

Algoritmos de Data Mining aplicados la enseñanza basada en la Web.

Ing. Cynthia Lorena Corso.

Sofía Lorena Alfaro.

Universidad Tecnológica Nacional, Departamento de Sistemas de Información,

Córdoba, Argentina, X5016ZAA

cynthia@bbs.frc.utn.edu.ar

soft@bbs.frc.utn.edu.ar

Abstract

Educational platform based on the use of web pages, presents an alternative for addressing the problems and needs of education.

The objective of this work is to present a compilation of research in data mining related to education based on virtual learning, specifically related to the alternative algorithms and software for its implementation.

Of the existing techniques for applying data mining, our work focuses on data mining mainly associated with virtual learning. The techniques that are implemented in this context are the algorithms for classification and clustering, and discovery of association rules and sequence patterns. Although other techniques are also described on the use of virtual platforms in relation to the educational field, such as web mining.

With respect to software, we have focused on Weka, one of the packages used to work with learning algorithms and whose license is open source.

Keywords: Data Mining, Learning Web, algorithms, Weka, Open Source.

Resumen

Las plataformas educativas basada en el uso de páginas web, presenta una alternativa para enfrentar las problemáticas y necesidades de la educación.

El objetivo de este trabajo es realizar una recopilación del estado presente en la investigación de data mining vinculado con la educación basada en el enseñanza virtual, específicamente relacionada con los algoritmos y una alternativa de software para su implementación.

De las técnicas existentes para aplicar data mining, nuestro trabajo se enfoca sobre data mining vinculado principalmente con la enseñanza virtual. Las técnicas que se implementan en este contexto son los algoritmos de clasificación y agrupamiento, descubrimiento de reglas, asociación y patrones de secuencia. Aunque también se califican otras técnicas relacionadas con el uso de plataformas virtuales en relación al ámbito educativo, como web mining.

Con respecto al software nos hemos enfocado en destacar las ventajas de Weka, uno de los paquetes más utilizados a la hora de trabajar con algoritmos del aprendizaje y cuya licencia es open source.

Palabras claves: Data Mining, Enseñanza Web, Algoritmos, Weka, Open Source.

1. INTRODUCCION

Las TICs han provocado un gran cambio en todos los niveles, el educativo no escapa a esta realidad. El uso de Internet ofrece a docentes como a estudiantes facilidad a la hora de acceder a nuevo conocimiento y potencia de manera significativa los canales de comunicación permitiendo el intercambio de opiniones y experiencias con personas de todo el mundo.

En consecuencia, se puede concluir que el paradigma de aprendizaje ha cambiado y en consecuencia surge la necesidad de adaptarse al nuevo paradigma educativo, para que los beneficios de este medio no se conviertan en un obstáculo. Las plataformas educativas basadas en el uso de páginas Web se presentan como buenas alternativas para encausar la problemática y las necesidades del contexto informático de la educación debido a la característica de hipertextualidad que comparten, a la capacidad multimedia y a las facilidades que brinda para la interactividad y el trabajo cooperativo, convirtiéndose en potenciales medios multiplicadores de ofertas académicas. [10]

[9] El uso de material para ser distribuido a través de Internet, incentiva a los alumnos a realizar trabajos cooperativos y colaborativos en pro de alcanzar los objetivos propuestos en sus actividades académicas. Sobre esta base, los estudiantes tienen la oportunidad de trabajar en forma colaborativa con otros estudiantes y obtener

beneficios de receptar las sugerencias e inquietudes de otros estudiantes o grupos de estudiantes sobre el mismo tema.

En este punto es importante destacar la importancia que adquiere el diseño de material para ser distribuido en la web ya que debe ser diseñado para ser usado de manera interactiva y ser capaz de brindar diversos tipos de información visual; usada para fortalecer el proceso de enseñanza y aprendizaje tanto en las modalidades presenciales como semipresenciales. Pero en la realidad muchos docentes no tienen conocimiento o información sobre los parámetros “recomendables” para obtener ventajas significativas en el uso de sistemas de enseñanza basada en la web.

El creciente uso de los sistemas de enseñanza basado en la web, ha avanzado de manera significativa, lo que ha llevado al estudio del comportamiento relacionado con el uso de los sistemas de enseñanza basados en web, como por ejemplo que recursos los alumnos acceden con mayor o menor frecuencias, cantidad de accesos etc.

Existen diversos trabajos de investigación que han aplicado técnicas de data mining, que intentan revelar aspectos como patrones de navegación, clasificación o agrupación de los alumnos teniendo en cuenta frecuencia de acceso y otros aspectos relacionados con la utilización de plataformas educativas.

El objetivo de este trabajo es presentar una recopilación de la investigación de técnicas de Data Mining aplicadas al ámbito educativo y en la enseñanza basada en la web. La importancia de este trabajo radica en brindar, a los docentes que administran y utilizan estas plataformas educativas, conocimiento relacionado con el uso por parte de los alumnos, para mejorar su administración y uso. Al adoptar medidas correctivas, como reestructurando los contenidos, favorece a los alumnos en el proceso de aprendizaje incrementando una mayor comprensión y accesibilidad.

2. DATA MINING

Existen diversas definiciones de data mining, una de ellas la considera como “extracción no trivial de información implícita, previamente no conocida y potencialmente útil desde los datos”. (William J. Frawley, Gregory Platestky-Shapiro and Christopher Matheus)

Si nos remitimos a una definición más técnica, se considera como un proceso analítico diseñado para explorar grandes cantidades de datos en búsqueda de patrones consistentes y/o relaciones sistemáticas entre variables, y luego validar el hallazgo aplicando los patrones encontrados a nuevos conjuntos de datos.

En conclusión lo que en verdad hace el data mining, es reunir los beneficios de varias áreas como la estadística, la inteligencia artificial, las bases de datos y el procesamiento masivo, usando principalmente como materia prima las bases de datos.

Es por ello que la tecnología de data mining puede generar nuevas oportunidades de negocios, brindando las siguientes capacidades:

Predicción de tendencias y comportamientos: Este tipo de modelo, toma como base una hipótesis definida por el usuario y posteriormente se efectúa una prueba de validación contra los datos, para evaluar la calidad de estos modelos, tales como simplicidad e interoperabilidad etc.

Modelos de descubrimiento: Este tipo de modelos intentan encontrar relaciones y patrones de comportamiento en el conjunto de datos para ofrecer conocimiento sobre un problema concreto. [6,8]

3. FUNDAMENTOS SOBRE WEB MINING

Una de las ramas del data mining, aplicadas a los entornos virtuales es el web mining.

Según Sankar, Varun, Mitra [15] “El Web Mining puede ser definido en grandes rasgos como el descubrimiento y análisis de información útil desde el denominado World Wide Web”.

La información a analizar proviene de Internet, y sus fuentes de origen pueden ser clasificadas desde la perspectiva de los servidores web, desde el lado de los clientes, desde los servidores proxy utilizados para mejorar el rendimiento, o desde los equipos de comunicación de datos. Esta información es obtenida desde bases de datos disponibles en el web que contienen información de negocios o datos consolidados de una empresa; es evidente por tanto que dependiendo de la fuente, la información a recolectar cambia.

Las técnicas de web mining analizan y procesan estos archivos logs, para lograr conocimiento significativo.

A continuación en la figura se visualiza cuales son las principales tareas del web mining:

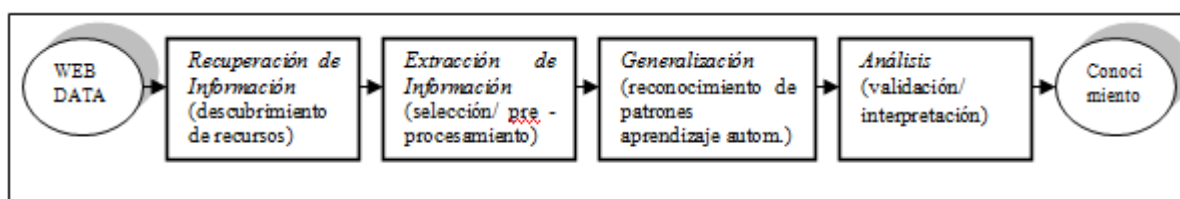


Figura 1. Etapas del Web Mining

Los datos obtenidos de los archivos “web log” tienen una diversidad de formatos, es por ello que estos datos que proceden del cliente, pueden ser recopilados con el empleo de un agente remoto. En general estas técnicas están enfocadas a lograr una mejor identificación de los clientes, pero generar un mayor consumo de recursos en el equipo usuario afectando el rendimiento del mismo.

[12]Una vez recolectados los datos de los archivos “web log” se procede a realizar una tarea de limpieza que incluye los siguientes pasos: filtrado, identificación de usuarios y determinación de sesiones.

Filtrado: El objetivo es eliminar registros que no son necesarios, como los log de solicitudes de imágenes que son solamente parte de la página HTML que las contiene, por lo tanto los accesos a ficheros cuyas extensiones son ‘.gif’, ‘.jpeg’, ‘.png’, también son eliminadas. En una primera etapa el filtrado da como resultado datos a nivel de página web.

Identificación de usuarios: En este paso se lleva a cabo dos niveles de identificación, en el primer nivel se identifican las peticiones de páginas realizadas por el mismo usuario durante una visita. La segunda etapa radica en reconocer a un usuario dentro de sus múltiples visitas a un determinado sitio web, con la finalidad de poder analizar el comportamiento del usuario a lo largo de días, meses o años.

Una estrategia óptima de solución para la identificación de usuario sería mediante un “nombre de usuario” y “contraseña”, pero como se conoce la navegación web se lleva a cabo normalmente de forma anónima, por lo resulta bastante complicado reconocer a un mismo usuario entre los diferentes servicios a los que accede dentro de una misma sesión y mucho más compleja resulta cuando se tiene en cuenta la evolución temporal.

Determinación de sesiones: Tiene en cuenta un conjunto de servicios solicitados por un mismo usuario a una única visita al sitio o portal web. Las sesiones son un factor importante ya a través de ellas se puede conocer la percepción del usuario con respecto a su visita al portal.

La mejor solución para realizar esta actividad es mediante una aplicación que cree un identificador de sesión la primera vez que un determinado usuario acceda al portal.

[7,17]Una vez efectuado las diferentes etapas de procesamiento de archivos “web logs”, Web mining tiene en cuenta, para su posterior análisis y generación de conocimiento los siguientes recursos:

- Día y a qué hora un usuario navega por el sitio,
- si es la primera vez que visita el sitio o si es reincidente,
- por cual archivo entra a visitar el sitio y por cual sale,
- cuánto dura la visita y cuanto tiempo pasa en cada página,
- de qué país es, qué navegador usa, qué sistema operativo,
- desde donde llega a nuestra página,
- Si llegó a través de un buscador, cuáles fueron las palabras clave usadas,
- etc.

En general los objetivos que persigue web mining son:

1. Buscar e identificar información, más relevante o específica o con relación entre sí.
2. Crear nueva información a partir de información existente.
3. Personalizar la información.
4. Aprender comportamientos de usuarios Web.

3.1 Tipos de web mining

El web mining con la finalidad de optimizar la explotación de data mining clasifica en tres categorías generales a partir de la definición anterior denominadas Web Content Mining (WCM), WebStructure Mining (WSM) y Web Usage Mining (WUM), con una estructura como se muestra en la figura 2.

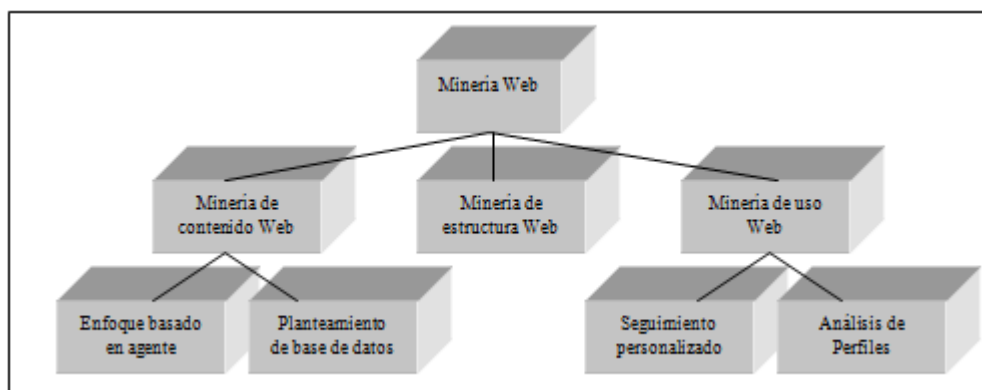


Figura 2. Ramas del Web Mining.

Web Content Mining (minería de contenido web): Se centra en los datos reales que contienen las páginas, es decir el contenido, estos datos consisten generalmente en textos y gráficos. Y se puede obtener datos que acerca de la forma de escribir que elementos o recursos son más atractivos para el usuario, de si la catalogación que usamos sirve para mejorar un ranking, si los temas que se tratan interesan o no.

Web Structure Mining (minería de estructura web): En esta variante de web mining, la información a emplear es aquella que describe la organización del contenido de la página web. La información de la estructura interna de una página incluye por ejemplo el arreglo de las varias etiquetas HTML o de XML dentro de esta. Esto se puede representar como una estructura de árbol en donde una etiqueta (HTML) se convierte en la raíz de este. La clase principal de información en la estructura interna de una página son los “hyper-links” que conectan una página con otra. El tipo de información acerca de si los usuarios encuentran la información, si la estructura de sitio es demasiado ancha o demasiado profunda, si los elementos están colocados en los lugares adecuados dentro de la página, si la navegación se entiende, cuáles son las secciones menos visitadas y su relación con el lugar que ocupan en la página central.

Web Usage Mining (minería de uso web): está orientada a descubrir el uso de las páginas web, direcciones IP, referencias a páginas, la fecha y la hora de accesos, datos que proporcionan la información demográfica sobre los usuarios del Sitio Web. Esto incluye a información de datos que permita determinar comportamiento de clientes o información de perfil del cliente.

En este trabajo nos hemos remitido a detallar fundamentos teóricos sobre web mining, ya que las investigaciones llevadas a cabo en el ámbito de la enseñanza basada en sistemas web, utilizan esta técnica que deriva del Data Mining. Si bien el área de Data Mining aplicada a la educación es muy reciente, cuenta con un número importante de investigadores que han realizado aportes en revistas y congresos. Y que cada año se incrementa denotando la importancia de esta área de investigación dentro del data mining. La gran mayoría de las investigaciones están enmarcadas dentro del ámbito de las universidades.

4. Técnicas de web mining implementadas en la enseñanza web.

Entre las técnicas de data mining que más se utilizan en web mining mencionemos a: las reglas de clasificación y agrupamiento, las de asociación y los sucesos frecuentes que nos permiten por ejemplo clasificar y agrupar a nuestros usuarios y asignarles patrones de comportamientos según la reiteración de acciones que se detecten como clave y así poder ofrecerles productos o servicios acordes a sus perfiles. [17]

4.1 Reglas de decisión

[3] Las reglas de decisión desempeñan un rol importante en el contexto de la nueva visión de la web. Estas técnicas se han implementado en comercio que se manejan de forma electrónica y permiten el desarrollo de estrategias voraces de marketing.

Esta técnica está relacionada con el uso de bases de datos transaccionales, donde cada transacción consiste en un conjunto de ítems. En este modelo, el problema consiste en descubrir todas las asociaciones y correlaciones de ítems de datos donde la presencia de un conjunto de ítems en una transacción implica la presencia de otros ítems.

Esta técnica generalmente está asociada con el número de ocurrencias de los ítems dentro del log de transacciones, por lo tanto, podemos identificar la cantidad de usuarios que acceden a determinadas páginas, también acceden a otra determinada página. Esto facilita mejorar considerablemente la estructura del sitio.

La extracción de reglas de asociación se ha aplicado con éxito en sistemas de enseñanza web para descubrir relaciones o asociaciones entre distintas páginas web visitadas, actividades realizadas, calificaciones obtenidas, etc. Uno de los primeros que ha utilizado estas técnicas en sistemas de enseñanza orientadas a la web es Osmar Zaine[], con la finalidad de proponer actividades de aprendizaje tomando como base historiales de acceso optimizando el proceso de aprendizaje.

Data mining utiliza estas reglas de asociación para preparar al docente y generar un modelo que representa el comportamiento de accesos.

Otro trabajo efectúa un análisis de archivos log en un sistema de enseñanza web, haciendo uso de técnicas de data mining para descubrir patrones de navegación y proponer un modelo de navegación, dicho modelo se basa en dos tipos de relaciones: las relaciones de asociación y relaciones de secuencia de documentos. [11]

Romero [13] en su trabajo describe la aplicación de técnicas de descubrimiento de reglas de descripción de subgrupos haciendo uso de algoritmos evolutivos, tomando como base datos utilizados en el sistema de enseñanza a distancia Moodle. Este trabajo ha implementado algoritmos evolutivos, desarrollados en el ámbito de descubrimiento de subgrupos, que son SIGDA [4] es un algoritmo evolutivo de inducción de reglas difusas que

utiliza reglas lingüísticas como lenguaje de descripción para la especificación de los subgrupos y adaptaciones de las medidas utilizadas en los algoritmos de inducción de reglas de asociación como medidas de calidad para la tarea de descubrimiento de subgrupos (confianza y soporte). SDMGA es una versión multiobjetivo del anterior.

Como complemento hace uso de algoritmos clásicos en el descubrimiento de grupos, que es A PRIORI-SD y CN2-SD, efectuando una comparación en términos de valores promedio: el número de reglas, el número de atributos en el antecedente de las reglas y los valores para las medidas soporte, cobertura, precisión y relevancia de las reglas generadas por los algoritmos clásicos y evolutivos.

Gabroeanu Mihai and Diaconescu Ion-Mircea en su trabajo enfoca el análisis de los archivos log que provienen del sistema de enseñanza web Moodle, efectúa el procesamiento de los mismos para generar reglas de asociación que faciliten la identificación de comportamiento de accesos y uso de los estudiantes. Para la generación de reglas de asociación utiliza una herramienta de aprendizaje automático Weka, que es open source y muy usada en los trabajos que se han descrito en los apartados anteriores. El algoritmo implementado para la generación de reglas es A PRIORI, además utiliza como complemento un framework Jenna Rules para inferir en otros aspectos complementarios de dichas reglas de asociación.

4.2 Clasificación y clustering.

Estas técnicas tienen como finalidad de ordenar o particionar fenómenos, calificados por un conjunto importante de objetos, en unidades más pequeñas que facilitan su administración, comprensión y entendimiento.

La implementación a sistemas de enseñanza web, facilita la agrupación de los usuarios por su comportamiento de navegación o reunir páginas por su contenido, tipo o acceso.

La implementación de algoritmos de clasificación y clustering es realizada por [14] en su trabajo muestra una serie de resultados preliminares y la finalidad es visualizar el resultado de una serie de experimentos sobre algoritmos relacionados con agrupamiento y clasificación, relacionado con información (básica y académica) que se obtiene a partir de la plataforma educativa en uso.

Una vez recolectada la información se utiliza algoritmos de agrupación o clustering, con el propósito de reunir los alumnos en grupos (buenos y malos), teniendo en cuenta el atributo calificación. El algoritmo usado es Fuzzy CMeans (Dunn 2003), como una extensión del popular algoritmo K-Means. El algoritmo de las Fuzzy C-Means es muy robusto y obtiene buenos resultados en muchos problemas de agrupamiento.

Una vez finalizada la tarea de agrupamiento con el objetivo de predecir futuras ocurrencias se utilizaron algoritmos de clasificación como J4.8, Naive Bayes, MPL, SMO, IBk e ID3. Para ello se utiliza el software weka y se hace una comparación de los algoritmos de clasificación visualizando cuál de ellos es el más confiable.

Los trabajos que se detallan a continuación si bien no están en el marco de la educación basados en los sistemas de enseñanza web, están muy relacionados y pueden generar aportes significativos en el marco de la enseñanza virtual.

[2] Otro trabajo en el ámbito educativo propone un clasificador de páginas Web basada en la frecuencia de términos. La clasificación de páginas Web permite mejorar la búsqueda mediante el filtrado de contenidos y la navegación asistida. Para el proceso de clasificación se han usado redes neuronales y algoritmos bayesianos que usa un clasificador ingenuo de Bayes; el mismo puede ser entrenado de manera eficiente y proporciona mejores resultados.

[1] En esta investigación sus autores proponen la utilización de SOM, que es un tipo de red neuronal, con la finalidad de identificar patrones y hábitos en el uso de acceso a páginas web. Como resultado de este trabajo han tomado dos sitios de diferente naturaleza y han efectuado una comparación entre los métodos SOM y K-MEANS uno de los algoritmos clásicos de agrupamiento; logrando visualizar la cantidad de clústeres generado con cada uno de los algoritmos; lo cual concluye para detectar hábitos de uso es más eficiente SOM que K-MEANS.

4.3 Análisis y Patrones de Secuencia.

Es una técnica de la minería de datos que permite revelar patrones de secuencia sobre un conjunto de datos, donde el factor tiempo es esencial.

En el log de transacciones de los servidores de Web, se guarda la fecha y hora en la que un determinado usuario realizó los requerimientos. Analizando estos datos, se puede determinar el comportamiento de los usuarios con respecto al tiempo.

En el ámbito de sistemas de comercio electrónico, se ha aplicado extrayendo de los logs información que puede ser útil para predecir las futuras visitas y así poder organizar mejor los accesos y publicidades para determinados períodos. Por ejemplo, utilizando esta técnica se podría descubrir que los días laborables entre las 9 y las 12 horas muchas de las personas que accedieron al servidor lo hicieron para ver las ofertas y en los siguientes días la mayoría compró productos. Entonces por la mañana debería facilitarse el acceso a las ofertas y brindar la publicidad más llamativa posible.

También puede ser utilizado para descubrir tendencias, comportamiento de usuarios, secuencias de eventos, etc. Esta información puede ser aprovechada tanto en el aspecto comercial (pensar una campaña de marketing) como en el aspecto técnico (mejorar los tiempos de acceso). [7,15]

[5] En su trabajo implementa una técnica de web mining denominada Click Stream Analysis, con el objetivo de descubrir cuál es el comportamiento de los usuarios al operar el portal de la Facultad de Ciencias Económicas, de la Universidad Nacional de Córdoba. En esta técnica el algoritmo usado es el Sequence clustering.

5. CONCLUSIONES

En este trabajo se ha detallado el concepto de data mining y las principales técnicas aplicadas al ámbito de la enseñanza web. Enfocándonos en una de las ramas del data mining que es el web mining, se han citado trabajos que usan estas técnicas mediante la implementación de diversos algoritmos que facilitan la obtención de conocimiento en el ámbito de la enseñanza basada en la web.

La gran mayoría de los trabajos se encuentran dentro del área del data mining usando web mining más precisamente web usage mining, implementando técnicas de clasificación y agrupamiento, descubrimiento de reglas o secuencia de patrones. En este caso hemos hallado más cantidad de trabajos en el ámbito de la enseñanza web, que usan técnicas de descubrimiento de reglas de asociación. En la gran mayoría de estos trabajos los algoritmos implementados para generar conocimiento son los tradicionales A PRIORI, CN2 y sus variantes.

Es importante destacar que una serie de trabajos ha implementados estas técnicas de data mining, haciendo uso del paquete de software Weka, esta implementado en Java y ofrece una serie de ventajas como por ejemplo es de libre distribución, multiplataforma, dispone de una serie de algoritmos de clasificación. Tiene incorporado herramientas para el preprocesado de datos, visualización de datos y resultados.

Este tipo de tecnologías implica un cuidado minucioso en su tratamiento, ya que de ello depende que se logre y afiance el proceso de enseñanza y aprendizaje.

La aplicación de web usage mining en la enseñanza basada en la web proporciona:

- ✓ Información relacionada con las preferencias personales de los estudiantes.
- ✓ Criterios y pautas para personalizar el sistema de enseñanza.
- ✓ Ajustar el servidor.
- ✓ Establecer cambios estructurales en el sistema de enseñanza, fundamentado en las preferencias de los alumnos.

En resumen, Data Mining se presenta como una tecnología innovadora, con una serie de ventajas: por un lado, resulta un buen punto de encuentro entre los investigadores y personas relacionadas al ámbito educativo; por otro, revela conocimiento y genera alternativas para la administración eficiente los sistemas de enseñanza basados en la web.

Referencias

[1] Britos Paola, Martinelli Damián, Merlino Hernán, García-Martínez Ramón, ,“Web Usage Mining Using Self Organized Maps”; IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.6, June 2007

[2] Carrera Enrique V., García María del Cisne, Pasmay Fausto; “Un Algoritmo Simple y Eficiente para la Clasificación Automática de Páginas Web”

[3] De Gyves Camacho Francisco Manuel, “Web Mining: Fundamentos Básicos”, Doctorado en informática y automática, Universidad de Salamanca.

[4] Del Jesús, M.J., Gonzalez, P., Herrera, F. and Mesonero, M., Evolutionary fuzzy rule induction process for subgroup discovery: a case study in marketing. IEEE Trans. Fuzzy Systems, In press.

[5] Díaz Cecilia, Gauna Eduardo; “Un portal universitario y su evaluación con la aplicación de una técnica de web mining, el clic stream analysis”; Centro de Computación y Tecnologías de Información; Facultad de Ciencias Económicas – Universidad Nacional de Córdoba

[6] Feliz Luis Carlos., Artículo. “Data Mining: torturando a los datos hasta que confiesen”, Noviembre 2002.

[7] Fuentes Reyes Sady Carina ., y Ing. Marina Ruiz Lobaina; “Minería de Textos: Aplicación de Web Mining, Instituto de Información y Tecnología”, La Habana Junio 2007.

- [8] Hipp, J., Guntzer, U., y Nakhaeizadeh, G., "Algorithms for Association Rule Mining: A General Survey and Comparison", SIGKDD Explorations, 2 (1), 58_64, 2000.
- [9] Honmy J. Rosario N.; Jesús Zambrano R.; Hyxia Villegas, "Material Instruccional basado en la Web. Impacto en ambientes educativos.", Unidad de Computación, Facultad de Ciencias de la Educación, Centro de Procesamiento de Imágenes, Facultad de Ingeniería, Universidad de Carabobo, Venezuela
- [10] Malbernat Lucía Rosario, "Educación basada en plataformas Web. El desafío tecnológico para los docentes del siglo XXI", copyright © CiberEduca.com 2002
- [11] Morales Cristóbal Romero, Ventura Soto Sebastián, Martínez Cesar Hervás; "Estado actual de la aplicación de la minería de datos aplicados a los sistemas de enseñanza basada en web"
- [12] Moreno Gimena, "Explotación de datos del web mining"; <http://gamoreno.wordpress.com/2007/08/24/explotacion-de-datos-del-web-mining/>
- [13] Romero C., P. González, S. Ventura, M.J. del Jesús, F. Herrera, "Aplicación de algoritmos evolutivos de descubrimiento de subgrupos en e-learning: un caso de estudio analizando cursos de Moodle"
- [14] Sanabria Garzón Jhon Alexis, "Sistema de personalización Web para el proceso de aprendizaje en una plataforma de educación virtual", Maestría en Ingeniería y Sistemas, Universidad Nacional de Colombia.
- [15] Sankar K. Pal, Varun Talwar, Pabitra M.; "Web Mining in Soft Computing Framework".
- [16 e] Sierra Araujo Basilio, Arbelaitz Olatz, Armañanzas Rubén, Arruti Andone, Bahamonde Antonio, "Aprendizaje automático: conceptos básicos y avanzados: aspectos prácticos utilizando el software WEKA", Pearson, 2006
- [17] "Web mining en el diseño de sitios web"; <http://www.interacciones.com.ar/web-mining-en-el-diseno-de-sitios-web/>, Octubre 2005