

Web Usage Mining, aplicado a servidores Web Apache

Julián Alberto Monsalve Pulido¹

Grupo de Investigación GIBRAN, Facultad de Ingeniería de Sistemas.

Universidad Santo Tomás Seccional Tunja Boyacá, Colombia.

jmonsalve@ustatunja.edu.co, jmsistemas@gmail.com

Resumen

El presente artículo muestra algunos avances de investigación en el análisis de patrones de comportamiento de usuarios en el uso de servidores Web Apache utilizando técnicas de minería de datos. Como resultado de la investigación, se presenta un análisis teórico y un proceso de ingeniería para la creación de un software que pretende mejorar los procesos de minería Web para la toma de decisiones en la aplicación de técnicas científicas y analíticas en la informática forense, principal objetivo de la investigación. La investigación se centró en el desarrollo de una aplicación web de libre distribución, donde carga el archivo log de conexión Apache y aplica los procesos de minería de datos para descubrir comportamientos de usabilidad en los usuarios y posteriormente realizar informes que ayudarán a la detección de problemas de seguridad, análisis de informática forense y análisis de web usage mining.

Palabras clave:

Minería web, usabilidad, Apache.

Abstract

This article shows some research advances in the analysis of user behavior patterns in the use of apache web servers, using data mining techniques. As a result of the investigation presents a theoretical analysis and software engineering process for creating software that aims to improve Web mining processes for decision making in the application of scientific and analytical techniques in computer forensics, main objective of the investigation. The research is centered on developing a web application for free distribution, which loads the apache log connection and applies data mining process to discover usability behaviors in users and then make reports to help detect security issues, computer forensics and analysis of web usage mining analysis.

Keywords:

Web mining, usability, Apache.

¹ Ingeniero de Sistemas, Magister en Software libre con especialidad en desarrollo de aplicaciones de la Universidad Autónoma de Bucaramanga - Universitat Oberta de Catalunya España.

Introducción

El análisis de usabilidad de los sitios Web en los portales de las empresas brinda al usuario poder participar en los procesos de diseño y desarrollo dentro del portal garantizando la calidad de la información e interacción. Esta investigación se centra en un portal universitario donde se tomó las muestras para la investigación y presenta reportes de usabilidad del mismo.

Se desarrolló un software para garantizar la aplicación de la técnica de la minería Web por medio de un algoritmo apriori donde identifico las reglas de asociación para la identificación de comportamientos de uso de los usuarios en el portal.

Algunos resultados de la investigación se presentan por medio de informes y gráficas estadísticas que genera la aplicación desarrollada después de hacerle minería de datos a la información recolectada por el Log de conexión del servidor Web Apache.

Este artículo se estructura en cuatro capítulos: el primero muestra la conceptualización básica de servidores web y de usabilidad; en la segunda parte, se describe la obtención de datos del servidor; en la tercera parte, se describe el proceso de desarrollo de la herramienta de libre distribución y, por último, se muestra las conclusiones de la investigación y referencias que se utilizaron en el proceso.

Conclusiones

Medir la usabilidad de un sitio web, brinda al Webmaster una gran ayuda para la administración, da los argumentos para generar estrategias de cambios de contenidos, estructura y uso.

Actualmente, los modelos de negocios incluyen el cumplimiento de los requerimientos, como publicidad, cantidad de usuarios perfilados, líneas de comunicación enmarcadas en la satisfacción del cliente y desarrollos basados en la opinión y criterios del usuario final.

La minería Web beneficiará a la evolución de los negocios que tiene presencia en Internet, ya que se puede mejorar cada día la usabilidad del mismo para que el usuario sea protagonista del proceso empresarial.

El desarrollo y el potencial del Web Mining permite detectar información invisible, pero de gran importancia para consolidar y ampliar el criterio de la W3. La determinación de los patrones de conducta se establece como redes de relaciones existentes que permiten identificar grupos homogéneos de usuarios para encauzar sus intereses comunes al desarrollo de grupos participativos y líneas de investigación con personas dedicadas a temáticas afines.

La informática forense es una ciencia que identifica delitos informáticos, pero los resultados son difíciles de identificar por la cantidad de información. Una solución es la aplicación de técnicas de minería de datos por medio de una herramienta de libre distribución.

Referencias

- [1] MapR Technologies, Inc.; MapR Technologies Signs Corporate Contributor License Agreement for Apache Software Foundation. (2011, June). Computers, Networks & Communications, 387. Retrieved June 14, 2011, from ProQuest Computing. (Document ID: 2369059991).
- [2] Hernández, J.; Ramírez, M.J.; Ferri, C. (2004); "Introducción a la Minería de Datos", Pearson Educación, Madrid. 2004.

- 
- [3] COOLEY, Robert. Discovery and Application of Interesting Patterns from Web Data. Ph. D. Thesis. University of Minnesota. May 2000.
 - [4] MOBASHER, Bamshad. Web Usage Mining and Personalitation. CRC. Estados Unidos: Press LLC, 2004, 129 p.
 - [5] Raymond Kosala, Hendrik Blockeel. Web Mining Research: A Survey. SIGKDD Explorations. ACM SIGKDD. July 2000.
 - [6] S. K. Madria et al. Research issues in web data mining. Proceedings of Data Warehousing and Knowledge Discovery, First International Conference, DaWaK '99. 1999.
 - [7] Sankar, Paul. Web Mining in soft Computing Framework. Estados Unidos: World Scientific Publishing Co Pte Ltd. 2004, 133 p
 - [8] Witten & Frank, Clark, P.; Boswell, R. (2000); "Data Mining. Practical Machine Learning Tools and Techniques with Java Implementations", Morgan Kaufmann Publishers, 2000.
 - [9] Jeimy J. Cano, Introducción a la informática forense, Revista Sistemas Seguridad y computación forense ACIS. Junio 2006.
 - [10] Botia Valderrama, Diego José Luis. Aplicación de las herramientas de software libre Sleuthkit y Autopsy a la informática forense, Revista Intekhnia Volumen 3 No 7 Diciembre 2008.