



EL DESCUBRIMIENTO DE CONOCIMIENTO EN BASES DE DATOS, CONTRIBUYE A LA EDUCACIÓN CONTINUA DE LOS INGENIEROS DE SISTEMAS DE LA INSTITUCIÓN UNIVERSITARIA CESMAG

José María Muñoz Botina, Carlos Fernando González Guzmán, Javier Alejandro Jiménez Toledo

**Institución Universitaria CESMAG
Pasto, Colombia**

Resumen

Las diferentes Instituciones de Educación Superior en Latinoamérica, pretenden siempre estar en contacto con las personas que egresan de sus diferentes programas académicos, ya sea por diferentes medios como: las redes sociales, correo electrónico, números de celulares o simplemente la dirección de residencia o de trabajo. Es así como el programa de Ingeniería de Sistemas de la Institución Universitaria CESMAG desde el año 2010, priorizó políticas, sistemas y procesos alrededor de las actividades, servicios y ubicación de sus graduad@s, de manera permanentemente, ya que con el tiempo tiende a crecer el número de ellos/ellas, y se incrementa la dificultad de realizar un seguimiento de sus vivencias tanto laborales, educativas y como personales, ya que la misma dinámica del individuo, el medio laboral y sus necesidades lo/la pudieron llevar a trabajar o vivir fuera del país, o en otras regiones o en casos lamentables hasta pudieron fallecer; en ese orden de ideas, se requiere conocer su situación real actual, además de sus necesidades de actualización y formación pos-gradual para que a través de la Oficina de Proyección Institucional, la Coordinación de Egresados y el Programa de Educación continuada, la Institución contribuya a solventar esta problemática. No obstante, la información de los/las graduados/as se encuentra en diferentes medios documentales, filmicos o digitales como archivos, bases de datos, entre otros, aumentando el nivel de complejidad para el acceso a dicha información, por tal motivo, el grupo de investigación Tecnofilia, adscrito al programa de Ingeniería de Sistemas, aprovechando el descubrimiento de conocimiento en bases de datos y el análisis multidimensional está realizando trabajos investigativos que le permitan cruzar diferentes variables relacionadas con los/las egresados/as y sus necesidades de formación, y por otra parte, analizar la coherencia, entre las competencias del Ingeniero/a de Sistemas que el medio laboral necesita con los que fue formado, de allí

que el proceso que se está desarrollando busca optimizar los recursos y un portafolio de servicios para ofrecer al graduado/da lo que realmente debe contar y de ésta forma contribuir a mejorar su calidad de vida impactando positivamente en el medio.

Palabras clave: graduados; educación continua; descubrimiento de conocimiento

Abstract

The various institutions of Higher Education in Latin America, they always get in touch with the people who graduate from their different academic programs, either by different media such as: social networks, email, cell phones, or simply the address of residence or work. This is how the Systems Engineering program of the institution University CESMAG since the year 2010, prioritized policies, systems and processes around the activities, services and location of their graduad@s, permanently, since over time tends to grow the number of them, and it increases the difficulty of monitoring their experiences both educational and labor, as well as personal, because the same dynamics of the individual could lead to work or live outside the country, or in other regions or in unfortunate instances were even able to die; in that order Of ideas, it is necessary to know the real situation, in addition to their current needs of post-training and step-by-step through the Office of Institutional Projection, the coordination of graduates and the Continuing Education Program, the Institution will contribute to solve this problem. However, the information of the graduates is found in different media documentaries, film or digital files, databases, among others, increasing the level of complexity for access to such information, for this reason, the research group Tecnofilia, attached to the Systems Engineering program, taking advantage of the knowledge discovery in databases and the multidimensional analysis is doing investigative work that will allow crossing different variables related to the graduates and their training needs, and on the other hand, analyze the coherence between the powers of the Engineer/systems that the working environment need with that was formed, so that the process that is developing seeks to optimize resources and a portfolio of services to offer the graduate/gives you what you really must count and in this way contribute to improve their quality of life having a positive impact on the environment.

Keywords: graduates; continuing education; knowledge discovery

1. Introducción

Es importante tener en cuenta que el ser humano, en todo momento se está aprendiendo, y mucho más, si en la disciplina en la que se ha formado, es tan dinámica, con lo es la Ingeniería de Sistemas. Por lo tanto, en este documento se pretende mostrar como desde el descubrimiento de conocimiento en base de datos, con y sus diferentes etapas que van desde la selección, el pre-procesamiento, la transformación, la minería de datos, y la interpretación y evaluación de datos; junto con un análisis multidimensional como herramienta computacional busca caracterizar las necesidades

de formación de los egresados del programa de Ingeniería de Sistemas, en cuanto a sus expectativas y formación post gradual o complementaria, aprovechando para ello el proceso realizado de autoevaluación del programa de Ingeniería de Sistemas con miras a la renovación del registro calificado; articulando acciones con la Oficina de Proyección Institucional y su Coordinación de Egresados, con el fin de apoyar los procesos de seguimiento de los mismos y de esta forma, poderles brindar un mejor servicio y además saber cuáles son las competencias del sector laboral con respecto a I@s Ingenier@s de Sistemas.

Por otra parte, los/las egresados/as hacen parte de la comunidad educativa y sus aportes, para la Institución como para ellos mismos, son bastante valiosos, sus opiniones y sugerencias sobre la formación que recibieron, y retroalimentación y participación es sumamente importante para identificar los aspectos en los cuales tanto el programa de ingeniería de sistemas como los diferentes programas de la Institución pueden mejorar, e igualmente que necesidades específicas requieren en materia de educación continuada y formación post gradual.

Debido a que los datos provenían de diferentes fuentes como: el Observatorio Laboral para la Educación, el Sistema de Información institucional ZEUS y el sistema que maneja la Coordinación de Egresados; se realizó el proceso de selección de dicha información, para posteriormente complementarla, depurarla y con el uso de la herramienta computacional WEKA, analizar y contrastar algoritmos, obteniendo información del cruce de diferentes variables asociadas a las necesidades de formación que conlleva la retroalimentación al interior del programa académico, generando los cambios pertinentes en el currículo (plan de estudios). Cabe señalar que esta investigación está en la etapa final, pero todavía no está concluida.

2. Problemática

Es un estudio realizado por la Universidad Católica del Norte (2016), se afirma que algunas universidades de la costa atlántica han señalado que uno de los factores más preocupantes en las IES en Colombia, se refiere al seguimiento de la situación de sus graduados, desde el punto de vista de su desarrollo personal, laboral, social y su impacto en el medio. Por otra parte, la Institución Universitaria CESMAG se demoró en fijar políticas y operacionalizarlas a través de sistemas de información eficientes alrededor de los graduados, trayendo como consecuencia la dificultad para comunicarse con ellos, ya sea por el cambio de trabajo, de ciudad, de dirección o de celular, o por fallecimiento, entre otros datos; generando una gran cantidad de información pero desarticulada, incompleta, a veces desactualizada o errónea; solo desde el año 2010, como plan de mejoramiento y luego incluido en plan de desarrollo institucional 2013-2019, apenas se comenzó a trabajar sobre la gestión con egresados, pero con todos los inconvenientes de manejo de información que se han señalado anteriormente.

En este orden de ideas, de continuar con esta situación, la institución y en particular el programa de Ingeniería de Sistemas, a pesar de que se realizan procesos de

autoevaluación periódicos, no cuenta con la información necesaria para ofertar cursos de actualización y complementación a través del programa de educación continuada, así como tampoco el ofrecimiento de formación de alto nivel, como maestrías o doctorados, alejando a los profesionales de oportunidades que el mundo está requiriendo en materia de Tecnología e Informática -TI.

3. Metodología

Para este proyecto se optó como metodología, las diferentes etapas del proceso de descubrimiento de conocimiento en bases de datos; así, inicialmente se seleccionaron, del Sistema de Información Institucional ZEUS de la I.U.CESMAG, la información relacionada con aspectos personales, socio-económicos, disciplinares e institucionales de los estudiantes que egresaron entre el 2010 y 2016, además de los datos, gestionados por la Coordinación de Egresados desde el Observatorio Laboral de la Educación del programa de Ingeniería de Sistemas. Posteriormente, con estos datos se creó un repositorio utilizando para ello el Gestor de base de datos MySQL. A estos datos se les aplicó las etapas de pre-procesamiento y transformación con el fin de obtener conjuntos de datos limpios y listos para aplicarles las técnicas y los algoritmos de minería de datos. Se obtendrán perfiles de necesidades de formación de los egresados con el empleo de la técnica de clasificación (algoritmo) basada en árboles de decisión con la herramienta libre de minería de datos Weka (García, s.f.) y finalmente, estos resultados serán analizados, evaluados e interpretados para determinar la validez del conocimiento obtenido. A continuación, se detalla cada una de ellas:

3.1 Etapa de selección de datos

El objetivo de esta etapa fue identificar las fuentes internas y externas de datos, que sirvieran de base para el proceso posterior de minería de datos. Como fuentes internas, se seleccionaron las bases de datos del sistema de información institucional ZEUS y también la base de datos que maneja la Coordinación de Egresados en la Oficina de Proyección Institucional. La primera almacena información personal y académica de 471 egresados entre 2010 y 2016; y la segunda almacena información actualizada de carácter personal y laboral de 451 graduados en el mismo lapso de tiempo. Se hizo una selección inicial de 60 atributos de estas dos bases de datos y, conjuntamente con sus registros, se almacenaron en un el repositorio de datos denominado REP_INTERNO, compuesto por 471 registros y 60 atributos.

Como fuentes externas principales, se seleccionó la base de dato del Observatorio Laboral para la Educación del Ministerio de Educación Nacional, e información de la Registraduría Nacional del Estado Civil Colombiano. De estos 317 registros, se escogieron únicamente aquellos datos de los egresados desde el 2010 y 2016, y que participaron del proceso de autoevaluación. Como resultados se obtuvieron 106 registros con 40 atributos, correspondientes a información socioeconómica, académica y laboral. Estos datos fueron almacenados en un nuevo repositorio denominado REP_EXTERNO. Con respecto a los 40 atributos, de éstos se

seleccionaron inicialmente 15, que se consideraron relacionados con las necesidades de formación. Finalmente se obtuvo la tabla REP_GENERAL, con 106 registros y 15 campos, para trabajar la siguiente etapa.

3.2 Etapa de pre-procesamiento de datos

En esta etapa se pretende obtener datos limpios. Por medio de consultas SQL ad-hoc o a través de histogramas, se analizó minuciosamente la calidad de los datos contenidos en cada uno de los atributos de la tabla REP_GENERAL. Al tener en cuenta la relevancia de ciertos atributos para la investigación, los valores nulos de éstos fueron actualizados con los valores encontrados en fuentes externas.

Adicionalmente, se agregaron nuevos atributos obtenidos en fuentes externas y también derivándolos de otros atributos existentes, como resultado de estos procesos, se obtuvo el repositorio final REP_FINAL, con 106 registros y 25 campos, que se utilizó en las etapas subsiguientes de transformación y minería de datos, cuyos atributos y distintos valores se muestran en la tabla 4.

3.3 Etapa de transformación de datos

Su objetivo fue transformar la fuente de datos en un conjunto listo para aplicar las diferentes técnicas de minería de datos.

3.4 Etapa de minería de datos

En esta etapa, es la más relevante del proceso de descubrimiento de conocimiento en bases de datos, cuyo objetivo es la búsqueda, extracción y descubrimiento de patrones inesperados pero de mucho interés, que a simple vista no se identifican; está constituida por las siguientes actividades, cada una de las cuales puede considerarse como un problema que puede ser resuelto a través de un algoritmo de minería de datos, como lo señala Adamo (2001) y Hernández, Ramírez y Ferri (2005) referenciados por Timaran(2015); la clasificación por árboles de decisión es por ejemplo una de ellas. El conocimiento obtenido en el proceso de aprendizaje, según Wang, Iyer y Scott (1998), se representa mediante un árbol, en el cual cada nodo interior contiene una pregunta sobre un atributo concreto (con una rama por cada posible respuesta) y cada hoja del árbol se refiere a una decisión (una clasificación).

Durante la etapa de construcción del árbol de decisión, en forma recursiva, cada conjunto de datos se divide en subconjuntos de acuerdo a un criterio de particionamiento, con el fin de escoger el atributo que mejor separe los ejemplos restantes en clases individuales. Por esta razón, se escogió la tarea de minería de datos, la clasificación por árboles de decisión para el proceso de determinar las necesidades más frecuentes de formación en los egresados del programa de Ingeniería de Sistemas de la I.U.CESMAG, al tener en cuenta que con los valores del atributo clase nivel_formacion_alcanzado, se puede construir un modelo de clasificación que determine las características de las necesidades de formación de los egresados.

4. Resultados parciales

Debido a que es una investigación que todavía está en curso, porque se está finalizando la etapa de minería de datos y por consiguiente todavía no se puede realizar la última fase de interpretación de datos, pero si se puede hablar de resultados parciales

Como resultado de interpretar el árbol de decisión, generado por el algoritmo J48 con el conjunto de datos REP_FINAL se obtuvieron las reglas de clasificación más representativas, de un total de 120 reglas analizadas a través de la herramienta WEKA, quien establece una asociación directa entre variables; para este caso preguntas con la misma cantidad de egresados a combinar, además de ser las reglas con mayor cruce de preguntas enfocadas hacia las necesidades con un nivel de confiabilidad mayor al 75%, las cuales se muestran en la tabla 1.

Tabla 1.

REGLA	% SOPORTE	%CONFIANZA	REGISTROS POR REGLA
Formación profesional	66.98%	100%	71
Desempeño como empleado maneja inglés	50.94%	100%	54
Formación profesional, manejan inglés, no han recibido distinción	48.11%	100%	51
Formación profesional, laboran en Pasto	44.34%	100%	47
Formación profesional, manejan idioma inglés e ingreso mensual 1 a 2 smmlv	36,79%	100%	39
Formación profesional, maneja idioma inglés y programa favorece proyecto de vida	30.18%	100%	32

El análisis de estas reglas dará como resultado la interpretación del cruce de preguntas de donde se obtendrán las conclusiones que ayuden a la institución a decidir mejor sus contenidos en el programa de Ingeniería de Sistemas, al igual que identificar las necesidades de los egresados en su vida laboral.

5. Referencias

- Adamo, J. (2001). Data mining for association rules and sequential patterns: Sequential and parallel algorithms. New York (USA): Springer-Verlag.
- Fundación Universitaria Católica Del Norte. (2016). Impacto del egresado del Programa de Ingeniería Informática y su desempeño con respecto a la industria del software. Consultado el 10 de marzo de 2017 en <http://www.ucn.edu.co>.
- García, D. (s.f.). Manual de Weka. Consultado 15 de mayo de 2017 en <http://www.xmarks.com/s/site/www.metaemotion.com/diego.garcia.morate/download/weka.pdf>.

- Hall, M., Frank, E. and Witten, I. (2011). Practical data mining: [Tutorials]. Consultado el 15 de mayo de 2017 en <http://www.micai.org/2012/tutorials/Weka%20tutorials%20Spanish.pdf>.
- Hernández, J., Ramírez, M. y Ferri, C. (2005). Introducción a la Minería de Datos. Madrid (España): Pearson Prentice Hall. pp. 651.
- Timarán, R. & Jiménez, J. (2015). Extracción de perfiles de deserción estudiantil en la Institución Universitaria CESMAG. Ivestigium IRE: Ciencias Sociales y Humanas, Vol. 6, pp. 30-44.
- Wang, M., Iyer, B. and Scott, J. (1998). Scalable Mining for Classification Rules in Relational Databases. Consultado el 10 de mayo de 2017 en http://www.research.ibm.com/people/m/minwang/publications/classification_journal.pdf.

Sobre autores

- **José María Muñoz Botina.** Ingeniero de Sistemas, Magister (en curso) en Educación en Universidad de Nariño, Especialista en Redes y Servicios Telemáticos de la Universidad del Cauca. Profesor tiempo completo. Grupo de investigación Tecnofilia. jmunoz@iucesmag.edu.co
- **Carlos Fernando González Guzmán.** Ingeniero de Sistemas. Magister (en curso) en Docencia Universitaria en Universidad de Nariño, Especialista en Gerencia Informática (en curso) en Corporación Universitaria Remington. Profesor tiempo completo. Grupo de investigación Tecnofilia. cgonzalez@iucesmag.edu.co
- **Javier Alejandro Jiménez Toledo.** Ingeniero de Sistemas. Doctor en computación (en curso) en Universidad del Cauca. Magister en Computación en la Universidad del Cauca. Especialista en Docencia Universitaria en Universidad Cooperativa de Colombia. Profesor tiempo completo. Director Grupo de investigación Tecnofilia. jajimenez@iucesmag.edu.co

Los puntos de vista expresados en este artículo no reflejan necesariamente la opinión de la Asociación Colombiana de Facultades de Ingeniería.

Copyright © 2017 Asociación Colombiana de Facultades de Ingeniería (ACOFI)