

B.6. Engañar a los buscadores

Por **Mari-Carmen Marcos**

Marcos, Mari-Carmen. "Engañar a los buscadores".

En: *Anuario ThinkEPI*, 2008, pp. 50-54.



Resumen: Conocer los criterios que tienen en cuenta los buscadores para determinar el ranking de resultados en las búsquedas es clave para los responsables de páginas web, de ahí la importancia del SEO (search engine optimization) en los últimos años. El conocimiento de estas técnicas y su aplicación son ventajosos tanto para los sitios web, que mejoran su posición en el listado, como para los buscadores, pues las técnicas que mejoran el SEO no pretenden sino proporcionar a los buscadores la máxima información sobre el contenido del sitio web. De esta forma, los sitios web evitarán el uso de imágenes y programación sin textos legibles por el robot que rastrea e indiza las páginas, e intentarán enlazar y ser enlazados por páginas web de contenidos similares. El buen uso de las técnicas de SEO se conoce como *white hat* y queda limitado al diseño de sitios web pensando en los usuarios, sin olvidar la forma de trabajo de los buscadores.

En cambio es posible diseñar sitios web pensando prioritariamente en el ranking que harán después los buscadores, e intentar forzarlo engañando a éstos ofreciendo información que las personas no verán (texto escondido o de pequeño tamaño, etc.). Estas técnicas no éticas denominadas *black hat* no sólo van en contra de dar un buen servicio a los usuarios sino que cuando son detectadas por los buscadores se penalizan con una bajada fuerte en el ranking.

Palabras clave: Buscadores, Spam en buscadores, SEO, Optimización de sitios web en buscadores, White-hat SEO, Black-hat SEO.

Title: Tricking search engines

Abstract: Knowing the factors that search engines consider in their ranking results is key for webmasters. This knowledge is an advantage for a website's owner, as the site improves its position in ranking results, and for the search engines because they have access to more information about the site's content to determinate the ranking. Websites that use search engine optimization techniques avoid the use of images without alternative text and try to link and be linked by related content sites. Good use of these search engine optimization (SEO) techniques is called "white hat SEO"; its main feature is that websites are designed with users in mind and without forgetting how search engines work. Nevertheless, web design oriented toward search engines rather than users is possible; it means websites are designed for ranking results, abusing SEO techniques to deceive search engines, such as with hidden text or small fonts, two of the most common strategies. When search engines discover these unethical techniques, called "black hat SEO", websites are punished with poor rankings.

Keywords: Search engines, Spam, SEO, Search engine optimization, White-hat SEO, Black-hat SEO.

COSTUMBRE, PEREZA O COMODIDAD, el caso es que, incluso cuando conocemos la URL del sitio web al que queremos ir, colocamos la palabra clave en el buscador en lugar de teclear el incómodo y largo www.loquesea.com. Esas búsquedas (que más que búsquedas son atajos para teclear menos) más todas las búsquedas "de verdad" hacen que alrededor del 80% de las visitas que reciben los sitios web procedan de un clic rea-

lizado en los resultados de búsqueda, en particular de uno de los tres grandes buscadores: **Google, Yahoo! o Live.**

Dicho esto, hay que añadir que la mayoría de los usuarios de la Web no son expertos en el uso de buscadores, por tanto no conocen las técnicas de búsqueda de las que nosotros podemos alardear; es más, a veces ni siquiera tienen claro cuál es el objeto de búsqueda, o éste es algo más complejo de lo que el bus-

cador es capaz de interpretar. Sea como sea, frente a una consulta el buscador ofrece una serie de resultados, casi siempre una cantidad mucho mayor de la que una persona es capaz de leer en años (a menudo en toda su vida), y el usuario revisará tan sólo los primeros antes de sucumbir a la tentación de pinchar en uno, dos, quizá tres de ellos. Si ese usuario nos interesaba y nuestro sitio no ha sido uno de los afortunados, hemos perdido la jugada.

Póngame medio kilo de primeras posiciones

La idea general que he podido oír de personas que no están en nuestro campo es que para estar en los primeros puestos hay que pagar a los buscadores. De poco o nada me ha servido explicarles qué es el *SEO*, ellos insisten en que las empresas grandes aparecen las primeras porque tienen dinero para comprar los puestos, y que eso es así como toda la vida lo fue. Mientras tanto, piden diseños en los que un enorme *flash* ocupa toda la pantalla, por supuesto sin texto legible en el código fuente, y mantienen un *title* del tipo "Bienvenido al sitio web de Fulanito" si es que con suerte el editor que usaron para hacer la página les pedía un título al guardar el fichero.

En el otro extremo están quienes se lo han aprendido todo sobre el *ranking* de resultados y utilizan esta información para forzar una posición óptima, usando técnicas que los buscadores persiguen y sancionan. De esto va a tratar el presente texto.

La ética del sombrero

En el mundillo del posicionamiento web o *SEO* (*search engine optimization*) se denominan técnicas *white-hat* a aquellas que se preocupan de que las páginas web puedan ser indizadas correctamente por los robots de los buscadores y que la información que contengan deje clara su temática. En el otro extremo se encuentran las llamadas técnicas *black-hat*, que son todas aquellas que pretenden engañar a los algoritmos de *ranking* para ocupar mejores puestos en los resulta-

Las técnicas *black-hat* son las que pretenden engañar a los algoritmos de *ranking* para ocupar mejores puestos en los resultados

dos. Entre unas y otras supongo que podríamos hablar de algunos *grey-hat*, de intensidad variable.

Estas oscuras técnicas comienzan en las más puras prácticas del diseño web. Me vienen algunos ejemplos a la cabeza (hablando de sombreros nunca mejor dicho), pero todos se pueden detectar de una forma bastante fácil:

- El famoso texto invisible: fuente de color blanco sobre un fondo blanco (os prometo que a día de hoy todavía existen casos). ¿Acaso creen que el robot se dedica a mirar la página como una persona? Nada de eso, rastrea el código fuente, y detectar que ambos colores coinciden es muy sencillo. Parecida a esta maldad está la de usar una fuente muy pequeña en ese texto, pero igualmente es fácil de detectar.

- La repetición de palabras clave en el *title* de la página. Es el caso contrario de los que todavía andan con el *untitled document*, más de 26 millones de páginas según *Google!* (*allintitle: "untitled document"*).

- La recepción de enlaces de forma masiva, por ejemplo por medio de programas que ponen comentarios en *blogs* e incluyen el enlace. Las páginas que reciben muchos enlaces son en principio páginas mejores; en cambio a los buscadores no les parece –y con toda la razón del mundo– que todos los enlaces tengan el mismo valor y han tomado



medidas: los que procedan de páginas de temática similar serán bien vistos, los que vengan de sitios no relacionados no se tendrán en cuenta. De esta manera se palia el efecto de estos enlaces que llegan de cientos en cientos, y además es fácil de detectar porque siempre llegan en bloque.

– Diseñar un sitio web atractivo para los usuarios y otro pensando en *SEO*, esto se llama *cloaking*. La página para los robots tendrá los elementos necesarios para un buen posicionamiento, mientras que la que ven los usuarios no necesariamente.

Lo cierto es que los buscadores están al tanto de las prácticas de abuso de estos sombreros negros y las van atajando lo mejor que pueden. Últimamente hemos oído hablar del *ranking* de confianza (*trust ranking*), que consiste en un grupo de factores que los buscadores tienen en cuenta para considerar si un sitio web es de fiar: los años que hace que se registró el dominio (para burlar esto los *spammers* tendrían que comprar dominios antiguos, si es que se los quieren vender...), el número de enlaces que lo enlazan y desde cuándo, y el *ranking* de confianza que a su vez tienen esas páginas que le han enlazado.

No podemos saber a ciencia cierta lo penalizadas que están esas prácticas, pues son demasiados factores los que intervienen en el *ranking* de los buscadores y no todos con el mismo peso.

Estar al día en *SEO*, blanco o negro

Una buena forma de estar al día en las novedades de los grandes buscadores es tan sencilla como leer los *blogs* que sus equipos escriben:

- *Official Google Blog*:
<http://googleblog.blogspot.com/>
- *Yahoo! Search Blog*:
<http://www.ysearchblog.com/>
- *Live Search's Weblog*:
<http://blogs.msdn.com/lmsnsearch/default.aspx>

Además de leer los *posts* que publican, es interesante ver los enlaces a otros *blogs*, tanto internos como externos. Una opción interesante es revisar *posts* antiguos; en los tres están disponibles a partir de 2004.

Lo cierto es que los buscadores están al tanto de las prácticas de abuso de estos sombreros negros y las van atajando lo mejor que pueden

Otros sitios web de referencia para no perderse en este tema tan cambiante son:

- **Aaron Wall's SEO Book**:
<http://www.seobook.com>
- Foro Buscadores:
<http://www.forobuscadores.com/foros/>
- Ojo Buscador:
<http://www.ojobuscador.com>
- *Online Marketing Blog*:
<http://www.toprankblog.com>
- *Search Engine Land*:
<http://searchengineland.com/>
- *Search engine ranking facts newsletter (Axandra)*:
<http://www.free-seo-news.com/index.php>
- *Search Engine Watch*:
<http://searchenginewatch.com/>
- *Search Engine Weblog*:
<http://www.searchengine-weblog.com/>

Y para profundizar en cuestiones más técnicas...

ACM Sigir. En: *Adversarial information retrieval on the web (Airweb 2006)*, 2006, v. 40, n. 2.

Castillo, Carlos; Donato, Debora; Gionis, Aristides; Murdock, Vanessa; Silvestri, Fabrizio: "Web spam detection using the web topology". *Spire 2007*, <http://www.citeulike.org/user/ChaTo/article/955052>

Castillo, Carlos. Todo por el rating.
<http://www.tejedoresdelweb.com/307/article-77584.html>

Castillo, Carlos. "WebSpam". En: *Conferencia en la Universidad de Salamanca*, 2007, Mayo.
http://www.tejedoresdelweb.com/slides/2007_talk_spam_salamanca.pdf

Codina, Lluís. "Posicionamiento web: conceptos y ciclo de vida". En: *Anuario Hipertext.net*, 2004, Mayo.
<http://www.hipertext.net>

Codina, Lluís; Marcos, Mari-Carmen. "Posicionamiento web: conceptos y herramientas". En: *El profesional de la información*, 2005, v. 14, n. 2, pp. 84-99.

Gonzalo, Carlos. "La selección de palabras clave para el posicionamiento en buscadores: conceptos y herramientas de estudio" En: *Anuario Hipertext net*, 2004, mayo.
<http://www.hipertext.net>

Gyöngyi, Zoltán; García-Molina, Héctor. "Spam: it's not just for inboxes anymore". En: *IEEE Computer magazine*, 2005, octubre, v. 38, n. 10, pp. 28-34.
<http://doi.ieeecomputersociety.org/10.1109/IMC.2005.352>

Gori, Marco; Witten, Ian. "The bubble of web visibility". En: *Comm. ACM*, 2005, marzo, v. 48, n. 3, pp. 115-117.
http://portal.acm.org/ft_gateway.cfm?id=1047715&type=pdf

Ntoulas, Alexandros; Najork, Mark; Manasse, Mark; Fetterly, Dennis. "Detecting spam web pages through content analysis". En: *World Wide Web Conference*, 2006, pp. 83-92.
<http://www2006.org/programm/item.php?id=3052>

Posicionamiento y visitas

Por José Luis Ortega



Es cierto que en torno al 60-80% de las visitas que se realizan a un web provienen de los buscadores, y por lo tanto tener una buena presencia en un buscador es muy importante para atraer tráfico y adquirir visibilidad. Esto es debido a que los buscadores son

enormes webs que generan una enorme cantidad de enlaces salientes y entrantes (*hubs*), lo que ocasiona que estos webs dirijan el 80% del tráfico en toda la Web.

Sin embargo hay un factor bastante obvio que pocas veces se discute cuando se habla de posicionamiento, y es el hecho de que internet es una red de páginas conectadas por enlaces. Esto nos obliga a considerar que las visitas que llegan a nuestro web pueden venir de múltiples lugares y no sólo desde los buscadores. De esta forma, no se puede olvidar la importancia de estas fuentes de visitas y, creo, sería aconsejable incorporar al mundo del posicionamiento medidas para "posicionar" nuestro web en el entorno más cercano además de mirar esos enlaces que ya se poseen. Por ejemplo, grandes y periódicas reestructuraciones de sedes web ocasionan la pérdida de los enlaces entrantes (*linkrots*), perdiendo todo el tráfico que provenía por esos enlaces. Así y desde mi opinión, obsesionarnos con subir un puesto en según qué consulta de un buscador, mientras que ignoramos los enlaces de otros webs, puede ser un actitud contraproducente para el nuestro.

Por otro lado los resultados de las medidas *SEO* se han basado en su posición en determinadas consultas en según qué buscador y al análisis cuantitativo de las visitas. Creo que estos resultados no justifican las medidas *SEO*. Porque un web bien posicionado no implica que sea relevante al usuario y menos si se ha conseguido con técnicas *black-hat*, como comentaba **Mari Carmen**. Así que una buena posición puede generar una gran cantidad de *hits* o visitas que computa nuestro contador web, pero no sabemos cuánto tiempo ha estado viendo nuestro web, cuántas páginas internas ha visitado, o cuántas veces nos ha visitado. En resumen, las visitas no nos permiten conocer la relevancia de nuestro web.

Por último, y al hilo de la cuantificación de visitas, un hecho relevante es medir la calidad de nuestros visitantes. Quizás sean más relevantes unos pocos visitantes que desean comprar nuestros productos que miles que sólo van a mirar. Es posible que según de donde provengan, de enlaces o desde los buscadores, nuestros visitantes sean más fieles y estén más interesados en nuestros contenidos. Simplemente, un enlace desde otro web nos aporta cuanto menos una descripción escueta de nuestro recurso y la autoridad de la fuente, mientras que un buscador tan sólo reproduce el título y un somero párrafo de nuestro web. Creo que los enlaces son una vía muy importante y a tener en cuenta para conseguir visitas de calidad que se pueden convertir en potenciales clientes o usuarios.

José Luis Ortega
jose_ortega@gmx.net

Granjas de enlaces

Añadir una técnica más que he visto a personas/instituciones hacer para lograr posicionamiento web.

Lo llaman "granjas" y consiste en tener una serie de webs creados (la parte estática siempre en xhtml) que se enlazan entre sí. Lo normal es tener un web de información de la institución, y varios *blogs* sobre los diferentes temas que tratan en ésta, siempre con buenos contenidos, enlazándose entre sí, pero en diferentes servidores.

Óskar Calvo Vidal
oskartsu@yahoo.es

Referencias

Siguiendo el hilo de la contribución de **Mari-Carmen Marcos** y el comentario de **Oskar Calvo**, os referencio un artículo muy interesante de **Carlos Gonzalo** sobre las granjas de enlaces que publicó la revista *BiD* hace un año. Contextualiza esta práctica fraudulenta en el ámbito del posicionamiento y el uso y motivación de los enlaces web.

Gonzalo, Carlos. "Tipología y análisis de enlaces web: aplicación al estudio de los enlaces fraudulentos y de las granjas de enlaces". En: *BiD: textos universitarios de biblioteconomía i documentació*, 2006, juny, n. 16. Consultado en: 07-07-2007.

http://www2.ub.edu/bid/consulta_articulos.php?fichero=16gonza2.htm

Ernest Abadal
abadal@ub.edu

doc6 en 30 segundos

La empresa

Creada en el año 1988, especializada en gestión de la información y la documentación.

Certificada con la ISO 9001:2000 por Bureau Veritas desde Septiembre del 2000 en las siguientes actividades:

- CONSULTORES EN RECURSOS DE INFORMACIÓN
- GESTIÓN DEL CONOCIMIENTO
- INTRANET/EXTRANET/INTERNET
- EDICIÓN ELECTRÓNICA
- EXTERNALIZACIÓN DE SERVICIOS DOCUMENTALES
- PROVEEDORES DE RECURSOS ELECTRÓNICOS DE INFORMACIÓN
- SOLUCIONES EN GESTIÓN DOCUMENTAL
- CURSOS DE FORMACIÓN

Nuestra misión

Contribuir a aumentar la eficiencia de empresas y organizaciones a través de una eficaz gestión de la información y el conocimiento.



Nuestros valores

Innovación permanente.
Satisfacción del cliente. Calidad
Profesionalidad / Cooperación con el cliente
Confidencialidad
Responsabilidad social y valores éticos

Los clientes

El principal activo de la empresa, junto con sus empleados.
Más de 900 clientes: empresas, laboratorios, administraciones públicas, universidades, hospitales, bufetes de abogados, museos, fundaciones, colegios profesionales.

Nuestros profesionales

Una eficaz suma de conocimientos: informáticos y documentalistas con experiencia
El 75% son licenciados superiores o diplomados en informática o documentación.



DOC6 S.A. ha recibido la etiqueta EUREKA de I+D para el proyecto COGNOS (n. 2734) de metodología para el desarrollo de mapas del conocimiento en las organizaciones.



DOC6 ha obtenido la Certificación ISO 9001 otorgada por Bureau Veritas Quality International (BVQI) con las acreditaciones ENAC (España) y UKAS (Reino Unido) para sus Centros de Barcelona y Madrid.

doc6
consultores en recursos de información

Mallorca, 272, planta 3 - 08037 Barcelona • Tel. 932 154 313 Fax 934 883 621
Orense, 14, 5º A - 28020 Madrid • Tel 915 535 207 Fax 915 346 112

mail@doc6.es www.doc6.es