

## ARTÍCULOS

# MINERÍA DE USO DE WEB PARA PREDICCIÓN DE USUARIOS EN LA UNIVERSIDAD

## WEB USAGE MINING OF USERS PREDICTIONS IN THE UNIVERSITY

Ulises Román<sup>1</sup>, Luis Alarcón<sup>1</sup>

### RESUMEN

En este trabajo se presentan las técnicas, herramientas y estrategias de la Minería de Uso de la Web para el análisis, procesamiento y predicción de comportamiento de las personas y organizaciones que accedan y usen la web para realizar alguna operación. La Minería Web (MW) se ha convertido en un proceso de identificación, extracción de información automática desde la web. Se ha empleado la técnica de reglas de asociación de la Minería de Uso de la Web (MUW) para el caso de estudio. Como resultado presentamos la aplicación del algoritmo *Apriori* –usando las reglas de asociación para determinar el comportamiento de los usuarios/docentes de la universidad que accedan y extraen cierto tipo de información.

**Palabras Clave:** Minería de datos, minería web, patrones, archivos logs, clustering, reglas asociación, servidor.

### ABSTRACT

This work presents the techniques, tools and strategies of web mining. Focus in analysis, processing and prediction of human and organizations behavior, based in access and use of web, in order to do some operation.

The web mining is a process of identification, extraction of automatic information from web. In this work we use the technique of rules association from web using mining. As result, we present the algorithm *Apriori* using the rules association to determine the behavior of users and professors form this university when they access and extract and kind of information.

**Key words:** Data mining, web mining, patterns files Logs, clustering, rules association, servers.

## 1. INTRODUCCIÓN

En la actualidad la información se ha convertido en un recurso estratégico y crítico para cualquier organización. La web ha experimentado un crecimiento exponencial desde su aparición en 1990, convirtiéndose

en el repositorio más grande y ampliamente conocido de hipertexto. A medida que crece una organización, el volumen de los datos va en aumento, siendo más dificultoso su manejo y posterior uso como información de apoyo a decisiones. Es por ello que surge la Minería de Datos (MD) para poder llevar un

<sup>1</sup> Docentes del Departamento de Ciencias de la Computación-FISI, UNMSM.  
E-mail: {alarconl, nromanc}@unmsm.edu.pe

mejor control de las transacciones basándose en patrones de búsqueda específicos. En este contexto, la Minería Web (MW) usa las técnicas de MD para descubrir y extraer información automáticamente desde el WWW, y nos ayuda a comprender las preferencias de navegación y el comportamiento de los usuarios [1].

La MW consta de tres métodos de extracción más usados: Minería de Minería de contenido de la Web o Web Content Mining (WCM), Minería de estructura de la Web o Web Structure Mining (WSM) y Minería de uso de la Web o Web Usage Mining (WUM) <http://www.cs.ualberta.ca/~tszhu/webminig/>

El presente estudio se centra en la Minería de Uso de la Web (MUW) para aplicaciones de predicción de comportamiento de usuarios de la universidad. En la sección 2 describimos la Minería Web y la Minería de Uso de la Web: Procesos, categorías, técnicas y aplicaciones. En la sección 3 se propone usar la reglas de asociación y algoritmo *a priori* para el caso de discusión. En las secciones 4 y 5, discusión y análisis de resultados. En la sección 6 se presentan las conclusiones; y en la Sección 7, las referencias bibliográficas.

## 2. FUNDAMENTACIÓN TEÓRICA

### 2.1. Minería Web-Web Mining (WM)

El término minería web (WM) fue introducido por Etzioni [2], quien definió la minería web como el uso de técnicas de minería de datos –Data Mining (DM)– para descubrir y extraer información automáticamente desde la web. Según Cooley [3], es la aplicación de la DM para grandes repositorios de datos web, donde se aplican técnicas de extracción de patrones de interés que serán útiles para descubrir las tendencias de una organización.

El descubrimiento de patrones de acceso del usuario de los servidores web lo podemos verificar cuando uno visita un sitio en internet, ya que el usuario deja gran cantidad de información (dirección de IP, el navegador, fecha, transacción, etc.). Cada operación del usuario se almacena en un archivo log.

#### a) Proceso de la Minería Web

La minería web tiene cuatro procesos para seleccionar y transformar los resultados:

- La primera etapa consiste en localizar documentos y servicios en la web (proceso de descubrimiento a las fuentes), recuperación de datos desde las fichas textuales de la web usando índices. Ejemplo: Google, Yahoo, Altavista, etc.
- Extraer automáticamente la información específica desde las fuentes (proceso de selección y pre-procesado). Incluye cualquier proceso de selección y de transformación de sus datos originales obtenidos. Ejemplo: los sistemas de extracción de Harvest [4]; capaz de encontrar título y autor de documentos Latex, FAO-Finders [5], que extrae las respuestas de preguntas más frecuentes.
- Descubrir patrones generales desde los sitios web (proceso de generalización). Se pueden usar técnicas de agrupamiento-*cluster*, reglas de asociación, recuperación de información (IR) [6, 7] y técnicas propias, como análisis de caminos, usadas para extraer secuencia de patrones de navegación desde archivos log.
- Interpretación y validación de los patrones minados (proceso de análisis). En esta etapa se deben desarrollar técnicas y herramientas que permitan el consumo humano del conocimiento minado.

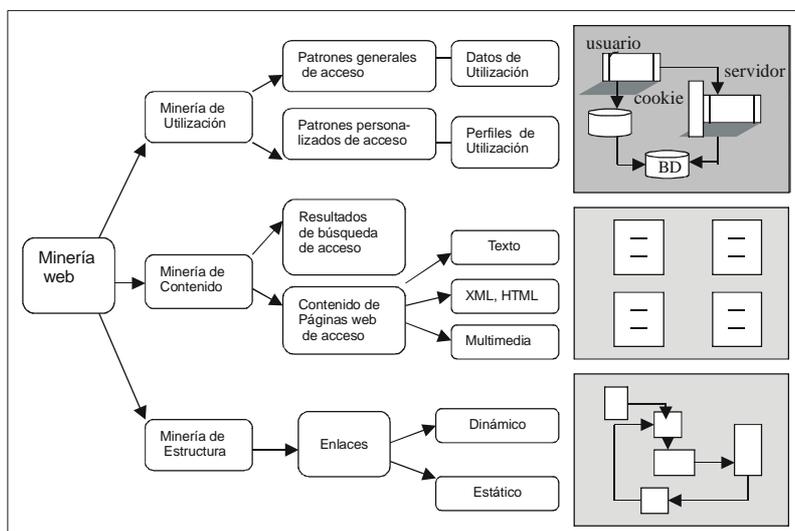


Figura N.º 1. Categorías de la Minería Web.

**b) Categorías de la Minería Web [1,8]**

En la Figura N.º 1 se muestran las categorías de la Minería Web y las operaciones que realizan cada una de ellas. En función de la parte que se minan, se pueden clasificar en tres áreas de interés:

- Minería de contenido de la Web–Web Content Mining (WCM)
- Minería de estructura de la Web–Web Structure Mining (WSM)
- Minería de uso de la Web–Web Usage Mining (WUM).

**Minería de contenido de la Web–Web Content Mining (WCM)**

Es el proceso que consiste en la extracción de conocimiento del contenido de documentos o sus descripciones. La localización de patrones en el texto de los documentos, el descubrimiento del recurso basado en conceptos de indexación o la tecnología basada en agentes también pueden formar parte de esta categoría.

**Minería de estructura de la Web – Web Structure Mining (WSM)**

Es el proceso de inferir conocimiento de la organización del WWW y la estructura de sus ligas. Su principal propósito es identificar documentos más preferibles al usuario.

**Minería de uso de la Web–Web Usage Mining (WUM)**

Aquí lo que se intenta es bucear en los registros de los servidores (logfiles) sobre las transacciones informáticas que se realizan en la web con el fin de encontrar patrones sobre el uso que se le da.

Ejemplo, páginas más visitadas, recorridos habituales, etc.; podemos distinguir también aquí: *Seguimiento de patrones generales de acceso* y *seguimiento personalizado de patrones de acceso*.

**c) Minería Web en los servidores y clientes web**

Existen varias herramientas que aplican WM en los servidores tanto en los clientes web como en los comerciales y académicos; todas aplican el proceso KDD sobre el log del servidor. Estas herramientas generan respuestas estadísticas y gráficas sobre el uso del servidor. Ejemplo: web trends, getstats, analog, msinterse, market focus, etc.

Los datos almacenados siguen un formato estándar designados por CERN y NCSA. Ejemplo: Dirección IP del cliente, ID, fecha de acceso, requerimiento, URL de la página ingresada, protocolo ID, código error, agente, N.º bytes transmitidos, etc.

Existen agentes web personalizados que obtienen las preferencias del usuario y descubren fuentes de información. Ejemplo: *tendencias del usuario, páginas ofrecidas, patrón de acceso*. En la Figura N.º 3 se muestran los paquetes enviados por el usuario, que son almacenados en los logs del servidor; se puede apreciar el IP de la computadora, fecha y hora exacta de acceso a la página, también qué es lo que pidió el usuario al servidor, el status, los bytes transferidos, el browser utilizado y el acceso a una página definida. Existen programas especializados para la recuperación de este tipo de información (*Cpanel, WebStats, AceFTP, otros*). Se tiene **Analog** (<http://www.analog.cx/>) una herramienta para el análisis de archivos log de servidores web.

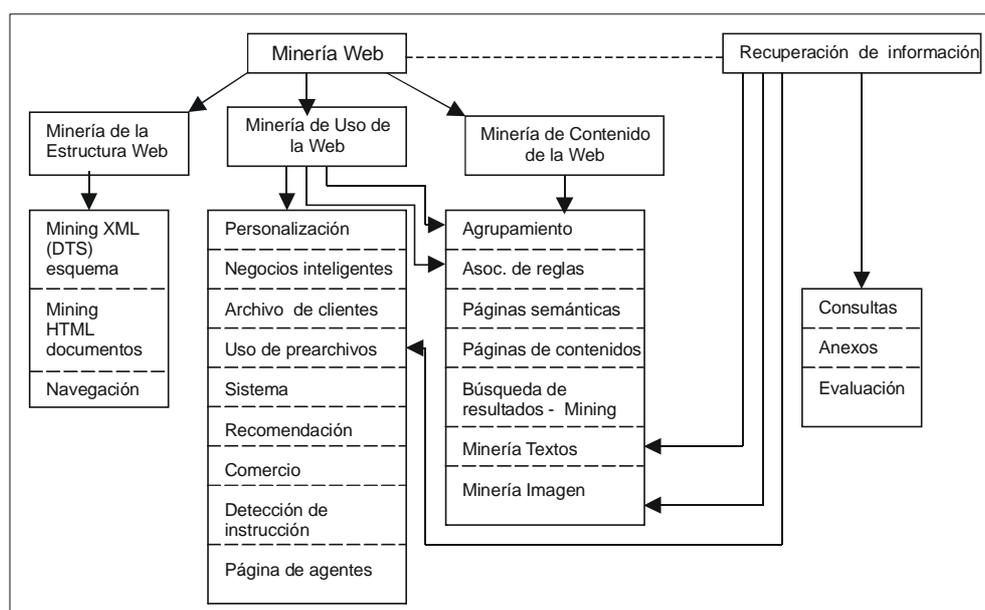


Figura N.º 2. Taxonomía de la Minería Web [8].

Originating IP: 198.81.129.99  
 Timestamp: (26/Jul/1999:  
 10:26:56 - 0400)  
 HTTP Command & Protocol Versión:  
 «GET / ido / images / id.gif. HTTP/1.0»  
 Status Code: 200  
 Bytes transferred: 660  
 Browser: « Mozilla /4.51 [ en]  
 (WinNT; U)»  
 Referer URI: «http: // www.company.com/»

## 2.2. Minería de uso de la web - Web usage mining (WUM)

La WUM es uno de los procesos que permite registrar las transacciones informáticas que se realizan en la web, también permite la personalización de archivo de clientes, uso de profiles, detección de intrusos, páginas agentes, etc.

### a) Técnicas empleadas en la Minería de uso de Web

Mencionaremos tres técnicas, las más usadas para la minería web:

#### 1. Agrupamiento y clasificación

Las técnicas de agrupamiento o *clustering* distribuyen comportamientos de individuos similares en grupos homogéneos. Esto es, dos elementos con características parecidas pertenecerán al mismo grupo y las características de un grupo (definidas por el elemento prototipo o ideal) serán diferentes a los de otro grupo. Dependiendo de la información *almacenada* en los *ficheros log*, es posible detectar grupos de usuarios como:

- Aquellos que visitan gran cantidad de páginas con un tiempo de estancia similar en todas ellas.
- Los que visitan un número pequeño de páginas en sesiones cortas.
- Los que visitan un número pequeño-mediano de páginas con tiempo variable en cada una de ellas.

Una vez descubiertos los prototipos o perfiles de cada grupo, se pueden usar las características de cada uno de ellos para realizar la *clasificación*.

En la WUM, las técnicas de clasificación permiten desarrollar un *perfil para clientes/usuario* que acceden a ficheros concretos del servidor, en función de sus patrones de acceso. El *agrupamiento* de clientes-usuario puede facilitar el desarrollo de la ejecución de *estrategias de mercado futuras*, tanto *on line* como *off line*. Ejemplo, envío de correos automáticos a aquellos clientes / usuario

que se encuentren dentro de cierto grupo, reasignación dinámica de servidor para un cliente (Ejemplo: menos sobrecargado, para darle un mejor servicio), o presentación de contenidos específicos según el tipo de cliente.

#### 2. Reglas de Asociación

Las reglas de asociación permiten capturar patrones referentes a los ítem sets sin distinción, en los que ocurre una transacción de datos. Con esta técnica podemos encontrar correlaciones como:

- El 45% de clientes / usuarios que accedieron a la página web con URL */entidad/productos/productos1.html* también accedieron a */entidad/productos/productos2.html*
- El 35% de clientes / usuarios que accedieron a la página web con URL */entidad/anuncio/oferta-especial.html* efectuaron un pedido interactivo */entidad/productos/productos1.html*

A través del análisis de asociaciones, podemos descubrir las relaciones sin que exista intervención alguna por parte del operador. El descubrimiento de estas reglas ayuda a las organizaciones dedicadas al e-commerce a definir sus estrategias de mercados efectivos.

El aprendizaje de Reglas de Asociación se divide normalmente en dos fases: *Fase 1*. Extracción de los conjuntos de los ítem que cumplen con la cobertura requerida desde los datos; y *Fase 2*. Generación de las reglas a partir de estos documentos.

#### 3. Secuencias Frecuentes

La minería de secuencias puede ser considerada como asociación minando sobre los *datasets temporales* y una secuencia de lista ordenada (con el paso del tiempo) de *itemsets* no vacíos. El objetivo de esta técnica es descubrir el tiempo de las secuencias ordenadas de URLs que ha sido seguido por usuarios, para predecir a futuros.

En general, en la BD Transaccionales están disponibles los datos en un periodo de tiempo y se cuenta con la fecha en que se realizó la transacción.

El descubrimiento de patrones de secuencia (*sequential patterns*) en el log puede ser utilizado para predecir las futuras visitas y así poder organizar mejor los accesos y publicidades para determinados periodos de tiempo. Ejemplo: Los días laborables entre las 9 a.m. y las 12 m., 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 las mañanas se debería facilitar el acceso a las ofertas

y brindar la publicidad más llamativa posible. Podremos ver más detalles en *WUM 7.0* (<http://www.iwi.nu.berlin.de>), un entorno integrado para analizar el comportamiento de navegación de los usuarios y descubrir patrones secuenciales.

## b) Aplicaciones de la Minería de Uso de la Web

Cuando el usuario interactúa con un sitio web, los datos que registran su comportamiento se almacenan en los *logs de los servidores web*; las aplicaciones (MUW) pueden clasificarse con:

- Aprendizaje de patrones de navegación (APN).
- Aprendizaje de perfiles de usuario –personalización (APU).

### 1. Aprendizaje de Patrones de Navegación (APN)

Podemos transformar los datos a una notación tabular (pares tributo-valor), en [10] se presenta un método consistente en representar las sesiones de navegación de usuario desde los archivos logs, como una gramática probabilística de hipertexto (GPH), tal que las cadenas generadas por la gramática con mayor probabilidad corresponden a los *caminos* preferidos por los usuarios.

La (GPH) es una gramática regular (una tupla  $\langle V, \Sigma, E, S, P \rangle$  con una relación 1 a 1 entre el conjunto de símbolos no terminales ( $V$ ) y el conjunto de símbolos terminales ( $\Sigma$ ).  $S$  y  $F$ , estados de inicio y fin de sesiones de navegación; cada símbolo no terminal corresponde a una página web y con regla de producción de la gramática ( $P$ ) a un enlace entre páginas.

Considerando:

$\Omega$  : Peso de un estado para ser la primera página en una sesión de navegación

$N$  : Número de URLs anteriores

$\rho$  : Probabilidad de cada producción

$A_i$  : Recorrido de las páginas

Cálculo de probabilidad de visitas de cada página:

$$\rho(S \rightarrow a_1 A_1) = (\Omega * N_{\text{visitas}_{A1}} / N_{\text{total\_visitas}}) + (\Omega * N_{\text{inicios}_{A1}} / N_{\text{total\_inicios}})$$

Donde:  $V = \{S, A_1, A_2, \dots, A_n, F\}$  y  $\Sigma = \{a_1, a_2, \dots, a_n\}$ , donde  $\Omega$  puede tomar valores  $[0, 1, >0]$  y  $N[1, \geq 1]$

### 2. Aprendizaje de Perfiles de Usuario-Personalizado (APU)

La personalización (interfaces adaptativos) puede definirse como cualquier acción que adapta la web al gusto del usuario. Existe un interés de personalizar cualquier software que interactúe con el usuario.

Existen métodos como: colaborativos, aprendizaje automático, basados en refuerzos y sistemas recomendadores, etc. [11,12,13]. A continuación mostramos los *módulos para la personalización de la web*:

- Módulo de categoría de contenido conceptual.
- Módulo de categoría de uso conceptual.
- Módulo búsqueda y adquisición de información.
- Módulo de administración de contenidos (BD contenido) .
- Módulo de uso de patrones (asociación de reglas, clusters, secuencia de patrones, etc.).
- Módulo de Minería de uso de la web (BD uso log, BD uso archivos, estructura).
- Módulo de inicio www.

Estos módulos nos permiten optimizar y personalizar la navegación de los usuarios, y de esta forma brindarlos para un mejor servicio. Se tiene *OK Log* (<http://www.oklog.biz/>), módulo que examina los portales web para obtener información de las visitas de los usuarios.

## 3. MÉTODOS

### 3.1. Reglas de asociación (RA) y dependencia

Las RA son una manera muy popular de expresar patrones de datos de una BD. Tiene aplicaciones prácticas en estudio de textos de búsqueda, búsqueda de patrones a páginas web, análisis de canasta de compra en un supermercado y otros. Estas aplicaciones llevan asociadas un gran volumen de datos. Ejemplo: podemos detectar patrones del tipo «si un cliente/usuario compra un reproductor de DVD, es probable que al mes siguiente compre varias películas de DVD». «El 45% de personas (alumnos) que consultan la página web de la universidad sobre el contenido de los cursos en un área específica, en menos de dos días extraen de la página web la información requerida».

Se tiene *WLS XP* (<http://www.webmining.cl>); esta herramienta descubre patrones de la actividad de los usuarios en un portal web, generando reglas de asociación a partir de archivos log.

#### a. Denotación de Reglas de Asociación (RA)

Sean  $X$  e  $Y$  un conjunto de atributos de los conjuntos de datos (*datasets*), podemos expresar una regla de asociación como:  $X \Rightarrow Y$ , implica que cualquier transacción contiene a  $X$  como a  $Y$  [14,15]. Esta expresión se puede codificar como:

**SI**<condiciones\_satisfechos>**ENTONCES**<predicen\_valores\_para\_atributos>

**SI X ENTONCES Y** o **IF X THEN Y**

b. *Medidas para conocer la calidad de la regla*

Dos son las medidas para conocer la calidad de la regla: cobertura/suporte (*support*) y confianza/precisión (*confidence*). La *cobertura/suporte* se define como el número de instancias que la regla predice correctamente (usa %), y la *confianza/precisión* mide el % de veces que la regla se cumple cuando se aplica.

c. *Algoritmos de aprendizaje de reglas de asociación*

El más conocido y popular es el algoritmo **Apriori**; se basa en la búsqueda de conjunto de items con determinada cobertura [16], y el algoritmo **AprioriAll** [17] para el aprendizaje de RA secuenciales –tiene como objetivo hallar las secuencias de conjuntos de items que cumplan una mínima cobertura.

Existen herramientas como WEKA y CLEMENTINE de minería de datos; el paquete «WEKA.Associations.Apriori» contiene la implementación del algoritmo de aprendizaje de reglas de asociación *Apriori*. Para más información visite <http://www.cs.waikato.ac.nz/~ml/WEKA>; y en el caso de CLEMENTINE, comercializada por SPSS, contiene una arquitectura distribuida (cliente/servidor) multiplataforma para la visualización de resultados de las reglas de asociación; para mayores detalles, visitar <http://www.spss.spssbi/clementine>

3.2. **Algoritmo Apriori: Búsqueda de conjunto de items**

En la Figura N.º 3 se muestra el algoritmo de búsqueda de conjunto de items (Apriori):

Para la búsqueda de un conjunto de items se ha empleado un conjunto de items formado por *X* items frecuentes; *Si y sólo si* cada uno de los *X* items es frecuente *por sí sólo*. A continuación describimos el algoritmo usado para validar la Tabla N.º 1:

```

ALGORITMO Apriori (D: Datos, MinC: Cobertura Mínima)
i ← 0
Rellenar_item(C1) // Incluye en C0 todos los items de tamaño 1
MIENTRAS Ci ≠ ∅
  PARA CADA X = elemento de Ci
    SI cobertura(X) ≥ MinC ENTONCES Li = Li U X
  FIN PARA
  Ci+1 ← Selecciona_candidatos(Li)
  i ← i + 1
FIN MIENTRAS
RETORNA C
FIN ALGORITMO
    
```

Figura N.º 3. Algoritmo apriori.

Donde:

- i : Tamaño del conjunto de ítem
- i+1 : Tamaño de posibles candidatos
- (X,Y) : Conjunto de items de Z

4. **DISCUSIÓN DEL CASO**

Sean las transacciones (T) y la información deseada (I<sub>d</sub>) por los *usuarios/docentes* de la UNMSM. La tabla contiene sólo valores binarios, un **1** en la posición ( i, j ) indica que la transacción requerida i incorpora la información deseada j; mientras que un cero **0** indica que el *usuario/docente* no ha logrado extraer la información deseada de un tema específico.

Tabla N.º 1

Transacción	I <sub>d</sub> 1	I <sub>d</sub> 2	I <sub>d</sub> 3	I <sub>d</sub> 4	I <sub>d</sub> 5	I <sub>d</sub> 6	I <sub>d</sub> 7
T1	1	1	0	0	0	1	0
T2	0	1	1	0	0	0	0
T3	0	0	0	1	1	1	0
T4	1	1	0	1	1	1	1
T5	0	0	0	0	0	1	0
T6	1	0	0	0	0	1	1
T7	0	1	1	1	1	0	0
T8	0	0	0	1	1	1	1
T9	1	1	0	0	1	0	1
T10	0	1	0	0	1	0	0

En la Tabla N.º 1 vamos extraer una muestra para el caso de análisis del uso de la Regla de Asociación; veamos: Si I<sub>d</sub>5 y I<sub>d</sub>4 ENTONCES I<sub>d</sub>6

5. **ANÁLISIS DE RESULTADO**

Sea la *cobertura mínima* igual a 2, en la tabla existen 7 conjuntos de sólo un ítem (*siete atributos*) y todos aparecen en la tabla, al menos 2 veces; por lo tanto tomaremos los siete (7). Si pasamos a los conjuntos formados por dos ítems tendremos:

Posibles casos de encontrar la información deseada, tendrían 15 conjuntos.

$$C^7 = \frac{7!}{2 \cdot 5!} = 42$$

Los ítems I<sub>d</sub>5 y I<sub>d</sub>4 tienen cobertura cuatro (4); continuando, tendríamos once (11) conjuntos de tres (3) ítems, y tan sólo dos (2) conjuntos de cuatro (4) ítems.

A continuación se extraen las reglas que tienen un *nivel de confianza mínimo*. Para el caso de *conjunto de ítems I<sub>d</sub>4 y I<sub>d</sub>5 y I<sub>d</sub>6*, podemos extraer la siguiente regla de asociación:

1.  $S_I I_{d5}$  y  $I_{d4}$  ENTONCES  $I_{d6}$  Cb = 4, Cf =  $\frac{3}{4}$
2.  $S_I I_{d5}$  y  $I_{d6}$  ENTONCES  $I_{d4}$  Cb = 3, Cf =  $\frac{3}{3}$
3.  $S_I I_{d6}$  y  $I_{d4}$  ENTONCES  $I_{d5}$  Cb = 3, Cf =  $\frac{3}{3}$
4.  $S_I I_{d6}$  ENTONCES  $I_{d5}$  y  $I_{d4}$  Cb = 6, Cf =  $\frac{3}{6}$
5.  $S_I I_{d5}$  ENTONCES  $I_{d4}$  y  $I_{d6}$  Cb = 6, Cf =  $\frac{3}{6}$
6.  $S_I I_{d4}$  ENTONCES  $I_{d5}$  y  $I_{d6}$  Cb = 4, Cf =  $\frac{3}{4}$
7.  $S_I \text{Æ}$  ENTONCES  $I_{d5}$  y  $I_{d6}$  y  $I_{d4}$  Cb = 10, Cf =  $\frac{3}{10}$

Para cada regla se ha incluido su cobertura (Cb) y confianza (Cf) de acuerdo a la Tabla N.º 1.

## 6. CONCLUSIONES

1. Dado el carácter cambiante de la propia web es de esperar que el área de la minería web evolucione y crezca dando nuevas técnicas propias, así como nuevos sistemas de minería web.
2. El crecimiento continuo en el tamaño y el uso de web mundial impone nuevas metodologías de diseño y desarrollo de los servicios de información en línea.
3. Mediante el uso de la WM se pueden descubrir tendencias, comportamientos de usuarios, patrones de navegación, secuencia de eventos, etc.

## 7. REFERENCIAS BIBLIOGRÁFICAS

- [1] José Hernández O., José Ramírez Q., César Ferri R. (2004), *Introducción a la Minería de Datos*. Pearson-Prentice Hall, España
- [2] Etzioni, O. (1996), *The World Wide Web: quagmire or gold mine?* Communications of the ACM, 39(11): 65-68.
- [3] Cooley, F. (2000), *Data mining*, Addison Wesley.
- [4] Brown, C.M.; Danzing, P.B.; Hardy, D.; Mamber, U. y Schwartz, F. (1994), *The Harvest information discovery and access system*, in Proceedings of the Second International World Wide Web Conference, pp. 763-771.
- [5] Hammond, K; Burke, R; Martin, C; Lytinen, S.(1995), *FAQ finder: A case -based approach to knowledge navigation*, in Working Notes of the AAAI Spring Symposium: Information Gathering from Heterogeneous, Distributed Environment, pp. 69-73, Stanford University, AA AI Press.
- [6] Salton, G; McGill, M.(1983), *Introduction to modern information*. McGraw-Hill, New York.
- [7] Baeza-Yates, R; Ribeiro-Neto, B.(2000), *Modern Information Retrieval*, Addison-Wesley.
- [8] Sushmita Mitra, Tinku Acharya (2003), *Data Mining: Multimedia, Soft Computing, and Bioinformatics*, Wiley Interscience USA.
- [9] D. Arotaritei; S. Mitra.(2003), *Web mining: A survey in the fuzzy frame-work*, Fuzzy Sets and systems (accepted).
- [10] Borges, J; Leven, M.(2000), *A fine grained Heuristic to Capture Web Navigation Patterns*, SIGKDD Explorations, 2(1): 1-11.
- [11] Resnick, P; Varian, H. (1997), *Recommender Systems-Introduction to the Special Section*, Communications ACM 40(3):56-58.
- [12] Hirsh, H; Basu, C; Davison, B. (2000), *Learnig to personalize*, Communications ACM 43(8):102-106.
- [13] Billsus, D; Pazzani, M. (1999), *A hybrid user model for news story classification*, in Proceedings of the 7<sup>th</sup> International Conference on User Modelling, pp. 99-108.
- [14] S. Mitra; R.K,De; S.K, Pal. (1997), *Knowledge-based fuzzy MLP for classification an rule generation*, IEEE Transactions on Neuronal Networks, vol. 8, pp. 237-250.
- [15] S. Mitra; S.K., Pal; P. Mitra (2002), *Data mining in soft computer frame-work: A survey*, IEEE Transactions on Neuronal Networks, vol. 13, pp.3-14.
- [16] Han, J; Camber, M. (2001), *Mining: Concept and Techniques*, Morgan Kaufman Publishers.
- [17] Agrawal, R; Srikant. (1995), *Mining Sequential Patterns*, in Proceedings of the 11<sup>th</sup> International Conference on Data Engineering, pp.3-14.