

## **Minería de uso Web: Presentación de Caso práctico para el mejoramiento de la educación, basada en el uso de plataformas virtuales.**

Esp. Ing. Corso Cynthia Lorena.  
mail:cynthia@bbs.frc.utn.edu.ar  
Universidad Tecnológica Nacional. F.R.C. Departamento de Sistemas

Alfaro Sofía Lorena.  
mail:soft@bbs.frc.utn.edu.ar  
Universidad Tecnológica Nacional. F.R.C. Departamento de Sistemas.

### **Abstract**

Este trabajo tiene como objetivo la presentación de un caso de aplicación relacionado con la minería web, más específicamente a la minería de uso web. En él se expone una metodología para la aplicación de la minería de uso web, como así también la proyección de que aspectos se pretenden descubrir con este caso práctico.

El contexto será aplicado será a nivel educativo; en una Asignatura que pertenece al área de Programación y al primer nivel de la carrera de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional Facultad Regional Córdoba. Si bien las carreras de grado en esta universidad no ofrecen la modalidad de clases a distancia, se han incorporado el uso de plataforma virtual como una alternativa de acompañamiento y refuerzo de las clases presenciales.

### **Introducción**

En los últimos años, e-learning supone una nueva concepción de la educación mediante la incorporación de nuevas tecnologías al proceso de enseñanza y aprendizaje. Las oportunidades y posibilidades que aportan son diversas y debidas al importante crecimiento de avance tecnológico han generado una oferta importante en lo relacionado a propuestas de herramientas educativas, que apoyan a las tareas realizadas en el aula, como así también generación de materiales de estudio. En los últimos años, se ha caracterizado por una fuerte tendencia por parte de las universidades e institutos de enseñanza en el uso de plataformas educativas basadas en la web, tanto para sus cursos a distancia, semipresenciales y presenciales.

La aparición de las plataformas educativas ha reunido en una sola alternativa, funcionalidades que facilitan la administración y organización de contenidos a medida de los alumnos.

En la Universidad Tecnológica Nacional F.R.C aproximadamente hace unos 5 años, se ha incorporado a las asignaturas de la carrera de Ingeniería en Sistemas de Información. La gran mayoría de las cátedras utiliza herramientas de Moodle, con los siguientes objetivos:

- ✓ Publicación de Contenidos Curriculares.
- ✓ Publicación de Novedades de la Asignatura.
- ✓ Foros.
- ✓ Automatización de Trabajos Prácticos.
- ✓ Automatización de Examen Parciales Teóricos/Prácticos.
- ✓ Automatización de actividades de autoevaluación.

Esta plataforma educativa proporciona tanto a docentes como alumnos una importante variedad de recursos, lo que en muchas ocasiones esto suele ser una desventaja. Esto hace que el alumno se encuentre desorientado frente a la elección de recursos diseñados. Este problema requiere una reestructuración de cómo el docente ha diseñado la asignatura, con el objetivo que ese diseño se adapte mejor manera a las características e intereses de los alumnos. [1]

La minería web es una metodología de recuperación de información que utiliza herramientas de la minería de datos. Y se enfoca en la aplicación de técnicas para el descubrimiento de patrones en la Web, esta herramienta nos podría orientar y ayudar a conocer como el alumno interactúa con la plataforma de educación virtual y en el caso de enfrentarse con alguna/s dificultades identificar cuáles son esos factores.

En este contexto lo ideal y deseable sería disponer de un sistema que cuenta con una interfaz que analiza el comportamiento y preferencias del usuario y realiza peticiones en su representación a una base de conocimientos que se encuentra estructurada y gestionada debidamente para la rápida y correcta recuperación de información; en donde dicha base de conocimientos devuelve las respuesta a la interfaz para que esta a su vez le brinde una respuesta al usuario.

Este tipo de sistemas puede implementarse en un servidor, en el cliente o ser un sistema que actúa como nexo que permita la localización de recursos en donde puede existir un componente en el cliente y otro en el servidor.

Atendiendo a los aspectos que se desean investigar, la minería web se clasifica en tres ramas:

**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.

**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.

**Minería de uso web:** La minería de uso en la web consiste en la aplicación de técnicas de minería de datos para descubrir los patrones de uso de la información web con el objetivo de entender y satisfacer las necesidades de los usuarios [2].

Las principales técnicas usadas son de análisis estadístico, reglas de asociación, agrupación de ítems, clasificación y modelos de dependencia aplicadas a las bitácoras de los servidores web y secuencia de páginas visitadas para realizar transacciones. [3]

Los aspectos en los que se enfoca están orientados 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. La fuente de datos típica en la minería de uso web son: logs de acceso en servidores, proxies, agentes y cookies.

Uno de los problemas que existen en torno del uso de plataformas de educación virtual, es que resulta sumamente necesaria la personalización de las mismas, la cual carece de apoyo en lo que respecta a la parte pedagógica.

Este trabajo pretende como objetivo, la elaboración de una propuesta basada en la personalización de plataformas educativas basadas en la web, haciendo uso de técnicas de la minería web, con el objetivo primordial de mejorar la calidad educativa, ofreciendo a los alumnos un servicio más personalizado acorde con sus necesidades e intereses.

## Metodología

Atendiendo a los objetivos que este trabajo pretende alcanzar, el mismo se focaliza en la rama de la minería web más precisamente sobre la minería de uso web. Las etapas planificadas para lograr los objetivos establecidos en el apartado anterior son:

**Preprocesado:** Esta etapa incluirá un conjunto de tareas como limpieza de datos, identificación de las páginas visitadas, de usuarios y sesiones.

La identificación de los recursos visitados nos permitirá obtener conocimiento relacionado con el tipo de actividad que los alumnos desarrollan en el uso de la plataforma.

Si bien existen diversas técnicas para identificar usuarios como creación de perfiles o cookies o la identificación de direcciones IP cada una de ellas con sus ventajas y desventajas, nosotros utilizaremos un recurso de Administración llamado “Informe”, que está incorporado a la plataforma virtual Moodle, solamente es accedido por el perfil Docente. De esta opción existen diversas variantes, la que nos interesa para nuestro estudio es “Estadística”.

Esta herramienta nos ofrece la posibilidad de seleccionar la asignatura, el tipo de Informe es decir si lo que se necesita disponer una vista de la actividad del Administrador, Profesor, Alumno o Invitados en nuestro caso nos interesa la vista de los alumnos. Además otra posibilidad que nos brinda es obtener información histórica de hasta dos años atrás.

A continuación se muestra un caso concreto de la actividad de los alumnos 9 meses atrás.



**Figura1.** Representación gráfica de la actividad de los alumnos.

Fecha	Dirección IP	Nombre completo	Acción	Información
lun 30 de noviembre de 2009, 22:11	190.230.170.3	Federico Herrero	resource view	Guía de ACCESS con Bases de Datos de ejemplos
lun 30 de noviembre de 2009, 22:11	190.230.170.3	Federico Herrero	course view	Informatica I
lun 30 de noviembre de 2009, 21:48	190.230.168.41	Lucas Alejandro Muñoz	course view	Informatica I
lun 30 de noviembre de 2009, 21:30	200.117.208.59	Agustin Andrés Pintado	quiz view all	
lun 30 de noviembre de 2009, 21:18	190.228.229.87	Federico Herrero	course view	Informatica I
lun 30 de noviembre de 2009, 21:12	190.227.249.220	Neri Gustavo Herrador Rojo	course view	Informatica I
lun 30 de noviembre de 2009, 20:57	201.252.163.105	Susana Virginia Benito	resource view	Información sobre SAP R/3
lun 30 de noviembre de 2009, 20:51	201.252.163.105	Susana Virginia Benito	resource view	Link - Administración de Proyectos de Sistemas
lun 30 de noviembre de 2009, 20:50	201.252.163.105	Susana Virginia Benito	resource view	Link - Administración de Proyectos de Sistemas
lun 30 de noviembre de 2009, 20:47	201.252.163.105	Susana Virginia Benito	resource view	Guía de ACCESS con Bases de Datos de ejemplos

**Figura2.** Información de la sesión de los alumnos con actividad en esa fecha.

Como se puede visualizar en la figura 2, no solo se obtiene información de manera gráfica como lo que se muestra en la Figura 1. Además se presenta un listado con toda la actividad de los alumnos relacionada con el uso de la plataforma Moodle en el periodo considerado. Los datos que se visualizan son la fecha, la dirección IP, nombre del alumno, acción que está relacionada con el tipo de actividad que hizo el alumno y al final información adicional relacionada con el recurso accedido. Si bien todos los datos proporcionados por la herramienta son pertinentes para nuestro estudio, planificamos una actividad de limpieza, en este caso la dirección IP no nos aportará ningún aspecto significativo en nuestro estudio.

La incorporación de la herramienta Estadística en Moodle, nos simplificaría bastante el trabajo de recolección de información para un posterior análisis, sin necesidad de incurrir a alguna herramienta externa para tal propósito.

Otra actividad planificada es proceder a efectuar un análisis y sumariación de la información obtenida en cada sesión de los alumnos considerando un determinado periodo de tiempo. Como máximo dos años, que es lo que plantea la herramienta. Es posible que esta información pueda ser migrada a otro formato con el objetivo de facilitar y agilizar su análisis como por ejemplo, a formato de texto u ODS o planilla de cálculo. El formato seleccionado para este trabajo será planilla de cálculo por lo potencia de las herramientas para el análisis estadístico y representación gráfica de los datos.

**Procesamiento:** La tarea principal en esta etapa será la creación de perfiles. Un perfil está compuesto por sesiones de alumnos usuarios similares entre sí, que representan un comportamiento de navegación específico de acuerdo con las páginas que fueron visitadas en cada una de las sesiones.

En este trabajo se aplicarán técnicas de agrupación, el objetivo es la identificación de grupos de alumnos que han tenido comportamientos similares de navegación y utilización de la plataforma. El algoritmo que usaremos será uno de las más simples y clásico dentro de este tipo de técnicas, que es el K-Means.

El procedimiento a aplicar teniendo en cuenta el algoritmo seleccionado será:

- ✓ Se considerarán todas las sesiones de los alumnos, extraídas en la etapa anterior y se seleccionará de manera arbitraria las “n” sesiones que se convertirán en los centros de los “n” grupos.

- ✓ De las sesiones restantes, a cada una de ellas se las asignará a un determinado grupo. La pertenencia estará fundamentada por un cálculo. El mismo resulta de la similitud de la sesión del alumno y cada uno de los centros seleccionados en el punto anterior. Será finalmente al grupo de mayor similitud.
- ✓ El centroide de cada grupo será recalculado.
- ✓ Las sesiones serán nuevamente reasignadas, teniendo en cuenta el establecimiento de los nuevos centroides.
- ✓ Los pasos 2,3 y 4 se repetirán, hasta que los nuevos centroides de cada grupo no hayan tenido una variación significativa.

La herramienta que se prevé para la implementación de este algoritmo será Weka. Esta herramienta es una de las herramientas para la implementación de minería de datos cuya filosofía es Software Libre, aunque a simple vista disponga una interfaz bastante sencilla y simple, es una herramienta muy potente que dispone de diversas herramientas para implementar la gran mayoría de las técnicas de minería de datos, aplicadas fundamentalmente al uso de la minería web.

**Descubrimiento y análisis de patrones:** Una vez obtenido los perfiles de los estudiantes, es necesario poder evaluar la calidad o nivel de confianza de los mismos. Las métricas Recall y Precisión utilizadas en el área de recuperación de información, han sido utilizadas para tal fin. El recall es una medida que corresponde al porcentaje de ítems que son correctamente recomendados, mientras que la precisión mide la calidad promedio de una recomendación individual. La validación de los resultados para cualquier técnica de minería de datos, es fundamental, ya que determinan el nivel de confianza de las inferencias resultantes de este estudio.

## Conclusiones

En este trabajo se propone una metodología de minería de uso web aplicada a la enseñanza acompañada y guiada en el uso de plataformas de educación virtual.

Con el resultado de aplicar esta metodología, se espera responder los siguientes interrogantes como:

¿Los alumnos usan la plataforma sólo como herramienta de comunicación con el docente?

¿Cuál es el nivel de acceso a las actividades académicas diseñadas por el docente?

¿Cuál es el nivel de acceso a las actividades de evaluación y autoevaluación, diseñadas por el docente?

¿Los alumnos sólo acceden a los contenidos curriculares de la asignatura, conceptualizando la herramienta como un repositorio de información, sin tener en cuenta otros recursos como foros, encuestas etc.?

Lograr responder a estos interrogantes, nos permitirá diseñar una alternativa de personalización de plataforma educativa web, que permita tener un nivel más alto de atención y establecer una guía que les facilite a los alumnos poder acceder a información pertinente y hacer uso adecuado de los recursos didácticos fortaleciendo el aprendizaje significativo y autónomo.

## Referencias

- [1] Sistema recomendador colaborativo usando minería de datos distribuida para la mejora continúa de cursos e-learning, Enrique García Salcines, Cristóbal Romero Morales, Sebastián Ventura Soto y Carlos de Castro Lozano, IEEE-RITA Vol. 3, Núm. 1, Mayo 2008
- [2] Web Usage Mining: Discovery and Applications of Usage Paterns from Web Data, Jaideep Srivastava, Robert Cooley, Mukund Deshpande y Pang-Ning Tan, SIGKDD Explorations, ACM. Enero 2000.
- [3] Minería web y personalización: Revisión bibliográfica y propuesta de un marco de referencia; Antonio González Torres.
- [4] Extracción de perfiles basado en agrupamiento genético para recomendación de contenido, Sandra Patricia Tocarruncho Tocarruncho, Fredy Andrés Aponte Novoa, Arturo Tocarruncho Tocarruncho, Conferencia IADIS Ibero-Americana WWW/Internet, 2007.
- [5] Excavando en la web, Baeza-Yates, Ricardo. En: El profesional de la información, 2004, enero-febrero, v. 13, n. 1, pp. 4-10.
- [6] Minería de la Web: Crawling the Web, Cecchini Rocío Luján.
- [7] Web Watcher: A tour guide for the World Wide Web; Thorsten Joachims. Universidad Dortmund. Departamento de Informática, Dayne Freitag. Universidad Carnegie Mellon. Departamento de Ciencias de Computación, Tom Michell. Universidad Carnegie Mellon. Departamento de Ciencias de Computación.
- [8] Minería Web: un recurso insoslayable para el profesional de la información, Lic. Sady C. Fuentes Reyes e Ing. Marina Ruiz Lobaina, VI Jornada Bibliotecaria del IDICT, 17 y 18 de julio de 2007 en el Capitolio Nacional, La Habana, Cuba.
- [9] Introducción al uso de la Minería Web: Data Mining sobre ficheros log, Blaya Juan Botía. Departamento de Ingeniería de la Información y las Comunicaciones. Universidad de Murcia.
- [10] Web Usage Mining and Personalization, Bamshad Mobasher, DePaul University.
- [11] Adaptive Web Sites: an AI Challenge, Mike Perko, Oren Etzioni, Department of Computer Science. University of Washington. Seattle.
- [12] Personalización de los servicios en la web, Mestras Juan Pavón, Departamento de Sistemas Informáticos y Programación, Universidad Complutense Madrid, 2001.