.
En lo que respecta al servicio a brindar, se proyecta difundir y ampliar el sitio a toda la zona comprendida por la Misión del Centro Oeste la República Argentina.

7. Bibliografía

1- Dirección de Marketing, Banco de Sangre. Resultado de estudios de mercado, motivaciones y satisfacción de donantes de sangre (4 jun 2013, Internet) http://www.donarsang.gencat.cat/media/upload/pdf/com_sou_els_donants_de_sang_motivacions_satisfaccio_editora_7_6_1_editora_8_12_1.pdf
2- https://www.adventistas.org/es/jovenes/proyecto/vida-por-vidas/
3- Blanca Cisterna, directora del Banco de Sangre de Hemoterapia del hospital Enrique Barros (2018, 17 de octubre).
4- http://www.bloodbankindia.net
5- https://bancosangrerioja.org/
6- http://mutualrh.com.ar/quienes-somos/
7- http://www.protecciondedatos.com.ar/ley25392.htm

Realidad Virtual y Realidad Aumentada aplicadas al Museo Histórico del Colegio Provincial Nº1 "Joaquín V. González" – La Rioja

Raúl Esteban Luna Mercado[1], Marta Elena Cabrera Villafañe[2]

[1] Universidad Nacional de la Rioja.
raulestebanlunamercado@gmail.com, marta.cabreravilla@gmail.com

Resumen El Museo Histórico del Colegio Provincial Nº 1 "Joaquín V. González" de la provincia de la Rioja, atraviesa una etapa de apatía en cuanto a las audiencias por parte de los alumnos y publico en general según los datos registrados por el personal encargado de la gestión, debido al estado de abandono en el que se encuentra el patrimonio histórico. Este trabajo se propone una estrategias para reactivar el funcionamiento del Museo a través de la implementación de recursos educativos emergentes como la Realidad Virtual y/o Realidad Aumentada, permitiendo crear en los asistentes una experiencia enriquecedora.

Palabras claves: Tecnologías emergentes, Realidad Virtual, Realidad Aumentada, educación.

1- Introducción

Realidad Aumentada y Realidad virtual: se trata de dos tecnologías "inmersivas" que ofrecen diferentes maneras de combinar el mundo digital con la realidad, su uso crece en las aulas de todo el mundo.
Un informe reciente del Observatorio de Innovación Educativa (México) asegura que estas herramientas definirán el rumbo de las innovaciones educativas del futuro.[1]

Fig 1. Realidad Aumentada vs Realidad Virtual

Para el presente proyecto se han seleccionados algunos desarrollos como puntos de referencias que se mencionan a continuación:
1: Desarrollo de una versión virtual del Museo de la Telecomunicación Vicente Miralles Segarra en [2].

2: Gilbert Adrián Mengana de la Fé en el 2019 con Realidad Aumentada, una herramienta para la gestión de los valores patrimoniales en [3].
3:Martí Testón Ana en el 2018 con Hacia una museografía 4.0. Diseño de experiencias inmersivas con dispositivos de realidad aumentada en [4], entre otros.
El sistema propuesto se encargará de capturar, procesar y administrar los datos contenidos en la ficha descriptiva sobre el patrimonio afectado a la exposición del Museo Histórico del Colegio Provincial Nº1 "Joaquín V. González" en la ciudad de la Rioja, para luego mediante un motor gráfico mostrar de forma interactiva y atractiva información sobre el objeto aplicando la técnica de Realidad Virtual o Realidad Aumentada. La información a mostrar sobre el objeto capturado podrá ser en formato texto, audio y/o un objeto renderizado mediante un motor gráfico 3D. El sistema solo contendrá una secuencia interactiva con Realidad Virtual y tres secuencias interactivas con Realidad Aumentada con la posibilidad de expansión en el futuro.
Este proyecto tiene como propósito, a su vez, difundir la potencialidad educativa que tiene el museo como también las tecnologías de Realidad Virtual y Realidad Aumentada. Por otro lado, permite una revalorización del patrimonio ayudando a encontrar formas adecuadas para que éste se mantenga vivo y renueve el diálogo con los individuos y la sociedad.

2- Línea de Investigación

La línea de investigación para el presente proyecto se encuentra respaldada bajo la Resolución CICYT N°044/11 del Consejo de Ciencia y Técnica de la Universidad Nacional de La Rioja, derivando para su aplicación las áreas de Agentes y Sistemas Inteligentes – Cibercultura- Realidad Virtual - Realidad Aumentada.

3- Modalidad

De acuerdo con la modalidad establecida en el Reglamento de Trabajo Final de la Carrera de Licenciatura en Sistemas de Información, el desarrollo se considera como un **Trabajo de extensión a la comunidad educativa.**

4- Metodología

Se optó por la metodología de desarrollo ágil Programación Extrema (XP); porque da mayor valor al individuo, a la colaboración con el cliente o usuario, al desarrollo incremental del software con interacciones muy cortas
Para las pruebas del sistema se prevé hacer uso de un teléfono móvil / smartphone con las siguientes características:

- Marca: Samsung
- Modelo: Galaxy J2 Prime SM-G532M
- Procesador ARM Cortex-A53
- GPU ARM Mali-T720 MP2
- Memoria 1,5 GB LPDDR3
- Almacenamiento 8 GB

Las pruebas del sistema también se llevarán a cabo en otros smartphones como Xiaomi Mi 8 Lite, Moto G5, Moto G5 Plus, Samsung Galaxy S3 y Samsung Galaxy J7 pero el dispositivo principal es el detallado anteriormente. Si bien no cuenta con el sensor giroscopio, servirá para probar el rendimiento del sistema en dispositivos de Gama baja.

Para el Software: Para el desarrollo del sistema se utilizará los siguientes programas:
El motor gráfico a utilizar será Unity con las siguientes características:
- Versión 2019.1.11
- Unity Personal

El modelador y animador a utilizar será Blender con las siguientes características:
- Versión 2.80

El editor de código fuente a utilizar será Visual Studio Code con las siguientes características:
- Versión 1.37.1

Para la transferencia de archivos por FTP se hará uso de Filezilla con las siguientes características:
- Versión 3.43

5- Resultados esperados

Con el desarrollo del mismo se espera que la aplicación permita preservar el patrimonio, además de ofrecer al visitante un ambiente atractivo, interactivo y dinámico que le permita un mejor aprendizaje sobre la cultura, arte, tradición e historia que procura presentar el museo.
Con la implementación de este proyecto se pretende mejorar la calidad en los procesos de enseñanza y aprendizaje y en la forma de trasmisión del mensaje del museo para que a los visitantes se los sorprenda con la utilización de esta tecnología novedosa, captando su atención y brindando un mejor servicio.

6- Referencias

1. Alfredo Dillon. Realidad aumentada y realidad virtual: el boom educativo que viene. http://eduprensa.com/realidad-aumentada-realidad-virtual-boom-educativo-viene/
2. VECINA DOLZ, Alejandro (2019). Desarrollo de una versión virtual del Museo de la Telecomunicación Vicente Miralles Segarra. Universidad Politécnica de Valencia. Departamento de Ingeniería Electrónica - Escuela Técnica Superior de Ingenieros de Telecomunicación
3. MENGANA DE LA FÉ, Gilbert Adrián (2019). Realidad Aumentada, una herramienta para la gestión de los valores patrimoniales. Universidad de Oriente. Santiago de Cuba, Cuba. Santiago 149, Mayo-Agosto
4. MARTÍ TESTÓN, Ana (2018). Hacia una museografía 4.0. Diseño de experiencias inmersivas con dispositivos de realidad aumentada. Universidad Politécnica de Valencia. Departamento de Comunicación Audiovisual, Documentación e Historia del Arte.

Grupos Focales en la Elicitación de Requisitos Desarrollo de Videojuego en Realidad Virtual destinado a Usuarios con Parálisis Cerebral

Cintia Ferrarini Oliver[1], Mónica Gonzalez[1], Alicia Aballay[1], Luís Olguín[2]

[1] Departamento de Informática, FCEFyN, UNSJ, [2]Instituto de Informática, FCEFyN, UNSJ.
{ferrarinicintia, gonzalez.monicagilda, prof.alicia, lolguinunsj}@gmail.com

Abstract. En este artículo se presentan resultados de la aplicación de Grupos Focales en la Elicitación de Requisitos para la obtención de un modelo de interacción entre niños con discapacidad (parálisis cerebral) y un videojuego en Realidad virtual, multijugador en línea. El objetivo del videojuego fue favorecer la evolución psicomotriz de los niños que asisten a la Fundación Abril (San Juan, Argentina), cuyas edades cognitivas están comprendidas entre 5 y 10 años. El videojuego se realizó en el marco del Proyecto de Desarrollo Tecnológico y Social (PDTS) denominado MoVeR-T: Entorno de videojuego en Realidad Virtual Telecontrolado para Rehabilitación Motriz, como propuesta de intervención terapéutica alternativa. La aplicación de Grupos Focales permitió recabar información cualitativa significativa para producir ajustes en el diseño del videojuego, a partir de revisiones, reflexiones y aportes de investigadores y expertos, en el marco de un proceso de construcción colaborativa de conocimientos.

Keywords: FocusGroup, Métodos empíricos de investigación, Ingeniería de Software, Ingeniería de Requisitos, videojuego en realidad virtual, tecnología para la salud.

1 Introducción

El grupo focal (GF) es un método empírico de investigación científica, que surge de las Ciencias Sociales. Mediante entrevistas grupales, da lugar a la construcción de espacios de debate focalizado de los que se puede obtener información de tipo cualitativa. [1, 2, 3]. Es un método recomendable para la Elicitación de Requisitos (ER), ya que permite obtener información de tipo cualitativa para la identificación de necesidades y definición de requisitos [3,4,5] .

En este artículo se presentan resultados de la aplicación de Grupos Focales en la Elicitación de Requisitos para la obtención de un modelo de interacción entre niños con discapacidad (parálisis cerebral) y un videojuego en Realidad virtual, multijugador en línea (Multiplayer Online Virtual RealityGame, o MOVER-Game). El objetivo del videojuego fue favorecer la evolución psicomotriz de los niños que asisten a la Fundación Abril (San Juan, Argentina), cuyas edades cognitivas están comprendidas entre 5 y 10 años. El videojuego se realizó en el marco del Proyecto de

Desarrollo Tecnológico y Social (PDTS) denominado MoVeR-T: Entorno de videojuego en Realidad Virtual Telecontrolado para Rehabilitación Motriz, y pretendió constituirse en una propuesta de intervención terapéutica alternativa a las terapias convencionales.

En el proceso de elicitación de requisitos se consideró pertinente indagar y focalizar la atención en aspectos que daban cuenta de la cultura organizacional de la Fundación Abril en la Elicitación de Requisitos. La Cultura Organizacional se la comprende como un: "sistema de creencias y valores compartidos que se desarrolla dentro de una organización o dentro de una subunidad y que guía y orienta la conducta de sus miembros." [5,6]. La cultura organizacional actúa a modo de marco de referencia para los miembros de la organización y señala el rumbo por medio de lineamientos generales acerca de cómo las personas deben conducirse en ésta[7].La Fundación Abril proyecta sus actividades en el ámbito de la salud y educación de personas con discapacidad. Cuenta con un equipo de profesionales y técnicos, agrupados según las necesidades para la atención diferenciada por edades y diagnósticos. En esta institución se visualizó un entramado de interrelaciones entre sus miembros y las familias de las personas con discapacidad, generado por los profesionales a partir de espacios de orientación y asesoramiento continuo al grupo familiar. Si bien la cultura organizacional es un factor de éxito en las organizaciones, es dable destacar que cuando es necesario implementar innovaciones debe ser administrada mediante un proceso de cambio planificado ya que es un elemento organizacional complejo, puesto que una cultura profundamente arraigada, puede ser difícil de cambiar en sus creencias, normas y actitudes.

En el marco de esta investigación, se adoptó la perspectiva del modelo social de discapacidad y se tomó como referencia el término "persona con discapacidad" en lugar de "persona discapacitada". Lo que permite poner en primer lugar a la persona con una característica especial que es la discapacidad. Mientras que el término "persona discapacitada", evoca una condición de la persona que la engloba, la define y la excluye de la sociedad [8]. Desde esta mirada, se afirma que las limitaciones de las personas con discapacidad no son individuales sino de la propia organización social que no presta servicios adecuados ni asegura que las necesidades de estas personas sean tenidas en cuenta[9]. Así, se considera a la discapacidad como una construcción relacional entre la sociedad y un sujeto individual o colectivo. Es decir, toma cuerpo en un espacio situacional, dinámico e interactivo entre alguien con cierta particularidad y la comunidad que lo rodea [10]

Por otra parte, se sabe que los Videojuegos en RV poseen un gran potencial en ámbitos educativo, de ocio y de salud. En [11,12,13,14] se presentan estudios con resultados aceptables respecto de la utilización de VR en el tratamiento de patologías discapacitantes. El mismo autor afirma que uso de tecnología de realidad virtual (VR) aumenta la motivación de los niños y promueve la participación en las actividades propuestas, en comparación con los tratamientos convencionales.

En relación a los grupos focales, cabe destacar que su aplicación en el proceso de elicitación, permitió cumplir el objetivo de crear un modelo de interacción entre niños con parálisis cerebral y Videojuego de RV. También, posibilitó la comprensión de los criterios que debería incluir el desarrollo tecnológico: singularidad, necesidad y expectativas de los destinatarios del videojuego.

A continuación se describe el proceso de implementación de Grupo Focal y sus resultados.

2. Proceso de aplicación de Grupo Focal: Instancias, Fases y Momentos

A continuación se describe el proceso metodológico llevado a cabo en la aplicación de Grupos Focales en la Elicitación de Requisitos. Dicho proceso se desarrolló conforme a las siguientes instancias y momentos, como se presenta en la Figura N°1:

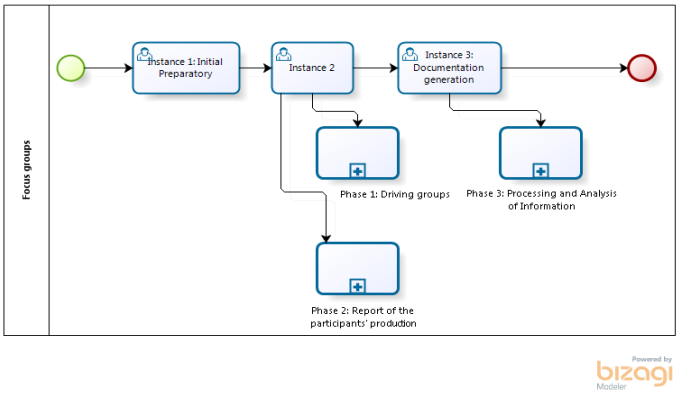


Fig. 1: Modelo del Proceso de Grupo Focal generado en el software Bizagi 2.8

Primera instancia: Inicial Preparatoria

En esta instancia se administraron entrevistas focalizadas a profesionales de las disciplinas Psicología y Kinesiología. A partir del procesamiento cualitativo de la información relevada fue posible extraer indicadores de la cultura organizacional de la Fundación Abril, como también, planificar las fases y momentos de la segunda instancia. Se diseñó el protocolo del debate y los documentos que incluían las consignas de trabajo e interrogantes disparadores y orientadores de la discusión.

Se cursó invitación a los profesionales de la institución demandante señalada. La convocatoria fue abierta para los integrantes de dicha institución y su participación voluntaria.

Segunda Instancia: Implementación de Grupos Focales.

Contextualización del evento: El Encuentro de Expertos, como se lo denominó, se llevó a cabo en la Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional San Juan. En una jornada de trabajo de 4 horas con la participación de 20 profesionales, distribuidos en tres grupos según su perfil. Las moderadoras fueron la primer y segunda autora de este artículo.

Esta instancia constituida por Fase 1 de Conducción de los grupos de discusión desarrollada en tres momentos y Fase 2 que incluyó las acciones desplegadas en el momento IV de Reportes de producción de los participantes

Fase 1: 2.1. Momento I (Duración aproximada 30 minutos)

Las moderadoras comenzaron la sesión con la explicitación del Acuerdo de Confidencialidad; la presentación del proyecto y sus objetivos; la exposición de los objetivos del encuentro, su dinámica de trabajo. La técnica lluvia de ideas y las preguntas disparadoras sobre la temática "Videojuegos de Realidad Virtual", fueron guiando el debate. A partir de preguntas disparadoras para niños y jóvenes con parálisis cerebral de la Fundación Abril.

2.2. Momento II (Duración aproximada 1:30hs): Se llevó a cabo la presentación de la metodología de trabajo. Se conformaron los grupos de participantes según criterios de segmentación estipulados y se distribuyeron las consignas de trabajo elaboradas para cada uno de ellos. Se entregaron carpetas con material impreso en hojas de diferentes colores que identificaban a cada grupo constituido por profesionales de disciplinas afines complementarias.

Los participantes se distribuyeron, según su especialidad, en los siguientes grupos:

1.Neuro-Fisiológico conformado por: Kinesiólogos, Psiquiatras, Profesor de Educación Física.

2.Psico-socio-pedagógico conformado por: Psicólogo, Trabajador Social, Sociólogo, Psicopedagogo, Profesor de Educación Especial, Terapista Ocupacional.

3.Psico-socio-afectivo conformado por: Acompañante terapéutico, Celador, Estudiantes Avanzados en carreras de Educación Especial, Familiares de niños y adolescentes que asisten a la Fundación Abril.

4.Técnico conformado por: Informáticos, Diseñadores Gráficos.

Los integrantes de los grupos constituidos comenzaron a producir sus narrativas conforme a pautas establecidas y expuestas por las moderadoras. Luego se comenzó con el debate e intercambio de experiencias y se elaboraron narrativas con conclusiones. Además del registro de información realizado por un actor externo (observador no participante), se implementó como técnica de captura la grabación de audio.

2.3.Momento III (Duración aproximada 1:30hs)

En este momento se procedió, mediante una técnica de dinámica grupal (del caracol) con el proceso reconfiguración de los grupos a fin de generar nuevos espacios de discusión y construcción colectiva de conocimientos, conforme a una guía con pregustas semiestructuradas. Luego se procedió a la construcción del:

- Árbol de Problemas: Producción grupal; Identificación, exposición de los problemas, consenso.
- Árbol de Soluciones o de Objetivos; Identificación de soluciones, consenso, conclusiones finales.

Fase 2: Reportes de la producción de los participantes

2.4. Momento IV (Duración aproximada 20 minutos)

Consistió en la recolección de la producción de los participantes; despedida / cierre; agradecimiento a los profesionales que participaron en el Encuentro de Expertos. Comunicación de la utilización de los datos y reafirmación del acuerdo de Confidencialidad. Se concluyó con la entrega de las certificaciones correspondientes.

Tercera Instancia: Generación de documentación

Fase 3 procesamiento y análisis de la información. Memorias del debate: recopilación de conclusiones. Reporte de resultados.

3. Resultados

El proceso de aplicación de Grupos Focales fue valorado positivamente por los investigadores, moderadoras, observador no participante y profesionales participantes en la "Reunión de Expertos".

El proceso de discusión y construcción colectiva de conocimientos guiado por las moderadoras, se llevó a cabo en un clima positivo de producción, evitando que los participantes se sintieran incómodos compartiendo con otros sus teorizaciones y pensamientos. Cada grupo de trabajo centró su atención en la especificidad de su área e intercambió opiniones y comentarios buscando acuerdos a partir de las estrategias propuestas por las moderadoras.

La aplicación de grupos focales en el proceso de elicitación permitió cumplir el objetivo de crear un modelo de interacción entre niños con ECNE y Videojuego de RV y posibilitó la comprensión de los criterios que debería incluir el desarrollo tecnológico: singularidad, necesidad y expectativas de los destinatarios del videojuego.

Del análisis e interpretación de la información relevada en cada instancia, fases y momentos del proceso de aplicación del método Grupo Focales, fue posible realizar una categorización de aspectos a tener en cuenta en la modelización de los procesos de interacción entre los usuarios mencionados y el videojuego de realidad virtual. Cada categoría cuenta con indicadores que permiten su valoración y que fueron consideradas en la evaluación del videojuego en Realidad Virtual desarrollado, como se señala a continuación en la Figura 1.

Tabla 1. Indicadores de Calidad del videojuego.

Categorías	Indicadores
Satisfacción	Gustos y preferencias del jugador con respecto a la dinámica y estética del videojuego. (Observación de gestos, expresiones verbales, opiniones de las personas de su círculo más cercano)
Aprendizaje	Facilidad para comprender la mecánica del videojuego, para interactuar con los objetos, reglas y acciones que este propone. (Tiempo de respuesta ante el

	estímulo acorde a su patología y posibilidad de realización.)
Efectividad	Captación de la atención (En un tiempo acorde con su patología)
Inmersión	Capacidad para involucrarse en el juego, permanecer realizando las actividades propuestas (Conexión del sujeto con el videojuego)
Motivación	Capacidad para mantener un determinado comportamiento que permita alcanzar los objetivos. (Se intentará propiciar emociones positivas que ayuden al usuario a persistir en el videojuego)
Emoción	Capacidad para expresar lo que el videojuego provoca (Alegría, tristeza, frustración, enojo, etc.)
Socialización	Capacidad para comunicarse con otros a través del videojuego (Por el momento sólo se consideró la comunicación con el terapeuta o especialista de la salud con quien interactúa en lo cotidiano.)

Una vez establecido el marco general en base a lo relevado durante la aplicación del método antes mencionado se generó un documento de requisitos del software. En la Figura 2 se muestra los aspectos a tener en cuenta en el proceso de desarrollo del videojuego.

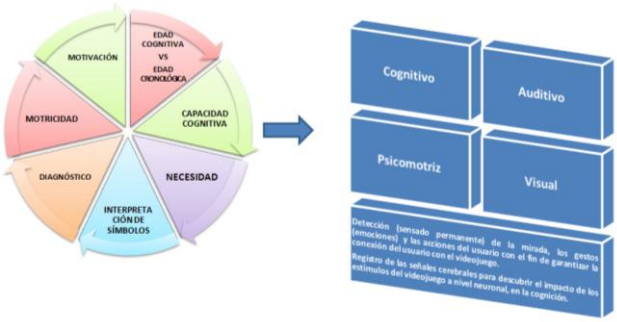


Fig. 2. Aspectos implicados en el diseño de un videojuego destinado a niños con parálisis cerebral.

A continuación se describen cada uno de los aspectos mencionados en la Figura 2.

Visual: Uso de pocos colores, sólidos y pasteles; texturas pronunciadas que permitan diferenciar un plano de otro; pocos objetos de gran tamaño y bien delimitados; que resulten familiares para el niño. Es decir, que materialicen los símbolos básicos cotidianos de su entorno (por ejemplo: tasa, manzana, cuchara, etc.) que contrasten con el fondo.Textos de gran tamaño, avatares o pictogramas (como

sistema alternativo o aumentativo de comunicación), que puedan emitir gestos, emociones, sonidos y señas, con las cuales el usuario se pueda identificar.

Auditivo: Lenguaje hablado para la emisión de consignas, constituido por oraciones breves, claras, concisas y sencillas. Las actividades deben ser transmitidas en forma pausada y en lo posible musicalizadas con sonidos suaves, nítidos, con las voces de sus profesores y de sus familiares (personas con quien ha establecido un vínculo).

Psicomotriz: Uso de dispositivos físicos de gran tamaño (por ejemplo palancas y/o botones) y softwares adaptados a fin de detectar movimientos mínimos dirigidos de aquellos que son involuntarios. Estos deben permitir realizar acciones con la cabeza, los ojos y/o una extremidad a fin de favorecer la interacción entre el usuario y el videojuego.

Conclusiones y Reflexiones finales

Respecto al método utilizado, se prevé aplicar nuevamente Grupos Focales en el Proyecto de Desarrollo Tecnológico y Social (PDTS) denominado "Videojuegos en realidad virtual como intervención terapéutica alternativa para estimular la cognición en niños con encefalopatía crónica no evolutiva" (Vivitan), siguiendo las mismas instancias propuestas en este caso, en la Elicitación de requisitos para el desarrollo de videosjuegos en Realidad Virtual.

En relación al proceso de discusión o de debate, se desarrollo en un marco organizado, sistemático y de libertad, sin restringir/ incidir en las intervenciones de los participantes.

La aplicación de Grupos Focales generó la necesidad de producir ajustes en el diseño de videojuegos, a partir de revisiones y reflexiones de los investigadores y aportes de expertos, en un espacio interdisciplinario de construcción colaborativa de conocimientos.

En cuanto a los requisitos tecnológicos funcionales del videojuego en realidad virtual se definió que el software debía:

- Permitir la creación de escenarios 3D colaborativos y en línea multiusuario (o multijugador)

- Estar basado en la plataforma de videojuegos Unity 3D.

- Permitir que la interacción en línea se lleve a cabo a través de dos servidores en la nube: Gamesparks (para almacenar las interacciones) y Photon Engine (para el trabajo sincronizado y colaborativo)

References

1 Mendoza-Moreno, M., González-Serrano, C., & Pino, F. J. Focus group como proceso en ingeniería de software: una experiencia desde la práctica. Dyna, 80(181), (2013), 51-60. https://revistas.unal.edu.co/index.php/dyna/article/view/36969

2 Kontio, J., Bragge, J. and Lehtola, L., The Focus Group Method as an Empirical Tool in Software Engineering.Springer London, London.(2008)

3 Kasirun, Zarinah Mohd, and Siti Salwah Salim. "Focus group discussion model for requirements elicitation activity." 2008 International Conference on Computer and Electrical Engineering. IEEE, 2008.

4 Onwuegbuzie, A. J., Dickinson, W. B., Leech, N. L.,& Zoran, A. G. . Un marco cualitativo para la recolección y análisis de datos en la investigación basada en grupos focales. Paradigmas, 3, 127-157. Disponible en: Dialnet-UnMarcoCualitativoParaLaRecoleccionYAnalisisDeDato-3798215.pdf

5 Focus group discussion model for requirements elicitation activity. In Computer and Electrical Engineering, 2008. ICCEE 2008. International Conference on (pp. 101-105). IEEE.

6 Ferreira, Miguel AV. Prácticas sociales, identidad y estratificación: tres vértices de un hecho social, la discapacidad. *Intersticios. Revista sociológica de pensamiento crítico*, 2007, vol. 1, no 2.

7 Chiavenato, Idalberto, "Introducción a la Teoría General de la administración"(7a. Ed)., Ed. Mc.Graw-Hill, México, 2006,pág. 319.

8 Pedraza-Álvarez, Lilibeth et al. Cultura organizacional desde la teoría de Edgar Schein: estudio fenomenológico. Clío América, [S.l.], v. 9, n. 17, p. 17-25, ene. 2015. ISSN 2389-7848. Disponible en: http://revistas.unimagdalena.edu.co/index.php/clioamerica/article/view/1462

9 Palacios, A. El modelo social de discapacidad: orígenes, caracterización y plasmación en la Convención Internacional sobre los Derechos de las Personas con Discapacidad. Primera edición Colección cermi.es editadas por Ediciones Cinca, S.A. Madrid. (2008)

10 Palacios, A.. El modelo social de discapacidad y su concepción como cuestión de derechos humanos. [Editorial]. Revista Colombiana de Ciencias Sociales, 8(1), pp. 14-18. DOI: http://dx.doi.org/10.21501/22161201.2190. (enero-junio, 2017)

11 Howcroft, Jennifer, et al. "Active video game play in children with cerebral palsy: potential for physical activity promotion and rehabilitation therapies." *Archives of physical medicine and rehabilitation* 93.8 (2012): 1448-1456.

12 Ustinova K. I., Ingersoll C. D., & Cassavaugh N.. "Short-term practice with customized 3D immersive videogame improves arm-postural coordination in patients with TBI," in 2011 International Conference on Virtual Rehabilitation (ICVR), pp. 1–7.(2011)

13 Raina P., O'Donnell M., Rosenbaum P., Brehaut J., Walter S. D., Russell D., Swinton M., Zhu B., & Wood E., "The health and well-being of caregivers of children with cerebral palsy,"Pediatrics, vol. 115, no. 6, pp. e626–e636. 2005

14 Biswas, P., & Robinson, P. . Modelling user interfaces for special needs. Accessible Design in the Digital World (ADDW). Disponible en :http://publicaciones.unitec.edu.co/ojs/index.php/PAR/article/view/20/16,2008

Modelo de Proceso incorporando Requisitos de Usabilidad y Experiencia de Usuario en el Desarrollo de Software.

Iván Balmaceda Castro[1], Luis Fuentes[1], Juan Magaquian[1]
Carlos Salgado[2], Mariela Fernandez[1], Sandra Rodriguez[3], Monica Gimenez[3]

[1]Sede Regional Chamical. Universidad Nacional de La Rioja
ibalmaceda89@gmail.com,
[2]Departamento de Informática Facultad de Ciencias Físico-Matemáticas y Naturales
Universidad Nacional de San Luis
[3]Departamento Académico de Ciencias Exactas, Físicas y Naturales
Universidad Nacional de La Rioja

Abstract. La Experiencia de Usuario (UX) puede considerarse como el proceso que lleva a cabo el usuario cuando interactúa con un producto. Como en cualquier desarrollo, el diseño, también debe centrarse en el usuario, ser usable y considerar las características y habilidades de los usuarios a la hora de interaccionar con la interfaz gráfica. Así mismo, la usabilidad es considerada hoy en día cómo un atributo de calidad fundamental de los mismos. Tanto la Usabilidad como la UX, no puede reducirse al estudio de lo que hace el usuario, además es necesario comprender por qué lo hace. En este sentido, se introdujeron requisitos de usablidad en un modelo de proceso en cada una de sus etapas del ciclo de vida.

Keywords: Modelo de Proceso, Experiencia de Usuario, Usabilidad, Requisitos.

1. Introducción.

El desarrollo de sistemas de información se ha transformado en un proceso que busca construir aplicaciones útiles y correctas para su uso. Uno de los grandes objetivos de la Ingeniería de Software (IS) es construir aplicaciones de calidad, útiles a los usuarios finales, aplicando distintos métodos y principios. Por su parte el diseño centrado en el usuario surge como un enfoque y método que consiste en conocer algunas particularidades del usuario, con el objetivo de hacer más familiares y efectivas las interfaces graficas que se diseñan para él. Esto permite encontrar una solución visual y funcional a diversos requerimientos comunicativos y funcionales en sistemas informáticos.

Dado que cada vez es mayor el uso de dispositivos, es importante planificar soluciones acordes a las necesidades de los usuarios asegurando el acceso y la usabilidad de las mismas.

El presente trabajo se enmarca dentro del proyecto de investigación “Modelo de Proceso y Evaluación Centrado en el Usuario incorporando requisitos de Usabilidad y Experiencia de Usuarios” - Centro de Investigación y Desarrollo Informático, Departamento Académico de Ciencias Exactas, Físicas y Naturales - Sede Regional Chamical Universidad Nacional de La Rioja. Cod. Proy. 27PIN/C0003, financiado por Consejo de Investigaciones Científicas y Tecnológicas UNLaR Res N°052/18. Y

con la colaboración del Proyecto de Investigación: Ingeniería de Software: Conceptos, Prácticas y Herramientas para el desarrollo de Software con Calidad – Facultad de Ciencias Físico-Matemáticas y Naturales, Universidad Nacional de San Luis. Proyecto Nº P-031516.

2. Diseño Centrado en el Usuario (DCU)

Mayormente, los sistemas son utilizados por personas a las que llamamos "usuarios", y generalmente, en el ciclo de vida de un software, estos, muy pocas veces suelen "contar" para los diseñadores y desarrolladores. Una vez llegado el momento de la puesta en funcionamiento del mismo, se lo involucra; por tanto su interacción se ve en la fase final de un proyecto cuando ya poco puede hacerse en su beneficio.
El Diseño Centrado en el Usuario o DCU, es el proceso de diseño de interfaces de usuario dando mucha atención a las necesidades, aspiraciones y limitaciones de los usuarios del producto diseñado en cada estado de su proceso. Se caracteriza por su proceso de múltiples etapas para resolver problemas que no solo requieren que los diseñadores analicen y prevean cómo los usuarios van a emplear el sistema, sino también que prueben y validen sus supuestos con respecto al comportamiento y experiencia del usuario en pruebas reales. Para ello, propone realizar un diseño basado en el pensamiento del usuario, convirtiéndole en el punto central del desarrollo, tratando de conseguir productos fáciles de usar, efectivos y eficientes. [1]
El estándar internacional ISO 13407, describe cómo un proceso de Diseño Centrado en Usuario (DCU) puede ser usado para conseguir sistemas usables. Este, provee un marco de trabajo para aplicar técnicas en el diseño y evaluación del DCU. La Usability Professionals' Association (UPA) lo define como un enfoque de diseño cuyo proceso está dirigido por información sobre las personas que van a hacer uso del producto. Los procesos DCU se centran en los usuarios a través de la planificación, el diseño y el desarrollo de un producto, a su vez especifica los tipos de actividades que deben realizarse, no obstante no pide técnicas o métodos particulares, tan solo las recomienda. [2, 3].
El DCU ayuda a los diseñadores de software a cumplir los objetivos del desarrollo, los requerimientos de los usuarios son considerados desde el principio del proceso y todo el ciclo de desarrollo del producto. Los requerimientos son refinados a través de varios métodos: Estudios etnográficos, investigación del contexto, pruebas con prototipos, pruebas de usabilidad y otros métodos.

3. Experiencia de Usuario y Usabilidad: Ingeniería de Usabilidad.

La Experiencia de Usuario (UX por sus términos en inglés User eXperience) puede considerarse como el proceso que lleva a cabo el usuario cuando interactúa con un producto. Se involucran factores sociales, culturales y de contexto que pueden generar una percepción positiva o negativa del producto [4, 5, 6, 7, 8].
UX es entendido como dinámico, da el estado emocional y cambiante de una persona y las diferencias en las circunstancias durante y después de una interacción con un

producto, buscando resolver el problema de la utilidad del producto en la satisfacción de su uso. [7, 9, 10, 11]

En cuanto a la Usabilidad, J. Preece, define como que son "Sistemas fáciles de usar y de aprender" [26]. La ISO 9241-11, dice que es "la medida en la que un producto se puede usar por determinados usuarios para conseguir objetivos específicos con efectividad, eficiencia y satisfacción en un contexto de uso especificado".

A su vez, el término de Ingeniería de la Usabilidad (IU) fue adoptado por profesionales de usabilidad de Digital Equipment Corporation [12]. Se comenzó a utilizar este término para referirse a los conceptos y técnicas para planificar, conseguir y verificar objetivos de la usabilidad de sistema. Como idea principal, se sostenía que los objetivos "medibles" de usabilidad deben ser definidos pronto en el desarrollo del software y después evaluarlos repetidamente durante el desarrollo para asegurar que se han conseguido [13][14].

Muchas son las ventajas que la usabilidad puede proporcionar y por ello debería ser tratada cómo un factor de calidad estratégico y relevante [15]. Para poder Diseñar un producto con una alta calificación respecto a su usabilidad, ésta debe ser considerada en todas las fases del desarrollo, desde el momento en que este comienza hasta el momento en que el producto o servicio es puesto en disposición del público.

La norma ISO/IEC 9126-1 [16], considera a la usabilidad como un parámetro de calidad del software y es una de las características de calidad relevante del mismo. El objetivo es satisfacer las necesidades del usuario a través de seis características principales o atributos que son: funcionalidad, fiabilidad, utilidad, eficiencia, mantenibilidad y portabilidad. Esta norma contempla la calidad interna, externa y en uso de un producto de software [17].

Mientras que la experiencia de usuario está ganando más interés en el campo de desarrollo de productos de software, por lo que es necesario emplear métodos de evaluación UX apropiados para evaluarlo en diferentes productos de software [8][18].

Es así que, la UX y la Usabilidad no puede reducirse únicamente al estudio técnico de lo que realiza el usuario. Por consiguiente, su medición no debe quedar en una mera descripción de lo que hace. Es necesario comprender por qué lo hace para determinar su experiencia. Para ello, es necesario saber cómo se sienten las personas cuando usan un sistema, considerando la usabilidad, la satisfacción, entre otros. La motivación y expectativas del usuario tienen un papel muy importante en la experiencia del usuario en lugar de la usabilidad tradicional. [11, 19, 20].

Los modelos de IU propuestos distan demasiado de los modelos de Ingeniería de Software (IS), con lo cual el desarrollador lo ve como un gran cambio radical en su metodología de trabajo y continua como siempre lo hizo. [13][14]. Por esta razón, el estudio de la interacción y de la usabilidad del software ha tomado un carácter relevante y debe ser tenido en cuenta por parte de la IS.

4. Requisitos en la Usabilidad

Un requisito de software es la capacidad que debe alcanzar o poseer un sistema o componente de un sistema para satisfacer un contrato, estándar, especificación u otro documento formal [21]. También, podemos definir como que un requisito de software

es la propiedad que un software desarrollado o adaptado debe tener para resolver un problema concreto.
Estos requisitos tienen atributos, los cuales son información complementaria que se almacena para la gestión posterior de los mismos. Éstos surgen generalmente en el momento de la especificación de los requisitos. Uno de los atributos importantes de las características de los requisitos es poder verificarlos a través de la "validación de los requisitos". Ésta es una propiedad intrínseca, por la que es posible comprobar que el sistema de software en forma integral o a través de alguno de sus componentes, satisface el requisito con el que fue especificado inicialmente. [22]

4.1 Requisitos No Funcionales

Según Ian Sommerville [23] define el concepto de requerimientos no funcionales como que: "son las limitaciones sobre servicios o funciones que ofrece el sistema. Incluyen restricciones tanto de temporización y del proceso de desarrollo, como impuestas por los estándares".
Establece también que los requerimientos no funcionales se aplican al sistema como un todo y no a características individuales del sistema en sí. Los requerimientos establecen características necesarias del sistema y determinan las funcionalidades que debe poseer el mismo.
Los requerimientos no funcionales suelen ser más críticos que los requerimientos funcionales, dado que la ausencia de uno de los mismos puede inutilizar al sistema.
Los Requerimientos No Funcionales están muy relacionados a los atributos de calidad del sistema y ciertas condiciones (restricciones) que debe cumplir el mismo para brindar las funcionalidades requeridas para las que fue diseñado. Cuando se habla de restricciones o condiciones, el término se refiere a las limitaciones que enfrenta el Ingeniero de software para llevar a cabo el desarrollo del sistema. Cuando se expresa el término calidad, dentro del ámbito de los requerimientos no funcionales, se refiere a todas aquellas características o cualidades que el cliente o usuario requiere del sistema para sentirse satisfecho con el mismo [24].
Se puede establecer que los requerimientos no funcionales surgen de:

1) Requerimientos del Producto: Que especifican el comportamiento o limitaciones del software, cuestiones tales como: seguridad, confiabilidad, eficiencia y usabilidad.
2) Requerimientos organizacionales: son requerimientos que se relacionan con normativas, procedimientos y políticas organizacionales. Se originan por políticas organizacionales, procedimientos internos de la organización, involucrando aspectos tales como: cuestiones operativas y organizacionales, lenguaje de desarrollo de los aplicativos, estándares a cumplir, metodología de modelado para desarrollo de software, etc.
3) Requerimientos externos: se refiere a requerimientos regulados por entidades externas a la organización y a su proceso de construcción (leyes o cuestiones éticas, interoperabilidad con otros sistemas, etc.).

Los requerimientos no funcionales no se deben proponer como metas generales, sino más bien como aspectos cuantitativos que puedan ser medidos y evaluados como por

ejemplo: la rapidez, el tamaño, la facilidad de uso, la fiabilidad, la robustez y la portabilidad. Los Requerimientos No Funcionales no se pueden considerar separados de los requerimientos funcionales del sistema, sino que se deben especificar cuando se desarrolla el proyecto, ya que son inherentes a la calidad del mismo [25].

5. Modelo de Proceso.

El DCU tiene etapas y fases que se suceden y se retroalimentan, es clave a la hora de planificar y llevar a cabo el diseño de un sistema interactivo. El cual permite detectar errores y desviaciones lo más temprano posible, e incorporar a los usuarios finales en todas las etapas del diseño.
El modelo de Proceso que proponemos, está basado en una serie actividades que añade a la Ingeniería de Software.

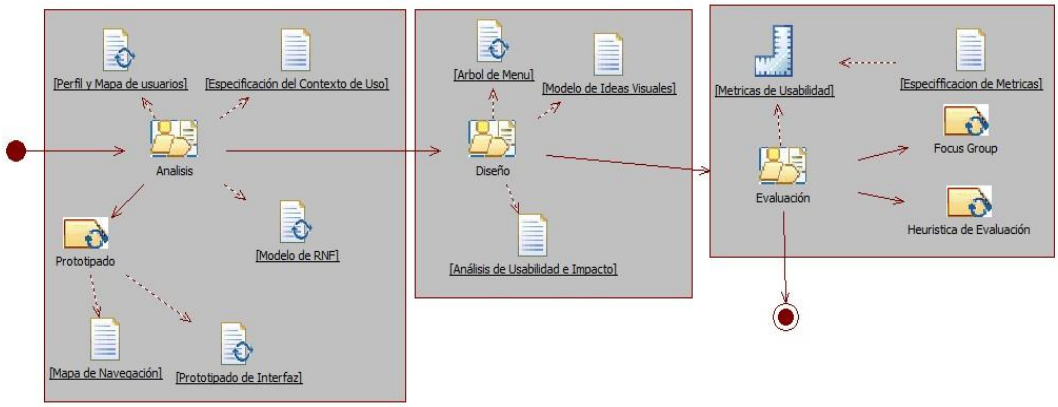


Fig.1 Muestra las diferentes fases en que se divide el Modelo de Proceso y como deben realizarse cada una de ellas.

5.1 Explicación del Modelo

- Análisis
    - Perfil y Mapa de Usuario: Comprende la recopilación de información de cada usuario con respecto a su experiencia de trabajo, nivel de educativo, edad, experiencia con el uso de computadoras, contexto de trabajo y social, entre otros.. Para de esta manera, familiarizarnos con sus actividades. Así mismo, se obtiene una visión mucho más amplia para la especificación de requisitos, de ésta manera permitiendo verificar todas las partes involucradas.

Este análisis nos sirve para adaptar el software para los futuros usuarios.

- Especificación del contexto de uso: Permite obtener las tareas que los usuarios realizarían, como así también el hardware, software y recursos necesarios para el normal funcionamiento.
- Casos de Uso Esenciales (CUE): Complementa a los Casos de Uso de la IS, propone un enfoque a procesos del problema a modelar, centrándose en las interacciones entre el usuario y las interfaces gráficas, mediante las respuesta que ofrece el usuario. Los CUE, describen procesos en términos de actividades esenciales.
- Prototipado: Versión reducida del sistema completo, para representar las funcionalidades que tendrá el producto, pero no ejecuta nada en realidad. Bocetos del sistema o interfaces que el usuario aspira utilizar.

- Diseño
  - Arbol de Menú: Permite definir la organización de los menús de las interfaces.
  - Análisis de Usabilidad e impacto: Identificación de problemas de usabilidad del sistema actual. Está técnica ayuda a justificar un esfuerzo de rediseño.
  - Modelo de Ideas Visuales: Modelo descriptivo y explicativo que se enfoca en los objetos y acciones de la interfaz que también permite captar las ideas a los usuarios.
- Evaluación
  - Métricas de Usabilidad: Establece los objetivos de usabilidad y diferentes niveles de rendimiento para los atributos..
  - Focus Group: En un lapso de tiempo determinado, un grupo de usuarios observa para encontrar temas de carácter relevante con respecto a la usabilidad,
  - Heuristica de Evaluación: Se basa en la experiencia de un experto o varios expertos en usabilidad que observan la interfaz para recopilar opiniones con respecto a lo bueno y malo.

6. Conclusión

En el presente trabajo se han analizado y comprendido los conceptos sobre que es la Experiencia de Usuario, Usabilidad, el Diseño Centrado en el Usuario y qué son los requerimientos y las actividades que involucran. Esto nos permitió definir e incluir criterios desde el punto de vista de la Ingeniería de Software donde se propone un modelo de proceso y evaluación centrado en el usuario, que incorpora aspectos y requisitos de usabilidad y experiencia de usuario en forma temprana durante el proceso de construcción del software.

Con el mismo se pretende disminuir costos de reproceso y mejorar la experiencia en usabilidad tempranamente para que los usuarios puedan interactuar con ellos de la forma más fácil, cómoda e intuitiva posible.

Referencias

1. Constantine, L.L., Loockwood L.A.D., "Software for Use: A practical guide to the Models and Methods of Usage-Centered Design". Addison-Wesley (1999).
2. Hassan, Y; Ortega Santamaria, S. (2008) Informe APEI sobre Usabilidad. APEI (versión online: http://www.nosolousabilidad.com/manual/3.htm).
3. UPA. Usability Professionals' Association. What is User-Centerd Design? versión online http://www.upassoc.org/usability_resources/about_usability/what_is_ucdhtml
4. Martin, G. Gaetaán, V. Saldaño, A. Pires, G. Miranda, S. Villagra, A. Carrizo, C. Cardozo, H. Sosa "Un enfoque integrador para diseñar y evaluar interfaces de usuario web" XIX Workshop de Investigadores en Ciencias de la Computación (WICC 2017, ITBA, Buenos Aires), 2007
5. ISO 9241-11, Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts, International Organization for Standardization, 2018.
6. J. Nielsen, "Usability inspection methods," in Conf. Human factors in computing systems, Boston, Massachusetts, United States, 1994.
7. J. Nielsen, and R. Molich, "Heuristic Evaluation of User Interfaces", in Proceeding SIGCHI '90 Conference on Human factors in Computing Systems, pp.249-256, 1990.
8. N. Bevan, "International Standards for HCI and Usability", International journal of human-computer studies, vol. 55, no. 4, pp. 533-552, 2001.
9. H. Hartson, and T. Andre, "Criteria for Evaluating Usability Evaluation Methods" In International Journal of Human-Computer Interaction, pp. 373-410, 2003.
10. ISO/IEC 9126-1, Software Engineering - Product Quality, 2001.
11. E. Sosa Bruchmann, G. Montejano, A. Garis "Análisis de la experiencia del usuario: relación entre el comportamiento emocional y la satisfacción de uso" XVII Workshop de Investigadores en Ciencias de la Computación, 2015.
12. Good M., Spine T.M., Whiteside J., George P., "User-derived impact análisis as a tool for usability engineering". In Proceedings of Human Factors in Computing Systems: CHI'86. New York: ACM (1986)
13. Bennet J., "Managing to meet usability requirements. In Visual Display Terminals: Usability Issues and Health Concerns". Prentice-Hall (1984).
14. Gilb T., "The Impact Análisis Table applied to human factors design". In Proceedings of the 1st IFIP Conference on Human-Computer Interaction-INTERACT '84 (Amsterdam) (1984)
15. Constantine, L.L., Loockwood L.A.D., "Software for Use: A Practical guide to the Models and Methods of Usage-Centered Design". Addison-Wesley (1999)
16. Norma ISO/IEC ISO9126-1 "Software Enginnering – Product Quality – Part 1" 2001.
17. Nigel Bevan "Quality and Usability: A new framework" Achieving Software product quality, 1997.
18. N. Bevan, J. Carter and S. Harker, "ISO 9241-11 revised: What have we learnt about usability since 1998?", in Human-Computer Interaction: Design and Evaluation, Lecture Notes in Computer Science, vol. 9169. M. Kurosu, Ed. Springer, pp. 143-151, 2015.
19. J. Nielsen, "Usability engineering", 1 edition ed., San Diego, CA 92101, UnitedStates: Academic Press, 1993.
20. Carreras, J. (2003). "Midiendo la experiencia de usuario".
21. IEEE (1990), *IEEE Standard glossary for software engineering terminology*.
22. SWEBOK, "IEEE Computer Society Professional Practices Committee, "Guide to the Software Engineering Body of Knowledge – 2004 Version"," Los Alamitos – California USA, 2004. Consultado: 09/05/2013

23. Ian Sommerville, *Ingeniería de Software*, Novena ed. México: Pearson Educación, 2011.
24. Salvador Sánchez, Miguel Ángel Sicilia, and Daniel Rodríguez, *Ingeniería del Software. Un Enfoque Desde La Guía SWEBOK*. México: Alfaomega, 2012.
25. I. Sommerville, *Ingeniería del Software. Septima Edición*. Madrid: Pearson Educación, 2005.
26. PREECE, J. (1994): Human-Computer Interaction. New York: Addison-Wesley. Academic Press, Inc.

Data Mining aplicado a la deserción estudiantil en la Facultad de Ciencias Exactas de la UNSa

Claudia Ibarra, Cristian Martínez

Departamento de Informática - Universidad Nacional de Salta,

Salta, 4400. Argentina

RESUMEN

Resumen. Es preocupante el porcentaje de deserción estudiantil en las universidades argentinas. El objetivo de esta investigación fue utilizar técnicas de Data Mining para obtener modelos predictivos que permitan identificar, a los estudiantes con mayor riesgo de deserción y los factores de abandono. Se utilizó la regresión logística binaria para predecir la probabilidad de la continuidad o no en los estudios de los estudiantes y el método de árboles de clasificación, para obtener un modelo que permita predecir los factores que propician la deserción de los estudiantes, basado en los datos disponibles. En este trabajo se utilizó datos correspondientes a la base histórica de los períodos 2015, 2016 y 2017, de los alumnos de la facultad de Ciencias Exactas de la Universidad Nacional de Salta y el lenguaje de programación R para la implementación de los métodos.

Palabras claves: deserción, árboles de clasificación, regresión logística binaria, R.

1 INTRODUCCIÓN

El alto porcentaje de deserción estudiantil es uno de los principales problemas que enfrentan las instituciones de educación superior.

En Salta, según los resultados de Estadísticas Universitarias Argentinas de la Secretaría de Políticas Universitarias del Ministerio de Educación de la Nación (SPU), durante el período 2005 – 2015, la tasa promedio de crecimiento anual de egresados de títulos de pregrado y grado es -1.7 [1].

En este trabajo, se propone analizar la problemática referida a la deserción, en las carreras de la Facultad de Ciencias Exactas, a través de la aplicación de distintas técnicas cuantitativas y cualitativas del Data Mining (o Minería de Datos). De esta manera lograr obtener modelos predictivos que permitan identificar, a los estudiantes con mayor riesgo de deserción, identificar los motivos de abandono, con el fin de aportar a las autoridades una guía que oriente la toma de decisiones preventivas, en pos de mejorar la situación de deserción.

La Minería de Datos es la aplicación de los métodos de aprendizaje y estadísticos para la obtención de patrones y modelos. Tiene por objetivo analizar los datos para producir nuevo conocimiento que pueda utilizar el usuario. Esto se realiza construyendo un modelo basado en los datos recopilados para este efecto [2].

2 METODOLOGÍA DE LA INVESTIGACIÓN

La presente investigación es propositiva ya que se fundamenta en la necesidad de reducir el índice de deserción estudiantil en la Facultad de Ciencias Exactas y tratar de orientar a los directivos a tomar mejores decisiones al implementar acciones para superar la problemática actual. Es experimental, pues utilizará un conjunto de datos correspondientes a la base histórica.

2.1 CASO DE ESTUDIO

La Universidad Nacional de Salta tiene su sede central de gobierno y administración en la ciudad de Salta. Está integrada por Facultades, Departamentos, Sedes Regionales, Institutos de Investigación y Establecimientos de Enseñanza Preuniversitaria.

La Facultad de Ciencias Exactas es una de las seis facultades de la Universidad. Está integrada por los Departamentos de: Física, Química, Matemática e Informática. En la misma se dicta quince carreras de grado en sus distintas Unidades Académicas.

2.2 PREPARACIÓN DE DATOS

Los datos analizados corresponden a los alumnos ingresantes a la Facultad en el año 2015. La principal fuente de datos la constituyen los registros correspondientes al periodo académico del año 2015 hasta 2017, obtenidos a través del sistema de gestión de alumnos de la Universidad. Inicialmente se rescataron 35 atributos.

Con los datos integrados, se realizó un análisis estadístico de las variables seleccionadas, para tener una mayor comprensión del registro e identificar las propiedades típicas de los datos. Luego se procedió con la limpieza de los datos a fin de eliminar problemas que afecten la calidad de los mismos. Finalmente, se realizaron las transformaciones necesarias en algunas de las variables seleccionadas.

Se obtuvo un repositorio compuesto por 780 registros y 19 atributos.

2.3 Regresión logística

La regresión logística es un método lineal que intenta modelizar la probabilidad de ocurrencia de un evento. La variable dependiente es categórica dicotómica o policotómica, a los efectos de facilitar la interpretación [3].

Se construyó el modelo de predicción considerando las siguientes variables:

- Variable dependiente: Estado, variable con valor igual a 1 si el estudiante continúa sus estudios ó valor igual a 0 si no continúa sus estudios.
- Variables independientes (o predictores): Carrera, Regulares (cantidad de materias regulares, Edad (al iniciar su carrera), EstMadreR (nivel de estudio de la madre), SitLaboralPadre (situación laboral del padre).

El ajuste del modelo de regresión logística se realizó dividiendo los datos en una proporción de 70:30, obteniendo dos conjuntos: conjunto

de entrenamiento (70%) y conjunto de prueba (30%). El conjunto de entrenamiento (train) se utilizó para capacitar el modelo de regresión logística en la fórmula dada, el que luego se validó usando el conjunto de prueba (test).

2.4 Árboles de decisión

La segunda técnica elegida para el diseño de un modelo predictivo fue el árbol de decisión de tipo clasificación, debido a que es una técnica predictiva y los datos disponibles para el desarrollo del modelo corresponden a variables categóricas y discretas, que se ajustan a las características de un árbol de decisión, además por la factibilidad de interpretar la información obtenida de forma gráfica [2].

Para construir un árbol de decisión es necesario definir una función que relaciona una variable categórica dependiente (clase) con n variables independientes que pueden ser categóricas o numéricas. En nuestro caso se trabajó con una variable clase, dependiente denominada 'Estado', que indica la continuidad o no en los estudios de un estudiante. Esta variable es de carácter dicotómica que toma valor 1 si el estudiante continúa sus estudios ó 0 si no continúa.

Las variables predictoras, independientes, fueron las variables Carrera, Regulares (cantidad de materias regulares), Edad (edad al iniciar su carrera), EstMadreR (nivel de estudio de la madre), SitLaboralPadre (situación laboral del padre).

El siguiente paso realizado, fue definir la muestra de entrenamiento y la muestra de prueba en una proporción 70:30. Luego se diseñó y graficó el modelo, a partir de las variables predictoras. Finalmente se analizó su capacidad predictiva.

2.5 Resultados obtenidos

Se describen a continuación los resultados obtenidos aplicando el método de regresión logística y árboles de clasificación, utilizando el

lenguaje de programación R.

Regresión Logística

Se utilizó la función genérica summary(), del lenguaje de programación R, para obtener un resumen de los coeficientes y de todo el proceso de ajuste de nuestro modelo. Se muestran a continuación las salidas del programa, del modelo de estudio, con formula1= Estado~ Carrera + Regulares + Edad + EstMadreR + SitLaboralPadre y utilizando los datos de entrenamiento. El modelo estimado fue:

$$Y = -2.30792 + 0{,}02416\ _{\text{Carrera}} + 1.18535\ _{\text{Regulares}} + 0.06229\ _{\text{Edad}} + (-0.05381)\ _{\text{EstudioMadreR}} + (-0.01779)\ _{\text{SituacionLaboralPadre}}$$

```
Call:
glm(formula = formula1, family = "binomial", data = train)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.69798  -0.81890   0.01286   0.78877   1.79027

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)    -2.30792    0.61206  -3.771 0.000163 ***
Carrera         0.02416    0.02109   1.145 0.252008
Regulares       1.18535    0.12972   9.138  < 2e-16 ***
Edad            0.06229    0.01905   3.269 0.001078 **
EstMadreR      -0.05381    0.05820  -0.925 0.355179
SitLaboralPadre -0.01779    0.16682  -0.107 0.915092
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 771.04  on 556  degrees of freedom
Residual deviance: 551.16  on 551  degrees of freedom
AIC: 563.16

Number of Fisher Scoring iterations: 6
```

Fig. 1. Resumen del modelo (datos de prueba)

Del resumen obtenido (Figura 1), se observa que las variables ‘Regulares’ y ‘Edad’ son estadísticamente significativas, la primera señala una fuerte asociación de la cantidad de materias ‘Regulares’ del estudiante con la probabilidad de continuar sus estudios.

En nuestro modelo el logaritmo de los odds de que un estudiante continúe con sus estudios está positivamente relacionado con las

materias 'Regulares' (coeficiente de regresión = 1.18535). El exponencial del coeficiente de regresión no es más que el odds ratio entre dos individuos que se diferencian en una unidad de la variable independiente (predictora). Esto implica que, por cada unidad que se incrementa la variable 'Regulares' (materias regularizadas), le corresponderá un odds ratio de continuar sus estudios más de tres veces mayor (3.27183) que otro estudiante, siempre que las variables restantes sean iguales en ambos estudiantes. Las variables 'Carrera', 'EstMadreR' y 'SitLaboralPadre', no son estadísticamente significativas, siendo la última la menos significativa de todas.

La evaluación de la bondad de ajuste del modelo elegido se realizó con el estadístico *razón de verosimilitud* (Likelihood ratio) resultando un valor p de p-value: 1.57074590596499e-45, lo que indica que el modelo sí es significativo. Se obtuvo además el número Pseudo R^2 de McFadden = 0.2851815, valores entre 0.2 y 0.4 indica un muy buen ajuste del modelo.

La matriz de confusión de los datos de entrenamiento, mostró que de un total de 551 observaciones, 427 fueron predicciones correctas y un total de 130 predicciones incorrectas. Esto nos da una precisión de predicción de 77% y un error de predicción de 23%. La matriz de confusión de los datos de prueba, mostró un total de 168 predicciones correctas y un total de 54 predicciones incorrectas. Esto nos da una precisión de predicción de 76% y un error de predicción de 24%.

Finalmente se graficó la curva ROC y se calculó el AUC (área debajo de la curva) sobre los datos de entrenamiento, el cual resultó igual a 0.7602941, lo que demostró una buena capacidad predictiva.

Árboles de decisión

Se diseñó el modelo de árboles de decisión a partir de las variables predictoras definidas. La Figura 2, muestra la gráfica del modelo de árbol de clasificación obtenido.

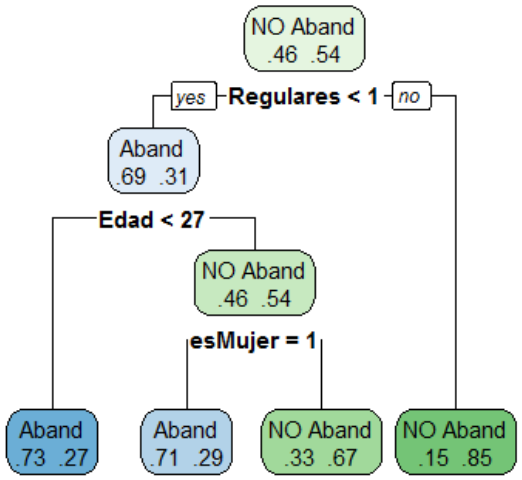


Fig. 2. Modelo de Árboles de Clasificación

En concreto, el Modelo genera 4 reglas de decisión. De estas, 2 conducen al abandono de los estudios y dos al No abandono de los estudios.

Si el alumno regularizó menos, o sea si no regularizó ninguna materia, y su edad es menor que 27, tiene un 73 % de probabilidad de abandonar sus estudios.

Si el alumno regularizó menos de una materia, es mayor que 27 años y es mujer, tiene un 71 % de probabilidad de abandonar sus estudios.

Si el alumno regularizó menos de una materia, es mayor que 27 años y no es mujer, tiene un 33 % de probabilidad de continuar sus estudios.

Si el alumno regularizó por lo menos una materia, tiene un 73 % de probabilidad de 15 % de no abandonar sus estudios

3 CONCLUSIONES

En la búsqueda de un modelo que sea capaz de predecir si un alumno continuará o no sus estudios se observó que las variables más significativas, en el modelo de regresión logística y en el de arboles de clasificación, resultaron ser: 'Regulares' (cantidad de materias regularizadas) y 'Edad' (edad del estudiante al iniciar su carrera), mientras que las variables relacionadas a indicadores económicos no resultaron ser relevantes para ninguno de los modelos.

La importancia que demuestra la variable 'Regulares' en ambos modelos, se condice con la realidad de alumnos ingresantes universitarios.

Los modelos obtenidos resultaron ser significativos para el análisis del caso de estudio, perfilando los factores que inciden en la deserción de alumnos de la facultad de Ciencias Exactas.

4 REFERENCIAS

[1] SPU, Estadística Universitaria Argentina. 2015. Disponible en: http://informacionpresupuestaria.siu.edu.ar/DocumentosSPU/diu/Anuario%202015%20-%20ESTAD%C3%8DSTICAS%20UNIVERSITARIAS%20ARGENTINAS%20-%20SPU%20.pdf

[2] Orallo, J. et al., *Introducción a la Minería de Datos*. Madrid: Pearson Prentice Hall, 2004.

[3] Britos, P., *Minería de Datos*. Buenos Aires: Nueva Librería. 2005.

[4] Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, and Keying Ye. *Probability & Statistics for Engineers & Scientists. Prentice Hall / Pearson Education*, Inc., Boston, MA, 9th edition, 2012. ISBN 10: 0-321-62911-6.

[5] José Luis Cañadas Reche Regresión logística. *Tratamiento computacional con R*. Junio 2013

Lista de autores

——/ A /——

Aballay, Alicia 195, 339
Aballay, Laura 61, 75
Aciar, Silvana 61
Alfonzo, Pedro 177, 237
Alonso, Nancy 215, 225
Amaro, M. 265
Aranda, César Omar 107
Aranda, Juan 225
Azcurra, D. 265

——/ B /——

Balmaceda Castro, Iván 347
Becerra, María del Carmen 195, 275
Berlles, Julio Jacobo 153
Bordón, José M. 237
Buemi, María Elena 153

——/ C /——

Cabrera Villafañe, Marta 255, 335
Campos, Matías 137
Candelero, Diego 331
Castro Rojas, Luis Fernando 297, 311
Castro, Isabel 49
Catania, Carlos 141
Chavez, Susana 321
Ciccarelli, Soledad 285
Cuero, Edwin Moreno 297
Cúneo, Juan 225, 255

——/ D /——

De Armas, Adrián 13
De La Jara, Facundo 245
Diaz Ogás, Mathias 215
Dávila, Gabriel 39
Díaz, Daniel 225, 255
Díaz, Mathias 61

——/ E /——

Eraso, Martín 293
Espitia Peña, Esperanza 297, 311

——/ F /——

Fernandez, Mariela 347
Ferrarini Oliver, Cintia 339
Flores, Lorena 185
Fuentes, Luis 347

——/ G /——

Gallardo, Juan 177
Gamba, Cristian 289
Garay, Leonardo 285
Garcia, Sebastian 141
Gel, Matías 159, 163
Gimenez, Monica 347
Giménez, Rosana 27, 289, 293
Gonzalez, Mónica 339
Gonzalez, Rodrigo 137
Guardia, Fernando 225
Guerra, Jorge 141
Guevara Atencio, Miguel José 127
Gutiérrez, Laura Viviana 245
Gómez, Pablo 117

——/ H /——

Herrera, Myriam 93

——/ I /——

Ibarra, Claudia 355
Ibañez, Francisco 225
Iturraspe, Rodolfo 159, 163
Izetta, Javier 205

——/ L /——

Lovos, Edith 75
Luconi, Eleonora 289
Luna Mercado, Raúl Esteban 335
Lund, María Inés 39, 49, 93

——/ M /——

Magaquian, Juan 347
Mamani, Enrique 285
Marisa, Haderne 289
Mariño Sonia 185
Mariño, Sonia 177, 237
Martin, Adriana 321
Martins, Sebastián 185
Martín Villareal, Emilio Izarra 159, 163
Martínez, Cristian 355
Martínez, Roxana 167
Marín, María Alejandra 289
Masanet, Maria Isabel 245
Migani, Silvina 1, 39
Molina, Ana 215
Montilla, Andrés 311
Moyano, Ezequiel 159, 163
Murazzo, María 61, 117, 127
Méndez-Garabetti, Miguel 27, 293

——/ O /——

Olguín, Luís 339

Oliva, Elisa 215
Orona, Priscila 331
Orozco, Carlos Ismael 153
Ortega, Manuel Oscar 245
Otazú, Alejandra 255
Oviedo, Sandra 225, 255

——/ **P** /——

Pacheco, Gabriel 285
Palau, Franco 141
Pandolfo, Pablo 13
Parkinson, Christian 167
Paz, Fabiola Patricia 205
Pinciroli, Fernando 83

——/ **R** /——

Rearte, Emilio 331
Rigaki, Maria 141
Rivadulla, Ariel 285
Rodriguez, Sandra 255, 331, 347
Rodríguez, Nelson 117, 127, 321
Rodríguez, Rocío 167
Romagnano, María 195
Rosenstein, Javier 137
Rossi, Bibiana 13

——/ **S** /——

Salgado, Carlos 347
Sandez, Pablo 27
Sarmiento, Adriana 195
Serra, Santiago 27

——/ **U** /——

Urciuolo, Adriana 159, 163

——/ **V** /——

Valdivieso, Eleonora 289
Vera, Cristina 1
Vera, Pablo 167
Vieira, Jhonatan 163
Villareal, Martín 159, 163

——/ **Z** /——

Zapata, Sergio 1
Zarate, Pedro 275
Zumel, Raquel 289
Zárate Alvarez, Nicolás 137