

Solapamiento en herramientas de búsqueda de información en la World Wide Web

Francisco Javier Vidal Bordés

Facultad de Filosofía y Letras. Biblioteca

José Antonio Salvador Oliván

Departamento de Ciencias de la Documentación e Historia de la Ciencia
Universidad de Zaragoza

0.1. Resumen

La búsqueda exhaustiva de información en la World Wide Web sobre un tema requiere la utilización de diversos motores de búsqueda. Conocer el solapamiento entre los buscadores de carácter general más utilizados en la web, puede ser útil cuando, en una búsqueda de información, queramos obtener el mayor número de referencias diferentes sobre un tema. La experiencia llevada a cabo demuestra que herramientas como Lycos y Yahoo, por su menor grado de solapamiento, son más recomendables que HotBot o Google para ser utilizadas de forma complementaria con otros buscadores, en búsquedas de este tipo.

Palabras clave: Motores de búsqueda. Solapamiento. World Wide Web.

0.2. Abstract

Searching exhaustive information in the World Wide Web about a specific topic requires the use of some search engines. Information retrieval evaluation techniques are needed to know their performance. In the study, overlapping between general search engines is analyzed to know which are the most complementary. The results showed that Lycos and Yahoo, both with little overlap, are more recommended than HotBot and Google to be utilised with other search engines to find more exhaustive information about a specific topic.

Keywords: Search Engines. Overlapping. World Wide Web.

1. Introducción

Existe una gran cantidad de información disponible en Internet (1), fundamentalmente en la World Wide Web y hay muchos usuarios que necesitan localizar

información, aunque, en este medio, la búsqueda de información relevante se ha convertido en una tarea difícil y complicada en comparación con la consulta en bases de datos en línea y CD-ROMs (Dong y Su, 1997). Ello es debido al procedimiento de formación de un tipo y otro de base de datos, ya que si bien las bases de datos de consulta en línea y CD-ROM contienen información seleccionada y tratada documentalmente, en las bases de datos de los buscadores de la web, la información es de muy desigual valor científico y mínimo tratamiento documental. Estos son algunos de los factores que determinan la efectividad de los resultados de las búsquedas de información en la web. Pero, como veremos más adelante, existen otros que también tienen su importancia.

La búsqueda exhaustiva de información en la web requiere la utilización tanto de buscadores generales y especializados como de metabuscadores. No obstante, hay que advertir, como señala Brodley (1999), que las bases de datos de los motores de búsqueda apenas contienen indizados el 16% de los recursos disponibles en Internet, y que dado el crecimiento exponencial de la información en la red, este porcentaje puede haberse reducido en la actualidad. Este y otros aspectos pueden ayudarnos a entender la frecuente aparición de nuevos buscadores, que en la actualidad se caracterizan por la especialización en determinadas materias, proliferación de buscadores en servicios de pago, con recursos sometidos a un mayor tratamiento documental, y de directorios especializados, que facilitan el acceso a una serie de recursos específicos de interés.

Dado que estas herramientas de búsqueda, han estado sometidas desde sus orígenes a un constante desarrollo en cuanto a capacidades de recuperación y dado que requieren por parte del usuario un buen conocimiento de sus características, prestaciones, y opciones de búsqueda, se hace necesario un estudio de evaluación para ver hasta qué punto pueden considerarse los buscadores más utilizados en la web complementarios y en qué medida hay o no solapamiento.

2. Herramientas de búsqueda: concepto y clases

Los buscadores de información Web hacen su aparición hacia mediados de los años noventa, siendo WebCrawler el primero de ellos, seguido de Lycos y en el año 1995 aparecen Open Text Index, Meta Crawler y Altavista. Chowdhury (1999) los define como un servicio de recuperación de información que consiste en una base de datos que contiene principalmente recursos disponibles en la web. Hock (1999) se refiere a “herramientas de búsqueda” (*finding tools*) como expresión que abarca tanto a Motores de búsqueda (*search engines*) como a directorios (*directories*).

En el ámbito español, los trabajos de Codina, Cordón García et al., Marcos Mora, Maldonado Martínez, Fernández Sánchez y Olvera Lobo permiten un acercamiento a la terminología sobre estos servicios y sus partes. Cordón García

y colaboradores (1999) definen el motor de búsqueda como una “herramienta Web que localiza de forma rápida información existente en Internet y que está formado por tres elementos bien diferenciados: un interface (página Web a la que accede el usuario y en la que realiza la búsqueda), un robot (programa que recorre la Web analizando páginas Web) y una base de datos (índice de palabras, frases y datos asociados con la dirección URL de las páginas Web)...”.

Los robots son programas encargados de visitar servidores de la red Internet y, fundamentalmente en los servidores web, extraer de los recursos que contienen aquella información necesaria para la formación o actualización de sus índices. Así, Lycos busca diariamente en la World Wide Web, Gopher y FTP. Altavista en páginas Web y en las News. Los robots se ocupan de actualizar la información de la base de datos, de forma constante. Algunos motores de búsqueda permiten que se les envíe una dirección URL para posteriormente ellos visitarla e incorporarla a su base de datos. La interfaz permite consultar la base de datos y representar en pantalla los datos y resultados de la búsqueda. Finalmente, la base de datos cuenta con sus propios programas de gestión que almacenan toda la información indizada, lista las direcciones web, títulos, cabeceras, palabras, primeras líneas, abstracts, partes del texto o el texto completo.

Según su contenido, podemos distinguir entre motores de carácter general, que indizan recursos de toda la web, y motores especializados, que se ocupan de formar sus bases de datos con recursos que tratan sobre un tema determinado. Dong y Su (1997) dividen los motores de búsqueda en tres categorías: motores de búsqueda individuales, directorios de materias y metabuscadores.

Los motores individuales pueden a su vez diferenciarse según sus bases de datos sean a texto completo o no. Pero existen otras clasificaciones basadas en los métodos utilizados por los robots para reunir la información, ya que algunos de ellos —como ALIWEB o W3— están programados para visitar determinados sitios web (Dong y Su).

Los directorios están formados por listas de materias organizadas jerárquicamente que agrupan recursos seleccionados en relación con un tema. El proceso no se realiza de forma automática como en los motores, sino mediante la intervención de personas. Algunos contienen además un buscador de palabras clave que los complementa como ocurre en Yahoo y LookSmart. Para Dong y Su (1997), estas herramientas están mejor organizadas y pueden ofrecer información más relevante dado que contienen información seleccionada.

Cada vez son más los servicios que integran ambas funciones de acceso permitiendo localizar información tanto a través de determinadas categorías de materias como mediante la consulta a la base de datos a utilizando la clásica ventana de búsqueda.

Los metabuscadores son aquellos que realizan una misma búsqueda simultáneamente en diferentes motores como MetaCrawler, SuperSeek, etc. Lancaster (1998) señala entre sus ventajas la exhaustividad en las búsquedas y su utilidad para evaluar motores, y como inconvenientes su mayor lentitud y la recuperación de un número de registros limitado. Otro aspecto a favor es que permiten de forma rápida conocer si existe información de un tema específico ya que evita utilizar varios buscadores.

Finalmente debemos señalar la existencia de directorios especializados en un tema como es el caso de LawCrawler que tiene muy presentes aspectos como la selección y calidad de los recursos (Schwartz, 1998).

En relación con el tipo de información que contienen, hay que señalar que algunos motores no sólo indizan páginas web, sino que también se ocupan de news y menús Gopher así como de determinados tipos de documentos en formato ASCII, VRML, SGML y PDF.

3. Funcionamiento de los motores de búsqueda

Taylor y Clemson afirman que para hacer una búsqueda efectiva se debe entender el funcionamiento del motor de búsqueda, el tamaño de su base de datos, y su estrategia de indización de documentos. Se hace por tanto necesario un pequeño estudio de estos aspectos. Este conocimiento de las herramientas es recomendable para el usuario, porque puede ayudarle a elegir un motor de búsqueda u otro, dependiendo de sus características y de los recursos que se quieran buscar.

Milstead y Feldman (1999) señalan como característica de los motores de búsqueda estadísticos, la utilización de algoritmos que dan peso o importancia a las palabras en función de la rareza en el índice, la frecuencia del término en el documento, la posición de la palabra y la proximidad de los términos de búsqueda en el texto. El resultado se añade al cálculo de relevancia de una búsqueda y de este modo se establece el orden de aparición de los registros. Excite realiza búsquedas conceptuales basadas en co-ocurrencias de términos establecidos en la base de datos. Así, si un término aparece próximo a otro, se supone una relación entre ellos.

Lancaster (1998) afirma que los motores de búsqueda construyen índices con los recursos de la red: "What this means, in effect, is the extraction of words or phrases from the text itself and the building of files that permit the efficient searching of these extracts using Boolean (and sometimes other) approaches". El software que contienen les permite visitar un gran número de servidores y, en unos casos extraer información para registrar en sus bases de datos, y en otros casos, actualizarla si ha sufrido variaciones desde la anterior visita. Su análisis se centra tanto en la información del recurso como de los enlaces de hipertexto que pueda contener dicho recurso.

Pero no todos los buscadores contienen el mismo software, ni las mismas características técnicas, lo que influye en el modo de operar y en las capacidades de tratamiento, almacenamiento y recuperación de la información de cada uno de ellos. Así, ocurre que determinados motores resumen el contenido informativo de los recursos y extraen palabras clave, mientras que otros recogen la información de etiquetas Meta que contienen los términos de carácter descriptivo más interesantes para la localización del recurso, e incluso otros aplican sofisticados criterios lingüísticos en la indización de textos. En este sentido, Excite utiliza un procedimiento denominado Intelligent Concept Extraction para encontrar y numerar documentos basándose en una correlación de sus conceptos y sus palabras clave. Otros motores como Swapy utilizan sofisticados procesos lingüísticos para la indización automática (2).

Lancaster (1998) señala una serie de capacidades comunes de los buscadores:

- Uso de la lógica booleana.
- Búsqueda manteniendo el orden de las palabras.
- Posibilidad de especificación de subconjunto dentro de un determinado conjunto, por ejemplo, buscar en “perros” lo referente a la raza “boxers”.
- Posibilidad de restringir la búsqueda a una parte determinada del recurso: título o componente de la URL, etc.
- Truncamientos y uso de comodines
- Adyacencia
- “Interrogación por ejemplo” que permite localizar recursos similares a uno conocido.

3.1 Ordenación y presentación de los resultados

Como señala Lancaster, la capacidad de establecer un baremo o ranking en los resultados de las búsquedas, depende del modo de indización del motor, y de la cantidad y tipo de información almacenada en los archivos de búsqueda. El motor de búsqueda asigna al recurso un valor que es más alto cuanto mayor es el número de términos que contiene que coinciden con los que se buscan. Pero esta afirmación de carácter general ofrece variaciones ya que cada motor trabaja según esté programado, y se vale para el cálculo de la relevancia del análisis de diferentes parámetros como pueden ser la proximidad de los términos de búsqueda, la frecuencia en la base de datos, la posición del término en el documento, etc.

Así, por ejemplo, el algoritmo que utiliza Altavista para el establecimiento de la clasificación considera tres aspectos: el lugar en que las palabras de búsqueda aparecen en el recurso —si aparecen al inicio del texto como en títulos o cabeceras, tienen alto valor—, la proximidad de los términos, y la frecuencia de

la ocurrencia de las palabras o frase de búsqueda en la fuente. HotBot, por su parte, ordena según diversos criterios: frecuencia de la palabra en el texto, posición en el texto, y la extensión del texto —las palabras que aparecen en un texto corto tienen mayor peso en el ranking que las que aparecen en textos largos—.

Respecto a la presentación de los resultados, algunos motores permiten modificarla, aunque, generalmente, los datos que aparecen para cada recurso son: título, relevancia, resumen breve del contenido, tamaño del archivo, fecha, fecha de entrada en la base de datos, URL, lengua y términos de búsqueda.

Algunos motores añaden nuevas prestaciones. Es el caso de Excite, que presenta registros relacionados con los localizados. Lycos añade el número de enlaces externos a los que apunta el recurso. HotBot identifica duplicados y los agrupa como alternativos (*alternates*).

4. Evaluación de los motores de búsqueda

Para realizar una búsqueda precisa, es recomendable no sólo conocer las características y el funcionamiento de las herramientas a utilizar, sino que, además, es necesario saber interrogar correctamente a la base de datos. Para orientar tanto a profesionales de la información como a usuarios en general sobre estos aspectos y sobre las prestaciones de los motores de búsqueda, en estos últimos años se han realizado un gran número de estudios que han evaluado en determinados motores algunas de sus capacidades. Dong y Su (1997) revisan los desarrollados entre 1995 y 1996.

Para Schlichting y Nilsen, (1996) al evaluar un buscador es más importante la calidad de la información que la cantidad, por ello recomiendan el uso de directorios, ya que ofrecen recursos que han sido previamente seleccionados, si bien no hemos de perder de vista que los administradores del directorio y los usuarios pueden utilizar diferentes estándares de calidad y además el usuario puede buscar información personal que puede no localizar en el directorio.

Jacob (1995) señala los siguientes criterios para evaluar motores de búsqueda:

- Que sea fácil de usar.
- Rápido.
- Que soporte tanto búsquedas simples como complejas, con indicadores booleanos.
- Que permita búsquedas por partes de palabras, uso de truncamientos y comodines.
- Que permita búsquedas por frases y palabras próximas.
- Que permita al usuario un control.

- Que tenga un tesoro para determinar el término de búsqueda.
- Que permita a los usuarios elegir el máximo número de registros en la recuperación.
- Que indice el texto completo de documentos.
- Que proporcione en el resultado el título y la URL.
- Que presente los resultados de forma fácil de interpretar con indicadores de relevancia o sistema de rankings.
- Que indique cuando se indizó el recurso.
- Actualización regular de la base de datos para borrar enlaces nulos.
- Permitir al usuario registrar URLs no incluidas en la base de datos.

Estos aspectos permiten una evaluación individual del motor de búsqueda. Pero la evaluación también puede ir dirigida a tratar de establecer hasta que punto estas herramientas pueden considerarse complementarias o hasta que punto podemos hablar de solapamiento entre ellas en cuanto a la recuperación de información.

5. Solapamiento en las de herramientas de búsqueda de la WWW

5.1 Metodología

Para la elección de los buscadores objeto del presente estudio, hemos seleccionado los más populares y aquellos cuyas características y prestaciones han sido más estudiadas. Nos hemos basado en los trabajos de Falk, Zorn et al., Schlichting et al., Schwartz, Chu y Rosental, Peterson y otros. Los servicios de búsqueda elegidos fueron los siguientes: Altavista, Excite, Fast, Google, HotBot, Infoseek, Lycos, Northern Light y Yahoo.

Una vez seleccionados los buscadores, se establecieron cuatro temas de búsqueda en inglés, relacionados con el campo de la Documentación, siendo el último más específico para determinar el comportamiento de los buscadores ante búsquedas de este tipo. Se realizaron las búsquedas de forma correlativa en un mismo motor, efectuándose posteriormente en el resto. Se utilizó la opción de búsqueda simple, insertando el término entre comillas, para forzar una búsqueda por frase. Los temas sobre los que se ejecutaron las búsquedas fueron cuatro: *Information Retrieval*, *Knowledge Management*, *Digital Library* y *Distributed Information Retrieval*.

Tras la operación de búsqueda, se analizaron los 50 primeros registros recuperados para calcular cuantos correspondían a recursos recuperados exclusivamente por un motor —es decir, registros únicos— y cuántos habían sido recuperados por más de una de estas herramientas de búsqueda.

5.2. Resultados

Los resultados fueron los que aparecen en la figura 1.

5.3 Análisis de los resultados

Con relación al solapamiento en las cuatro búsquedas, podemos observar que la búsqueda número tres (*Digital Library*) es la que ofrece un mayor solapamiento con 211 registros solapados de los 450 analizados. Este número descien- de a 178 en la segunda búsqueda (*Knowledge Management*); y en la cuarta (*Distributed Information Retrieval*), que trataba de recuperar recursos más espe- cíficos, el número de recursos solapados fue de 129. Finalmente, el solapamien- to es menor en la búsqueda número uno (*Information Retrieval*) con 99 recursos de este tipo. De estos resultados, en primer lugar llama la atención el alto por- centaje de solapamiento que se observa en los motores analizados, y que va des- de el 22% en la primera búsqueda al 46,8% en la tercera.

Esta diferencia puede tener su explicación en el modo de funcionamiento de los robots que forman parte de los buscadores, a los que les será más fácil encontrar

		Búsqueda 1		Búsqueda 2		Búsqueda 3		Búsqueda 4		Total fila	
		Únicos	Solapados	Únicos	Solapados	Únicos	Solapados	Únicos	Solapados	Únicos	Solapados
Altavista	nº registros	35	15	21	29	15	35	39	11	110	90
	porcentaje	70%	30%	42%	58%	30%	70%	78%	22%	55%	45%
Excite	nº registros	46	4	33	17	31	19	47	3	157	43
	porcentaje	92%	8%	66%	34%	62%	38%	94%	6%	78,5%	21,5%
Fast	nº registros	39	11	24	26	18	32	39	11	120	80
	porcentaje	78%	22%	48%	52%	36%	64%	78%	22%	60%	40%
Google	nº registros	32	18	29	21	20	30	29	21	110	90
	porcentaje	64%	36%	58%	42%	40%	60%	58%	42%	55%	45%
HotBot	nº registros	28	22	24	26	17	33	23	27	92	108
	porcentaje	56%	44	48%	52%	34%	66%	46%	54%	46%	54%
Infoseek	nº registros	37	13	30	20	25	25	38	12	130	70
	porcentaje	74%	26%	60%	40%	50%	50%	76%	24%	65%	35%
Lycos	nº registros	48	2	43	7	44	6	43	7	178	22
	porcentaje	96%	4%	86%	14%	88%	12%	86%	14%	89%	11%
Northern Light	nº registros	44	6	30	20	34	16	39	11	147	53
	porcentaje	88%	12%	60%	40%	68%	32%	78%	22%	73,5%	26,5%
Yahoo	nº registros	42	8	38	12	35	15	24	26	139	61
	porcentaje	84%	16%	76%	24%	70%	30%	48%	52%	69,5	30,5%
Total columna	nº registros	351	99	272	178	239	211	321	129		
	porcentaje	78%	22%	60,5%	39,5%	53,2%	46,8%	71,4%	28,6%		

Figura 1. Resultados de las búsquedas en número de registros y porcentaje

enlaces que apunten hacia bibliotecas digitales, ya que son recursos de un gran interés y que aparecen muy a menudo referenciados en distintas páginas web, que otro tipo de recursos que traten de temas específicos como pueda ser el de la recuperación de la información, que a su vez contengan enlaces a recursos relacionados.

El total acumulado de las cuatro búsquedas señalan, por este orden, a HotBot y Altavista como los motores que recuperan mayor número de recursos solapados, seguidos por Google y Fast. Esto nos indica que su uso combinado puede no ser una buena solución para localizar información exhaustiva sobre un tema. Por el contrario, Lycos es el que recupera un menor número de registros solapados, ya que de los 200 recursos analizados en las cuatro búsquedas sólo 22 fueron recuperados también por otro motor. Ello indica que este buscador es una buena herramienta en búsquedas exhaustivas de documentación, y que, en este sentido, es recomendable utilizarlo junto a cualquiera de los demás. En próximos estudios trataremos de analizar en que medida existe solapamiento entre unos motores y otros, para ver que combinación entre ellos puede resultar más efectiva.

El análisis de la recuperación en cada una de las búsquedas nos lleva a apuntar los siguientes aspectos:

- *Búsqueda 1: Information Retrieval.* Los motores con menor solapamiento son Lycos con sólo dos (4%) registros solapados, Excite con cuatro (8%), Northern Light con seis (12%) y el Directorio Yahoo con ocho (16%). Los motores con mayor número de registros solapados son HotBot con 22 (44%), Google con 18 (36%), Altavista con 15 (30%), Infoseek 13 (26%) y Fast 11 (22%).
- *Búsqueda 2: Knowledge Management.* Lycos es de nuevo el motor con menor número de recursos solapados, aunque la cifra, en esta ocasión asciende a 7 (14%). Yahoo acumula 12 (24%) registros solapados, Excite 17 (34%) y Northern Light e Infoseek 20 (40%). En esta búsqueda es Altavista el que mayor número de recursos solapados recupera con 29 (58%) seguido de HotBot y Fast, ambos con 26 (52%), a los que sigue Google con 21 (42%).
- *Búsqueda 3: Digital Library.* En esta búsqueda es en la que se observa un mayor solapamiento entre motores. Lycos se mantiene en primer lugar con tan sólo seis registros solapados (12%), seguido de nuevo por Yahoo con 15 (30%). Northern Light pasa de nuevo a formar parte de este grupo con 16 (32%) y Excite se mantiene con 19 (38%). Con mayor solapamiento tenemos a Infoseek —25 (50%)—, Google —30 (60%)—, Fast con 32 (64%) y HotBot con 33 (66%). Altavista es el motor que registra mayor número de registros solapados con un total de 35 (70%).

- *Búsqueda 4. Distributed Information Retrieval.* La búsqueda de información más específica da en determinados motores resultados sorprendentes. Así, en esta ocasión es Excite el que menor solapamiento ofrece con tan sólo tres registros (6%), seguido de Lycos con siete (14%), de Northern Light, Fast y Altavista con 11 (22%), siendo de destacar el mejor comportamiento de este último buscador respecto a las búsquedas anteriores. Infoseek presenta 12 registros solapados (24%). En el grupo de mayor solapamiento, tenemos a Google con 21 (42%) y a Yahoo que en esta búsqueda recuperó 26 (52%) registros solapados, viéndose tan sólo superado por HotBot con 27 (54%). Estos últimos resultados son de un gran interés, ya que indican, una vez más, la necesidad de analizar el solapamiento no sólo individual sino el que se da entre diferentes motores, ya que permitiría explicar, en ciertos casos a que obedecen estas cifras, ya que en el caso de HotBot y Yahoo, en esta última búsqueda se observa que entre ambos hay un gran solapamiento, pues de los 26 registros solapados en Yahoo, 22 aparecen en HotBot, y a la inversa, de los 27 solapados de HotBot, 22 los recuperó también Yahoo. Como hemos señalado, este es un aspecto muy importante que habrá que analizar en próximos estudios.

En resumen, podemos señalar como Lycos ofrece un buen comportamiento en las cuatro búsquedas, mientras que Yahoo, en la cuarta búsqueda se descuelga del grupo de buscadores con menor solapamiento. A Northern Light le ocurre más o menos lo mismo en la segunda búsqueda. Excite presenta también una gran regularidad, estando siempre en el grupo con menor solapamiento, aunque en la búsqueda con mayor solapamiento, su número de recursos solapados aumenta considerablemente. Los buscadores en los que se observa un importante solapamiento en las cuatro búsquedas son HotBot y Google, ya que en todas ellas aparecen en el grupo de mayor solapamiento. Al motor Altavista le salva la cuarta búsqueda, al igual que le ocurre a Fast e Infoseek.

6. Conclusiones

La siguiente experiencia muestra un desigual solapamiento en los motores de búsqueda dependiendo del tipo de búsqueda que se realice. Es sorprendente en este sentido que un mismo buscador, como es el caso de Altavista, dé en la primera de las búsquedas 35 registros únicos (70%) y 15 (30%) solapados y estos valores se vean invertidos en la tercera búsqueda, con 15 (30%) registros únicos y 35 (70%) solapados. A pesar de ello, hemos podido constatar que existe cierta uniformidad en los resultados de las cuatro búsquedas. Así, Lycos es en todas ellas el motor con menor solapamiento, uniformidad que se rompe en el caso de Yahoo en la cuarta búsqueda, debido al gran solapamiento observado en esta ocasión con HotBot. Esto mismo le ocurre a Northern Light que aumenta su solapa-

miento en la segunda búsqueda, aunque en este caso el solapamiento no es debido a la mayor coincidencia de los resultados con un determinado motor. Además, se ha podido constatar la necesidad de analizar en este tipo de estudios el solapamiento entre motores, aspecto que abordaremos en próximos trabajos.

También existe uniformidad en cuanto a los resultados que ofrecen los motores con mayor solapamiento que son HotBot y Google, ambos entre los que mayor solapamiento registran en las cuatro búsquedas, seguidos de Altavista, Infoseek y Fast, que destacan en tres búsquedas. El menor solapamiento observado en buscadores como Lycos y Yahoo hace que sean especialmente recomendables, en combinación con otros motores para recuperación exhaustiva de información en la web.

7. Notas

- (1) Como señala Phil Bradley (1999), se estima que el número de páginas web en Internet supera los 800 millones.
- (2) Podemos encontrarlo en la siguiente URL: <http://www.iconovex.com/SWAPI/SWA-PIWP.HTM>

8. Bibliografía

- Andrews, W. Searching questions. // *Web Week*. 3 : 28. 1, 44.
- Bradley, P. (1999). The great search engine con-trick. // *Online Information 99 Proceeding*. 259-262.
- Codina, L. (1997). Cómo funcionan los servicios de búsqueda en Internet: un informe especial para navegantes y creadores de información (Parte I). // *Information World en Español*. 6 : 5 (mayo 1997) 22-27.
- Codina, L. (1997). Cómo funcionan los servicios de búsqueda en Internet: un informe especial para navegantes y creadores de información (Parte II). // *Information World en Español*. 6 : 6 (junio 1997) 19-27.
- Courtois, M.P.; Baer, W.M.; Stark, M. (1995). Cool tools for searching the Web: a performance evaluation. // *Online*, 19 : 6 (1995) 15-32.
- Chowdhury, G.G. (1999). *Introduction to modern information retrieval*. London, Library Association Publishing, 1999.
- Dempsey, L. Meta Detectors. URL: <<http://www.ariadne.ac.uk/issue3/metadata>>
- Dong, X. ; Su, L. (1997). Search engines on the World Wide Web and Information retrieval from the Internet: a review and evaluation. // *Online & CDROM review*. 21 : 2 (1997) 67-82.
- García Marco, J.; Tramullas Sanz, J. (1996). *World Wide Web: fundamentos, navegación y lenguajes de la red mundial de información*. Madrid : RA-MA, 1996.
- Hock, R. (1999). *The extreme searcher's guide to web search engines: a habdbook for serious searcher*. Medford (New Jersey): CyberAge Books, 1999.

- Lancaster, F.W. (1998). *Indexing and abstracting*. 2nd. ed. Londres: Library Association Publishing, 1998.
- Maldonado Martínez, A.; Fernández Sánchez, E. (1999). Comparing Internet Search Tools. // *Online Information 99 Proceedings*. (1999) 263-266.
- Marcos Mora, P. (1998). Motores de recuperación de información: un análisis comparativo. (parte I). // *El profesional de la información*. 7 : 1-2 (en.-feb.1998) 18-22.
- Marcos Mora, P. (1998). Motores de recuperación de información: un análisis comparativo. (parte II). // *El profesional de la información*. 7 : 3 (mar. 1998) 13-20.
- Milstead, J.; Feldman, S. (1999). Metadata: Cataloging by Any Other Name. // *ONLINE* (January 1999). URL: <<http://www.onlineinc.com/onlinemag/OL1999/milstead1.html>>
- Notess, Greg R. (2000). Search Engine Statistics: Unique Hits Report. // URL: <<http://www.notess.com/search/stats/unique.shtml>>.
- Taylor, Arlene, G.; Clemson, P. (1999). Acces to Networked documents. Catalogs? Search Engines? Both? URL: <<http://www.oclc.org/oclc/man/colloq/taylor.htm>>.
- Venditto, G. (1996). Search engine showdown: IW labs test seven Internet search tools.// *Internet World*. (May. 1996)79-86.
- Winship, I. R. (1995). Word Wide Web searching tools-an evaluation.// *VINE*. 99 (1995).49-54. URL=<<http://bubl.bath.ac.uk/http://bubl.bath.ac.uk/BUBL/IWinship.html>>