

Aplicación de tecnologías inteligentes de explotación de información para el análisis de perfiles de tesis de las carreras de grado de Informática de la Universidad de Morón

Marisa Panizzi¹, Iris Sattolo¹, Oscar Bravo¹, Javier Lafont¹, Nicolás Armilla¹

¹Facultad de Informática, Ciencias de la Comunicación y Técnicas Especiales

Universidad de Morón - Argentina

marisapanizzi@outlook.com; iris.sattolo@gmail.com;

oscarbravo2006@gmail.com; lafontjavier@hotmail.com; nicolasarmilla@hotmail.com

Resumen

En esta comunicación se comparte un Proyecto de Investigación que se enmarca en la línea de investigación de Informática Aplicada y Nuevas Tecnologías según la convocatoria 2017 de nuestra Universidad. Se emplearán tecnologías inteligentes de explotación de información para el análisis de los perfiles de los tesis de las carreras de grado de informática (Licenciatura en Sistemas e Ingeniería en Informática) que se imparten en la Universidad de Morón. En una primera instancia se diseñará un instrumento de recolección de datos, se realizarán las pruebas de validez, fiabilidad y consistencia del instrumento como su prueba en campo. Para el desarrollo metodológico del proyecto de explotación de información, se utilizará la metodología CRISP-DM. Se realizará la evaluación de distintos ambientes de trabajo para la realización del proceso de minería de datos y en función de la evaluación se optará por la más conveniente dependiendo de los modelos a construir. A partir del conocimiento obtenido, las cátedras reflexionarán y elaborarán un conjunto de recomendaciones, las cuales permitirán redefinir las líneas de investigación a proponer a los alumnos, brindar el apoyo académico necesario para que los alumnos concluyan sus trabajos de fin de carrera, identificar áreas de interés de los alumnos para diseñar una oferta de cursos de posgrado y cursos de extensión entre otras.

Abstract

In this communication a Research Project is shared that is part of the research line of Applied Computing and New Technologies according to the 2017 call of our University. Information Mining (also known as Knowledge Discovery Process) will be used for the analysis of the profiles of the students of the computer science degree programs (Bachelor of Systems and Computer Engineering) taught at the University of Morón. In the first instance a data

collection instrument will be designed, the tests of validity, reliability and consistency of the instrument will be carried out as its field test. For the methodological development of the information exploitation project, the CRISP-DM methodology will be used. The evaluation of different work environments will be carried out to carry out the data mining process and, depending on the evaluation, the most convenient one will be chosen depending on the models to be built. From the knowledge obtained, the chairs will reflect and develop a set of recommendations, which will redefine the lines of research to be proposed to the students, provide the necessary academic support for the students to complete their final year projects, identify areas of interest of the students to design an offer of postgraduate courses and extension courses among others.

Palabras clave

Perfiles de tesis, carreras de informática, explotación de información, minería de datos.

1. Introducción

De la revisión de antecedentes respecto a la explotación de la información, se han encontrado los siguientes avances; la ingeniería de explotación de información entiende los procesos y las metodologías utilizadas para: ordenar, controlar y gestionar la tarea de encontrar patrones de conocimiento en masas de información [8].

La Explotación de Información consiste en la aplicación de herramientas de análisis y síntesis con el objetivo de extraer conocimiento no trivial que se encuentra distribuido en forma implícita en los datos disponibles de diferentes fuentes de información dentro de una organización [19]. Este conocimiento es previamente desconocido y puede resultar útil para

la toma de decisiones dentro de una organización [21].

De la observación de los antecedentes, se ha encontrado que la explotación de la información en el contexto académico de carreras de informática se aplica para resolver fundamentalmente los problemas relacionados a la deserción y desgranamiento.

Podemos mencionar que a nivel Nacional se encuentra trabajando el Grupo Investigación en Sistemas de Información del Desarrollo Productivo y Tecnológico de la Universidad Nacional de Lanús [10], el Departamento de Informática de la Facultad de Ingeniería de la Universidad Nacional de la Patagonia San Juan Bosco [12] y el Departamento de Formación Básica de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca [2]. Existen otros grupos de investigación como, por ejemplo, el Grupo de Estudio en Metodologías de Ingeniería en Software (GEMIS) de la UTN-Facultad Regional Buenos Aires que se encuentra trabajando en el análisis del seguimiento de perfiles de alumnos y sus percepciones respecto a la carrera Ingeniería en Sistemas de Información [1].

En el contexto Internacional, se han encontrado publicaciones previas, las cuales han demostrado que la minería de datos se puede utilizar para detectar alumnos en riesgo de deserción. En su trabajo, Luan [8] aplicó exitosamente técnicas de minería de datos para predecir qué grupos de alumnos podrían abandonar los estudios. En un trabajo relacionado, Lin [11] usó técnicas de minería de datos para optimizar los esfuerzos para retener estudiantes. Investigadores de la Universidad Estatal de Bowie, Chacon, F. et al. [6] desarrollaron un sistema basado en minería de datos que permite a dicha Institución, identificar y atender a estudiantes en riesgo de abandono de estudios.

La Minería de Datos Educativos (Educational Data Mining -EDM) como expone Romero et al. [17] es una disciplina relacionada con el desarrollo de métodos para extraer información útil a partir de los datos que se generan en los entornos educativos, y utilizarla para mejorar dicho entorno. La información así obtenida se convierte en el insumo indispensable para la toma de decisiones.

2. Formulación del problema e Hipótesis

Desde las cátedras de tesis, se ha observado que el mayor inconveniente que posee el alumno al comenzar la materia es la definición del tema, ocasionando un retraso en la finalización de sus estudios, y en algunos casos el abandono de la carrera en su última materia.

Por otro lado, las asignaturas de tesis de las carreras de grado de informática en la Universidad de Morón no cuentan con un instrumento formal que, a los docentes, les permita identificar los saberes que poseen los alumnos, su experiencia laboral, sus tiempos de dedicación a la academia, sus características personales, las líneas de investigación, desarrollo e innovación preferidas, tiempos de permanencia para el desarrollo de sus trabajos de tesis, entre otros. En la actualidad, las cátedras de tesis poseen una planilla Excel que contiene la información sobre datos del alumno (apellido, nombre, matrícula y carrera), fecha de defensa de la tesis, docente-tutor o docente-director del trabajo, título del trabajo, línea de investigación en la cual se inserta el trabajo, resumen, objetivos y futuros trabajos. Este archivo ha permitido a las cátedras llevar un registro sistemático de un total de 290 tesis, desde el año 2005 hasta el mes de marzo del 2018.

Nuestra pregunta problema es: ¿Cuáles son los patrones de comportamiento de los tesisistas de las carreras de grado de informática en la UM (Universidad de Morón)?

La inteligencia de negocio propone un abordaje interdisciplinario (dentro del que se encuentra la Informática), que tomando todos los recursos de información disponibles y el uso de herramientas analíticas y de síntesis con capacidad de transformar la información en conocimiento, se centra en generar a partir de éstos, conocimiento que contribuya con la toma de decisiones de gestión y generación de planes estratégicos en las organizaciones. La Explotación de Información es la subdisciplina de los Sistemas de Información que aporta a la Inteligencia de Negocio las herramientas para la transformación de información en conocimiento. Ha sido definida como la búsqueda de patrones interesantes y de regularidades importantes en grandes masas de información.

3. Objetivos propuestos

El Objetivo General de este proyecto de investigación consiste en analizar los perfiles de los tesisistas de las de grado de informática en la UM (Universidad de Morón).

Los objetivos específicos son:

- Identificar variables descriptoras del tesisista de las carreras de grado de informática de la UM a partir de un instrumento de recolección de datos diseñado en el marco del proyecto.
- Identificar en el conjunto de variables descriptoras, el subconjunto de variables relacionadas con la definición de perfiles de los tesisistas de las carreras de grado de informática de la UM.

- Identificar patrones de comportamiento de los tesisistas de las carreras de grado de informática de la UM.
- Formular un conjunto de recomendaciones, entre ellas, la redefinición de las líneas de investigación de las cátedras, brindar el apoyo académico necesario para que los alumnos concluyan sus trabajos de tesis, el ofrecimiento a los alumnos información de eventos científicos-tecnológicos de acuerdo con sus áreas de interés y la identificación de áreas de interés de los alumnos para diseñar una oferta de cursos de posgrado y cursos de formación profesional.

4. Metodología de trabajo

Para construir el conocimiento asociado al presente Proyecto, se seguirá un enfoque de investigación clásico propuesto por Riveros *et al.* [16] y Creswell [7] con énfasis en la producción de tecnologías según Sábato y Mackenzie [18]; identificando métodos y materiales necesarios para desarrollar el proyecto:

- **Métodos.** Se utilizarán revisiones sistemáticas y Prototipado Evolutivo Experimental (Método de la Ingeniería). Las revisiones sistemáticas propuestas por Argimón, de artículos científicos siguen un método explícito para resumir la información sobre determinado tema o problema. Se diferencia de las revisiones narrativas en que provienen de una pregunta estructurada y de un protocolo previamente realizado [3]. El prototipado evolutivo experimental expuesto por Basili, que consiste en desarrollar una solución inicial para un determinado problema, generando su refinamiento de manera evolutiva por prueba de aplicación de dicha solución a casos de estudio (problemáticas) de complejidad creciente. El proceso de refinamiento concluye al estabilizarse el prototipo en evolución [4].
- **Materiales.** Para el desarrollo del proceso de minería de datos propuestos se analizarán y utilizarán los siguientes ambientes de trabajo: Tanagra [20], WEKA [22] y Rapid Miner [14].
- **Abordaje Metodológico.** Para el Objetivo Específico 1, se propone la aplicación de una metodología que permita la construcción de un instrumento de recolección de los datos. Para esta instancia se emplearán los lineamientos propuestos por Hernández Sampieri R, *et al.* [9]. Para los Objetivos Específicos 2 y 3, se propone como metodología del proyecto de explotación utilizar CRISP-DM [5] que es una metodología de desarrollo de proyectos de explotación de información, que se encuentra definida en base a un modelo jerárquico de procesos. El foco se

pondrá en los procesos del nivel superior que son lo suficientemente genéricos como para cubrir todas las posibles aplicaciones de explotación de información.

Para catalogar las áreas de trabajo se considerarán las líneas de investigación propuestas en el Congreso Argentino de Ciencias de la Computación (CACIC), el cual es organizado por la Red de Universidades Nacionales con carreras en Informática (RedUNCI), el cual cubre temas de importancia en Ciencias de la Computación [15].

5. Resultados logrados

En el primer estadio del proyecto se ha logrado la construcción del instrumento de recolección de datos. Se ha realizado una prueba piloto del instrumento diseñado, la cual tuvo varios propósitos, someter a prueba el instrumento de recolección de datos, evaluar las condiciones de aplicación como así también los procedimientos involucrados. También ha permitido analizar si las instrucciones han sido comprendidas por los graduados (encuestados) y si los ítems funcionaron de manera adecuada. La prueba se compuso de tres instancias: una prueba piloto inicial y dos pruebas a las que denominamos prueba piloto 1 y prueba piloto 2.

En la prueba piloto inicial se convocó a un graduado y se suministró el cuestionario con el objetivo de aplicar la estrategia de validación de caso único. La prueba nos permitió evaluar el lenguaje y la redacción utilizada en el instrumento, como resultado de esta se han realizado algunas modificaciones en la redacción de algunas preguntas del cuestionario y se obtuvo una segunda versión del instrumento de recolección de datos.

Para probar la confiabilidad inicial del instrumento de recolección se desarrollaron la prueba piloto 1 y la prueba piloto 2, empleando el método de estabilidad (test – retest), esto permitió comparar las respuestas de la prueba piloto 1 y de la prueba piloto 2 y se encontró un aceptable nivel de similitud entre los datos recolectados.

6. Conclusiones

La obtención de un conocimiento más acabado sobre los tesisistas de nuestras carreras permitiría a los docentes de las asignaturas orientar más rápidamente a los alumnos en cuanto a la selección del tema de su trabajo de fin de carrera, redefinir periódicamente las líneas de investigación a ofrecer, descubrir alumnos que potencialmente les interese incorporarse como investigadores de la UM, proponer la participación en eventos científicos de acuerdo a sus intereses personales y profesionales, entre otros.

Luego de someter a prueba el instrumento construido, la cual arrojó resultados satisfactorios, el

grupo de investigación se encuentra desarrollando la recolección de los datos de los 290 graduados.

Referencias

- [1] M. Acosta, L. Straccia, L. Bernal, C. Vegega, P. Pytel, Ma. F. Pollo-Cattaneo, "Avances en el Análisis del Seguimiento de Perfiles de Alumnos y de sus Percepciones sobre la carrera de Ingeniería en Sistemas de Información mediante Explotación de Información" 3er. Congreso Nacional de Ingeniería en Informática/Sistemas de Información. (CONAIIISI 2015). Área Educación en Ingeniería. Argentina, Ciudad Autónoma de Buenos Aires. Noviembre 2015. ISBN: 978-987-1896-47-9.
- [2] H. Ahumada, Hugo R. Dip, C. Herrera, Almendra J. Leguizamón. "Minería de datos para un Sistema de alerta temprana de deserción en carreras de Ingeniería" XVII Workshop de Investigadores en Ciencias de la Computación (WICC 2015). Área de Bases de Datos y Minería de Datos. Argentina, Ciudad de Salta. Abril 2015. ISBN: 978-987-633-134-0.
- [3] J. Argimón. 2004. Métodos de Investigación Clínica y Epidemiológica. Elsevier España. 84-8174-709-2.
- [4] Basili. The Experimental Paradigm in Software Engineering. En *Experimental Software Engineering Issues: Critical Assessment and Future Directions* (Ed. Rombach, H., Basili, V., Selby, R.). Lecture Notes in Computer Science, Vol. 706. (1933). ISBN 978-3-540-57092-9.
- [5] P. Chapman, J. Clinton., R. Keber, T. Khabaza, T. Reinartz, C. Shearer, R. Wirth. CRISPDM 1.0 Step by step BIguide . Edited by SPSS.(2000).
- [6] F. Chacon, D. Spicer & A. Valbuena. Analytics in support of student retention and success. *Research Bulletin*, 3, 1-9. (2012).
- [7] J. Creswell. *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Prentice Hall. (2002). ISBN 10: 01-3613-550-1.
- [8] R. García-Martínez, P. Britos, P. Pesado, R. Bertone, M. F. Pollo Cattaneo, D. Rodríguez, P. Pytel, J. Vanrell. Towards an Information Mining Engineering. En *Software Engineering, Methods, Modeling and Teaching*. Sello Editorial Universidad de Medellín. ISBN 978-958-8692-32-6. 2011; pps 83-99, 2011.
- [9] R. Hernández Sampieri, C. Fernández Collado y L. Baptista. *Metodología de la investigación* (Cuarta ed.). México: Mc Graw Hill. (2006).
- [10] R. Lelli., R. García-Martínez, R. Charczuk, S. Martins, E. Baldizzoni. "Identificación de causales de deserción y desgranamiento de los estudiantes de la Licenciatura en Sistemas utilizando Ingeniería de explotación de Información". XVII Workshop de Investigadores en Ciencias de la Computación (WICC 2015). Área de Tecnología Informática Aplicada a la Educación. Argentina, Ciudad de Salta. Abril 2015. ISBN: 978-987-633-134-0.
- [11] S. H. Lin. Data mining for student retention management. *Journal of Computing Sciences in Colleges*, 27(4), 92-99. (2012).
- [12] C. López de Munáin., A. Sandoval, M. Torrent. "Sistemas de apoyo a la toma de decisiones. Una aplicación en el área de Gestión Universitaria". XVII Workshop de Investigadores en Ciencias de la Computación (WICC 2015). Área de Tecnología Informática Aplicada a la Educación. Argentina, Ciudad de Salta. Abril 2015 ISBN: 978-987-633-134-0.
- [13] J. Luan. Data mining and its applications in higher education. *New directions for institutional research*, 2002 (113), 17-36.
- [14] RapidMiner Management Team (S/A). RapidMinerStudio. <https://rapidminer.com/products/studio/> Disponible online en enero de 2017.
- [15] Red de Universidades con Carreras en Informática(RedUNCI). <http://redunci.info.unlp.edu.ar>
- [16] H. Riveros y L. Rosas. 1985. El Método Científico Aplicado a las Ciencias Experimentales. Editorial Trillas. México. ISBN 96-8243-893-4.
- [17] C. Romero, S. Ventura, M. Pechenizkiy & R.S. Baker. *Handbook of educational data mining*. CRC Press. Eds. (2010).
- [18] J. Sábato y M. Mackenzie. *La Producción de Tecnología*. Editorial Nueva Imagen. México. (1982). ISBN 968-429-348-8.
- [19] J. Schiefer, J. Jeng, S. Kapoor & P. Chowdhary. *Process Information Factory: A Data Management Approach for Enhancing Business Process Intelligence*. Proceedings IEEE International Conference on E-Commerce Technology, pp. 162-169, 2004.
- [20] Tanagra. Universidad de Lyon. Laboratorio ERIC (Entrepôts, Représentation et Ingénierie des Connaissances). <http://eric.univ-lyon2.fr/> Disponible online en enero de 2017.
- [21] E. Thomsen., BI's Promised Land. *Intelligent Enterprise*, 6(4), pp. 21-25, 2003.
- [22] Weka. University of Waikato. Machine LearningGroup <http://www.cs.waikato.ac.nz/ml/weka/downloading.html/> Disponible online enero de 2017.