RETOS DE LA INVESTIGACIÓN EN INGENIERÍA DE SISTEMAS:

APLICACIONES, HERRAMIENTAS Y DESARROLLOS

SELLO EDITORIAL UNIVERSITARIO Americana

CORPORACIÓN UNIVERSITARIA AMERICANA
INSTITUCIÓN UNIVERSITARIA

RETOS DE LA INVESTIGACIÓN EN INGENIERÍA DE SISTEMAS: APLICACIONES, HERRAMIENTAS Y DESARROLLOS

Compilación y Dirección Editorial
Mg. Jovany Sepúlveda Aguirre
Director Editorial y de Publicaciones – Sede Medellín

Libro resultado de investigación a partir de la actividad colaborativa y de cohesión entre los grupos de investigación de la Corporación Universitaria Americana y diferentes grupos de investigación del ámbito nacional e internacional.

620.0011
C822

Corporación Universitaria Americana. (2018). Retos de la investigación en Ingeniería de Sistemas: aplicaciones, herramientas y desarrollos. Jovany Sepúlveda-Aguirre (Comp.). Medellín: Sello Editorial Coruniamericana.

173 Páginas: 16X23 cm.
ISBN: 978-958-5512-03-0

1. SOFTWARE - 2. GESTIÓN DEL CONOCIMIENTO - 3. SEGURIDAD INFORMÁTICA - 4. REDES NEURONALES - 5. APLICACIONES EMPRESARIALES.

CORPORACIÓN UNIVERSITARIA AMERICANA-CO /SPA /RDA
SCDD 21 /CUTTER - SANBORN

Corporación Universitaria Americana©
Sello Editorial Coruniamericana©
ISBN: 978-958-5512-03-0

Sello Editorial Coruniamericana
selloeditorialcoruniamericana@coruniamericana.edu.co

Diagramación y carátula:
EDUARDO A. MURILLO PALACIO

1ª edición: Octubre de 2018

PARES EVALUADORES

Mg. John Anthony Jaramillo Betancur
Megister en Ingeniería Universitat Oberta de Catalunya.
Docente Investigador. Fundación Universitaria María Cano.

Mg. Silvia Marcela Henao Villa
Ingeniera Informática. Magister en Educación.
Ph(c). en Educación por Competencias.
Instructora Cisco Networking Academy.
Docente Universitaria adscrita a la
Fundación Universitaria María Cano.

CONTENIDO

10 Descripción de los elementos asociados a un software web, orientado al desarrollo de Planes Individuales de Ajustes Razonables (PIAR) para población con discapacidad
Jovany Sepúlveda Aguirre
César Felipe Henao Villa
David Alberto García-Arango
Elkin Darío Aguirre-Mesa
Gustavo Andrés Araque-González
Christian Hernán Obando Ibarra

23 Buenas prácticas de seguridad informática para mitigar las amenazas persistentes avanzadas en el sector salud en Colombia
Yexid Montenegro García
Christian Hernán Obando Ibarra
Gloria Amparo Lora
Enevis Rafael Reyes Moreno

47 Caracterización de elementos para la creación de una herramienta computacional para la gestión del conocimiento en las organizaciones
Diana María Montoya Quintero
Jovani Alberto Jiménez Builes

70 Líneas de producción de software para la construcción de un sistema de biblioteca, a través de frameworks basados en componentes
César Felipe Henao Villa
David Alberto García Arango
Jovany Sepúlveda Aguirre
Elkin Darío Aguirre Mesa
Gustavo Andrés Araque González
Christian Hernán Obando Ibarra

84 Metodologías de detección de intrusos (IDS) basadas en anomalías de red aplicando redes neuronales SOM y GHSOM
Eduardo de la Hoz Correa
Johan Mardini Bovea
Emiro de la Hoz Franco

115 ETL (extracción, transformación y carga) o intercambio de información entre aplicaciones empresariales con el fin de crear un prototipo
Yexid Montenegro García
Christian Hernán Obando Ibarra
Hugo Eduardo Pérez Muñoz
Enevis Rafael Reyes Moreno

139 Implicaciones de la neutralidad de la red y el diseño para la inclusión en las políticas del Sistema Nacional de Competitividad, Ciencia, Tecnología e Innovación en Colombia
David Alberto García Arango
Jovany Sepúlveda Aguirre
César Felipe Henao Villa
Elkin Darío Aguirre Mesa
Gustavo Andrés Araque González
Laura Isabel Bedoya Corrales

157 Uso de tecnologías inmersivas en educación: realidad aumentada, realidad virtual y smartroom
Mauricio Hincapié Montoya
Christian Andrés Díaz León

PRESENTACIÓN

Descripción de los elementos asociados a un software web, orientado al desarrollo de Planes Individuales de Ajustes Razonables (PIAR) para población con discapacidad

Jovany Sepúlveda Aguirre[1]
César Felipe Henao Villa[2]
David Alberto García Arango[3]
Elkin Darío Aguirre Mesa[4]
Gustavo Andrés Araque González[5]
Christian Hernán Obando Ibarra[6]

Resumen

Se presentan los resultados de una plataforma web como solución de seguimiento de lo que en el decreto 1421 se propone como Planes Individuales de Ajustes Razonables (PIAR) que en su esencia desde la ley se constituye en

1 Magister en Gestión de la Innovación Tecnológica, Cooperación y Desarrollo Regional. Investigador Junior integrante del Grupo de Investigación AGLAIA de la Corporación Universitaria Americana. ORCID: https://orcid.org/0000-0002-1047-6673. E-mail: jasepulveda@americana.edu.co.

2 Docente-Investigador del grupo AGLAIA - Corporación Universitaria Americana. Ingeniero de Sistemas de la Universidad Nacional de Colombia Sede Medellín, magíster en entornos virtuales de aprendizaje. Correo electrónico: chenao@coruniamericana.edu.co . ORCID: https://orcid.org/0000-0001-7426-2589

3 Docente-Investigador del grupo AGLAIA - Corporación Universitaria Americana. Licenciado en Matemáticas y Física de la Universidad de Antioquia, Magíster en Matemáticas Aplicadas de la Universidad EAFIT. Escuela de Ciencias, departamento de ciencias Matemáticas, doctorando en Educación de la Universidad Nacional de Rosario – Argentina. Correo electrónico de contacto: dagarcia@coruniamericana.edu.co . ORCID: https://orcid.org/0000-0002-0031-4275

4 Docente- Institución Universitaria Pascual Bravo. Ingeniero de sistemas de la Fundación Universitaria María Cano, Magíster en Gestión de la Tecnología Educativa de la Universidad de Santander. Correo electrónico: elkin.aguirre@pascualbravo.edu.co. ORCID: https://orcid.org/0000-0003-2521-6003

5 Docente-Investigador Corporación Universitaria Americana. Ingeniero Industrial, con especialización en Gestión Logística Integral, magíster en Ingeniería de Producción con énfasis en transporte y logística. Correo electrónico: garaque@americana.edu.co. ORCID: https://orcid.org/0000-0001-8627-8924.

6 Ingeniero en electrónica y telecomunicaciones, Especialista en seguridad informática y Magister en tecnologías de la información y comunicación. Director del Programa de Ingeniería de Sistemas de La Corporación Universitaria Americana.

una "herramienta utilizada para garantizar los procesos de enseñanza y aprendizaje de los estudiantes, basados en la valoración pedagógica y social, que incluye los apoyos y ajustes razonables requeridos, entre ellos los curriculares, de infraestructura y todos los demás necesarios para garantizar el aprendizaje, la participación, permanencia y promoción. Son insumo para la planeación de aula del respectivo docente y el Plan de Mejoramiento Institucional (PMI), como complemento a las transformaciones realizadas con base en el Diseño Universal de Aprendizaje (DUA)." (Ministerio de Educación Nacional, 2017). Es importante resaltar que además de presentar aspectos de diseño e implementación del software en su etapa inicial, se plantean igualmente perspectivas de avance para una segunda etapa del software en el marco de un análisis de las implicaciones de este como agente articulador TIC del sistema educativo, las regulaciones nacionales y la praxis docente (Bourdieu, 1996) desde un enfoque de adaptaciones curriculares para la inclusión de estudiantes con discapacidad en el marco de la administración curricular. Con el presente escrito, se pretende realizar una descripción de aspectos relacionados con un sistema de información desarrollado por la Corporación Universitaria Americana en Medellín denominado PIAR web el cual esta direccionado al seguimiento de lo que el Decreto 1421 propone como Planes Individuales de Ajustes Razonables (PIAR) para población con discapacidad.

Palabras clave: PIAR, software, inclusión educativa, diseño universal de aprendizaje.

Introducción

El consejo económico y social de las Naciones Unidas, propuso dentro de su agenda a 2030, los Objetivos de Desarrollo Sostenible (ODS), los cuales contienen metas para el ámbito mundial que deben obtenerse favoreciendo un entorno propicio para la vida, la sostenibilidad y sustentabilidad. En sus objetivos, vale la pena resaltar los siguientes: "4. Garantizar una educación inclusiva, equitativa y de calidad y promover oportunidades de aprendizaje durante toda la vida para todos" y "10. Reducir la desigualdad entre los países". (Naciones Unidas, 2016, pág. 4). Analizando a profundidad lo que allí se plantea, implica que, aunque en los Objetivos de Desarrollo del Milenio (ODM) se propuso la importancia de lograr la enseñanza primaria universal, esta tarea quedó incompleta, en tanto que, para afrontar los retos propuestos en la época

actual, se hace necesario, no solo ampliar cobertura, sino mejorar la calidad y acoger a las minorías y a personas con discapacidad. Es así como en informes finales de obtenidos de los ODM, se plantea que:

> con base en las lecciones aprendidas de los ODM, las intervenciones tendrán que adaptarse a las necesidades de grupos específicos de niños, y en particular de niñas, de niños que pertenecen a minorías y de comunidades nómadas, de niños que participan en el trabajo infantil y de aquellos que viven con discapacidades, en situaciones de conflicto o en zonas urbanas marginales (Naciones Unidas, 2015).

Se puede interpretar entonces que, hay una tendencia creciente a incluir a personas con discapacidad en el sistema educativo. La creciente tendencia de la población mundial hacia la cantidad de personas con discapacidad está creciendo, es así como "…debido al envejecimiento de la población -las personas ancianas tienen un mayor riesgo de discapacidad- y al incremento global de los problemas crónicos de salud asociados a discapacidad, como la diabetes, las enfermedades cardiovasculares y los trastornos mentales." (Organización Mundial de la Salud, 2011, pág. 8). De hecho, en Colombia, se identifica que "…la evolución de la matrícula de estudiantes con NEE ha aumentado ostensiblemente, de 12.417 matriculados en 1996 a 62.092 en 2004. Resulta interesante observar el incremento de centros que integran a estudiantes con NEE, de 12,8% en 1996 a 59,6% en 2004" (Samaniego de García, 2009, p. 272).

Por estas y otras razones, es necesario considerar la importancia de fortalecer el sistema educativo con herramientas tecnológicas y objetivos estratégicos que propendan por el mejoramiento de la información, seguimiento y atención a las personas con discapacidad.

En el presente artículo, se muestran los aspectos relacionados con la investigación desarrollada para la concepción, diseño e implementación de un software orientado a la formulación, apoyo y control de los planes individuales de ajustes razonables (PIAR) propuestos en el Decreto 1421 de 2017 del Ministerio de Educación Nacional en una institución educativa del municipio de Caldas en los niveles de básica primaria, básica secundaria y media.

El capítulo, se presenta en los cuatro momentos principales: políticas de inclusión educativa en Colombia, metodología de elaboración e investigación

respecto al software, resultados del desarrollo e implementación del software, implicaciones prácticas y conclusiones de la investigación.

La metodología de desarrollo de investigación se realizó con un enfoque mixto, donde por medio del análisis de las necesidades del entorno de la Institución y entrevistas a los actores representativos e intervinientes en el sistema educativo, se levantaron los requisitos y se propuso un diseño que posteriormente fue llevado a la institución como prototipo funcional, para que luego de ser validado pueda ser utilizado en el grueso de la población de la Institución.

El proceso de investigación, presentó como conclusión fundamental la importancia de propiciar la implementación de este tipo de herramientas como una forma de apoyar en la medida de lo posible el proceso de seguimiento a las adaptaciones curriculares para atender a estudiantes con necesidades educativas especiales.

Políticas de Inclusión Educativa en Colombia

Las políticas de inclusión educativa en Colombia se han venido desarrollando desde el momento en que se genera el enfoque adoptado por la Asamblea General de las Naciones Unidas el 13 de diciembre de 2006, que posteriormente se hace visible en la Ley 1346 de 2009 para finalmente materializarse en la Ley 1618 de 2013 "Por medio de la cual se establecen disposiciones para garantizar el pleno ejercicio de los derechos de las personas con discapacidad" (Ministerio de Educación Nacional, 2013). Es desde ese momento en que se van desarrollando estrategias que propenden por incluir a la población con discapacidad en los entornos áulicos regulares tanto a nivel privado como oficial.

Siguiendo un rastreo de la efectividad de estas políticas para el período 2009-2017, se evidencia que aún deben materializarse estos esfuerzos, de tal suerte que sean efectivos y eficientes desde un enfoque de inclusión para el aprendizaje y para la convivencia en el marco de la promoción social para este sector de la población. En la búsqueda de la efectividad, el Ministerio de Educación Nacional (MEN), promulgó el Decreto 1421 del 29 de agosto de 2017, el cual hace operativo el ámbito de la ley para el caso educativo "Por el cual se reglamenta en el marco de la educación inclusiva la atención educativa

a la población con discapacidad" (Ministerio de Educación Nacional, 2017). En el Decreto, se reglamenta la prestación del servicio educativo a cinco años (2018-2022), para niños, niñas, adolescentes y jóvenes de educación primaria, secundaria y media con cualquier tipo de discapacidad.

Los parámetros a través de los cuales el Decreto opera para comprometer al sector educativo en el acceso a educación inclusiva, pertinente y de calidad desde los primeros meses de vida, se resumen en cinco aspectos:

> 1. Condiciones para el acceso sin barreras, la permanencia y procesos educativos de calidad para la continuidad educativa. 2. Oferta educativa pertinente. 3. Herramientas pedagógicas e institucionales. 4. Corresponsabilidad de actores (entidades territoriales certificadas, colegios, rectores, docentes) para su implementación. 5. Plan progresivo de implementación (Ministerio de Educación Nacional, 2017).

En este mismo documento, se propone que desde el acompañamiento docente:

> Los docentes de aula estarán acompañados y asesorados por un docente de apoyo con experiencia en la atención de población con discapacidad y educación inclusiva. Estos docentes de apoyo capacitarán a sus pares de aula y realizarán visitas de seguimiento a la implementación de las prácticas de enseñanza y articulación de las prácticas pedagógicas para todos los niños y jóvenes (Ministerio de Educación Nacional, 2017).

Metodología

Los aspectos metodológicos se fundamentaron en dos componentes, uno científico y otro técnico. Desde el aporte científico, se realizó una investigación con enfoque mixto, en la cual, mediante una interpretación de las necesidades de una institución educativa pública en la cual se llevó a cabo el desarrollo, se realizaron entrevistas a los actores del proceso de tal forma que se identificaron los casos de uso y el alcance del desarrollo.

El desarrollo del software se realizó al interior de la industria de software que nace en la Corporación Universitaria Americana, donde se atendieron los criterios de Concepción, Diseño, Implementación y Operatividad del software según las necesidades identificadas. Previo al desarrollo, se realizó un análisis de viabilidad en el cual se obedecieron las etapas de análisis de los siguientes aspectos: estado de desarrollo, políticas de financiamiento, aspectos regulatorios, mecanismos de propiedad intelectual y análisis de mercado. Las etapas anteriores se desarrollaron en el marco del software como un resultado de investigación que se puede clasificar en la etapa de prototipado que cumple las características de Producto Mínimo Viable (PVM) para ser llevado a implementación en la institución como mecanismo de transferencia tecnológica. Para los aspectos regulatorios se obedece al marco normativo de diseño para todos o diseño para la inclusión según las normas establecidas en el ámbito colombiano. El análisis del impacto del software se realizó desde la utilización de técnicas discursivas y el análisis de las modificaciones de hábitos docentes que implican la adaptación al uso de este en el quehacer del aula. La población beneficiada hasta el momento han sido 124 estudiantes con discapacidad, clasificados en 14 categorías diferentes de la Institución Educativa José María Bernal del municipio de Caldas. La elección de la institución se hizo considerando que cuenta con alrededor de cuatro mil estudiantes, lo cual hace que la población de estudiantes con necesidades educativas especiales requiera de el apoyo de este tipo de herramientas de software, puesto que el personal de apoyo psicopedagógico no es suficiente para tantos estudiantes.

En cuanto a los aspectos técnicos, se inició a través del lenguaje PHP la aplicación PIARweb, utilizando el framework de laravel puesto que al usarlo se trabaja de una manera más eficiente en la creación de una aplicación web. Para cumplir con el conjunto de requisitos se realizaron varias entrevistas, llevando a diseñar un sistema que consiste en una aplicación web con un backend e implementado características integradas de Laravel como Composer, Eloquent, Blade y Artisan y de servidor XAMPP. La aplicación web se creó utilizando la versión 5 del framework de Laravel, que tiene como objetivo hacer que el desarrollo de PHP sea más fácil, más rápido y más intuitivo. La aplicación web fue construida siguiendo el patrón de arquitectura MVC. La figura 2 presenta una ventana con las categorías de discapacidad que se tratan en la institución, desde allí pueden consultarse, editarse, eliminarlas o crear nuevas.

Resultados

Como hallazgos y resultados de la investigación y desarrollo del software, se crearon paneles de administración para actualizar y dar una administración intuitiva a todos los módulos.

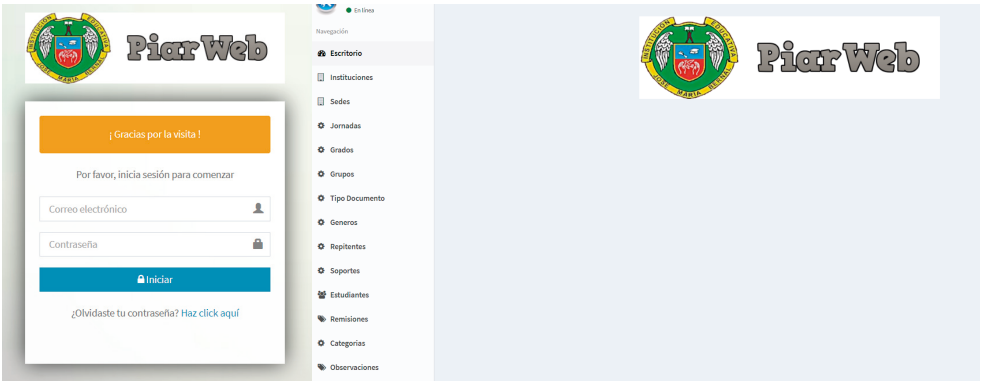


Figura 1. Menú de acceso (izquierda) y menú de navegación (derecha) – Perfil de administrador

En la figura 1, se puede observar el menú de acceso con contraseña según el perfil y luego de ingresar, presenta una ventana con un menú desplegable en el cual se pueden administrar todos los procesos asociados al software.

El software permite guardar los datos del estudiante y dar un diagnóstico psicológico. También se puso a disposición una interfaz pública para que los usuarios registrados puedan iniciar sesión y validar la información. Se crearon perfiles de administrador, docente y estudiante. La aplicación es fácilmente escalable y las características se pueden agregar o eliminar de una manera fácil gracias a la capacidad de Laravel de administrar paquetes a través de Composer. Los resultados demostraron que la Aplicación PiarWeb creada en Laravel 5 es efectivamente una opción de un framework en PHP que ayuda a los desarrolladores a crear rápidamente aplicaciones web seguras y actualizables.

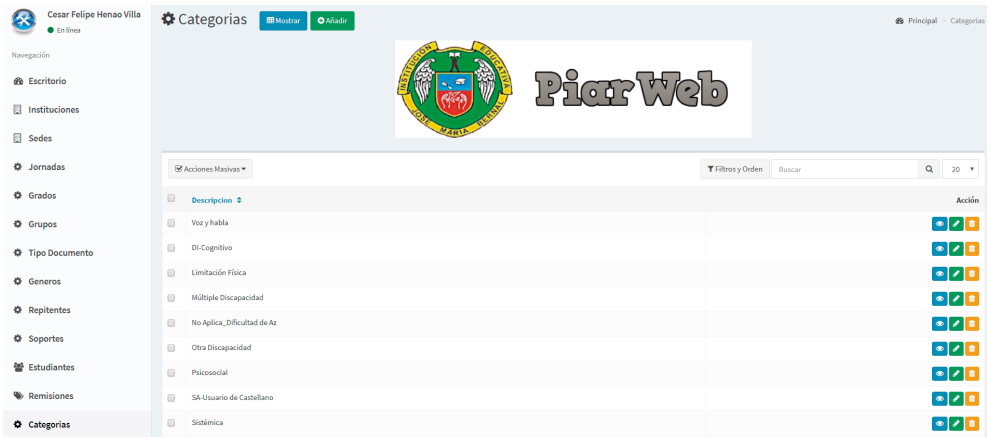


Figura 2. Sección del software en la cual se consultan las categorías de discapacidad.

Mediante una entrevista semiestructurada, se identificaron los aspectos de impacto del software respecto a la implementación los Planes Individuales de Ajustes Razonables, presentando un resultado positivo en su primera etapa, la cual está relacionada con la administración de la información por parte de las docentes de apoyo y psicoorientación, mostrando que el software tiene un gran potencial respecto a la promoción, prevención y tratamiento de situaciones asociadas a los planes de inclusión de población en situación de discapacidad.

Es así como se llevó a cabo la elicitación de requerimientos, respecto a este concepto, Zowghi y Coulin (2005) lo definen como "el proceso de buscar, descubrir, adquirir y elaborar requerimientos para sistemas basados en computador". Los autores plantean la actividad de elicitación de requerimientos como un proceso complejo que como tal implica el relacionamiento de entes con propiedades de conexión, interdependencia, diversidad, adaptabilidad, dependencia de la ruta elegida y emergencia (no reducible a sus partes individuales).

De esta forma puede inferirse que efectivamente, el mecanismo funcional de la ingeniería de requisitos es en principio de corte cualitativo para, posteriormente generar indicadores preferiblemente cuantitativos del desempeño

del sistema. No está de más resaltar que las dos actividades para las cuales la tabla anterior plantea el posible uso de todas las técnicas son "Análisis de los interesados" y "Elicitación de requerimientos". Lo cual sugiere una interesante relación subjetiva en la toma de requerimientos.

Respecto a las tareas propuestas para la elicitación de requisitos, Durán y Bernárdez (2000), proponen una serie de tareas relacionadas con la elicitación de requisitos que se concretan en un producto denominado *Documento de Requisitos del Sistema* (DRS). Las tareas que proponen son:

Tarea 1: Obtener información sobre el dominio del problema y el sistema actual.
Tarea 2: Preparar y realizar las reuniones de elicitación/negociación.
Tarea 3: Identificar/revisar los objetivos del sistema.
Tarea 4: Identificar/revisar los requisitos de almacenamiento de información.
Tarea 5: Identificar/revisar los requisitos funcionales.
Tarea 6: Identificar/revisar los requisitos no funcionales.
Tarea 7: Priorizar objetivos y requisitos (Toro y Jiménez, 2000).

Se identifica que, si bien es cierto, las tareas están relacionadas con la intencionalidad del productor de software, requieren de la información que provee el adquisidor de software.

Figura 3. Ventana de flexibilización

Con base en la elicitación, como puede observarse en la figura 3, el software se desarrolló para la escritura de flexibilizaciones curriculares propuestas por las docentes de apoyo para ser tenidas en cuenta por el docente de grado o en su defecto ser reevaluadas en el marco de su práctica docente. Los resultados obtenidos de la entrevista a la psicóloga y docentes de apoyo, permiten identificar que se han logrado los objetivos planteados en la primera etapa, puesto que se agiliza el proceso de identificación y recepción de casos relacionados con discapacidad en el aula, estos casos pasarían desapercibidos en los entornos áulicos de no haber una estrategia TIC de divulgación segura y confiable que posibilite involucrar y comprometer a todos los actores del proceso.

Implicaciones prácticas

Con base en los resultados obtenidos de la implementación del software en su primera etapa, se identifica la posibilidad de escalarlo de tal forma que se consolide como una solución de seguimiento, evaluación y control de los planes de apoyo que complementan el diseño curricular, se tiene un camino por recorrer en términos de su apropiación por parte de la planta docente de la institución, no obstante, se tiene una gran prospectiva respecto a las implicaciones en la gestión de la información de educación para la inclusión. El desarrollo del software que en principio fue concebido como una solución a la ausencia de docentes de apoyo, se ha transformado en una herramienta viable de implementación para la ley de inclusión educativa, que posibilita el conocimiento y posterior intervención de procesos de inclusión en los entornos educativos.

La implementación del software, abre la posibilidad de repensar el papel de las Tecnologías de la Información y la Comunicación (TIC) en los entornos áulicos y cómo éstas deben migrar hacia la consolidación de comunidades de conocimiento y el empoderamiento para la participación y la colaboración en comunidades de aprendizaje. De aquí se plantea el paso de las TIC a las Tecnologías para el Aprendizaje y la Comunicación (TAC), y posteriormente migrar hacia las Tecnologías para el Empoderamiento y la Participación (TEP), como puede verse en Soler Fernández (2016).

Conclusiones

La concepción del desarrollo del software PIARweb, pasó por un proceso de verificación y validación de las necesidades institucionales en el marco de la norma legal colombiana de inclusión escolar, para tal efecto, el diseño ha sido pensado desde el protocolo propuesto por la ley, constituyéndose así en un Producto Mínimo Viable que resuelve una necesidad de carácter social relacionada con la administración de ajustes curriculares e información relacionada con las características y tratamiento de estudiantes con discapacidad de tal forma que se facilite su inclusión en el aula.

El desarrollo de un sistema de información de Planes Individuales de Ajustes Razonables (PIAR), es necesario en la medida en que el apoyo de TIC a la administración curricular, posibilita una comunicación ubicua (anytime/anywhere) (Sakamura y Koshizuka, 2005; Burbules, 2012, Ángel y De Castro, 2013) que permite adecuar las necesidades sociales a los aspectos de habitus docente que surgen de forma habitual y espontánea en el entorno educativo.

El enfoque de inclusión en educación no solo depende del rol de docente, también implica una articulación administrativa y una voluntad institucional que vaya mas allá de considerar los aspectos legales, implica tomar acciones de seguimiento y mejoramiento continuo, entre los cuales, un software orientado en esta línea de trabajo se constituye en una herramienta potente de cara un proceso de inclusión más integral.

Mediante la utilización del software se ha intervenido una población de 124 estudiantes, con los cuales se tiene información en tiempo real de sus discapacidades, tratamiento, grado y docentes que tienen que ver con su proceso, en este sentido, el software debe garantizar un acceso seguro y cómodo a los actores directamente implicados, de tal forma que sea una opción de utilidad en el proceso.

Referencias

Ángel, M., y De Castro, C. (2013). La Información digital actual, un nuevo modelo de contenido educativo para un entorno de aprendizaje. *RED-Revista de Educación a Distancia ubicuo*.

Bourdieu, P. (1996). *Raisons practiques.* París: Points.

Burbules, N. (2012). El aprendizaje ubicuo y el futuro de la enseñanza . *Encounters/Encuentros/Rencontres on Education*, 3-14.

Carrizo, D., y Ortiz, C. (2016). Modelos del proceso de educción de requisitos: Un mapeo sistemático. *Ingeniería y desarrollo*, 184-203.

Consejo Nacional de Política Económica y Social. (2016). *Documento Conpes - Política Nacional de Ciencia, Tecnología e Innovación 2016-2025.* Bogotá: Departamento Nacional de Planeación.

Cornell, U. (2015). *Global Innovation Index.* Geneva: OMPI.

Departamento Nacional de Planeación. (2016). *Políticas de desarrollo productivo, ciencia, tecnología e innovación.* Recuperado de: http://investincauca.com/sites/default/files/descargables/politicas_dsllo_productivo_ctei_pnd_0.pdf

Durán, A., y Bernárdez, B. (2000). *Metodología para la Elicitación de Requisitos de Sistemas Software.* Sevilla: Universidad de Sevilla.

Granhag, P. A., Kleinman, S. M., y Oleszkiewicz, S. (2015). The Scharff Technique: On How to Effectively Elicit Intelligence from Human Sources. *International Journal of Intelligence and CounterIntelligence*, 132-150.

Lee, K. (2013). *Schumpeterian Analysis of Economic Catch-up.* Cambridge: Cambridge University Press.

Maloney, W., y Bitran, E. (2013). *Outline - Innovación para la Competitividad.* Bogotá D.C.

Ministerio de Educación Nacional. (27 de Febrero de 2013). *Educación Bogotá.* Recuperado de: http://www.educacionbogota.edu.co/archivos/Temas%20estrategicos/banco_oferentes/2013/2012%20LEY%201618%20Derechos%20Personas%20con%20Discapacidad.pdf

Ministerio de Educación Nacional. (Agosto de 2017). *mineducación.* Obtenido de Decreto de Educación Inclusiva para población con discapacidad. Recuperado de: https://www.mineducacion.gov.co/1759/articles-362988_abc_pdf.pdf

Ministerio de Educación Nacional. (29 de Agosto de 2017). *Presidencia de la República de Colombia.* Recuperado de: http://es.presidencia.gov.co/normativa/normativa/DECRETO%201421%20DEL%2029%20DE%20AGOSTO%20DE%202017.pdf

Naciones Unidas. (2015). *Objetivos del Desarrollo del Milenio. Informe de 2015.* Nueva York: Naciones Unidas.

Naciones Unidas. (2016). *Agenda 2030 y los Objetivos de Desarrollo Soste-*

nible. Una oportunidad para América Latina y el Caribe. Santiago de Chile: Naciones Unidas.

Organización Mundial de la Salud. (2011). *Resumen. Informe Mundial sobre la Discapacidad.* Malta: Organización Mundial de la Salud.

RAE. (16 de 10 de 2017). Obtenido de Real academia de la lengua española: http://dle.rae.es/srv/search?m=30yw=educci%C3%B3n

Sakamura, K., y Koshizuka, N. (2005). Ubiquitous coputing technologies for ubiquitous learning. *Wireless and Mobile Technologies in education*, 11-20.

Samaniego de García, P. (2009). *Personas con discapacidad y acceso a servicios educativos en Latinoamérica.* Madrid: Grupo editorial CINCA.

Soler, M. (2016). *De las TIC a las TEP pasando por las TAC.* Valencia: Universidad Jaime I.

Taskin, F., y Zaim, O. (1997). Catching-up and innovation in high- and low-income countries. *Economics Letters*, 93-100.

Zowghi, D., y Coulin, C. (2005). *Requirements Elicitation: A Survey of Techniques, Approaches, and Tools.* New York: Springer.

Buenas prácticas de seguridad informática para mitigar las amenazas persistentes avanzadas en el sector salud en Colombia

Yexid Montenegro García[1]
Christian Hernán Obando Ibarra[2]
Gloria Amparo Lora[3]
Enevis Rafael Reyes Moreno[4]

Resumen

El crecimiento de la economía digital sobre las actividades diarias de la sociedad, ha traído consigo un conjunto de riesgos, amenazas, vulnerabilidades e incidentes tales como malware, troyanos, suplantación de identidad y amenazas persistentes avanzadas (APT), de acuerdo al último informe del Departamento Nacional de Planeación (2016), en Colombia se pasó de gestionar un total de 4.640 incidentes digitales en 2014 a un total de 7.323 en 2015. La gran mayoría de este tipo de ataques preocupa por la efectividad de los mismos y muchas veces se dificulta su detección de forma oportuna.

Para minimizar los riesgos de que las amenazas se materialicen y puedan afectar el sector hospitalario en Colombia con posibles fugas de información, se propuso un Framework de seguridad informática orientada al sector hospitalario, que permite mitigar la fuga de información ocasionada por APT que se propaga a través de correo electrónico.

1 Ingeniero de Sistemas y Computación, candidato a Magister en seguridad Informática. Docente de tiempo completo de la Corporación Universitaria Americana. fmontenegro@coruniamericana.edu.co. ORCID ID 0000-0003-2708-4497.

2 Ingeniero en Electrónica y Telecomunicaciones, Especialista en Seguridad en Informática, Magister en Tecnologías de la Información y Comunicación. Docente de tiempo completo de la Corporación Universitaria Americana. cobando@americana.edu.co. ORCID ID 0000-0003-2326-8934.

3 Ingeniera en Sistemas, Candidata a Magister en Seguridad Informática. Docente hora catedra del Politécnico Jaime Isaza Cadavid. gloriamaparo@gmail.com. ORCID ID 0000-0002-9802-9615.

4 Director de Ingenierías de la Corporación Universitaria Americana. Magister en Entornos Virtuales de Aprendizaje, Universidad de Panamá. Especialista en ciencias Electrónicas e informáticas de la Universidad de Antioquia, Especialista en Entornos Virtuales de aprendizaje. Correo electrónico de contacto: diringenieriasmed@americana.edu.co. ORCID: 0000-0003-4145-1898

Palabras clave: riesgo, amenaza, vulnerabilidad, salud, framework, fuga, APT.

Proposal of a set of good IT security practices to mitigate advanced persistent threats in the health sector in Colombia

Abstract

The growth of the digital economy on the daily activities of society, has reached the set of risks, threats, vulnerabilities and incidents such as malware, Trojans, identity theft and advanced persistent threats (APT), according to the latest report of the National Department of Planning (2016), in Colombia it became a total of 4,640 digital incidents in 2014 a total of 7,323 in 2015. The vast majority of this type of attacks concerned about the effectiveness of these and in many cases it is difficult to detect these in a timely manner.

To minimize the risks of the threats materializing and affecting the hospital sector in Colombia with possible information leaks, a computer security framework was proposed aimed at the hospital sector that allows mitigating the leakage of information caused by APT that is propagated through of e-mail.

Key words: risk, threat, vulnerability, health, framework, fugue, APT.

Introducción

Los nuevos tipos de amenazas conocidos como ataques dirigidos duraderas con alto impacto, llamados amenazas persistentes avanzadas (APT), han ido en crecimiento, enfocándose en diferentes sectores de la economía colombiana. El área de la salud ha comenzado a ser víctimas de diferentes tipos de ataques, los ataques se enfocan en un objetivo en común dentro éstas, haciendo un análisis de sus vulnerabilidades, para luego iniciar el proceso de infección; hasta quebrantar los esquemas de seguridad estipulados por la organización. Estas infraestructuras están vulnerables ante un medio que crece a pasos agigantados y donde la formación de los usuarios y/o pacientes para

solucionar, actuar o mitigar estos tipos de amenazas, no es suficiente para contrarrestarlas. Por ejemplo, frente a la fuga de información proveniente de correo electrónico; los riesgos emergentes cotidianos y las amenazas de los cibercriminales han visto en el correo la oportunidad de recopilar información confidencial y sensible de una organización, puesto que la comunicación por correo electrónico puede ser explotada por intrusos para recopilar la información confidencial y sensible de la organización, los atacantes utilizan esta falla de seguridad para enviar información inapropiada, sensible y publicitaria mediante el correo, haciendo uso de una identidad legitima dentro del servidor de correo. Siendo así, las organizaciones del sector salud están a merced de los nuevos tipos de amenazas a la seguridad de la información, sin embargo, algunas no asumen el riesgo o piensan que no les afecta, solo que en algunos casos prefieren esperar que suceda una vulneración de algún sistema para después actuar.

La información es un recurso muy importante para el funcionamiento de cualquier empresa y del buen manejo que se haga de ella depende su nivel de competitividad. No basta solo con tener almacenada una gran cantidad de información para lograr la eficiencia en los procesos, también es necesario que esta sea gestionada, administrada y asegurada oportunamente de tal manera que sirva de apoyo a la toma de decisiones convenientes para la entidad. Por tal motivo para obtener el mayor beneficio y seguridad de la información es necesario emplear mecanismos adecuados para que esta fluya sin ningún obstáculo, tanto interna como externamente, uno de ellos es la Arquitectura de Seguridad orientada al sector hospitalario que permita detectar la fuga de información causada por APT.

Contar con un modelo de seguridad informático orientado al sector hospitalario mejora considerablemente la seguridad de estos sistemas de comunicación de las entidades que prestan servicios de hospitalario, porque esto le permite asegurar y minimizar los ataques a los que se encuentra expuesta la información de empleados, pacientes y personal administrativo usuarios del correo electrónico. Todo esto debido a un crecimiento significativo en las amenazas cibernéticas que cada vez son más difíciles de detectar. Por otra parte, se ha demostrado en los últimos años que las amenazas a las infraestructuras críticas nacionales son una realidad y presentan una tendencia creciente. En 2015, la OEA y Trend Micro llevaron a cabo una encuesta entre los jefes de seguridad de infraestructuras críticas nacionales, donde hicieron

algunos hallazgos: El 53% de los encuestados había observado un incremento de los incidentes en sus sistemas de cómputo durante el 2014, y el 76% de los encuestados respondió, que dichos incidentes contra las infraestructuras críticas nacionales se han vuelto más sofisticados (Departamento Nacional de Planeación, 2016).

Los nuevos tipos de amenazas que, en los últimos años, conocidos como ataques dirigidos duraderas con alto impacto, llamados amenazas persistentes avanzadas (APT), han ido en crecimiento, enfocándose en estas infraestructuras, los ataques se enfocan en un objetivo en común dentro éstas, haciendo un análisis de sus vulnerabilidades, para luego iniciar el proceso de infección; hasta quebrantar los esquemas de seguridad estipulados por la organización. Estas infraestructuras están vulnerables ante un medio que crece a pasos agigantados y donde la formación para solucionar, actuar o mitigar estos tipos de amenazas, no es suficiente para contrarrestar estas amenazas, como por ejemplo la fuga de información, no existen reglas para actuar frente a una amenaza persistente avanzada, que tenga como finalidad el robo de información o cualquier otro fin oscuro, muchos de los esfuerzos de los países es mejorar sus propios servicios hospitalarios y de productos para reducir los costos de las enfermedades, esto podría prever un aumento de la orientación de los grupos APT, en miras de obtener algún tipo de beneficio con esta información. Siendo así, que las organizaciones del sector salud están a merced de los nuevos tipos de amenazas a la seguridad de la información, sin embargo, algunas empresas no asumen el riesgo o piensan que no les afecta, solo que en algunos casos prefieren esperar que suceda una vulneración de algún sistema para después actuar.

El sector salud en Colombia es víctima de brechas de seguridad, fugas o pérdidas de información ocasionada por APT, debido a que sus servicios entre los diferentes prestadores se encuentran aislados entre sí, como también debido al volumen de información y a los medios en las que la información puede ser almacenada y transferida, sin el consentimiento previo del usuario.

Las fugas de información, han ido creciendo; amenazando seriamente la solidez del sector salud y la privacidad de sus pacientes, al no contar con los planes de mitigación que puedan frenar está y así evitar una posible inactividad total o parcial de los servicios, donde el usuario final es el más perjudicado.

Framework de seguridad de la información

La información es uno de los principales activos de cualquier organización, para el normal funcionamiento y la consecución de sus objetivos. Debido a esto, las empresas y organizaciones necesitan protegerla, para asegurarse que esta sea fiable, a la hora de gestionar grandes volúmenes de ella. La seguridad de la información, es el proceso mediante el cual la empresa mantiene y alcanza unos niveles apropiados de confidencialidad, integridad, disponibilidad y autenticidad, desde ese punto de vista existen varios, estándares para garantizar los niveles de seguridad mencionados anteriormente, entre ellos están:

Norma ISO 27000

Dedicada a la seguridad de la información, específica los ítems para la seguridad de la información, más no las directrices para ello. El proceso se describe en los siguientes ítems: Estructura para la seguridad de la información, lo que tiene que ver con terceros, el control de acceso, adquisición y desarrollo de los sistemas de información; sirviendo de apoyo para el proyecto en cuanto a lo que ISO 27000 puede ofrecer para el mismo:

- El alcance que tendrá el SGSI sobre los procesos de la entidad hospitalaria.
- La política general de seguridad de la información.
- La identificación y valoración de los activos de la información.
- Los riesgos a los cuales los activos identificados se encuentran expuestos.
- La selección de los controles para mitigar los riesgos que se han detectado (Lopez & Ruiz, 2005).

ISO 27799: 2016 e ISO / IEC 27002

Estas normas al ser tomadas en conjunto, definen lo que se requiere en términos de seguridad de la información en el sector hospitalario. La norma ISO 27799: 2016 es neutra desde el punto de vista tecnológico. Esta neutralidad se da con respecto a la implementación de tecnologías. Asimismo, proporciona directrices para las normas de seguridad de la información organizacional y

las prácticas de gestión de la seguridad, incluyendo la selección, implementación y administración de controles, teniendo en cuenta el ambiente de riesgos de seguridad de la información de la organización, en un entorno donde las amenazas y vulnerabilidades únicas, deben considerarse con especial atención (ISO, 2016).

ISO 31000

Es la norma que brinda los principios y las directrices genéricas sobre la gestión del riesgo. Esta norma se puede aplicar durante toda la duración de una organización y a un amplio rango de actividades, incluyendo estrategias y decisiones, operaciones, procesos, funciones, proyectos, productos, servicios y activos.

Esta norma se puede aplicar a cualquier tipo de riesgo, cualquiera sea su naturaleza, bien sea que tenga consecuencias positivas o negativas.

El numeral 5 contempla lo siguiente: Comunicación y consulta (5,2), Establecimiento del contexto (5.3), Valoración del Riesgo (5.4), Identificación del Riesgo (5.4.2), Análisis del Riesgo (5.4.3), Evaluación del riesgo (5.4.4), Tratamiento del riesgo (5.5), Monitoreo y revisión (5.6) (Ramírez & Ortiz, 2011).

Desarrollo de la temática

Para lograr los resultados de este capítulo de libro, se hizo un análisis de información de las buenas prácticas en seguridad informática más reconocidas, teniendo en cuenta los autores más citados hasta el momento. Adicionalmente, para la propuesta se tienen en cuenta estándares internacionales como COBIT, Margerit, HIPA, ISO 27000, ISO 31000, NIST SP 800-30, ITIL, OCTAVE entre otros.

Para la propuesta de la guía de buenas prácticas se utilizó el modelo Kill Chain. El modelo Kill Chain (figura 1), patentado por Lockheed Martin, es uno de los modelos más aceptados al momento de detectar cualquier tipo de

amenaza, ya sea una amenaza tradicional o una amenaza persistente avanzada. Este fue seleccionado por que permite aplicar estrategias proactivas de prevención y detección temprana para dar respuesta, contención y mitigación a los incidentes APT y también se eligió teniendo en cuenta la trayectoria del autor Lockheed Martin en seguridad informática, el cual ha sido citado en más de 235 artículos.

De esa forma se presentan la descripción de las fases del APT usando el modelo Kill Chain, para mitigar la fuga de información, donde se aprecia el resumen las técnicas de prevención y detección que más se mencionaron y coincidieron por los autores que realizaron publicaciones de 2013 hasta 2017, y teniendo en cuenta también los autores más citados.

En la figura 1 se muestra el Modelo Kill Chain, el cual consta de 7 fases:

Figura 1. Ciclo de vida de amenazas persistentes avanzadas - Modelo Kill Chain.
Fuente: Serrano (2017)

Fase de reconocimiento

En esta fase se realiza una recopilación de información de los objetivos que se desea infiltrar identificando los puntos débiles de la organización y/o de los empleados. El reconocimiento se puede desglosar identificando los objetivos, luego mirando e identificando los perfiles de usuarios en la organización. (Bhatt & Yano, n.d.; Marchetti et al., 2016). También se valen de las relaciones con otras entidades en donde se pueden apoyar los atacantes para alcanzar el objetivo. Analizan la red para buscar servicios abiertos o desprotegidos, identifican los sistemas de defensa que la organización utiliza, y analizan que empleados tienen acceso a la información específica que les sirva para lograr su cometido, utilizando información de las redes sociales públicas de la que los empleados pueden ser miembros (por ejemplo, LinkedIn, Facebook, entre otras) (Giura & Wang, 2013).

El reconocimiento es una de las fases en la que el atacante se toma el tiempo necesario para poder infiltrar el sistema y así poder lograr su objetivo. Para poder realizar dicha intrusión en el sistema, el cibercriminal busca fuentes de información (correos corporativos, sitios más visitados, reconocimiento de puertos, sistemas operativos, aplicaciones, entre otros) que sirvan como estrategia para lanzar ataques dirigidos (Bhatt & Yano, n.d; Marchetti et al., 2016), es un paso importante de preparación antes del ataque. Los atacantes identifican y estudian la organización, recopilan toda la información posible sobre el entorno técnico y personal clave de la organización (Chen, Desmet, & Huygens, 2014), para luego en la fase de preparación usar la información vulnerable de la organización y preparar la operación de ataque.

Fase de preparación de la operación

Los atacantes en esta fase preparan el entorno a atacar, haciendo uso de malware, el cual diseñan y desarrollan para explorar vulnerabilidades identificadas en la fase 1. El código malicioso se desarrolla de tal forma que tiene la capacidad de acoplarse a formatos insospechados como pdfs, docs y ppts (Chen et al., 2014; Bhatt et al., 2014). También ponen en marcha servicios maliciosos, para engañar a los usuarios y que así hagan uso de estos (Luh et al.,

2017). La preparación puede consistir en un correo electrónico de phishing dirigido, utilizando información que reunieron en la etapa de reconocimiento.

En este caso, el correo electrónico de phishing, podría contener una invitación a un evento, rifa, ofertas, pago de servicios, que son programados por una organización en el que el empleado objetivo confíe, y proceda a realizar un clic sobre la URL y descargar un archivo con documentos o archivos adjuntos infectados. En la gran mayoría de los casos el correo electrónico es el vector de infección, más usado en la fase de distribución para entregar el malware, pero también se pueden usar otros canales, tales como medios extraíbles USB y sitios Web de baja reputación (Giura & Wang, 2013).

Fase de distribución

En ésta fase los atacantes tienen conocimiento fuerte de su objetivo y de los empleados, que se identificaron en la fase 1. La organización criminal tiene todo lo que necesita para empezar a buscar un punto de ingreso a la red de la compañía y establecer uno o varios accesos permanentes. Generalmente los atacantes se aprovechan de vulnerabilidades que la víctima no ha identificado (Crowe, 2015; Chen et al., 2014; Bhatt et al., 2014). En esta fase para realizar la entrega, el atacante utiliza varios canales de entrega, ya sea correos, sitios web, medios extraíbles USB, etc. (Luh et al., 2017), para engañar a los usuarios de la organización, usando tácticas de engaño, mencionadas en la fase 2.

Fase de explotación

Esta fase de centra en la entrega de carga útil a la víctima anfitrión, la explotación desencadena código malicioso para realizar la intrusión. (Luh et al., 2017). La explotación se dirige a una vulnerabilidad de software de aplicación o software operativo y aprovecha una característica del sistema operativo que permite auto ejecutar el malware o código malicioso (Yadav & Rao, 2015). Esta es la fase más crítica, ya que mediante vulnerabilidades existentes en el software permite acceder al sistema y tener control total de este. La vulnerabilidad es el error de software que puede resultar en una amenaza potencial

para el sistema. Un error de software es una condición inesperada en la que el software se porta mal (Medina, 2014). Una vez el malware ingrese en la red de la organización por medio del empleado seleccionado, el malware descargado por engaño, se instala y se activa, para más tarde crear una conexión de comando y control (C & C), desde la máquina víctima hasta la computadora del atacante. Una vez asegurada la conexión de C & C, los ciberdelincuentes continúan en silencio recopilando información sobre las configuraciones de seguridad de la computadora infectada y de los equipos conectados en red, también recopila información relacionada del sistema, las contraseñas, mensajes de correo electrónico de usuario para soportar futuros ataques (Giura & Wang, 2013).

Fase de instalación

Después de explotar la vulnerabilidad en la fase anterior, el atacante puede acceder al sistema de la víctima y lograr la persistencia en la máquina infectada y así tener acceso a la información objetivo de su ataque, mediante el uso de programas que permiten instalar el malware, tales como:

- **Dropper:** Diseñado para instalar malware (virus, backdoors, otros) a un sistema de destino y evitar ser detectados por los programas antivirus. Una vez activado dropper puede ser utilizado para robar la identidad o para dañar el rendimiento de los equipos (Yadav & Rao, 2015).
- **Downloader:** Programa que permite descargar automáticamente caratulas de música con código maligno y posteriormente instalar el malware y ocultarlo para evitar ser detectado por el sistema antivirus (Yadav & Rao, 2015). En esta fase se valen de las vulnerabilidades explotadas para realizar inyección de código, dejar gusanos o troyanos, buscar otros kits de Exploits y realizar suplantación de identidades o realizar fraude. En esta fase, en la que la carga útil entregada aprovecha una vulnerabilidad y se instala en la máquina de la víctima. (Crowe, 2015), también se considera la fase donde los atacantes usan técnicas de instalación de malware en secreto, para explotar las vulnerabilidades a nivel de sistema operativo, de software aplicativo y a nivel de red, para hacerse con los sistemas, causando denegación de servicios, ejecución

de código remoto o local, escalar privilegios, deficiencia en el proceso de negociación en el protocolo TLS, SSL, corrupción de memoria, error de entradas invalidas, buffer overflow, dangling pointer, use after free y bypass.

Fase comando y control

En esta fase el atacante obtiene el comando y control de la máquina infectada, denominado C&C, después de haber explotado una vulnerabilidad del sistema, mediante el uso de troyanos, botnets y denegación de servicios. El sistema de comando y control es usado para dar instrucciones remotamente a una maquina comprometida (Yadav & Rao, 2015). El comando y control se puede obtener usando estructuras de comunicación. Entre ellas, se destacan tres tipos de estructuras de comunicación de comando y control, la estructura tradicional centralizada, la peer-to-peer (método para intercambio de archivos, programas, aplicaciones, vídeos o fotos) descentralizada y las más reciente en redes sociales basadas en última estructura. En este caso la carga útil se instala y establece la conexión saliente con el entorno del atacante para permitir la interacción con el adversario malintencionado (Crowe, 2015). Hay que tener en cuenta que los atacantes se valen de canales anónimos de comunicación de malware como IRC Chats, TCP, HTTP, FTP, estenografía, TOR, DNS Fast Flux, DNS como medio y algoritmos de generación de dominios para sustraer información sin ser detectados.

Fase de acciones sobre objetivos

Después de lograr configurar la comunicación con el sistema de destino, el atacante ejecuta los comandos necesarios de acuerdo con los intereses que dieron inicio al plan de ataque (Yadav & Rao, 2015). Esta es la fase final de un ataque APT, donde este está en posición para hacerse con los datos objetivos. Dependiendo del tipo de objetivo, esta actividad puede incluir robo de información de cuentas bancarias, correos electrónicos, redes sociales, credenciales de administrador y manipulación de información confidencial o secreta, extracción de datos, entre otros (Crowe, 2015).

Tabla 1. *Resumen del ciclo de vida del APT con el modelo Kill Chain y las técnicas de prevención y detección más recomendadas por Autores.*

Fase	Técnicas de Ingeniería Social	Riesgos	Técnicas de prevención	Técnicas de Mitigación	Autores														
					1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Reconocimiento	Técnicas presenciales no agresivas	Fisgoneo o buscar en la basura	Pruebas de seguridad, políticas de destrucción de documentos	Router Logs, Firewall logs															
			Recopile registros de visitantes																
	Pasivas	Escaneo de red	Proteger la red para evitar el reconocimiento de servicios innecesarios	Políticas DROP para firewall			✓					✓			✓	✓	✓		✓
	Técnicas no presenciales	Rootkit	Actualizaciones de software y parcheo de vulnerabilidades	Antivirus															
		Troyano	Registro de logs, http logs (Análisis)	Autenticación (Análisis de trafico DNS logs malicioso), Análisis de Malware		✓	✓		✓	✓		✓		✓	✓				
	Técnicas agresivas	Robo de Identidad, Chantaje	Políticas de contraseñas seguras y cambio de contraseñas	Controles de acceso lógico y físico										✓				✓	
	Pasivas	Perfil de los empleados (Redes Sociales, Sitios Web)	Educación en seguridad									✓	✓	✓	✓		✓	✓	✓
	Técnicas no presenciales	DOS	HIDS	NIDS		✓		✓			✓	✓							

Fase	Técnicas de Ingeniería Social	Riesgos	Técnicas de prevención	Técnicas de Mitigación	Autores														
					1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Reconocimiento	Técnicas no presenciales	DOS	Prevenir inundaciones (floods) en los protocolos TCP/ UDP	Limitar el número de conexiones concurrentes al servidor. Restringir el uso del ancho de banda por aquellos hosts que cometan violaciones.		✓		✓			✓	✓			✓				
		Redes botnets	Filtrado de protocolos innecesarios.	Monitoreo de logs y de las conexiones TCP/ UDP que se llevan a cabo en el servidor. Limitar la tasa de tráfico proveniente de un único host.		✓		✓			✓	✓					✓		
		Malware	Análisis de navegador	Detección de firmas								✓			✓				✓
				Máquinas de aprendizaje		✓			✓	✓	✓	✓				✓			
				Detección virtual de sambox		✓						✓				✓			
		Sondeo de Servicios	Evaluación de la vulnerabilidad utilizando un equipo azul y Rojo			✓						✓				✓			
		Waltering hole	Correlación de eventos, herramientas y técnicas de generación de gráficos de ataque	Análisis contextual (Enfoques Bayesianos), Análisis Web		✓	✓		✓	✓	✓	✓		✓					
			Revisar la configuración de Routers y Firewalls ACL	Caracterización de reputación (Listas Blancas, MET, SpamFlow)			✓		✓	✓	✓	✓				✓			
	Técnicas no presenciales	Ataque de Ofuscación		Identificación basado en ataques de evasión	✓														
			Segmentación de red, DMZ									✓	✓						

Fase	Técnicas de Ingeniería Social	Riesgos	Técnicas de prevención	Técnicas de Mitigación	Autores														
					1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Militarización	Técnicas no presenciales	Archivos infectados	Recolecte archivos y metadatos para análisis futuros	Filtros de reputación (anti spam, antivirus, Listas blancas, Listas negras y listas de bloque en tiempo real basadas en DNS, ClamAV, SpamAssassin), filtro proxy, Detección de reglas, Políticas basadas en reglas.															
		URL maliciosa	Identificar e integrar las actividades de evaluación de controles de seguridad				✓	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓
		Envió de correo spear phishing	Enumerar vulnerabilidades descubiertas																
	Técnicas no presenciales	Malware	Analice la línea de tiempo de creación de malware	Análisis de artefactos de malware.															
		Exploit	Determinar artefactos de armamento comunes a las campañas de APT				✓	✓	✓	✓	✓	✓	✓			✓			✓
		Fuerza Bruta / Diccionario																	
Entrega	Técnicas no presenciales	Phishing	Filtros de ejecución de contenido dinámico					✓	✓	✓	✓	✓	✓	✓			✓	✓	
	Técnicas no presenciales	Botnets	NIPS	NIDS				✓			✓	✓							
	Técnicas no presenciales	Archivos infectados	Computación Segura y confiable basada en controles de software	Analice el medio de entrega									✓	✓				✓	✓

Fase	Técnicas de Ingeniería Social	Riesgos	Técnicas de prevención	Técnicas de Mitigación	Autores													
					1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Entrega	Técnicas no presenciales		Usuario vigilante								✓								
	Técnicas no presenciales	Troyano	In-line AV					✓			✓			✓			✓		
	Técnicas no presenciales	Malware	Queuing								✓								
			Comprenda la infraestructura ascendente	Consumo de Recursos (para filtrar correo electrónico)			✓	✓	✓	✓			✓	✓				✓	✓
			Entender a los servidores y personas específicos, sus roles y responsabilidades	Random Forests Algorithm				✓	✓										✓
			Recopile registros de correo electrónico y web para la reconstrucción forense	Dataset del email (Vortex-IDS, WEKA).				✓	✓				✓						✓
			Detectar nuevas cargas maliciosas en el punto de entrega	Herramientas de Clasificación de correos				✓					✓	✓					✓
Explotación	Técnicas no presenciales	Rootkit	Capacitación de usuarios	HIDS															
		Rootkit	Pruebas de correo electrónico para los empleados	Restringir privilegios de administrador		✓		✓			✓							✓	✓
		Troyano	Actualizaciones (Patch), hardening, Corregir las Vulnerabilidades	Monitorear Alertas del sistema		✓					✓			✓		✓	✓		✓
		Exploits 0 Day	DEP(Prevención de ejecución de datos de Windows)	Bloquear ejecución de programas sin autorización.							✓			✓			✓		✓

Fase	Técnicas de Ingeniería Social	Riesgos	Técnicas de prevención	Técnicas de Mitigación	Autores														
					1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
		Inyección sql	Capacitación de codificación segura para desarrolladores web	Auditoría del proceso del punto extremo para análisis forense y conocer origen del exploit															
		Malware	Análisis de vulnerabilidades y pruebas de penetración regulares-	Medidas de refuerzo del punto final- Reglas del punto final personalizado para bloquear la ejecución del código shell															
Instalación	Técnicas no presenciales	Instalación de malware adicional	Restringir Privilegios	HIDS															
			Entender el malware	Auditoría de proceso de punto final		✓	✓	✓				✓	✓						✓
				Extrae certificados de cualquier archivo ejecutable firmado															
		Ocultamiento de malware	Chroot jail	Crear nuevas mitigaciones de punto final.							✓							✓	✓
		Backdoor	AV	HIPS				✓			✓			✓	✓				
Comando y control	Técnicas no presenciales	Escalada de Privilegios	Firewall ACL	NIDS		✓	✓	✓	✓	✓	✓	✓					✓	✓	✓
		Propagación																	
		Pivotin																	
		Transferencia no autorizada de datos en resguardo	DLP	Bloquear el canal de comando y control	✓												✓		
			Software para ocultar las pulsaciones de teclado	Análisis completo de malware en la infraestructura de comando y control											✓	✓			✓
			Usar un enfoque de las Ersatz Passwords para mejorar la seguridad de los hashes	Bloquear la categoría proxy, incluidos los dominios “ninguno” o “sin categoría”.											✓				✓
			Cifrado de datos																
		Rootkit	NIPS	Intoxicación del receptor de DNS y envenenamiento del servidor de nombres				✓			✓								✓

Fase	Técnicas de Ingeniería Social	Riesgos	Técnicas de prevención	Técnicas de Mitigación	Autores														
					1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Comando y control	Técnicas no presenciales	Instalación de malware adicional	Tarpi	Red Harden							✓						✓		✓
			DNS redirect	Llevar a cabo investigaciones de código abierto para descubrir nuevas infraestructuras C2 adversas			✓	✓	✓	✓	✓								
			Personalice bloques de protocolos C2 en proxies web																
Acción sobre los objetivos	Técnicas no presenciales	Canal encubierto	Qualility of service (calidad del servicio)	Audit log															
				Estrategias de respuesta ante incidentes		✓	✓	✓			✓						✓	✓	✓
				Plan de comunicaciones															
		Destruir o modificar datos	Honeypot	Capturas de paquetes de red							✓								✓
				Realizar una evaluación de daños															
		Robo de datos	Incluir las lecciones aprendidas	Entender el ataque (Análisis Forense en puntos finales)											✓	✓		✓	✓

Fuente: (1, Mustafa, 2013; 2, Luh et al., 2017; 3, Giura & Wang, 2013; 4, Deshmukh, Shelar y Kulkarni, 2014; 5, Oprea, Li, Yen, Chin, & Alrwais, 2015; 6, Hutchins, Cloppert, & Amin, n.d.; 7, Hutchins et al., n.d.; 8, Moon, Im, Lee, & Park, 2014; 9, Virvilis & Gritzalis, 2013; 10, Crowe, W, 2015; 11, Al-Mohannadi, Mirza, Namanya et al., 2016; 12, Christensen, 2013; 13, Ioannou, Louvieris, Clewley, & Powell, n.d.; 14, Bodeau & Graubart, 2013; 15, Lockheed, 2014).

Resultados para la guía de buenas prácticas

Tenga en cuenta las buenas prácticas anteriores, para mitigar los ataques que causan fuga de información en cada fase del ciclo de vida APT, sin embargo, algunas se repiten en alguna de las fases, si dentro de la organización ya las tiene y los sistemas fueron vulnerados, debe configurarlas de manera adecuada para que mitigue la extracción de datos en el sector hospitalario, a continuación, se listan.

- **Fase de reconocimiento:** Detectar las amenazas en esta etapa puede ser muy difícil, pero cuando los defensores descubren su funcionamiento, puede revelar la intención de los adversarios. En la fase de reconocimiento se deben tener en cuenta las buenas prácticas que se mencionan en la figura 2.

Figura 2. Buenas prácticas que debe usar el encargado de la seguridad del sector hospitalario en Colombia en la fase de reconocimiento.

- **Fase 2 de preparación de la operación:** es la fase esencial para entender cómo defenderse, aquí se puede inferir mediante el análisis de malware que herramientas pueden usar para vulnerar las barreras de seguridad existentes. Es necesario implementar las buenas prácticas que se mencionan en la figura 3.

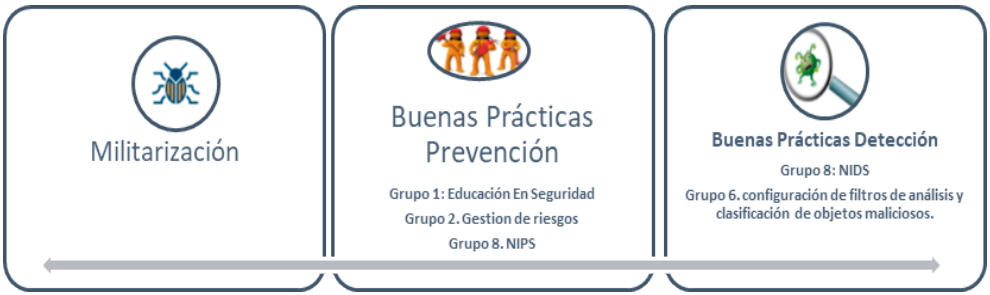


Figura 3. Buenas prácticas que debe usar el encargado de la seguridad del sector hospitalario en Colombia en la fase de militarización.

- **Fase 3, Distribución o entrega:** en esta fase es donde realmente los sistemas deben actuar y es la oportunidad para que los defensores bloquen la operación maliciosa. La eficacia para bloquear estos intentos de intrusión en la etapa de entrega nos permitirá prevenir la carga del paquete con contenido malicioso en los sistemas del sector hospitalario. Utilice las buenas prácticas de la figura 4 para mitigar la fuga de información que usa el vector correo como medio de infección.

Figura 4. Buenas prácticas que debe usar el encargado de la seguridad del sector hospitalario en Colombia en la fase de Entrega.

- **Fase 4, Fase de explotación:** la técnica más usada en esta fase es la ingeniería social y ataques basados en navegador, que traen consigo amenazas. Estos usan vectores de infección muy conocidos: USB, mensajería instantánea, redes P2P y e-mail. Aquí usando medidas de endurecimiento del grupo 4, permite tener capacidad de recuperación y evitar ataques conocidos, con las buenas prácticas de grupo 9 previene ataques de día cero.

Figura 5. Buenas prácticas que debe usar el encargado de la seguridad del sector hospitalario en Colombia en la fase de explotación

- **Fase 5, Fase de instalación:** las buenas prácticas aquí permiten detectar y registrar los intentos de instalación de software malicioso.

Figura 6. Buenas prácticas que debe usar el encargado de la seguridad del sector hospitalario en Colombia en la fase de instalación.

- **Fase 6, Comando y control:** el administrador tiene la última oportunidad de defensa y bloquear la operación de fuga de información, si bloquear el canal de comando y control, el atacante no puede emitir comandos para hacerse con la máquina.

Figura 7. Buenas prácticas que debe usar el encargado de la seguridad del sector hospitalario en Colombia en la fase de comando y control.

- **Fase 7, Fase de acciones sobre objetivos:** las buenas prácticas en esta etapa son para analizar mediante equipo forense e identificar lo sucedido.

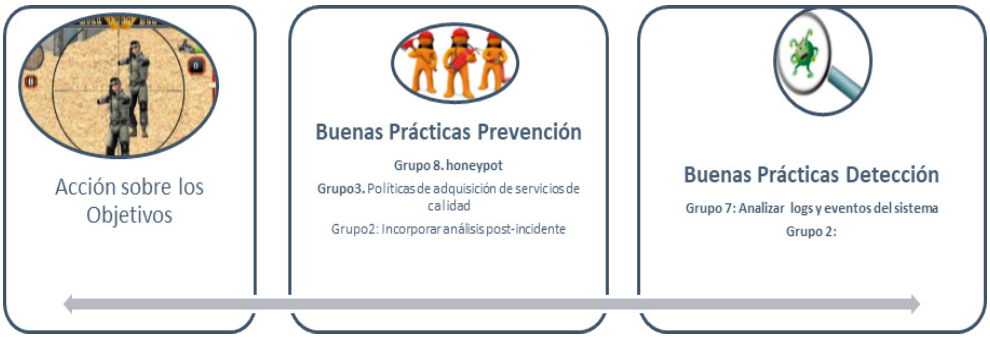


Figura 8. Buenas prácticas que debe usar el encargado de la seguridad del sector hospitalario en Colombia en la fase de acción sobre los objetivos.

Conclusiones

La revisión sistemática de portales como el Ministerio de las TIC, Departamento de administración pública y la Supersalud, ofrecen a las diferentes entidades del Estado buenas prácticas basadas en ISO 27000 y ISO 31000, muchas de ellas ofrecen artefactos construidos a partir de estos lineamientos, sin embargo, ninguno de ellos se adaptaba a las necesidades propias del sector salud y que se enfoquen en mitigar las amenazas persistentes avanzadas (APT).

Las buenas prácticas propuestas para el sector hospitalario de Colombia, han sido adaptadas de una lista de estándares, recomendaciones de organismos internacionales y nacionales, normas y autores que estudiaron a fondo la forma de operar de las APT, lo cual permite ser asertivos a la hora de combatir los peligros asociados a la fuga de información causada por las APT, ya que este es un proceso continuo. Al igual que en los procesos de desarrollo,

despliegue y mantenimiento, los errores ocurren, pero se debe reconocer su existencia y estar preparados para su aparición y mitigación.

Los administradores del área de sistemas y encargados deben realizar: vigilancia constante mediante el uso de herramientas automatizadas de monitorización en tiempo real, que permitirán en cierta medida adelantarse a las acciones que pueda realizar un atacante.

Los artefactos propuestos para identificación de activos sensibles a fuga de información pueden ser desarrollados como insumo para trabajo de pregrado en la carrera de Ingeniería de Sistemas.

Referencias

Abad, C. (2015). *Aplicación de metodología de Análisis de Malware al caso de estudio de la Amenaza Avanzada Persistente (APT)* "Octubre Rojo". Recuperado de: http://reunir.unir.net/handle/123456789/2841

Ammar, M., Rizk, M., Abdel, A., & Aboul, A. K. (2016). A Framework for Security Enhancement in SDN-Based Datacenters. In *2016 8th IFIP International Conference on New Technologies, Mobility and Security (NTMS)* (pp. 1–4). IEEE. https://doi.org/10.1109/NTMS.2016.7792427

Bhatt, P., & Yano, E. T. (n.d.). *Analyzing Targeted Attacks using Hadoop applied to Forensic Investigation*. https://doi.org/10.5769/C2013004

Bhatt, P., Yano, E. T., & Gustavsson, P. (2014). Towards a framework to detect multi-stage advanced persistent threats attacks. In *Proceedings - IEEE 8th International Symposium on Service Oriented System Engineering, SOSE 2014*. https://doi.org/10.1109/SOSE.2014.53

Bodeau, D., & Graubart, R. (2013). Intended effects of cyber resiliency techniques on adversary activities. In *2013 IEEE International Conference on Technologies for Homeland Security, HST 2013*. https://doi.org/10.1109/THS.2013.6698967

CAPEC. (2017). CAPEC Lista Versión 2.9. Retrieved April 29, 2017. Recuperado de: https://capec.mitre.org/data/index.html

Caralli, R. A., Stevens, J. F., Young, L. R., & Wilson, W. R. (2007). *Introducing OCTAVE Allegro: Improving the Information Security Risk Assess-*

ment Process. Recuperado de: http://www.sei.cmu.edu/publications/pubweb.html

Charlotte, U. N. C., Covington, M. J., Hu, H., Jin, J., Ahn, G.-J., & Zhang, X. (2009). Patient-centric authorization framework for sharing electronic health records. *Policy*, 125–134. https://doi.org/10.1145/1542207.1542228

Chen, P., Desmet, L., & Huygens, C. (2014). A study on advanced persistent threats. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-66244885-4_5

Crowe, W. (2015). *Cybersecurity Kill Chain*. Recuperado de: http://www.isaca.org/chapters2/jacksonville/events/Documents/CyberSecurity Kill Chain 8.19.15.pdf

Departamento Nacional de Planeación. (2016). *Conpes 3854 - Política Nacional De Seguridad Digital.* Recuperado de: https://colaboracion.dnp.gov.co/CDT/Conpes/Econmicos/3854.pdf

Giura, P., & Wang, W. (2013). A context-based detection framework for advanced persistent threats. In *Proceedings of the 2012 ASE International Conference on Cyber Security, CyberSecurity 2012*. Recuperado de: https://doi.org/10.1109/CyberSecurity.2012.16

Hutchins, E. M., Cloppert, M. J., & Amin, R. M. (n.d.). *Intelligence-Driven Computer Network Defense Informed* by Analysis of Adversary Campaigns and Intrusion Kill Chains.

Ioannou, G., Louvieris, P., Clewley, N., & Powell, G. (n.d.). *A Markov Multi-Phase Transferable Belief Model: An Application for predicting Data Exfiltration APTs*.

ISO. (2016). ISO 27799:2016 - *Health informatics -- Information security management in health using ISO/IEC 27002*. Recuperado de: https://www.iso.org/standard/62777.html

Lockheed, M. (2014). *The modern day attacker*. Recuperado de: https://www.youtube.com/watch?v=Lyn50b_n0CY&list=UUJWcF0ex7_doPdIQGbVp DsQ

López, A. y Ruiz, J. (2005). *ISO27000.es - El portal de ISO 27001 en español. Gestión de Seguridad de la Información*. Recuperado de: http://www.iso27000.es/

Luh, R., Marschalek, S., Kaiser, M., Janicke, H., Schrittwieser, S., & Luh, B. R. (2017). Semantics-aware detection of targeted attacks: a survey.

Journal of Computer Virology and Hacking Techniques, *13*, 47–85. https://doi.org/10.1007/s11416-016-0273-3

Marchetti, M., Pierazzi, F., Colajanni, M., & Guido, A. (2016). Analysis of high volumes of network traffic for Advanced Persistent Threat detection. *Computer Networks*. Recuperado de: https://doi.org/10.1016/j.comnet.2016.05.018

Moon, D., Im, H., Lee, J. D., & Park, J. H. (2014). MLDS: Multi-layer defense system for preventing advanced persistent threats. *Symmetry*. Recuperado de: https://doi.org/10.3390/sym6040997

Mustafa, T. (2013). Malicious Data Leak Prevention and Purposeful Evasion Attacks: An approach to Advanced Persistent Threat (APT) management. *2013 Saudi International Electronics, Communications and Photonics Conference, SIECPC 2013*, 1–5. Recuperado de: https://doi.org/10.1109/SIECPC.2013.6551028

Oprea, A., Li, Z., Yen, T. F., Chin, S. H., & Alrwais, S. (2015). Detection of Early-Stage Enterprise Infection by Mining Large-Scale Log Data. In *Proceedings of the International Conference on Dependable Systems and Networks*. Recuperado de: https://doi.org/10.1109/DSN.2015.14

Serrano, L. (2017). Uso del análisis estadístico e inteligencia para detectar amenazas en la red. Recuperado de: http://reedlatam.com/sadmoweb/files/modulos/ConferenciasTalleres/infosecurity/2017/workshops-salon-1/presentacion/presentacion_solcomp_-_hillstone.pdf

Virvilis, N., & Gritzalis, D. (2013). The big four - What we did wrong in advanced persistent threat detection? In *Proceedings - 2013 International Conference on Availability, Reliability and Security, ARES 2013*. Recuperado de: https://doi.org/10.1109/ARES.2013.32

Caracterización de elementos para la creación de una herramienta computacional para la gestión del conocimiento en las organizaciones[1]

Diana María Montoya Quintero[2]; Jovani Alberto Jiménez Builes[3]

Resumen

Se hace un estudio minucioso sobre algunos elementos que son caracterizados para la extracción de conocimiento humano, los cuales tienen como propósito el desarrollo de una herramienta que permita conservar el conocimiento del capital intelectual dentro de las organizaciones. Surge una problemática en la gestión del conocimiento, cuando se quiere llevar a la transferencia tecnológica las experiencias y buenas prácticas en la labor de cada individuo dentro y fuera de estas organizaciones, esto debido a que cada experto en un área determinada se le dificulta expresar sus saberes y aplicaciones vividas a lo largo de su permanencia en la organización, y tan pronto se retiran de estas, se llevan ese saber. Es necesario proponer herramientas inteligentes que simulen el conocimiento de los expertos humanos para que permanezcan dentro de las organizaciones y no se vayan con el individuo.

Palabras clave: extracción, conocimiento, ingeniería, sistemas, gestión.

1 Capítulo de libro resultado del proyecto de investigación "Modelo de extracción del conocimiento humano aplicando razonamiento basado en casos para sistemas basados en conocimiento". Producto del doctorado realizado en la Universidad Nacional de Colombia sede Medellín por Diana Montoya.

2 Lic. Doc. en Computadores, Magister en Ingeniería de Sistemas, Doctora en Ingeniería de Sistemas e Informática. Docente investigadora asociada al Grupo de Investigación AGLAIA de la Corporación Universitaria Americana. ORCID: https://orcid.org/0000-0003-4195-9480.E-mail:dmmontoya@americana.edu.co.

3 Lic. Doc. en Computadores, magister en ingeniería de sistemas, doctor en Iingeniería de sistemas e informática. Docente investigador de la Universidad Nacional de Colombia. correo electrónico: jajimen1@unal.edu.co

Characterization of elements for the creation of one computational tool for knowledge management in at organization

Abstract

A meticulous study is made of some elements that are characterized for the extraction of human knowledge, which has as purpose the further development of a tool that allows to conserve the knowledge of the intellectual capital within the organizations. A problem arises in the management of knowledge when it is wanted to take to the technological transfer the experiences and good practices in the work of each individual inside and outside the organization, where these experts find it difficult to express their knowledge and applications lived throughout of their permanence in the institutions, and as soon as they are removed from them, they take that knowledge. It is necessary to propose intelligent tools that simulate the knowledge of human experts so that they remain within the organizations and do not leave with the individual.

Key words: extraction, knowledge, engineering, systems, management.

Introducción

En el recorrido de la lectura de este capitulo, se puede observar cómo los sistemas que se basan en el conocimiento humano (SBC), son una promesa de transferencia tecnológica en las teorías de la gestión del conocimiento; a la medida que estos sistemas hacen un gran aporte en la sistematización, administración, procesamiento y control del conocimiento de un área o dominio dentro de una organización. Los SBC son considerados una técnica de la Inteligencia Artificial (IA), los cuales se especializan en simular el conocimiento que tiene un experto humano Bjornson (2008).

Aquí se presenta una perspectiva de elementos que han sido caracterizados para la creación de una herramienta computacional en gestión del conocimiento dentro de una organización, dichos elementos serán la base para un nuevo modelo de extracción de conocimiento humano dentro de las mismas.

Lo anterior, con el interés de dar repuesta a la continua problemática centrada en la necesidad de análisis, abstracción, comprensión y conservación del capital intelectual de los seres humanos en las organizaciones.

Entre las tendencias del desarrollo de software, existe un enfoque de crear herramientas que realmente se acerquen más al conocimiento; este tipo de sistemas proponen resolver problemas del mismo modo como lo hace el experto (rol de la persona que tiene el conocimiento dentro o fuera de una organización). Los sistemas de este tipo necesitan de la capacidad de aprendizaje para alcanzar nuevos conocimientos, eliminar errores, mejorar el conocimiento existente y ordenar el almacenamiento, lo cual puede crear condiciones de ventaja para el razonamiento real del dominio para el cual se va a desarrollar.

La Ingeniería del Conocimiento (IC), es una fase encargada de representar formalmente el conocimiento de un experto humano y de gestionar el proceso para la implementación de un SBC.

Teng, Chen, y Xian (2016) definen como diferentes precursores de la Inteligencia Artificial han venido trabajando con otras técnicas de análisis, mencionando el Razonamiento Basado en Casos; el cual posee un estimado como una técnica de solución de nuevos problemas que se apoya en las soluciones de problemas anteriores. Ambas disciplinas (IC & RBC) trabajan de manera independiente para generar resultados dentro de cada necesidad. Dentro de este proyecto es un desafío tomar elementos que las componen para el futuro desarrollo de un modelo de extracción.

El RBC es un método de apoyo para elicitación de requisitos, ya que esta técnica se inicia desde un caso en particular sin tener como fuente principal el experto humano, contrario a la IC que se fundamenta en este principio, pero no tiene en cuenta los casos dentro de las metodologías existentes.

Para la caracterización de los elementos, se involucraron cuatro tipos de conocimiento que pueden ser procesados por una computadora como son: conocimiento procedimental, declarativo, cognoscitivo y de transferencia tecnológica *(desde un conocimiento tácito a un conocimiento explicito definido por Tseng, y Lee (2014))*

La adquisición de conocimiento es tradicionalmente considerada como una de las actividades más complejas dentro de la IC (Kendal, y Creen, 2007).

¿Cómo hacer eficiente la extracción de conocimiento de un experto humano para transferirlo a una computadora?

Para comprender que es el conocimiento dentro de los seres humanos y el como transferirlo a una computadora, partimos de algunos antecedentes históricos sobre la teoría del conocimiento, teniendo en cuenta un acercamiento de algunos filósofos y literarios, donde se comparten concepciones diferentes:

El conocimiento

Como dice Broyncote (sf) mencionando a Kant: "Aunque todo nuestro conocimiento empieza con la experiencia, no es procedente (pensar) que todo él surja de la experiencia". Los cambios en el paradigma del liderazgo organización de una región son incúmbete frente a sus procesos, estos cambios son conocidos de acuerdo a Kendal, y Creen (2007) como "*catch-up cycles*", el cual considera que la innovación tecnológica desempeña un papel clave en el proceso de avance tecnológico, el autor Teece, (2016) expone como elementos básicos de la innovación, y el explorar la relación entre la acumulación del conocimiento y la innovación tecnológica ha sido uno de los puntos clave en la investigación académica.

Klausinger (2007), con respeto a las teorías de Hegel, publicado en la página palmera (sf), considera que el conocimiento empieza con la percepción sensorial, la cual se vuelve más subjetiva y racional a través de una purificación dialéctica de los sentidos, y finalmente alcanza la etapa del "Espíritu absoluto" que es tener conocimiento. El concepto anterior nos da una perspectiva del elemento de "razón", manera de que cada individuo pueda identificar conceptos, cuestionarlos, hallar coherencia o contradicción entre ellos de acuerdo con las relaciones pertinentes para conectar actividades prácticas y restricciones cognitivas que se obtienen en la categorización del saber.

El conocimiento es un proceso humano y dinámico que se orienta a algún fin con intención y perspectiva, es específico y atiende al contexto donde se genera; es individual antes que grupal y que se asocia con la pericia, la competencia y la capacidad de actuar de cada individuo. Puede ser según definiciones Ahmad et al. (2017) tácito, dinámico, delimitado y movible.

Syed (2014) menciona a Marshall indicando que: "En gran parte el capital consiste en conocimiento y organización. El conocimiento es la máquina de producción más poderosa a nuestro alcance, la organización ayuda al conocimiento". En la interpretación de Tseng, y Lee (2014) ratifican el valor del conocimiento para ser llevado y manipulado en la sociedad y en la naturaleza de lo laboral, en el hacer que se da dentro de una organización. Existen otros autores que se unen a esta teoría como la administración científica de Taylor, la cual se encarga de la conversión de las habilidades tácitas y las experiencias de los trabajadores en conocimiento científico equitativo. Sin embargo, no se tuvo dentro de esta corriente, consideraciones de las experiencias y los juicios de los trabajadores como una fuente de nuevo conocimiento. En esta teoría se da una reflexión sobre la clasificación, tabulación y reducción del conocimiento a reglas y a fórmulas que deberían aplicarse en el trabajo diario.

Otra orientación al conocimiento es dada por el paradigma cognitivo que centra sus esfuerzos en entender los "procesos mentales" y las "estructuras de la memoria" con el fin de comprender la conducta humana. La imagen que plantea proyectar el cognitivismo en la mente humana es, "el aprendizaje", como en la vida cada persona es el arquitecto de su propio conocimiento, interpretación de Marano (2012).

El conocimiento es un poder social que mueve organizaciones y grupos de seres humanos con intereses propios y grupales, para beneficio común. La importancia del conocimiento genera productos indispensables para el progreso y avance de la humanidad.

El autor Kruger y Johnson (2011) en su análisis de las teorías de Hayek, clasificó el conocimiento en conocimiento científico (por ejemplo, reglas generales) y de las circunstancias particulares de tiempo y espacio, y sostuvo que las circunstancias cambiantes redefinen continuamente la ventaja relativa que un individuo pude tener en cuanto a conocimiento.

Una vez el conocimiento es tratado por el mismo ser humano, crea ciencia y principios propios para iniciar procesos, innovación, cultura, generaciones y nuevas tecnologías.

El conocimiento ha sido estudiado por diferentes disciplinas, tales como la sicología, la filosofía, la epistemología, la hermenéutica, entre otras mencio-

nadas por Tseng y Lee (2014) una de las teorías filosóficas que afirma que el conocimiento emana de dos fuentes principales, en intervenciones de Kendal y Creen (2007), se descifra la capacidad de recibir impresiones sensibles, así como la facultad del entendimiento de conocer ese objeto por medio de estas, nuestra mente conoce por medio de intensiones puras que son las formas a priores de la sensibilidad y por medio de las categorías del entendimiento que son los conceptos puros e independientes de la experiencia .

De las hipótesis anteriores, se relacionan las búsquedas de la selección de elementos para su caracterización. El cómo se acerca la concepción a partir del hecho, a la acción, y la permanencia, dejando como objeto participante en el conocimiento la experiencia de cada vivencia.

Extracción y taxonomía del conocimiento

De otro lado, Anderson y Krathwohl (2001) estipulan el conocimiento como la capacidad de recordar hechos específicos universales, métodos, procesos, esquemas, estructuras o marcos de referencia, puesto que cualquier cambio ya implica un proceso de nivel superior. Tienen varios elementos base dentro del conocimiento humano, como los que se pueden observar en la figura 1.

Figura 1.
Taxonomía del conocimiento de acuerdo a Bloom.
Fuentes:
Elaboración del autor a partir de Anderson & Krathwohl (2001) y Richard (2004)

Conceptualización cognoscitiva

En la recopilación de diferentes conferencias de autores como Cambridge y Dartmouth contemporáneos de Herbert Simón, Marvin, y John McCarthy (siendo estos, precursores en áreas diversas; la lingüística, la psicología, la neurología y la inteligencia artificial), se basan en una misma hipótesis cognitivista: "la mente es una forma lógica asimilable al comportamiento de un computador" (Varela, 1989).

A la corriente cognitiva se le considera como un referente de las ciencias del conocimiento. Basándose en la psicología para exponer los diferentes procesos mentales, bajo un patrón de la mente humana como un sistema. La potestad de la computación y los sistemas basados en el conocimiento humano pueden compartir con las teorías de Von Neuman, Newell, Simón, donde se manifiesta como "la memoria humana se representa tal espacio direccionable y organizado, en el que se almacenan unidades discretas de conocimiento que pueden ser recuperadas automáticamente o mediante una búsqueda premeditada" (Newell, 1983).

En otra instancia, es importante vincular el conocimiento con la IA quien ha venido incorporándose sobre los diferentes desenlaces lógicos que puede causar la mente de los seres humanos, para representar el razonamiento a través de un proceso simbólico, donde se aplican diferentes métodos, metodologías y técnicas para ser codificado y procesado por una computadora. Sin embargo, a la hora de la adquisición de conocimiento, sólo tiene sentido si posteriormente se planea hacer algo con ese conocimiento, como lo determina Milton (2007).

En el *Institute for Human & Machine Cognition* dentro de sus investigaciones considera como la ciencia cognitiva es relativamente joven, redefiniéndose a medida que evoluciona y que las nuevas ideas y puntos de vista revelan nuevos métodos y técnicas constantemente. En este instituto se observa como profesores de filosofía se interesan en las elucidaciones de las bases de investigación cognitiva, la computación y la información.

Uno de los objetivos de la Inteligencia Artificial era visto como la creación de mecanismos de inteligencia, que imitan la conversación humana. Para las investigaciones que se realizan dentro de la institución IHMC, se centran en

lo humano teniendo como alternativa la computadora, quien complementa como prótesis el centro del conocimiento que es el sujeto.

Cada precursor del cognitivismo, lleva a buscar la conservación del conocimiento humano para beneficio social, organizacional y cultural, determinando como la tecnología y las maquinas computacionales pueden brindar una solución.

Existen técnicas y métodos naturales que emergen de igual forma dentro de la extracción y taxonomía del conocimiento así:

Ontología: técnica encargada de hacer una descripción de una conceptualización; es decir, una distribución contextual normalizada y de aceptación no sólo para acumular la información, sino también para examinarla y rescatarla.

La estadística bayesiana: accede a la información subjetiva sobre arquetipos ignorados dentro de una manifestación bajo estudio, con el fin de realizar dichas apreciaciones. Los parámetros de la información subjetiva se realizan mediante un proceso conocido como "adquisición".

Las redes semánticas: se enfocan en la concepción de que los objetos (cualquier ente o cosa) o los conceptos logran ser incorporados por algún tipo de analogía. Estas correspondencias se simbolizan utilizando una unión que conecte dos o más impresiones. Peng (2014).

Herramientas para representar el conocimiento

PROforma y Tallis: es un lenguaje de representación del conocimiento que se puede utilizar para establecer descripciones de métodos, los cuales se desarrollan con el tiempo y requieren la cooperación de diversos actores. Uno de los beneficios del lenguaje es la facilidad para su uso dentro de un entorno gráfico. El proceso se compone de objetos extraídos de las clases establecidas en un diagrama de clases UML, cada clase de objeto tiene un conjunto de propiedades, y cada instancia de una clase tiene diferentes negocios. Peng (2014).

Lenguaje de Modelado Unificado (UML, 2015): aplicado en la etapa de análisis y diseño de sistemas de software convencionales para hacer el modelado o abstracción de requerimientos funcionales y no funcionales del sistema. Contiene varios diagramas con la intención de representar de forma dinámica y estática los procesos y funcionalidades del producto final.

Técnica Delphi: Gilson, Brown, Faulkner, Cena, Murphy, Pringle, Proper Puig, y Stathi (2009) definen esta técnica como un método de estructuración de un proceso de informes grupales positivos a la hora de permitir a un grupo de personas, tratar un problema complejo. El método emana de la interpelación a expertos humanos con la ayuda de sondeos sucesivos, a fin de colocar en manifiesto tendencias de veredictos y concluir casuales aprobaciones. La encuesta se lleva de una manera anónima y su equitativo es "disminuir el espacio intercuartil precisando la mediana".

Las herramientas anteriores son parte del diseño del sistema y son procesos de la ingeniería de software, para crear un esquema que facilite al desarrollador la codificación. Si bien se observa su aporte en un producto de software para el cumplimiento de estándares de calidad, no es orientado a la IC, ni a lenguajes declarativos. No tienen una orientación en el saber del experto humano sino en la funcionalidad de los requerimientos para el desarrollo de un sistema.

Técnicas, modelos y métodos que se acercan al proceso de adquisición de conocimiento

En algunos escritos de Borrajo et al. (1993), se narra cómo en la década de los 80 salieron lenguajes formales de programación y sistemas que también simbolizaban el conocimiento. Se codificaron grandes planes del conocimiento en lenguajes como Prolog, el cual representa estipulaciones y lógica básica, y puede derivar conclusiones de premisas conocidas. Otro lenguaje es el KL-One creada en los 80 más orientado a la representación del conocimiento en sí, trabaja a través de redes semánticas y marcos ontológicos para incorporar explícitamente información conceptual como una red de herencia.

Desarrollos en la representación del conocimiento sistemático

XML: Esta es una tecnología que busca dar solución al problema de expresar información estructurada de la manera más abstracta y reutilizable. Cuando la información es estructurada se compone de partes que son definidas, y estas se componen de otras partes. Entonces se tiene un árbol de trozos de información. El XML proviene de un lenguaje creado por IBM en los años 70. El lenguaje de IBM se llama GML (*General Markup Language*) y surge por la necesidad que tenían las organizaciones de almacenar grandes y diversos temas de información (IBM, 2014).

Web semántica: la cual se ratifica en la idea de añadir metadatos semánticos y ontológicos a la *World Wide Web*. La información adicional que describen el contenido, el significado y la relación de los datos, se deben proporcionar de manera formal, para que así sea posible evaluarlas automáticamente por máquinas de procesamiento. El objetivo es mejorar Internet ampliando la interoperabilidad entre los sistemas informáticos usando "agentes inteligentes" (Berners, Hendler, y Lassila, 2001).

Resource Description Framework, (W3C, 2014), (RDF): el Marco de Descripción de Recursos es un modelo estándar para el intercambio de datos en la Web, tiene características que facilitan la fusión de datos, incluso si los esquemas son diferentes, apoyando la evolución para intercambios de esquemas. Fue creado para acciones de la World Wide Web Consortium (W3C) originalmente diseñado como un modelo de datos (Lenguaje orientado a base de datos) para metadatos (datos que describen otros datos). Se utiliza como un método general para la descripción conceptual o modelado de la información que se implementa en los recursos de la web, utilizando una variedad de notaciones de sintaxis y formatos de series.

DARPA Agent Markup Language (DAML) (CSL, 2014): Fue un programa para la Agencia de Proyectos de Investigación Avanzados de la Defensa en los Estados Unidos conocido por sus siglas en ingles DARPA. El programa enfocó la creación de representaciones legibles por máquina para la Web. En el momento tiene mayor capacidad que XML describiendo objetos y relaciones entre los objetos, expresando semánticas que permiten crear un mayor nivel de interoperabilidad entre sitios Web. En conclusión, actualmente su objetivo es crear tecnologías que permitirán a agentes de software identificar y entender

de forma dinámica las fuentes de información, proporcionando interoperabilidad entre los agentes de una manera semántica.

Web Ontology Language (OWL) (W3C, 2014): es un lenguaje de marcado que permite compartir datos aplicando ontologías (formulación de esquema conceptual, fundamentado en la representación gráfica o simbólica de un concepto) dentro de uno o varios dominios dados; con la finalidad de facilitar la comunicación y el intercambio de información entre diferentes sistemas y entidades, el objetivo central es facilitar un modelo de marcado construido sobre RDF y codificado en XML.

Sistemas de Extracción de Información: cuyo propósito consiste en detectar la información que es relevante dentro de un conjunto de textos, ignorando la no relevante, y estructurarla para su almacenamiento en una base de datos.

Los sistemas de Búsqueda de Respuestas: que tienen como objeto dar una respuesta concreta a la pregunta formulada por el usuario.

Los sistemas de Generación de Resúmenes: que se centran en condensar la información más relevante de un texto.

Finalmente se tiene la concepción de diferentes autores quienes proponen predicciones cualitativas con datos cuantitativos, usando técnicas bayesianas. La estadística Bayesiana no exige numerosos datos históricos, se puede basar en conocimientos a priori o de expertos frente al tema, facilitando el pronóstico, como se muestra en diferentes trabajos, una técnica bayesiana parte de realizar un producto entre una distribución a priori para parámetro (s) y una función de posibilidad en los datos, y así obtener una distribución conjunta a posteriori; se completa el producto entre esta y la función de los datos sobre el rango de medidas para finalmente obtener la distribución predictiva para pronosticar.

Ingeniería del conocimiento

Centrándonos en los procesos de la IA para obtener productos codificados pasamos a la Ingeniería del Conocimiento (IC) quien, tiene como objetivo,

analizar el saber específico de un experto humano, para luego ser representado en un diseño que permita la comprensión del codificador en el desarrollo de un SBC o un sistema experto (SE) de tal forma que cuando se implemente el sistema, simule el conocimiento que fue transferido en un modelo dinámico (primera etapa del desarrollo de un SBC), quien es una respuesta al análisis realizado. La IC, verifica el conocimiento como la ingeniería de software los requerimientos funcionales y no funcionales en el ciclo de vida de su desarrollo.

En esta primera etapa de análisis y representación de conocimiento intervienen dos actores principales conocidos como agentes: un ingeniero de conocimiento, quien es el encargado de analizar y representar el conocimiento para el desarrollo y construcción del SBC y un experto humano, que es la persona humana que tiene el conocimiento para ser representado.

Dentro de la literatura se concluye que la adquisición de conocimiento es el punto que plantea una mayor dificultad a la hora de crear una base de conocimiento como se observa en los cuadros de texto de la figura 2.

<table>
<tr><th colspan="3">Cuello de botella en la extracción de conocimiento
de un humano aplicando ingeniería del conocimiento</th></tr>
<tr><td>El ingeniero del conocimiento no es un experto en el campo que intenta modelar, mientras que el experto en el tema no tiene la experiencia modelando su conocimiento</td><td rowspan="2">El proceso de desarrollo de un sistema basado en conocimiento se ve constantemente amenazado por los problemas que aparecen al extraer el conocimiento del experto para luego codificarlo en un sistema computacional.
(Motta, 1999), (Davis & Sarkani, et. al 2011)(Kerschberg L. 2005) (Nalepa & Adrian, 2012).</td><td rowspan="2">Evolución tanto en la fase de inicio como en la de uso de un sistema basado en conocimiento, por incorporación de nuevos conocimientos o porque los expertos reconsideren la forma en la que se ha expresado su conocimiento</td></tr>
<tr><td>Dar forma automática y manipulable a eso que dijo por el experto; ya que la adquisición del conocimiento no puede ser un problema de entrevistas informales, debido a que debe desarrollar métodos explícitos con propósitos específicos (Iwazum & Kaneiwa, 2013).</td></tr>
<tr><td>El proceso de desarrollo de un sistema basado en conocimiento se ve constantemente amenazado por los problemas que aparecen al extraer el conocimiento del experto para luego codificarlo en un sistema computacional.
(Motta, 1999), (Davis & Sarkani, et. al 2011)(Kerschberg L. 2005) (Nalepa & Adrian, 2012).</td><td>El experto describe lo que hace, pero no sobre lo que realmente pasa.
El Ingeniero del conocimiento puede no estar inmerso en el dominio del problema
(Davis & Sarkani, 2011) (Kerschberg L. 2001) (Nalepa & Adrian, 2012)</td><td>Estructuras de requerimientos en los que se pueden encontrar requisitos subjetivos, entradas inconsistentes por parte del experto humano o incompletas con un alto grado de incertidumbre, y no pueden ser resueltos por algoritmos clásicos o la investigación operativa (Guida, & Tasso, 1994)</td></tr>
</table>

Figura 2. Cuello de botella en la extracción de conocimiento de un humano aplicando ingeniería del conocimiento. **Fuente:** elaboración de los autores

Sistemas Basados en Conocimiento

Al pasar del proceso de la ingeniería del conocimiento, este es transferido al desarrollo de un SBC, el cual cumple con varias etapas como se pueden observar en la siguiente figura 3.

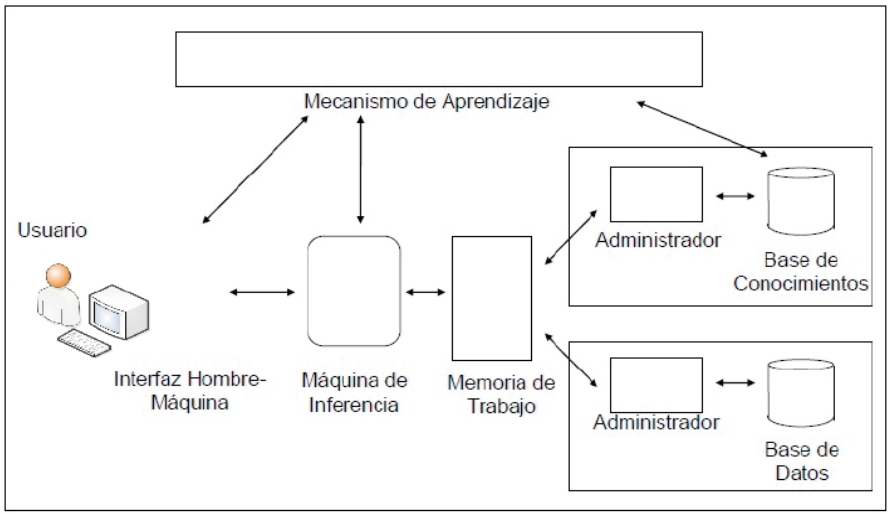


Figura 3. Elementos que componen un sistema basado en conocimiento humano.
Fuente: adaptación de los autores.

Cada uno de los elementos de la figura 3, está conformado por un modelo de negocio de intervención del conocimiento que tiene un experto humano y una computadora, con la intervención de un ingeniero encargado de extraer el conocimiento, para así generar un mecanismo de aprendizaje a la maquina en su codificación. Ese aprendizaje se encarga de almacenar el conocimiento analizado y pre procesado por aquellos que intervienen en cada una de los componentes de la estructura anterior.

De acuerdo a Schreiber et al. (1999), entre las metodologías que se emplean para el desarrollo de un SBC se aborda el diseño y desarrollo de algunas ya existentes como: CommonKads, MIKE y PROTÉGÉ , en este capítulo, se hace un resumen de la metodología CommonKads por considerarse actualmente más adecuada y precisa para adquirir el conocimiento para crear un SBC.

CommonKads es considerada como una metodología completa para el desarrollo de SBC. La metodología describe las principales técnicas, lenguajes de modelado, y estructuras argumentadas para ayudar en tres fases de la construcción de un sistema basado en el conocimiento.

La fase de "análisis contextual" se centra en la organización que con el tiempo utilizará el sistema, describiendo el proceso de negocio de la misma, tiene vigente los recursos y activos de conocimiento de la organización, así como la descripción de los efectos que el SBC tendrá. Uno de los objetivos de esta fase contextual es demostrar la viabilidad propuesta y los beneficios que traen a la organización la adopción e implementación, al igual que su beneficio en costos. También proporcionan más detalles sobre las tareas que la organización y el personal llevan a cabo, o algunos sistemas, que pueden hacer esta labor.

En la segunda fase de "análisis conceptual" la metodología se utiliza para aclarar el conocimiento que el SBC requiere para ser representado, al igual que el razonamiento necesario para llevar a cabo, y las interacciones necesarias con los usuarios internos y externos.

La tercera fase de "diseño" de la metodología se utiliza para crear el diseño del SBC para ser interpretado por un ingeniero del conocimiento para la codificación en la computadora por fases así: Fase de contexto, fase conceptual, y una fase de diseño.

Sin embargo, esta metodología tiene sus contras, a la medida que es bastante compleja y extensa, además no existe suficiente información relevante en su uso, lo que pone en riesgo su acceso y comprensión, si bien se encuentran ejemplos son superficiales y no se muestran observaciones completas de su aplicación.

Caracterización de elementos para la creación de una herramienta computacional para la gestión del conocimiento en las organizaciones un software basado en conocimiento

La representación del conocimiento se convierte en un componente decisivo para el desarrollo de un SBC. Como se indicó anteriormente, existen algunos modelos de representación adecuado en algunos casos según el tipo de programación o codificación en la que se va a crear. Teniendo en cuenta

los esquemas generados por los productos de software se analizan dos tipos en particular entre ellos: el procedimental y el declarativo (*cuyo motivo se da a que se acercan a la lógica que lleva consigo la técnica de RBC y los SBC y las concepciones teóricas del conocimiento en sí*).

Representación del conocimiento procedimental

Se refiere a las habilidades cognitivas o motrices (acciones, actividades, hechos, y experiencias que son realizadas por el hombre). Por su carácter dinámico y difícilmente interpretado en el lenguaje humano, este conocimiento es más difícil de modelar. El lenguaje natural, en ocasiones determina las acciones para admitir un hecho o un suceso como cierto, con la seguridad completa del significado, naturaleza o explicación de alguna cosa o fenómeno. Este tipo de conocimiento se visualiza en uno de los procesos de la ingeniería del software en la etapa de adquisición de requerimientos para la implementación de un sistema de software (cabe anotar que los sistemas inteligentes o pertenecientes al área de la IA son productos también de la ingeniería de software, solo que el desarrollo difiere en su programación).

Representación del conocimiento declarativo

Hace una descripción detallada de los hechos y eventos que pueden ser recordados conscientemente. Este conocimiento puede ser representado por medio de listas proposicionales, y de redes semánticas o de marcos "frames" Woolfolk, et al. (2010).

Al hablar de extracción de conocimiento implica: la representación de ese conocimiento, y el procesamiento que hace la computadora a través de representaciones compuestas por objetos explícitos encargados de permitirles sacar conclusiones de conocimiento almacenados en un medio de información.

Para representar cualquier tipo de conocimiento en una computadora, es indispensable tener presente los pasos, metodologías, técnicas e instrumentos que la IC pueda otorgar para apoyar el proceso. Paras pasar a la IC se hace necesario destacar la propuesta de varios autores que se dedican a los estudios y teorías de la Gestión del Conocimiento (GC) Se coinciden con quienes coinciden en los siguientes procesos:

El conocimiento es un proceso humano y dinámico que es específico y atiende al contexto donde se genera; que es individual antes que grupal y que se asocia con la pericia, la competencia y la capacidad de actuar de cada individuo.

Del análisis y clasificación de las teorías y antecedente anteriores se obtienen la caracterización de elementos que componen la extracción de conocimiento de un experto humano para la creación de una herramienta computacional de gestión del conocimiento en las organizaciones.

Caracterización de los elementos

Para seleccionar los elementos, se tuvo en cuenta cada numeral de este capítulo, de acuerdo al grado de importancia dada por cada autor y la similitud de elementos que presentan entre sí. Basados en los criterios dados por cada uno de ellos, se unifican varios de estos elementos (a los cuales les damos en adelante el nombre de datos o variables) para obtener un conjunto aglomerado para su validación, observar el grupo en la tabla 1.

Tabla 1. *Elementos seleccionados para la caracterización.*

Memorizar	Inventar	Gestión del conocimiento	Lenguaje inferencia
Aprender	Diseñar	Organización	Recordar
Comprender	Clasificar	Reglas	Reconocer
Analizar	Pensamiento lógico	Seleccionar	Información
Aplicar	Mejorar	Justificar	Evaluar
Crear	Predecir	Caso	Razonar
Explicación	Modelar	Optimizar	Efecto
Transferencia tecnológica	Derivar	Calcular	Modelo
Proponer	Interpretar	Resolver	Actividad
Reutilización	Representación	Determinar	Procedimiento

Fuente: elaboración de los autores

Los elementos anteriores nos acercan más a una aproximación de la importancia de la gestión del conocimiento humano, relacionado con las teorías propias de esta disciplina del conocimiento dentro de las organizaciones: desde lo tácito y lo explícito. Diferentes autores entre ellos Inkinen (2016), reafirma las teorías sobre la clasificación de estos dos conocimientos:

El conocimiento Tácito: es el conocimiento de la persona, resultado de la experiencia, qué es versado debido a que se utiliza para actuar, está incluido en cada ser humano es quien implica ideales, valores y emociones de cada persona. No se transporta con facilidad entre las personas.

El conocimiento explícito: es aquel que se puede codificar, sistematizar, con resultado del procedimiento y la racionalidad es secuencial y teórico, puede adaptar la forma de programas informáticos, patentes, diagramas o similares. Es transferible entre las personas, por lo tanto, es trascendental en la generación de conocimiento.

La GC es vista en este apartado, como un ciclo de administración y tratamiento de la información, para que sea recreada dentro de la organización, mediante mecanismos de asimilación y captación que presente soluciones prácticas y generen un nuevo conocimiento. La GC "encarna el proceso organizacional que busca la combinación sinérgica del tratamiento de datos e información, a través de las capacidades de las tecnologías de información y de creatividad e innovación de los seres humanos" Andreeva y Kianto (2012) en esta misma línea se evidencia un proceso sistemático para organizar, filtrar y presentar la información con el objetivo de mejorar la comprensión de las personas en un área específica de interés.

Estos acercamientos conceptuales de la ingeniería y la gestión del conocimiento, derivan una gran fortaleza para las organizaciones, ya que, si la gestión del conocimiento se encarga de administrar y transferir el conocimiento con sus diferentes teorías para la mejora continua dentro de los procesos de una organización, la IC interfiere a la hora de tratar el conocimiento del saber de un experto humano para ser modelado y simulado en la computadora. Un análisis más detallado de esta evolución histórica, y de otras metodologías basadas en el modelado se puede ver en Joo, et al. (2016).

La figura 4, muestra la construcción a una aproximación de un nuevo modelo para el desarrollo de una herramienta de gestión de conocimiento en una

organización, la cual tiene como objetivo el apoyo adecuado para la vinculación y uso de los sistemas de gestión de conocimiento, cuando se involucre la relación del conocimiento y el aprendizaje relacional con una variable de cooperación en cada uno de los elementos seleccionados, de acuerdo a las causales que acompañen un efecto.

Estos elementos relacionan diferentes procesos: uno de ellos es el negocio del conocimiento para la captura del dominio del un experto humano en un tema específico, otro es el almacenamiento del conocimiento en bases de datos, las cuales buscan el dominio a través de diferentes fuentes del saber de las organizaciones desde el capital intelectual, hasta bases de datos existentes (En sus diferentes tipos: relaciones, orientadas a objetos, transaccionales, entre otras), con el objetivo de identificar el conocimiento para su clasificación y relación para los conceptos del dominio, de igual forma con el almacenamiento y recuperación de la indexación de datos para generación de inferencias con los algoritmos basados en reglas.

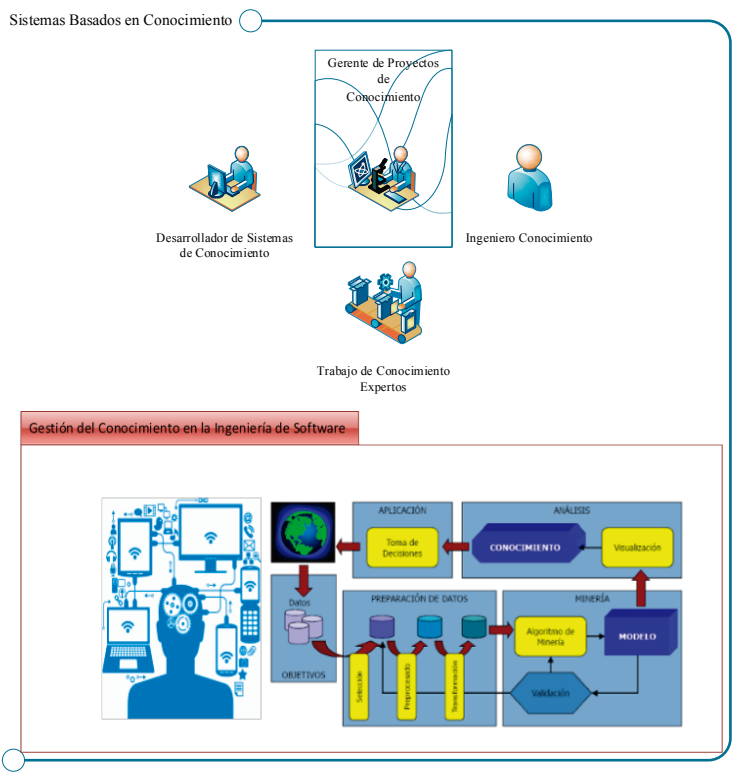


Figura 4. Elementos para el desarrollo de una herramienta de gestión de conocimiento en una organización. **Fuente:** Adaptación de los autores.

Dentro de estos elementos, se observa a su vez una tendencia en la vinculación de las tecnologías de la información y la comunicación como estrategia para hacer a través de un portal corporativo, su divulgación, incluyendo; mensajería electrónica sincrónica o asincrónicamente, buscando los intereses de cada miembro de la organización, y dándose una colaboración entre estos para compartir el conocimiento tácito y combinarlo con el explícito en la resolución del dominio. Donde se reconozca la heterogeneidad de las fuentes del conocimiento y se establecen diferentes componentes que integren nuevos conocimientos.

En conclusión, la figura 4 integra los elementos de la tabla 2, en un conjunto de procesos orientados a un modelo conceptual, el cual muestra los componentes de un sistema basado en la gestión del conocimiento humano dentro de las organizaciones, donde va a la mira de los servicios que se definen en una combinación de un modelo de negocio y las metas establecidas de forma estratégica para conservar el conocimiento y la información no estructurada, al igual que servicios basados en los procesos normalizados e información estructurada.

Conclusiones

La gestión del conocimiento, controla y administra el juicio humano para preservarlo o transformarlo de tácito a explicito, y la ingeniería del conocimiento entra de forma potencial, para permitir procesar y codificar ese conocimiento con estándares propios de transferencia tecnológica, lo que puede generar el éxito de los procesos y liderazgo en el capital intelectual de las organizaciones. Este capítulo nos incita a desarrollan un soporte metodológico que precise la representación del conocimiento para ser leído e interpretado por un ingeniero de software, o el mismo experto humano, basados en los elementos seleccionados para el mismo.

Cada desarrollo de un producto de software o un SBC, debe pasar por la etapa de su ciclo de vida tecnológico, enmarcado desde el análisis de conocimientos, su modelado y diseño para ser codificados por una computadora. Este ciclo sucede una vez se identifica un problema que no ha sido resuelto dentro de la organización, aun dentro de las organizaciones no se tiene determinada la forma de retener el conocimiento del Capital Humano, que com-

prende la competencia, conocimiento, valores y potencial innovador de los individuos que están dentro de ellas. Día a día se busca una reestructuración necesaria, para implementar departamentos con una reingeniería que genere continuamente estrategias para su buen funcionamiento y así establecer negocios para un mejor desarrollo en su economía.

Referencias

Ahmad, N., Lodhi, M. S., Zaman, K., & Naseem, I. (2017). Knowledge management: A gateway for organizational performance. *Journal of the Knowledge Economy, 8*(3), 859–876.

Anderson, L., y Krathwohl (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

Andreeva, T. y Kianto, A. (2012). Does knowledge management really matter? Linking knowledge management practices, competitiveness and economic performance. *Journal of Knowledge Management, 16*, 617e636

Berners, L. Hendler, J. Lassila, O. (2001). The Semantic Web. *Scientific.33*, pp. 600 – 684.

Bjornson, F., y Dingsoyr, T. (2008). "Knowledge management in software engineering: A systematic review of studied concepts, findings and research methods used," *Inf. Softw. Technol 50*, (11), (pp. 1055–1068).

Borrajo, D. , Juristo, N. , Martínez V. y Pazos, J. (1993) *Inteligencia Artificial. Métodos y Técnicas*. Madrid: Centro de Estudios Universitarios Ramón Areces.

Davis, S. Sarkani, S. & Mazzuchi, T. (2011). What's at STEAK? Exploring engineering methodologies to identify existing generational boundaries impeding the strategic transfer of engineering and architectural knowledge (STEAK). *Computer Supported Cooperative Work in Design (CSCWD), 15th International Conference* . Digital Object Identifier: 10.1109/CSCWD.2011.5960214, (pp. 827 – 834)

Eriksson, H. Shahar, Y. Tu, S. Puerta, R. y Musen, M (1995). Task modeling with reusable problem solving methods. *Artificial Intelligence, 79*(2):3-26.

Gilson, N., Brown, W., Faulkner, G., Mckena, J., Murphy, M., Pringle, A., Proper K., Puig, A., Stathi, A. (2009). The International Universities Walking Project: development of framework for workplace intervention using Delphi technique. *J Phys Act Health. 6*(4), 520-8.

Guida, G, Tasso, (1994) Design and Development of Knowledge Based Systems. From: Wiley. J. et al. *Life cicle to methodology*. England.

Inkinen, H. (2016). Review of empirical research on knowledge management practices and firm performance. *Journal of Knowledge Management, 20*(2), 230–257.

Iwazum, M. and Kaneiwa, K. (2013) Community-Driven and Ontology-Based Biological Knowledge Management: A Hybrid Approach to Harnessing Collective Intelligence Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed 140 Computing *(SNPD), 14th ACIS International Conference.* Digital Object Identifier: 10.1109/SNPD.2013.113, pp 387 – 393.

Joo, B. K., Park, J. G., & Lim, T. (2016). Structural determinants of psychological wellbeing for knowledge workers in South Korea. *Personnel Review, 45*(5), 1069–1086.

Kendal, S., y Creen M. (2007) *An Introduction to Knowledge Engineering,* London: Springer Press.

Kerschberg, L. (2005) *Knowledge Management in Heterogeneous Data Warehouse Knowledge Base*. Recuperado de: http://eceb.gmu.edu/pubs/KerschbergDaWak2001.pdf.

Klausinger, H. (2007). The Making of F. A. Hayek´s Monetary Theory and the Trade Cycle. Working Paper presentado en la *Conferencia de la ESHET*. Estrasburgo, el 5 de julio de 2007.

Kruger, C., y Johnson, R. (2011). Is there a correlation between knowledge management maturity and organizational performance? *Vine 41*, 265e295. http://dx.doi.org/10.1108/03055721111171618.

Lee, K. y Malerba, F. (November, 2016) Catch-up cycles and changes in industry leadership: Windows of opportunity and responses of firms and countries in the evolution of sectoral systems. *Research Policy, 46*, 338-351, doi: 10.1016/j. respol.

Marano, E., (2012), Conductismo, Cognitivismo y Constructivismo. Cuadro Resumen realizado con fines didácticos. *Performance Improvement Quarterly, 6*(4), 50-72. Recuperado de: http://www.

docstoc.com/docs/110266467/ Conductismo-Cognitivismo-y-Constructivismo.

Milton, N. (2007) *Knowledge acquisition in practice: a step by step guide*. London: Springer. Recuperado de: http://www.gestiondelconocimiento.com/modelos_kpmg

Motta, E., & Hara, N. (1996). Solving VT in VITAL. *Human-computer Studies., 34*.

Nalepa, G.J.; Adrian, W.T.; Bobek, S.; Maslanka, P.(2012) Combining AceWiki with a CAPTCHA System for Collaborative Knowledge Acquisition. *Tools with Artificial Intelligence (ICTAI), IEEE 24th International Conference*. 405 – 410.

Newell, A. (1983). The Knowledge Level. *Artificial Intelligence, 18*.

Peng, H. (2014) Counter Cyber Attacks By Semantic Networks, In Emerging Trends in ICT Security. In B. Akhgar, H. Arabnia, M. Schreiber, J. M., Akkermans, A. Anjewierden, R. Hoog, N.R., Shadbolt, W. Van del Velde & B.J. Wielinga. (2000). *Knowledge Engineering and Management. The Common KADS Methodology*. Massachusetts, USA: MIT Press.

Schreiber, A. T., Akkermans, J. M., Anjewierden, A., de Hoog, R., Shadbolt, N. R., Van del Velde, W. & Wielinga, B. J. (2000). *Knowledge Engineering and Management. The CommonKADS Methodology*. Massachusetts , USA: MIT Press.

Shouming H., Yongxian L., Lijuan, H., Wei, Z. & Wei, W. (2010). Research on knowledge-based engineering system for rapid response design of machine tool. *Control and Decision Conference (CCDC), Chinese*. Digital Object Identifier: 10.1109/CCDC.2010.5498376, (4310, 4314)

Syed, V. (2014) *Next Generation Knowledge Machines*. Oxford: Elsevier. ISBN 9780124166295, http://dx.doi.org/10.1016/B978-0-12-416629-5.00010-4.

Teece, D. (July, 2016). Dynamic capabilities and entrepreneurial management in large organizations: Toward a theory of the (entrepreneurial) firm. *European Economic Review,86,* 202-216, doi: org/10.1016/j.euroecorev.2015.11.006

Teng, Z., Chen, J., y Xia, H. (2016). Study on case-based reasoning-inspired approaches to machine-learning. In 2015 *Int. Conf. Intell. Transp. Big Data Smart City, ICITBS 2015*, (760–763).

Tseng, S., y Lee, P. (2014). The effect of knowledge management capability

and dynamic capability on organizational performance. *Journal of Enterprise Information Management, 27*, 158e179. http://dx.doi.org/10.1108/JEIM-05-2012- 0025.

Varela, F. (1989) *Connaître: les Sciences Cognitives. Tendances et Perspectives*. París: Ed. SEUIL (Título original: Cognitive Sciences. A Cartography of Current Ideas. 1988)

Woolfolk, A., Winne, P., Perry, N., y Shapka, J. (2010). *Educational Psychology*. Toronto: Pearson Canada. ISBN 978-0-205-75926-2

Líneas de producción de software para la construcción de un sistema de biblioteca, a través de frameworks basados en componentes

César Felipe Henao Villa[1]; David Alberto García Arango[2];
Jovany Sepúlveda Aguirre[3]; Elkin Darío Aguirre Mesa[4];
Gustavo Andrés Araque González[5]; Christian Hernán Obando Ibarra[6]

Resumen

Algunos estudios han demostrado mejoras en la calidad del producto y reducciones de tiempo de desarrollo cuando la Línea de productos de software (SPL) se introduce en la ingeniería. Sin embargo, existe un desconocimiento sobre la ayuda que muchos framework traen para apoyar su adopción, a través del reusó de componentes a través de framework.

1 Docente-Investigador del grupo AGLAIA - Corporación Universitaria Americana. Ingeniero de Sistemas de la Universidad Nacional de Colombia Sede Medellín, magíster en entornos virtuales de aprendizaje. Correo electrónico: chenao@coruniamericana.edu.co . ORCID: https://orcid.org/0000-0001-7426-2589

2 Docente-Investigador del grupo AGLAIA - Corporación Universitaria Americana. Licenciado en Matemáticas y Física de la Universidad de Antioquia, Magíster en Matemáticas Aplicadas de la Universidad EAFIT. Escuela de Ciencias, departamento de ciencias Matemáticas, doctorando en Educación de la Universidad Nacional de Rosario – Argentina. Correo electrónico de contacto: dagarcia@coruniamericana.edu.co . ORCID: https://orcid.org/0000-0002-0031-4275

3 Magister en Gestión de la Innovación Tecnológica, Cooperación y Desarrollo Regional. Investigador Junior integrante del Grupo de Investigación AGLAIA de la Corporación Universitaria Americana. ORCID: https://orcid.org/0000-0002-1047-6673. E-mail: jasepulveda@americana.edu.co.

4 Docente- Institución Universitaria Pascual Bravo. Ingeniero de sistemas de la Fundación Universitaria María Cano, Magíster en Gestión de la Tecnología Educativa de la Universidad de Santander. Correo electrónico: elkin.aguirre@pascualbravo.edu.co. ORCID: https://orcid.org/0000-0003-2521-6003

5 Docente-Investigador Corporación Universitaria Americana. Ingeniero Industrial, con especialización en Gestión Logística Integral, magíster en Ingeniería de Producción con énfasis en transporte y logística. Correo electrónico: garaque@americana.edu.co. ORCID: https://orcid.org/0000-0001-8627-8924.

6 Ingeniero en electrónica y telecomunicaciones, Especialista en seguridad informática y Magister en tecnologías de la información y comunicación. Director del Programa de Ingeniería de Sistemas de La Corporación Universitaria Americana.

Este documento muestra la implementación de un sistema de biblioteca en una Institución Educativa, la cual no posee un sistema de información que les entregue datos en tiempo real de los libros que pueden prestar a sus estudiantes. El desarrollo de aplicaciones web a través de frawework como Laravel proporciona la posibilidad de crear aplicaciones en corto tiempo y por medio de la utilización de componentes.

Palabras clave: líneas de producto de software, framework, laravel, programación basada en componentes, reutilización.

Lines of Production of Software for the Construction of a Library System, through frameworks based on components.

Abstract

Some studies have shown improvements in product quality and reductions in development time when the Software Product Line (SPL) is introduced into engineering. However, there is a lack of knowledge about the help that many frameworks bring to support their adoption, through the reuse of components through the framework.

This document shows the implementation of a library system in an Educational Institution, which does not have an information system that gives them real-time data of the books they can provide to their students.

Although students can use the library, they generally lack the information necessary to use library resources effectively.

The development of web applications through frawework such as laravel provides the possibility to create applications in a short time and through the use of components.

Key words: Software product lines, framework, Laravel, Component-based Programming, Reuse.

Introducción

Este trabajo tiene por objetivo mostrar las ventajas al escoger la metodología de programación orientada a componentes aplicando la reutilización de código al desarrollar un sistema para la administración de una biblioteca en la Institución Educativa Sol de Oriente ubicada en la ciudad de Medellín. Esta aplicación se desarrolló mediante el framework de laravel y la metodología SCRUM. El software se desarrolla reutilizando código de otras aplicaciones que se están creando desde la unidad de Industria de Software. Estas aplicaciones son *ConjuridicoWeb* y *Piarweb*. La unidad de Industria de software está creando un repositorio propio para el reúso de código fuente que se pueda implementar en aplicaciones que se desarrollen en la facultad de Ingeniería, buscando justamente lograr un enfoque de reutilización sistemático dentro de la unidad de desarrollo, aplicando así el concepto de "*Líneas de Productos de Software*". De acuerdo con Addison-Wesley Longman Publishing Co. (2001) una línea de productos de software se refiere a un conjunto de sistemas de software (producto) que comparten características y que son desarrollados a partir de un conjunto común de bienes núcleo (partes reutilizables).

En la biblioteca de la Institución Educativa Sol de Oriente, la cual es un espacio de aprendizaje y transferencia de conocimiento para los estudiantes, donde además se proporciona un gran apoyo en el proceso de enseñanza y aprendizaje, los docentes y estudiantes no tienen un sistema de información de los libros que existen en la biblioteca. El sistema *BibliotecaWeb* permite organizar los recursos de tal forma que sean fácilmente accesibles y utilizables para los estudiantes, teniendo la posibilidad de buscar la información detallada de los libros, como categorías, temas, autores, etc, dentro de un sistema de información web. Esto motivara a los estudiantes a leer los libros que se encuentran en la Institución Educativa. La aplicación se realiza mediante el framework de laravel y bajo la línea de software de Aplicaciones Web de la facultad de Ingeniería, del programa de Ingeniera de Sistemas.

Antecedentes

La aplicación de componentes de software es una extensión de la programación orientada a objetos, esta última fue pensada para desarrollar sistemas cerrados, sin embargo, la programación orientada a objetos es insuficiente para abordar desarrollos de sistemas abiertos. Los sistemas abiertos presen-

tan una problemática mucho más compleja de abordar que la de los sistemas transaccionales, ya que la evolución de la aplicación no se puede definir en un estado global externamente entendible. Lo cual soluciona la programación orientada a componentes debido a que tiene el poder de desplegar independientemente y ser dinámicamente extensible.

Arquitectura de Laravel

La arquitectura del framework de Laravel está compuesta por pequeños componentes, en nuestro caso, orientados a crear sistemas de información web. Como resultado, aceleramos nuestras aplicaciones en más del 30% y también conseguimos una mayor capacidad de mantenimiento. Estos principios pueden, ser reutilizados y aplicarse fácilmente a otros proyectos.

El software evoluciona con el tiempo y los framewok actualmente se utilizan para crear aplicaciones de última tecnología y en muy poco tiempo. Gracias a los frameworks no es necesario comenzar desde cero, se puede reutilizar componentes previamente desarrollados y es así como luego se logra obtener una baja deuda técnica a nivel de mantenibilidad, por el alto componente en el backend. Una estructura modular con un enfoque de bajo nivel de principios sólidos.

Elementos de la arquitectura básica de una app en Laravel

Mientras Routes and Controllers permanecen en la capa HTTP, los componentes incluyen toda la lógica de negocios que está separada en comandos, eventos resultantes y sus manejadores, así como entidades / modelos / repositorios. Un beneficio adicional en este framework es que en las carpetas de componentes también pueden tener subcomponentes. Los componentes también pueden comunicarse entre sí.

La carpeta de aplicaciones contiene múltiples instancias de Laravel. Todo lo relacionado con las implementaciones: los archivos, configuraciones y migraciones. Los archivos de entorno y las secuencias de comandos se basan en el directorio raíz del back-end. La carpeta compartida contiene nuestros componentes compartidos y proveedores de servicios.

Al igual que todos los frameworks PHP populares como Symfony, Yii, Codeigniter y otros, Laravel es un framework MVC (Model-View-Controller). Es un modelo de arquitectura de aplicaciones que separa una aplicación en tres componentes lógicos

- **Modelo:** Este componente maneja toda la lógica de la aplicación.
- **Vista:** Este componente maneja toda la interfaz de usuario (interfaz web de usuario) y los elementos de presentación de la aplicación.
- **Controlador:** Este componente actúa como interfaz entre el Modelo y la Vista. Controla las interacciones entre el Modelo y la Vista.

Estructura de directorios básicos

Aquí hay una imagen con estructura de directorios de Laravel 5.5, Figura 1:

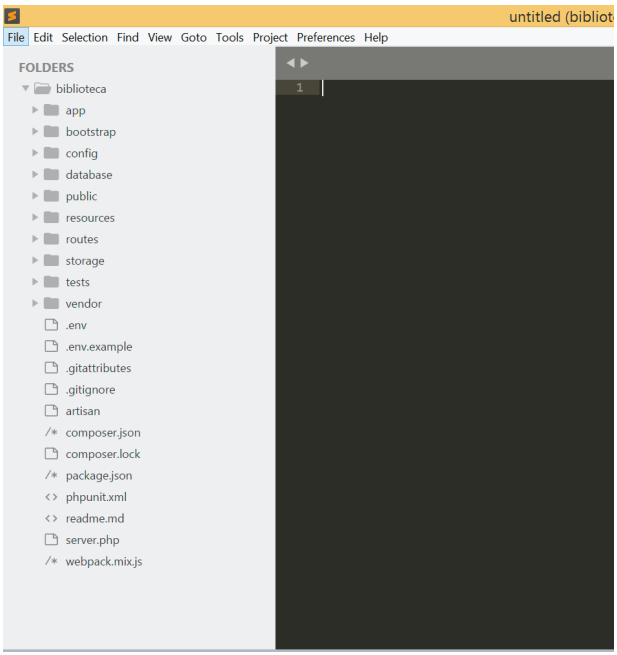


Figura 1. Estructura de directorios de Laravel

La carpeta de la aplicación en la Laravel contiene los Modelos y Controladores de la aplicación.

Los modelos se crean en la raíz de la carpeta de la aplicación, mientras que los Controladores y Middlewares se crean en sus respectivas carpetas dentro de la carpeta Http.

Las vistas en Laravel (las plantillas que se representan como HTML) se crean en la carpeta de vistas dentro de la carpeta de recursos.

El directorio de recursos contiene activos en bruto como los archivos LESS y Sass, archivos de localización y lenguaje, y plantillas que se representan como HTML.

El directorio de almacenamiento contiene almacenamiento de aplicaciones, como cargas de archivos, entre otros. Almacenamiento de marcos (caché) y registros generados por aplicaciones.

El directorio del proveedor contiene dependencias de composer.

El enrutamiento para los controladores se maneja mediante el archivo Web.php ubicado dentro de la carpeta de rutas. El directorio del proveedor contiene dependencias del manejador de paquetes comper.

En el archivo. env puede agregar sus datos para conectarse a una base de datos mysql (DB_DATABASE, DB_USERNAME, DB_PASSWORD).

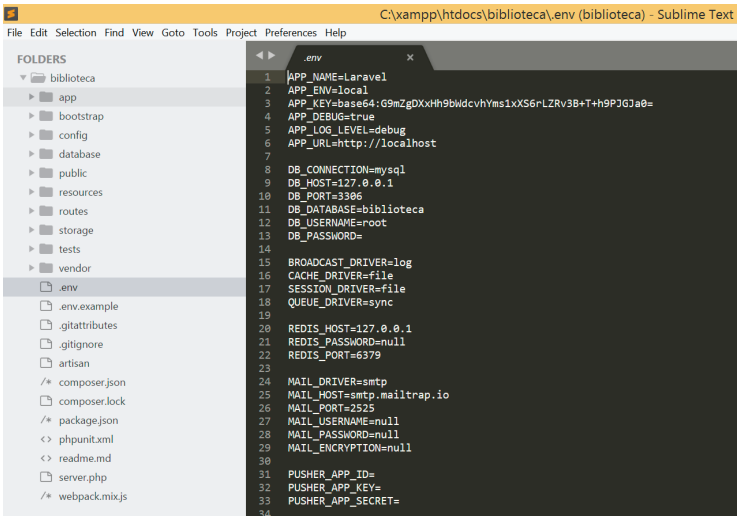


Figura 2. Estructura de archivo env. En Laravel

También puede configurar la configuración regional, la zona horaria, etc. de la aplicación en el archivo config / app.php.

La carpeta pública es la raíz del documento de la aplicación. Comienza la aplicación Laravel. También contiene los activos de la aplicación como JavaScript, CSS, Imágenes, etc. Para ejecutar su aplicación, debe ir a la carpeta pública de su aplicación Laravel.

Sistema Bilioteca*Web*

La aplicación ***BibliotecaWeb*** está creada en el lenguaje php, bajo el framework laravel utilizando componentes. A continuación, se presenta la Arquitectura general como está estructurado Laravel.

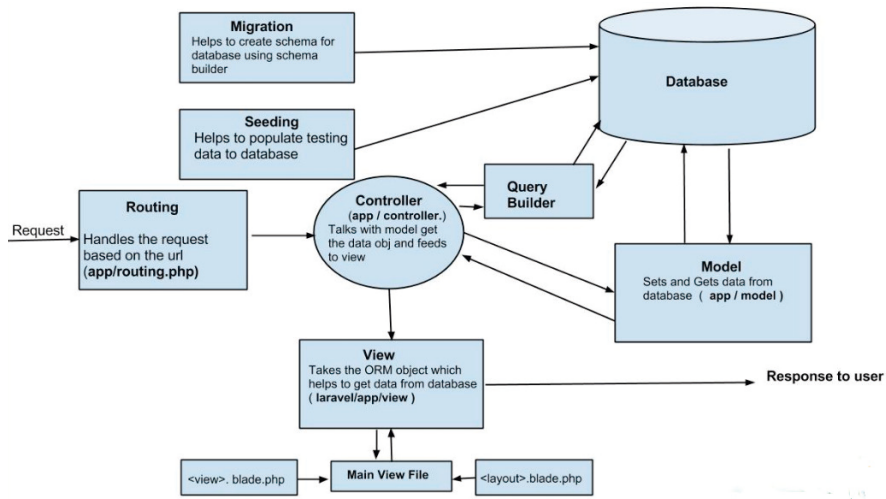


Figura 3. Arquitectura General de Laravel.

Los componentes pueden estar en cualquiera de las partes de la estructura.

Existen muchas otras opciones como Yii, CodeIgniter, Phalcon, etc., sin embargo, Laravel se destaca por su sintaxis simple, elegante y expresiva. Esto se hace simplificando las tareas repetitivas utilizadas en el desarrollo de la

mayoría de las aplicaciones web y permite utilizar mediante componentes el reúso de código fuente. La documentación para Laravel ha sido escrita por su fundador y desarrollador principal, Taylor Otwell. Este framework permite crear diseños profesionales con contenido dinámico, debido a sus plantillas y widgets livianos que incluyen código JS y CSS con estructuras sólidas.

Laravel proporciona una herramienta integrada para la línea de comando que crea y maneja el entorno del proyecto Laravel, esta herramienta se llama Artisan. Se puede usar para crear un código esqueleto, su estructura de base de datos y construir la migración que luego hace que sea muy fácil administrar la base de datos.

Los componentes se pueden utilizar dentro de: controladores, vistas, modelos, sistemas request y response, etc. A continuación, se muestra el ciclo de vida de un desarrollo de un sistema basado en componentes (Figura 4).

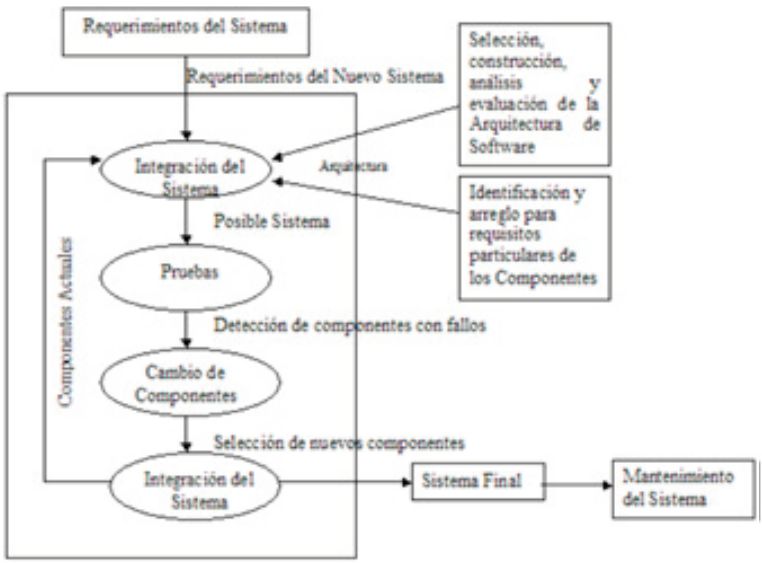


Figura 4. ciclo de vida de un desarrollo de un sistema basado en componentes

La aplicación ***BibliotecaWeb***, fue generada por el paquete que trae por defecto Laravel, que nos permite crear las vistas, el controlador y las rutas correspondientes.

Además, tiene 5 módulos, los cuales son:

- Autores
- Categorías
- Editoriales
- Libros
- Prestamos

Todos los módulos tienen la opción de añadir la información necesaria, solicitada por cada módulo.

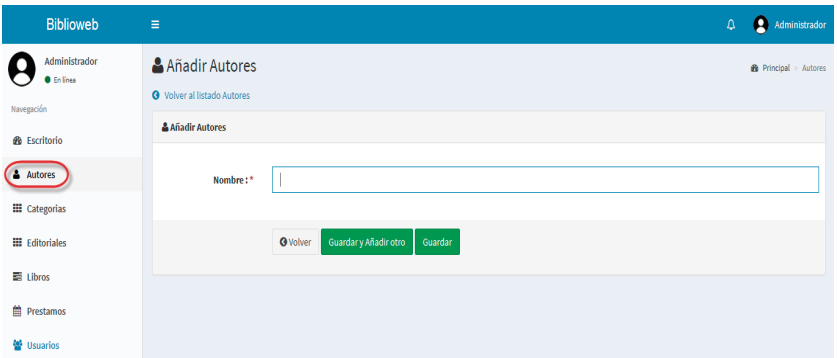


Figura 5. Módulo Autores.

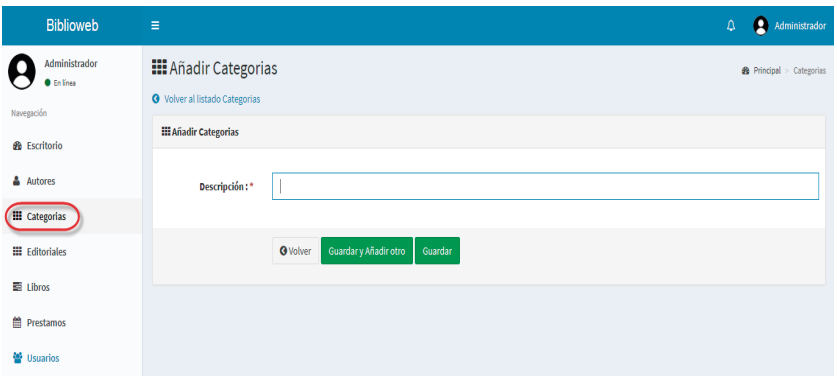


Figura 6. Módulo Categorías.

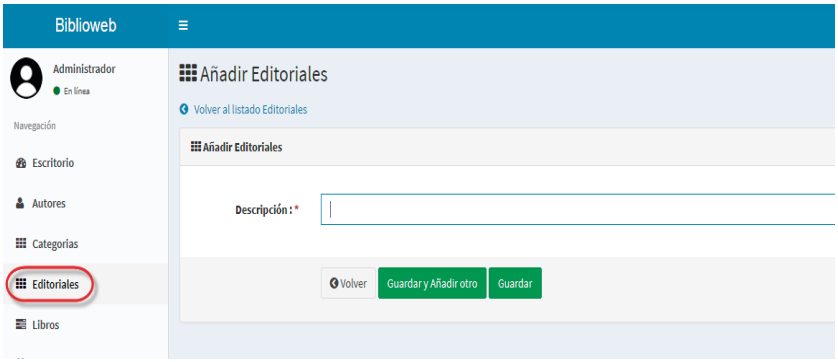


Figura 7. Módulo Editoriales.

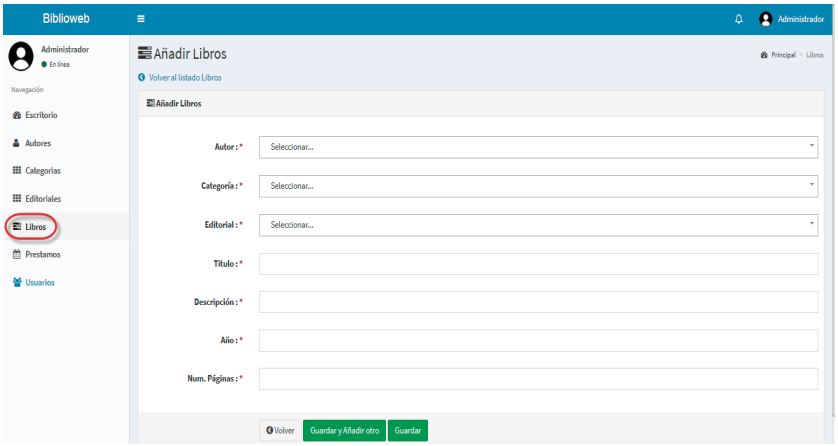


Figura 8. Módulo Libros.

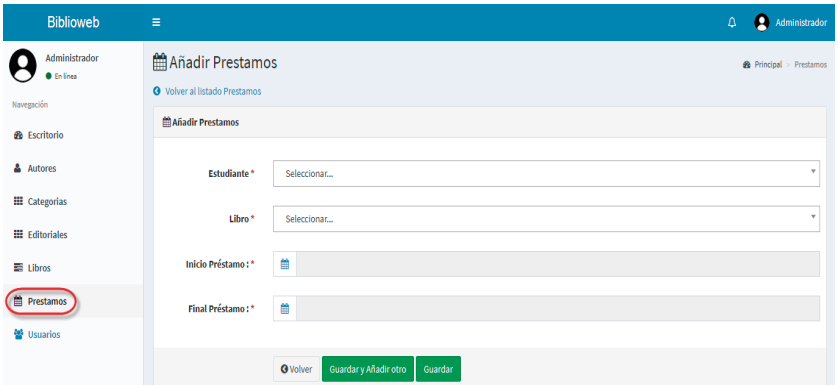


Figura 9. Módulo Préstamos

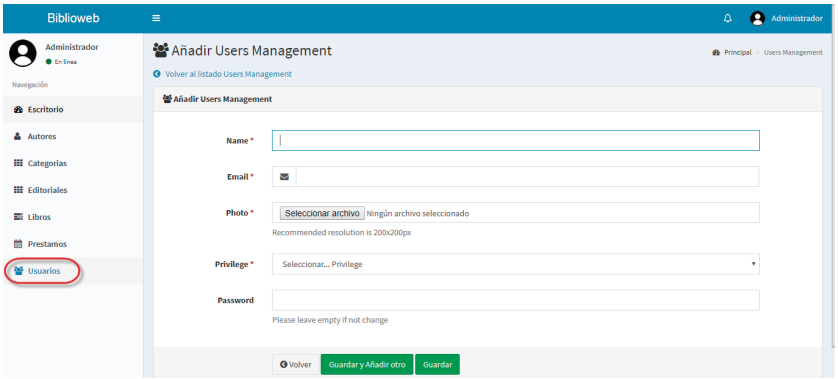


Figura 10. Módulo Usuarios.

La aplicación ***BibliotecaWeb***, permite obtener Reportes para préstamos y libros en formato PDF y Excel.

Figura 11. Módulo Reportes

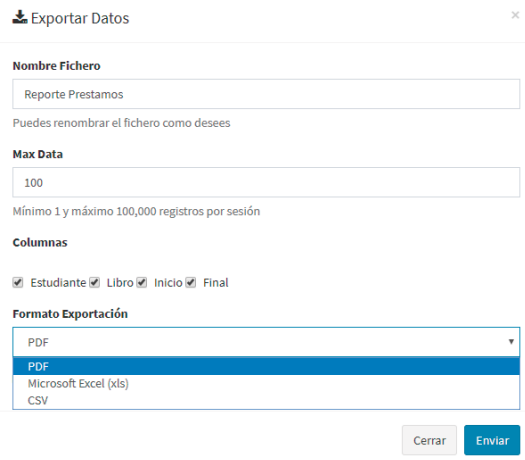


Figura 12. Módulo exportar Reportes

Manejo de componentes y reúso de código mediante Laravel:

Actualmente muchas empresas de desarrollo de software están adoptando enfoques dirigidos a la reutilización proactiva de los componentes de software y desarrollando ciclos de planificación para construir productos de alta calidad de forma más rápida y reduciendo los costos de desarrollo de software, este enfoque se está integrando a la metodología de desarrollo de la unidad de industria de software.

La investigación se llevó en la Institución Educativa Sol de Oriente, a través de la Unidad de Industria de Software de la Facultad de Ingeniería de la Corporación Universitaria Americana, se realiza mediante una serie de metodologías para la planeación, construcción y retroalimentación de los proyectos de software llegando a obtener resultados satisfactorios dentro del desarrollo basado en componentes.

Se logra determinar, que en muchas ocasiones la existencia de componentes reutilizables no garantiza su reutilización, sino se tiene claridad en su arquitectura.

Los componentes reutilizables ayudan a los desarrolladores de software a pensar en niveles más altos de abstracción, ayudando a aumentar el conocimiento sobre el software y la comunicación al interior del mismo. Los componentes reutilizables expanden la expresividad de desarrolladores de software y contribuyen a la reducción de la complejidad de desarrollo de software. Sin embargo, los desarrolladores de software deben aprender la sintaxis y semántica de los componentes si escogen esta opción.

En el desarrollo de software basado en componentes, decimos que un sistema está integrado por un conjunto de mecanismos, que permiten la interconexión de componentes de software, junto a una colección de servicios cuya utilidad principal es permitir de forma fácil las tareas de los componentes que se encuentran al interior del sistema. La definición de componente se formuló en la edición de ECCOP. (Workshop on componentoriented programming, 1996) por primera vez: "Un componente software es una unidad de composición de aplicaciones software con interfaces especificadas mediante contrato y un conjunto de requisitos que solo posee dependencias de contexto explicitas. "Un componente de software puede ser desarrollado, distribuido inde-

pendientemente de otros, incorporado al sistema de forma independiente y es propenso a la composición con otros componentes desarrollados por terceras partes, en tiempo y espacio" (Szyperski,1998).

A continuación, se muestra la tabla 1, donde se observa el tiempo en desarrollo de la Aplicación ***BibliotecaWeb*** vs las demás aplicaciones desarrolladas en la unidad de Industria de Software. Aplicando la reutilización de componentes.

Tabla 1. Tiempo en meses de desarrollo de las Aplicaciones en Industria de Software

Nombre de la Aplicación	*Tiempo en Desarrollo de la Aplicación.*
Consultorio Jurídico	12 meses
PiarWeb	6 meses
BibliotecaWeb	5 meses

Además, también se muestra en la figura 13, el porcentaje de código fuente común entre las aplicaciones.

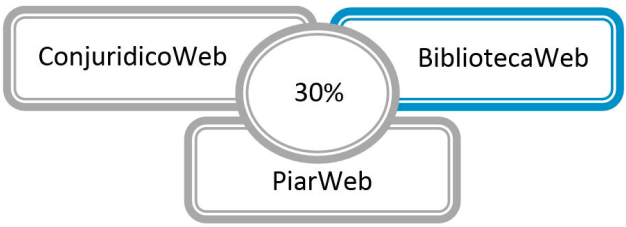


Figura 13. Código Fuente Común

Existe código reutilizado en un porcentaje correspondiente al 30% en cada una de las aplicaciones, siendo ***ConjuridicoWeb*** el primero que se desarrolló, seguido de ***PiarWeb*** y ***Bilioteca Web***. Todos fueron construidos en el framework laravel.

Metodología

La investigación previa al desarrollo del software ***BibliotecaWeb*** fue desarrollada mediante una recopilación de datos relacionados con la aplicación de la programación orientada a componentes mediante framework como laravel.

Particularmente investigamos sobre metodologías de reúso y su relación con los componentes dentro del framework, permitiendo lograr un conocimiento especializado en el reúso de código para el desarrollo ágil de software.

Conclusiones

Se construye la aplicación ***BibliotecaWeb***, que soluciona las necesidades que se tenía al interior de la Institución.

Se observa que la mayor parte de las aplicaciones que se construyen en la actualidad son de una gran complejidad, por lo que requieren de metodologías que faciliten el desarrollo, utilizando componentes ya sea como paquetes de software, servicios web, o módulos que encapsulen conjuntos de datos.

Laravel permite mediante la adopción de componentes maduros preexistentes, realizar aplicaciones de forma rápida.

El framework de laravel permite la reusabilidad, donde un mismo componente puede ser usado en varias aplicaciones, facilitando la actualización y la mejora del sistema, realizando la sustitución de componentes.

Referencias

Veryard, R. (2001). *Component-based business : plug and play*. London : Springer.

Szyperski, C. (2002). *Component Software: Beyond Object-Oriented Programming*. Addison-Wesley Professional, Boston

ECOOP. (1996). *Workshop on componentoriented programming*. Recuperado de https://link.springer.com/chapter/10.1007/3-540-47853-1_8

Metodologías de detección de intrusos (IDS) basadas en anomalías de red aplicando redes neuronales SOM y GHSOM

Eduardo de la Hoz Correa[1]; Johan Mardini Bovea[2]; Emiro de la Hoz Franco[3]

Resumen

La acelerada aparición de nuevas tecnologías de comunicaciones, así como de múltiples protocolos de transferencia de datos, ha provocado un desbordamiento en el crecimiento de la internet, y por ende en la interconectividad de computadores. Esta interconectividad, sumado al aumento constante de personas maliciosas, y la evolución de las técnicas de ataques a infraestructuras informáticas, dificulta considerablemente el proceso de distinción entre lo que puede ser o no, trafico normal en una red de datos. Lo anterior, ha conllevado al estudio de nuevas prácticas que propendan una detección eficaz y eficiente de dichos flujos de datos potencialmente peligrosos. En la actualidad se han propuesto métodos que van desde la aplicación de la Inteligencia Artificial en distintos ámbitos, entre las que encontramos la implementación de técnicas estadísticas. En este trabajo, se presenta una comparativa de los métodos de selección de características *CHI-SQUARE e INFO.GAIN* al momento de hacer procesos clasificatorios de tráfico en redes de datos usando las redes neuronales SOM y GHSOM con el fin de distinguir entre conexiones normales y anormales.

Palabras clave: *CHI-SQUARE*, Info Gain, Redes Neuronales, SOM, GHSOM

1 PhD. en Tecnologías de la Información y la Comunicación. Docente Investigador. Corporación Universitaria Americana, Medellín, Colombia. Grupo de Investigación AGLAIA. ORCID: https://orcid.org/0000-0001-7468-6058. Mail: emdelahoz@americana.edu.co

2 Magister en Ingeniería de sistemas y Computación. Universidad del Atlántico, Barranquilla, Colombia. Grupo de investigación 3I+D. ORCID: https://orcid.org/0000-0001-6609-1687. Mail: johanmardini@mail.uniatlantico.edu.co

3 PhD. en Tecnologías de la Información y la Comunicación. Jefe departamento de Ingenierías. Universidad de la Costa, Barranquilla, Colombia. Grupo de Ingeniería de Software y Redes. ORCID: https://orcid.org/0000-0002-4926-7414. Mail: edelahoz@cuc.edu.co.

Intruders Detection Methodologies (IDS) based on network anomalies applying SOM and GHSOM neural networks

Abstract

The accelerated appearance of new communications technologies, as well as multiple data transfer protocols, has caused an overflow in the growth of the Internet, and therefore in the interconnectivity of computers. This interconnectivity, added to the constant increase of malicious people, and the evolution of the techniques of attacks to computer infrastructures, makes the process of distinguishing between what may or may not be normal traffic in a data network considerably difficult. The foregoing has led to the study of new practices that favor effective and efficient detection of such potentially dangerous data flows. At present, methods have been proposed that range from the application of Artificial Intelligence in different areas, among which we find the implementation of statistical techniques. In this paper, we present a comparison of the *CHI-SQUARE* and *INFO.GAIN* feature selection methods when making traffic classification processes in data networks using the SOM and GHSOM neural networks in order to distinguish between normal and abnormal connections.

Key words: *CHI-SQUARE*, Info Gain, Neural Networks, SOM, GHSOM.

Introducción

Con el pasar de los años hemos visto cómo las redes de comunicaciones han pasado a ser parte de nuestra vida cotidiana. Estas redes han dejado de ser un medio de comunicación para pequeños grupos de personas, a ser utilizadas por la mayoría de los ciudadanos y empresas. Solo en el ámbito de la seguridad informática, los intentos de intrusión en redes de computadores han crecido en los últimos años. A diario aparecen programas maliciosos que buscan afectar equipos, ya sea para realizar daños locales o para perjudicar toda una red informática. Es por ello, que es necesario entender los ataques informáticos y encontrar formas no solo eficaces sino eficientes de contrarres-

tarlos, ya sea previniéndolos o detectándolos a tiempo para que su impacto sea menor al esperado por el atacante.

Para contrarrestar este tipo de problemas hoy se dispone de varias herramientas (Vesanto, Himberg, Alhoniemi, & Parhankangas, 2000) que les dan a las redes un apoyo para mantener la seguridad. Entre estas herramientas encontramos los llamados Sistemas de Detección de Intrusos o IDS. Los IDS supervisan y registran los eventos que ocurren en una computadora o en una red de computadoras y buscan patrones que permitan identificar intrusiones para responder de la forma más efectiva posible (Dain & Cunningham, 2001), además de evitar malas prácticas, como en el caso de los usuarios autorizados que intentan sobrepasar sus límites de restricción de acceso a la información (Girardin, 1999), con el ánimo de poder dar con los responsables del ataque y tomar acciones conducentes a mejorar la vulnerabilidad y castigar, si se puede, a los responsables del ataque. Es por ello que los IDS han ganado terreno en la mayoría de organizaciones que buscan darle un poco más seguridad en sus sistemas informáticos.

Podemos encontrar dos tipos de IDS, los primeros, los *IDS de uso indebido*, los cuales requieren de una base de datos como sistema de apoyo, por lo que necesitan un mantenimiento regular, como son las actualizaciones periódicas. Ejemplos de ellos son los antivirus. Los segundos, son los llamados *IDS basados en anomalías*, que no cuentan con ningún soporte, ya que aprenden mediante técnicas estadísticas y/o de inteligencia artificial. Este tipo de algoritmos aprenden a diferenciar entre un comportamiento normal y un comportamiento anormal de los usuarios. Sin embargo, ningún IDS puede detener totalmente la gran cantidad de anomalías que existen hoy día. Por consiguiente, el fin de los IDS actual es tratar prevenir en gran medida el efecto de cualquier ataque.

Consecuentemente con lo anterior, en los últimos años la aplicación de estrategias de Inteligencia Computacional (Computacional Intelligence o Soft Computing) se ha convertido en uno de los campos de investigación más prominente dentro de los IDS como lo expresa Xiaonan & Banhzaf (2010) y Selvakani & Regan (2010). Sadkhan (Sadkhan, 2009) muestra que este nuevo paradigma se ha venido desarrollando para solucionar problemas que no han podido ser resueltos a través de los métodos tradicionales. Como resultado

de estas nuevas técnicas o estrategias en los IDS se han propuesto métodos de detección de anomalías basadas en redes neuronales (Xiaonan & Banhzaf, 2010), conjuntos de lógica difusa, computación evolutiva, sistemas inmunológicos artificiales y sistemas de colonias de hormigas, por señalar algunos. Muchos de estos enfoques son capaces de adquirir e integrar autónomamente conocimiento y pueden ser implementados a través de un modo de aprendizaje supervisado, o no supervisado.

La detección de anomalías no es tarea sencilla en un entorno real, algunos autores como Ghorbani, Lu, & Tavallae (2009), Levin (2000), Pfahringer (2000) y Tavallaee, Stakhanova, & Ghorbani (2010) plantean problemas interesantes relacionados con la clasificación y la selección de características. Actualmente, conjuntos de datos como KDD99 (y también el NSL-KDD) han sido construidos para proporcionar entrenamiento y subconjuntos de prueba con diferentes distribuciones estadísticas como se espera en las tareas de detección de anomalías reales (Zargari & Voorhis, 2012). En Nziga (2011), se proporciona un análisis a bajo nivel de la adopción que ha tenido la industria en los procedimientos de detección de intrusos que se proponen en la literatura académica, enfocándose en el informe de alto rendimiento que estas puedan tener en la actualidad.

En este trabajo, se usaron las técnicas *CHI-SQUARE* e *INFO.GAIN* las cuales son descritas en la sección 2, y se muestra su incidencia en el conjunto de datos implementado (NSL-KDD), esto con el fin de identificar las características que tienen mayor relevancia en el proceso de clasificación. Estas características se seleccionan según la calidad de los atributos y de acuerdo con lo apropiado de sus valores y la ocurrencia intrínseca de cada método. Posteriormente, el proceso clasificatorio es llevado a cabo usando redes neuronales SOM y GHSOM, con la que se logra, dado un grupo de características, representar distribuciones conjuntas de modo que permitan calcular la probabilidad a posteriori de un cumulo de clases, y así clasificar los objetos en la clase más probable como se evidencia en Sebe y otros (2004) y Gu & Ji (2004). No obstante, detectar no sólo un ataque sino también el tipo al que este pertenece, no es una tarea facil, y es precisamente en este punto donde se ahonda esfuerzos por mejorar la calidad de las detecciones.

Materiales y Métodos

Existen diferentes técnicas para prevenir y corregir acciones intrusivas maliciosas, entre ellas: Listas de Control de Acceso (ACLs), encriptamiento de mensajes, bloqueo de puertos, Redes Privadas Virtuales (VPNs) y Cortafuegos (firewalls). Estos últimos restringen el tráfico de servicios desconocidos, mediante el bloqueo de puertos. Si bien, son útiles para contrarrestar una gran variedad de ataques, queda aún un hueco de seguridad desde el exterior, cuando se encapsulan los ataques en el tráfico de servicios permitidos por el dispositivo, además, tanto los firewalls como las otras técnicas mencionadas, no controlan los ataques que se generan desde el interior de la red.

Para solventar estos inconvenientes se han desarrollado Sistemas de Detección de Intrusos (IDS) que identifican tráfico malicioso en la red para un posterior proceso de bloqueo y documentación que contrarreste las acciones del atacante. Los IDS pueden detectar ataques con una metodología basada en firmas (comparando los ataques con una base de datos de firmas o rules) o con una metodología basada en anomalías (empleando un algoritmo de aprendizaje), los primeros se han implementado ampliamente en IDS comerciales y en software libre (Snort y Prelude), sin embargo, no detectan ataques nuevos; los segundos detectan ataques nuevos con cierto porcentaje de exactitud.

Producto de las experimentaciones encontradas en el estado del arte, se ha detectado que una variable que incide directamente en la eficiencia del algoritmo de aprendizaje al crear un IDS, es la identificación de las características que se van a evaluar durante la fase de pre-procesamiento, debido a que la escogencia de la totalidad de características o algunas de ellas que no sean las apropiadas, generará largos tiempos de respuesta computacional, incidiendo negativamente en la evaluación final del algoritmo de aprendizaje.

En este trabajo se propone un modelo que entrena una red neuronal que, de forma automática, efectúa el proceso de clasificación de flujos de datos identificando el tipo de tráfico, independientemente de aquellos nuevos ataques que se puedan generar.

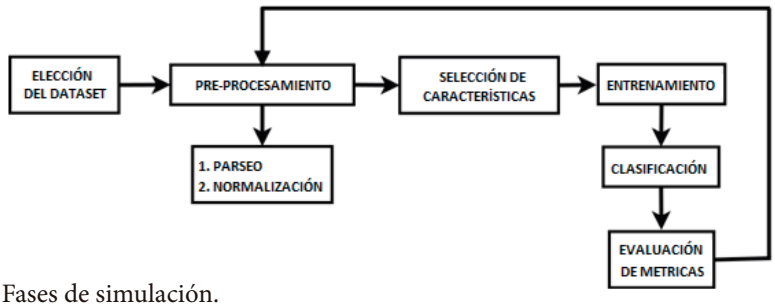


Figura 1. Fases de simulación.
Fuente: Construcción propia.

Para la validación del modelo se implementaron escenarios de simulación, los cuales comprenden tres fases (entrenamiento, clasificación y cálculo de métricas). Se realizó variación de la cantidad de características usadas para probar el modelo en cada simulación, con el fin de analizar la eficiencia computacional. En la fase de entrenamiento se cargó el dataset DARPA KDD-NSL. Después de ello, se efectuó la reducción de características aplicando técnicas de selección, con el propósito de categorizarlas por orden de relevancia, para efectuar posteriormente el entrenamiento de la red neuronal con las características seleccionadas.

En la fase de clasificación se carga otro dataset (DARPA NSL-KDD TEST), diferente al utilizado en el proceso de entrenamiento, a dicho dataset se le realiza una reducción de las características, usando técnicas de selección (teniendo en cuenta las mismas características seleccionadas en la fase de entrenamiento), por último, se clasifican los datos, basándose en el mapa generado en el proceso de entrenamiento y en el nuevo subconjunto de datos.

En la fase final se calcularon diferentes métricas de desempeño (sensibilidad, especificidad, precisión y exactitud), esto permitió determinar la eficiencia del modelo planteado.

Conjunto de Datos (DataSet)

El conjunto de datos utilizados en las experiencias realizadas para la consecución de los resultados es el NSL-KDD (DARPA, 2018). El NSL-KDD surge

como una mejora para evitar los problemas del conjunto de datos KDD´99. Una de las mejoras sustanciales que tiene este nuevo conjunto de datos es la eliminación del gran número de registros redundantes. El incremento era de alrededor del 78% y 75% para los subconjuntos de entrenamiento y prueba respectivamente. Al no incluir registros redundantes en la colección de datos para el entrenamiento, los clasificadores no se sesgarán hacia los registros más frecuentes ya que los registros redundantes en el conjunto de entrenamiento causan un aprendizaje inapropiado. Otras de las mejoras que se pueden mencionar del NSL-KDD respecto a sus predecesores es que no existen registros duplicados en las colecciones de datos propuestas para el conjunto "*test*", haciendo que el rendimiento del aprendizaje no este sesgado por los métodos que tienen mejores tasas de detección de registros frecuentes. Además, el número de registros seleccionados de cada grupo de nivel de dificultad es inversamente proporcional al porcentaje de registros en el conjunto original de datos KDD, dando como resultado que las tasas de clasificación de los distintos métodos de aprendizaje varíen en un rango más amplio, lo que hace que sea más eficiente para tener una evaluación precisa de las diferentes técnicas de aprendizaje. Por último, cabe mencionar que el número de registros en las colecciones de datos para el entrenamiento y de test es razonable haciendo posible realizar experimentos con el conjunto de datos completo sin necesidad de seleccionar al azar una pequeña porción de este.

El dataset NSL-KDD contiene tipos de ataques los cuales se clasifican en cuatro categorías:

- ***Usuario a Root*** (User to Root attack, U2R). Consiste en hacer una autenticación en el sistema como usuario normal, y luego tratar de cambiar a una credencial de superusuario mediante la explotación de vulnerabilidades presentes en el sistema. Después de hacer la explotación de la o las vulnerabilidades, el atacante obtiene acceso de administrador en el sistema. Las clasificaciones de los ataques de U2R pueden apreciarse en la Tabla 1.
- ***Denegación de servicio*** (Denial of Service Attack, DoS). En este caso, el atacante intenta sobrecargar la máquina de la víctima para que esté demasiado ocupada para responder a nuevas peticiones legitimas. Así, la maquina atacada negará estas solicitudes. Este ataque se puede hacer por agotamiento de la memoria, la interfaz de red u otros recursos computacionales del servidor destino. Las clasificaciones de los ataques

de DoS pueden apreciarse en la Tabla 2.

- ***Ataque de Sondeo*** (Probbing Attack, PROBE). Este ataque se basa en tratar de recuperar información acerca de los ordenadores conectados a una red. El escaneo de puertos con el fin de tratar de descubrir puertos abiertos en los equipos que pertenecen a una red, es un ejemplo claro de este ataque. Las clasificaciones de los ataques de PROBE pueden apreciarse en la Tabla 3.
- ***Remoto a Local*** (Remote to Local Attack, R2L): En este caso, el atacante utiliza vulnerabilidades del sistema al que no tiene acceso para iniciar una sesión en como un usuario normal. Este ataque puede ser un paso previo del ataque U2R. Las clasificaciones de los ataques de R2L se dan en la Tabla 4.

Tabla 1. *Ataques involucrados en U2R (DARPA NSL-KDD)*

Tipo de ataque	Descripción del ataque
loadmodule	Consta de un ataque oculto que reinicia la IFS para los usuarios comunes y genera una Shell de root.
rootkit	Constituido por un escenario de varios días donde el usuario común instala algún componente de un rootkit.
buffer overflow	Implementación de desbordamiento en la pila del buffer.
perl	Establece la identificación de usuario común como root implementando un script de perl y crea un Shell en modo root.

Tabla 2. *Ataques más importantes involucrados en DoS (DARPA NSL-KDD)*

Tipo de ataque	Descripción del ataque
smurf	El atacante envía un ping que suele proceder de la víctima, y en el broadcast a una gran parte de la red en donde todos los host responderán al host objetivo.
neptune	Consta de una Inundación por envíos de paquetes TCP/SYN en uno o múltiples puertos de red.
pod	Es un ataque conocido como el Ping de la Muerte, el cual envía muchos paquetes ICMP muy cargados dejando fuera de servicio a un dispositivo de red.

land	Constituido por él envió de paquetes TCP/SYN falso conteniendo la dirección de la víctima como origen y destino causando así que se auto responda continuamente.
back	Es un tipo de ataque informático hacia un web services basado en Apache el cual consiste en un cliente pide una URL que contiene muchas barras.
teardrop	Consta de la aplicación de un algoritmo capaz de fragmentar paquetes IP para enviar paquetes corruptos a la víctima.

Tabla 3. *Ataques involucrados en Probing (DARPA NSL-KDD)*

Tipo de ataque	Descripción del ataque
portsweep	Análisis tipo barrido de puertos en un host específico para determinar los servicios que se apoyan en el mismo.
Nmap	Análisis de redes aplicando un escaneo utilizando la herramienta NMAP.
Satan	Es una herramienta que se caracteriza por hacer sondeo en redes con el propósito de encontrar debilidades desconocidas.
Ipsweep	Es una herramienta que se caracteriza por hacer sondeo en redes con el propósito de encontrar debilidades desconocidas.

Tabla 4. *Ataques involucrados en R2L (DARPA NSL-KDD)*

Tipo de ataque	Descripción del ataque
phf	Implementación de un Script CGI que permite la ejecución de comandos en un servidor web mal configurado.
guess passwd	Intenta adivinar la contraseña utilizando TELNET para las cuentas de visitantes.
multihop	Escenario de intrusión que toma múltiples días en donde el atacante inicialmente accede a una máquina para luego usar como punto inicial para atacar a otras máquinas.
ftp write	Constituido por la creación de un archivo .rhost creado por un usuario FTP remoto y obtiene un logeo local.
warezclient	Aplica cuando los usuarios descargan archivos de forma ilegal atreves de FTP anónimo usando Waresmaster.
warezmaster	Subida de archivos usando FTP anónimo de Warez (software de copias ilegales).

spy	Analizador de los protocolos de red LAN por la interfaz de red.
imap	Impacto por desbordamiento remoto de bufer implementado el puerto correspondiente a IMAP

El conjunto de datos KDD en su nueva versión contiene características que se dividen en tres clases: *características básicas*, *características de contenido*, y *características de tráfico*. La principal razón para tener estos tres grupos de características es que la detección y la identificación de algunos ataques requiere el uso de más de una clase de características. Por ejemplo, las funciones basadas en el tiempo son necesarias para detectar los ataques que requieren algunas estadísticas calculadas durante un cierto periodo de tiempo. Por lo tanto, el primer paso que se realiza para poder trabajar con el conjunto de datos consiste en analizarlo detalladamente con el fin de extraer las características de los archivos de texto y construir los vectores que comprenden el espacio de características. En las *características básicas* encontramos todas las categorías que se pueden extraer de una conexión TCP / IP. Para el caso de las *características de contenido*, encontramos características calculadas con relación a un intervalo de la ventana y se divide en dos grupos. El primero de ellos es el grupo de los atributos de *mismo host*, que tienen en cuenta solo las conexiones en los dos últimos segundos que tengan el mismo destino que la actual, y las estadísticas relacionadas con el protocolo, los servicios, etc. El segundo grupo es el de los atributos de *mismo servicio*, que examinan sólo las conexiones en los dos últimos segundos que tienen el mismo servicio que la conexión actual. Este tipo de tráfico mencionado en las dos características anteriores se le conoce como características de tráfico basadas en tiempo. En la actualidad existen otro tipo de ataques llamados de *sondeo* que escanean los puertos de los hosts en un intervalo de tiempo mucho mayor que 2 segundos. Este tiempo mayor podría ser de 1 minuto, lo que da como resultado que no hay patrones de intrusión en una ventana de tiempo de 2 segundos. Para resolver este inconveniente, las características del *mismo host* y del *mismo servicio* se recalculan en una ventana de observación de 100 conexiones dando como resultado las características de tráfico. Por último, encontramos a las *características de contenido*, en este tipo de características encontramos a los ataques R2L y U2R. Estos ataques son sustancialmente diferentes que los ataques DoS y de sondeo ya que no cuentan con una frecuencia notoria de patrones secuenciales. Esto se debe a la cantidad de conexiones que hacen los ataques DoS y Probing a algún host

en un periodo muy corto de tiempo. Sin embargo, los ataques R2L y U2R se encuentran incrustados en las porciones de datos de los paquetes, implicando en la mayoría de las veces una sola conexión. Para poder hacer una detección exitosa de este tipo de ataques se recurre de algunas características que hacen posible detectar algún comportamiento sospechoso en la parte de los datos. Un ejemplo claro es el número de intentos de acceso fallidos. A partir de aquí se generan las llamadas *características de contenido*.

Procesamiento de datos

El pre-procesamiento es una fase previa a la selección o extracción de características en el proceso de simulación realizado. Esta fase permite homogenizar la presentación de los datos provenientes del dataset e integrar esos datos contenidos en un formato diferente a la herramienta de simulación que se va a utilizar para efectos de procesamiento. El pre-procesamiento implica la ejecución de dos (2) sub fases llamadas *Parseo* o *Parsin*g y la *Normalización*. El *parsing* se refiere, básicamente a presentar los datos en un formato que sea fácil de procesar por la herramienta en la cual se implementa la simulación, que en este caso es Matlab™. Ello requerirá recorrer iterativamente cada registro de la colección de datos del archivo *.txt*, que tiene una estructura del tipo CVS (con separación de los datos por comas), extrayendo registro a registro de datos no estructurados a un formato de datos estructurados, de tal forma que estos se representen como una matriz de datos donde cada fila corresponde a un patrón o una conexión de red y cada columna corresponderá a las características o atributos que permiten identificar la conexión de datos. Una vez ejecutado el parseo, se procede a *normalizar* los datos, el proceso de normalización implica presentarlos en el mismo formato, dado que éstos inicialmente toman valores disimiles o heterogéneos. Algunos atributos toman valores simbólicos, otros toman valores discretos, numéricos enteros y otros valores continuos numéricos representados en formato de coma flotante. Algunos valores son excesivamente grandes, valores numéricos representados en millones o millones de millones, y otros en valores decimales muy pequeños, de uno o dos (2) símbolos decimales. El propósito de la normalización, es, por tanto, representar todos los atributos lo más homogéneo posible, para que luego de ser procesados en el proceso de entrenamiento, los datos tengan la misma representatividad en el modelo que va a efectuar el proceso de clasificación de la información. En Vesanto, Himberg, Alhoniemi, & Parhankangas (2000)

definen la normalización como la necesidad de que ninguna característica contribuya más que otra a la medida de la distancia y se presentan seis implementaciones de métodos de normalización, de la SOM toolbox de Matlab™: *var, range, log, logistic, histD e histC*.

Para esta investigación se implementó el método de normalización *var*. La decisión fue tomada debido a la naturaleza de las características del conjunto de dados utilizado, que son tanto numéricas como categóricas, el proceso de normalización se aborda de manera diferente en cada caso. Las variables continuas se normalizan a media cero y varianza única. Se trata de una transformación lineal simple como se observa en la ecuación 1:

$$\hat{x} = \frac{x - \bar{x}}{\sigma}$$

(Ecuación 1)

donde x y σ son la media y la desviación estándar de la variable x, respectivamente. Esto equivale a la expresión de la variable x como el número de desviaciones estándar de distancia de la media. Sin embargo, las variables categóricas requieren un tratamiento diferente. Específicamente, estas variables se codificaron antes de la normalización de acuerdo con la medida de similitud (Choi, Cha, & Tappert, 2010).

Selección de características

La fase de selección de características, documentada en Hota & Shrivas (2014), se define como el proceso de optimización que trata de encontrar el mejor subconjunto de características de un conjunto fijo de ellas. El objetivo principal de la selección es reducir el tamaño de los datos de entrada para facilitar el procesamiento y análisis, descartando datos que no contribuyen en mayor medida al posterior proceso de clasificación. Lo cual genera un ahorro de tiempo en el procesamiento de los datos, sin desestimar la generación de

resultados óptimos. Según Mendoza Palechor (2013), la capacidad de emplear la selección de características, es primordial para realizar un análisis eficaz, debido a que los datos contienen información que no es necesaria para la generación del modelo.

Existen dos enfoques principales en la selección de características (Bolón, Sánchez, & Alonso, 2013), *evaluación individual*, también conocida como ranking que evalúa las características individuales asignándoles pesos de acuerdo a su grado de relevancia y el enfoque de *subconjunto individual*, el cual produce subconjuntos candidatos de características, en base a una cierta estrategia de búsqueda.

Además de esta clasificación, los métodos de selección de características también pueden dividirse en tres modelos según Newton, Monard, & Tsoumakas (2014), el de *filtros,* que se basan en las características generales de los datos de entrenamiento y llevan a cabo el proceso de selección de características como una etapa de pre-procesamiento con independencia del algoritmo de inducción. Los segundos, *envoltorios*, que consiste en la optimización de un predictor como parte del proceso de selección. Por último, los *m*étodos *embebidos* realizan la función de selección en el proceso de aprendizaje y suelen ser usados en la implementación de técnicas de aprendizaje de máquina.

Para esta investigación se han seleccionado las técnicas *CHI-SQUARE* e *INFO.GAIN* ya que en la exploración del estado del arte se observó que, al implementarlas en temas relacionados a la detección de anomalías, sus resultados fueron prometedores más sin embargo se implementaran en conjunto con redes neuronales SOM y GHSOM con el objetivo de analizar las métricas de desempeñado que arroje el modelo propuesto.

A continuación, se presentan en detalle las técnicas de selección de características *CHI-SQUARE* e *INFO.GAIN*.

Chi-Square

CHI-SQUARE es definida por Namik & Othman (2011) como una técnica de análisis útil para determinar las reglas de asociación estadística. Téngase presente que las reglas de asociación son una técnica popular para producir

calidad en las detecciones basadas en mal uso (*misused-based*); sin embargo, la debilidad de las reglas de asociación radica en el hecho de que a menudo se producen miles de normas lo que reduce el rendimiento de los IDS.

El cálculo estadístico de *CHI-SQUARE* depende de una pareja de variables (A y B), esto implica la construcción de tablas de contingencia las cuales se utilizan para examinar la relación entre las variables (A y B) o bien explorar la distribución que posee una variable categórica entre diferentes muestras. *CHI-SQUARE* está representado con la siguiente formula (ecuación 2) :

$$\hat{x} = \frac{x - \bar{x}}{\sigma}$$

(Ecuación 2)

Esta técnica ha sido ampliamente usada por distintos autores con el fin de identificar y seleccionar información relevante en conjuntos de datos de información. En Namik & Othman (2011) se muestra una aplicación de post-minería para reducir el número de reglas y en consecuencia mejorar la calidad para producir firmas, utilizando dos conjuntos de datos recogidos de KDD-cup 99, dividiendo 4 grupos de datos en función del tipo de ataque (*PROBE, U2R, R2L y DOS*). A su vez, cada partición utilizo *CHI SQUARE* como técnica de computación, en donde se midió la calidad de normas aplicando la ecuación (2). Los resultados del experimento lograron reducir las reglas hasta el 98% permaneciendo la calidad de las normas.

En Muraleedharan, Parmar, & Kumar (2010) se presenta un sistema basado en el flujo de datos IP en una red computacional, para detectar anomalía utilizando *CHI-SQUARE*, este sistema ofrece una solución para identificar actividades anómalas y de inundación aplicando un análisis del comportamiento automático del tráfico de la red.

En Saganowski, Goncerzewicz, & Andrysiak, (s.f) se propone un pre-procesador para detección de anomalías de red para el sistema basado en software

IDS SNORT. El preprocesador fue examinado en la red real. Los resultados presentados demuestran que los algoritmos empleados pueden ser utilizados para mejorar la ciber-seguridad y la resiliencia de las infraestructuras de red.

Gong, Fang, Liu, & Li (2014) proponen una arquitectura de *Sistema de Detección de Intrusiones Multi-agente - MIDS* y un enfoque de selección de características para proteger un *Sistema de Control Industrial - ICS*. La arquitectura propuesta está diseñada para la detección de intrusos y el control descentralizado de la prevención en grandes redes conmutadas, por lo que puede hacer que el IDS sea eficaz y escalable, además, el enfoque de detección de características propuesto mejora la fiabilidad de detección.

Info.Gain

Esta técnica también conocida como *Information Gain*, es utilizada para identificar el nivel de relevancia o ranking de las características de una colección de datos (Hota & Shrivas, Advanced computing, Networking and Informatics, 2014). El atributo con la mayor ganancia de información se elige como el atributo de división para el nodo ***N***. Este atributo minimiza la información necesaria para clasificar las duplas en la partición resultante y refleja la menor aleatoriedad o impureza en estas particiones. La ecuación (3) define este nivel de relevancia.

Autores como Pal & Parashar (2014) han propuesto esta técnica para la selección de características en sistema de detección de intrusos, donde al realizar esta combinación permiten al sistema generar un subconjunto óptimo de atributos en medio de una enorme cantidad de información que viaje a través de una red.

Spolaˆ propuso en Newton, Monard, & Tsoumakas (2014) un modelo alternativo LCFS, que construye nuevas etiquetas en base a las relaciones de las etiquetas originales, en sus experimentos utilizan el método de selección de características *INFO.GAIN* como medida para evaluar las características. Los resultados muestran que la fijación de LCFS con estrategias sencillas utilizando pares de etiquetas da lugar a mejores clasificadores que las construidas utilizando el estándar enfoque en el conjunto de datos original.

Enache & Sgârciu (2014) realizaron un estudio de un sistema de detección de anomalías basado en *Support Vector Machine - SVM*, donde la principal contribución es la propuesta de un *NIDS* el cual combina IG (*INFO.GAIN*) y SVM, como resultado se obtuvo que *IG-BA-SVM*, obtiene la tasa de detección (95.76%), precisión (94,16%) y la tasa de falsas alarmas más baja (0.0408) con respecto a *SVM*.

Entrenamiento

Redes Neuronales SOM (Self-Organizing Map)

Se han presentado evidencias de que hay neuronas en el cerebro capaces de organizarse en muchas zonas de forma tal que la información captada del entorno a través de los órganos sensoriales se representa internamente en forma de mapas bidimensionales (Hilera & Martínez, 2000). Aunque esta organización neuronal está predeterminada genéticamente, se presume que parte de ella se origina mediante el aprendizaje, sugiriendo que el cerebro lograría tener la posibilidad de formar mapas topológicos de la información recibida del exterior. Es precisamente en estas ideas en las que se basa *T. Kohonen* (Kohonen, 1982) para proponer un modelo de red neuronal con capacidad de formar mapas de características de manera similar a como ocurre en el cerebro donde un estímulo externo o datos de entrada por sí solo, con estructura propia y descripción funcional del comportamiento de la red, puede originar mapas topológicos de semejanza con los creados por el cerebro. Según Kohonen (2013) la característica más preponderante de los SOM es que aprenden a clasificar los datos mediante un algoritmo de aprendizaje NO supervisado (un SOM aprende a clasificar los datos de entrenamiento sin ningún tipo de control externo).

Funcionamiento de SOM

Una red SOM trata de establecer una correlación entre los datos de entrada y un espacio bidimensional de salida, creando mapas topológicos de dos dimensiones, con miras a que, si existen datos de entrada con características comunes, se activen neuronas situadas en zonas próximas de la capa de salida. Una representación visual de este procedimiento se muestra en la Figura 2,

Corresponde a una arquitectura típica de un mapa auto organizado en las que neuronas de salida se disponen en forma bidimensional para representar los mapas de características. De ahí que los SOM también se les llame *Kohonen Feature Maps*.

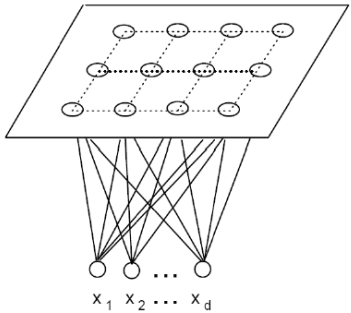


Figura 2. Arquitectura típica de un mapa auto organizativo (Kohonen, 2001)

Esta visualización del conjunto de datos es generada de forma automática y sin intervención humana, lo que categoriza al algoritmo SOM como un algoritmo de redes neuronales de aprendizaje no supervisado a los que nos hemos referido anteriormente, que no utiliza valores a priori, ni tiene retroalimentación. Este modelo es uno de los más populares que se utilizan en redes neuronales artificiales y pertenece a la categoría de redes con aprendizaje competitivo. En este aprendizaje, las células reciben de manera idéntica la información de entrada sobre la cual compiten, siendo esta competencia una lucha por determinar cuál de las neuronas es la que mejor representa a un estímulo de entrada dado. Como resultado de esta competición, solo una neurona es activada en cada momento, como se muestra en la Figura 3.

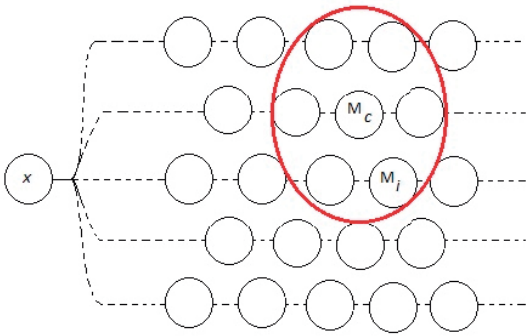


Figura 3. Neurona ganadora = Mc y Neurona vecina = Mi (Kohonen, 2001)

Redes Neuronales GHSOM (Growing Hiererchical Self-Organizing Map)

Según Rauber, Pampalk, & Merkl, (2002), GHSOM es una estructura jerárquica y dinámica, desarrollada para superar las debilidades y problemas que presenta SOM. La estructura GHSOM consiste en múltiples capas compuestas de varias SOM independientes cuyo número y tamaño se determinan durante la fase de entrenamiento. El proceso de crecimiento de adaptación está controlado por dos parámetros que determinan la profundidad de la jerarquía y la amplitud de cada mapa. Por lo tanto, estos dos parámetros son los únicos que tienen que ser fijados inicialmente en GHSOM. Según Dittenbach, Merkl, & Rauber (2002) hay dos propósitos para la arquitectura de GHSOM, el primero que sostiene que un SOM tiene una arquitectura de red fija, es decir el número de unidades de uso, así como la distribución de las unidades tiene que ser determinada antes del entrenamiento, y el segundo que dice que los datos de entrada que son de naturaleza jerárquica deberían estar representados en una estructura jerárquica para mayor claridad de la representación.

GHSOM utiliza una estructura jerárquica de varias capas, donde cada capa está formada por un número de SOM independientes. Solo un SOM se utiliza en la primera capa de la jerarquía según Dittenbach, Merkl, & Rauber (2000) y por cada unidad del mapa, una SOM podría añadirse a la siguiente capa de la jerarquía. Este principio se repite con el tercer nivel del mapa y las demás capas de la GHSOM, tal como se muestra en la Figura 4.

Para llevar a cabo el proceso de inserción de columnas o filas en un GHSOM deben seguirse los siguientes pasos: los pesos de cada unidad se inicializan con valores aleatorios, seguidamente se aplica el algoritmo estándar de SOM. Luego, la unidad con la mayor desviación entre el vector de pesos y los vectores de entrada es elegida para representar la unidad de error, y por último, una fila o una columna se inserta entre la unidad de error y la unidad vecina más distinta en términos de espacio de entrada. Los pasos 1 al 3 se repiten hasta que el Error de Cuantificación Medio (*MQE*) alcance un determinado umbral, una fracción del error de cuantificación promedio de la *Unidad i*, en la capa de procedimiento de la jerarquía tal como se aprecia en la Figura 5.

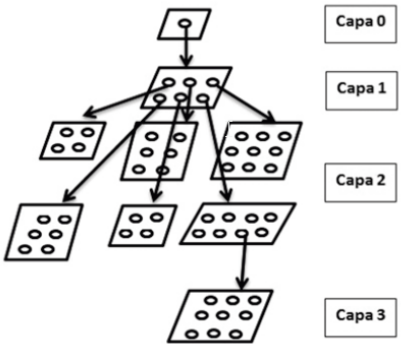
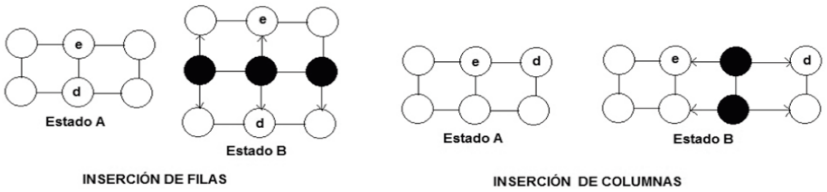


Figura 4. Estructura de Una Red GHSOM. *(Dittenbach, Merkl, & Rauber, The GHSOM Architecture and Training Process, 2016)*

Figura 5. Proceso de inserción de filas de una red GHSOM. *(Dittenbach, Merkl, & Rauber, 2016)*



El aprendizaje en GHSOM

Según Dittenbach, Merkl, & Rauber (2000), la etapa de entrenamiento del algoritmo GHSOM empieza con la configuración inicial, en este paso se crea el "mapa" en el nivel 0 con una sola unidad. El vector m_0 de pesos de esta unidad es inicializado, con la media de todos los vectores de entrada y también se calcula el error medio de cuantificación mqe_0. Posterior a este procedimiento de inicialización, viene el proceso de formación y crecimiento del mapa. El crecimiento de la estructura, inicia creando un nuevo SOM debajo de la capa 0 del mapa, con un tamaño inicial de 2x2 unidades. El proceso de crecimiento continúa hasta que el error medio de cuantificación *MQE*, en mayúsculas, alcanza una cierta fracción τ_1 del mqe_u de la unidad correspondiente en la capa superior; es decir, la unidad que constituye la capa 0 del mapa para la primera capa de mapa.

Métricas de desempeño

- **Sensibilidad**: Define sensibilidad como la capacidad que tiene un IDS para identificar resultados "verdaderos positivos":

$$Sensibilidad = \frac{VP}{VP + FN} \qquad \textit{(Ecuación 4)}$$

- **Especificidad***:* Define especificidad como la capacidad que tiene un IDS de medir la proporción de "verdaderos negativos" que se han identificado correctamente:

$$Especificidad = \frac{VN}{VN + FP} \qquad \textit{(Ecuación 5)}$$

- **Exactitud**: Define exactitud como el grado de cercanía de las mediciones de una cantidad (X) al valor de la magnitud real (Y); Es decir la proporción de resultados verdaderos (tanto verdaderos positivos como verdaderos negativos). Una exactitud del 100% significa que los valores medidos son exactamente los mismos que los valores dados:

$$Exactitud = \frac{VP + VN}{VP + FP + FN + VN} \qquad \textit{(Ecuación 6)}$$

- **Precisión:** Define la proporción de verdaderos positivos contra todos los resultados positivos:

$$Precisión = \frac{VP}{VP + FP} \qquad \textit{(Ecuación 7)}$$

Escenarios experimentales

En los escenarios se utilizó variación de selección de características (aplicando las técnicas *CHI-SQUARE* e *INFO.GAIN*). De manera complementaria para la clasificación se utilizaron las técnicas de SOM y GHSOM, lo cual permitió comparar su comportamiento respecto a otros estudios similares.

En todos los escenarios de simulación se utilizó el dataset DARPA NSL-KDD. Para el proceso de entrenamiento, específicamente se utilizó el KDD-Train y para el proceso de clasificación (o prueba) se utilizó el KDD-Test, ambos al 100%, es decir con todos los registros con los que viene el dataset original de DARPA. Una vez implementadas las técnicas de selección y clasificación, antes mencionadas, se procedió a evaluar una serie de métricas de desempeño, con el objeto de valorar la eficiencia del modelo propuesto.

Escenario experimental No.1 (Conjunto completo de características)

En este escenario se consideran las 41 características del dataset KDD-Train, variando las técnicas de entrenamiento SOM y GHSOM de forma separada, y realizando la clasificación utilizando el dataset KDD-Test con todas sus características.

Tabla 5.
Resultados escenario experimental SOM y GHSOM sin selección de atributos

Escenario	No Características	Exactitud	Sensibilidad	Especificidad	Precisión
SOM	41	65.05%	32.31%	93.80%	82.50%
GHSOM	41	60.27%	29.02%	93.31%	81.6%

Como se observa en la Tabla 5, el método SOM plantea una mejora sustancial en todas sus métricas con respecto a GHSOM, trazando unos índices de precisión en 82.5%, exactitud de 65.05%, sensibilidad en 32.31% y especificidad con 93.80%. Esto último significa que la clasificación con SOM, utilizando 41 características, detecta el trafico normal en un porcentaje del 93.80%.

Escenario experimental No.2 (Conjunto de características seleccionadas y clasificando con SOM)

En este escenario experimental se utilizaron las técnicas de selección de características *CHI-SQUARE* e *INFO.GAIN*, al aplicarlas se pudo efectuar una reducción de las características, obteniendo un dataset depurado, con el cual

se realizó el entrenamiento del SOM. El dataset específicamente usado en este proceso ha sido el KDD-Train al 100%.

En este escenario experimental se plantearon dos simulaciones, en la primera se efectuó la selección de características mediante la técnica *CHI-SQUARE* y la clasificación aplicando SOM, en la segunda, la selección de características fue mediante la técnica *INFO.GAIN*, utilizando también como clasificador a SOM. Luego de analizar los resultados obtenidos en estas dos simulaciones, se procedió a consolidar los resultados analizando las métricas antes mencionadas.

Simulación CHI-SQUARE + SOM

Se desarrolló una simulación tomando el dataset KDD-Train 100% y aplicando la técnica de selección de característica *CHI-SQUARE*. Se pudo identificar el orden de prioridad de las características del dataset, lo que permitió variar el número de ellas, generando pruebas con 5, 10, 11, 15-20, 30 y 41 características. Se decidió enfatizar las pruebas en el intervalo de 15-20 debido a que producto del análisis del estado del arte, se ha evidenciado que un número importante de propuestas generan buenos resultados al utilizar un número de características contenidas en este rango.

En cada una de estas pruebas de simulación, el dataset con un número específico de características, permitió el entrenamiento del SOM. Posteriormente se efectuó el proceso de clasificación, utilizando el KDD-Test al 100%. Los mejores resultados en cada una de las métricas se obtuvieron con 10, 11, 30 y 41 características. Dado que los resultados obtenidos con 10 y 11 características, 30 y 41 características, son respectivamente muy similares (casi idénticos), y dado que el propósito de este ejercicio ha sido la disminución del número de características, se ha decidido escoger como las mejores opciones, las pruebas efectuadas con 10 y 30 características.

Basados en el hecho de que la exactitud es la proporción de resultados verdaderos (tanto verdaderos positivos, como verdaderos negativos) y teniendo en cuenta los resultados obtenidos en las pruebas efectuadas con 10 y 30 características, los cuales representan valores de exactitud del 63,73% y 65,05% respectivamente con una diferencia porcentual equivale a 1,32%, se logra evi-

denciar que si bien el mejor resultado lo arroja la evaluación con 30 características, los resultados obtenidos con 10 características son muy interesantes, dado que con un número muy bajo de características el diferencial porcentual de la exactitud respecto a 30 características, es inferior al 2%. Y con sólo 10 características muy probablemente se optimizará más, el rendimiento del clasificador, que al evaluar 30 características (Tabla 6).

Simulación INFO.GAIN + SOM

Se simuló tomando el dataset KDD-Train 100% y aplicando la técnica de selección de característica *INFO.GAIN*. Con esta técnica, se pudo identificar el orden de prioridad de las características del dataset, lo cual permitió variar el número de ellas, generando pruebas con 5, 10, 11, 15-20, 30 y 41 características. En coherencia con los anteriores escenarios de simulación, se decidió enfatizar las pruebas en el intervalo de 15-20 dando como mejor resultado la utilización de 15 características generando una exactitud del 92,10% notablemente superior a las demás pruebas como se evidencia en la Tabla 7.

Tabla 6.
Resultados de las pruebas de simulación CHI-SQUARE + SOM con reducción de características

Escenario	No Características	Exactitud	Sensibilidad	Especificidad	Precisión
CHI-SQUARE + SOM	5	62,09%	27,47%	88,53%	77,30%
	10	63,73%	31,81%	94,55%	83,00%
	11	63,72%	31,81%	94,55%	83,00%
	15	57,52%	25,58%	88,57%	77,00%
	16	57,52%	25,58%	88,57%	77,00%
	17	57,52%	25,58%	88,57%	77,00%
	18	57,05%	25,30%	88,41%	76,80%
	19	57,05%	25,30%	88,41%	76,80%
	20	57,05%	25,30%	88,41%	76,80%
	30	65,05%	32,31%	93,80%	82,50%
	41	65,05%	32,31%	93,80%	82,50%

Escenarios experimental No.3 (Conjunto de características seleccionadas y clasificando con GHSOM)

e desarrolló una simulación tomando el dataset KDD-Train 100% y aplicando la técnica de selección de característica CHI-SQUARE.Con esta técnica, se pudo identificar el orden de prioridad de las características del dataset, lo cual permitió variar el número de ellas, generando pruebas con 5, 10, 11, 15-20, 30 y 41 características.

En cada una de estas pruebas de simulación, se utilizó el dataset con un número específico de características, lo cual permitió el entrenamiento del GHSOM. Posteriormente se efectuó el proceso de clasificación, utilizando elKDD-Test al 100%.

Como se observa en la Tabla 7 al combinar la técnica de selección *CHI-SQUARE* con la técnica de clasificación GHSOM utilizando 5 características, se obtuvieron los mejores porcentajes en las métricas (exactitud 70,14%, sensibilidad 36,81%, especificidad 96,69% y precisión 85,80%).

Tabla 7.
Resultados de las pruebas de simulación INFO.GAIN + SOM con reducción de características

Escenario	No Características	Exactitud	Sensibilidad	Especificidad	Precisión
INFO.GAIN + SOM	5	62,08%	27,47%	88,52%	77,30%
	10	63,72%	31,81%	94,55%	83,00%
	11	63,72%	31,81%	94,55%	83,00%
	15	92,10%	93,52%	90,54%	92,10%
	16	57,52%	25,58%	88,57%	77,00%
	17	57,52%	25,58%	88,57%	77,00%
	18	57,05%	25,30%	88,41%	76,80%
	19	57,05%	25,30%	88,41%	76,80%
	20	57,05%	25,30%	88,41%	76,80%
	30	65,06%	32,31%	**93,80%**	82,50%
	41	65,05%	32,31%	93,80%	82,50%

Tabla 8.
Resultados de las pruebas de simulación CHI-SQUARE +GHSOM con reducción de atributos

Escenario	No Características	Exactitud	Sensibilidad	Especificidad	Precisión
CHI SQUARE + GHSOM	5	70,14%	36,81%	96,69%	85,80%
	10	64,16%	31,69%	94,46%	83,10%
	11	64,69%	32,03%	94,51%	83,20%
	15	58,74%	25,94%	89,02%	77,60%
	16	58,71%	25,92%	89,02%	77,60%
	17	58,84%	25,99%	89,04%	77,60%
	18	62,28%	30,75%	94,85%	83,20%
	19	62,26%	30,74%	94,85%	83,20%
	20	62,26%	30,74%	94,85%	83,20%
	30	60,26%	29,06%	93,43%	81,70%
	41	60,26%	29,06%	93,43%	81,70%

Simulación INFO.GAIN + GHSOM

Se desarrolló un escenario experimental tomando el dataset KDD-Trainal 100%, variando el número de características seleccionadas debido a la identificación de la relevancia de las mismas, mediante la técnica *INFO.GAIN*. Permitiendo identificar las características más importantes con el fin de entrenar la red GHSOM, posteriormente se realizó el proceso de clasificación usando el dataset KDD-Test al 100%. Al implementar *INFO.GAIN*+GHSOM con 5 características se obtuvieron los mejores resultados, tal como se observa en la Tabla 9.

Tabla 9.
Resultados de las pruebas de simulación INFO.GAIN+GHSOM con reducción de atributos

Escenario	No Características	Exactitud	Sensibilidad	Especificidad	Precisión
INFO.GAIN + GHSOM	5	70,14%	36,81%	96,69%	85,80%
	10	64,16%	31,69%	94,46%	83,10%
	11	64,16%	31,68%	94,45%	83,10%
	15	58,74%	25,94%	89,02%	77,60%
	16	58,70%	25,92%	89,02%	77,60%
	17	58,84%	25,99%	89,04%	77,60%
	18	62,28%	30,75%	94,85%	83,20%
	19	62,26%	30,74%	94,85%	83,20%
	20	62,26%	30,74%	94,85%	83,20%
	30	60,26%	29,06%	93,43%	81,70%
	41	60,26%	29,06%	93,43%	81,70%

Consolidación de resultados

En la Tabla 10 se comparan los mejores resultados obtenidos de los escenarios de simulación anteriormente expuestos. A partir de ello se deduce que si bien la propuesta *CHI-SQUARE*+SOM utilizando 10 características, genera mejores resultados en cuanto a la especificidad (94,55%), la propuesta *INFO. GAIN*+SOM utilizando 15 características, le supera considerablemente en las otras métricas (exactitud 92,10%, sensibilidad 93,52% y precisión 92,10%).

Por tanto, dado que la exactitud es la métrica que mejor define la capacidad de acierto de un clasificador, se concluye que la propuesta *INFO.GAIN*+SOM es considerablemente superior a la propuesta *CHI-SQUARE*+SOM.

Tabla 10.
Mejores resultados de las pruebas de simulación aplicando SOM

Escenario	No Características	Exactitud	Sensibilidad	Especificidad	Precisión
CHI- SQUARE + SOM	10	63,73%	31,81%	94,55%	68,27%
INFO.GAIN + SOM	15	92,10%	93,52%	90,54%	92,10%

La Tabla 11. resume los resultados al implementar *CHI-SQUARE* e *INFO. GAIN* con GHSOM de la cual se obtienen los mismos resultados en cada una de las métricas evaluadas, para todas las pruebas de simulación con variación del número de características.

Tabla 11.
Mejores resultados de las pruebas de simulación aplicando GHSOM

Escenario	No Características	Exactitud	Sensibilidad	Especificidad	Precisión
CHI SQUARE + GHSOM	**5**	**70,14%**	**36,81%**	**96,69%**	**85,80%**
INFO.GAIN + GHSOM	**5**	**70,14%**	**36,81%**	**96,69%**	**85,80%**

Conclusiones

Producto de un riguroso análisis del estado del arte, se evidencia que diferentes investigadores vienen haciendo uso de las Redes Neuronales

Artificiales – RNA, en temas inherentes a la seguridad informática, específicamente en Sistemas de Detección de Intrusos - IDS. Si bien existen una serie de técnicas estadísticas basadas en inteligencia artificial y minería de datos tales como: Máquinas de Soporte Vectorial (Support Vector Machine – SVM), Inducción de Reglas, Lógica Difusa, Algoritmos genéticos, sistemas inmune artificial, entre otras, y que se han venido aplicando en los IDS, se observa una tendencia a utilizar redes neuronales artificiales, específicamente SOM y GHSOM, en la implementación de IDS basados en anomalías debido a su inherente capacidad de clasificación sin supervisión (Nagaraja & Jagadeesh Chandra Bose, 2006). Esto ha llevado a plantear en esta investigación el uso de las técnicas SOM y GHSOM, precisamente por los óptimos resultados obtenidos al evaluar estas técnicas y sus altas tasas de exactitud. Por ello se decidió integrarlas con las técnicas de selección de características (*CHI-SQUARE* e *INFO.GAIN*) utilizadas en una fase previa del clasificador, debido a que, pese a la obtención de buenos resultados con ellas, en otros ámbitos de investigación, no existe evidencia de la aplicación de estas técnicas en el ámbito de los IDS.

Se realizó una evaluación y estudio comparativo de las técnicas *CHI-SQUARE* e *INFO.GAIN*, a partir de ello, se identificó que no se han utilizado e hibridado, específicamente las técnicas de selección de características *CHI-SQUARE* e *INFO.GAIN*, con las técnicas de entrenamiento y clasificación basadas en Redes Neuronales Artificiales SOM y GHSOM en Sistemas de Detección de Intrusos. La implementación de estas técnicas en este trabajo se da como mecanismo previo al proceso de entrenamiento y clasificación con SOM y GHSOM, para documentar y analizar los resultados obtenidos, con miras a futuras implementaciones en IDS basados en anomalías.

Posteriormente y basados en los resultados investigativos, se diseñó y desarrolló un modelo funcional que fue simulado mediante el desarrollo de una serie de scripts en Matlab™, utilizando las técnicas de selección de características *CHI-SQUARE* e *INFO.GAIN* y las técnicas de clasificación SOM y GHSOM. Paso seguido, se efectuó un diseño de experimentos y luego se ejecutaron cada uno de los escenarios de experimentación propuestos. Los resultados obtenidos en cuanto a las tasas de detección de tráfico normal y de ataques, en cada uno de estos escenarios, se evaluaron mediante las métricas de desempeño (exactitud, especificidad, sensibilidad y precisión).

El modelo arroja resultados prometedores para la correcta identificación de intrusiones y de tráfico normal, utilizando solo 15 características de las 41 características posibles que contiene el conjunto de datos. Por lo tanto, se concluye que las técnicas de selección de características *INFO.GAIN* y *CHI-SQUARE* integradas con las redes neuronales artificiales basadas en SOM y GHSOM prometen un real acercamiento porcentual de aciertos en el desempeño de la detección del tráfico en redes de datos.

Referencias

Andersen, J., Glasdam, S.-M., Larsen, D., & Molenaar, N. (2016). New Concepts of Quality Assurance in Analytical Chemistry: Will They Influence the Way We Conduct Science in General? *19th Romanian International Conference on Chemistry & Chemical Engineering* (RICCCE 19), 203, 1582-1590.

Bolón, V., Sánchez, N., & Alonso, A. (2013). *A review of feature selection methods on synthetic data. Knowledge and information systems.*

Bouckaert, R. R. (2008). Practical bias variance decomposition. Advances in *Artificial Intelligence - LNCS, 5360*, 247-257.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. (1984). *Classification and Regression Trees* (Wadsworth Statistics/Probability) (Vol. 1). Boca Raton London New York Washington, DC.: Chapman and Hall/CRC; Edición: New Ed (1 de enero de 1984).

California, U. o. (28 de October de 1999). (Irvine) Recuperado el 04 de 2018, de KDD Cup 1999 Data: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

Choi, S.-S., Cha, S.-H., & Tappert, C. (2010). A survey of binary similarity and distance measures. 8(1), 43-48.

Dain, O., & Cunningham, R. (2001). *Fusing Heterogeneous Alert Streams into Scenarios*. Springer.

DARPA. (7 de April de 2018). Obtenido de Canadian Institute for Cybersecurity: http://www.unb.ca/cic/datasets/nsl.html Dittenbach, M., Merkl, D., & Rauber, A. (2002). Organizing and exploring high-dimensional data with the Growing Hierarchical Self-Organizing Map. FSKD, 626-630.

Dittenbach, M., Merkl, D., & Rauber, A. (2016). *The GHSOM Architecture and Training Process.* Obtenido de Department of Software Technology. Vienna University of Technology: http://www.ifs.tuwien.ac.at/~andi/ghsom/description.html

Dittenbach, M., Merkl, D., & Rauber, A. (July de 2000). The Growing Hierarchical Self-Organizing Map. *International Joint Conference on Neural Networks - IJCNN,* 15-19.

Eid, H., Hassanien, A., Kim, T., & Banerjee, S. (2013). Linear correlation-based features selection for network intrusion detection model. Advances in *Security of Information and Communication Networks*, 240-248.

Enache, A.-C., & Sgârciu, V. (2014). Anomaly intrusions detection based on support vector machines with bat algorithm. *System Theory, Control and Computing (ICSTCC), 2014 18th International Conference*, (págs. 856-861).

Ghorbani, A., Lu, W., & Tavallae, M. (2009). *Network Intrusion Detection and Preven- tion: Concepts and Techniques*. Springer-Verlag.

Girardin, L. (1999). *An Eye on Network Intruder-Administrator Shootouts*. Santa Clara.

Gong, Y., Fang, Y., Liu, L., & Li, J. (2014). *Multi-agent Intrusion Detection System Using Feature Selection Approach.* Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP).

Gu, H., & Ji, Q. (2004). An automated face reader for fatigue detection. In Automatic Face and Gesture Recognition. *Sixth IEEE International Conference,* (págs. 111-116).

Hilera, J., & Martínez, V. (2000). *Redes neuronales artificiales: fundamentos modelos y aplicaciones*. Madrid: Alfaomega Ra-Ma.

Hota, H., & Shrivas, A. (2014). Advanced Computing, Networking and Informatics. 1(27).

Hota, H., & Shrivas, A. (2014). Advanced computing, Networking and Informatics. Advanced Computing and Informatics Proceedings of the *Second International Conference on Advanced Computing*, Networking and Informatics (ICACNI-2014), 1(27).

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics, 43*(1), 59-69.

Kohonen, T. (2001). *Self-Organizing Maps.* Springer-Verlag Berlin Heidelberg.

Kohonen, T. (2013). Essentials of the self-organizing maps. *Neural Networks*, 52-65.

Lakshmanan, V., Fritz, A., Smith, T., Hondl, K., & Stumpf, G. (2007). An automated technique to quality control radar reflectivity data. *Journal of applied meteorology and climatology, 46*(3), 288-305.

Levin, I. (2000). KDD-99 classifier learning contest. *LLSoft's results overview, 1*(2), 65-66.

Mendoza, F. (2013). A*plicación de selección de características, metricas de aprendizaje y reducción de dimensión en sistemas de detección de intrusos.*

Muraleedharan, N., Parmar, A., & Kumar, M. (2010). *A flow based anomaly detection system using Chi-square*.

Nagaraja, G., & Jagadeesh Chandra Bose, R. (January de 2006). Adaptive conjugate gradient algorithm for perceptron training. *Neurocomputing, 69*(4-6), 368-386.

Newton, S., Monard, M., & Tsoumakas, G. (2014). Label construction for multi-label features selection. *2014 Brazilian Conference on Intelligent Systems. IEEE*.

Namik, A., & Othman, Z. (2011). Reducing network intrusion detection association rules using Chi-Squared pruning technique. In *Data Mining and Optimization (DMO). 3rd Conference*, págs. 122-127. IEEE.

Nziga, J. (2011). Minimal dataset for network intrusion detection systems via dimensionality reduction. *6th International Conference on Digital Informa- tion Management (ICDIM)*.

Pal, D., & Parashar, A. (2014). Improved genetic algorithm for intrusion detection system. *International conference on computational intelligence and communication networks,* (págs. 835-839).

Panda, M. A. (2010). Discriminative multinomial naive bayes for network intrusion detection. I*nformation Assurance and Security* (IAS, 5-10.

Pfahringer, B. (2000). Winning the FDD99 classification cup: bagged-boosting. *SIGKDD Explor, 1*(2), 67-75.

Sadkhan, S. (2009). On artificial intelligence approaches for network intrusion detection systems. MASAUM *Journal of Computing*, 236-243.

Saganowski, Ł., Goncerzewicz, M., & Andrysiak, T. (s.f.). Anomaly Detection Preprocessor for SNORT IDS System. Image Processing and Communications Challenges 4. 184, págs. 225-232. Springer.

Sebe, N., Sun, Y., Bakker, E., Lew, M., Cohen, I., & Huang, T. (2004). Towards authentic emotion recognition. In *Systems, Man and Cybernetics, IEEE International Conference* (págs. 623-628). IEEE.

Selvakani, S. K., & Regan, S. R. (2010). Integrated Intrusion Detection System Using Soft Computing. *International Journal of Network Security*, 87-92.

Tavallaee, M., Stakhanova, N., & Ghorbani, A. (5 de September de 2010). Toward credible evaluation of anomaly-based intrusion-detection methods. *IEEE Trans. Syst. Man Cybern., C*(40), 516–524.

Vesanto, J., Himberg, J., Alhoniemi, E., & Parhankangas, J. (April de 2000). (SOM Toolbox Team - Helsinki University of Technology) Recuperado el 6 de January de 2016, de SOM Toolbox for Matlab 5: http://www.cis.hut.fi/projects/somtoolbox/

Xiaonan, S., & Banhzaf, W. (2010). The use of computational intelligence in intrusion detection systems: A review. *Applied Soft Computing, 10*(1), 1-35.

Zargari, S., & Voorhis, D. (2012). Feature selection in the corrected KDD-dataset, in: *Proceedings of the 3rd International Conference on Emerging Intelligent Data and Web Technologies*. 174-180.

ETL (extracción, transformación y carga) o intercambio de información entre aplicaciones empresariales con el fin de crear un prototipo

Yexid Montenegro García[1]; Christian Hernán Obando Ibarra[2];
Hugo Eduardo Pérez Muñoz[3]; Enevis Rafael Reyes Moreno[4]

Resumen

Las compañías actuales recolectan grandes volúmenes de datos, pero muchas veces no son capaces de aprovecharlos. Para convertirlos en información, estos volúmenes de datos son transportados mediante diversos softwares de ETL. Los procesos de extracción, transformación y carga (ETL, por sus siglas en inglés) son responsables de las operaciones que se ejecutan en el núcleo de una bodega de datos. En primer lugar, los datos son extraídos de una fuente de datos que puede ser una base de datos transaccional, archivos de cualquier formato, una página web, documentos de cualquier clase, entre otros. En segundo lugar, los datos extraídos son trasladados a un área especial de la bodega llamado área temporal (o área staging) donde son transformados, homogeneizados y limpiados. Las transformaciones más comunes incluyen filtros y validaciones sobre los datos, para asegurarse de que éstos cumplan las reglas de negocio y las restricciones de integridad que la bodega o el sistema de destino necesita. Finalmente, los datos son cargados al sistema o a la bodega de datos. En una bodega de datos tradicional, el proceso de ETL refresca periódicamente la bodega de datos durante los periodos de baja operación,

1 Ingeniero de Sistemas y Computación, candidato a Magister en seguridad Informática. Docente de tiempo completo de la Corporación Universitaria Americana. fmontenegro@coruniamericana.edu.co. ORCID ID 0000-0003-2708-4497.

2 Ingeniero en Electrónica y Telecomunicaciones, Especialista en Seguridad en Informática, Magister en Tecnologías de la Información y Comunicación. Docente de tiempo completo de la Corporación Universitaria Americana. cobando@americana.edu.co. ORCID ID 0000-0003-2326-8934.

3 Administrador de Sistemas informáticos, Analista de desarrollo Intergrupo. hperez@intergrupo.com ORCID ID https://orcid.org/0000-0001-8481-4332.

4 Director de Ingenierías de la Corporación Universitaria Americana. Magister en Entornos Virtuales de Aprendizaje, Universidad de Panamá. Especialista en ciencias Electrónicas e informáticas de la Universidad de Antioquia, Especialista en Entornos Virtuales de aprendizaje. Correo electrónico de contacto: diringenieriasmed@americana.edu.co. ORCID: https://orcid.org/0000-0003-4145-1898

por ejemplo, en la noche. Las demandas y necesidades del negocio requieren que una bodega de datos esté actualizada y sea confiable mediante cargas frecuentes de procesos ETL.

Palabras clave: Automatización de ETL, bodega de datos, extracción, limpieza de datos, transformación, proceso ETL.

ETL process (Extract, Transform and Load) or exchange of information between business applications in order to create a prototype

Abstract

Current companies collect large volumes of data, but often they are not able to take advantage of them. To convert them into information, these data volumes are transported through various ETL softwares. The processes of extraction, transformation and loading (ETL, for its acronym in English) are responsible for the operations that run in the core of a data warehouse. First, the data is extracted from a data source that can be a transactional database, files of any format, a web page, documents of any kind, among others. Secondly, the extracted data are transferred to a special area of the winery called temporary area where they are transformed, homogenized and cleaned. The most common transformations include filters and validations on the data, so that the business rules and the integrity restrictions that the warehouse or the destination system needs are met. Finally, the data is loaded into the system or into the data warehouse. In a traditional data warehouse, the update process of the data warehouse during periods of low operation, for example, at night. The demands and needs of businesses require an esthetic and reliable data warehouse through the frequent loading of ETL processes.

Key words: ETL Automation, data warehouse, extraction, transformation, data cleaning, ETL process.

Introducción

El Proceso de ETL es un conjunto de programas encargados de la extracción, transformación y carga de datos. Es uno de los componentes más importantes de una estrategia de bodega de datos en una compañía, puesto que debe asegurar la calidad de datos final de la bodega de datos. Existen muchas herramientas, licenciadas y gratuitas que se pueden usar para desarrollar aplicaciones (o soluciones) de tipo ETL.

Este prototipo se desarrollará mediante el componente de SQL Server llamado Integration Services[5] (Servicios de Integración de Microsoft SQL Server). Integration Services es una plataforma para la construcción de Soluciones de integración y transformación de datos de nivel empresarial. Puede utilizarse Integration Services para resolver problemas complejos de negocios mediante la copia o la descarga de archivos, el envío de mensajes de correo electrónico en respuesta a eventos, la actualización de los almacenes de datos, limpieza y extracción de datos, y la gestión de objetos y datos de SQL Server. Los paquetes pueden trabajar solos o en concierto con otros paquetes para hacer frente a las necesidades empresariales complejas. Servicios de integración pueden extraer y transformar datos de una amplia variedad de Fuentes, tales como archivos XML de datos, archivos planos, y Fuentes de datos relacionales, y luego cargar los datos en uno o más destinos.

Como parte de Integration Services se incluye un amplio conjunto de tareas y transformaciones; herramientas para la construcción de paquetes; y servicios para ejecutar y administrar paquetes. Puede utilizar las herramientas gráficas de Integration Services para crear soluciones sin escribir una sola línea de código; o se puede programar el modelo extensivo de objetos de Integration Services para crear paquetes mediante programación y tareas personalizadas de código y otros objetos de paquete.

Para efectos de la investigación y con el objeto de facilitar la generalización del concepto de ETL, en este capítulo, se usará indistintamente el término Bodega de Datos o Sistema Destino para referirnos a cualquier repositorio don-

5 La versión SQL Server utilizada en este prototipo es la versión SQL Server 2012 Express Edition. Esta es una edición gratuita con muchas funcionalidades disponibles y que puede usarse para diseñar ETLs desde el computador personal y publicar en un servidor de producción.

de se pueda llevar los datos extraídos y transformados. En concreto, la etapa de carga (load) de ETL puede bien aplicarse al término usado por los autores de Data Warehousing (por ejemplo, Inmon y Kimball), pero también se refiere a cualquier sistema (base de datos, archivo plano, archivo XML, entre otros) donde se pueda almacenar de manera persistente los datos.

Contexto sobre la problemática

En el mercado tecnológico empresarial es muy común encontrar herramientas de tipo software que realizan estos procesos ETL. Lo que no es común es encontrar un software que automatice estos tres procesos. La motivación de presentar o proponer esta solución es la de desarrollar una automatización de procesos ETL.

¿Cómo intercambia información una compañía?

Todas las compañías actuales hacen uso de aplicaciones para este fin, algunas llamadas sistemas core (base o fundamentales), otras llamadas aplicaciones de apoyo.

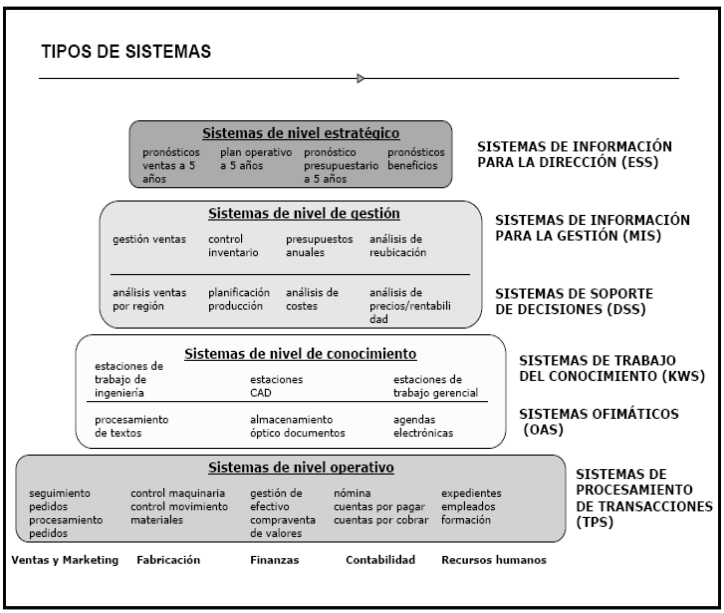


Figura 1. Tipos de sistemas de información empresarial de acuerdo a Kenneth C. Laudon Jane P. Laudon.
Fuente: Sistemas de información gerencial 12° Edición - Kenneth C. Laudon Jane P. Laudon

Tales aplicaciones de Gestión de Ventas, Inventarios, Cartera, Nómina, de Relacionamiento con el cliente (CRM). También usan sistemas más robustos, llamados ERP (Planeación de Recursos Empresariales). Todos estos sistemas intercambian información.

- Esta información es intercambiada de forma manual.
- Los computadores personales de los empleados contienen información sensible.
- Al estar dispersos los datos, es compleja la consolidación de ellos en un solo lugar.
- Producen bloqueos en la recuperación de la información.

Una afirmación que es innegable, es que todos los sistemas requieren información de otro sistema. Por ejemplo, el sistema Nómina requiere información de Ventas para procesar los bonos de los empleados; Inventario necesita conocer cuáles son los productos de mayor demanda para planear la compra a proveedores y definir su bodegaje.

Se tienen muchos casos de compañías que mueven millones de registros, por ejemplo, compañías de telecomunicaciones móviles, donde registran cada llamada, que sirve a su vez para realizar análisis de comportamientos y gustos de clientes, servicios más utilizados, entre otros.

Entonces, la respuesta a la pregunta: ¿Cómo intercambia información una empresa?, puede resumir su sentimiento en la Figura 2:

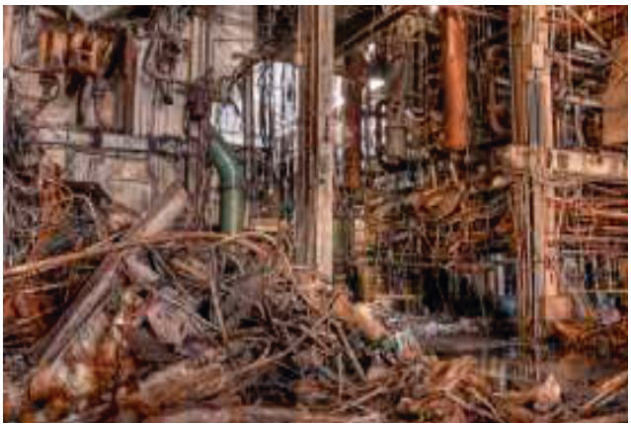


Figura 2. reflejo de la forma como intercambia información entre sus aplicaciones una empresa.

Basado en la Figura 2 y en el párrafo anterior, produce preocupación, identificar que las compañías manejan sus datos de forma inadecuada.

Como caso particular sobre la forma inadecuada de recolectar información, una compañía del sector de servicios públicos, por ejemplo, cuando algún directivo de mando superior solicita un informe a su área de tecnología y luego solicita el mismo informe al área de información financiera. El gerente queda alarmado por las discrepancias de resultados entre los dos informes. Los datos, tomados para el informe, son recolectados inadecuadamente; son costosos porque se encuentran en lugares dispersos y se recolectan de forma poco productiva; se los entregan en formatos heterogéneos (lo que conlleva a que el esfuerzo del analista de información sea arduo y tedioso). Lo más preocupante, es que se obtienen de forma manual, lo que puede producir manipulación sesgada y de dudosa calidad.

Preguntas tales como: ¿Cuánto tiempo gasta un operador de información en recolectar los datos para un informe ejecutivo? O, por ejemplo, ¿cuáles son las fuentes de datos para generar un informe de Pareto?[6]

A través de la estrategia propuesta de automatizar los procesos ETL al interior de la compañía que la use, para tener datos limpios que tengan validez; con calidad tal, que va a ser posible usarlos de un lado al otro; confiables, es decir, que cuando lleguen a la gerencia, a operaciones o al cliente se puede confiar que lo que se está entregando es lo que la compañía tiene.

Figura 3. refleja la manera adecuada de intercambiar información: Datos limpios, con calidad y confiables.

6 Es aquel que permite identificar que el 80% de algo es producido por el 20% de otro. Por ejemplo, que el 20% de mis clientes me produce el 80% de las ganancias.

Propuesta: ETL automático para uso general

Así que la Figura 3 presenta el objetivo general de la propuesta de este artículo: la construcción de un prototipo, que permita intercambiar datos entre los diferentes sistemas informáticos mediante procesos ETL y de forma automatizada. De tal manera que pueda identificar los procesos ETL sin ser un experto en la materia. Este objetivo se logra cubriendo cuatro aspectos: primero, mediante un prototipo de una aplicación realizar la extracción de datos; segundo, transformarlos aplicando las reglas de negocio que la compañía solicite; tercero cargarlos al sistema destino; y cuarto, todo ello de forma automática, es decir, en pasos simples y sin altos conocimientos en las metodologías de ETL. Cualquier compañía puede beneficiarse de este modelo de intercambio de información, como se muestra en la Figura 4.

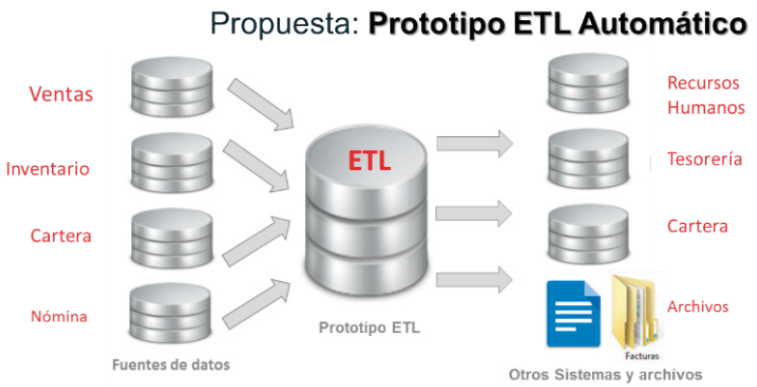


Figura 4. Propuesta: ETL Automático para uso general

Como parte de la investigación realizada se han identificado tres metodologías de inteligencia de negocios, que incluye procesos de integración de datos mediante ETL.

En la literatura técnica que se encuentra en internet y en libros de tecnología, se habla mucho de Bill Inmon (1999), Ralph Kimball(1996) y Douglas Hackney (2001) como los autores de metodologías de uso de datos para hacer inteligencia de negocios (Castillo, Morata, & del Arbol, 2015).

En nuestra propuesta se usa mayormente la metodología Kimball, es decir, la que propone que la información se puede alojar en un sitio centralizado

llamado bodega de datos (data warehouse). Para llegar a la bodega, se usan los procesos ETL. Como variación de la metodología Kimball, proponemos permitir que la información viaje de un sistema a otro, no necesariamente a una bodega, pero que cada sistema se beneficie de la información que cada sistema puede ofrecer, como se muestra en la Figura 4.

Antecedentes históricos a las bodegas de datos

La primera generación de sistemas de bodegas de datos (almacenes de datos) fue construida sobre ciertos principios establecidos por líderes de la industria. Se reconoce a dos grandes pioneros en el área de bodegas de datos: Bill Inmon y Ralph Kimball. Estos dos científicos han proporcionado las definiciones y los principios de diseño que la mayoría de los profesionales utilizan hoy en día. Aunque sus guías no sean seguidas exactamente, es común hacer referencia a la definición de bodegas de datos de Inmon y a las reglas de diseño de Kimball (Kimball & Caserta, The Data Warehouse ETL Toolkit, 2004).

El autor del ciclo de vida del modelo de negocios dimensional, Ralph Kimball define una bodega de datos (Data Warehouse) como (Kimball & Caserta, The Data Warehouse ETL Toolkit, 2004): «Un Data Warehouse es un sistema que extrae, limpia, ajusta y entrega las fuentes de información y solo entonces soporta e implementa herramientas de consulta y análisis con el propósito de tener una correcta toma de decisiones».

En los años 90, Inmon (William, 1999) llama a esta metodología Data Warehouse o Information Warehouse (Bodega de Datos o Bodega de Información, DW por sus siglas en inglés) y lo define como: "una colección de datos orientados por temas, integrados, variables en el tiempo y no volátiles para el apoyo de la toma de decisiones". Inclusive, el mismo autor lo cataloga como el "corazón del procesamiento de información de un DSS" (Decision Support System, por sus siglas en inglés). A su vez indica que una Bodega de Datos es la evolución de la información de un DSS. Mencionando la complejidad de la transformación y la integración de las fuentes de datos y la bodega de datos, aborda la necesidad de unos procesos formales de ETL. Una bodega de datos requiere procesos de integración, porque los datos que se introducen en el almacén se obtienen de una variedad de fuentes de datos (sistemas heredados, bases de datos relacionales, archivos COBOL, etc.). Inmon es consiente que

"El primer software ETL era crudo, pero rápidamente fue madurando hasta el punto donde casi cualquier transformación podría ser manejada" (William, 1999).

Los procesos ETL vienen en dos variedades: software que produce código y software que produce módulos en tiempo de ejecución que han sido parametrizados. El primero tiene la particularidad de ser más robusto, porque puede hacer que las antiguas aplicaciones (sistemas legacy o heredados) puedan usarse sin tener que cambiar su estructura o tener que adaptarse al proceso ETL, entre tanto que el ETL que produce módulos en tiempo de ejecución generalmente requiere que los datos heterogéneos de un sistema heredado sean "aplanados" o ajustados a la forma de leerlos. Este aplanamiento supone una pérdida de información valiosa.

En cualquiera de los dos casos, para lograr la integración de esa variedad de fuentes se utilizan los procesos ETL. Dichos procesos son los responsables de la extracción de los datos a partir de las diversas fuentes de datos heterogéneas, de la transformación de estos (conversión, limpieza, etc.) y de su carga en la bodega de datos. En este y otros autores (Muñoz, Mazón, & Trujillo, 2007), se reconoce ampliamente que el diseño y mantenimiento de los procesos ETL son factores claves en el éxito de proyectos de bodega de datos, como también advierte, que los procesos ETL podrían consumir hasta el 80% de los recursos de desarrollo de un proyecto de bodega de datos, y que tiene la ventaja (dependiendo de la perspectiva en que se vea), de que se puede usar para descubrir problemas ocultos de los Sistema Fuente, que deben ser subsanados en el destino. En este sentido, los procesos ETL son un componente clave, porque los datos incorrectos producirán decisiones incorrectas; por esto, un esquema correcto en la fase de diseño de la bodega de datos es absolutamente necesario (Muñoz, Mazón, & Trujillo, 2007).

Para Kimball y Ross (2002), un proceso ETL es el fundamento de las bodegas de datos. Considera que un proceso ETL bien diseñado extrae datos de las fuentes, hace cumplir estándares de calidad, a fin de que esos datos puedan ser utilizados por los desarrolladores para las aplicaciones y, de modo que los usuarios finales (también llamados gestores de información o information workers) puedan tomar decisiones estratégicas. Es decir, los datos son extraídos de los sistemas fuentes, los cuales pasan por una secuencia de validaciones

y transformaciones antes de que se carguen el sistema destino o la bodega de datos. El repositorio de los sistemas fuentes que contienen datos para una bodega de datos puede variar desde hojas de cálculo (tipo Excel o archivos csv) hasta sistemas mainframe.

Dicho proceso requiere de mucho cuidado y a menudo las características de las transformaciones y validaciones implican un alto nivel de detalle y debe tener en cuenta procedimientos para integración, limpieza, estandarización y homologación de los diferentes conceptos de negocio utilizados. Un proceso de ETL debe ser lo menos invasivo posible sobre los sistemas que soportan las fuentes de datos transaccionales. Es decir, deben consumir la menor cantidad de recursos en los sistemas de origen que proporcionan la información para la bodega. Adicionalmente, debe tener un buen control y manejo de datos que no cumplen con las reglas de integridad ni con las reglas del negocio y un buen registro de auditoría y trazabilidad sobre los datos y los procesos (Kimball & Caserta, 2004).

Conceptos clave como *cliente*, con frecuencia no se encuentran unificados o estandarizados en una organización. Es común encontrar múltiples sistemas de información con datos básicos del cliente, con información desactualizada y con un bajo nivel de sincronización. El nombre, la dirección, el correo, la fecha de nacimiento, la categorización del cliente, son datos que tienen muchas variaciones de un sistema a otro y genera contratiempos a la hora de relacionarse con el cliente o de tomar decisiones sobre la forma de acompañarlo en los procesos de fidelización, por mencionar alguno. Las organizaciones invierten ingentes cantidades de recursos en procesos para obtener la unificación y estandarización de la información del cliente. Los procesos de ETL son una herramienta para lograr esta unificación y sincronización, no solo para el cliente, sino también para las posibilidades de negocio que lo requieran.

Los procesos de ETL se pueden usar para poblar un ODS (Almacén de Datos Operacional, según Inmon), un esquema dimensional (Almacén de Datos dimensional según Kimball) o ambas de forma simultánea. Aunque por lo general puede ocurrir que los datos que se cargan a un ODS incluyen las transacciones realizadas, mientras que los datos para cargar un esquema de estrella pueden ser fotos del estado final de una tabla en la base de datos transaccional. También puede utilizarse un ODS como fuente para cargar los esquemas de estrella de un modelo dimensional (Castillo , Morata, & del Arbol , 2015).

La limpieza de datos cómo etapa separada de los procesos ETL.

Aunque podría entenderse como una acción integrada en la fase de transformación de datos, en la actualidad la tendencia es considerar la limpieza de datos como una fase separada del proceso ETL.

Esta visión corresponde a una concepción más moderna y práctica del proceso. Para ahorrar tiempo y ganar en efectividad es conveniente unificar criterios, por ejemplo, introduciendo "av" en vez de "avenida" en todos los registros de una base de datos de direcciones postales, antes de empezar el proceso ETL.

Tan importante es tener la información consolidada, como que todos los datos sean correctos y con una visión única para todos los usuarios. Solo así se pueden lograr circuitos de trabajo y análisis de los mismos, realmente óptimos y efectivos.

¿Qué sistemas se pueden integrar en un proceso ETL?

Los procesos ETL pueden incluir:

- Sistemas legacy. Es decir, heredados o antiguos.
- Sistemas nuevos. Basados en Windows, Linux y también en las redes sociales modernas: Facebook, Twitter, Linkedin, entre otros.

Los sistemas legacy o heredados se caracterizan, generalmente, por ser cerrados, no permitir cambios y tener un difícil acceso (Normalmente se necesita algún tipo de driver especial). Son sistemas que procesan hacia dentro y, por lo tanto, no permiten la agregación de una computadora que trabaje en paralelo.

Por el contrario, los sistemas nuevos o modernos (basados en Windows o Linux) son abiertos, amplios e interconectados. Un ejemplo lo constituiría una granja de servidores Linux, la cual permite la interconexión de los distintos nodos entre sí. A cualquier empresa u organización le beneficia poner en marcha un proceso ETL para mover y transformar los datos que maneja.

Algunas compañías, que compran nuevos sistemas, tales como un MDM (Master Data Management, es decir, un repositorio central estandarizado), necesitan, unificar todos los datos de la organización. Por ejemplo, si tenemos un objeto cliente en una base de datos de créditos y otro objeto cliente en la base de datos de tarjetas de crédito, lo que haría el MDM sería definir, de forma concreta e inequívoca, un registro de cliente único con su nombre y apellidos para la organización. El ejemplo anterior posibilita a los directivos tomar decisiones estratégicas basadas en el análisis de los datos cargados en las bases nuevas y actualizadas: MDM entonces es el proveedor de datos centralizados para una buena bodega de datos. Algunas bodegas de datos organizan su información por temas, lo que se conoce en la metodología BI (Inteligencia de Negocios o Business Intelligence) de Kimball, como Datamart.

Las organizaciones crecen de forma orgánica y cada vez se van agregando más fuentes de datos. Esto provoca que comiencen a surgir nuevas necesidades, como tener una visión global de todos los datos consolidados en una Bodega de datos, que se puede realizar mediante el proceso ETL: Un sistema efectivo, pero con retos y cuestiones a resolver. Los procesos ETL son muy útiles y beneficiosos para las organizaciones por su capacidad para integrar grandes bases de datos, logrando así una visión única global, que permite a los analistas y directivos, tomar las decisiones estratégicas adecuadas. La implantación de un sistema ETL bien definido supone todo un reto puesto que, para que sea realmente efectivo, debe permitir integrar los sistemas legacy (algunos ya muy obsoletos) con los más modernos. Además, el acceso a todos estos sistemas se debe producir no solo en modo de lectura, sino también como escritura.

El sistema de ETL es la base sobre la cual se alimenta la bodega de datos, pero también cualquier otro sistema (llámese MDM, OLTP o OLAP). Si el proceso ETL se diseña adecuadamente, puede extraer los datos de los sistemas de origen de datos, aplicar diferentes reglas para aumentar la calidad y consistencia de los mismos, consolidar la información proveniente de distintos sistemas, y finalmente cargar (grabar) la información en un formato acorde para la utilización por parte de las herramientas de análisis o simplemente para proveer información válida para el funcionamiento de otro sistema.

Evolución de las bodegas de datos y los procesos ETL

Los procesos ETL en este siglo (XXI) han sido un aliado de la tecnología de base de datos desde su nacimiento. Durante ese período, el software ETL estaba oculto, subyacente, como una tarea de programación de rutina más, sin ningún nombre en particular o de importancia particular. Así que, desde sus inicios hasta ahora, cualquier tipo de software de procesamiento de datos, que forma nuevos datos o usa filtros sobre registros, que calcula nuevos valores y, luego rellena otro sistema destino con esos datos, en su forma más primigenia es una forma de programa ETL.

Como ejemplifica Villanueva (2011), la problemática principal de las bodegas de datos surgió desde el momento en que múltiples medios de almacenamiento estuvieron a disposición dentro de las organizaciones. Bajo los primeros esquemas de almacenamiento como las cintas magnéticas, se tenían problemas como la sincronización de archivos y la complejidad de mantener aplicaciones desarrolladas y de desarrollar nuevas. Al surgir los actuales medios de almacenamiento como los dispositivos de almacenamiento de acceso directo, se dio pie al desarrollo de los sistemas gestores de bases de datos. Esto dio lugar al paradigma «una sola fuente de datos para todo tipo de procesamiento (transaccional y analíticos)», según la investigación sobre Inmon.

De acuerdo a la misma investigación (William, 1999), a lo largo de la evolución de los sistemas de almacenamiento presentada en la Figura 5, un sin fin de fuentes de datos fueron desarrolladas sobre los esquemas de datos iniciales y posterior a la división de los sistemas en OLTP[7] (Online Transaction Processing) y OLAP (Online Analytics Processing) se han desarrollado nuevos sistemas bajo los esquemas actuales. La evolución de los sistemas de almacenamiento, y muchos de ellos orientados a apoyar la toma de decisiones (DSS) se puede ver en la figura anterior, difundida por Inmon (William, 1999).

7 Un sistema OLTP también se conoce como sistema transaccional, debido a que sirve para almacenar las operaciones de una aplicación operativa de una compañía.

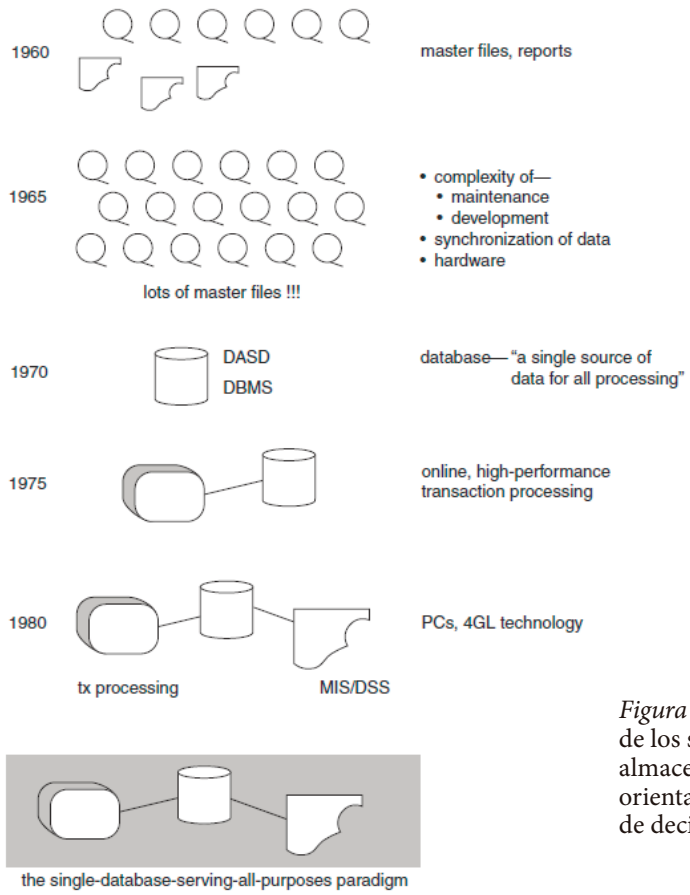


Figura 5. Evolución de los sistemas de almacenamiento orientado a la toma de decisiones (DSS).

Esto produjo el surgimiento del desarrollo de aplicaciones de extracción (a las que en años posteriores se sub-dividieron, además, en transformación y carga), con el objetivo de alimentar sistemas OLAP con datos provenientes de sistemas OLTP. Este programa extractor, como también se le denomina, busca y analiza a través de una hoja de cálculo, un archivo o una base de datos, criterios de selección de datos y su transporte a otro sistema igual o con características diferentes. Es innegable el hecho de que los datos son dinámicos y tienen vida propia (en el sentido de su adaptación y su maleabilidad ante las condiciones y usos) dentro de un sistema transaccional (OLTP), además, el uso de programas extractores enfrenta serios desafíos: en primer lugar, falta de credibilidad de los datos, por la multitud de fuentes no sincronizadas o

consolidadas, por la variedad de niveles de detalle (granularidad) o porque la variación en el tiempo en el que es extraída no está definida: en segundo lugar, la disminución en productividad, especialmente en grandes empresas, que poseen numerosos sistemas, haciendo complejo localizar los datos para analizar o entregar los datos al programador o analista de recursos de manera fácil y rápida; y en tercer y último lugar, por la imposibilidad de transformar los datos en información, cuando los datos tienen nombres diversos entre las distintas fuentes. Los cambios de concepción de las bodegas de datos han dado vida independiente a los procesos ETL.

Características de los procesos ETL

Como se ha referido anteriormente, el proceso ETL es el fundamento de una Bodega de Datos (Data Warehouse), pero también es una filosofía de manejo de datos más allá del término puro. Un proceso ETL diseñado y llevado a cabo de forma apropiada, extrae datos de las fuentes de información, refuerza la calidad y consistencia de los mismos y finalmente entrega los datos en una presentación y formato listo para ser consumidos por aplicaciones para la toma de decisiones.

El proceso ETL agrega un valor significativo a los datos. Éste va más allá de la transportación de los datos de las fuentes orígenes a la carga dentro de la bodega de datos. En específico el proceso ETL se encarga de:

- Remover errores y corregir datos faltantes.
- Proporcionar medidas documentadas de la calidad de los datos.
- Supervisar el flujo de los datos transaccionales.
- Ajustar y transformar los datos de múltiples fuentes para poder unificarlos.
- Estructurar los datos para ser usados por las herramientas y usuarios finales.

El proceso ETL es intuitivo y fácil de comprender. La idea básica del proceso ETL es: tomar los datos de las fuentes de información y depositarla en un destino, generalmente asociado a una bodega de datos; sin embargo, la limpieza y transformación de la información son procesos mucho más complicados de lo que se puede apreciar a simple vista. De hecho, estos procesos generales

suelen dividirse en un sin fin de tareas específicas, dependiendo de las características de las fuentes de datos, las reglas de negocios, las herramientas existentes y las características de la bodega de datos final.

El diseño y alcances del proceso ETL están siempre en función de las necesidades de negocio, es decir, los propósitos específicos por los que se quiere consolidar la información dentro del Data Warehouse. Puesto que un Data Warehouse es para propósitos meramente de análisis, se debe tener en cuenta el tipo de información que se quiere obtener, y con base a ello determinar los elementos y consideraciones de las fuentes de información que estarán involucradas en el análisis.

Procesos de Extracción y Carga

El reto para un correcto desarrollo del proceso ETL es planificar adecuadamente la gran cantidad de tareas, para lo cual es indispensable conservar la perspectiva simple e intuitiva de la misión del proceso.

Resulta indispensable antes de comenzar con el proceso ETL, comprender las necesidades de las personas involucradas en el análisis que se realizará sobre la bodega de datos, principalmente para saber qué información va a formar parte del proceso, qué transformaciones se tendrán que hacer sobre la misma y en qué forma será consultada y entregada la información final.

Una vez conocidas las necesidades que dan pie al proceso ETL, se puede entonces determinar los elementos de las fuentes de información que participarán en el proceso, los mecanismos de acceso y extracción de los datos, las transformaciones y políticas de negocio que habrá que aplicar sobre los mismos y determinar así la forma, presentación y estructura en que finalmente serán almacenados para su consumo final.

Otros aspectos a considerar posteriores a la realización del proceso ETL son:

- Grado de integración de los datos: cuál será el volumen o unidad mínima.
- Información sobre la cual se harán las cargas de información.
- Latencia de los datos: con qué frecuencia se harán adiciones de infor-

mación al Data Warehouse.
- Requerimientos de seguridad: cuales son las políticas de seguridad que se aplicarán para la carga de información.
- Perfil de los datos: Tipo y naturaleza de las fuentes de información que intervienen.
- Linaje (heredado) y archivado: Cuáles mecanismos se usarán para dar seguimiento a los datos y sus transformaciones, desde su sistema origen hasta su destino final dentro del Data Warehouse.

La arquitectura general sobre la cual será desarrollado el proceso ETL es un elemento primordial para garantizar el éxito de la implementación. Un mal diseño y elaboración de la arquitectura involucraría la nueva implementación total del proceso. Algunos aspectos a considerar para el desarrollo de la arquitectura del proceso son los siguientes:

- La arquitectura se implementará sobre una herramienta ETL o se desarrollará codificando los módulos del proceso ETL que sean requeridos de forma manual.
- Elaborar la implementación sobre una herramienta ETL hace el desarrollo más rápido, pero requiere de un perfecto entendimiento de la política de negocio y de los objetivos que se persiguen.
- De manera adicional, resuelve problemas tediosos que se tendrían en una implementación manual. Entre estos podemos citar: la gestión de meta-datos, la sincronización de los repositorios de información y los mapeos entre fuentes de datos origen y destino. Por el contrario, si se decide realizar la codificación de forma manual, se puede tener un control más específico sobre las transformaciones y los meta-datos.
- La carga de información será a través de procesos batch o sobre un flujo de datos. La arquitectura estándar del proceso ETL está basada en cargas batch de información periódicas, éstas suelen ser lentas debido al gran volumen de información que tienen que transportar. Sin embargo, si en la práctica se requiere que un Data Warehouse sea actualizado de forma constante y rápida, el esquema de actualización por batch no es funcional y se puede recurrir a la carga por flujo de datos constante. Dado que la naturaleza de estos dos esquemas son muy diferentes, su impacto en la arquitectura general del proceso es muy grande, puesto que todas las rutinas de extracción, limpieza, integración y entregas se implementarían de maneras radicalmente opuestas.

- La dependencia entre tareas será vertical u horizontal. Si se opta por hacer que las tareas sean independientes sobre un flujo de trabajo horizontal, implica el hecho de que los datos de dos fuentes de información pueden ser procesados de forma independiente. Por el contrario, si se hacen tareas independientes sobre un flujo vertical, los trabajos de extracción, limpieza y carga de las diversas fuentes de información estarán sincronizados y la carga de información se hará de manera simultánea.

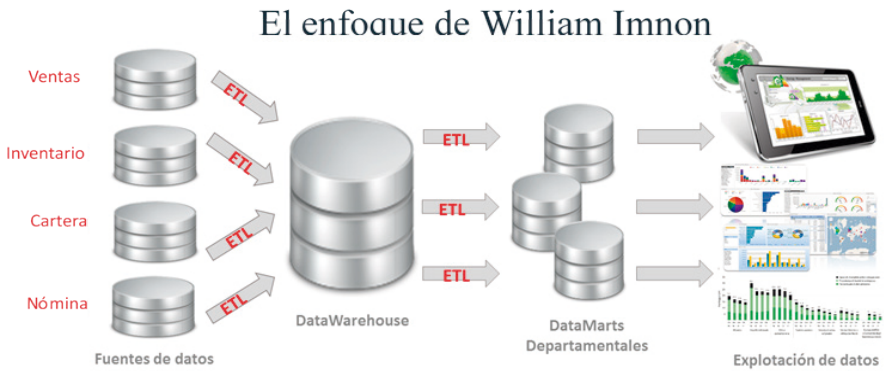


Figura 6. Creación de un Almacén de datos Operacional según Inmon.

La carga de información será a través de procesos batch o sobre un flujo de datos. La arquitectura estándar del proceso ETL está basada en cargas batch de información periódicas, éstas suelen ser lentas debido al gran volumen de información que tienen que transportar. Sin embargo, si en la práctica se requiere que un Data Warehouse sea actualizado de forma constante y rápida, el esquema de actualización por batch no es funcional y se puede recurrir a la carga por flujo de datos constante. Dado que la naturaleza de estos dos esquemas son muy diferentes, su impacto en la arquitectura general del proceso es muy grande, puesto que todas rutinas de extracción, limpieza, integración y entregas se implementarían de maneras radicalmente opuestas.

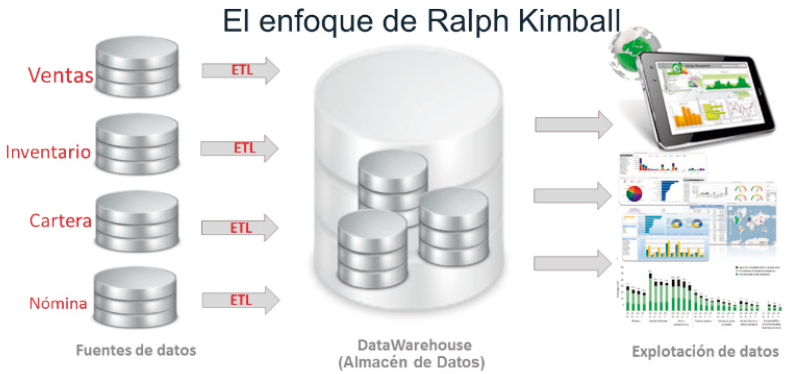


Figura 7. Conjunto de herramientas para un Data Warehouse, según Ralph Kimball

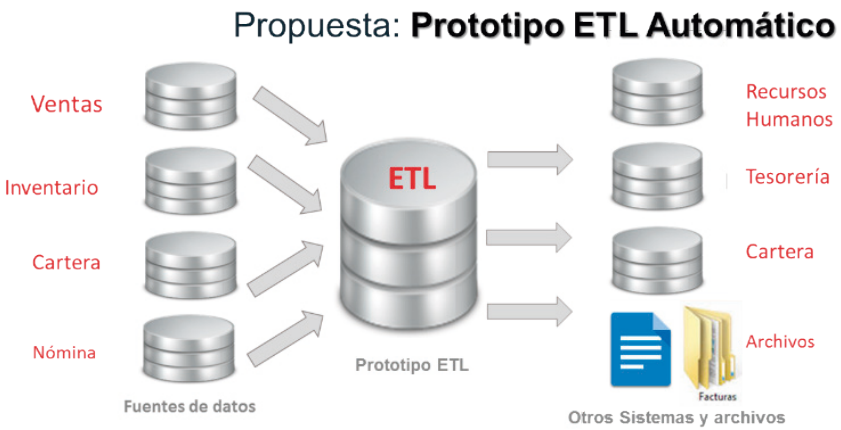


Figura 8. Propuesta: ETL Automático para uso general

El proceso ETL involucra la interacción con diversas estructuras de datos que se leen y escriben en diversos dispositivos de almacenamiento, por tal motivo revisamos las estructuras de datos más frecuentes que se presentan a lo largo del proceso. Revisaremos aspectos puntuales como: estructura, métodos de extracción e inserción.

- Archivos de texto plano, con una estructura definida.
- Archivos XML
- Base de datos relacionales

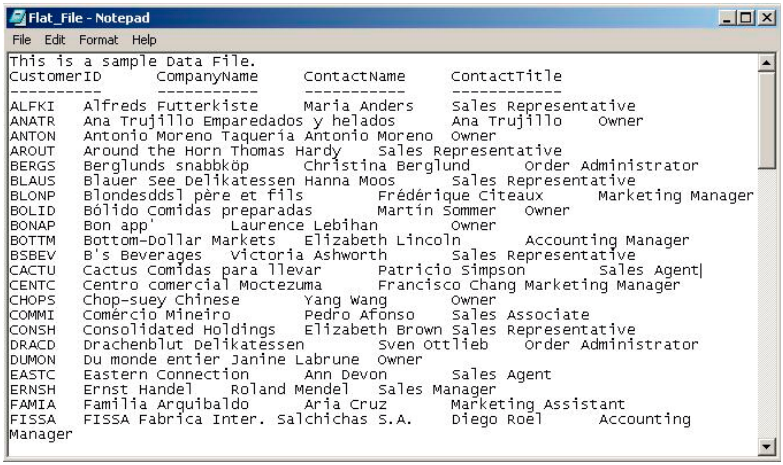


Figura 9. Muestra de un archivo plano listo para cargar.

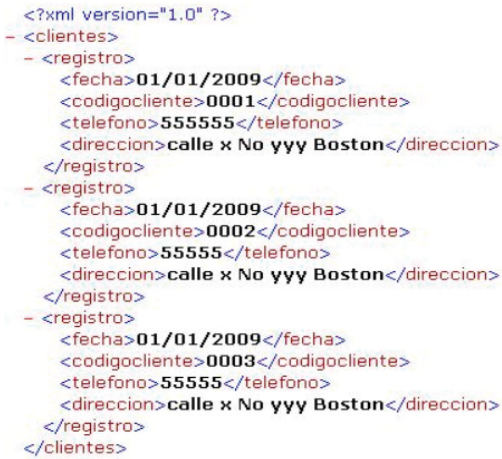


Figura 10. Muestra de un archivo XML listo para cargar

Proceso de transformación

Estos son algunos ejemplos de transformaciones que se deben hacer a los datos que se extraen de algunos sistemas de información de origen:

- Seleccionar sólo ciertas columnas para su carga (por ejemplo, que las columnas con valores nulos no se carguen).
- Traducir códigos (por ejemplo, si la fuente almacena una “H” para Hombre y “M” para Mujer, pero el destino tiene que guardar “1” para

Hombre y "2" para Mujer).

- Codificar valores libres (por ejemplo, convertir "Hombre" en "H" o "Sr" en "1").
- Obtener nuevos valores calculados (por ejemplo, total venta = cantidad * precio).
- Unir datos de múltiples fuentes (búsquedas, combinaciones, entre otras).
- Calcular totales de múltiples filas de datos (por ejemplo, ventas totales de cada región). (Kimball & Caserta, 2004).

El proceso de transformación incluye los siguientes pasos:

1. Estudio del sistema origen
2. Ubicación donde ocurrirán las transformaciones
3. Validación de (selección) los datos extraídos
   a. Detección de deltas (variaciones en los datos)
   b. Frecuencia de deltas
   c. Comprobación y conteo de registros
   d. Captura de datos erróneos
4. Mapeo (relación de campos de origen y destino)
   a. Preparación de documento de mapeo
5. Reformateo de datos
   a. Conversión de tipos y longitudes
6. Limpieza de datos
   a. Comprobación de la calidad de los datos
7. División, unión y combinación de datos
8. Integración y cálculo de campos
   a. Resumen y totales calculados
9. Mejoramiento de los datos
10. Aplicación de reglas de negocios.

Todos estos pasos se realizan en una zona segura llamada *staging*, con el objetivo de dejar trazabilidad y permitir la auditoria de la información procesada.

Avances de la propuesta

El prototipo de este capítulo usa algunas características propias de SQL Server 2012: El asistente de importación y exportación. Este componente lla-

mado desde una aplicación web permite extraer y cargar de forma rápida y fácil (Figura 12).

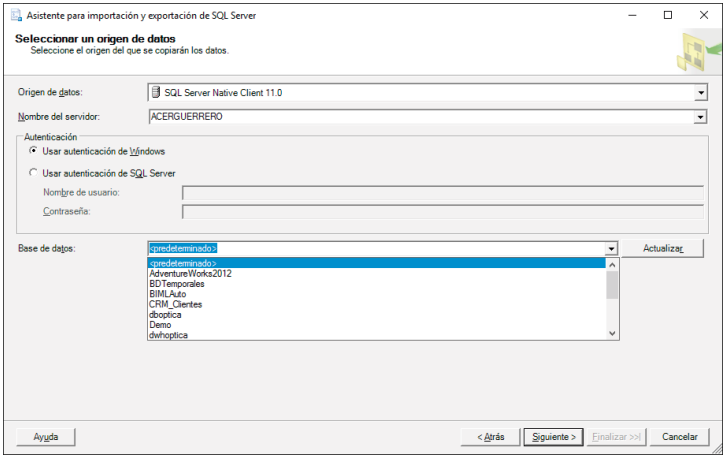


Figura 12. Proceso de extracción con la conexión a la fuente de origen.

Durante este paso de extracción se produce la conexión a la fuente de datos, que incluye el servidor y la base de datos (o la ruta donde se encuentra el archivo, en caso de archivos planos). En segundo lugar, se selecciona las tablas de origen y destino, como se aprecia en la Figura 13. Las estructuras de destino, generalmente son ubicadas en un área llamada Staging (o área intermedia y temporal).

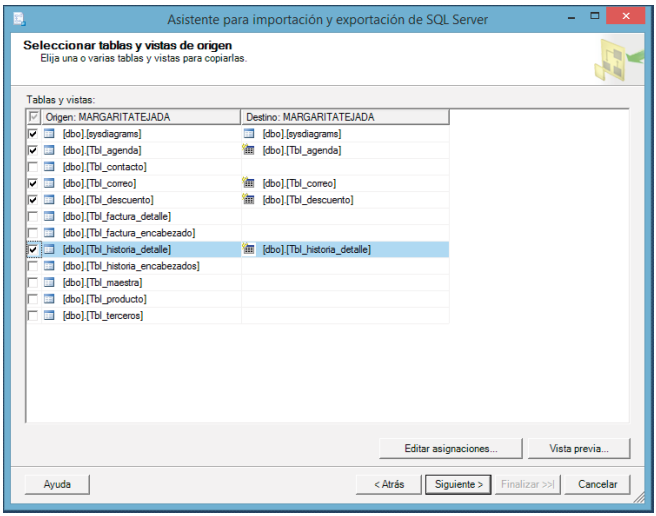


Figura 13. Proceso de Extracción con la definición de los campos

En la Figura 14, se mapea cada campo origen con la estructura de destino, donde se revisa los tipos de datos, nombres y tamaños básicos. En la metodología Kimball se propone que se lleve al área *staging* sin ninguna modificación, con el objetivo de facilitar el proceso de transformación.

Hasta la entrega del capítulo, se ha elaborado el 50% del desarrollo, incluyendo temas de Extracción y Carga. El proceso de transformación se encuentra en las etapas de especificación de requisitos y diseño.

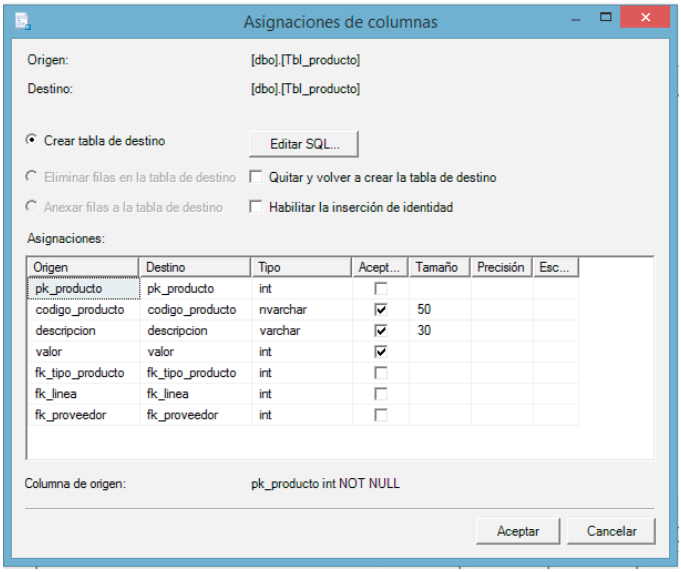


Figura 14. Refleja la manera adecuada de intercambiar información: datos limpios, con calidad y confiables.

Conclusiones

La metodología de Kimball proporciona una base empírica y metodológica adecuada para las implementaciones de almacenes de datos, dada su gran versatilidad y su enfoque ascendente, que permite construir modelos de transferencia de datos entre sistemas en forma escalonada. Un proceso ETL bien diseñado y acoplado brinda una estrategia segura y confiable de movimiento de datos que garantiza que todos los sistemas se surtan de la información necesaria para su funcionamiento. La propuesta de este artículo ha buscado la manera fácil de cumplir este propósito, al construir una aplicación que oculte

la complejidad de un proceso ETL y brinde una visión panorámica del intercambio de información simple, dejando al usuario un flujo sencillo de seguir.

Referencias

Castillo , R., Morata, J., & del Arbol , L. (2015). *Operational Data Store (ODS)*. España. Recuperado de: http://www.lsi.us.es/redmidas/CEDI/papers/933.pdf

Kimball, R., & Caserta, J. (2004). *The Data Warehouse ETL Toolkit*. Indianapolis.: Wiley Publishing, Inc. Recuperado de: http://users.itk.ppke.hu/~szoer/DW/Kimball%20&%20Caserta%20-The%20Data%20Warehouse%20ETL%20Toolkit%20%5BWiley%202004%5D.pdf

Kimball, R., & Ross, M. (2002). *The Data Warehouse Toolkit*. John Wiley & Sons, Inc. Recuperado de: http://users.itk.ppke.hu/~szoer/DW/Kimball%20&%20Ross%20-%20The%20Data%20Warehouse%20Toolkit%202nd%20Ed%20%5BWiley%202002%5D.pdf

Villanueva, J. (2011.). Marco de trabajo basado en ontologías para el proceso ETL.En: J. Villanueva. *Marco de trabajo basado en ontologías para el proceso ETL*. México, D.F.: Centro de Investigación y de Etudios Avanzados del Instituto Politécnico Nacional. pp.17 - 47.

William, I. (1999). *Building the Data Warehouse*. John Wiley & Sons, Inc. Recuperado de: http://fit.hcmute.edu.vn/Resources/Docs/SubDomain/fit/ThayTuan/DataWH/Bulding%20the%20Data%20Warehouse%204%20Edition.pdf

Implicaciones de la neutralidad de la red y el diseño para la inclusión en las políticas del Sistema Nacional de Competitividad, Ciencia, Tecnología e Innovación en Colombia

David Alberto García Arango[1]; Jovany Sepúlveda Aguirre[2];
César Felipe Henao Villa[3]; Elkin Darío Aguirre Mesa[4];
Gustavo Andrés Araque González[5]; Laura Isabel Bedoya Corrales[5]

Resumen

La ejecución de políticas de desarrollo social y económico en Colombia, pasa por diversos estados de concepción, diseño e implementación que se reconfiguran según la dinámica de institucionalidad y gobernabilidad en el marco de la intervención de las Tecnologías de Información y la Comunicación en el fenómeno de la globalización. Las Instituciones de Educación Superior, se ven inmersas en el amplio entramado de conexión Universidad-Empresa-Estado que busca relacionar ciencia, competitividad, tecnolo-

1 Docente-Investigador del grupo AGLAIA - Corporación Universitaria Americana. Licenciado en Matemáticas y Física de la Universidad de Antioquia, Magíster en Matemáticas Aplicadas de la Universidad EAFIT. Escuela de Ciencias, departamento de ciencias Matemáticas, doctorando en Educación de la Universidad Nacional de Rosario – Argentina. Correo electrónico de contacto: dagarcia@coruniamericana.edu.co. ORCID: https://orcid.org/0000-0002-0031-4275

2 Magister en Gestión de la Innovación Tecnológica, Cooperación y Desarrollo Regional. Investigador Junior integrante del Grupo de Investigación AGLAIA de la Corporación Universitaria Americana. ORCID: https://orcid.org/0000-0002-1047-6673. E-mail: jasepulveda@americana.edu.co.

3 Docente-Investigador del grupo AGLAIA - Corporación Universitaria Americana. Ingeniero de Sistemas de la Universidad Nacional de Colombia Sede Medellín, magíster en entornos virtuales de aprendizaje. Correo electrónico: chenao@coruniamericana.edu.co. ORCID: https://orcid.org/0000-0001-7426-2589

4 Docente- Institución Universitaria Pascual Bravo. Ingeniero de sistemas de la Fundación Universitaria María Cano, Magíster en Gestión de la Tecnología Educativa de la Universidad de Santander. Correo electrónico: elkin.aguirre@pascualbravo.edu.co. ORCID: https://orcid.org/0000-0003-2521-6003

5 Docente-Investigador Corporación Universitaria Americana. Ingeniero Industrial, con especialización en Gestión Logística Integral, magíster en Ingeniería de Producción con énfasis en transporte y logística. Correo electrónico: garaque@americana.edu.co. ORCID: https://orcid.org/0000-0001-8627-8924.

5 Ingeniera Mecánica, Magister en Ingeniería Agroindustrial. Docente de la Corporación Universitaria Americana. Correo electrónico: libedoya@americana.edu.co

gía e innovación sin dejar de lado las exigencias del medio respecto a políticas de acceso, accesibilidad y adaptabilidad a las Tecnologías de la Información y la Comunicación. Los esfuerzos de esta conexión, se materializan en el Sistema Nacional de Ciencia, Competitividad, Tecnología e Innovación, propuesto por el Consejo Nacional de Política Económica y Social y que, para efectos de las Instituciones de Educación Superior, se aplican en las funciones de docencia, investigación y extensión. Esta investigación hizo una mirada de este sistema aplicado a las Instituciones de Educación Superior considerando la óptica de los conceptos de neutralidad de la red y diseño para la inclusión. La metodología de aproximación al objeto de estudio fue desde un razonamiento abductivo con enfoque mixto donde se contextualizaron las unidades de análisis y se identificaron aspectos de eficiencia en la Ciencia, Tecnología e Innovación en Colombia desde la producción técnica y tecnológica de los grupos de investigación. Como resultado, se proponen rutas de intervención respecto a la forma en que las Instituciones de Educación Superior aportan al sistema desde las ópticas anteriormente especificadas.

Palabras clave: neutralidad de la red, diseño para todos, instituciones de educación superior, gobernanza del internet.

Implications of the neutrality of the network and the design for the inclusion in the policies of the National System of Competitiveness, Science, Technology and Innovation in Colombia

Abstract

The execution of social and economic development policies in Colombia, passes through various states of conception, design and implementation that are reconfigured according to the dynamics of institutionality and governance in the framework of the intervention of Information and Communication Technologies in the phenomenon of the globalization. The Higher Education Institutions are immersed in the broad connection framework University-Company-State that seeks to relate science, competitiveness, technology and innovation without neglecting the demands of the environment regarding

policies of access, accessibility and adaptability to Technologies of Information and Communication. The efforts of this connection are materialized in the National System of Science, Competitiveness, Technology and Innovation, proposed by the National Council of Economic and Social Policy and that, for the purposes of the Institutions of Higher Education, are applied in the teaching functions , research and extension. This investigation made a look of this system applied to Higher Education Institutions considering the optics of the concepts of net neutrality and design for inclusion. The methodology of approach to the object of study was from an abductive reasoning with a mixed approach where the units of analysis were contextualized and aspects of efficiency in Science, Technology and Innovation in Colombia were identified from the technical and technological production of the research groups. As a result, intervention routes are proposed regarding the way in which Higher Education Institutions contribute to the system from the previously specified optics.

Key words: net neutrality, design for all, higher education institutions, internet governance.

Introducción

Las Instituciones de Educación Superior (IES), son organizaciones que tienen por principal activo el conocimiento, el cual se dinamiza a través del relacionamiento de sus actores con las políticas de enseñanza y aprendizaje en el marco de la solución de problemáticas en un contexto social. Entre los cambios que supone los retos asociados al auge de la tecnología y la consecuente proliferación de la información, el rol de las instituciones educativas pasa a ser a organizaciones "no sólo que deben aprender, sino que deben también generar y gestionar conocimientos en este escenario" (Domínguez, 2001, p.497). El área de influencia de las IES va actualmente mas allá de las aulas de clase, en el sentido que se tiene una Responsabilidad Social Universitaria (RSU) que debe ir más allá de las políticas de autonomía universitaria propuestas por la Ley 30 de 1992. Las IES, concebidas desde esta RSU, deben pasar de adaptadoras a transformadoras donde "Formación, investigación, liderazgo social y compromiso son los elementos sustantivos que determinan el formato de esta relación para hacer efectiva su incidencia social" (Beltrán, Íñigo, y Mata, 2014, p.16).

Tomando como base las consideraciones anteriores, vale la pena interpretar el papel de las IES en el Sistema Nacional de Competitividad, Ciencia, Tecnología e Innovación en Colombia. Es menester identificar cómo las políticas de Ciencia, Tecnología e Innovación, desde la perspectiva estratégica del gobierno nacional, se han enfocado gradualmente en intervenir la productividad de las organizaciones y así mismo medir la forma en que tal intervención influye en el desarrollo de las regiones y en el fortalecimiento de las políticas de investigación para el fortalecimiento del emprendimiento.

En este presente escrito, se describen los aspectos relacionados con una investigación encaminada a identificar las implicaciones del concepto de neutralidad de la red y las políticas de diseño para la inclusión en el desarrollo del Sistema Nacional de Competitividad, Ciencia, Tecnología e Innovación (SNCCTI) en Colombia.

El estudio se realizó con una metodología con enfoque mixto, donde mediante el análisis de informes y estudios técnicos, se midió el índice de efectividad para los grupos reconocidos por COLCIENCIAS respecto al informe generado en el año 2016 acerca del estado de la ciencia en Colombia para la convocatoria 693 de 2014 de medición de grupos de investigación. Con base en los resultados obtenidos, se realizaron interpretaciones en el marco de la discursividad y la hermenéutica de la forma en que los conceptos de neutralidad de la red y el diseño para la inclusión, proponen líneas de trabajo para las IES en la búsqueda del fortalecimiento del SNCCTI.

El escrito se desarrolla en las siguientes etapas: Presentación de aspectos relacionados con el SNCCTI, conceptualizaciones respecto a la neutralidad en la Red y el diseño para la inclusión, aspectos metodológicos de la investigación, resultados de la medición de la eficiencia para grupos de investigación, implicaciones para las IES y conclusiones.

Como resultado del proceso investigativo, se resalta la importancia de que las IES promuevan el desarrollo y posterior acceso a productos de investigación con enfoque de inclusión y neutralidad en el acceso para los sectores sociales más vulnerables de Colombia.

Contextualización

A continuación, se presentarán los principios del SNCCTI y conceptualizaciones generales en relación a la neutralidad en la Red y el diseño para la inclusión.

El Sistema Nacional de Competitividad, Ciencia, Tecnología e Innovación

El SNCCTI, se crea en el artículo 186 de la Ley 1753 de 2015 en la cual se expide el Plan Nacional de Desarrollo 2014-2018. En este se establece "Intégrese el Sistema de Competitividad e Innovación con el Sistema de Ciencia, Tecnología e Innovación para consolidar un único Sistema de Competitividad, Ciencia, Tecnología e Innovación". El sistema está integrado con la finalidad de garantizar la participación conjunta de instancias de investigación, emprendimiento y comités de articulación Universidad-Empresa-Estado. El sistema aún no está reglamentado, pero ya está en funcionamiento desde la Consejería Presidencial del Sistema Nacional de Competitividad e Innovación. Específicamente, en el eje de Ciencia, Tecnología e Innovación, el sistema tiene como proyectos estratégicos el diseño e implementación de la Política de Ciencia, Tecnología e Innovación de largo plazo, diseño e implementación del portal de innovación, estructuración de Colombia Bio, lineamientos de política para estimular la inversión privada en Ciencia, Tecnología e Innovación a través de deducciones tributarias, piloto de calificación automática para empresas altamente innovadoras y parques en beneficios tributarios para CTI y diseño e implementación de Programa Ecosistema Científico.

Neutralidad en la Red y diseño para la inclusión

En Colombia, la Comisión de Regulación de Comunicaciones (CRC), establece en el artículo 56 de la Ley 1450 de 2011 las condiciones regulatorias para la neutralidad en internet estableciendo los principios de:

1. Libre elección (por parte de los usuarios);
2. No discriminación (por parte de proveedores de acceso y servicios);
3. Transparencia respecto de políticas de gestión de tráfico; y
4. Información asociada a las condiciones de prestación del servicio de acceso a Internet (Subsecretaría de Telecomunicaciones, 2015).

La neutralidad de la red ha sido un tema de estudio reciente, en tanto que se han desarrollado determinaciones para pasar de una red neutral en su acceso a una red empaquetada en la cual se paga según la cantidad de banda ancha que consuma la aplicación. Respecto a la regulación en materia internacional, en Estados Unidos, se pone en manos de la Comisión Federal de Comunicaciones (FCC por sus siglas en inglés), la responsabilidad de verificar que efectivamente se lleven a cabo los principios de neutralidad en la red, para tal efecto se creó un cargo denominado ombudsperson que "servirá como un punto de contacto para asuntos de internet abierto en la Comisión para ayudar a los consumidores y proveedores de contenido sus quejas y requerimientos a las partes interesadas" (Federal Communications Commission, 2015).

Respecto al desempeño de esta política, se han identificado falencias respecto a la atención en las quejas, pero las denuncias de estas falencias no se han reconocido por parte del gobierno, insistiendo que es necesaria la regulación de internet tanto por motivos de calidad como por motivos económicos. Las quejas van desde deficiencias en el servicio de internet hasta el corte intencionado del servicio para favorecer a las compañías o sectores que pagan más por el servicio.

Aunque en general se considere que la red debe ser neutral, se han venido implementando algunos tratamientos diferentes con aplicaciones que podrían poner en peligro este concepto o que al menos podrían en entredicho el concepto. Para tal efecto se puede citar el caso relacionado con Netflix en Estados Unidos respecto al pago que esta compañía debió hacerle a las operadoras de red para que pasen sus contenidos de alta definición o considerar por ejemplo la forma en que las compañías telefónicas promueven el uso gratuito diferenciado de WhatsApp y Facebook siempre que se utilicen los planes de pago que proponen.

En cuanto al diseño para todos, se define como: "La intervención sobre entornos, productos y servicios con el fin de que todos, incluidas las generaciones futuras, independientemente de la edad, el sexo, el género, las capacidades o el bagaje cultural, puedan disfrutar participando en la construcción de nuestra sociedad" (Instituto de Mayores y Servicios Sociales, 2006)

Vale la pena considerar que las IES tienen el compromiso de establecer en su currículo el diseño para todos en tanto que la inclusión de los sectores

sociales marginados, menos favorecidos y vulnerables, se constituye en uno de los Objetivos para el Desarrollo Sostenibles propuestos por las Naciones Unidas.

Es así como la convergencia de políticas regulatorias de la neutralidad de la red, al ser orientadas desde el diseño para la inclusión, posibilitan un marco regulatorio de las Over-the-Top Technologies, (tecnologías que se sirven de la red para ofrecer sus servicios) que interviene positivamente en fortalecimiento de las relaciones que se pretenden con la implementación del SNCCTI. Respecto al diseño para todos, puede afirmarse que surge como un resultado de la reflexión sobre el desarrollo de la accesibilidad, que es un medio determinante para la integración de los ciudadanos independientemente de la edad, el género, sus capacidades. En los últimos años, el concepto de "Diseño Para Todos" se ha ido extendiendo por los diferentes países europeos. En este orden de ideas, vale la pena repensar el papel de la universidad en la formación de profesionales con compromiso social y mentalidad empresarial en proyectos en ciencia, tecnología e innovación.

Metodología

La búsqueda y posterior identificación de las implicaciones prácticas fue desarrollada mediante una lógica de razonamiento abductivo, donde la delimitación de las unidades de análisis se realizó por medio de los resultados obtenidos del estudio de las características del *habitus* institucional de las IES, se llevó a cabo el estudio de los factores de calidad propuestos por el Comité Nacional de Acreditación (CNA), relacionándolos con los indicadores propuestos por el SNCCTI. Mediante un enfoque mixto de tratamiento de los datos y el uso de técnicas de análisis discursivo, se propuso desde un nivel paradigmático pragmático, una postura epistemológica que posibilitó la intervención de procesos hermenéuticos de recolección de información del quehacer de las IES mediante encuestas de valoración y entrevistas a profundidad a los actores implicados en los procesos áulicos que son inevitablemente intervenidos por políticas regulatorias de neutralidad de la red. Para tal efecto, se identifican aspectos comunes en diversos diseños curriculares de programas de ingeniería comparándolos con las características del diseño para todos e identificando puntos críticos de los planes estratégicos de facultad que serían susceptibles a la formación en materia de políticas de neutralidad de la red.

Por lo anterior, el proceso de investigación, constó de las siguientes etapas: Identificación de las políticas en materia de ciencia, competitividad, tecnología e innovación en el marco del sistema de aseguramiento de la calidad educativa, relacionamiento de las políticas con el quehacer institucional universitario, identificación de elementos de diseño para todos en programas de ingeniería y finalmente el reconocimiento del carácter intervencionista de las *over-the-top technologies* en los diversos aspectos curriculares de la vida universitaria (se utilizaron herramientas y técnicas para detectar y caracterizar prácticas no neutrales en un diseño experimental de corte transversal). Todo lo anterior transversalizado por las implicaciones que a nivel práctico supone el cambio en materia de regulación en neutralidad de la red tanto a nivel nacional como internacional.

Resultados de la medición de la eficiencia para grupos de investigación

Con base en el informe del estado de la ciencia en Colombia, planteado por COLCIENCIAS, se analizó la relación entre el porcentaje del PIB de inversión en Actividades de Ciencia, Tecnología e Innovación (ACTI), porcentaje del PIB inversión en I+D e investigadores reconocidos con la producción científica y tecnológica y en especial con los productos de desarrollo tecnológico e innovación. Un estudio similar se realizó para estudiar el impacto de la cualificación docente y la inversión en educación en los resultados de las pruebas saber en el departamento de Bolívar. El estudio realizado por Maza-Ávila, Quesada-Ibargüen, y Vergara-Schmalbach (2013), obtuvo el nivel de eficiencia de los recursos con base en los resultados de un Análisis Envolvente de Datos (DEA) por sus siglas en inglés, y generó un índice de Malmquist (Färe, Grosskopf, Norris, y Zhang, 1994) para verificar la evolución temporal de la productividad de la calidad educativa.

Debido a que solo se tiene un informe generado por COLCIENCIAS, para obtener un porcentaje de relación, se utilizó únicamente el DEA, el cual utilizó el modelo CCR con orientación al input. Información respecto al modelo, puede hallarse en Cooper, Seiford, y Tone (2007). Los datos obtenidos para la aplicación del modelo se presentan en la tabla 1:

Tabla 1.
Datos porcentuales obtenidos para cada departamento según el informe del estado de la Ciencia en Colombia generado por COLCIENCIAS.

Departamento	Inversión ACTI {I}	Inversión ID {I}	Investigadores Reconocidos {I}	Producción {O}	Desarrollo TeI {O}
Antioquia	0.207	0.252	0.2068	0.2097748	0.00713234
Amazonas	0.0028	0.0051	0.0008	0.00089049	0
Arauca	0.0008	0.0015	0.0001	7.9079E-05	0
Atlántico	0.0129	0.0006	0.0429	0.04315627	0.00332303
Bolívar	0.0147	0.0182	0.0274	0.03046931	0.0011883
Boyacá	0.0026	0.0019	0.0168	0.01402441	0.00029451
Caldas	0.0289	0.0436	0.036	0.03682654	0.00099432
Caquetá	0.0008	0.0013	0.0018	0.00181537	3.4492E-05
Casanare	0.0002	0.00001	0.0004	0.00017535	1.7184E-05
Cauca	0.0105	0.0094	0.011	0.00967509	9.6751E-05
Cesar	0.0007	0.0008	0.004	0.00437339	0.00010933
Chocó	0.0092	0.0015	0.002	0.00168128	7.5658E-05
Córdoba	0.0027	0.0028	0.012	0.00964071	0.00013497
Cundinamarca	0.0348	0.0481	0.016	0.01635207	0.00058867
Bogotá	0.549	0.483	0.36	0.36873646	0.01032462
Guainía	0.0004	0.0002	0.0001	8.9393E-05	6.8833E-06
Huila	0.0022	0.0028	0.0068	0.00550799	7.7112E-05
La Guajira	0.0016	0.0015	0.016	0.00245487	7.1191E-05
Magdalena	0.0056	0.007	0.01	0.00917311	0.00022015
Meta	0.0017	0.001	0.0046	0.00422898	5.0748E-05
Nariño	0.0079	0.0073	0.0114	0.01137012	0.00015918
Norte de Santander	0.0032	0.0031	0.013	0.01088877	0.00042466
Putumayo	0.0004	0.0003	0.0001	5.1573E-05	0
Quindío	0.0031	0.0042	0.013	0.00734399	0.00021298
Risaralda	0.005	0.0044	0.0176	0.02036101	0.00081444
San Andrés y Providencia	0.0027	0.0024	0.0005	0.00057074	0
Santander	0.0172	0.0178	0.0471	0.0522503	0.00235126
Sucre	0.0003	0.0004	0.0042	0.00344164	5.1625E-05
Tolima	0.0031	0.0006	0.0136	0.01181709	0.00028361
Valle del Cauca	0.0734	0.0675	0.0973	0.09437855	0.03850645

Fuente: elaboración propia.

Para efectos de estandarización, se pasaron los valores porcentuales a su respectivo equivalente decimal. Los valores {I} representan las entradas del sistema y los valores {O}, las salidas. Se realizaron dos aplicaciones del mode-

lo, variando las salidas para calcular el nivel de eficiencia respecto a los recursos destinados. Se exceptuaron los datos relacionados con los departamentos de Vaupés y Vichada puesto que no registraron insumos o inputs desde la inversión en CTeI.

La aplicación del modelo a los datos, se llevó a cabo con el software Efficiency Measurement System (EMS) desarrollado por Holger (2000). Los resultados obtenidos para la eficiencia, fueron los presentados en las figuras 1 y 2.

De los resultados obtenidos, puede inferirse que Atlántico, Cesar, Risaralda y Sucre tienen más eficiencia en el sentido en que a pesar de que no tienen un apoyo significativo desde la inversión en PIB para sus proyectos, generan una producción que supera esta inversión comparativamente con otros departamentos que tienen una mayor inversión. Si se desagrega la producción en la relacionada con Desarrollo Tecnológico e Innovación, entonces se tiene que también el Valle del Cauca genera una producción significativa y eficiente.

	DMU	Score	Invers {I}{V}	Invers {I}{V}	Inves {I}{V}	Produ {O}{V	Benchmarks	{S} Invers {I}	{S} Invers {I}	{S} Inves	{S} Produ
1	Antioquia	87,68%	0,00	0,00	1,00	1,00	25 (10,30)	0,13	0,18	0,00	0,00
2	Amazonas	96,22%	0,00	0,00	1,00	1,00	25 (0,04)	0,00	0,00	0,00	0,00
3	Arauca	68,36%	0,00	0,00	1,00	1,00	25 (0,00)	0,00	0,00	0,00	0,00
4	Atlántico	100,00%	0,00	1,00	0,00	1,00	5				
5	Bolívar	96,12%	0,00	0,00	1,00	1,00	25 (1,50)	0,01	0,01	0,00	0,00
6	Boyacá	87,80%	0,21	0,16	0,63	1,00	4 (0,07) 11 (1,24) 28 (1,59)	0,00	0,00	0,00	0,00
7	Caldas	88,42%	0,00	0,00	1,00	1,00	25 (1,81)	0,02	0,03	0,00	0,00
8	Caquetá	87,18%	0,00	0,00	1,00	1,00	25 (0,09)	0,00	0,00	0,00	0,00
9	Casanare	43,27%	0,00	0,02	0,98	1,00	4 (0,00) 25 (0,00)	0,00	0,00	0,00	0,00
10	Cauca	76,03%	0,00	0,00	1,00	1,00	25 (0,48)	0,01	0,01	0,00	0,00
11	Cesar	100,00%	0,28	0,00	0,72	1,00	6				
12	Chocó	72,66%	0,00	0,00	1,00	1,00	25 (0,08)	0,01	0,00	0,00	0,00
13	Córdoba	71,57%	0,12	0,00	0,88	1,00	11 (1,16) 25 (0,22)	0,00	0,00	0,00	0,00
14		88,34%	0,00	0,00	1,00	1,00	25 (0,80)	0,03	0,04	0,00	0,00
15	Bogotá	88,54%	0,00	0,00	1,00	1,00	25 (18,11)	0,40	0,35	0,00	0,00
16	Guainía	77,27%	0,00	0,00	1,00	1,00	25 (0,00)	0,00	0,00	0,00	0,00
17	Huila	70,02%	0,00	0,00	1,00	1,00	25 (0,27)	0,00	0,00	0,00	0,00
18	La Guajira	17,96%	0,15	0,15	0,70	1,00	4 (0,01) 11 (0,04) 28 (0,58)	0,00	0,00	0,00	0,00
19		79,29%	0,00	0,00	1,00	1,00	25 (0,45)	0,00	0,00	0,00	0,00
20	Meta	80,93%	0,00	0,12	0,88	1,00	4 (0,01) 25 (0,18)	0,00	0,00	0,00	0,00
21	Nariño	86,21%	0,00	0,00	1,00	1,00	25 (0,56)	0,00	0,00	0,00	0,00
22	Norte de	73,81%	0,13	0,00	0,87	1,00	11 (0,83) 25 (0,36)	0,00	0,00	0,00	0,00
23	Putumayo	44,58%	0,00	0,00	1,00	1,00	25 (0,00)	0,00	0,00	0,00	0,00
24	Quindío	49,98%	0,12	0,00	0,88	1,00	11 (0,68) 25 (0,21)	0,00	0,00	0,00	0,00
25	Risaralda	100,00%	0,00	0,02	0,98	1,00	23				
26	San	98,67%	0,00	0,00	1,00	1,00	25 (0,03)	0,00	0,00	0,00	0,00
27	Santander	95,89%	0,00	0,00	1,00	1,00	25 (2,57)	0,00	0,01	0,00	0,00
28	Sucre	100,00%	1,00	0,00	0,00	1,00	3				
29	Tolima	91,21%	0,31	0,06	0,63	1,00	4 (0,19) 11 (0,10) 28 (0,89)	0,00	0,00	0,00	0,00
30	Valle del	83,84%	0,00	0,00	1,00	1,00	25 (4,64)	0,04	0,04	0,00	0,00

Figura 1. Resultados obtenidos de eficiencia considerando la producción científica en su totalidad. **Fuente:** elaboración propia.

	DMU	Score	Invers {I}{V}	Invers {I}{V}	Inves {I}{V}	Produ {O}{V	Desa	Benchmarks	{F} Invers {I}	{F} Invers {I}	{F} Inves	{S} Produ	{S} Desar
1	Antioquia	65,12%	0,20	0,20	0,20	0,24	0,01	4 (0,84) 11 (39,66)	3,66%	2,79%	4,17%	0,00	0,00
2	Amazonas	51,95%	0,25	0,25	0,25	0,27	0,00	25 (0,04)	7,81%	3,77%	5,22%	0,00	0,00
3	Arauca	42,98%	0,25	0,25	0,25	0,18	0,00	25 (0,00)	2,43%	1,14%	3,36%	0,00	0,00
4	Atlántico	100,00%	3,47	0,72	32,12	26,72	9,59	14					
5	Bolívar	75,11%	0,20	0,20	0,57	0,68	0,04	4 (0,09) 11 (2,60) 25 (0,75)	5,66%	3,90%	3,00%	0,00	0,00
6	Boyacá	87,33%	0,20	0,20	0,41	0,68	0,00	4 (0,03) 28 (3,67)	3,34%	3,30%	3,00%	0,00	0,00
7	Caldas	65,97%	0,20	0,20	0,20	0,25	0,01	4 (0,03) 11 (8,10)	1,08%	4,90%	3,87%	0,00	0,00
8	Caquetá	70,82%	0,20	0,20	0,20	0,31	0,00	11 (0,42)	5,32%	3,54%	2,24%	0,00	0,00
9	Casanare	63,97%	0,20	0,20	0,20	0,00	0,24	4 (0,01)	3,35%	1,03%	5,46%	0,00	0,00
10	Cauca	62,80%	0,20	0,20	0,20	0,23	0,00	11 (2,21)	4,75%	3,83%	3,45%	0,00	0,00
11	Cesar	100,00%	0,41	0,33	2,77	3,43	0,08	10					
12	Chocó	58,12%	0,20	0,20	0,20	0,18	0,00	4 (0,04)	5,46%	1,56%	3,56%	0,00	0,00
13	Córdoba	73,84%	0,20	0,20	0,20	0,34	0,00	28 (2,80)	1,12%	3,02%	3,04%	0,00	0,00
14		61,44%	0,20	0,20	0,20	0,21	0,00	11 (1,00) 25 (0,59)	3,47%	7,05%	3,71%	0,00	0,00
15	Bogotá	63,73%	0,20	0,20	0,20	0,23	0,00	4 (0,49) 11 (79,45)	1,29%	3,22%	4,15%	0,00	0,00
16	Guainía	59,23%	0,20	0,20	0,20	0,17	0,02	4 (0,00)	5,68%	3,62%	3,86%	0,00	0,00
17	Huila	68,71%	0,20	0,20	0,20	0,29	0,00	28 (1,60)	1,82%	2,86%	3,85%	0,00	0,00
18	La Guajira	50,78%	0,20	0,20	0,20	0,10	0,01	4 (0,01) 28 (0,55)	3,71%	5,24%	7,94%	0,00	0,00
19		66,82%	0,20	0,20	0,20	0,27	0,00	11 (2,10)	5,22%	3,97%	3,90%	0,00	0,00
20	Meta	74,33%	0,20	0,20	0,20	0,34	0,00	4 (0,10)	4,36%	3,88%	1,39%	0,00	0,00
21	Nariño	68,55%	0,20	0,20	0,20	0,29	0,00	11 (2,60)	3,04%	3,49%	1,22%	0,00	0,00
22	Norte de	75,87%	0,20	0,20	0,20	0,34	0,02	4 (0,10) 28 (1,94)	7,55%	3,91%	4,88%	0,00	0,00
23	Putumayo	37,87%	0,25	0,25	0,25	0,13	0,00	25 (0,00)	3,17%	3,71%	4,58%	0,00	0,00
24	Quindío	62,86%	0,20	0,20	0,20	0,21	0,02	4 (0,04) 28 (1,65)	1,98%	5,28%	3,05%	0,00	0,00
25	Risaralda	100,00%	0,39	0,33	2,60	3,19	0,13	7					
26	San	52,25%	0,25	0,25	0,25	0,27	0,00	25 (0,03)	5,19%	5,14%	3,67%	0,00	0,00
27	Santander	84,42%	0,20	0,20	0,93	1,12	0,06	4 (0,26) 11 (2,27) 25 (1,53)	3,11%	3,00%	3,00%	0,00	0,00
28	Sucre	100,00%	1,43	0,20	2,37	3,91	0,10	7					
29	Tolima	90,39%	0,55	0,20	0,20	0,85	0,00	4 (0,23) 28 (0,59)	3,00%	2,20%	3,77%	0,00	0,00
30	Valle del	100,00%	0,31	0,31	0,78	0,35	1,05	0					

Figura 2. Resultados obtenidos de eficiencia, separando la producción total y los productos de desarrollo tecnológico. **Fuente:** elaboración propia.

Profundizando más en los resultados, puede observarse como el Valle del Cauca tiene siete centros de investigación y de desarrollo tecnológico reconocidos, casi comparable con Antioquia que tiene once, pero que comparativamente posee más inversión en PIB.

Igualmente, el diálogo de relacionamiento entre Universidad-Empresa es mucho más fluido en estos sectores observándose una producción orientada a las necesidades de las organizaciones e industrias.

Implicaciones para las Instituciones de Educación Superior

Como hallazgo se considera que a la par con el desarrollo de las políticas educativas y de calidad en las IES, se hace fundamental considerar los aspectos regulatorios de la Ciencia, Tecnología e Innovación en Colombia, de cara al fortalecimiento de la relación tríadica Universidad-Empresa-Estado. A este respecto, es esencial en primer lugar identificar las etapas del desarrollo de las políticas de CTeI en Colombia, relacionadas en la figura 3.

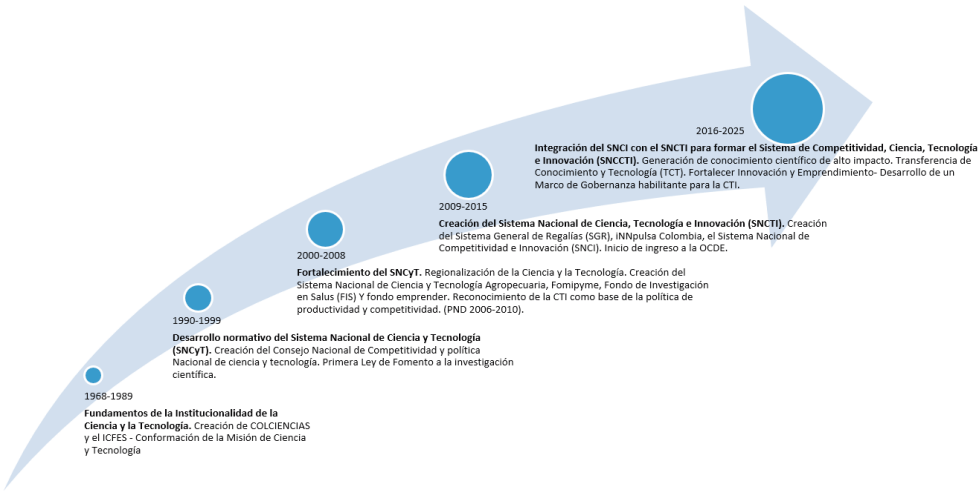


Figura 3. Etapas del desarrollo de las políticas de CTeI en Colombia. Adaptado de Documento CONPES.
Fuente: Consejo Nacional de Política Económica y Social (2016).

Con base en los elementos aportados por la figura 3, puede visualizarse cómo las políticas de desarrollo de CTeI en Colombia, pasan por el establecimiento de fundamentos de la institucionalidad en ciencia y tecnología para luego generar un sistema nacional que posteriormente reconoce la innovación y el emprendimiento como elementos habilitantes para el fortalecimiento de la productividad en el marco de la transferencia de conocimiento y tecnología, lo cual es impactado fuertemente por las regulaciones sobre la neutralidad de la red. Se puede visualizar cómo se pretende regular e institucionalizar la función productiva de la ciencia y la tecnología para de esta forma habilitar a las organizaciones y el sector productivo hacia el "catch-up" el cual involucra fuerte atención a la captura por parte del Estado y las organizaciones, de conocimiento que posibilite la evolución de varias capacidades y habilidades, dicha captura se realiza mediante la identificación de tecnologías específicas blanco, objetivo o target que como característica principal tienen el hecho de tener una alta velocidad de transición entre el producto y sus procesos (Tecnologías de ciclos cortos). En resumidas cuentas, consiste en desarrollar tecnologías de nivel medio con un ciclo de producción corto para luego propender por tecnologías de alto nivel. Por el contrario, estados que propenden por tecnologías de bajo nivel con un largo ciclo de producción (como la fabricación de acero o

el ensamblaje de automóviles) se quedan estancadas en este nivel tecnológico (Lee, 2013, p. 14).

Metodológicamente, los elementos diagnósticos para la concepción de la política en materia de CTeI, fueron interpretados a la luz del Índice Departamental de Innovación para Colombia (IDIC), el cual a su vez fue construido a raíz de la metodología de construcción del Global Innovation Index (Cornell, 2015), para 25 de los 32 departamentos de Colombia. El CONPES resalta acerca de este índice que

> los distintos subíndices que componen el Global Innovation Index son: (i) de insumo o entrada, en el que se miden cinco pilares que apoyan y facilitan las actividades de innovación (instituciones, capital humano e investigación, infraestructura, sofisticación del mercado y sofisticación de las empresas); y (ii) de producto o salida, que evalúan resultados científicos, tecnológicos y creativos (Consejo Nacional de Política Económica y Social, 2016).

El índice establece la relación cuantitativa entre productividad e innovación mediante promedio simple y la consideración de los anteriores numerales que como subíndices se escriben en forma de fracción para conformar la razón de eficiencia en innovación. Otras expresiones como el índice de Malmquist que identifica el nivel de "catching-up" en los estados permiten identificar la relación entre el índice de productividad y el índice de innovación. Para ver estudios al respecto se puede remitir a Taskin y Zaim (1997).

Con base en los elementos anteriores, la política nacional de CTeI comprende los objetivos presentados en la tabla 2.

Desde las implicaciones prácticas de la investigación en materia de formación, de acuerdo con el Observatorio Colombiano de Ciencia y Tecnología, en su portal de datos abiertos, la formación en programas de pregrado en Bogotá y Antioquia ha tenido un crecimiento constante, hasta tal punto de estar cercana a la formación tecnológica para el año 2015, la cual ha estado decreciendo en los últimos años tal y como se observa en la figura 4.

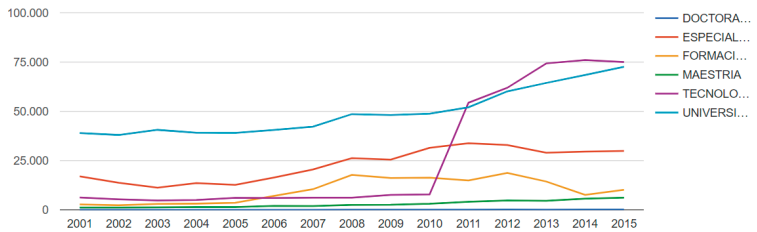
En cuanto a la producción científica en materia de Ingeniería y Tecnología para Antioquia (figura 5), se observa un aumento constante para publicacio-

nes tipo SCOPUS/SciELO/WoS durante los períodos 2006-2015, que decae fuertemente para el año 2016. La interpretación de esta caída se debe en principio por el cambio en los modelos de medición y en la destinación de los recursos para actividades de Ciencia, Tecnología e Innovación (ACTI), donde incluso, se redujo la inversión privada.

Tabla 2. *Objetivos Generales, Específicos y Estratégicos de la Política Nacional de CTeI.*

Objetivo General	Objetivos Específicos	Objetivos Estratégicos
Impulsar el desarrollo económico, social y ambiental del país y sus regiones a través de la ciencia, tecnología e innovación.	• Aumentar la generación de conocimiento científico de alto impacto. • Generar las condiciones para la cooperación entre los sectores productivo, público y privado por medio de la TCT. • Aumentar la actividad innovadora y de emprendimiento en el aparato productivo. • Desarrollar un marco de gobernanza habilitante para la CTeI.	• Fortalecer las capacidades para realizar I+D • Incrementar el impacto de la producción científica colombiana. • Incrementar el capital humano altamente calificado para la realización de ACTI con criterios de calidad y pertinencia. • Fortalecer la mentalidad y cultura de ciencia y la tecnología en la sociedad colombiana. • Preparar el aparato productivo para el aprovechamiento de la TCT • Fortalecer a las instituciones generadoras de conocimiento para la TCT • Fortalecer los servicios de apoyo a la TCT y la vinculación de actores. • Mejorar las capacidades y condiciones del entorno para innovar y emprender • Fortalecer y crear mecanismos de apoyo financiero a la innovación y el emprendimiento • Consolidar la arquitectura institucional actual de la CTeI • Fortalecer la generación de insumos para el diseño, seguimiento y evaluación de la Política de CTeI • Aumentar el esfuerzo público y privado para financiar la CTeI • Desarrollar el marco normativo para CTeI y promover su uso

Fuente: Consejo Nacional de Política Económica y Social, 2016, pp.74-75.



MODALIDAD	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
DOCTORADO	15	10	24	23	18	40	42	69	77	101	113	137	133	180	197
ESPECIALIZACION	17.002	13.765	11.274	13.547	12.646	16.418	20.494	26.257	25.480	31.495	33.776	32.880	28.988	29.575	29.868
FORMACION TECNICA PROFESIONAL	2.760	2.361	2.995	3.132	3.604	7.074	10.474	17.739	16.140	16.317	14.883	18.735	14.368	7.568	10.151
MAESTRIA	1.135	1.200	1.226	1.444	1.463	1.976	1.962	2.538	2.602	3.108	4.080	4.781	4.628	5.636	6.191
TECNOLOGICA	6.213	5.367	4.741	5.017	6.088	6.008	6.147	6.200	7.624	7.800	54.341	61.948	74.310	76.032	74.979
UNIVERSITARIA	38.975	38.014	40.591	39.136	39.064	40.560	42.220	48.581	48.133	48.815	52.031	60.128	64.402	68.435	72.569

Figura 4. Graduados por niveles de formación en Bogotá. **Fuente:** Observatorio de Ciencia y Tecnología.

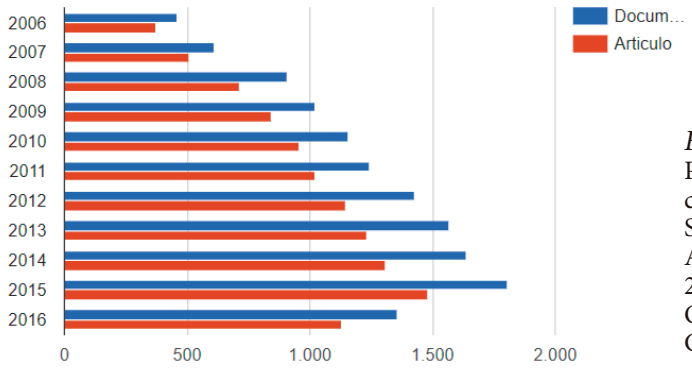


Figura 5. Publicaciones científicas SciELO/ SCOPUS/WoS para Antioquia en los años 2006-2016. **Fuente:** Observatorio de Ciencia y Tecnología.

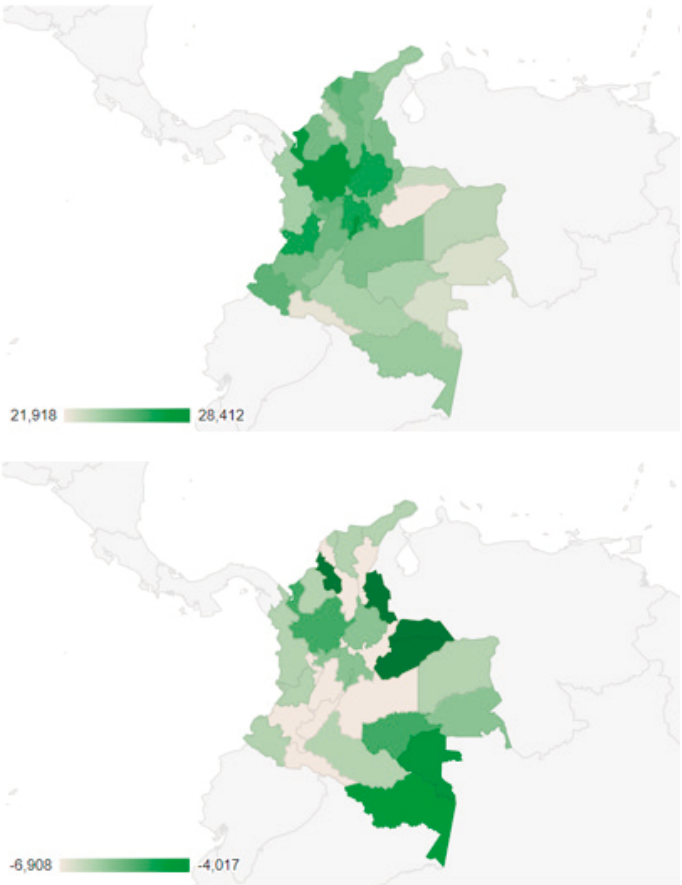


Figura 6. Ejecución tipo ACTI (arriba). % de PIB invertido (abajo). **Fuente:** Observatorio de Ciencia y Tecnología.

Según lo observado en la figura 6, los departamentos o localidades mejor posicionados en el IDIC (Antioquia y Bogotá), no necesariamente recibieron mayores recursos del PIB pero tuvieron mayor ejecución en ACTI, ante este escenario, las políticas de CTeI, pretenden lograr una mayor efectividad y pertinencia de la ejecución de los recursos públicos que promueva igualmente la inversión privada, tales perspectivas deben plantearse desde el diseño para todos, de tal forma que se posibilite un marco de trabajo estable que impacte positivamente a la comunidad tal y como lo establece la Comisión de Regulación de Comunicaciones en la Resolución 3502 de 2011: "…la regulación es un instrumento de intervención del Estado en el sector de Tecnologías de la Información y las comunicaciones, y debe atender a las dimensiones social y económica de las mismas, debiendo para el efecto velar por la libre competencia y la protección de los usuarios, por lo que aquella debe orientarse a la satisfacción de sus derechos e intereses".

Conclusiones

A manera de conclusión, se considera que si bien es cierto es necesario y justificado regular aspectos relacionados con neutralidad de la red, debe ponerse a consideración el papel de las IES de tal suerte que se tenga la preparación y las herramientas adecuadas para formar profesionales aptos para potenciar el desarrollo económico y social de la región. Estos profesionales deberían estar capacitados en la vigilancia tecnológica de los marcos regulatorios, logrando así un impacto significativo en la competitividad del país.

El crecimiento de la producción científica en las IES, debe estar motivado por la articulación con el sector empresarial e industrial, de esta forma se genera una mayor producción tecnológica con investigación de base; para llegar a este nivel de articulación. Las IES deben asumir su rol de organizaciones gestoras y productoras de conocimiento, orientando sus planes de enseñanza y aprendizaje en la reflexión conjunta de las necesidades locales, regionales y nacionales.

Referencias

Beltrán, J., Íñigo, E., y Mata, A. (2014). La responsabilidad social universitaria, el reto de su construcción permanente. *Revista Iberoamericana de Educación Superior*, 3-18.

Comisión de Regulación de Comunicaciones. (2011). Resolución 3502. Bogotá. Colombia.

Consejo Nacional de Política Económica y Social. (2016). *Documento Conpes - Política Nacional de Ciencia, Tecnología e Innovación 2016-2025.* Bogotá: Departamento Nacional de Planeación.

Cooper, W. W., Seiford, L. M., & Tone, K. (2007). *Data Envelopment Analysis. A Comprehensive Text with Models, Applications, References and DEA-Solver Software.* Netherlands: Kluwer Academic Publisher.

Cornell, U. (2015). *Global Innovation Index.* Geneva: OMPI.

Departamento Nacional de Planeación. (2016). *Políticas de desarrollo productivo, ciencia, tecnología e innovación.* Recuperado de: http://investincauca.com/sites/default/files/descargables/politicas_dsllo_productivo_ctei_pnd_0.pdf

Domínguez, G. (2001). La Sociedad del Conocimiento y las organizaciones educativas como generadoras de conocimiento: el nuevo "continuum" cultural y sus repercusiones en las dimensiones de una organización. *Revista Complutense de Educación*, 485-528.

FÄRE, R., Grosskopf, S., Norris, M., y Zhang, Z. (1994). Productivity growth, technical progress, and efficiency change in industrialized countries. *The American Economic Review*, 66-83.

Federal Communications Commission. (2015). In the Matter of Protecting and Promoting the Open Internet. Report and order on remand, declaratory ruling, and order. Washington: FCC.

Holger, S. (15 de 08 de 2000). *Manual de programa EMS.* Recuperado de: http://www2.ulpgc.es/hege/almacen/download/5/5728/Manual_programa_EMS_modelos_DEA.pdf

Instituto de Mayores y Servicios Sociales. (2006). *Libro Blanco del Diseño para Todos en la Universidad.* Madrid: Grup 3. Fundación ONCE.

Lee, K. (2013). *Schumpeterian Analysis of Economic Catch-up.* Cambridge: Cambridge University Press.

Maloney, W., y Bitran, E. (2013). *Outline - Innovación para la Competitividad.* Bogotá D.C.

Maza, F. J., Quesada, V. M., y Vergara, J. C. (2013). Eficiencia y productividad de la calidad educativa en municipios del departamento de Bolívar-Colombia. *Entramado*, 28-39.

Subsecretaría de Telecomunicaciones. (2015). *Grupo de trabajo. Neutralidad de la Red y Gobernanza de Internet.* Santiago de Chile: Regulatel.

Taskin, F., y Zaim, O. (1997). Catching-up and innovation in high- and low-income countries. *Economics Letters*, 93-100.

Uso de tecnologías inmersivas en educación: realidad aumentada, realidad virtual y smartroom

Mauricio Hincapié Montoya[1], Christian Andrés Díaz León[2]

Resumen

En los últimos años se han generado diferentes preocupaciones sobre la calidad de las prácticas de enseñanza, en el caso de Colombia, debido a los bajos niveles de rendimiento que los estudiantes han obtenido en pruebas nacionales e internacionales. Algunos autores han propuesto el estudio y el uso de nuevas tecnologías inmersivas con el propósito de mejorar las experiencias de aprendizaje y enseñanza. Algunas de estas tecnologías son la realidad aumentada y virtual, hasta evolucionar a conceptos integrales como el aula de clase inteligente. En este capítulo los autores proponen explorar diferentes trabajos de investigación relacionados con las tecnologías inmersivas, en particular aquellos realizados por los autores, para discutir respecto a elementos de diseño e implementación que facilitan la utilización de estos conceptos en el aula de clase.

Palabras clave: aula de clase inteligente, realidad virtual, realidad aumentada, tecnologías inmersivas.

1 Ingeniero en Instrumentación y Control, Magister en Matemáticas aplicadas, Doctor en Ingeniería. Docente investigador adscrito al Grupo de Investigación AGLAIA de la Corporación Universitaria Americana. Correo: emhincapie@americana.edu.co

2 Ingeniero Biomédico, Magister en Informática, Doctor en Ingeniería. Docente investigador de la Corporación Universitaria Americana. Correo: christiandiazleon@gmail.edu.co

Using inmersive technologies in education: virtual and augmented reality and smartroom

Abstract

In recent years, different concerns have been raised about the quality of teaching practices, in the case of Colombia, due to the low levels of achievement that students have obtained in national and international tests. Some authors have proposed the study and use of new immersive technologies with the purpose of improving learning and teaching experiences. Some of these technologies are the augmented and virtual reality, until evolving into integral concepts such as the smart classroom. In this chapter of the book the authors propose to explore different research works related to immersive technologies, in particular those made by the authors, to discuss about design and implementation elements that facilitate the use of these concepts in the classroom.

Key words: Augmented reality, immersive technologies, smart room, virtual reality.

Introducción

En los últimos años se han generado diferentes preocupaciones sobre la calidad de las prácticas de enseñanza. Para el caso de Colombia, debido a los bajos niveles de rendimiento que los estudiantes han obtenido en pruebas nacionales e internacionales como PISA (Programa de Evaluación Internacional de Estudiantes). En esta prueba, Colombia se ha ubicado en las últimas posiciones entre 2006 y 2015 (Redacción 2013; Altablero 2008). De manera similar, en otros exámenes internacionales como PIRLS (Progreso en Lectoescritura Internacional) los resultados fueron inferiores al promedio (PIRLS 2011) y TIMSS (Tendencia en el estudio internacional de matemática y ciencias) donde Colombia ha participado por segunda vez, permaneciendo en la posición 36 de 40 países que participaron (Quevedo 2012). En general, la mayoría de los países de América Latina han obtenido puntajes muy bajos en estas pruebas internacionales, algunos casos son México, El Salvador, Chile, Perú, Brasil, entre otros.

Hay muchos factores, tanto internos como externos, que influyen en el rendimiento académico de los estudiantes. Los factores internos son factores directamente relacionados con los procesos pedagógicos, como las deficiencias de los docentes, los currículos inadecuados, el tipo de contenido, entre otros (Salcedo, 2010). Por el contrario, los factores externos son aquellos que no están directamente relacionados con el proceso pedagógico, como el nivel socioeconómico y el nivel de educación de los padres de los estudiantes, las condiciones ambientales en el aula, entre otros (Gaviria y Barrientos, 2001).

Asimismo, el uso de la tecnología ha mejorado la enseñanza y las experiencias de aprendizaje dentro del aula de clase (Kesim & Ozarslan, 2012; Radu, 2014). Una de las tecnologías que ha impactado de forma positiva es la realidad aumentada. La realidad aumentada crea varias capas de información virtual sobre la realidad con el objetivo de incrementar la percepción que el usuario tiene respecto a dicha realidad (Azuma et al. 2001). La realidad aumentada posee varias ventajas desde el punto de vista educativo tales como las describen Cuendet, Bonnard, Do-Lenh & Dillengourg (2013):

- Posee la habilidad de promover el aprendizaje kinestésico.
- Mejora la compresión de cierta información ya que permite la visualización e inspección de objetos 3D o material de clase desde una gran variedad de perspectivas.
- Incrementa los niveles de compromiso y atención de los estudiantes producto de un mayor estado de motivación
- Permite agregar información contextual estática y dinámica a los objetos reales, relacionada con las actividades de aprendizaje.

Por otra parte, Radu (2014) propone otros beneficios que tiene la realidad aumentada en educación:

-
- Incrementa la comprensión del contenido visualizado: Varios artículos han demostrado que la realidad aumentad es más efectiva para enseñar a los estudiantes cuando es comparada con libros, videos o experiencias tipo computadores.
- Aprender la estructura y función espacial: se ha comprobado su efectividad para entender formas geométricas, estructuras químicas, maquinaria mecánica, configuraciones astronómicas, entre otros.

- Aprender asociaciones del lenguaje: facilita la asociación simbólica del lenguaje, como aprender el significado de las palabras escritas.
- Retención en la memoria de largo plazo: se ha demostrado a través de varios estudios que los estudiantes que aprendieron usando realidad aumentada memorizan más fuertemente los conceptos que usando experiencia que no aplican realidad aumentada.
- Mejora el desempeño de tareas físicas: diferentes investigaciones han demostrado que la realidad aumentada es más efectiva para entrenar tareas físicas de alta precisión que otro tipo de medios.
- Incrementa la atención del estudiante.
- Mejora el aprendizaje colaborativo.

Sin embargo, este autor también plantea que la realidad aumentada tiene ciertos perjuicios para el aprendizaje. Por ejemplo, el "*attention tunneling*" donde los estudiantes reportan que los contenidos de realidad aumentada requieren de altas demandas en la atención, lo que puede producir que el estudiante ignore elementos importantes de la experiencia. Entre otros factores que no benefician el aprendizaje en realidad aumentada están:

- Dificultades en la usabilidad.
- Es difícil de integrar al salón de clases.
- Algunos tipos de aprendizaje no se adaptan bien.
- No se conoce el impacto en la memoria y el aprendizaje de la realidad aumentada.

Tanto en realidad virtual como en realidad aumentada, los contenidos que pueden ser desplegados y que soportan la actividad de aprendizaje pueden ser de dos tipos, estáticos o dinámicos (Nincarean, Ali, Dayana, Abdul & Abdul, 2013). Textos, pistas visuales o modelos 3D cuya apariencia no varía durante la interacción del usuario son definidos como contenidos estáticos. Por otra parte, los contenidos dinámicos son aquellos que varían con o sin la interacción del usuario. Visualizaciones dinámicas tales como animaciones o videos son descripciones del concepto que cambian continuamente sobre el tiempo y representan un flujo continuo de movimiento, mientras que las visualizaciones estáticas no muestran el flujo y por lo tanto especifican estados particulares del concepto que se está aprendiendo (Lowe & Schnotz, 2008). Qué tipo de contenido debe ser desplegado en realidad virtual o aumentada depende de la

aplicación y la experiencia que se desea proporcionar al estudiante (Aisnworth, 1999; Aisnworth, 2006).

Por otra parte, la realidad virtual se define como la creación de un ambiente virtual sintético, cuyo propósito inicial es engañar físicamente cada uno de los sentidos del cuerpo humano, visión, tacto, gusto, olfato y audición, para como último propósito engañar a nivel psicológico al ser humano, y cambiar su percepción de estar en la realidad para estar en el ambiente sintético desarrollado. La fidelidad de un ambiente virtual se puede evaluar a partir de dos conceptos. El primero de ellos se define como inmersión, y que varios autores han asociado al nivel de engaño físico logrado, en el cual, cada uno de los sentidos es engañado por el ambiente sintético generado y la forma que es desplegado a los dispositivos (hardware) que interactúan con el usuario. El segundo es definido como presencia, y se refiere al estado psicológico de engaño, donde el usuario siente estar presente en otro escenario diferente a la realidad.

Esta inmersión y presencia lograda por la realidad virtual ha demostrado que facilita los procesos de enseñanza y aprendizaje (Ai-Lim Lee, Wai Wong & Che Fung, 2010). Por otra parte, estudios como los propuestos por Bell y Foyler (1997) han demostrado que la realidad virtual tiene grandes beneficios para todos los estilos de aprendizaje, los cual significa que proporciona las cuatro características de aprendizaje de Kolb's (Kolbs, 2005).

En la próxima sesión describiremos dos experiencias que los autores han tenido desarrollando proyectos de investigación que involucran estos dos conceptos: realidad virtual y realidad aumentada.

Uso de tecnologías inmersivas dentro del aula de clase

En esta sesión se describe la experiencia que han tenido los autores desarrollando tres proyectos de investigación relacionados con el área de educación y que utilizan estas tecnologías inmersivas enunciadas en la sección anterior. El primer proyecto de investigación utiliza la realidad aumentada para enseñar conceptos de electrónica básica y explora el tipo de contenido que puede ser mostrado, ya sea estático o dinámico. El segundo proyecto utiliza realidad virtual para la enseñanza de conceptos básicos de anatomía y fisiología, y ex-

plora cómo la realidad virtual puede complementar el aprendizaje cuando el profesor esta ubicado de manera remota.

Contenidos dinámicos y estáticos en realidad aumentada

El objetivo de este proyecto de investigación era evaluar cómo la naturaleza de los contenidos proyectados en realidad aumentada pueden afectar la comprensión y el aprendizaje de ciertos conceptos.

Se han explorado diferentes estrategias de aprendizaje o actividades cognitivas aplicadas por los estudiantes cuando usan texto o contenidos basados en diagramas (Cromley, Snyder-Hogan y Luciw-Dubas, 2010). Para medir o evaluar estos procesos, se ha propuesto el uso del protocolo de pensar en voz alta y las actividades cognitivas codificadas, tales como: inferencia, conocimiento previo, vocabulario, entre otros. A partir de pruebas experimentales realizadas en asignaturas de aprendizaje como biología, los autores encontraron que los estudiantes realizan actividades cognitivas más elaboradas cuando se aprenden mediante diagramas que utilizando texto, sin embargo, no determinaron si el rendimiento o percepción del aprendizaje fue mejor en algunos de los dos modos (Cromley et al. al., 2010).

Otros trabajos se han centrado en evaluar si hay un efecto en el aprendizaje cuando el alumno usa contenidos estáticos o dinámicos. Se ha descrito un análisis de cómo las diferentes habilidades, habilidades y conocimiento del alumno afectan el proceso de comprensión del contenido dinámico (Hegarty & Kritz, 2008). Además, los autores de este trabajo de investigación informaron ocho estudios en los que se evalúa la comprensión de un sistema mecánico complejo que utiliza diagramas estáticos y animados, con y sin instrucciones verbales. A partir de los resultados, pudieron determinar que la capacidad del espacio no tiene un efecto significativo en la comprensión del contenido, y posiblemente este tipo de habilidad es más útil cuando el contenido es textual o verbal y el estudiante tiene que crear mentalmente una representación visual de es (Hegarty y Kritz, 2008). Finalmente, los autores determinan que no hay un impacto significativo en el aprendizaje cuando se usa contenido estático o dinámico.

En esta propuesta de investigación se plantean dos enfoques para configurar el contenido estático y dinámico en aplicaciones de realidad aumentada. Una basada en características visuales, es decir, cómo las visualizaciones se muestran como dinámicas y estáticas, complementadas con audio y texto, y cómo la interacción se realiza utilizando widgets comunes y dispositivos con pantalla táctil. En este primer enfoque, la visualización es un modelo 3D definido como contenido estático o animación 3D definida como contenido dinámico. Este primer enfoque es similar al propuesto en el estado de la técnica en otro tipo de aplicaciones. El segundo enfoque se basa en los resultados experimentales obtenidos utilizando la primera propuesta. En este enfoque, el contenido estático y dinámico se ve de manera integral, es decir, considerando lo visual, lo interactivo y lo verbal como estático o dinámico. La Tabla 1 muestra las diferencias de configuración entre el contenido estático y dinámico propuesto para la aproximación propuesta.

Tabla 1.

Descripción de las dos aproximaciones propuestas para definir contenidos estáticos y dinámicos.

Aproximación	Type of Content	Sensorial Channel		Interaction*
		Visual	Verbal	
Primera aproximación: contenido basado en elementos visuales	Estático	Modelos 3D no animados	Descripción del contenido basada en texto y audio	Resaltado de diferentes partes de los modelos utilizando widgets, como botones de opción.
	Dinámico	Modelos 3D Animados	Descripción del contenido basada en texto y audio	Resaltado de diferentes partes de los modelos utilizando widgets, como botones de opción.
Segunda aproximación: contenido basado en diferentes elementos de forma integral.	Estático	Modelos 3D no animados	Descripción del contenido basada en texto	Reenvío de descripción de texto utilizando widgets, como botones.
	Dinámico	Modelos 3D Animados	Descripción del contenido basada en audio	Etiquetas seleccionables en el modelo 3D para activar la descripción basada en audio.

* Todos los enfoques permiten el cambio del punto de vista de la visualización cambiando la posición de la cámara del dispositivo móvil con respecto a la posición del objetivo.

Para el desarrollo y la implementación de la aplicación en el dispositivo móvil se utilizó Unity3D. Unity3D es un motor de juegos que se puede integrar con ARCore permitiendo el desarrollo de aplicaciones de realidad aumentada (Hocking, 2015; Murray, 2014). Utilizando las funcionalidades pro-

porcionadas por Unity3D: (i) el contenido estático y dinámico se asociaron a los objetivos, y (ii) la interfaz de usuario se creó con botones para alternar la primera aproximación o la colisión con etiquetas para el segundo enfoque, que permitió al alumno interactuar con el contenido, mostrar textos y reproducir audios explicativos y visualizaciones 3D. Por ejemplo, a través de un conjunto de botones, el alumno puede elegir una estructura del átomo, y la aplicación muestra y explica la estructura, cambiando el contenido estático o dinámico, y el audio y el texto que se muestran. La aplicación de realidad aumentada desplegada en el dispositivo móvil se puede observar en la Figura 1.

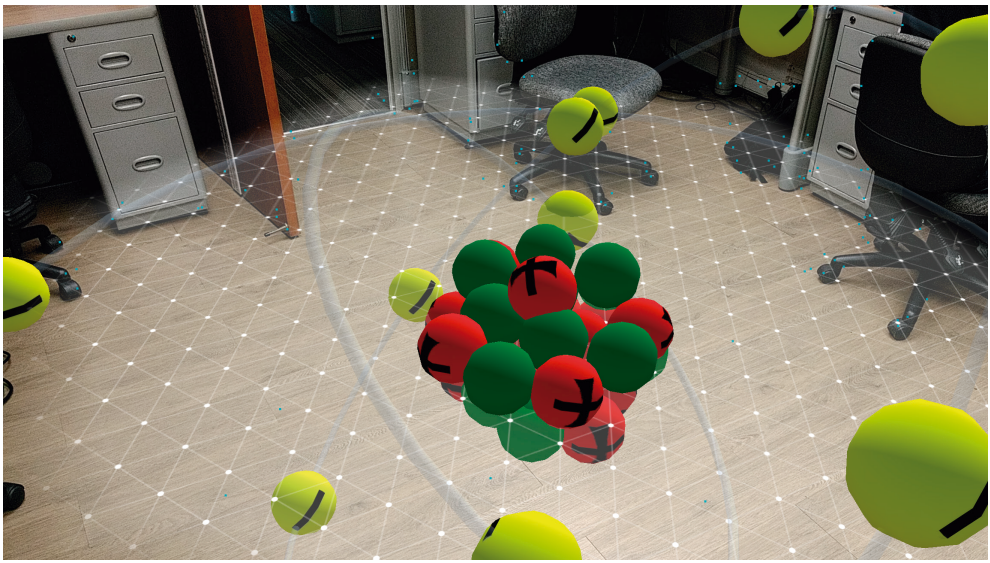


Figura 1. Estudiante interactuando con la aplicación de realidad aumentada desarrollada.

Dos pruebas experimentales fueron propuestas, una utilizando la primera aproximación y la otra usando la segunda aproximación de contenidos estáticos y dinámicos descrita en la tabla 1. En los experimentos un conjunto de 41 estudiantes fueron evaluados utilizando las primera y la segunda aproximación propuesta en la tabla 1. Cada estudiante, realizó un pret-test para determinar su conocimiento inicial en los conceptos de electrónica básica, luego experimentaba y exploraba la aplicación de realidad aumentada por 10 minutos, y finalmente realizaba un post-test con el propósito de evaluar la mejora en la comprensión y aprendizaje del concepto estudiado.

En la tabla 2 se pueden observar los resultados de análisis estadístico realizado para los datos obtenidos durante las pruebas experimentales.

Tabla 2.

Análisis estadístico de cada una de las variables medidas en los dos experimentos.

Abreviación	Variable	Min	Max	Media	Mediana	Desv. Est.
FA_{pr}	Pre-test usando FA*	0.0	3.25	0.7031	0.5	0.8622
FA_{po}	Post-test usando FA	3.0	5.0	4.297	4.0	0.690
SA_{pr}	Pre-test usando SA*	0.0	3.0	0.94	1.0	0.820
SA_{po}	Post-test usando SA	4.0	6.0	5.47	6.0	0.674
FA_{sc}	Incremento en la calificación usando un contenido estático y FA	0	2	1.359	1.5	0.6829
FA_{dc}	Incremento en la calificación usando un contenido dinámico y FA	1.0	3.0	2.234	2.0	0.760
SA_{sc}	Incremento en la calificación usando un contenido estático y SA	0.5	3	1.73	2	0.6919
SA_{dc}	Incremento en la calificación usando un contenido dinámico y SA	1.25	3	2.80	3	0.4448
FA_{Diff}	Diferencia entre el post-test y el pre-test usando FA	1.750	5	3.594	4	1.102
SA_{Diff}	Diferencia entre el post-test y el pre-test usando SA	2.25	6	4.53	5	0.971

*FA es la abreviación para la primera aproximación y SA es la abreviación de la segunda aproximación

A partir de los resultados se puede afirmar que la calificación obtenida por los estudiantes cuando aprendieron usando un contenido dinámico es mayor que cuando aprendieron usando un contenido estático. Para determinar si hay un aumento estadísticamente significativo entre la calificación obtenida por los estudiantes que utilizan un contenido estático y dinámico, se aplicó una prueba de hipótesis estadística. Con este propósito se aplicó una prueba de Shapiro para determinar si las muestras provenían de una población distribuida normalmente, los valores p obtenidos para las variables $SA_{sc}SA_{sc}$, and SA_{dc} SA_{dc} fueron 0.2071 y 7.887 x 10-8, respectivamente. Estos valores p sugieren, con un intervalo de confianza del 95%, que la variable SA_{dc} SA_{dc} no tiene

una distribución normal, pero la variable $SA_{sc}SA_{sc}$ si tiene un distribución normal. Por esta razón, se aplicó una prueba emparejada de Wilcoxon para determinar si existe una diferencia significativa entre las calificaciones obtenidas usando contenido estático y dinámico. El valor p calculado para comparar el contenido estático y dinámico es 2.586 x 10-5. Estos valores p sugieren, con un intervalo de confianza del 95%, que las calificaciones obtenidas por los estudiantes que utilizan contenido estático y dinámico no son similares. Además, las diferencias en la media (Tabla 2) de las variables sugieren que el aumento en la calificación causado por el contenido dinámico es mayor (M = 2.80, SD = 0.4448) que el aumento en la calificación causado por el contenido estático (M = 1.73, SD = 0.6919). Esta afirmación puede contradecir lo que se reporta en la literatura donde los autores han comparado visualizaciones estáticas y dinámicas, pero vale la pena tener en cuenta que la aproximación propuesta en el articulo es diferente porque usa realidad aumentada y utiliza una propuesta integral de contenido dinámico en la que lo visual, lo verbal y la interacción son dinámicos.

Disminuyendo el impacto de la presencia remota mediante realidad virtual

En esta segunda investigación se planteó el uso de avances en las tecnologías de la información y en el área de la realidad virtual para el desarrollo de una herramienta de enseñanza remota que aplique conceptos concernientes al sistema músculo-esquelético y de esta manera llegar a proponer una solución a dos importantes problemas en la educación como los son la limitante geográfica y la limitante didáctica e interactiva, cuando se da una clase de manera no presencial a un grupo de estudiante. Para hacer frente a este problema, se propuso el uso de un ambiente que utiliza la realidad virtual como elemento didáctico para describir a profundidad el concepto de estudio dentro de la clase, y un ambiente de videoconferencia para complementar la charla del docente de manera remota.

Existen diferentes trabajos desarrollados con el fin de mejorar la enseñanza de conceptos anatómicos en general, a continuación resumiremos algunos de ellos. El AnatLine es una aplicación desarrollada por la National Library of Medicine la cual es un buscador anatómico y una base de datos en línea. El buscador anatómico proporciona imágenes prerenderizadas de modelos 3D

que el usuario puede usar para navegar a través de la región del torax del cuerpo (AnatLine, 2018).

Por otra parte, el sistema "Web-based three-dimensional Virtual Body Structures" (W3D-VBS) es un sistema de entrenamiento anatómico basado en realidad virtual para la enseñanza de la anatomía humana sobre Internet cuya arquitectura es de la forma cliente servidor. Los usuarios pueden realizar un recorrido virtual del cuerpo completo mientras exploran y manipulan rebanadas y un volumen de interés determinado. Además, a través de un dispositivo háptico como el phantom permite palpar la estructura visualizada (Temkin, et al. 2002).

En Alverson, Saiki, & Caudell (2005) fue desarrollado un ambiente de simulación y entrenamiento médico. En dicho ambiente se realizó una simulación de un paciente interactivo que permitía a los estudiantes dinámicamente determinar los resultados del escenario, tratando de crear situaciones comunes en una sala de emergencias. El ambiente permitía desarrollar colaborativamente las tareas necesarias sobre el paciente virtual, y además la evaluación remota de un experto.

Por último, el ambiente de simulación "Virtual interactive musculoskeletal system" VISM es un de los primeros ambientes de realidad virtual donde se emula el comportamiento y la funcionalidad del sistema músculo esquelético. Este sistema esta orientado a la investigación, educación del área de biomecánica y ortopedia y además como herramienta clínica de planeación (Chao, Armiger, and Yoshida, 2007). Existen otras investigaciones que hacen uso de realidad virtual y realidad mixta con el fin de facilitar el aprendizaje de ciertos conceptos, una buena revisión puede ser encontrada en Pan et al. (2006). Por otra parte, en el laboratorio de realidad virtual de la Universidad EAFIT, se desarrolló una herramienta que aplica realidad aumentada para la enseñanza de conceptos espaciales de Cálculo (Alvarez, Jaramillo y Trefftz, 2003; Orozco, Esteban y Trefftz, 2006).

Para la enseñanza de los conceptos anatómicos y funcionales del sistema músculo esquelético, se desarrolló un contenido en Java 3D el cual era cargado, visualizado y distribuido con la ayuda de la aplicación Telepresencia (Restrepo y Trefftz, 2005). La herramienta de Telepresencia es una aplicación desarrollada en Java que hace uso de ambientes Virtuales Colaborativos, como

herramienta de apoyo para procesos de enseñanza y aprendizaje en el ámbito universitario.

Básicamente la aplicación está compuesta por los siguientes tres ambientes:

- Ambiente de diapositivas: este ambiente permite utilizar ayudas visuales tipo diapositivas para apoyar el desarrollo de la sesión. Cuando el ambiente de diapositivas está activo, el ambiente de Realidad Virtual no y viceversa.
- Ambiente de teleconferencia: el ambiente de teleconferencia implementa la captura y reproducción del video y audio local y remoto. Además, la telepresencia permite que el usuario pueda variar la calidad del video y el audio dependiendo de la conexión de la que se disponga.
- Ambiente de realidad virtual: este módulo consiste en un ambiente virtual colaborativo en el cual tanto el profesor como el estudiante pueden interactuar con objetos 3D compartidos, los cuales dependen del contenido específico a utilizar. Los participantes tienen conocimiento de la ubicación y/o atención de su contraparte en el ambiente virtual gracias a objetos teleapuntadores (flechas 3D). El profesor puede interactuar con el ambiente virtual con un sensor de posición electromagnético (Polhemus). Cada contenido que se carga en el ambiente de Realidad Virtual consta de dos partes: un panel donde se muestran los objetos 3D (panel de realidad virtual) y otro panel el cual contiene los elementos que controlan el ambiente virtual.

El contenido desarrollado consiste de un modelo 3D de estructuras óseas y musculares que interactúan entre sí. Los conceptos que deseaban ser enseñados por medio de este contenido son los tipos de contracción (isométrica, excéntrica y concéntrica), coordinación muscular (motores primarios, motores secundarios, neutralizadores y fijadores) y finalmente el significado funcional de antagonismo y agonismo.

Los conceptos mencionados anteriormente serían tratados a partir del movimiento de flexión y extensión de la articulación del codo, por tal razón se cargaron modelos de la escápula, el húmero, el cúbito y el radio, que eran las estructuras directamente involucradas con el movimiento. Sin embargo, para mejorar la relación espacial que el estudiante captaba del cuerpo humano en la simulación y por lo tanto su calidad, se adicionaron modelos de las vértebras,

costillas, esternón y clavícula. Cada estructura ósea era fácilmente distinguida por un color diferente, como puede ser observado en la figura 2.

Por otra parte, la interacción con el contenido desarrollado se llevaba a cabo mediante un teleapuntador cuyo movimiento es descrito por medio de un rastreador electromagnético. Igualmente, el panel de control relacionado con el contenido estaba compuesto por tres subpaneles los cuales controlaban las estructuras musculares, conceptos funcionales y finalmente el estado de la simulación.

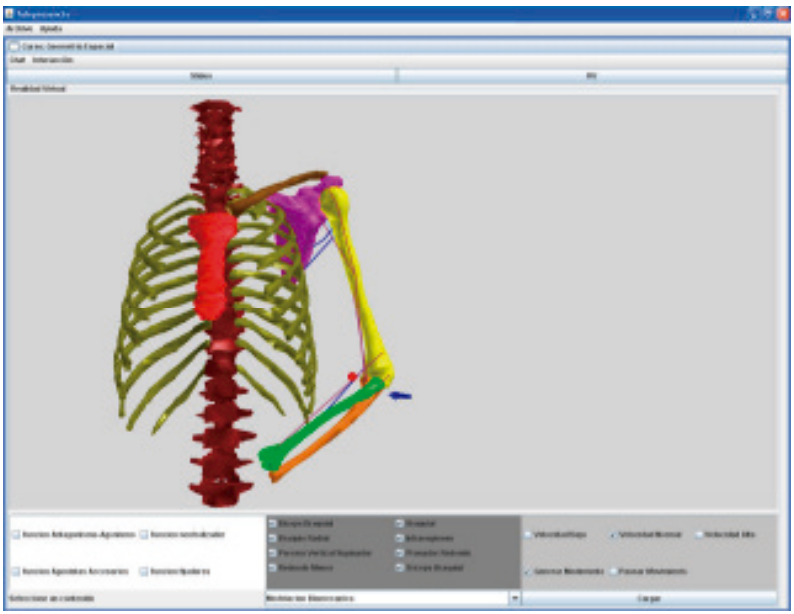


Figura 2. Estructura ósea simulada

La hipótesis que se propone probar a través del contenido desarrollado para Telepresencia, es determinar si por medio de una clase no presencial se logra obtener el mismo nivel de aprendizaje que con una clase presencial que haga uso de métodos de enseñanza tradicionales. Para probar dicha hipótesis, se desarrolló una prueba experimental, en la cual se conformaron dos grupos de estudiantes de Ingeniería Biomédica que no tenían conocimientos previos referentes a conceptos relacionados al sistema músculo-esquelético. En total se evaluaron 25 estudiantes, donde un conjunto de estudiantes hacían parte

del grupo control, los cuales realizaron la sesión de enseñanza con el profesor de manera presencial y utilizando recursos tecnológicos tradicionales, como un tablero, marcador y un proyector. El otro grupo de estudiantes hizo parte del grupo experimental, los cuales recibieron la sesión de enseñanza realizada a través de la herramienta de telepresencia descrita anteriormente y el profesor estaba localizador remotamente.

La evaluación de los conocimientos adquiridos por los estudiantes se realizó de la misma manera como se planteo en el proyecto usando realidad aumentada. Se aplicó un pre-test para analizar el conocimiento previo de los estudiantes, y luego un post-test después de que cada grupo de estudiantes fue expuesto a la sesión de enseñanza, con el propósito de determinar el nivel de aprendizaje logrado en el área.

Con respecto a la configuración física. Cada una de las estaciones de trabajo la del instructor y el grupo de estudiantes contaba con un computador, una cámara Web, un micrófono y un par de altavoces. A diferencia de los estudiantes el profesor además contaba con un rastreador electromagnético el cual le servía como apuntador dentro del ambiente virtual. Los estudiantes se encontraban en un salón y la imagen de la aplicación era proyectada por un videobeam. La conexión entre las dos estaciones de trabajo era realizada por medio de la red de área local.

Como fue mencionado en la configuración experimental, los estudiantes de ambos grupos, el de control y el experimental presentaron un pre-examen antes de realizar la clase. Ambos exámenes, el pre-examen y el pos-examen consistían de un conjunto de 20 preguntas de selección múltiple. Las preguntas evaluaban cada uno de los conceptos tratados durante la prueba educativa. El promedio de calificación del grupo de control y experimental después de realizar el pre-examen fue de 2,0 y 1,8 respectivamente, lo que mostró que ambos grupos estaban en un nivel de cono- cimiento similar con respecto al tema. Posteriormente, se realizó la sesión de aprendizaje bajo cada una de las modalidades dependiendo del grupo. En la figura 4 se puede observar el porcentaje de estudiantes que obtuvo cada una de las calificaciones, en los dos grupos. El promedio de calificación del grupo de control y experimental después de realizar el pos-examen fue de 3,2 y 3,8 respectivamente.

Después de la sesión, los estudiantes del grupo experimental llenaron unos cuestionarios acerca de la herramienta y acerca de su impacto en la comprensión de cada uno de los conceptos. Varios estudiantes reportaron que la herramienta y el contenido les ayudó a entender más rápido y fácil cada uno de los conceptos que el profesor trataba durante la sesión.

Conclusiones

En el capítulo se introduce el tema de tecnologías inmersivas en educación, y se proporciona una descripción detallada de cómo se puede investigar y aplicar la realidad aumentada y virtual como dos tecnologías inmersivas que tienen mucho por aportar en el área de educación.

Dos proyectos de investigación son descritos. En el primero de ellos se aborda la evaluación de la realidad aumentada como medio para entregar un contenidos y conocimiento en el aula de clase a manera de trabajo individual. Particularmente, se profundiza en como la naturaleza del contenido desplegado a través de realidad aumentada puede afectar el nivel de aprendizaje del estudiante. Dos aproximaciones son propuestas, utilizando contenidos de tipo dinámico y estático. A partir de las pruebas realizadas se puede afirmar que debido a la naturaleza dinámica que ofrece la interacción en realidad aumentada, donde el usuario es libre de cambiar de puntos de vista y analizar el contenido con una rica interacción, es más apropiado usar contenidos de tipo dinámico, complementados por una rica interacción, y notificaciones por audio o texto.

Finalmente, en el segundo proyecto de investigación descrito, aunque no se pudo afirmar, al no tener resultados estadísticamente significativos, si se puede sugerir que la realidad virtual puede enriquecer el ambiente de aprendizaje en conceptos abstractos o complejos visualmente, cuando el instructor o docente esta localizado remotamente. El grupo de estudiantes, que usaron el contenido de realidad virtual para aprender los conceptos anatómicos y fisiológicos del movimiento del codo, lograron una calificación promedio más alta en el post-test. Por otra parte, como parte de los comentarios de los estudiantes y el instructor indicaron que requirió menos esfuerzo explicar y comprender los conceptos o las preguntas, a pesar de estar localizados remotamente.

Referencias

Ainsworth S. (1999). The functions of multiple representations. *Computers & Education*, *33*(2-3), 131–152.

Ainsworth S. DeFT (2006). A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*, 183–198.

Ai-Lim Lee, E.; Wai Wong, K.; Che Fung, C. (2010). udents with Different Learning Styles. *Transactions on Edutaiment IV*, pp. 79 – 90.

Altablero. (Enero-Marzo de 2008). Colombia: Qué y cómo mejorar a partir de las pruebas PISA. Ministerio de Educación Nacional. Recuperado de http://www.mineducacion.gov.co/ 1621/article-162392.html

Azuma, R.; Baillot, Y.; Behringer, R.; Feiner, S.; Julier, S. & MacIntyre B. (2001). Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, *21*, 34–47.

Bell, J. T.; Foyler, H. S. (1997). Virtual reality applications for engineering education. Paper presented at erican Society for Engineering Education Milwaukee, WI

Bétrancourt M. (2005). *The animation and interactivity principles in multimedia learning. The Cambridge handbook of multimedia learning*. Cambridge University Press, 287–296.

Cromley J, Snyder-Hogan L, Luciw-Dubas U. (2010). Cognitive activities in complex science text and diagrams. *Contemporary Educational Psychology*, *35*(1), 59 -74.

Cuendet S, Bonnard Q, Do-Lenh S, Dillengourg P. (2013). Designing augmented reality for the classroom. *Computer & Education*, *68*, 557-569.

Cushnan, D., El-Habbak, H. (2013). *Developing AR Games for iOS and Android*, Packt Publishing.

Diaz, C., Hincapié, M., Moreno, G. (2015). How the type of content in educative augmented reality application affects the learning experience. *Procedia Computer Science: 2015 International Conference Virtual and Augmented Reality in Education*, 75, pp. 205-212,

Díaz. C., Hincapie, M. and Moreno, G. (2017). Evaluating the effect on user perception and performance of static and dynamic contents dep3loyed in augmented reality based learning application. *EURASIA*

Journal of Mathematics Science and Technology Education, 13(1), 1-22.

Gaviria, A. y Barrientos, J. (2001). Determinantes de la calidad de la educación en Colombia. *Archivos de Economía*, Departamento Nacional de Planeación. No. 159.

Grubert, J., Grasset, R. (2013). *Augmented Reality for Android Application Development*, Packt Publishing.

Hegarty M, Kritz S. (2008). Effects of knowledge and spatial ability on learning from animation. Learning with animation: Research implications for design, pp. 3 – 29.

Hocking, J. (2015). *Unity in Action: Multiplatform Game Development in C# with Unity 5* (1st Edition), Manning Publications.

Kesim M, Ozarslan Y. (2012). Augmented reality in education: current technologies and the potential for education. *Procedia – Social and Behavioral Sciences*, *47*, 207 – 302.

Kolb, A.Y., Kolb, D.A. (2005). The kolb learning style inventory - version 3.1 2005 technical specifications

Kuhl T, Scheiter K, Gerjets P, Gemballa S. (2011). Can differences in learning strategies explain the benefits of learning from static and dynamic visualizations? *Computers & Education*, *56*, 176-187.

Lowe R, Schnotz W. (2008). A unified view of learning from animated and static graphics. In Learning with animation: Research implications for design, ed. Richard Lowe and Wolfgang Schnotz, pp. 304 – 356, Cambridge University Press.

Mayer, R. (2009). *Multimedia learning* (2nd ed.), Cambridge University Press.

Murray, J. (2014). C# Game Programming Cookbook for Unity 3D, CRC Press.

Nincarean D, Ali M, Dayana N, Abdul N, Abdul M. (2013). Mobile Augmented Reality: the potential for education. *Procedia – Social and Behavioral Sciences*, *103*, pp. 657 – 664.

Radu, I. (2014) Augmented reality in education: a meta-review and cross – media analysis. Pers Ubiquit Comput, 18:1533-1543.

Salcedo, A. (2010). "Deserción Universitaria en Colombia". Revista Academia y Virtualidad. INSEDI. Volumen 3-N°1-ISSN2011-0731-2010.

Schnotz W. (2005). An integrated model of text and picture comprehension. The Cambridge handbook of multimedia learning, Cambridge University Press, pp. 49–69.