

Rutas de navegación empleadas por los usuarios para acceder a los contenidos disponibles en el sitio web de Infomed

[Adrián Coutin Domínguez](#)¹

Resumen

A partir de los datos aportados por Webalizer, un software para el procesamiento de los ficheros logs del servidor, se identificaron las rutas utilizadas por los usuarios para navegar en las estructuras jerárquicas del sitio web de Infomed y se realizó su diagramación, sin embargo, se detectó un grupo de URLs, que si bien poseían un alto número de accesos, no era posible relacionarlos mediante sus enlaces con otros URLs del mismo sitio. Con la hipótesis de que estos se accedían desde enlaces externos, que apuntan a determinados URLs del sitio de Infomed, incluidos los motores de búsquedas, se procesaron los ficheros logs del servidor web, en los cuales es posible identificar los sitios desde donde se accedían a los URLs que no mostraban acceso por ninguna de las estructuras jerárquicas del sitio web. Finalmente, se reelaboró el diagrama donde es posible apreciar 3 rutas de navegación en el sitio: una, que corresponde con la estructura jerárquica del sitio, otra donde los accesos se realizan mediante enlaces externos, y una tercera, relacionada con los accesos a partir de los motores de búsqueda. El análisis de ese diagrama permitió conocer las rutas de navegación que siguen los usuarios para acceder a los contenidos y servicios, cuáles de estas son las más utilizadas en los diversos niveles que integran la estructura jerárquica del sitio, entre otros resultados, lo cual argumenta que los estudios estadísticos de la navegación ofrecen indiscutibles ventajas al personal responsable de la administración y el desarrollo de los sitios.

Descriptor (DeCS): INTERNET

Descriptor (DeCI): SITIOS WEB; INFOMED; RUTAS DE NAVEGACION; FICHEROS LOGS; MOTORES DE BUSQUEDA

Abstract

Starting from the contribution made by Webalizer, a software for processing log files from the server, the routes used by users to sail in the hierarchical structures of Infomed website were identified and diagraming was carried out. Nevertheless a group of URL with a high number of accesses was noticed but was imposible to relate them using their links with other URL of the same site. Having the hypothesis that they were accessed from external links that pointed out to determined URL from Infomed website, including search engines, the log files from web server were processed, in which is possible the identification of the sites that made possible the access to the URL that did not show access by any hierarchical structure of the website. Finally a diagram showing the three routes of navigation in the site was devised: one that corresponds with the hierarchical structure of the site, another where the access is fulfilled by means of external links, and third related to the access from search engines. The diagram analysis showed the routes of navigation followed by the users to access contents and services, which are the most frequently used in the levels that integrate the hierarchical structure of the site among other results which argues that estatistical studies of navigation offer evident advantages to the personal responsible for the administration and development of the sites.

Subject headings (DeCS): INTERNET

Subject headings (DeCI): WEB SITE; INFOMED; NAVIGATION WAY; LOGS FILE; SEARCH ENGINE

Los estudios soportados en métodos cuantitativos para la evaluación de los sitios web se han concentrado en el análisis estadístico de sus patrones de uso, a partir de los ficheros logs de los servidores web.¹⁻⁴ De acuerdo con los supuestos establecidos por estos trabajos, muchos arquitectos de información, webmasters, evaluadores e investigadores en el tema de las e-métricas, han realizado diferentes estudios para mejorar su conocimiento sobre la usabilidad y efectividad de los sitios mediante el uso de las trazas dejadas por los clientes en los ficheros logs de los servidores. De una forma u otra, dichos estudios buscan obtener resultados rápidos y prácticos que faciliten la toma de decisiones relacionadas con la arquitectura de la información, así como sobre la usabilidad y el impacto económico de los contenidos y servicios disponibles en un sitio.

El presente trabajo, continuación de otros relacionados con la e-métrica del sitio de Infomed,⁵ se propone identificar cómo los usuarios acceden a los contenidos del sitio, cuáles son los niveles jerárquicos que más se utilizan y los URLs preferidos. Asimismo, se identificarán los esquemas de navegación jerárquica empleados por la audiencia y los URLs externos donde se encontraban ubicados los usuarios, antes de acceder al sitio en estudio. De esta forma, pueden obtenerse las rutas de navegación que establecen los usuarios para acceder a sus contenidos y servicios.

Métodos

Webalizer, el software para el análisis de los ficheros logs de los servidores web que se emplea para el sitio infomed.sld.cu desde hace 3 años,⁵ ofrece mensualmente un conjunto de datos claves para conocer su funcionamiento, por ejemplo: el número de accesos, los 30 URLs más accedidos en el sitio, las principales páginas de entrada y salida, entre otros. Estos datos se procesan y almacenan mensualmente en una base de datos *Access Microsoft* para facilitar su posterior análisis. Para este trabajo se analizó el período de un año, mayo del 2001 a abril del 2002. Posteriormente, se seleccionó una muestra de ficheros logs para identificar los URLs externos desde los cuales accedieron los usuarios al sitio.

Uno de los datos estadísticos mensuales que ofrece Webalizer es una relación de los 30 URLs más accedidos entre los disponibles en el sitio. Dicho informe incluye todos los tipos de ficheros que forman la página web disponible en el URL, como los de imágenes y los de sonido. A partir de la información estadística almacenada en la base de datos en Access, se seleccionaron los URLs, con contenidos y servicios, accedidos por la audiencia del sitio. Debido al diseño de Infomed, fue necesario incluir URLs que albergaran ficheros con extensiones html, htm, cgi y php. Los resultados se organizaron según el número de accesos de cada URL y se elaboró un diagrama para representar los diferentes niveles jerárquicos.

En el diagrama cada caja representa un URL específico del sitio y las líneas muestran los enlaces utilizados por los usuarios para navegar desde estos o hasta ellos. Con esta representación, es posible observar cómo los usuarios navegan en el sitio y cuáles URLs, dentro del conjunto de los más accedidos mensualmente, se acceden desde enlaces externos al sitio. Las cajas se colocaron en el diagrama de acuerdo con los números totales de accesos a los URLs. Aquellos colocados al inicio de cada nivel son los URLs con mayor número de accesos.

Debido a que Webalizer no ofrece la información de los sitios de procedencia de la audiencia, se hizo necesario procesar los ficheros logs del sitio web donde quedan registrados los URLs de donde provienen los usuarios que acceden al sitio, incluidos los motores de búsqueda.

El procesamiento de los ficheros logs no es una tarea sencilla. El volumen de los datos y de los ficheros a procesar es alto. Un fichero log, en dependencia del tamaño, diseño y accesos del sitio, puede tener más de 15 megas de tamaño y se generan diariamente, de ahí la complejidad de su procesamiento. En consideración a estas dificultades y el tipo de búsqueda que se debía realizar en estos ficheros: identificar cadenas de textos correspondientes a URLs no pertenecientes al dominio *sld.cu*, se decidió utilizar el software *Perl* para desarrollar un programa que facilitara procesar los ficheros logs de interés.

El programa, que utiliza poderosas instrucciones relacionadas con las expresiones regulares de Perl,

- Totales de accesos
- Totales de páginas
- Total de visitas
- Total de URLs

Se aprecia un paulatino incremento del número accesos, páginas consultadas, visitas y del total de URLs accedidos a partir de agosto del 2001, momento en que se produce una transformación total de las facilidades de conexión de Infomed con la red nacional (tabla). Por esta fecha, la conexión a la red mediante líneas telefónicas pasó de 53 a más de 250 debido a la instalación de un access server. Esta transformación tecnológica amplió el radio de acción de Infomed en el Sistema Nacional de Salud, para el que está diseñada y provocó, entre otros resultados, un incremento en el uso del sitio web por parte de su comunidad. Esta afirmación puede comprobarse mediante los resultados que muestran los incrementos del total de URLs accedidos a partir del mes de agosto del 2001.

Tabla Datos estadísticos generales del sitio web de Infomed. Período: mayo del 2001- abril del 2002

Mes/ año	Accesos	Páginas	Visitas	Total de URLs
Mayo /01	1610887	514707	134797	37456
Jun /01	1542320	418474	103730	21236
Jul /01	1441034	414154	67679	19998
Agos/01	1479019	369451	69360	23028
Sept/01	2253105	435100	79187	22000
Oct/01	1869587	385841	62009	66894
Nov/01	2986489	545379	94489	142047
Dic/01	3996688	748780	67767	386802
En/02	3893107	701907	86410	204048
Feb/02	3511218	665761	96156	26605
Mar/02	3986424	598690	106123	11824
Abr/02	4702985	729664	112134	11649
Totales	33272863	6527908	1079841	973587

Principales URLs del sitio

Como se mencionó al comienzo de este artículo, Webalizer ofrece la posibilidad de conocer mensualmente los 30 URLs con un mayor número de accesos. Pues bien, de acuerdo con el método descrito anteriormente, fue posible identificar los principales URLs según el total de usuarios que accedieron al sitio en el período analizado y organizarlos según su posición en el diagrama de navegación jerárquica del sitio.

La página principal del sitio posee el mayor número de accesos, 2 332 238, como es característico en este tipo de sitio. Luego, en el primer nivel, con un total de 457 216 accesos, se encuentra en el primer lugar el URL de las *Noticias de la Prensa*, seguido por el motor de búsqueda interno que utilizaba el sitio en esa fecha, HTSEARCH. Otro URL con un alto número de accesos, y por ende asumido como preferido entre la comunidad de usuarios, es el de *Otras Redes*, que posibilita el acceso hacia otros sitios web nacionales no relacionados con el sector salud. Se destaca también la opción *Navegar*, que facilita la navegación según un orden jerárquico en los contenidos y servicios disponibles en el sitio.

En el segundo nivel, con un total de 75 677 accesos, se halla en primer lugar, según el total de accesos, el servicio de información especializada, *Al Día y Diaria*, que recoge diariamente información noticiosa relacionada con la salud, publicada por las principales agencias de prensa del mundo. Luego se destacan un

conjunto de URLs de ese nivel, vinculados con la navegación dirigida en el sitio, como son *navegar/recursos*, que facilita el acceso a un inventario de recursos de información especializados en salud, y *navegar/temáticas* que posibilita la navegación en el sitio mediante un esquema temático. En todos los casos, se aprecia una navegación desde el primer nivel y los usuarios siguen la navegación jerárquica establecida por los diseñadores del sitio.

En el tercer nivel, se encuentran contenidos y servicios cuyo acceso requiere, en muchos casos, mayor destreza por parte del usuario. Se accede a estos mediante la navegación jerárquica pero, en muchas ocasiones, también a partir de los mecanismos internos de búsqueda, la selección del usuario mediante la opción de *Favoritos* de su navegador, etcétera. El URL con mayor número de accesos es el del servicio de *Preguntas al Experto*, en este caso, la versión disponible para el servicio *WebTV*. Luego se destacan tres URLs a los que se llega mediante la navegación desde el nivel anterior.

No obstante a lo descrito por este diagrama, debe señalarse, que si bien es posible visualizar la navegación jerárquica, existe un grupo de URLs en los cuales no es posible apreciar la procedencia o la forma en que los usuarios llegan a ellos. Al procesar los ficheros logs del servidor web fue posible obtener nuevos resultados.

Al igual que en la figura 1, esta vez, los URLs se organizaron de mayor a menor, según el número de accesos, en cada uno de los niveles correspondientes. La diferencia con la figura 2 radica en la utilización de distintos sombreados que permiten identificar las 3 rutas fundamentales empleadas por la audiencia para acceder al sitio.

El primer fondo, de color gris, se utilizó para aquellos URLs que se accedieron de acuerdo con el esquema de navegación jerárquico diseñado para el sitio web de Infomed. El segundo fondo, de líneas horizontales, se aplicó a los URLs que se accedieron desde enlaces externos. El tercer fondo, de líneas perpendiculares, se utilizó para aquellos, cuyo acceso se produjo a partir de los motores de búsquedas disponibles en Internet.

Como puede apreciarse la navegación jerárquica, fondo gris, es mayoritaria y cubre completamente el primer nivel y llega hasta el segundo y tercero sólo en aquellos URLs con indicaciones expresas de navegación en sus etiquetas. La navegación mediante enlaces externos, fondo de líneas horizontales, es minoritaria en todos los niveles del sitio, con mayor presencia en el tercer nivel. Los URLs que se accedieron a partir de motores de búsqueda, fondo con líneas perpendiculares, es mayoritario entre las formas de acceso al segundo nivel. Además, es la que más se utiliza en los dos URLs con mayor número de accesos en este nivel, el servicio de noticias *Al Día*. En el tercer nivel, es el más utilizado para acceder al URL de *Preguntas al Experto* y al buscador del servicio noticioso de *Al Día*.

Consideraciones finales

Los estudios de sitios web, como el realizado en este trabajo, permiten:

- Perfeccionar los esquemas de navegación jerárquicos a partir de la identificación de las mejores rutas de navegación a los contenidos y servicios dispuestos en el sitio.
- Evaluar los contenidos y servicios de acuerdo con los niveles jerárquicos del sitio. Así es posible tomar decisiones relacionadas con cuáles recursos pueden ser necesarios mover a otros niveles y cuáles no se han consultado de un mismo nivel y con calidad similar.
- Identificar rutas de navegación externas utilizadas por la audiencia. Esto permite, entre otras cosas, establecer estrategias de visualización del sitio en Internet, tanto en motores de búsquedas como en sitios donde se pueda captar audiencia, conociendo la importancia de algunos de estos enlaces externos de acuerdo con el número de accesos que provienen a través de estos.

Referencias bibliográficas

1. Pirolli P, Pitkow J. The scent of a site: A system for analysing and predicting information scent, usage, and usability of a web site. Disponible en: <http://www-users.cs.umn.edu/~echi/papers/chi2000/scent.pdf> Acceso: 13 de febrero del 2003.
2. Drott MC. Using web server logs to improve site design. In ACM 16TH International Conference on

- Systems Documentation. New York: ACM Press, 1998. pp 43-50.
3. Sullivan T. Reading reader reaction: A proposal for inferential analysis of web server log files. En: Proceedings of the Human factors and the Web. 3rd Conference on Human Factors & the Web. Denver: US West Communications, 1997.
 4. Theng YL, Marsden G. Authoring tools: Towards continuous usability testing of web documents. Disponible en: <http://www.cs.mdx.ac.uk/staffpages/yinleng/htworkshop.pdf> Acceso: 10 de febrero del 2003.
 5. Coutin A, Peña Escobio R, Núñez Gudás M. Comportamiento de la navegación en el sitio web de Infomed durante el período 2000 - 2001. ACIMED 2002; 10(1):18-44. Disponible en: http://bvs.sld.cu/revistas/aci/vol10_1_02/aci04102.htm Consultado: 4 de abril del 2003.

Recibido: 3 de julio del 2003. Aprobado: 21 de julio del 2003

MsC. *Adrián Coutin Domínguez*. Grupo de Desarrollo de Software. Red Telemática de Salud en Cuba (Infomed).

Calle 27 No. 110 e/ M y N. El Vedado. Ciudad de La Habana, Cuba. CP 10 400. AP 6520. Correo electrónico: acoutin@infomed.sld.cu

1 Licenciado en Información Científico-Técnica y Bibliotecología. Máster en Ciencias. Grupo de Desarrollo de Software. Red Telemática de Salud en Cuba (Infomed). Centro Nacional de Información de Ciencias Médicas.

© 2004 2000, *Editorial Ciencias Médicas*

Calle E No. 452 e/ 19 y 21, El Vedado, La Habana, 10400, Cuba.



acimed@infomed.sld.cu