

# WEB USAGE MINING PARA LA IDENTIFICACIÓN DE PATRONES DE COMPORTAMIENTO DE USUARIOS, MEDIANTE EL USO DE HERRAMIENTAS TECNOLÓGICAS



**Autor:**  
**Álvaro Pérez Niño**

## **Palabras Clave:**

Web Usage Mining, Patrones de comportamiento, Website.

## **Introducción**

Las herramientas de minería y análisis de datos, toman mayor relevancia cuando se afrontan grandes volúmenes de datos (Smyth, 2000) . En la actualidad los sitios y aplicaciones web ofrecen una gran oportunidad de contacto con los usuarios. Los servidores web registran de manera permanente los datos relacionados con las acciones realizadas por sus usuarios. Teniendo de esta manera registros que incluyen valores relacionados con la navegación de los clientes, tiempos de acceso, fechas, recursos buscados, respuestas del servidor ante las peticiones y orígenes de navegación. Es allí donde debemos aplicar los beneficios de la minería de datos y minería web y buscar el fortalecimiento a sistemas de información que ayuden a fortalecer las decisiones de la alta gerencia en las organizaciones (Y. M. Chae, 2003).

El sitio web representa un canal de alto impacto en la medición de niveles de comunicación con la comunidad y en el Plan de Medios de la Organización se constituye como componente clave de la estrategia de difusión y de igual forma, contempla valores que pueden ayudar a comprender indicadores. Se resalta la importancia de los valores éticos y profesionales en el desarrollo del estudio (Drury, 2003)

Abordamos el presente estudio procesando los datos almacenados en el log del servidor Web de un Hospital Público de Alta Complejidad del Municipio de Manizales Caldas y mediante el software libre Analog versión 6.0 y la versión de evaluación de Advanced Log Analyzer, visualizaremos información estadística que permitirá a la Institución tomar acciones frente a los patrones y comportamiento de los usuarios que navegan el website; mediante la generación de reglas de decisión y técnicas de agrupamiento.

## **Metodología**

De acuerdo al objetivo de la Investigación, se procede en primera fase a la extracción de información contenida en un log de acceso de los usuarios en un periodo de una semana; compuesta por la selección de 94.895 registros iniciales, seguido del pre procesamiento de los datos: donde se genera un log libre de ruido de 12.650 correspondiente a la limpieza y preparación de los mismos. Posteriormente se procede a la construcción de un modelo descriptivo que permite identificar patrones en los datos sin indicadores externos que guíen al algoritmo es decir, sin conocer la realidad “a priori”.

En la fase final se ejecutó el análisis de la información mediante técnicas de minería de datos las cuales son aplicadas de acuerdo al análisis realizado y las técnicas definidas después del proceso de extracción; en donde se generaron

diez reglas de asociación teniendo en cuenta el acceso a los usuarios y su comportamiento, en donde se valida el modelo obtenido contra los objetivos de negocio definidos para el estudio.

### Análisis de Resultados

Mediante el uso de las herramientas tecnológicas para el análisis estadístico y teniendo como resultado un promedio de 1,220 peticiones diarias exitosas y de las horas más afluentes de acceso, que se encuentran entre las 11:00 y 14:00 Horas; se evidencia estabilidad en el servicio del website, sujeta a su arquitectura empresarial y a la organización jerárquica de las secciones, generando diversidad de contenidos que influye en el interés de los usuarios en su navegabilidad y en la descarga de contenidos, potencializando el website del hospital como epicentro de divulgación de información pública relacionada con contratación, convocatorias y documentos de legislación entre los usuarios internos y externos. Para estudios posteriores, se recomienda segmentar la vista minable y realizar un mayor entrenamiento de los datos; de tal manera que

se puedan analizar casos particulares de la tendencia encontrada y de igual forma realizar agrupamientos de secciones de acuerdo a su objetivo, de usuarios y de estado del web services que permitan profundizar en las peticiones de los usuarios y los contenidos navegados, debido a que son rezagados por la tendencia dominante.

### Referencias

P. Smyth, “Data mining: data analysis on a grand scale?,” *Stat Methods Med Res*, vol. 9, no. 4, pp. 309–327, 2000

Y. M. Chae, H. S. Kim, K. C. Tark, H. J. Park, and S. H. Ho, “Analysis of healthcare quality indicator using data mining and decision support system,” *Expert Systems with Applications*, vol. 24, no. 2, pp. 167–172, Feb. 2003

J. Ma and C. G. Drury, “The Human Factors Issues in Data Mining,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 47, no. 4, pp. 716–720, Oct. 2003

**Autor:**

Álvaro Pérez Niño  
majash29@hotmail.com

**Grupo de Investigación**

**GESICOM**

Integrador del área de virtualización  
Centro de Comercio y Servicios  
Regional Tolima