

Estudio webmétrico de la revista electrónica Avanzada Científica.

Webmetric Study of Advanced Scientific electronic journal.



MSc. Lic. Silvio Curiel Lorenzo

Licenciado Especialidad Ruso, Español y Literatura.
Técnico en Ciencias Computacionales. Profesor Adjunto.
Categoría Docente: Instructor.
Centro de Información y Gestión Tecnológica de Matanzas.
(CIGET). Cuba
Telf: (45) 242483 E-mail: silvio@cigetmtz.atenas.inf.cu



Ing. Alayna Pantoja Trincado

Especialista Informática
Centro de Información y Gestión Tecnológica de Matanzas.
(CIGET). Cuba
Telf: (45) 242483

Recibido: 14-11-14

Aceptado: 05-01-15

Resumen:

Se realiza un estudio *webmétrico de uso* a la revista electrónica Avanzada Científica a través de sus ficheros log entre los años 2011 y 2014. Concretamente se investiga desde el 22 de septiembre de 2013 al 23 de septiembre de 2014 y se compara con un anterior trabajo que se enmarcó entre el 5 de septiembre de 2010 al 28 de julio de 2011. Se realiza una amplia búsqueda de información en la web. Se aplica el software Sawmill 7. La comparativa se realiza a los indicadores de visibilidad: análisis de enlaces, accesos a páginas, por usuario, por país de origen, petición, visitas, tipo de navegador y sistema operativo utilizado para consultar la revista, cantidad de hits, tipos de errores y sus códigos de identificación, ciudades, días de la semana y horarios entre otros no menos importantes.

Palabras clave: Revista electrónica, Internet, Indicadores de visibilidad

Abstract:

One study webmetric use the electronic journal Advanced Scientific is done through its log files between 2011 and 2014. Specifically it investigates since September 22, 2013 to September 23, 2014 and compared with a previous work that framed between September 5, 2010 July 28, 2011. An extensive search of information is done on the web. Number link analysis, access to pages, by user, by country of origin, request, visits, type of browser and operating system used to consult the journal: Sawmill software applies 7. The comparison is made visible indicators of hits, types of errors and their identification codes, cities, days of the week and times among other no less important.

Keywords: Electronic magazine, Internet, Visibility Indicators

Introducción:

Al final de los años noventa Cibermetrics y Webometrics entraron en la existencia aplicando la bibliometría, la cienciometría y la infometría a los estudios de Internet. (Arroyo Vázquez, Ortega Priego et al. 2005).

Por el 1997, los investigadores norteamericanos Micheal Cox y David Ellsworth, afirmaron “que el ritmo de crecimiento de los datos empezaba a ser un problema para los sistemas informáticos actuales” y le llamaron “el problema del Big Data”. La Wikipedia, define el *Big Data* como “el término que se refiere a un conjunto de datos tan grande y complejo, que resulta difícil de procesar usando los sistemas de gestión de bases de datos disponibles o las aplicaciones tradicionales de procesamiento de datos”. (Gil 2014)

Se plantea que para el 2020, el Big Data siga creciendo y llegue al 4300% en la generación de datos anuales, “gracias a los cambios tecnológicos acelerados de transición de lo analógico a lo digital, y al rápido aumento en la generación de datos, tanto por particulares como por grandes empresas”. (Gil 2014).

Hoy se mide el peso de la información hasta el **exabyte** y ya se manejan conceptos hasta el **zetabyte** (Gil 2014) y se plantea que para el 2020, la producción de datos será 44 veces superior a la de 2009. Debido a esta exponencial producción, se han desarrollado herramientas para “entender” esta información que circula en internet. Una de ellas en la webmetría.

Según (Medina Pagola 2005) y (Zhu S/A), la Minería de Web es la aplicación de la minería de datos o de otras técnicas de procesamiento de informaciones a la WWW, para encontrar patrones útiles que se espera ayuden a las personas a acceder a la WWW más eficientemente.

La Webmetría es el estudio de los aspectos cuantitativos de la construcción y uso de los recursos de información, estructuras y tecnologías de una parte concreta de Internet, por regla general a una web o portal, desde perspectivas bibliométricas e informétricas.

El llamado Web mining o minería de web, es una herramienta que se utiliza para el estudio de aquellas “huellas digitales”, entiéndase, cookies, navegadores, direcciones IP, páginas web visitadas, etc, las cuales son almacenadas en los ficheros .log.

La Web mining tiene entre otros objetivos: conocer el comportamiento de los cibernautas, buscar e identificar información que es consumida por los internautas y crear valor añadido.

La literatura reconoce tres tipos de minería web. Estas son: **Minería web de contenido**: La minería de contenido del web trata de extraer información relevante sobre el contenido de la web de manera que pueda ayudar clasificarlo, aumentando la organización de ese contenido, para posteriormente mejorar el acceso y la recuperación de la información en él contenida. La **Minería web de estructura** que se usa para conocer cómo está organizada una web, cómo está estructurada y cómo es la navegación a través de ella y la **Minería de uso del web** que trata de extraer patrones de uso del web por parte de los usuarios. Para ello se utilizan los archivos .log de los servidores Web de forma que aplicando minería de textos sobre ellos se pueda extraer información útil. Este tipo de minería tiene como objetivos principales: identificar patrones generales de uso de un sitio web de manera que se pueda reestructurar para que sea más fácil de utilizar y mejore el acceso por parte de los usuarios, y obtener perfiles de los distintos tipos de usuarios a través de su comportamiento y navegación, para poder atender de forma más personalizada.

La contribución al desarrollo de la sociedad, lo podemos hacer posible, entre otros elementos o interfaces, a partir de las revistas electrónicas basadas en la plataforma Open Journal System. El Open Journal Systems(OPJ) es un sistema de administración y publicación de revistas científicas y académicas en Internet. El sistema está diseñado para reducir el tiempo y energías dedicadas al manejo exhaustivo de las tareas que involucra la edición de una publicación seriada. Este sistema permite un manejo eficiente y unificado del proceso editorial, con esto se busca acelerar el acceso en la difusión de contenidos e investigación producido por las universidades y centros de investigación productores del conocimiento. Así mismo, busca consolidarse como una herramienta con innovaciones que permiten el acceso en texto completo de los documentos publicados. (Curiel Lorenzo 2008).

Según la Asociación Americana de Bibliotecas (ALA) la revista científica es una publicación periódica que publica artículos científicos y/o información de actualidad sobre investigación y desarrollo acerca de un campo científico determinado. Por su parte las normas ISO (International Standardization Organization)_consideran que las revistas científicas son una publicación en serie que trata generalmente de una o más materias específicas y contiene información general o información científica y técnica.(Aguirre Cabrera 2006).

La revista "Avanzada Científica" fue fundada en diciembre de 1997, como la revista del Polo Científico Productivo y es editada por el Centro de Información y Gestión Tecnológica (CIGET) de la provincia de Matanzas. Desde el año 1998 es una revista electrónica con frecuencia cuatrimestral, arbitrada, a ciegas y por revisión por pares. Esta registrada como revista científica por resolución del CITMA y se encuentra ubicada en grupo III, en la clasificación que hace el Ministerio de Educación Superior (MES) de las revistas para publicar.

Cuando hablamos de impacto en la ciencia nos podemos referir a cuatro efectos ventajosos entre los que se encuentra el efecto demostrativo, el efecto orientador, el motivador y el integrador a través del estudio de la producción científica de las revistas electrónicas. Teniendo en cuenta estos efectos, consideramos que dentro de los principales destinos a considerar para medir el impacto, podemos mencionar que "Avanzada" contribuye al desarrollo de la sociedad cubana y matancera a través de sus recomendaciones sobre aspectos importantes del desarrollo de la sociedad cubana.(Curiel Lorenzo 2010).

El objetivo de este trabajo es hacer un estudio **webmétrico de uso** a la revista electrónica Avanzada Científica a través de sus ficheros log entre los años 2011 y 2014. Concretamente se investiga desde 22 de septiembre de 2013 al 23 de septiembre de 2014 y se compara con un anterior trabajo que se enmarcó entre el 5 de septiembre de 2010 a 28 de julio de 2011

Materiales y métodos:

Se realizó una amplia búsqueda de información en la web. Se aplica el software Sawmill para realizar el estudio webmétrico de uso y se realizó el respectivo análisis de información utilizando Excel y Endnote X7.

Resultado y discusión:

La comparativa se realiza a los indicadores de visibilidad: análisis de enlaces, accesos a páginas, por usuario, por país de origen, petición, visitas, tipo de navegador y sistema operativo utilizado para consultar la revista, cantidad de hits, tipos de errores y sus códigos de identificación, ciudades, días de la semana y horarios entre otros no menos importantes.

En el estudio realizado en el año 2011, pudimos comprobar que:

Fueron visitadas **110,451** páginas por **8873** visitantes, siendo el mes de **abril**, el más visitado y el **jueves** el día de mayor visita y el período de tiempo más activo el de **12:00 PM y las 3:00 PM**. La revista fue visitada por **77 países** de los cinco continentes, donde se destacan por cantidad de visitas: Cuba, Estados Unidos, México, España, Federación Rusa, Perú, Venezuela, Canadá, Chile y Colombia. Se destacaron Cuba, la ciudad de Ciego de Ávila con 28,878 páginas visitadas, Camagüey (6,164), Matanzas (130), Holguín(59), Santi Spiritus(45), Pinar del Río(12) y Santiago de Cuba(6). De Estados Unidos, el estado de California es el que más nos visitó; de México, el Distrito Federal, Jalisco, Nuevo León y de España, Madrid, Cataluña, Andalucía y el País Vasco entre otros. **Desde nuestro país** se enlazaron a través de la url de la propia revista, <http://avanzada.idict.cu>, www.google.com.cu, www.google.com y www.cubasi.cu. Los **motores de búsqueda** por donde más nos localizaron fueron el Google y Yahoo.

Las frases que más utilizaron los internautas para buscar en la revista fueron: evaluación del desempeño, Valle del Yumurí, sexualidad responsable, pitiriasis rosada, precipitaciones en Matanzas, cultura organizacional, Avanzada Científica, revista Avanzada Científica y tesis doctorales, economía y Cuba. También: pie diabético, gestión de proyectos, algoritmo y sus aplicaciones, Enerplant y estudio de azúcares reductores **Los web browsers o navegadores** más utilizados por los internautas cubanos fueron: Internet Explorer, Firefox, Opera, Mozilla, Safari, PKP-OJS. **Los sistemas operativos** desde los que nos visitaron: Windows XP, Windows NT 6.1, Linux

Los códigos de respuestas del servidor más significativos son:

200: 46.2% de respuesta correctas.

304: 38.8% No modificado. Indica que la petición a la URL no ha sido modificada desde que fue requerida por última vez. Típicamente, el cliente HTTP provee un encabezado como If-Modified-Since para indicar una fecha y hora contra la cual el servidor pueda comparar. El uso de este encabezado ahorra ancho de banda y procesamiento tanto del servidor como del cliente.

404: 14.9% No encontrado Recurso no encontrado. Se utiliza cuando el servidor web no encuentra la página o recurso solicitado.

Desde el extranjero, los internautas interactúan de la forma siguiente:

Descripción de los enlaces de origen.

Se enlazan a nosotros o nos ven a través de sitios y páginas web tales como: <http://avanzada.idict.cu>, www.google.es, <http://scholar.es>, <http://dialnet.uniroja.es>, <http://scholar.google.com>,

Los dominios que más nos visitan son:

- a) 169.158.180.10: Cuba. Centro Nacional de Intercambio Automatizado
- b) 67.195.110.168: Estados Unidos. Illinois, Bloomington
- c) 190.6.84.21. Cuba. Red de Cultura
- d) 67.195.115.107: Estados Unidos. California.Sunnyvale
- e) 87.250.254.241: Federación Rusa. Yandex LLC.

En el estudio realizado este año 2014, se debe tener en cuenta, que por un período de casi 6 meses del 2013, por problemas técnicos en el servidor donde se encuentra hospedada, la revista se mantuvo inoperable.

Teniendo en cuenta lo anterior, el 2013 se presenta con los meses de octubre y noviembre como los de mayor visita.

Las entradas aceptadas o hits del período son: 450 909, para un 2013 con 18 092 y un 2014 con 432 617. Para el 2014, se presenta un total de 432.617 hits y son los meses de marzo y julio en los que más nos visitan. Los jueves, viernes y sábados son los días de la semana que más nos visita y de 9: a las 11 de la mañana el horario del día más visitado.

Entre los países que más nos visitan en este periodo y en comparación con el 2011, los Estados Unidos, pasan al primer lugar, desplazando a nuestro país. Alemania aparece por primera vez en un cuarto lugar, con Méjico, España, Canadá, China, Perú y Colombia entre los 10 primeros. Nos visitan 77 países de los 5 continentes y todos del llamado primer mundo.

De los EE.UU, los Estados que más nos visitan son: California, New York, Washington DC, Colorado, Arizona y Florida.

De las ciudades de España las que más nos visitan son: Madrid, Cataluña, Andalucía, Murcia, Valencia, Canarias, Castilla y León, País Vasco y Galicia.

De las ciudades de nuestro país se mantiene como la que más nos visita la ciudad de Ciego de Ávila, seguida de Camagüey, Cienfuegos, Holguín, Pinar del Río, Santiago de Cuba, Matanzas y Santis Spiritus.

El código de errores que más se manifiesta es el 404 No encontrado, que se da cuando el servidor web no encuentra la página o recurso solicitado. Los países y/o IP que más se destacan coinciden con los que más nos visitan.

Conclusiones:

En el ámbito de la ciencia y la innovación tecnológica, la medición de sus resultados por los llamados indicadores de impacto –cuyo principal objetivo es evaluar el beneficio tangible, la repercusión del resultado y no el resultado en sí– constituye un estadio superior en la evaluación de la producción científica e innovadora de investigadores y tecnólogos cubanos. Por lo que este estudio webmétrico de uso refuerza estos indicadores de impacto como el de autores y su calidad profesional, cantidad de visitante, cantidad de países que nos visitan, a la cantidad de empresas

que publican las cuales se relacionan con las empresas priorizadas en el territorio, las temáticas tratadas, también se relacionan con los proyectos de desarrollo científicos productivos, aunque no sean objetivo de este estudio, pero que están contenido.

Teniendo en cuenta, que el principal objetivo de la revista electrónica Avanzada Científica es lograr visualizar la producción científica y la ciencia matancera y cubana, los resultados de las investigaciones que constituyen novedad al menos para el país y cuya publicación concite el interés de los científicos, nacional e internacionalmente, sin importar que su aplicación sea o no inmediata, podemos considerarla una fuente principal del futuro impacto de esas propias investigaciones y resultados, por lo que este tipo de análisis nos aseguran y nos ayudan a direccional el trabajo de visualización, considerando que Avanzada contribuye a generar información útil que permite mejorar el proceso de toma de decisiones, tanto en el desarrollo de la ciencia y tecnología, como el desarrollo social, visualizando más allá de nuestras fronteras el desarrollo de la provincia.

De esta manera, el estudio presentado nos ayuda a implementar las conocidas "medidas de visibilidad e impacto" que exploran los patrones de enlace entre páginas y sedes distintas El número y diversidad de enlaces externos recibidos, el volumen de los mismos respecto los contenidos objeto de enlace (el famoso factor de impacto Web) o índices que se construyen de acuerdo al peso relativo de las sedes de origen de los enlaces (PageRank de Google del cual Avanzada Científica muestra el 6 de 10).

Referencias bibliográficas:

Aguirre Cabrera, M. (2006). Definición y Gestión de una Revista Científica. I Taller para Editores y Autores Científicos. La Paz, CONICYT.

Arroyo Vázquez, N., et al. (2005). Cibermetría. Estado de la cuestión.

Curiel Lorenzo, S. (2008). Algunos apuntes sobre el Open Journal Systems. Matanzas. Notas de trabajo.

Curiel Lorenzo, S. L. (2010). "La Unidad de Inteligencia Empresarial, la toma de decisiones su y impacto social." Revista Arquitectura e Ingeniería. Vol. 4, No. 1, 2010. 4(No. 1).

Gil, D. (2014). "Big Data y marketing digital: una relación imprescindible." from <http://www.educacionline.com/instituto-de-marketing-online/>.

Medina Pagola, J. E. (2005). Estado del Arte del Web Mining . Reporte Técnico Minería de Datos. S. Gris. La Habana, Centro de Aplicaciones de Tecnologías de Avanzada.

Zhu, T. (S/A). "Web mining is the application of data mining or other information process techniques to WWW, to find useful patterns which are expected to help people access WWW more efficiently."." Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada. Home page: <http://www.cs.ualberta.ca/~tszhu/>