

CONSIDERACIONES GENERALES DEL WEB MINING

**Sandra Milena Leal
Elizabeth Castiblanco Calderón***

RESUMEN: el presente artículo describe los conceptos básicos para la utilización del Webmining, dentro de los cuales se encuentran técnicas (Reglas de Asociación, Patrones secuenciales, clasificación y Clustering) y áreas (minería de contenido web, minería de estructura web y minería de uso web) de mayor utilización para el descubrimiento de información.

También se pretende exponer la importancia de la utilización de esta técnica dentro de las unidades de información, teniendo en cuenta que esta trabaja a partir de las necesidades de los usuarios, parte fundamental para el funcionamiento de las mismas.

PALABRAS CLAVES: Webmining, Logs, Técnicas, Minería de Datos, Áreas.

GENERAL CONSIDERATIONS OF WEBMINING

ABSTRACT: This article describes the basic concepts for the use of Webmining, within which are techniques (Rules of Association, sequential patterns, sorting and clustering) and areas (web content mining, mining structure and mining use Web site) of greater use for the information discovery.

It also aims to expose the importance of the use of this technique within units of information, bearing in mind that this works from the users' needs, essential to the operation.

KEYWORDS: Webmining, Logs, Techniques, Datamining, Areas.

INTRODUCCIÓN

Con la llegada de las tecnologías el ser humano ha evolucionado hacia una nueva sociedad, la sociedad del conocimiento, en la cual Las Tecnologías de Información se hacen presentes en todas las actividades del hombre; por ello es necesario que exista una actualización permanente con información oportuna que satisfaga sus necesidades.

El usuario actualmente se enfrenta a una serie de cambios rápidos debido al acelerado crecimiento de la información en la web y a la globalización de la sociedad.

Este crecimiento ha traído consigo información no relevante la cual lleva al usuario a una saturación de información, todo esto conlleva a la creación de técnicas que permitan la adecuada estructuración del contenido en la web, para que la recuperación de la información sea de forma ágil y oportuna.

Dentro de estas técnicas se encuentra el webmining que se deriva de la minería de datos, la cual se “usa para el estudio de varios aspectos esenciales de un sitio y ayuda a descubrir tendencias y relaciones en el comportamiento de los usuarios que sirven como pistas para, por ejemplo, mejorar la usabilidad de un sitio”¹. Esta a su vez se divide en diferentes áreas y técnicas las cuales se presentan a continuación.

ÁREAS Y TÉCNICAS

La minería es una técnica aplicada al descubrimiento de “patrones interesantes en la estructura, el contenido y la utilización de los sitios web”² con el fin de identificar las tendencias de los usuarios en el uso y acceso a la información.

Dentro de la aplicación de esta técnica se encuentran diferentes áreas que permiten el estudio de algunos componentes de la web, con el fin de proporcionar información personalizada a los usuarios teniendo en cuenta su perfil; dentro de las cuales encontramos:

¹ <http://www.webtaller.com/maletin/articulos/web-mining-diseno-sitios-web.php>

² <http://www.infovis.net/printMag.php?num=172&lang=1>

1. Web Content Mining (minería de contenido web)

Este pretende obtener datos sobre el contenido y las búsquedas realizadas en los website, con el fin de extraer patrones a partir de información on-line, incluida en los diferentes formatos del contenido de la web tales como: texto libre, HML, HTML, multimedia y su vez identificar los “patrones en los resultados de los motores de búsqueda”³

Esta técnica permite utilizar métodos para presentar los resultados de las búsquedas de forma tal que permita que el usuario seleccione las más relevantes para satisfacer sus necesidades de información, teniendo en cuenta que los motores de búsquedas “no proveen información estructurada ni filtran o interpretan los documentos.”⁴

Como se expuso anteriormente estos problemas han llevado a la “creación de herramientas inteligentes para la recuperación de información como es: agentes Web inteligentes, técnicas de bases de datos y datamining, con el fin de garantizar un mayor nivel de organización a los datos semiestructurados disponibles en la Web”.⁵

Dentro de las herramientas más conocidas se encuentran:

| Índices | Motores De Búsqueda | Multibuscadores | Representación De Contenido | Agentes Inteligentes |
|---|--|--|--|---|
|  Buscadores que mantienen una organización de las páginas incluidas en su base de datos por categoría. |  Permite la recolección de páginas a través de su contenido para facilitar su clasificación.  Recorre permanentemente las |  Se relaciona con varios buscadores con el fin de realizar una búsqueda simultánea.  Ej: My search (que busca simultáneam |  Se hace a través de palabras clave y se pueden encontrar de forma manual o automática que permiten |  Estos trabajan teniendo en cuenta un perfil particular de un usuario y lo pueden hacer de forma |

³ Ibíd.

⁴ Schiaffino Silvia. Descubrimiento de Conocimiento a partir de Datos. [en línea] disponible en : <<http://www.exa.unicen.edu.ar/catedras/dbdiscov/clase4.pdf>> consultada 27 octubre 2007.

⁵ Ibíd. Pág 7

| | | | | |
|--|---|---|--|--|
| <p>☐ La información se encuentra contenida en un directorio que permite la organización de las páginas por temas y a su vez cada página posee interrelaciones con otras páginas que contienen el mismo tema.</p> | <p>páginas Web en busca de posibles cambios. ☐ Ej: Google y Alta vista.</p> | <p>ente en Google, Yahoo, Altavista y muchos otros)</p> | <p>buscar en todo el documento o por partes. ☐ Ej: título, y primer párrafo.</p> | <p>autónoma o semi - autónoma en base a un dominio dado.</p> |
|--|---|---|--|--|

2. Web Structure Mining (minería de estructura web)

“Se centra en usar el análisis de la estructura de links de la Web y uno de sus propósitos es identificar los documentos más importantes y más buscados para un tema”.⁶ Esta determina el grado de popularidad de un documento a través de links, este permite que el usuario pueda remitirse a otras páginas que le amplíen la información que esta buscando.

Uno de los mecanismos relevantes para determinar la importancia de las páginas es el PageRank, este permite que los sitios más visitados y con mayor importancia sean ubicados en los primeros lugares.

⁶ Schiaffino Silvia. Descubrimiento de Conocimiento a partir de Datos. [en línea] disponible en : < <http://www.exa.unicen.edu.ar/catedras/dbdiscov/clase4.pdf>> consultada 27 octubre 2007.

3. Web Usage Mining (minería de uso web)

Identifica los patrones que determinan la forma en que los usuarios interactúan con la Web, estos patrones se determinan a través de los logs que están ubicados en los servidores Web.

Beneficios:

- ☐ Mejorar la navegación en un sitio.
- ☐ Permite realizar campañas de publicidad efectivas.
- ☐ Estructurar mejor el contenido en un sitio Web.
- ☐ Dirigir avisos ofertas a grupos de usuarios específicos.

Para la implementación de estas áreas es necesaria la aplicación de técnicas, las cuales provienen del datamining, dentro de las más concidas se encuentran:

Reglas de Asociación

“Descubrir las correlaciones entre referencias a varios archivos disponibles en el servidor hechas por un mismo cliente”.⁷

A través de esta regla se conocen las transacciones realizadas por cada usuario, las cuales están contenidas en el Log. Teniendo en cuenta que cada “transacción esta compuesta por un conjunto de URL”⁸. Esto permite encontrar las similitudes entre los diferentes ítems como por ejemplo:

·El 60% de los clientes que acceden a la página con URL /company/products/, también acceden a la página /company/products/product1.html.”⁹

Patrones secuenciales:

“Permite a las organizaciones basadas en la Web predecir los patrones de visita de los usuarios y los ayuda a dirigir las publicidades a grupos de usuarios basándose en estos patrones”¹⁰.

⁷ Ibíd. Pág 21

⁸ http://www2.ing.puc.cl/gescopp/Sergio_Maturana/SAG/Webmining.html

⁹ Ibíd

¹⁰ Schiaffino Silvia. Descubrimiento de Conocimiento a partir de Datos. [en línea] disponible en : < <http://www.exa.unicen.edu.ar/catedras/dbdiscov/clase4.pdf>> consultada 27 octubre 2007.

Esta técnica analiza la información contenida en el Log determinando el “comportamiento de los usuarios con respecto al tiempo”¹¹ con el fin de “predecir las futuras visitas”¹² y así promover con anticipación la publicidad de la página.

Clasificación

“las técnicas de clasificación permiten definir los perfiles de los clientes que acceden a los archivos del servidor”.

Ejemplo:

Cientes de agencias del estado o del gobierno que visitaron el sitio tienden a interesarse en la página /company/products/product1.html;
50% de los clientes que colocaron una orden online en /company/products/product2, pertenecen al grupo 20-25 años y viven en la capital.

Clustering:

“El clustering permite agrupar clientes y elementos de datos que tienen características similares”.¹³ A través de esta técnica se analizan los logs de transacciones con el fin de desarrollar y ejecutar futuras estrategias de marketing, como por ejemplo el envío de correo electrónico a los clientes.

CONCLUSIÓN

Como se evidenció la Webmining es una técnica que sirve en la actualidad para el descubrimiento de conocimiento en el uso de la web. Uso que refleja las tendencias de los usuarios en la búsqueda y recuperación de información, por lo tanto esta técnica es importante por que trabaja a partir de las necesidades propias de los usuarios.

Sería importante que las unidades de información conocieran esta técnica con el fin de ofrecer mejores servicios en pro de generar mayor acceso a los recursos, teniendo en cuenta que estos se encuentran en diferentes formatos. Es por esta

¹¹ http://www2.ing.puc.cl/gescopp/Sergio_Maturana/SAG/Webmining.html

¹² *Ibíd.*

¹³ Schiaffino Silvia. Descubrimiento de Conocimiento a partir de Datos. [en línea] disponible en : <<http://www.exa.unicen.edu.ar/catedras/dbdiscov/clase4.pdf>> consultada 27 octubre 2007.

razón que los profesionales de la información deben conocer y aplicar estas técnicas para garantizar que la información se ajuste a las necesidades de los usuarios.

BIBLIOGRAFÍA

Dürsteler, Juan C. Minería Web [en línea] disponible en:
<<http://www.infovis.net/printMag.php?num=172&lang=1>>

Explotación de datos del webmining [en línea] disponible en:
<<http://gamoreno.wordpress.com/2007/08/24/explotacion-de-datos-del-web-mining/>>

Koblinc, Gustavo D. Web Mining: Estado Actual de Investigación. [en línea]
Disponible en:
http://www2.ing.puc.cl/gescopp/Sergio_Maturana/SAG/Webmining.html

MENESES, Esteban. Web Mining [en línea] disponible en:
<http://www.geocities.com/emenesesr/recursos/webMining.pdf>****

Webmining en el diseño de sitios web [en línea] disponible en:
<<http://www.webtaller.com/maletin/articulos/web-mining-diseno-sitios-web.php>>

Schiaffino Silvia. Descubrimiento de Conocimiento a partir de Datos. [en línea]
disponible en : < <http://www.exa.unicen.edu.ar/catedras/dbdiscov/clase4.pdf>>
consultada 27 octubre 2007.