

Tipo de artículo: Artículo original
Temática: Soluciones Informáticas
Recibido: 3/11/2019 | Aceptado: 5/03/2020 | Publicado: 02/04/2020

Solución informática para la extracción de conocimiento del módulo Resultados de la colección El Navegante

Computing solution for knowledge extraction module collection results The Navigator

Gilberto Benítez Gautier^{1*}, Frank Fernández Escalona, Noel Enamorado Selema

1 Departamento de Informática, Facultad 1, Universidad de las Ciencias Informáticas. Carretera a San Antonio de los Baños km 2 ½ Reparto Torrens. Boyeros.

*Autor para correspondencia: gbenitez@uci.cu

Resumen

Con el adelanto de la informatización en varias esferas de la sociedad y el perfeccionamiento acelerado de las Tecnologías de la Información y las Comunicaciones (TIC), el desarrollo de software se ha convertido en un elemento de gran importancia para la sociedad actual. Su impacto en la educación ha transformado los métodos tradicionales de enseñanza, aprovechando los beneficios que brindan las TIC para un mejor desarrollo del proceso de enseñanza-aprendizaje (PEA). Existen soluciones que recopilan en sus registros la interacción de los estudiantes con aplicaciones en un entorno colaborativo dedicado a este fin. El volumen de los registros almacenados dificulta la obtención de información necesaria para la toma de decisiones de los profesores que representan a los estudiantes dentro del sistema. Por tal motivo la minería de datos se presenta como una tecnología de apoyo para extraer información y conocimiento útil contenido en grandes volúmenes de datos. En este sentido, la presente investigación aborda el desarrollo de una solución informática que permita a una persona, con conocimientos mínimos de minería de datos, realizar un proceso de extracción de conocimiento (KDD, Knowledge Discovery in Databases), que sirva de ayuda en la toma de decisiones. Se utilizó como metodología para guiar el KDD a Crisp-DM, apoyados en la herramienta Weka

para la extracción de conocimientos. El proceso de desarrollo fue conducido por la metodología Proceso Unificado de Rational, transitando por los flujos de trabajo: modelo del negocio, requisitos, análisis y diseño, implementación y pruebas; obteniéndose cada uno de los artefactos propuestos.

Palabras claves: conocimiento, enseñanza-aprendizaje, minería de datos, toma de decisiones.

Abstract

With the advancement of information technology in various spheres of society and the rapid development of Information Technology and Communications (ICT), software development has become a very important element for today's society. Its impact on education has transformed traditional teaching methods, taking advantage of the benefits offered by ICTs for better development of the teaching-learning (PEA). There are solutions that gather in their records the interaction of students with applications in a collaborative environment dedicated to this purpose. The volume of records that are stored difficult to obtain information necessary for decision-making of teachers representing students within the system. Therefore data mining is presented as an enabling technology for extracting useful information and knowledge contained in large volumes of stored data. In this sense, this research addresses the development of a software solution that enables a person with minimal knowledge of data mining, through a process of knowledge extraction (KDD, Knowledge Discovery in Databases), that it will help in the decision making. Such methodology was used to guide the KDD Crisp-DM, supported by the extraction tool Weka knowledge. The development process was led by the RUP methodology, passing through workflows: business model, requirements, analysis and design, implementation and testing, obtaining each of the proposed devices.

Keywords: knowledge, teaching-learning, data mining, decision making.

Introducción

El desarrollo de software para la educación ha transformado los métodos tradicionales de enseñanza, aprovechando los beneficios que brindan las TIC, para un mejor desempeño del proceso de enseñanza-aprendizaje y la obtención de resultados superiores en el ámbito educacional. La Universidad de las Ciencias Informáticas cuenta con centros productivos dedicados a las diversas esferas del conocimiento y el desarrollo de software. El centro de Tecnologías para la Formación (FORTES), se encarga del desarrollo de soluciones informáticas que apoyen el proceso de formación. Un ejemplo lo constituye el proyecto Multisaber-Navegante cuyo objetivo es la creación de las

colecciones de software educativo Multisaber, dirigida a los estudiantes de la enseñanza primaria y El Navegante destinada a la enseñanza secundaria, del Sistema Educativo Cubano.

El Navegante, específicamente en su versión multiplataforma, está basado en el concepto de hiperentorno de aprendizaje definido como: "...un sistema informático basado en tecnología hipermedia que contiene una mezcla o elementos representativos de diversas tipologías de software educativo que, utilizados de forma oportuna y acertada, pueden contribuir al manejo de diferentes estructuras lingüísticas" (Rizzo, 2006). Uno de los componentes más importantes de esta colección es el módulo Resultados, encargado de administrar la información para medir el avance de los estudiantes a través de reportes, mostrando una serie de datos e informaciones sobre cómo se ha comportado la interacción de los alumnos con cada uno de los restantes módulos de la colección. A través de estos reportes se realizan análisis estadísticos utilizando gráficas comparativas de efectividad. De esta manera el profesor puede consultar las evaluaciones y mostrar el recorrido de sus estudiantes.

La colección El Navegante permite almacenar grandes volúmenes de registros de la interacción de los estudiantes con cada módulo de la aplicación. Dichos registros pueden contener información valiosa como patrones de comportamiento y características de los estudiantes, pero imperceptible a las personas debido a los grandes volúmenes de información. A pesar de las funcionalidades presentes en el módulo Resultados, este no cuenta con una funcionalidad que facilite la realización de un KDD, que permita extraer patrones de comportamiento y características a partir de grandes volúmenes de registros presentes en la aplicación. Esto no les permite a los educadores realizar un análisis exhaustivo de los registros de los estudiantes en la aplicación, que les posibilitaría tener un mayor conocimiento del comportamiento de los alumnos, lo cual serviría de apoyo para la preparación de los profesores en el proceso docente educativo.

Ante la problemática existente el equipo de trabajo decide centrar la investigación en darle solución a: ¿Cómo realizar un proceso de extracción de conocimiento, que permita obtener patrones de comportamiento y características a partir de grandes volúmenes de registros presentes en la colección El Navegante? El estudio de la problemática planteada se realizará haciendo énfasis en desarrollar una solución informática para la extracción de conocimiento en los registros del módulo Resultados de la colección El Navegante.

Materiales y métodos

Una vez planteada la problemática surge la necesidad del estudio de los conceptos de Módulo Resultados como punto fundamental dentro de la colección donde se encuentran almacenados los registros generados por los estudiantes en su

interacción con la aplicación. La minería de datos surge también como concepto fundamental en la presente investigación, mostrándose como un proceso para lograr extraer información relevante que se encuentra oculta en los registros de datos y es capaz de generar resultados no perceptibles a través de un análisis trivial.

Módulo Resultados

El módulo Resultados de la colección El Navegante fue desarrollado con el fin de brindar funcionalidades para facilitar un mejor seguimiento y control de la trayectoria de los estudiantes en su interacción con la aplicación. Con el desarrollo de este módulo se logra obtener información importante sobre los estudiantes que puede ser utilizada por el profesor para una mejor estructuración del proceso de aprendizaje. Se brinda la posibilidad de acceder a las evaluaciones obtenidas por los estudiantes en su interacción con la aplicación, así como, mostrar la evolución que han alcanzado en las asignaturas y contenidos especificados en el sistema. El módulo Resultados posibilita llevar un control detallado de las actividades realizadas por los estudiantes a través de cuatro reportes: Trayectoria del estudiante, Análisis de los contenidos específicos, Historial del estudiante y Análisis integral del estudiante.

Minería de datos

La minería de datos es un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos (López, y otros, 2008). Se puede definir además la minería de datos como un proceso completo de aplicar métodos basados en la informática para descubrir conocimiento oculto en grandes cantidades de información (Baena, 2006).

Se adopta como concepto de la investigación el expresado por Gregory Piatetsky-Shapiro, cuando se refieren a la minería de datos como:

“... un conjunto de técnicas y herramientas aplicadas al proceso no trivial de extraer y presentar conocimiento implícito, previamente desconocido, potencialmente útil y humanamente comprensible, a partir de grandes conjuntos de datos, con objeto de predecir, de forma automatizada, tendencias o comportamientos y descubrir modelos previamente desconocidos” (Piatetski-Shapiro, 1991).

Análisis de soluciones similares existentes

En la actualidad el desarrollo de aplicaciones informáticas para guiar los procesos de enseñanza y aprendizaje constituye uno de los grandes aportes tecnológicos a la esfera de la educación a nivel mundial. La relación existente entre las tecnologías informáticas y la educación surge desde los primeros momentos en la etapa evolutiva del software, especialmente como apoyo a los procesos de la educación a distancia, de forma tal que el software

educativo evolucione a la par de las tecnologías informáticas, posibilitando una integración más adecuada y potente que posibilite un enriquecimiento enorme de los procesos instruccionales.

Minería de datos aplicada a entornos educativos

Durante las últimas décadas, las innovaciones más importantes en los sistemas educativos están relacionadas con la introducción de nuevas tecnologías. Los sistemas educativos basados en la web constituyen una forma de instrucción asistida por computadora prácticamente independiente de una ubicación específica y plataforma de hardware. La minería de datos aplicada a estos entornos educativos ha resultado ser una herramienta muy útil para la toma de decisiones, permitiendo el análisis automático y profundo de la información generada por los estudiantes al interactuar con los sistemas educativos (Ventura, et al., 2007). Este proceso es un ciclo iterativo en el cual se establecen hipótesis, se prueban y se refinan tal como se muestra en la figura 1.

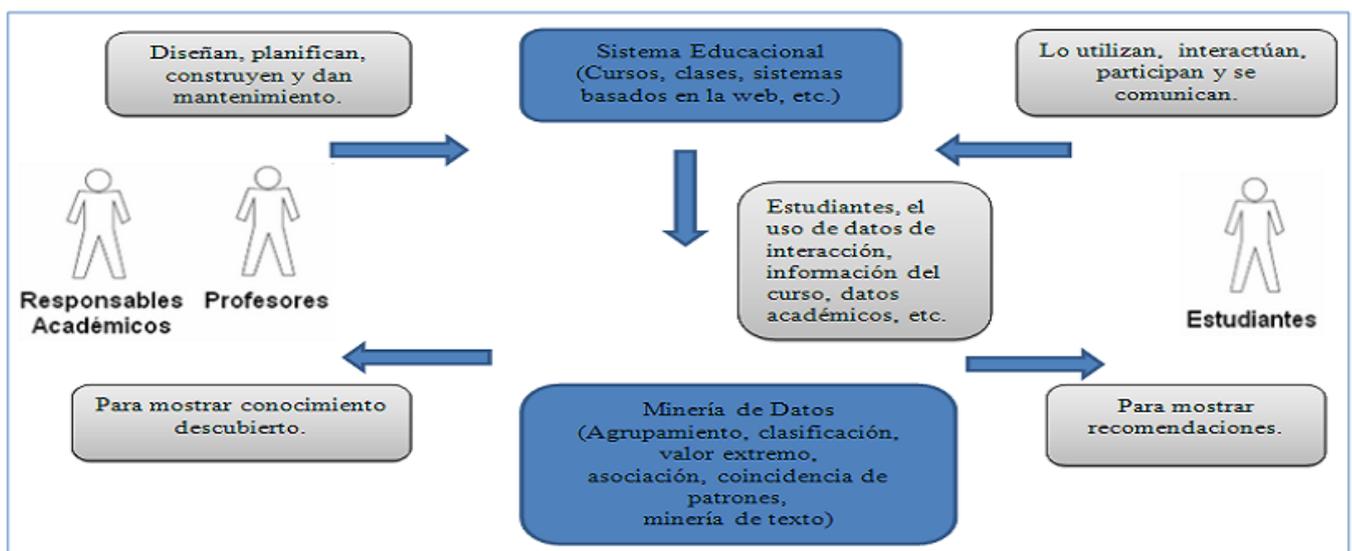


Figura 1: Ejemplo de fichero ARFF

La aplicación de técnicas de minería de datos en la educación se puede ver desde dos puntos de vista distintos (Ventura, et al., 2007):

Orientado hacia los autores: Tiene como objetivo ayudar a los profesores y/o autores de los sistemas de e-learning para que puedan mejorar el funcionamiento o rendimiento de estos sistemas a partir de la información de utilización de los alumnos. Sus principales aplicaciones son: obtener una mayor retroalimentación de la enseñanza, conocer más sobre cómo los estudiantes aprenden en la web, evaluar a los estudiantes por sus patrones de navegación, así como,

reestructurar los contenidos del sitio web para personalizar los cursos, clasificar a los estudiantes en grupos, entre otros. (Morales, et al., 2005)

Orientado hacia los estudiantes: Con el objetivo de ayudar o realizar recomendaciones a los alumnos durante su interacción con el sistema de e-learning para poder mejorar su aprendizaje. Sus principales aplicaciones son: sugerir buenas experiencias de aprendizaje a los estudiantes, adaptación del curso según el progreso del aprendiz, ayudar a los estudiantes dando sugerencias y atajos, recomendar caminos más cortos y personalizados, etc. (Ventura, et al., 2007)

Las principales aplicaciones de las técnicas de minería de datos en educación, son sistemas de personalización, sistemas recomendadores, sistemas de modificación, sistemas de detección de irregularidades, sistemas predictivos, entre otras. Dichas aplicaciones se basan en sus capacidades para: el descubrimiento de patrones de navegación regulares e irregulares, realización de clasificaciones de alumnos y de los contenidos, construcción adaptativa de planes de enseñanza, descubrimiento de relaciones entre actividades, diagnóstico incremental de los estudiantes.

Las tareas más utilizadas en la minería de datos aplicada a los sistemas de e-learning son: clasificación y agrupamiento, descubrimiento de reglas de asociación, y análisis de secuencias (Galindo, et al., 2010).

Minería de datos relacionada a Entornos Educativos en el mundo

En el mundo existen disímiles sistemas que hacen uso de técnicas de minería de datos. En un estudio realizado por Sebastián Ventura en el año 2008 (Soto, 2008) sobre la minería de datos en sistemas educativos se destacan el uso de las tareas de clasificación, regresión, clustering y asociación. Las tareas de clasificación y agrupamiento o clustering aplicadas a sistemas de e-learning permiten agrupar a los usuarios por su comportamiento de navegación, agrupar a las páginas por su contenido, tipo o acceso y agrupar los comportamientos de navegación similares. Las reglas de asociación descubren relaciones entre atributos de un conjunto de datos que superan unos determinados umbrales. Su aplicación más típica ha sido en los sistemas de comercio electrónico para informar sobre las preferencias de compra de los clientes. Su aplicación a sistemas de e-learning permite descubrir relaciones o asociaciones entre distintas páginas Web visitadas.

En el trabajo “Minería de Datos aplicada a entornos educativos: estudio e implementación de técnicas libre de parámetros” (Saiz, 2011) se emplean las tareas de clasificación y reglas de asociación con el objetivo de ofrecer a los profesores la posibilidad de extraer información y modelos de minería de datos, a partir de los datos de actividades de cursos virtuales registrados en plataformas e-learning, sin que estos tengan conocimientos de dichas tareas y las

técnicas que comprenden. En este trabajo se hace uso de los algoritmos Apriori y Predictive Apriori presentes en la herramienta Weka para la extracción de reglas de asociación.

Otro ejemplo lo constituye la utilización conjunta de clustering con otras técnicas como secuenciación, analizando el comportamiento de navegación de los usuarios para la personalización de sistemas e-learning (Miguillón, 2005). Existen otros trabajos que analizan los ficheros log de entornos de aprendizaje web utilizando técnicas de minería de reglas de asociación y filtrado, con el fin de descubrir patrones de navegación útiles y proponer un modelo de navegación (Morales, et al., 2005). El trabajo realizado por Porcel Dapozo (Dapozo, et al., 2010) en el que se hizo uso de técnicas de minería de datos que les permitió, a través de un clasificador, determinar el rendimiento académico de los alumnos de la carrera de Licenciatura en Sistemas de Información, para cumplir con esto llevaron a cabo un estudio comparando los diferentes algoritmos clasificadores disponibles en Weka, analizando los resultados que se obtuvieron de la aplicación de cada uno de ellos. Utilizaron la base de datos que alojaba toda la información sistematizada de los alumnos de la Facultad de Ciencias Exactas de la UNNE tomando particularmente en cuenta los siguientes datos:

- La cantidad de exámenes finales rendidos (número de intentos).
- La cantidad de exámenes finales aprobados.
- La cantidad de exámenes finales reprobados.

En base a esos valores, ellos generaron tres categorías que sirvieron para identificar a los alumnos:

- En el año de ingreso no rindieron ninguna materia.
- Los que sí rindieron pero no aprobaron ninguna.
- Los que rindieron y aprobaron por lo menos una materia.

Minería de datos relacionada a Entornos Educativos en Cuba

En Cuba existen pocos trabajos de minería de datos relacionadas a entornos educativos. Uno de ellos es el realizado en la UCI (Díaz, et al., 2009), cuyo principal resultado es el uso de técnicas de agrupamiento y asociación guiado por la metodología Crisp-DM, para determinar patrones entre la procedencia del origen social y los resultados académicos en los estudiantes de la UCI.

Existe otro trabajo que fue desarrollado en el Instituto Superior Politécnico José Antonio Echeverría CUJAE (Acosta, et al., junio, 2007), con el empleo de técnicas de clasificación para predecir el resultado de los estudiantes en su primer año, en dependencia de sus características de ingreso y la especialidad que cursaban. Otro trabajo realizado en

la CUJAE (Sarasa, 2008) fue desarrollado por el Ing. Raycos Brito Sarasa. Su principal objetivo es descubrir conocimiento en las bases de datos de la universidad y ayudar a la toma de decisiones de la Vicerrectoría Docente a través del uso de técnicas de minería de datos, auxiliándose en la metodología Crisp-DM y la herramienta Weka. En dicho trabajo se emplean tareas de asociación, agrupamiento y clasificación con árboles de decisión, para agrupar a los estudiantes atendiendo a sus características, determinar cuáles son las características de los estudiantes que pueden influir en sus resultados docentes de cada año de estudio y determinar cómo influyen las características de los estudiantes en el promedio de cada año de estudio.

En un trabajo realizado en la CUJAE por varios autores (Suárez, et al., 2008), se describen las actividades realizadas siguiendo el modelo propuesto por la metodología Crisp-DM y enfocados en la herramienta de análisis Weka para la extracción de conocimientos de sus bases de datos.

Resultados y discusión

Durante el desarrollo de la investigación no se encontró alguna solución que se adaptara a las características del sistema a implementar. El estudio realizado demostró que Weka se puede utilizar como una de las herramientas de apoyo a los procesos de extracción de conocimientos más efectiva y a Crisp-DM como metodología para guiar las distintas fases del proceso, presente en la mayoría de los trabajos analizados. Se identificaron las tareas de agrupamiento, clasificación y reglas de asociación como las más comunes en los sistemas educacionales. La metodología Crisp-DM estructura el ciclo de vida de un proyecto de Data Mining, que interactúan entre ellas de forma iterativa (no lineal) durante el desarrollo del proyecto.

Las herramientas de minería de datos constituyen para los mineros unos potentes instrumentos para facilitar los procesos de minería y a su vez agilizan y facilitan estos procesos. Weka (Waikato Environment for Knowledge Analysis) es un conjunto de librerías Java para la extracción de conocimientos desde bases de datos. Es un software desarrollado bajo la licencia GPL, código abierto e incluye interfaz gráfica compuesta por diversos entornos, desarrollada por un grupo de investigadores de la Universidad de Waikato en Nueva Zelanda en 1993. Se puede acceder a todas las funcionalidades de minería de datos a través de una interfaz de líneas de comandos (CLI), de manera que las aplicaciones puedan sacar el máximo partido de las funciones disponibles. Constituye un entorno de experimentación de análisis de información, formado por una serie de paquetes de código abierto con diferentes técnicas de pre-procesado, clasificación, asociación, y visualización de datos (Waikato, 2010). Se encuentra pública en los repositorios de los sistemas operativos Unix y por tanto se tiene acceso a su código fuente.

Para el desarrollo de la solución se identificaron las diferentes tecnologías y herramientas tanto del lado del cliente como del servidor que serán el soporte de la solución resultante de la presente investigación. Se identificó como frameworks del lado del cliente: JQuery 1.5 y Symfony 1.4.3 como framework del lado del servidor. Como herramienta para el modelado del diseño del sistema se seleccionó UML 1.4, como herramienta CASE Visual Paradigm 8.0. Como lenguajes del lado del cliente se utiliza HTML 5, CSS 3 y JavaScript 1.5 y como lenguajes del lado del servidor PHP 5.3. El servidor web seleccionado es Apache 2.0, como Sistema de Gestión de Base de Datos PostgreSQL 8.4 y como Entorno de Desarrollo Integrado NetBeans 7.2.

Para guiar el proceso de extracción de conocimiento se utilizó la metodología Crisp-DM. A continuación se detalla el empleo de cada una de las fases que plantea esta metodología:

La primera fase, **Comprensión del negocio**, se centra en entender los objetivos y requerimientos del proyecto desde una perspectiva de negocio, plasmando todo esto en una definición del problema de minería de datos (Orallo, y otros, 2004).

Establecimiento de los objetivos de negocio: Se pretende incluir entre los reportes generados por el módulo Resultados de la colección El Navegante, uno que sea capaz de generar información importante para los profesores tomando como punto de partida los datos almacenados en dicho módulo y como base la aplicación de algoritmos de minería de datos.

Establecimiento de los objetivos de la minería de datos: Se pretende con la aplicación de técnicas y algoritmos de minería de datos la obtención de reglas y patrones que puedan ser utilizados por los profesores para el perfeccionamiento del proceso de aprendizaje.

La segunda fase, **Comprensión de los datos**. Esta fase se centra en recopilación y familiarización con los datos, identificar los problemas de calidad y ver las primeras potencialidades o subconjuntos de datos que puedan ser interesantes analizar (Orallo, y otros, 2004).

Recopilación y descripción de los datos: Los datos que serán analizados se encuentran almacenados en el módulo Resultados de la colección, que son muestra de la interacción de los estudiantes con el sistema. Se analiza la información de las diferentes tablas de la base de datos, seleccionando la que más aspectos significativos posea y que sean más interesantes para analizar.

La colección El Navegante cuenta con cinco módulos que pueden ser accedidos por los estudiantes, por lo que resulta interesante analizar la interacción de estos con cada uno de los módulos, con el fin de encontrar patrones de navegación y otras informaciones importantes a tener en cuenta por los profesores a la hora de estructurar los elementos que integran la colección. A su vez, el módulo Contenido es el eje central del hiperentorno pues tiene la aspiración de cubrir el currículo de las asignaturas de la colección. Es de gran interés para los profesores controlar el

acceso de los estudiantes a este módulo ya que contiene la información necesaria para adquirir los conocimientos para poder interactuar con los restantes módulos, es decir, controlar cuáles son los temas más visitados y así estructurar los conocimientos con el fin de perfeccionar el proceso de aprendizaje de los estudiantes.

La tercera fase, **Preparación de los datos**. Esta fase cubre todas las actividades necesarias para construir la colección de datos que finalmente será minada a partir del grupo inicial. La herramienta Weka posee varias formas para definir el origen de los datos; especificándole la dirección de un fichero con extensión .ARFF localizado en el propio ordenador, o especificándole la dirección de este fichero a través de una dirección URL, o finalmente definiéndole la dirección de una base de datos existente y detallando los parámetros para la conexión.

Fichero ARFF: Los datos que se definen para el trabajo con la herramienta Weka se incluirán en un archivo de extensión .ARFF que se construirá siguiendo la estructura definida por el programa Weka y con los datos más importantes para la aplicación de los algoritmos de minería de datos.

Estructura del fichero ARFF: El fichero ARFF está compuesto fundamentalmente por tres partes: cabecera, declaración de atributos y sección de los datos.

En la cabecera se define el nombre de la relación siguiendo como estructura:

@relation <nombre_relación>.

En la sección de declaración de los atributos se declaran todos los atributos que contendrá el archivo, unido al tipo de datos del atributo. Se definirán los atributos de la siguiente forma:

@attribute <nombre_atributo> <Tipo_atributo>

Los tipos de atributos que define Weka son:

STRING: Para expresar cadenas de texto.

INTEGER: Para expresar los datos que representan números enteros.

DATE: Expresa fechas, para ello debe ir precedido de una etiqueta de formato entrecomillada la cual está compuesta por caracteres separadores y unidades de tiempo: dd Día, MM Mes, yyyy Año, HH Horas, mm Minutos, ss Segundos.

La última sección es donde se declaran las instancias que forman la relación. El inicio de la sección viene dado de la forma @data. Los atributos que forman las instancias se separan por coma y las instancias por saltos de líneas.

En la figura se muestra un ejemplo de un fichero ARFF que cumple con la estructura propuesta por Weka.

```
@relation Estudiantes_navegante

@attribute nombre STRING
@attribute apellido STRING
@attribute edad INTEGER
@attribute fecha_nacimiento date "dd-MM-yyyy"
@attribute evaluaciones nominal {B,R,M}

@data
Gilber,Benitez,23,8-5-1989,B
Yunior, Baez, 25,11-12-1989,B
Lili,Gautier,22,20-5-2004,M
Yadira,Diaz,7,20-1-2012,R
Yelena,Rodriguez,15,22-8-2007,M
```

Figura 2: Ejemplo de fichero ARFF

La cuarta fase es el **Modelado**. En esta fase se aplican las técnicas de modelado o de minería de datos seleccionadas teniendo en cuenta las características de los datos o vista minable previamente seleccionada. También se incluye una evaluación desde el punto de vista de precisión de los modelos generados. Es uno de los aspectos más importantes dentro del proceso debido a que constituye la base sobre la cual se van a sustentar las posteriores actividades. Se selecciona la técnica de minería que más se ajusta a las características de los datos con los que se trabajan y además teniendo en cuenta el tipo de problema a resolver y los resultados que se esperan. Como tareas del proceso de minería de datos se seleccionaron agrupamiento (clustering) y reglas de asociación.

La quinta fase es **Evaluación**. Al llegar a esta fase se analizarán los modelos seleccionados para medir su precisión y rendimiento para decidir si cumplen apropiadamente con los objetivos del negocio propuestos en la primera fase. Para lograr esto la evaluación se apoya de dos subfases: la evaluación de los resultados (evaluación de los resultados de la aplicación de la minería de datos, modelos aprobados) y la revisión del proceso (Orallo, y otros, 2004).

Evaluación de los resultados: Una vez aplicados los modelos se obtuvieron resultados que aportan información interesante a los profesores acerca del acceso de los estudiantes a los distintos temas del módulo Contenidos y sobre la interacción de los mismos con los restantes módulos de la colección. A través de estas informaciones los profesores pueden determinar cuáles son los temas más visitados y cuáles son los menos visitados, lo que brinda la posibilidad de incorporar a los temas que más visitan los alumnos aquellos conocimientos que por su importancia resultan relevantes y necesarios a tener en cuenta en el proceso de aprendizaje. Además resulta interesante analizar el acceso a los distintos módulos con el fin de estructurar el funcionamiento de la colección en general.

Revisión del proceso: Con la fase de la comprensión del negocio, que constituye su etapa inicial, centrada en entender los objetivos y metas fundamentales del negocio se establece como objetivo fundamental del negocio, generar un reporte que sea capaz de exponer de forma clara y sencilla elementos importantes y de relevancia almacenados en el módulo Resultados de la colección El Navegante.

Se dio paso a una de las fases más importantes del proceso, la selección de los datos, que por su relevancia se deben tener en cuenta a la hora de realizar el proceso; el acceso de los estudiantes al módulo Contenidos y los restantes módulos de la colección.

Una vez seleccionados los datos se procede a la preparación de estos para poder ser interpretados correctamente por la herramienta de minería de datos, transformándolos siguiendo la estructura del fichero de extensión ARFF que propone Weka.

En la siguiente fase se aplican las tareas y algoritmos de minería de datos seleccionados teniendo en cuenta las características de los datos que se eligieron. Se decide la realización de tareas basadas en agrupamientos y reglas de asociación. Se seleccionaron como algoritmos para la aplicación de estas tareas, relacionados con el agrupamiento a K-medias y DBScan, y por parte de las reglas de asociación los algoritmos Apriori y PredictiveApriori.

La sexta y última fase es **Despliegue**. En esta fase es donde se trata de explotar las potencialidades y garantías obtenidas de la aplicación de los modelos y la forma más idónea de integrar los conocimientos descubiertos a los sistemas de las organizaciones interesadas. Se utilizan los datos generados por los modelos para mostrar información más importante relacionada con los estudiantes. Se muestra los resultados de las tareas de agrupamiento a través de gráficos que reflejan los distintos grupos que resultaron de la aplicación de los algoritmos y también las reglas generadas por las tareas de reglas de asociación, todo esto de la forma más representativa posible.

Pruebas de software

Una vez desarrollado el software, es necesario realizar pruebas con el objetivo de minimizar la ocurrencia de errores y garantizar la calidad del producto. Para determinar si el sistema cumple con las expectativas del usuario final se deben realizar pruebas de software, definiendo niveles, tipos de pruebas, métodos y técnicas, para verificar que todos los requisitos fuesen implementados correctamente. Las descripciones de las pruebas realizadas se encuentran documentadas en el Plan de prueba del proyecto Multisaber-Navegante describiendo el propósito, alcance y visión de los tipos de pruebas identificados para el desarrollo de la colección.

La siguiente tabla relaciona con más detalles la estrategia de prueba diseñada:

Tabla 1: Estrategia de pruebas

Niveles de prueba	Tipos de prueba	Métodos	Técnicas
Integración	Funcionalidad (función)	Caja negra	Partición de equivalencia y (AVL)
Sistema	Funcionalidad (seguridad)	Caja negra	Partición de equivalencia

Sistema	Usabilidad (usabilidad)	Caja negra	Partición de equivalencia
----------------	-------------------------	------------	---------------------------

A partir del desarrollo de los casos de prueba se realizaron tres iteraciones, guiados por las técnicas expuestas anteriormente, se identificaron un total de 18 no conformidades (NC), las cuales fueron resueltas garantizando así el correcto funcionamiento del sistema. A continuación se muestra una tabla que resume las iteraciones realizadas y los tipos de no conformidades encontradas.

Tabla 2: No conformidades encontradas

Categorías	Iteración 1	Iteración 2	Iteración 3
Funciones incorrectas o faltantes	4	2	
Errores de interfaz	3	2	1
Errores en estructuras de datos	1		
Errores de comportamiento	3	2	

Concluida cada iteración de pruebas se analizaron, por parte del equipo de desarrollo, las (NC) encontradas para determinar cuáles realmente constituyeron defectos del sistema según la descripción de los casos de uso. Las (NC) generadas han sido solucionados en su totalidad, por tanto no afectan el desarrollo de la aplicación, por lo que se encuentra lista para ser desplegada.

Impacto de la Solución

Visión de Profesor:

La funcionalidad desarrollada permite a los profesores obtener una mayor retroalimentación de la enseñanza, conocer más sobre cómo los estudiantes aprenden en la plataforma. Les brinda además, una herramienta de apoyo para definir los métodos más idóneos de evaluación de los estudiantes, basándose en sus patrones de navegación dentro de la plataforma y así obtener un resultado más exacto del estado de los estudiantes. Les brinda a los profesores criterios para estructurar los contenidos de la plataforma en función de un mayor aprendizaje por parte de los estudiantes. Permite clasificar a los estudiantes teniendo en cuenta los resultados alcanzados en los ejercicios realizados, con la finalidad de poder hacer un trabajo diferenciado con los estudiantes. Permite identificar las posibles causas de malos resultados en las diferentes evaluaciones realizadas por los estudiantes, así como, posibles soluciones que pudieran mitigar estas con el fin de ayudar y perfeccionar el Proceso de Enseñanza y Aprendizaje.

Visión del Estudiante:

La funcionalidad desarrollada brinda a los estudiantes recomendaciones durante su interacción con el sistema que los ayuda a mejorar su proceso de aprendizaje. Facilita a los estudiantes orientaciones y buenas experiencias de aprendizaje basadas en los datos procesados de otros estudiantes. Recomienda caminos más cortos y personalizados con la finalidad de alcanzar los objetivos.

Conclusiones

Se identificaron las tareas agrupamiento y extracción de reglas de asociación, apoyado en la herramienta Weka y guiado por la metodología Crisp-DM para el minado de los datos en el desarrollo de la solución. La utilización de las metodología Crisp-DM para guiar el proceso de minado y RUP para guiar el proceso de desarrollo de software garantizaron el desarrollo exitoso de la aplicación. La funcionalidad desarrollada posibilita a los profesores realizar de forma sencilla un proceso de extracción de conocimiento que les permite identificar patrones de comportamiento de los estudiantes, posibilitando una mejor estructuración de los contenidos, con la finalidad de perfeccionar el proceso de aprendizaje de los estudiantes y mejorar sus resultados docentes. La solución propuesta fue validada funcionalmente utilizando pruebas de caja negra, basadas en los casos de pruebas desarrollados, lo que permitió erradicar deficiencias encontradas.

Referencias

- Acosta, Rolando y Vázquez, Luis. junio, 2007. *Obtención de patrones y reglas en el Sistema Docente del Instituto Superior Politécnico José A. Echevarría (CUJAE) utilizando Minería de Datos*. Habana, Cuba : s.n., junio, 2007.
- Baena, Domingo Savio Rodríguez. 2006. *Análisis de datos de Expresión Genética mediante técnicas de Biclustering*. Sevilla, España : s.n., 2006.
- Dapozo, G y Porce, E. 2010. *Aplicación de minería de datos con una herramienta de software libre en la evaluación del rendimiento académico de los alumnos de la carrera de Sistemas de la FACENA-UNNE*. Argentina : s.n., 2010.
- Díaz, Ernesto González, y otros. 2009. *Obtención de patrones y reglas en el proceso académico de la Universidad de las Ciencias Informáticas utilizando técnicas de minería de datos*. Habana, Cuba : s.n., 2009.
- Galindo, Álvaro Jiménez y Álvarez García, Hugo. 2010. *Minería de Datos en la Educación*. Madrid, España : s.n., 2010.
- López, César Pérez y Santín González, Daniel. 2008. *Minería de Datos Técnicas y Herramientas*. s.l. : Paraninfo, 2008.

- Miguillón, Julia. 2005. *E-learning Personalization based on Itineraries and Long-term Navigational Behavior*. Barcelona : s.n., 2005.
- Morales, Cristobal Romero, Ventura Soto, Sebastián y Hervás Martínez, Cesar. 2005. Estado actual de la aplicación de la minería de datos a los sistemas de enseñanza basada en web. Córdoba : s.n., 2005.
- Orallo, José Hernández, Ramirez Quintana, Maria José y Ferri Ramírez, César . 2004. *Introducción a la Minería de Datos*. Madrid : Pearson, 2004.
- Piatetski-Shapiro. 1991. *Knowledge discovery in databases: an overview*. California : s.n., 1991.
- Rizzo, César Labañino. 2006. Maestría en Ciencias de la Educación. La Habana : s.n., 2006.
- Saiz, Diego García. 2011. *Minería de Datos aplicada a entornos educativos: estudio e implementación de técnicas libre de parámetros*. Cantabria, España : s.n., 2011.
- Sarasa, Raycos Brito. 2008. *Minería de Datos aplicada a la Gestión Docente*. Habana, Cuba : s.n., 2008.
- Soto, Sebastián Ventura. 2008. *Minería de Datos en sistemas educativos*. Córdoba : s.n., 2008.
- Suárez, Alejandro Rosete, Brito Sarasa, Raycos y Acosta Sánchez, Rolando. 2008. *Desarrollo de un proceso de KDD en el ámbito docente: Preparación de los datos*. La Habana : s.n., 2008.
- Ventura, Sebastian y Romero, Cristobal. 2007. *Educational data mining: a survey from 1995 to 2005. Expert Systems with Applications*. 2007. 33:1.
- Waikato, Department of Computer Science The University of. 2010. [En línea] 2010. [Citado el: 13 de 1 de 2013.] <http://www.cs.waikato.ac.nz/>.