

NEODOCUMENTA

JOURNAL OF

DOCUMENT ENGINEERING

Vol. 13
Year 2014



Serendipity
Editions

Minería de Uso de la Web

Fuensanta M^a Guerrero Carmona
Manuel Marcos Aldón

Resumen. Debido al extraordinario crecimiento de las fuentes de información disponibles en la Web, es necesario utilizar herramientas que, de alguna manera, automaticen parte del proceso realizado por los usuarios para encontrar lo que desean, llevando a cabo un seguimiento y análisis de los patrones más comunes que emplean en su proceso de búsqueda, así como el estudio de su comportamiento en las visitas a las páginas Web. La minería de uso de la Web utiliza diversas técnicas para recopilar datos sobre los usuarios y su interacción con el sistema, que se utilizan en la adaptación y personalización de los sitios Web según sus necesidades e intereses, mejorando de esta manera su experiencia en la navegación y capturando información útil, tanto para los servidores y empresas como para los usuarios.

1. Introducción. Definición y objetivos de minería de uso de la Web

El incremento continuo de las transacciones comerciales a través de Internet y de los servicios y sistemas de información disponibles en la Web, ha motivado que el volumen de datos almacenados sobre los perfiles de los usuarios y las operaciones que realizan haya llegado a alcanzar cifras desmesuradas. El análisis de estos datos puede ayudar a las empresas y organizaciones a diseñar estrategias de marketing adecuadas, optimizar la funcionalidad de las aplicaciones y servicios ofrecidos, personalizar el contenido y encontrar la estructura más adecuada para su sitio Web [17]. Para ello es necesario utilizar herramientas capaces de descubrir, de forma automática, los patrones de comportamiento más significativos y las relaciones que se establecen a partir de grandes colecciones de datos semi-estructurados. Estos datos están almacenados tanto en la Web como en los registros de acceso de los servidores [17].

La minería de uso de la Web es el proceso de extracción de información a partir de los patrones de navegación y acceso de los usuarios [5], utilizando técnicas de minería de datos para descubrir patrones de uso a partir de la información proporcionada por la Web [26]. Esta operación se realiza de forma automática, y su objetivo es capturar, modelar y analizar los patrones de comportamiento y los perfiles de los usuarios en su interacción con el sitio Web [17]. Este descubrimiento automático de patrones se lleva a cabo a partir de colecciones de páginas, objetos o recursos a los que acceden con frecuencia grupos de usuarios con necesidades o intereses comunes.

Utilizando diversas herramientas de análisis de la Web es posible determinar el número de accesos al servidor, el tiempo que el usuario permanece en el sitio, los nombres de dominio y URLs de los usuarios y el análisis de las relaciones entre los datos. Estas herramientas se pueden dividir en dos categorías [5]:

- Herramientas de descubrimiento de patrones: utilizan técnicas de inteligencia artificial, minería de datos, psicología y teoría de la información para extraer conocimiento a partir de los datos recopilados. Los algoritmos de Chen [2] para encontrar las *referencias máximas* y las *secuencias de referencia más largas* se pueden utilizar para realizar varios análisis del camino recorrido, como por ejemplo identificar las rutas más utilizadas en los sitios Web.
- Herramientas de análisis de patrones: después de haber descubierto los patrones, los analistas necesitan las herramientas y técnicas adecuadas para entender, visualizar e

interpretar estos patrones. Algunas de estas técnicas son OLAP [7], que utiliza *data cubes*, y WEBMINER [6], que utiliza mecanismos de consulta tipo SQL para examinar el conocimiento descubierto, empleando reglas de asociación y patrones secuenciales.

2. Etapas de procesamiento

El análisis de la minería de uso puede utilizar distintos tipos de fuentes de datos [4][26]:

- Contenido: datos de la página que se quieren transmitir al usuario.
- Estructura: describe la organización del contenido, los hipervínculos internos y externos.
- Uso: datos que describen los patrones de utilización de los usuarios, tales como direcciones IP, referencias a páginas y fecha/tiempo de acceso. Esta información es obtenida a partir de los registros de servidores Web.
- Perfil del usuario: datos que no se pueden conseguir automáticamente de los registros, como los datos de acceso, información demográfica, etc.

La recopilación de estos datos se hace tanto en el cliente como en el servidor, así como a través de servidores proxy y *packet sniffers*, poniendo especial cuidado en proteger la información personal de todos los usuarios [19]. Estos datos serán procesados en las distintas etapas, extrayendo conocimiento útil para un amplio rango de aplicaciones.

2.1 Preprocesamiento

El preprocesamiento de uso es posiblemente la parte más difícil de esta fase [26], ya que se pueden encontrar problemas a la hora de identificar a los usuarios y las sesiones del servidor, puesto que pueden darse problemas de duplicación de IPs o de sesiones. Además, a veces se producen dificultades con las referencias de páginas en la caché, siendo necesario monitorizar el uso desde el lado del cliente. Los datos obtenidos a partir de los enlaces pulsados por el usuario son filtrados y divididos [17] en un conjunto de transacciones que representan las actividades que ha realizado durante sus diversas visitas al sitio Web.

El preprocesamiento del contenido, además de convertir el texto, las imágenes, los scripts, etc., en información utilizable en el proceso de minería de uso de la Web, implica realizar una clasificación o agrupamiento [26] de los datos utilizando los algoritmos apropiados. El contenido de páginas estáticas se lleva a cabo analizando el código HTML, mientras que las dinámicas requieren la utilización de herramientas para extraer la información necesaria de las bases de datos dónde está almacenado el contenido [26].

El preprocesamiento de la estructura se crea a partir de los hipervínculos entre las páginas. Naturalmente, los enlaces referidos a contenido dinámico plantean más problemas que los estáticos. Para cada sesión del servidor se construye una estructura del sitio diferente [26].

Por lo tanto, las tareas básicas que se deben realizar en esta fase son las siguientes [5]:

- Depuración de datos: desarrollar técnicas para depurar/filtrar los datos con el objetivo de eliminar valores atípicos y/o elementos irrelevantes. Para ello se puede implementar un modelo para registrar los datos de acceso, empleando mecanismos como cookies, almacenamiento en memoria caché, etc.
- Identificación de la transacción o sesión del usuario: agrupar accesos a páginas individuales en unidades semánticas o transacciones.

2.2 Inferencia de patrones (pattern discovery)

En esta fase se realizan operaciones estadísticas sobre los datos y procesos de aprendizaje automático, con el objetivo de obtener patrones ocultos [17] que reflejen el comportamiento característico de cada usuario, así como descubrir reglas de asociación y patrones secuenciales [5] para realizar agrupamientos y clasificaciones. También se elaboran diversas estadísticas

recopilatorias sobre los recursos utilizados, las sesiones y los usuarios. Las tareas que se llevan a cabo en esta etapa son [26]:

- **Análisis estadístico:** en la Web, un grafo representa algunas relaciones definidas entre las páginas. El análisis de las trayectorias de este grafo se puede utilizar para determinar las páginas del sitio que se visitan más frecuentemente, las relaciones entre las rutas seguidas para llegar a una página concreta, o el tiempo que tarda un usuario en abandonar el sitio [5].
- **Reglas de asociación:** las técnicas de descubrimiento de reglas de asociación y correlaciones en un entorno Web implican tener en cuenta las referencias entre las páginas y los archivos del servidor a los que se accede. Cada transacción está compuesta por el conjunto de URLs a las que accede un cliente cada vez que visita el sitio Web. Además, debido a la gran cantidad de datos existente, las reglas de asociación tratan de reducir el espacio de búsqueda.
- **Agrupamiento y clasificación:** las técnicas de agrupamiento establecen grupos de usuarios que utilizan patrones de búsqueda similares y determinan grupos de páginas con contenido relacionado. Las herramientas de clasificación permiten diseñar el perfil de los usuarios que acceden a determinados archivos en un servidor, teniendo en cuenta sus patrones de acceso y la información demográfica que se tiene almacenada sobre ellos [5].
- **Patrones secuenciales:** permiten predecir los patrones que siguen los usuarios teniendo en cuenta la secuencia de páginas visitadas [5]. También se pueden determinar las relaciones temporales entre los datos, investigando el acceso a páginas o recursos en un periodo de tiempo determinado.
- **Modelado de dependencia:** establece las dependencias entre ciertas variables del dominio Web, y es capaz de predecir los recursos que más se van a requerir en un futuro próximo.

2.3 Análisis de patrones

Una vez descubiertos los patrones de comportamiento y efectuadas las estadísticas oportunas sobre ellos, los datos obtenidos se procesan y se filtran para elaborar modelos de usuario generales. Dichos modelos se pueden utilizar en aplicaciones tales como motores de recomendación, herramientas de visualización, sistemas de análisis Web y herramientas de generación de informes [17].

Debido a que el tamaño de los registros en el servidor crece muy rápidamente, no es posible proporcionar un análisis online de todos estos datos [5], por lo que deben reducirse los datos del registro o enfocar el análisis de tal forma que se seleccione qué datos se van a extraer. Esta información se encuentra almacenada en distintas fuentes, como registros del servidor Web, registros de referencia, archivos de registro y registros del servidor de índices. Los tipos de datos que se pueden utilizar para construir modelos incluyen [13]: características del usuario (edad, género, localización, etc.), preferencias e intereses, conocimiento y habilidades y patrones de comportamiento.

2.4 Algunas herramientas existentes

- **WEBMINER** [5] es uno de los primeros sistemas desarrollados para la minería de uso de la Web [26]. Presenta una arquitectura general de análisis basada en el descubrimiento automático de reglas de asociación y patrones secuenciales, obtenidos a partir de los registros de acceso de los servidores.
- **WebViz** [20] es un sistema para visualizar patrones de acceso a la Web que permite analizar de forma selectiva los registros de acceso, controlar los grafos de enlaces, reproducir los eventos del registro de acceso, seleccionar los nodos y enlaces que

representan mejor la estructura de la base de datos y examinar el grafo en cualquier instante. De esta forma, la Web se visualiza como un grafo dirigido con ciclos, dónde los nodos son páginas y las aristas son hipervínculos [5].

- WebSIFT [4] está diseñado para realizar minería de uso de la Web a partir de archivos de registro con formato NSCA (National Center for Supercomputing Applications) extendido, que incluye campos de referencia y agentes [26]. Identifica usuarios, sesiones y referencias a páginas en la caché, además de preprocesar el contenido y la estructura de los sitios Web. Almacena los datos de cada sesión en un *episodio*, que contiene un análisis de la secuencia de patrones seguidos por el usuario, las reglas de asociación descubiertas, el agrupamiento realizado, así como estadísticas que pueden resultar interesantes.
- SpeedTracer [28] hace uso de información de referencia y de agentes en el procesamiento de rutinas para identificar sesiones de usuarios y de servidor [26], mediante la reconstrucción de las trayectorias del recorrido realizado por el usuario. Lleva a cabo un análisis de los usuarios, de las rutas y de los grupos para estudiar su comportamiento de navegación.
- WebLogMiner [29] es capaz de descubrir reglas de asociación, realizar clasificaciones y análisis de secuencias de eventos, de transiciones y de tendencias [26]. Una vez recopilados, los datos son filtrados para eliminar información no relevante y se crea una base de datos con los datos válidos. Utiliza OLAP (On-Line Analytical Processing) para agilizar las consultas y técnicas de minería de datos para predecir, clasificar y descubrir correlaciones.
- WUM [25] especifica características de las rutas más frecuentes que pueden ser interesantes para el análisis de la minería de uso de la Web [26], creando árboles de estructura con la combinación de la información obtenida. Incluye el análisis de las secuencias de eventos y de las rutas que los conectan.
- Shahabi et al. [23][30] han realizado un sistema que se basa en la recopilación de datos en el lado del cliente [26]. La técnica utilizada, llamada *path-mining*, detecta y codifica la ruta de navegación de los usuarios, agrupando los que navegan por caminos similares. La información que obtiene incluye el orden en que se accede a las páginas, accesos a los enlaces, caché de referencia y el tiempo exacto que el usuario permanece en la página.

3. Técnicas de aprendizaje aplicadas a la minería de uso

En la minería de uso Web se utilizan técnicas de aprendizaje automático para generalizar los comportamientos de los usuarios. Estos métodos se pueden clasificar en aprendizaje supervisado y no supervisado. Este último incluye métodos de clustering (k-means, vecinos más próximos, etc.), visualización (correlaciones, asociaciones y dependencias) y estadísticos, que se pueden utilizar como paso previo para el aprendizaje supervisado.

Los sistemas estadísticos utilizados en el aprendizaje supervisado incluyen [17][26]:

- Modelos Ocultos de Markov (HMMs): evolucionan en el tiempo pasando de un estado a otro, que se pueden representar en un grafo dirigido para modelar las propiedades observadas. A cada transición y a cada estado les asigna una probabilidad.
- Redes Bayesianas Dinámicas (DBNs): consisten en un grafo dirigido acíclico que modela las estructuras de datos secuenciales o series temporales.
- Modelos de Entropía Máxima (MEMs): estiman la probabilidad de que cierto hecho ocurra en un contexto determinado, según las restricciones impuestas por las observaciones.

- Campos Aleatorios Condicionales (CRFs): se utilizan para etiquetar y segmentar secuencias de datos o extraer información de documentos.
- Máquinas de Vectores de Soporte (SVMs): resuelven problemas de clasificación y regresión. Dado un conjunto de muestras, se pueden etiquetar las clases y entrenar una SVM para construir un modelo de predicción de los nuevos datos obtenidos.

Con la utilización de todas estas técnicas y métodos se pretende analizar el comportamiento de los usuarios en la Web para modelar sistemas de recomendaciones, estudiar los hábitos de uso para establecer tendencias de mercado, reestructurar un sitio Web para que el acceso a las páginas más visitadas sea rápido y fácil, etc.

4. Sitios Web adaptativos

4.1 Definición y objetivos

Un sitio Web adaptativo es aquel que mejora automáticamente su estructura y presentación aprendiendo de los patrones de acceso de los usuarios. El análisis del contenido, estructura y uso de la Web puede servir para personalizar y adaptar los sitios Web, mejorando la presentación de los contenidos en base a los datos recopilados. Se pueden emplear también herramientas de inteligencia artificial para transformar los sitios Web tradicionales en entidades inteligentes y evolucionadas [18], más accesibles, con enlaces destacados, conectadas con páginas relacionadas y con documentos similares agrupados, todo ello con el objetivo de facilitar y adecuar la navegación y la búsqueda de información a cada usuario según sus necesidades e intereses.

Los métodos utilizados más frecuentemente en el descubrimiento de patrones de uso para la construcción de sitios Web adaptativos son la minería de clustering, la minería de reglas de asociación y la minería de patrones secuenciales [13].

4.2 Aproximaciones

Varios factores dificultan el diseño de un buen sitio Web [18]: un visitante puede buscar información distinta en momentos diferentes, muchos sitios cambian su diseño original acumulando enlaces y páginas de forma descontrolada, un sitio diseñado para un uso determinado puede utilizarse de formas distintas. Estos problemas hacen que la navegación y búsqueda de información se complique, por lo que es necesario utilizar alguno de los enfoques que existen para llevar a cabo un sitio Web adaptativo [18]: personalización y optimización.

La personalización consiste en ajustar la forma de presentar la información para un usuario individual, que puede realizarse de forma manual (especificando el usuario las opciones elegidas) o mediante predicciones automáticas [15]. Se puede dividir en los siguientes módulos [8]: perfil de usuario, análisis de registro, adquisición de información, gestión del contenido y publicación del sitio.

Por su parte, la optimización se centra en mejorar el sitio como un todo, aprendiendo de todos los usuarios, no de un solo individuo [18]. Se trata de buscar en el espacio de todos los posibles diseños aquél que es preferible. Para ello, se mide el esfuerzo que el visitante emplea de media para encontrar lo que busca en el sitio Web y, según este resultado modificar, reordenar o crear los enlaces oportunos para que la navegación sea más fácil para los usuarios.

Algunas investigaciones realizadas sobre sitios Web adaptativos son:

- Programación genética lineal [1] para analizar los hábitos de los usuarios en el descubrimiento de patrones de uso. En las pruebas realizadas se utilizan técnicas de optimización para mejorar la extracción de información útil.

- Construcción de árboles de decisión [3] para clasificar los datos que se utilizarán en un sistema de recomendación con agentes inteligentes.
- Sistemas de e-learning para proporcionar a los alumnos recomendaciones automáticas en línea: [12][21][22].
- Algoritmo de aprendizaje Fuzzy-LZ [24] para extraer el historial de páginas a las que ha accedido el usuario e identificar patrones de acceso recurrentes. Esta nueva técnica se utiliza para hacer predicciones de aproximación sobre las URLs que visitará el usuario.
- Modelo de semántica latente probabilístico [10] para descubrir variables y relaciones ocultas entre los usuarios y las páginas visitadas, basado en patrones de co-ocurrencia en distintas sesiones del usuario. El conocimiento semántico del contenido Web permite aprovechar los datos para realizar un análisis más efectivo de los registros almacenados.

Muchas investigaciones actuales en este campo se refieren tanto al análisis semántico mencionado arriba como a la integración de minería de uso, de estructura y de contenido para aumentar la efectividad de los modelos resultantes [13]. La utilidad, calidad y validez de la adaptación se debe evaluar para comprobar la eficacia de la mejora de usabilidad introducida. Para lograrlo se pueden realizar pruebas [18], analizar los archivos de registro, observar la interacción de los usuarios con el sitio, etc.

4.3 Algunos sistemas existentes

- WebWatcher [11] es un sistema que orienta al usuario en la navegación Web. Aprende a predecir qué enlaces seguirá el usuario en función de sus intereses específicos por medio de la observación directa de las secuencias de enlaces que visita [18], resaltándolos y colocándolos al principio de la página.
- AVANTI [9] se centra en la personalización dinámica basada en las necesidades y preferencias de los usuarios [18], intentando predecir tanto sus posibles objetivos como el siguiente paso que va a dar.
- SiteHelper [27] es un agente que ayuda al servidor y al usuario para encontrar información relevante en el sitio Web, aprendiendo de forma interactiva e incremental sobre las áreas de interés del usuario.
- Letizia [14] es un agente inteligente que explora el comportamiento de búsqueda de los usuarios (enlaces pulsados, búsquedas realizadas, peticiones de ayuda) y trata de predecir qué elementos pueden resultarles interesantes. Utiliza modelos heurísticos para modelar el perfil del usuario y proponer recomendaciones.

5. Áreas de investigación relacionadas con la minería de uso de la Web

- Inteligencia artificial.
- Aprendizaje automático.
- Web semántica.
- Interacción hombre-máquina.
- Minería de datos y análisis de textos.
- Recuperación y extracción de información.
- Teoría de la información.
- Clustering y clasificación.
- Reconocimiento de patrones.

6. Conferencias internacionales donde se aborda la minería de uso de la Web

- 1st International Workshop on Usage Analysis and the Web of Data (USEWOD2011):
<http://data.semanticweb.org/usewod/2011/>
- 4th International Conference on Web Intelligence, Mining and Semantics (WIMS), 2014:
<http://www.wikicfp.com/cfp/servlet/event.showcfp?eventid=32493©ownerid=54714>
- International Conference on Web Intelligence, Mining and Semantics (WIMS), 2012:
<http://software.ucv.ro/Wims12/cfp-p.html>
- IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS), 2013:
<http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=6588691>
- 10th International Conference on User Modeling (UM'05):
<http://gate.ac.uk/conferences/um2005/um05.html>
- 6th Asia Pacific Web Conference (APWEB'04):
<http://www1.se.cuhk.edu.hk/~apweb04/accepted.htm>

7. Lista de referencias

- [1] ABRAHAM, A., RAMOS, V. “Web usage mining using artificial ant colony clustering and linear genetic programming”. En *Evolutionary Computation, 2003. CEC'03. The 2003 Congress on* [en línea]. IEEE, (2003). pp. 1384-1391. Disponible en: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1299832>
- [2] CHEN, M.-S., PARK, J. S., YU, P. S. “Efficient data mining for path traversal patterns”. En *Knowledge and Data Engineering, IEEE Transactions on* [en línea]. Vol. 10, n° 2, (1998) pp. 209-221. Disponible en: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=683753>
- [3] CHO, Y. H., KIM, J. K., KIM, S. H. “A personalized recommender system based on web usage mining and decision tree induction”. En *Expert Systems with Applications* [en línea]. Vol. 23, n° 3, (2002) pp. 329-342. Disponible en: <http://www.sciencedirect.com/science/article/pii/S0957417402000520>
- [4] COOLEY, R., TAN, P.-N., SRIVASTAVA, J. "Discovery of interesting usage patterns from web data". En *Web Usage Analysis and User Profiling* [en línea]. Springer Berlin Heidelberg, (2000). pp. 163-182. Disponible en: http://link.springer.com/chapter/10.1007/3-540-44934-5_10
- [5] COOLEY, R., MOBASHER, B., SRIVASTAVA, J. “Web mining: Information and pattern discovery on the world wide web”. En *Tools with Artificial Intelligence, 1997. Proceedings., Ninth IEEE International Conference on* [en línea]. IEEE, (1997). pp. 558-567. Disponible en: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=632303>
- [6] COOLEY, R., MOBASHER, B., SRIVASTAVA, J. “Grouping web page references into transactions for mining world wide web browsing patterns”. En *Knowledge and Data Engineering Exchange Workshop, 1997. Proceedings* [en línea]. IEEE, (1997). pp. 2-9. Disponible en: <http://www.computer.org/csdl/proceedings/kdex/1997/8230/00/82300002-abs.html>
- [7] DYRESON, C. E. “Using an incomplete data cube as a summary data sieve”. En *IEEE Data Eng. Bull* [en línea]. Vol. 20, n° 1 (1997), pp. 19-26. Disponible en:

- <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.39.7388&rep=rep1&type=pdf>
- [8] EIRINAKI, M., VAZIRGIANNIS, M. "Web mining for web personalization". En *ACM Transactions on Internet Technology (TOIT)* [en línea]. Vol. 3, nº 1, (2003) pp. 1-27. Disponible en: <http://dl.acm.org/citation.cfm?id=643478>
- [9] FINK, J., KOBASA, A., NILL, A. "User-oriented adaptivity and adaptability in the AVANTI project". En *Designing for the Web: empirical studies* [en línea]. (1996). Disponible en: http://bauhaus.ece.curtin.edu.au/~iain/PhD%20BU/A_PhD%20docs/To%20read/Accessibility%20info/citeseer/fink96useroriented.pdf
- [10] JIN, X., ZHOU, Y., MOBASHER, B. "Web usage mining based on probabilistic latent semantic analysis". En *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* [en línea]. ACM New York, (2004). pp. 197-205. Disponible en: <http://dl.acm.org/citation.cfm?id=1014076>
- [11] JOACHIMS, T. et al. "Webwatcher: A tour guide for the world wide web". En *IJCAI (1)* [en línea]. (1997). pp. 770-777. Disponible en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.29.4400&rep=rep1&type=pdf>
- [12] KHRIBI, M. K., JEMNI, M., NASRAOUI, O. "Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval". En *Advanced Learning Technologies. ICALT'08. Eighth IEEE International Conference on* [en línea]. IEEE, (2008). pp. 241-245. Disponible en: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4561676>
- [13] KOUTRI, M., AVOURIS, N., DASKALAKI, S. "A survey on web usage mining techniques for web-based adaptive hypermedia systems". En *Adaptable and Adaptive Hypermedia Systems, Idea Publishing Inc., Hershey* [en línea], (2005). Diponible en: http://acadmedia.wku.edu/Zhuhadar/eBooks/Survey-web-usage-mining-techniques-2004-Koutri_Avouris_Daskalaki_2004.pdf
- [14] LIEBERMAN, H. et al. "Letizia: An agent that assists web browsing". *IJCAI (1)* [en línea]. Vol. 1995, (1995) pp. 924-929. Disponible en: <http://www.aai.org/Papers/Symposia/Fall/1995/FS-95-03/FS95-03-016.pdf>
- [15] MOBASHER, B., COOLEY, R., SRIVASTAVA, J. "Creating adaptive web sites through usage-based clustering of urls". En *Knowledge and Data Engineering Exchange, 1999. (KDEX'99) Proceedings. 1999 Workshop on* [en línea]. IEEE, (1999). pp. 19-25. Disponible en: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=836525>
- [16] MOBASHER, B. "Web usage mining and personalization". En *Practical Handbook of Internet Computing* [en línea], (2004), pp. 264-265. Disponible en: <http://dl.acm.org/citation.cfm?id=345169>
- [17] MOBASHER, B., NASRAOUI, O. "Web Usage Mining". En B. LIU, *Web Data Mining. Exploring Hyperlinks, Contents, and Usage Data*. New York: Springer, 2007. pp. 527-603.

- [18] PERKOWITZ, M., ETZIONI, O. "Adaptive web sites: an AI challenge". En *IJCAI (1)* [en línea]. (1997). pp. 16-23. Disponible en: <http://www.ijcai.org/Past%20Proceedings/IJCAI-97-VOL1/PDF/003.pdf>
- [19] PIERRAKOS, D. et al. "Web usage mining as a tool for personalization: A survey". En *User modeling and user-adapted interaction* [en línea]. Vol. 13, n° 4, (2003) pp. 311-372. Disponible en: <http://link.springer.com/article/10.1023/A:1026238916441>
- [20] PITKOW, J. E., BHARAT, K. A. "Webviz: A tool for world wide web access log analysis". (1994). Disponible en: <https://smartech.gatech.edu/handle/1853/3593>
- [21] ROMERO, C. et al. "Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems". En *Computers & Education* [en línea]. Vol. 53, n° 3, (2009) pp. 828-840. Disponible en: <http://www.sciencedirect.com/science/article/pii/S0360131509001146>
- [22] SANTOS, O. C., BOTICARIO, J. G., BARRERA, C. "aLFanet: An adaptive and standard-based learning environment built upon dotLRN and other open source developments". En *Foro hispano de. LRN. Spain*, (2005).
- [23] SHAHABI, C. et al. "Knowledge discovery from users web-page navigation". En *Research Issues in Data Engineering, 1997. Proceedings. Seventh International Workshop on* [en línea]. IEEE, (1997). pp. 20-29. Disponible en: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=583692>
- [24] SOW, D. M. et al. "Prefetching based on web usage mining". En *Middleware 2003* [en línea]. Springer Berlin Heidelberg, (2003). pp. 262-281. Disponible en: http://link.springer.com/chapter/10.1007/3-540-44892-6_14
- [25] SPILIOPOULOU, M., FAULSTICH, L. C. "Wum: A web utilization miner". En *International Workshop on the Web and Databases, Valencia, Spain* [en línea]. (1998). Disponible en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.137.1712&rep=rep1&type=pdf>
- [26] SRIVASTAVA, J. et al. "Web usage mining: Discovery and applications of usage patterns from web data". En *ACM SIGKDD Explorations Newsletter* [en línea]. Vol. 1, n° 2, (2000) pp. 12-23. Disponible en: <http://dl.acm.org/citation.cfm?id=846188>
- [27] WENG NGU, D. S., WU, X. "SiteHelper: a localized agent that helps incremental exploration of the World Wide Web". En *Computer Networks and ISDN Systems* [en línea]. Vol. 29, n° 8, (1997) pp. 1249-1255. Disponible en: <http://www.sciencedirect.com/science/article/pii/S016975529700055X>
- [28] WU, K.-L., YU, P. S., BALLMAN, A. "Speedtracer: A web usage mining and analysis tool". En *IBM Systems Journal* [en línea]. Vol. 37, n° 1, (1998) pp. 89-105. Disponible en: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5387142>

- [29] ZAIANE, O. R., XIN, M., HAN, J. “Discovering web access patterns and trends by applying OLAP and data mining technology on web logs”. En *Research and Technology Advances in Digital Libraries, 1998. ADL 98. Proceedings. IEEE International Forum on* [en línea]. IEEE, (1998). pp. 19-29. Disponible en: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=670376>
- [30] ZARKESH, A. M. et al. “Analysis and design of server informative WWW-sites”. En *Proceedings of the sixth international conference on Information and knowledge management* [en línea]. ACM New York, (1997). pp. 254-261. Disponible en: <http://dl.acm.org/citation.cfm?id=266906>

8. Lista de enlaces

- IEEE Xplore Digital Library: <http://ieeexplore.ieee.org>
- Springer Link: <http://link.springer.com>
- ACM Digital Library: <http://dl.acm.org>
- ScienceDirect: <http://www.sciencedirect.com>
- CiteSeer: <http://citeseerx.ist.psu.edu>
- Scholarly Materials and Research at Georgia Tech Library (SMARTech): <https://smartech.gatech.edu>
- International Joint Conferences on Artificial Intelligence (IJCAI): <http://www.ijcai.org>