

Título: Evaluación del posicionamiento web en sistemas de información terminológicos online

Autor: Mari-Carmen Marcos et al.

Citación recomendada: Mari-Carmen Marcos et al.. *Evaluación del posicionamiento web en sistemas de información terminológicos online* [en línea]. undefined [Consulta: 21 ene. 2007]. undefined.

1. Introducción
2. Metodología de la investigación
3. Elección de palabras clave para el posicionamiento web
4. Análisis de los factores del posicionamiento web
 - 4.1. Frecuencia de aparición y posición de las palabras clave
 - 4.2. Metadatos
 - 4.3. Popularidad, textos de anclaje y tráfico de visitas
5. Conclusiones
6. Nota final
7. Referencias

1. Introducción

Hoy en día la mayoría de los accesos a sitios web se realiza a través de motores de búsqueda, por lo tanto, es fundamental para sus responsables asegurarse de aparecer bien posicionados en los resultados de búsqueda, tanto desde el punto de vista del márketing como para dar mejor servicio a sus usuarios.

En consecuencia, el lugar que un sitio web ocupa en el listado de resultados de un motor de búsqueda, cuando un usuario realiza una consulta en el buscador, es un aspecto de gran relevancia para los responsables de sitios web, y las acciones para su mejora vienen explicadas por el “posicionamiento web” (Codina, 2004; Codina, Marcos 2005; Arbildi, 2005). El eje central está en preguntarse cuáles serán las palabras que previsiblemente utilizarán los potenciales usuarios en sus búsquedas; una vez determinado esto, podrá hacerse uso "lícito" de las técnicas de optimización para que un determinado sitio web aparezca en una buena posición cuando los usuarios busquen información relacionada con los contenidos de dicho sitio web.

Los sitios web que albergan bases de datos terminológicas también pueden –y deben – hacer uso de las técnicas de mejora del posicionamiento para facilitar a sus posibles usuarios encontrar estos recursos a través de los buscadores. Nuestra investigación ha tomado como muestra diez bases de datos terminológicas cuya consulta es libre en la web, presentan multilingüismo y pertenecen a diferentes temáticas (tabla 1).

<i>Bases de datos</i>	<i>URL</i>
Base de Terminologie	http://www.cilf.org/bt.fr.html
CercaTerm	http://www.termcat.net
Eurodicautom	http://europa.eu.int/eurodicautom
EuskalTerm	http://www1.euskadi.net/euskalterm
OncoTerm	http://www.ugr.es/~oncoterm/
TerminoBanque	http://www.cfwb.be/franca/bd/bd.htm
TIS: Terminological Information System	http://tis.consilium.eu.int
UBTerm	http://www.ub.edu/slc/ubterm
UNTerm	http://unterm.un.org
WTOTerm	http://wtoterm.wto.org

Tabla 1. Bases de datos terminológicas estudiadas

En este artículo presentamos los principales resultados obtenidos del análisis de su posicionamiento web y las propuestas de mejora, que pueden aplicarse a otras herramientas de búsqueda lingüística como diccionarios online, tesauros, etc. La metodología de análisis empleada ha sido la observación empírica de los criterios que los buscadores tienen en consideración (véase el siguiente apartado) y el análisis estadístico multivariante, para poner en relación los distintos aspectos analizados con el posicionamiento obtenido para cada palabra clave.

2. Metodología de la investigación

Para facilitar ser encontrado en buscadores, los sitios web deben mejorar todos los aspectos que estos tienen en cuenta a la hora de establecer el ranking de resultados. Sin entrar a explicar en detalle las técnicas para la optimización de sitios web, nombramos los criterios que parecen estar usando los buscadores en sus rankings de resultados (Codina; Marcos, 2005):

Frecuencia (absoluta y relativa) de la expresión o término buscado en la página web (al que llamaremos "palabra clave"), siempre que no se caiga en una repetición abusiva que sea considerada *spam* por los buscadores. En este estudio se han planteado tres posibles palabras clave para cada sitio web.

Posición : lugar donde se encuentra el término dentro de la página; se tendrá en cuenta que los metadatos y el primer párrafo tienen más peso que otras partes de la página.

Metadatos : además de los metadatos de la sección *head* (principalmente *title* , *description* y *keywords*), otras etiquetas también proporcionan información descriptiva de utilidad para los buscadores, por ejemplo el *title* de los enlaces y de las imágenes, así como el texto alternativo (*alt*) de las imágenes.

Popularidad : número de enlaces externos que recibe la página web. Este criterio está relacionado con el *PageRank* determinado por la barra de herramientas de Google.

Anclaje : texto que sirve como enlace para llegar a esta página web.

Tráfico de visitas que recibe la página web, considerando a la vez el tiempo de permanencia de los usuarios en ella. Viene dado por el *TrafficRank* de la barra de herramientas de Alexa.

De estos factores, nuestra investigación está centrada fundamentalmente en los metadatos y la popularidad. Los demás factores también se han estudiado, si bien no han arrojado resultados que nos ayuden a estimar el grado de importancia que tienen para el posicionamiento web de los sitios analizados.

El estudio ha tenido en cuenta dos tipos de análisis: en primer lugar se ha llevado a cabo un análisis empírico-descriptivo atendiendo a los aspectos que la bibliografía sobre posicionamiento web indica que deben tenerse en cuenta. En base a los resultados obtenidos en esta fase, se han planteado algunas hipótesis que se han puesto a prueba mediante el análisis estadístico ANOVA; se trata de un análisis multivariable en el que una variable dependiente es cruzada con algunas variables independientes.

Para su aplicación se ha utilizado el programa Statgraphics Plus 5.0. La variable dependiente considerada ha sido el posicionamiento web de los sitios web, y las variables independientes con las que se ha cruzado han sido cada uno de los sitios web estudiados, los buscadores sobre los que se han hecho las consultas, las palabras clave de las consultas, y los metadatos *title* , *description* y *keywords* . El cruce de datos da como resultado el p-valor, que es el nivel de significación empírico en un contraste de hipótesis. Se considera que una hipótesis es nula en los casos en que el p-valor es inferior a 0'05 y que es alternativa siempre que supera esta cifra.

3. Elección de palabras clave para el posicionamiento web

Antes de analizar las causas del posicionamiento en estos sitios web, y puesto que este dato constituye la variable dependiente en el análisis estadístico, se ha comprobado cuál es éste para 3 consultas para las que estos sitios web consideramos que deberían aparecer bien posicionados, entendiendo por una buena posición los 10 primeros resultados de los buscadores más utilizados hoy en día. Se ha valorado de forma especialmente positiva (2 puntos) que el sitio web apareciera en el primer o el segundo puesto, de forma parcial (1 punto) si aparece entre el puesto 3 y el 10, y no se ha valorado (0 puntos) si está más allá del resultado número 10.

Las palabras de búsqueda elegidas se han traducido al idioma principal de la interfaz de cada herramienta terminológica

siguiendo estos criterios:

Palabra clave 1 (PC1): nombre de la base de datos.

Palabra clave 2 (PC2): sintagma que describe el tipo de recurso terminológico del que se trata en cada caso.

Palabra clave 3 (PC3): sintagma que describe una cualidad específica del recurso terminológico del que se trata en cada caso.

<i>PC1</i>	<i>PC2</i>	<i>PC3</i>
Base de Terminologie	CILF terminologie	Terminologie spécialisée
CercaTerm	Base de dades terminològica	Terminologia catalana
WTOTerm	Terminology database	World trade terminology
TerminoBanque	Banque de donées	Terminologie spécialisée
Eurodicautom	Terminology database	Multilingual specialized terminology
EuskalTerm	Banco terminológico	Terminologia euskera
TIS	Terminological information system	Consilium terminological database
UBTerm	Base de dades terminològica	Terminologia catalana
UNTerm	Terminology database	United Nations terminology
OncoTerm	Base terminológica de oncología	Terminología oncológica

Tabla 2. Palabras clave utilizadas para buscar en los 6 motores

En el mes de enero de 2006 se realizaron las tres búsquedas en 6 de los buscadores más utilizados hoy en día (Google, Yahoo! Search, MSN Search, Altavista, Teoma y Vivísimo). El resultado obtenido es muy diferente para cada una de las palabras clave (tabla 3):

Las búsquedas por la PC1 obtienen con más frecuencia el sitio web esperado entre los 10 primeros resultados, en especial con Yahoo! Search y Google, aunque no tanto con Teoma y Vivísimo, que sólo consiguen localizar estos sitios web en un 50% de las búsquedas.

El resultado cambia para la PC2, pues dos sitios web no se han encontrado entre los 10 primeros resultados, y otros no ocupan las primeras posiciones. Google y Yahoo! Search muestran estos sitios web en mejores posiciones que los otros buscadores. Teoma en cambio sólo muestra en este primer grupo de resultados 4 de los 10 sitios web.

La búsqueda por la PC3 muestra resultados diferentes a los anteriores: aunque sigue en la línea de la PC2 y algunos buscadores no colocan estos sitios web entre los 10 primeros resultados, el buscador Teoma esta vez ofrece las mejores posiciones, seguido de Vivísimo y Google.

Bases de datos	PC1	PC2	PC3	Promedio de cada base de datos
Base de Terminologie	50,0	100,0	100,0	83,3
Eurodicautom	100,0	50,0	100,0	83,3
TIS	33,0	91,6	100,0	74,8
WTOTerm	100,0	25,0	91,6	72,2
EuskalTerm	100,0	83,3	0,0	61,1
CercaTerm	91,6	0,0	50,0	47,2
OncoTerm	41,6	100,0	0,0	47,2
UNTerm	91,6	25,0	16,6	44,4
UBTerm	75,0	33,0	0,0	36,0
TerminoBanque	100,0	0,0	0,0	33,3

Promedio de cada PC	72,3	50,8	45,8	83,3
---------------------	------	------	------	------

Tabla 3. Clasificación de las bases de datos en función del promedio del posicionamiento que ocupa cada sitio web al buscar por la PC1, PC2 y PC. El valor máximo, 100, indica que se encuentra en el primer o el segundo puesto; los valores intermedios, cercanos a 50, indican que se encuentra entre el puesto 3 y el 10, y el valor 0 indica que no está entre los 10 primeros resultados

El análisis estadístico corrobora que existe una diferencia significativa entre el posicionamiento de los sitios web estudiados y las palabras clave para las que se ha probado su posicionamiento, pues para la PC1 el posicionamiento obtenido es mucho mejor que para las PC2 y PC3 (figura 1). Al mismo tiempo, muestra que el uso de un buscador u otro no provoca diferencias significativas en los resultados de posicionamiento, si bien Google, Yahoo! Search y Vivísimo presentan estos sitios web mejor posicionados que los otros 3 buscadores utilizados.

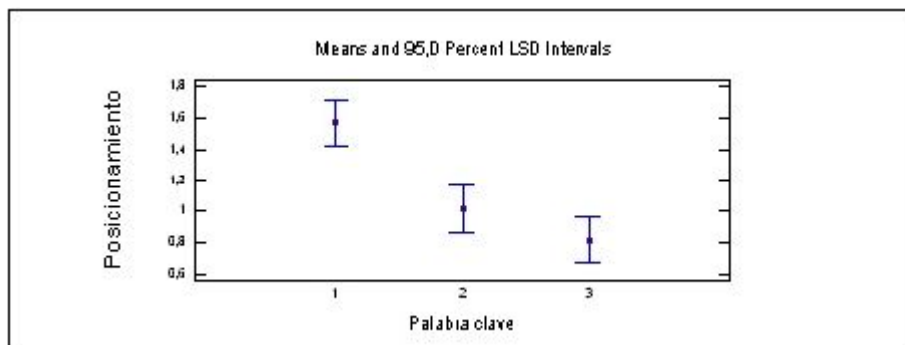


Figura 1. Posicionamiento de los sitios web en relación con PC1, PC2 y PC3

4. Análisis de los factores del posicionamiento web

A partir de los aspectos indicados en el apartado 2, presentamos los resultados más relevantes obtenidos en el análisis empírico-descriptivo y en el análisis estadístico.

4.1. Frecuencia de aparición y posición de las palabras clave

A pesar de que estamos de acuerdo en la importancia de estos factores para mejorar el posicionamiento web, se ha decidido no tenerlo en consideración para este estudio debido a dos motivos: en primer lugar, las interfaces de las herramientas de búsqueda de información terminológica no cuentan con un volumen de texto suficiente para poder determinar valores de frecuencia óptimos; y en segundo lugar, a excepción del nombre de la base de datos, las otras palabras clave establecidas como consultas (PC2 y PC3) no suelen aparecer reflejadas en los sistemas estudiados, por lo que este criterio no nos permitiría establecer comparaciones.

4.2. Metadatos

En lo referente a los metadatos de los sitios web estudiados, hemos comprobado que el 90% poseen el campo title relleno, el 40% incluyen el campo keywords y sólo el 20% presenta el campo description (tabla 3). No se han considerado las etiquetas title y el alt de los enlaces y las imágenes, pues en los sitios estudiados su número era muy bajo o incluso nulo, por lo tanto no afectaría al posicionamiento web de estos sitios.

<i>Bases de datos</i>	<i>Title</i>	<i>Keywords</i>	<i>Description</i>
Base de Terminologie	X	-	-
CercaTerm	X	X	-
Eurodicautom	X	X	-
EuskalTerm	X	X	X

OncoTerm	X	X	-
TerminoBanque	X	-	-
TIS	X	-	X
UBTerm	X	-	-
UNTerm	-	-	-
WTOTerm	X	-	-

Tabla 4. Uso de metadatos en los sitios web de las bases de datos estudiadas

En el análisis estadístico, el cruce del posicionamiento de los sitios web con la información de metadatos pone de manifiesto que sólo el campo description implica diferencias significativas de posicionamiento entre sitios web (figura 2), mientras que la existencia de keywords no influye tanto en los resultados obtenidos (figura 3). No se ha podido obtener un resultado fiable sobre la repercusión de la etiqueta title, pues 9 de los 10 sitios web contaban con ella, lo que no nos deja margen para establecer comparaciones.

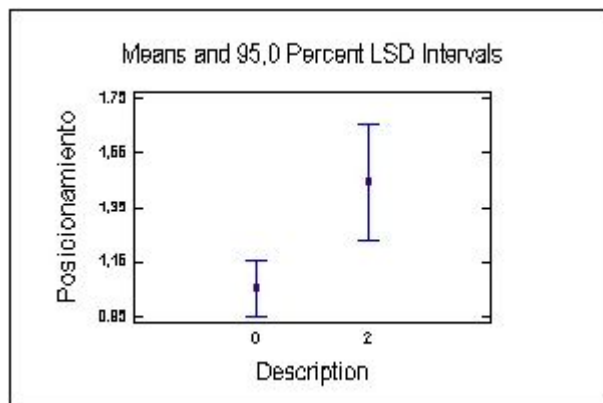


Figura 2. Posicionamiento de los sitios web considerando si tienen la etiqueta meta description (2) y si no la tienen (0)

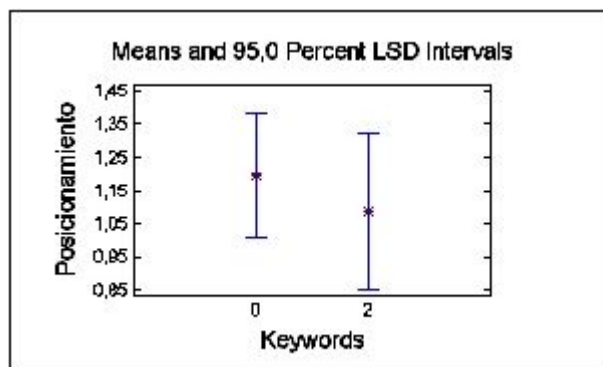


Figura 3. Posicionamiento de los sitios web considerando si tienen la etiqueta meta keywords (2) y si no la tienen (2)

4.3. Popularidad, textos de anclaje y tráfico de visitas

La popularidad, entendida como el número de enlaces que apuntan hacia un sitio web, puede conocerse parcialmente a través de la búsqueda con el limitador link: que ofrecen algunos buscadores. Mostramos los valores que da Yahoo! Search (tabla 5), que son una cifra mucho más alta que la que proporciona Google.

<i>Bases de datos</i>	<i>Enlaces de llegada (Yahoo! Search)</i>
Eurodicautom	2.340

CercaTerm	1.290
Base de Terminologie	1.030
TIS	458
EuskalTerm	243
TerminoBanque	163
WTOTerm	104
UNTerm	79
UBTerm	49
OncoTerm	9

Tabla 5. Clasificación de las bases de datos en función del número de enlaces que apuntan a cada sitio web según Yahoo! Search

Si comparamos los resultados que obtenemos en las tablas 3 y 5, podemos ver cómo existe una relación bastante directa entre el número de enlaces de llegada de cada sitio web y los resultados medios que han dado para el posicionamiento por las tres palabras clave. Las bases de datos con más citas son al mismo tiempo las mejor posicionadas para estas palabras. En el extremo contrario, UBTerm y UNTerm, que obtienen menos citas también aparecen peor posicionadas. El caso de CercaTerm, que no está tan bien posicionada y en cambio recibe muchas citas, se debe a que la cifra del número de enlaces de llegada que tiene no se refiere a la página principal de la base de datos, sino a la de su institución (TermCat), pues para llegar a la base de datos es necesario pasar por un formulario en la página de inicio de la institución, lo que dificulta el acceso directo a los buscadores. En el caso de este sistema, si no cambia el modo de acceso a la base de datos, deberá optimizarse la página de inicio de la institución para mejorar el posicionamiento en las búsquedas relacionadas con su herramienta terminológica.

En cuanto a los textos usados en los anclajes de esos enlaces, se han revisado los 5 primeros de la lista que ofrece Google y el resultado obtenido ha sido que, a excepción de la PC1 (el nombre de la base de datos), no se han utilizado las palabras clave escogidas en este estudio. Por lo tanto es un criterio que no nos va a servir para establecer comparaciones y detectar la relevancia que tiene en el posicionamiento de sitios web.

También se anotó el valor que otorga Google como PageRank, si bien finalmente no lo hemos considerado de interés para este estudio pues se trata de un valor obtenido con un cálculo poco transparente. El mismo criterio nos ha llevado a no tener en cuenta el valor que da Alexa como TrafficRank, que además no es específico para la página web que se analiza sino que toma el valor de su página de inicio.

5. Conclusiones

Aunque las técnicas de optimización para buscadores suelen enmarcarse dentro del márketing; nuestra visión de este concepto es amplia, pues no sólo creemos que se trata de una cuestión de publicidad, sino de servicio a los usuarios. En el caso de servicios de uso gratuito, como es el de las bases de datos terminológicas que estudiamos, también deberán preocuparse de poner los medios necesarios (y lícitos) para posicionarse bien, pues es el primer paso para poder ser usado.

En cuanto a la metodología empleada, el análisis empírico-descriptivo ha resultado de utilidad para comprobar que ninguno de los sitios web estudiados ha realizado una campaña de posicionamiento. De hecho, en la mayoría de los casos no se han rellenado los campos de metadatos. De los tres campos básicos que hemos mencionado, title es el más frecuente (está en el 90% de los sitios web analizados), seguido de keywords (40%) y, de forma muy lejana, por el de description (20%). No se ha seguido una política de enlaces para conseguir mayor visibilidad, tanto para los posibles usuarios como para los buscadores. Por otro lado, el análisis estadístico multivariable (ANOVA) ha resultado válido para comprobar las observaciones realizadas y verificar hipótesis. La aplicación conjunta de ambas metodologías nos lleva a afirmar que el metadato description juega un papel importante en el posicionamiento de sitios web, mientras que el metadato keywords no presenta esta evidencia (figura 4). De todas formas, podemos afirmar que el uso de metadatos no es decisivo para el posicionamiento web, pues se da el caso de sitios bien posicionados que no han hecho uso de metadatos, por lo que no hay que dejar de lado otros factores que influyen en el posicionamiento, como la popularidad.

De hecho, se ha podido observar que existe una correlación entre la popularidad de los sitios web analizados (el número de enlaces que apuntan hacia ellos) y su posicionamiento para las palabras clave escogidas, lo que indica que el criterio de popularidad juega un papel importante en la posición que los sitios web ocupan en los resultados de las búsquedas en

motores. Esta popularidad debería venir potenciada por unos textos de anclaje apropiados, de manera que se afiancen las palabras clave que las instituciones escojan para su posicionamiento en la web. En el caso de los sitios estudiados no se ha dado esta política.

En lo que se refiere a los buscadores usados en el estudio, Google es el que ha presentado un comportamiento más regular en relación con el uso de los metadatos en los sitios web, y Google, Yahoo! Search y Vivísimo son los 3 motores que han mostrado mejores posicionamientos para las palabras clave escogidas en estos 10 sitios web. En la mayoría de los buscadores los sitios web aparecen posicionados cuando se realizan consultas por palabras más específicas (PC1), a excepción de Teoma y Vivísimo, que contrariamente a lo dicho, los sitios web obtienen mejores posicionamientos en búsquedas más generales (PC2 y PC3). Por ejemplo, en el caso de Eurodicautom, la búsqueda por la PC3 no presenta este sitio web entre los primeros resultados de Yahoo! y, en cambio, para Teoma ocupa el segundo puesto.

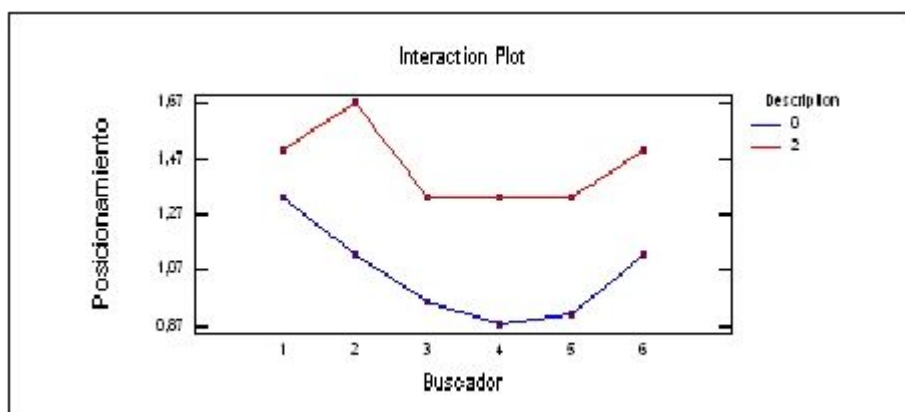


Figura 4. Posicionamiento de los sitios web en 6 buscadores considerando si tienen la etiqueta meta description (línea roja) o no la tienen (línea azul): Google (1), Yahoo! Search (2), MSN Search (3), Altavista (4), Teoma (5) y Vivísimo (6)

6. Nota final

Los autores queremos mostrar nuestro agradecimiento a las instituciones que han participado de alguna manera en nuestra investigación:

Instituto Universitario de Lingüística Aplicada de la Universitat Pompeu Fabra, que ha dado soporte a la realización de este estudio y ha concedido una beca a Albert Morales y Juan Manuel Pérez (profesor a tiempo completo en la Universidad de Antioquia-Colombia).

Ministerio de Educación y Ciencia (2004-2007), proyecto "Web semántica y sistemas de información documental", nº HUM2004-03162/FILO del que Mari-Carmen Marcos es investigadora.

Fundación Caixa Galicia – Claudio San Martín, que ha colaborado con la beca otorgada a Paulo Malvar.

Programa Alban (Becas de Alto Nivel de la Unión Europea para América Latina), que ha colaborado con la concesión de la beca nº E05M059270MX a Fernanda López.

Agencia Española de Cooperación Internacional del Ministerio de Asuntos Exteriores, que ha otorgado una beca a Hajar Benmakhlouf.

Fundación Ford, que ha otorgado una beca a Pedro Hernández.

7. Referencias

Arbildi, Iñigo (2005). "Posicionamiento en buscadores: una metodología práctica de optimización de sitios web". El profesional de la información, v. 14, n. 2, pp. 108-124.

Azlor, S. (2003). Posicionamiento en buscadores: guía básica, <<http://www.guia-buscadores.com/posicionamiento>> [Consulta: 11/04/06].

Bruce Clay Inc. Internet Business Consultants. Search engine optimization, <http://www.bruceclay.com/web_rank.htm> [Consulta: 11/04/06]

Calishain, T.; Dornfest, R. (2004). Google: los mejores trucos. Madrid: Anaya.

Codina, Lluís (2004). "Posicionamiento web: conceptos y ciclo de vida". Anuario Hipertext.net, <<http://www.hipertext.net>> [Consulta: 11/04/06].

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

Gonzalo, Carlos (2004). "La selección de palabras clave para el posicionamiento en buscadores: conceptos y herramientas de estudio" Anuario Hipertext.net <<http://www.hipertext.net>> [Consulta: 11/04/06].

Kent, Peter (2004). Search engine optimization for dummies. Hoboken, Wiley.

Sullivan, Danny, Search Engine Watch, autoridad en la materia, <<http://www.searchenginewatch.com>> [Consulta: 11/04/06].

Thurow, Shari (2003). Search Engine Visibility. Indianapolis: New Riders.