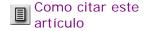


ACIMED ISSN 1024-9435 versión impresa

ACIMED v.11 n.3 Ciudad de La Habana Mayo-jun. 2003



El uso de los buscadores en Internet

Lic. Ania Torres Pombert1

Resumen

Se definen, describen y caracterizan un grupo de los principales motores de búsqueda que operan en Internet. Se analizan los métodos reconocidos para su evaluación y selección, así como se ofrecen algunos consejos para perfeccionar el proceso de búsqueda con el fin de mejorar sus resultados. Se exponen, de forma general, sus semejanzas y diferencias, ventajas y desventajas y se analizan sus tendencias actuales de uso.

Clasificación: Artículo de revisión

Descriptores (DeCS): INTERNET/clasificación

Descriptores (DeCI): INTERNET; MOTORES DE BUSQUEDA; TECNICAS DE EVALUACION ESTRATEGIAS DE BUSQUEDA/tendencias; METABUSCADORES

Abstract

A group of the main search tools operating in Internet are defined, described and characterized. The methods recognized for their assessment and selection are analyzed, and some pieces of advice are given to improve the selection process in order to obtain better results. Their difference and similarities, advantages and disadvantages are broadly expounded, and their current usage trends are analyzed.

Classification: Review article

Subject headings (DeCS): INTERNET/classification

Subject headings (DeCI): INTERNET; SEARCH ENGINE; EVALUATION TECHNIQUES SEARCH ESTRATEGIES/trends; META-SEARCH ENGINE

Desde hace muchos años, ya en la era moderna, cuando alguien necesitaba información de carácter científico, comercial o de entretenimiento solía encaminarse hacia una biblioteca pública, especializada o académica, en la que un bibliotecario o referencista lo orientaba; él podía también consultar los tradicionales catálogos de autor, título, materia u otro que describiera los documentos existentes. En el peor de los casos, el problema se resolvía cuando se remitía el usuario a otra biblioteca, pero inevitablemente se produjo un crecimiento exponencial de la literatura, sobre todo científica, que aun cuando coloca, a disposición de la comunidad académica, una gran variedad de recursos, requiere de una inversión importante de tiempo y esfuerzo para su consulta, evaluación y asimilación.

El desarrollo científico y tecnológico, con su crecimiento agigantado, ha generado, entre otros fenómenos,

el incremento y perfeccionamiento acelerado de las nuevas tecnologías de información y comunicación, justamente en función de un mejor registro, procesamiento, búsqueda y diseminación de la información; sin embargo, el problema para acceder sólo a la información relevante persiste.

Sin necesidad de analizar la evolución de las tecnologías de información, está claro que su resultado más importante es Internet. Si se retoma la idea inicial, puede pensarse que ahora, cuando alguien necesita realizar una búsqueda, incluso en el tema más sencillo, piensa en Internet y no en una biblioteca tradicional, y es que Internet, es como una gran biblioteca, con múltiples departamentos especializados en diferentes materias, es una biblioteca tan grande, como una ciudad que es muy fácil perderse.

Sucede que Internet es indiscutiblemente un medio de publicación rápido, libre de arbitraje, requisitos y normas, con cobertura internacional, que hace de ella el soporte preferido, súmese a esto el hecho de que muchos documentos que se editan en soporte impreso, tienen versiones en Internet precisamente a favor de su acceso y visibilidad, estos aspectos explican por sí solos el hecho de que la cantidad de información disponible en el web sea abrumadora.

Afortunadamente, a la par del crecimiento de Internet se han desarrollado y perfeccionado los motores de búsqueda, dirigidos a facilitar la navegación y el hallazgo de la información necesaria.

Evidentemente, si se comparan los motores de búsqueda de hace unos años atrás con los actuales será fácil percatarse de que la cantidad de información procesada en sus bases de datos es mucho mayor, debido precisamente a que la información en la red se multiplica a diario. Por otra parte, se estima que, mientras en 1995, apenas existía una docena de motores de búsqueda, hoy se calculan en alrededor de 2000, cada uno con características diferentes, facilidades particulares, formas de funcionamiento e interfaz propia. Si bien es cierto que en el inicio los motores de búsqueda, la preocupación de los navegantes era encontrar alguno cuyo host estuviera disponible en el momento en que fuera a hacerse uso de él o simplemente saber cuál realizaría la búsqueda de manera más fácil, en la actualidad el primer problema está en identificar, seleccionar y decidirse por uno de ellos. 1

Indiscutiblemente, aunque ellos constituyen un importante paso de avance, no son la solución al problema. El propio incremento de los motores de búsqueda, disponibles en la red, ha impuesto la necesidad, para la mayoría de los navegantes, de "hacer búsqueda de buscadores" con el fin de determinar cuál es el mejor para un determinado tema, incluso antes de formular la búsqueda que necesita para resolver su problema de investigación.

Mucho se considera el tiempo que los navegantes requieren para encontrar información relevante en Internet sobre un tema determinado, aun cuando poseen una ruta más o menos específica para buscarla; sería demasiado emplear el tiempo en hacer búsquedas en uno y otro buscador con el propósito de evaluar sus resultados particulares cada vez que se navegue y, a partir de ello, determinar cuál es el más relevante para responder a su demanda.

Para solucionar este problema, surgieron entonces los metabuscadores. Ellos tratan de simplificar el proceso, al consultar, de manera simultánea, varios buscadores; dicho proceso es transparente para el usuario. Incluso cuando se consideran un avance, debido al ahorro de tiempo y su amplia cobertura, estas herramientas tienen también sus desventajas, determinadas en lo fundamental, por sus diferentes formas de funcionamiento que generan resultados de diferentes niveles de relevancia a pesar de responder a una misma pregunta.

El cuestionamiento actual está en saber qué tipo de buscador utilizar para obtener una mayor eficiencia en el proceso de búsqueda; como no existe una respuesta definitiva al respecto, la única alternativa es seguir muy de cerca su desarrollo, así como estudiar sus ventajas y desventajas para determinar cuál emplear en correspondencia con lo que se busca.

Hacia esa línea está dirigido el presente trabajo, es así que se pretende actualizar los conocimientos acerca del desarrollo de los buscadores en Internet y realizar un análisis de las tendencias actuales de su uso, todo a partir de los diferentes puntos de vista de algunos expertos en la materia.

Métodos

La búsqueda en el web mediante Google.com, por términos como "motores de búsqueda" y "search engines", devolvió entre los resultados más importantes sitios como: http://www.allsearchengines.com y http://www.searchenginewatch.com, ambos con una gran cantidad de enlaces importantes a trabajos sobre estos temas. Producto de una lectura crítica de la bibliografía consultada, se obtuvo un grupo de conclusiones importantes en cuanto a definiciones y tipos de los buscadores, funcionamiento, estrategias para un uso más efectivo, ventajas y desventajas.

Buscadores en Internet. Conceptos, clasificación y características

Muchos son los nombres que han recibido los buscadores de acuerdo con su estructura y funcionamiento. Algunos autores los clasifican en índices, robots y metabuscadores,2 otros los agrupan en directorios y buscadores,3 o bien en índices y motores de búsqueda.4,5 Todos estos términos tienen su equivalente en inglés, sin embargo, en esta lengua además de "index", "meta- search engine" "directory" y "search engine", para los 2 últimos aparecen en la literatura sinónimos como "crawler o spider" y "human-powered" respectivamente. Veamos a continuación algunas definiciones.

En el artículo "Buscadores: Los puntos de partida en la red"2 se plantea que se dividen básicamente en índices, robots y metabuscadores. Aquí se consideran los índices como los pioneros en Internet, ellos funcionan mediante la incorporación constante de nuevas páginas a su base de datos, generalmente disponen de un evaluador que minimiza el número de páginas que ingresan al sistema y asegura su calidad; los robots (llamados por la gran mayor ía motores de búsqueda) disponen de programas "arañas" que recorren la red e incluyen de forma automática "todo cuanto encuentran", esto genera grandes dificultades para los navegantes al momento de seleccionar los recursos, sobre todo en el caso de búsquedas con un alto recobrado; por último, los metabuscadores se limitan a redirigir la pregunta que reciben a varios buscadores a la vez, y según el estado de opinión prevaleciente, suele ser difícil pero recomendable el trabajo con ellos, cuando no se encuentra información en los índices y motores.

Efectivamente, de acuerdo con las diferentes formas de funcionamiento, es acertado establecer estas 3 categorías para clasificar los buscadores, pero cabría preguntar qué significa "todo lo que encuentran" los robots, o será que realmente no hay límites en lo que ellos recogen en la red.

En otra clasificación, se dividen en directorios y buscadores. Se dice que los directorios son catálogos de direcciones que tienen relativamente pocos sitios (cientos o miles), pero ellos se seleccionan y verifican; esto supone cierta calidad, incluso, algunos directorios que contienen una evaluación del sitio que indica cuán bueno es, otros presentan un breve resumen, escrito por un reseñador o evaluador, que explica el contenido (la mayoría sólo tienen una descripción que proporciona el propio autor del sitio); por otra parte, se definen los buscadores como enormes índices que comúnmente registran millones de páginas y donde las búsquedas producen miles de sitios como respuesta.3

En estas clasificaciones, si bien es cierto que la definición de los directorios es muy amplia y clara, el concepto de buscador se queda sin explicar ¿cómo funciona esta herramienta?, además ¿por qué llamarle buscador?, ¿acaso no son buscadores también los directorios?; por otra parte, estos llamados "buscadores" no deben considerárseles índices precisamente, porque ellos no muestran listas de términos o categorías a los navegantes, el índice no es más que una parte interna donde se almacena una copia de cada una de las páginas identificadas.

Otras 2 clasificaciones, tal vez menos comunes, son buscadores de categorías y buscadores de contenido (o de palabras por contexto), al primero se le llama también índices, que está claro que se refieren a los directorios y aunque no se describen conceptos los ejemplos hablan por sí solos: por categorías, Yahoo y por contexto, Altavista.6

Es oportuno señalar que aun cuando los motores de búsqueda no se estructuran por categorías temáticas como los directorios, existen también motores especializados en áreas geográficas y temáticas, por tipos de recursos, entre otros.

Sullivan,7 un importante autor vinculado al tema de los motores de búsqueda, editor de la revista Search Engine Watch, plantea que el término "motor de búsqueda" se emplea indistintamente para los motores basados en los programas "crawler" o "spider", y para los directorios accionados por humanos, pero enfatiza

en las diferencias entre ambos justamente porque los primeros trabajan automáticamente y los segundos dependen del factor humano.

Un aspecto importante, tratado por este autor, son las partes de los motores de búsqueda: el primer elemento es el "spider" o "crawler", encargado de visitar las páginas web, leerlas y analizar sus enlaces a otras páginas, esta operación se repite cada cierto tiempo en busca de cambios. El segundo elemento es el índice, tambi én llamado catálogo donde se almacena una copia de cada página web que se encuentra, cada vez que cambia la página se actualiza con la nueva información, claramente transcurre un tiempo desde que se encuentra la página hasta que se indiza y no es hasta entonces que está disponible en el motor de búsqueda. El tercer elemento es el software, programa que se encarga de examinar los millones de páginas web contenidas en el índice, busca aquellas que contengan las palabras iguales a las solicitadas en la búsqueda, para luego ordenarlas según su relevancia. Para muchos navegantes es una interrogante saber si realmente el orden de los registros es aleatorio o responde a un determinado método de ordenamiento, más adelante se ofrecer án algunas consideraciones al respecto.7

Como puede apreciarse, en cada una de las definiciones, la función de cada una de estas herramientas es la misma: "buscar en la red", todos entonces se sitúan en una gran categoría general: buscadores. Son páginas web que analizan la red con el objetivo de que los navegantes encuentren los sitios donde se encuentra la información que ellos necesitan, ponen a su disposición "alternativas" para moldear las búsquedas con ayuda de sus propios lenguajes de interrogación, la diferencia está en el funcionamiento de cada uno; según esto, se dividen en 3 subcategorías, que en la opinión de la autora la manera más acertada de agruparlos es en:

- Directorios (se refiere a los llamados índices o buscadores de categorías).
- Motores de búsqueda (que algunos le llaman buscadores, robots, arañas o buscadores de contenido).
- Metabuscadores (que se abordarán más adelante).

Si se analizan detenidamente los conceptos referentes a las 2 primeras categorías es fácil encontrar ventajas y desventajas en cada una, por ejemplo, los directorios son excelentes para buscar por temas o categorías y los resultados se ordenan por dichas categorías, pero ellos no catalogan tantas direcciones como un motor de búsqueda; por su parte, los motores de búsqueda son muy útiles para buscar datos específicos y permiten encontrar "casi todo", pero esto, a su vez, produce como resultado demasiados sitios de todo tipo y calidad (miles) en una respuesta típica.

Pero hablar en la actualidad de buscadores, obliga necesariamente a hablar de los metabuscadores, innumerables trabajos, en la literatura especializada, abordan el tema, debaten sobre aspectos como funcionamiento, ventajas, desventajas y eficiencia, entre otros. A pesar de los beneficios indiscutibles de los motores de búsqueda y de los directorios temáticos, su crecimiento condujo a la creación de estas nuevas herramientas. 1,8-12

Según *Tyner*, 11 se conoce como metabuscadores, a aquellos que permiten interrogar varias bases de datos simultáneamente desde una única interfaz; aunque ellos no ofrecen el mismo nivel de control sobre la lógica y la interfaz de búsqueda que los motores y directorios, la mayoría son bastante rápidos. Apunta, además que la tendencia es mejorar sus capacidades con la incorporación de opciones de búsqueda al uso de los operadores convencionales, entre ellas, se destaca la posibilidad de poder escoger los motores de búsqueda donde debe buscarse, la búsqueda por campos (título, URL, etc.), entre otras.

Otras cualidades se agregan a la definición anterior, se plantea que al no tener sus propias bases de datos, a diferencia de los motores de búsqueda y los directorios temáticos, los metabuscadores no almacenan páginas web, no agregan direcciones, ni clasifican y reseñan sitios web, sino que envían una solicitud de búsqueda a varios buscadores a la vez.8

Stanley, 1 considera que los metabuscadores constituyen el próximo eslabón dentro de la cadena alimenticia de la información en los motores de búsqueda convencionales, se consideran son superiores. Muy claramente expone aquellos elementos que considera sus ventajas esenciales:

- Se accede a una sola página web para formular la búsqueda.
- Se necesita sólo conocer la interfaz de una página para la búsqueda.

- Se formula sólo una vez la estrategia de búsqueda.
- Los resultados permiten redireccionar la búsqueda a otros buscadores.
- Se obtienen los resultados integrados, a partir de varios buscadores.

Una importante reflexión de la autora apunta al siguiente hecho: que los metabuscadores no dispongan de sus propias bases de datos sino que interroguen las de otros buscadores, significan una importante reducción en los costos por concepto de hardware para los productores. Ciertamente podría considerarse esta, una causa para el auge de los metabuscadores y su aceptación entre los navegantes.

No es menos cierto que los elementos descritos por *Stanley*1 sobre los metabuscadores, constituyen ventajas para el proceso de búsqueda; sin embargo, otros autores, que enfatizan en los problemas que puede causar su uso, llaman la atención de los navegantes para que los tengan siempre muy en cuenta, reconocen, por tanto, que son una herramienta muy útil.

Liu8 expone algunos elementos que son válidos anotar, sobre todo porque otros autores, que han escrito sobre el tema de los metabuscadores, no los tratan. Él plantea que es totalmente imposible que estas herramientas puedan unificar todas las ventajas de cada uno de los motores y que, por consiguiente, las búsquedas booleanas pueden generar resultados diferentes en diversos buscadores, las búsquedas por frases puede que no se ejecuten en alguno de ellos, y otros elementos como el uso de limitadores pueden sacrificarse. Apunta que los metabuscadores no devuelven, desde cada buscador, todas las páginas que se corresponden con la solicitud sino que toman un rango entre 10 y 100 registros de cada uno, por lo que efectivamente permiten redireccionar la búsqueda, como apunta *Stanley*,1 una vez que se escoja el motor, que más resultados relevantes arrojó, para continuar la búsqueda.

Algunos autores aseguran que la mayoría de los grandes buscadores utilizan lenguajes similares para interrogar sus bases de datos. Por esto, los resultados de la búsqueda en los metabuscadores pueden verse favorecidos si se conocen, en alguna medida, los lenguajes utilizados por los buscadores que estos procesan. Afortunadamente algunos metabuscadores ofrecen al navegante la opción de escoger dentro de un grupo determinado de buscadores disponibles, cuáles incluir en su metabúsqueda. Si se escogen aquellos conocidos por su confiabilidad, potencia y rapidez, los resultados de la búsqueda serán mejores. 1

Cierto es, que cada categoría difiere de la otra sustancialmente por el funcionamiento y por las herramientas para la búsqueda en la red, pero incluso dentro de una misma categoría se diferencian unos de otros. Son eBlast y Yahoo directorios temáticos, pero no son iguales, son Altavista y Northern Light motores de búsqueda, pero son diferentes, de igual forma son Metacrawler y Savvy Search metabuscadores, y difieren uno del otro.

El uso de estos 3 mecanismos no presenta grandes diferencias a primera vista, por lo que es posible que los navegantes no distingan en cuál de ellos realizan habitualmente sus búsquedas, todos muestran un recuadro para introducir las palabras claves que se desean encontrar, solo que varía la forma en que se debe plantear la cuestión, porque, en algunos casos, puede complicarse con elaboradas reglas lógicas. Esta tabla resume brevemente la naturaleza de cada categoría de buscador.

Tabla. Caracterización de las diversas categorías de buscadores

Directorios temáticos	Motores de búsqueda	Metabuscadores
Bases de datos más pequeñas, menos actualizadas, y más elaboradas gracias a la presencia del factor humano.	Bases de datos más amplias y actualizadas.	No tiene bases de datos propias, sino que buscan automáticamente en las de otros buscadores.
Colocan la información por temas y categorías, una vez		No almacenan información porque no

recopilada, de forma manual en sus índices.	índices sin ordenarlas por temas, de manera automática y periódica.	dependen de bases de datos propias.
No realizan las búsquedas en Internet "en vivo", almacenan los datos de los sitios y ofrecen enlace a éstos.	No realizan las búsquedas en Internet "en vivo", sino en las copias de las páginas que almacenan en sus índices.	Envían su búsqueda a varios motores, sus resultados dependen de que estos estén disponibles en el momento de la búsqueda, o se descarguen en el período de tiempo permisible.
Son fáciles de usar, permiten, en primer lugar, ubicar la búsqueda en un tema determinado.	Son más difíciles, se requiere explotar al máximo las opciones de búsqueda porque contienen más información.	Son difíciles de usar para búsquedas muy precisas, porque tienen menos control de la búsqueda al interrogar varias bases de datos con interfaces diferentes.
Son convenientes para buscar información general, institucional porque devuelve resultados a las páginas principales.	Se utilizan para buscar información más escasa, especializada, actualizada o incluida en páginas personales.	Se recomienda para temas "oscuros", difíciles de encontrar.
LookSmart (http://www.looksmart.com) Open Directory (http://dmoz.org)Yahoo (http://www.yahoo.com)	Ejemplos Altavista (http://www.altavista.com) Google (http://www.google.com) HotBot (http://hotbot.lycos.com)	Metacrwler (http://metacrawler.com) NetLocator (http://nln.com) StartingPoint (http://www.sptp.com)

Mucho se ha debatido sobre la necesidad de conocer las herramientas con que se busca en la red, precisamente por la gran variedad que existe, donde unos suelen ser más eficientes que otros. Un aspecto importante a considerar para seleccionar los buscadores es la evaluación de sus bases de datos. Como se ha explicado anteriormente, los resultados de las búsquedas, obtenidos de cada buscador no son más que la "salida en pantalla" de la información de sus bases de datos, después que esta se localiza, procesa, evalúa y ordena. Por supuesto, no es comparable la evaluación de una base de datos en CD con otra en Internet, esta última es mucho más difícil.

Notess 13 expone, de manera muy sencilla, sus puntos de vista al respecto, plantea que el mayor problema para evaluar una base de datos en Internet es que sus registros nunca son estáticos, porque existen muchas páginas "efímeras" que se indizan en el tiempo en que se crea o actualiza la base de datos y poco tiempo después se mueven, cambian o rediseñan completamente; sin embargo, aun cuando el contenido de los registros es variable, plantea el autor, que las bases de datos en Internet pueden evaluarse en cuanto a alcance, estructura y actualidad. Se explica en este artículo, que para determinar el alcance de la base de datos se debe consultar la ayuda o las FAQs, se sugiere también, que el examen de la estructura es efectivo, por ejemplo, si el único acceso a los registros individuales es por una búsqueda de palabras claves, cabría entonces, pensar que es una base de datos que se genera automáticamente, mientras que, si tiene una estructura jerárquica de materias, debe construirse selectivamente. Ciertamente, estos son aspectos importantes, pero no se pueden obviar otros más elementales, sobre todo, para los buscadores de tipo directorio, como el título de la página, que se supone sea claro y refleje, con exactitud, el contenido del sitio

y la consulta del "about" que ofrece información general.

Wighton 14 concede también, una especial importancia a la consulta de las FAQs de los buscadores, antes de formular una búsqueda, enfatiza, que esto permite saber cómo se puede buscar y hasta dónde se puede llegar y, a partir de aquí, elegir o desechar el buscador.

Notess, 13 explica que la estructura de los registros de la base de datos, que se muestra en los resultados, puede parecer un elemento difícil de evaluar, sobre todo, porque a diferencia de otras, en las bases de datos en Internet, los campos no aparecen explícitamente, sin embargo, generalmente son fáciles de identificar, incluso existen buscadores como Google que explican detalladamente en su ayuda, a qué se refiere cada dato expuesto en los resultados de la búsqueda, la mayoría de los buscadores entregan sus resultados con los siguientes campos: título, descripción del contenido, categoría de materia y URL.

Realmente el navegante nunca podrá ver la estructura de la base de datos, por esto resulta un poco difícil saber cómo se maneja una búsqueda, una vez formulada la pregunta a la base de datos, pero conocer a qué se refiere cada elemento mostrado en los resultados puede ser un factor determinante a la hora de decidir cuáles son los registros más relevantes.

Otro aspecto enunciado por este autor para evaluar las bases de datos, es su modo de actualización, es difícil encontrar un motor que actualice constantemente cada uno de los sitios; para los grandes motores, porque tendrían una gran demanda en cuanto al ancho de banda y para los directorios, aunque más pequeños en volumen de información, porque exigiría una sobrexplotación del recurso humano. Plantea este autor, que es muy difícil determinar la frecuencia de actualización de una base de datos, pero el análisis de algunos de los resultados devueltos puede ser una vía para esto. Por ejemplo, en un intento de acceso a los enlaces que luego no se pueden alcanzar, o que te ofrecen un mensaje de que la página se ha movido a otra dirección sería conveniente fijarse en la fecha de la página si es posible, esto proporciona una medida del tiempo que lleva sin actualizarse; la última fecha de actualización de los sitios que forman parte de los resultados de la búsqueda muchas veces son de un año, 2 o más, ello indica que el creador del sitio cambió de empleo o sencillamente perdió el interés, un motor de búsqueda que se respete no debe incluir este tipo de sitios en sus índices.13

Existen varios trabajos donde se exponen diversos análisis y comparaciones entre buscadores, estos pueden ayudar a una mejor comprensión de su funcionamiento.6,11,12,15-22

Internet es una gran biblioteca, solo que a diferencia de la biblioteca tradicional no tiene libros y revistas impresas, materiales audiovisuales o diapositivas sino que sus documentos son sitios web sin control bibliográfico, no disponen de un equivalente al ISBN, que identifica los libros, tampoco existen sistemas análogos para su catalogación y clasificación, como los establecidos por la Biblioteca del Congreso, mucho menos, existe un catálogo central que registre todas sus páginas y, ni siquiera lo más elemental, muchos de los documentos que se encuentran en Internet carecen de autor y fecha de publicación.11

Evidentemente, como Internet es un medio libre para publicar, no existen normas estándares para hacerlo ni para indizar la información que fluye por ella. Las bases de datos de los buscadores no reconocen e indizan los mismos campos, y su funcionamiento para recuperar la información es diferente. De todo lo anterior, se infiere que para que los buscadores se conviertan en una herramienta eficaz es necesario que se conozcan muy bien.

Como se ha explicado anteriormente, la tendencia inevitable es la especialización de las herramientas de búsqueda en el web. Esto ha provocado la aparición de una gran variedad de buscadores, que generalmente resultan difíciles de evaluar por el navegante; por otra parte, en muchos casos han devenido, incluso en portales, añaden entonces a su oferta, un conjunto más o menos homogéneo de servicios, como noticias, bolsas de vivienda, chats o correo electrónico. Son gratuitos y se financian mediante la publicidad, debido a esto, el acceso a sus páginas principales es, a veces, la primera causa para que el navegante se sienta perdido entre tantos efectos especiales, comerciales y propagandas.

Muchos son los expertos que han sugerido los aspectos que deben considerar los navegantes para buscar en el web. En algunos casos, se apunta a que el navegante solo debe preocuparse por buscar la caja de diálogo, que permite introducir las palabras claves por las que pretende realizar la búsqueda, el cual tiene al lado un

botón o un gráfico con las palabras "buscar", "search", "go", "go to get it", donde el navegante mediante un click inicia el proceso de búsqueda, es recomendable buscar, cerca de esta caja de diálogo, opciones para seleccionar el número de registros a mostrar por página, los dominios donde buscar, los tipos de recursos o el tipo de lógica a emplear.6

Gresham23 opina que encontrar resultados relevantes en Internet es una mezcla de experiencia, técnicas, aptitudes, inventiva y buena suerte, todas combinadas con la habilidad de proceder en una dirección clara y metódica, en otras palabras, es simplemente "navegar con un propósito", apunta además que depende más de la estrategia de búsqueda que de la propia herramienta a utilizar. Sugiere la siguiente secuencia lógica de pasos:

- Determinar el tipo de información que se necesita (artículos científicos, estadísticas, documentos gubernamentales) y luego determinar qué tipo de organizaciones pueden ofrecer estos documentos.
- Crear una lista de todas las palabras claves posibles y de sus sinónimos.
- Determinar qué tipo de herramienta utilizar en la búsqueda (directorio, motor general o especializado, metabuscador) en dependencia de lo que se busca.
- Construir la estrategia de búsqueda y conducirla (en dependencia del buscador, se definirán las combinaciones de búsquedas, se debe ser lo más preciso posible y explotar las opciones disponibles).
- Evaluar los resultados de la búsqueda (si los primeros 15 registros no se consideran relevantes debe considerarse replantear la estrategia varias veces o cambiar de buscador, si el resultado persiste, si los resultados son relevantes debe valorarse la actualidad de los registros y si el sitio proviene de una fuente confiable).

Flanagan, 5 de manera más breve, apunta que para buscar en Internet existen 2 pasos claves a dar, en primer lugar, se ha de comprender correctamente qué se va a buscar, esto implica definir conceptos, determinar sinónimos, palabras relacionadas y, en segundo lugar, cómo usar las herramientas de búsqueda disponibles.

*Tyner*11 señala también, aspectos que considera importantes, y al igual que *Flanagan*,5 se refiere a formular la búsqueda y su alcance, mediante la identificación de los conceptos que abordan la cuestión, las palabras claves, sus sinónimos y variaciones, así como la preparación de la estrategia de búsqueda, sin embargo, en esta reflexión no se considera la herramienta que se utilizará, sin la cual no queda bien expresada la estrategia (aunque más adelante hace alusión a ello), ni se trata la evaluación de los resultados.

Realmente, el primer cuestionamiento del navegante debe ser siempre, ¿qué quiero buscar?, ¿cómo?, y ¿dónde puedo encontrarlo?. Los planteamientos anteriores, sobre todo de *Gresham*,23 constituyen una guía muy acertada de los pasos a seguir.

Una vez elaborada esta guía, deben considerarse otros elementos para lograr mejores resultados, estos aspectos sugeridos por varios autores,3,6,11,13,15,24 se consideran aplicables a la mayoría de los buscadores:

- Escribir en minúsculas y sin acentos.
- No usar una sola palabra en su búsqueda porque se obtendrán demasiados resultados.
- Emplear varias palabras claves que definan o determinen específicamente lo que se busca.
- Encerrar "entre comillas" las palabras que se desean encontrar juntas (frases).
- Usar los operadores más comunes (and, or, not).
- Buscar en las horas de menos tráfico, en consideración a que la mayoría de los motores están ubicados en los Estados Unidos.
- Consultar siempre la información disponible de cada buscador para conocer qué opciones puede utilizar para interrogar sus bases de datos.

Uso de operadores lógicos y funciones especiales de búsqueda

Como se expuso anteriormente, un buen planteamiento de la búsqueda es un factor decisivo para lograr resultados relevantes. La búsqueda por una sola palabra, o por un grupo de ellas escritas consecutivamente, sin una relación lógica sería un paso seguro a una búsqueda con bajísimos, casi nulos, índices de relevancia, pertinencia y precisión. Por eso, el uso de la lógica matemática es determinante para formular una búsqueda en la red, sea mediante signos matemáticos, símbolos o conjunciones en idioma inglés.6

La mayoría de los buscadores reconocen el uso de los operadores lógicos AND, OR y AND NOT para combinar palabras, el uso de comillas o paréntesis para buscar frases completas, e incluyen también, lo que algunos llaman funciones especiales de búsqueda, limitadores, entre otras herramientas.2,5,6,11,25,26 Otros usan los operadores de exactitud, proximidad, truncado y los posicionales.4,11 A continuación, se ofrece una breve descripción de la naturaleza de cada uno, según los criterios presentes en los trabajos referidos anteriormente.

Operadores booleanos

AND: indica que se recuperarán los documentos que contengan todas las palabras indicadas en la solicitud de búsqueda. Por esto, se considera muy útil para limitar una búsqueda y reducir el número de registros recuperados, mientras más términos se utilicen más específicos serán los resultados.

OR: ordena a la base de datos que devuelva todos los documentos que contengan, al menos, una de las palabras claves solicitadas. En este caso, el resultado puede ser: registros, con una, dos o todas las palabras incluidas en el planteamiento de la búsqueda. Es común utilizarlo cuando se puede buscar un término por sus sinónimos. En algunos buscadores se sustituye por el símbolo "|".

NOT o AND NOT: se utiliza entre 2 términos claves y se traduce por "no". Excluye de la búsqueda aquellos documentos que contengan la palabra clave a la que se refiere el operador. Es muy útil para eliminar los problemas causados por la polisemia; generalmente, se utiliza después de haber realizado una primera búsqueda, donde se obtengan resultados irrelevantes con las palabras solicitadas, que aparecen en un contexto diferente al que se busca. En algunos buscadores se sustituye por el símbolo "-" o "!".26

XOR: En la práctica no es un operador de uso muy frecuente, ni se conocen muchos buscadores que lo utilicen. Pide a la base de datos que devuelva todos los documentos que contengan una de las 2 palabras claves, pero excluye aquellos que contengan las 2 a la vez.26

Signos matemáticos: más y menos

En algunos buscadores se utilizan en sustitución de los operadores AND (+) y AND NOT (-). Se colocan al igual que los operadores delante de las palabras a la que se refiera sin dejar espacios.

Operadores posicionales

Los operadores posicionales definen, de alguna forma, cuál es la posición de las palabras claves dentro del documento y las interrelaciona entre ellas, según criterios de proximidad u orden.

NEAR: En español significa "cerca" y solicita los documentos que contengan ambas palabras claves, pero que no estén separadas por más de 10 palabras o 100 caracteres (aunque este número puede variar según el buscador). En algunos buscadores se puede sustituir por "~" o por "[]".26

FOLLOWED BY: En español significa "seguido de". Sus resultados son muy parecidos a los que produce Near, pero marca claramente cuál ha de ser el orden de las palabras claves. No es usado por muchos buscadores.26

Operadores de exactitud o truncado

Estos operadores tratan a las palabras claves como cadenas de caracteres, no como palabras completas. El sistema devuelve aquellos documentos que contengan a la palabra clave, pero también aquellas en la que la palabra clave sea raíz o sufijo. Se indican con símbolos como: *, #, ?, aunque hay buscadores que emplean diferentes símbolos en dependencia de la cantidad de caracteres que estos representen, por ejemplo Northern Light utiliza el * para representar varios caracteres, mientras que usa el % para indicar un solo carácter. Aunque la mayoría de los buscadores permiten solo el truncamiento a la derecha, los símbolos se colocarán al inicio, en el medio o al final de la palabra clave en dependencia de las facilidades permisibles. A esta facilidad de hacer búsqueda por términos truncados, algunos autores la llaman "uso de comodín o wildcards". *Tyner*11 ofrece ejemplos muy concretos para el uso de estas herramientas en varios buscadores.

Búsqueda de frases

En este caso, se considera a las palabras claves como si fueran una frase, es decir, deben aparecer en los registros recuperados como se introdujeron, una al lado de la otra, en el mismo orden. Es muy útil para hacer búsquedas muy concretas sobre un tema determinado, sobre todo cuando se considera una frase como una palabra y se combina con otras, a partir de los operadores lógicos, o los signos + y -.

Búsqueda por campos

La búsqueda por campos es una herramienta tradicional en la búsqueda en bases de datos convencionales. Se considera, en Internet, una de las técnicas más efectivas para restringir los resultados de la búsqueda y aumentar la relevancia.

Por título

Mediante cualquiera de las técnicas anteriores se indica a la base de datos que debe buscar solo en el campo título, como se muestra en el siguiente ejemplo:

title: "Panamerican Health Organization", nótese que no se deben dejar espacios entre (:) y la palabra clave. En este caso, se devolverán todos los sitios que incluyan esta frase en el título.

Por dominio

En este caso, se restringe la búsqueda a determinados dominios de países o tipo de instituciones (educacionales, no gubernamentales, militares, etc.). Por ejemplo: domain:cu +sld

La base de datos devolverá los sitios sobre ciencias de la salud en Cuba.

Por Host

Generalmente cuando los sitios son muy grandes los buscadores no los rastrean completamente sino que se limitan a buscar en las bases de datos propias de éstos. Esta técnica se utiliza cuando se necesita encontrar información en un sitio muy grande que no tiene un motor de búsqueda interno. Con esta técnica, puede especificarse al motor que busque en todas las páginas de determinado sitio, las palabras claves de interés. Un ejemplo podría ser: host:www.fda.gov +"clinical guides". En este caso el motor de búsqueda rastreará el sitio de la FDA completo en busca de la frase "clinical guides".

Por URL

Se limitan los resultados a páginas web donde las palabras claves se busquen en la URL. Por ejemplo, para conocer las páginas cuya dirección incluyen el término Infomed, se enuncia la búsqueda de la siguiente forma:

url:infomed

Por enlaces

Se utiliza para conocer las páginas que ofrecen enlaces a un sitio en particular. Por ejemplo, para conocer qué sitios tienen enlaces a Infomed, la búsqueda se formula de la siguiente forma: Link:www.infomed.sld.cu

Estas opciones de búsqueda constituyen las más comunes y pueden encontrarse indistintamente en uno u otro buscador, y aunque ninguno es capaz de conjugarlas todas, la combinación de búsqueda por diferentes campos, con la búsqueda de frases y el uso de los operadores lógicos puede ofrecer resultados muy precisos, por lo que es muy recomendable para restringir las búsquedas y obtener resultados más relevantes.

Como se explicó anteriormente, se sugiere la consulta de la ayuda, las FAQs y el "about" al iniciar el trabajo con un buscador, además de la consulta de literatura que recoge estudios completos al respecto.

Ordenamiento de los resultados de las búsquedas: ¿Mito o realidad?

Muchos navegantes se preguntan cómo es posible que en una búsqueda con miles de registros como resultado, estos puedan ordenarse por su grado de relevancia; los más incrédulos dudan, incluso, que los buscadores tengan esta capacidad.

Realmente resulta asombroso, pero lo es mucho menos, si se piensa que detrás de esta función está el conocimiento humano, creador de un software, que puede basarse en diferentes algoritmos, lo suficientemente inteligente para que el motor de búsqueda sea capaz de realizar esta tarea. Conocer cuál es el algoritmo que utiliza cada buscador en particular es casi un secreto pero la gran mayoría se basan en mecanismos similares.27

Chakrabarti,28 expone que el primer problema que enfrentan los buscadores para ordenar sus resultados es que no existe una fórmula matemáticamente precisa que permita el "mejor" ordenamiento. Señala además, que la cuestión más difícil es que tienen que asumir no solo el ordenamiento de una búsqueda muy precisa de 20 registros como resultado, lograda por una perfecta combinación de términos y frases, sino que también deben ser capaces de ordenar una búsqueda realizada por una sola palabra que arroje millones de registros. Entonces, ¿cómo se ordenan los resultados?.

Muchos buscadores ordenan los resultados por la relación localización/frecuencia, es decir, por la cantidad de veces que aparece la palabra clave, el valor es mayor cuando estas aparecen al principio del texto, en los títulos o en los encabezamientos,27,28 sin embargo, este se considera un método no muy útil para algunos tipos de búsqueda, por ejemplo al buscar la palabra clave "hernia", tal vez uno de los registros más relevantes se refiera al libro Kandy-Kolored Tangerine-Flake Streamline Baby de Tom Wolfe, porque este repite la palabra una docena de veces al principio, cuando realmente no se corresponde con el tema buscado, en estos casos los problemas de sinonimia y polisemia atentan contra los resultados.28

Desafortunadamente, hasta donde ha llegado la evolución de los buscadores, estos no conciben vocabularios controlados para considerar las relaciones de este tipo entre las palabras mediante referencias, 20 es válido mencionar que se han registrado algunos intentos al respecto, como el proyecto Wordnet de la Universidad de Princeton y el proyecto Clever de IBM, pero sin resultados muy alentadores. 8

Otro método, para el ordenamiento de los resultados de las búsquedas en el web, es el análisis de los enlaces, se basa en la relación entre el registro y el resto de estos, con enlaces a él, se supone que los webs de mayor calidad se citan con mayor profusión, así el sistema muestra primero los que más veces aparecen en otras páginas.27 El algoritmo conocido comoGoogleRank, constituye una joya inapreciable, también el proyecto Clever mencionado anteriormente, propone un método sumamente interesante para el análisis de los enlaces.

Verdaderamente, en la función de ordenar los resultados de las búsquedas es tal vez, donde los buscadores logran una menor eficiencia, y en ocasiones no es recomendable comparar los resultados entre uno y otro para determinar cuál los ordena mejor, sencillamente porque no indizan las mismas páginas, algunos indizan más que otros, con menor o mayor frecuencia, por esto, los resultados, aun sin ordenar, nunca serían los mismos.

Consideraciones finales

El análisis de cada uno de los aspectos tratados anteriormente permite realizar algunas consideraciones generales acerca del desarrollo y la utilización de los buscadores en Internet.

El uso de los buscadores es como una cadena que va desde lo más general a lo más específico. Los metabuscadores se utilizan generalmente para temas muy difíciles o cuando no existe idea alguna de los beneficios de uno u otro buscador. Los buscadores se emplean casi siempre en los inicios del uso de Internet, cuando el navegante tiene solo algún conocimiento sobre ellos, bien para buscar temas en los que no se ha investigado con frecuencia o muy particulares que suelen encontrarse en los directorios temáticos. En la medida en que crezca la cultura de Internet y la experiencia de los navegantes, estos dejarán de recurrir a los buscadores en beneficio de la creación de la marca, se identificarán más con los portales que se adecuen a sus necesidades de información sin necesitar intermediarios.

En cualquiera de los casos, el uso de los buscadores es sumamente útil, no es aceptable decir categóricamente que uno es mejor que otro. Para explotar sus potencialidades es necesario estudiarlos y

conocerlos, ello permitirá perfeccionar el uso de sus opciones de búsqueda que, cuanto más precisas, producirán resultados más relevantes.

Referencias bibliográficas

- 1. Stanley T. Search engines corner meta-searching on the Web [texto en línea], Febrero 2001. Disponible en: <URL: http://www.ariadne.ac.uk/issue14/search-engines/>. [Acceso 25 de Diciembre 2001].
- 2. Buscadores: Los puntos de partida en la red. Revista Consumer [publicación en línea], 2001:abr (43).Disponible en: <URL:http://www.consumer-revista.com/web/es/20010401/internet/28193.jsp>. [Acceso 2 de Noviembre 2001].
- 3. Cómo Buscar en Internet? (Explicaciones y tips) [texto en línea]. Disponible en: <URL: http://pp.terra.com.mx/~hugalde/comobusc.html#ini>. [Acceso 9 de Noviembre 2001].
- 4. Buscadores: Pequeños trucos para usar los buscadores. [texto en línea]. Disponible en: <URL:http://personal.redestb.es/necora/buscador/3.htm>. [Acceso 12 de Noviembre 2001].
- 5. Flanagan D. Finding it online: web search strategies [texto en línea] 1999. Disponible en: <URL: http://home.sprintmail.com/~debflanagan/main.html>. [Acceso 25 de Diciembre 2001].
- 6. Principales motores de búsqueda [texto en línea], 1999. Disponible en: <URL: http://www.uco.es/investiga/grupos/rea/search/descripcion.htm>. [Acceso 10 de Diciembre 2001].
- 7. Sullivan D. How Search engines rank web pages [texto en línea],. Disponible en: <URL: http://www.searchenginewatch.com/webmasters/rank.html>. [Acceso 10 de Diciembre 2001].
- 8. Liu J. Guide to meta-search engines [texto en línea], 1999. Disponible en: <URL: http://www.indiana.edu/~librcsd/search/meta.html>. [Acceso 25 de Diciembre 2001].
- 9. Multi search engines: a comparison [texto en línea]. Disponible en: <URL: http:// http://www.philb.com/msengine.htm >. [Acceso 25 de Diciembre 2001].
- 10. Notess G. Toward more comprehensive web searching: single searching versus megasearching. Online [publicación en línea], 1998:22(2). Disponible en: <URL:http:// http://www.onlineinc.com/onlinemag/OL1998/net3.html >. [Acceso 25 de Diciembre 2001].
- 11. Tyner R. Sink or swim: Internet search tools & techniques [texto en línea]. Disponible en: <URL: http://www.sci.ouc.bc.ca/libr/connect96/search.htm>. [Acceso 10 de Diciembre 2001].
- 12. Vernon H. Performance of four World Wide Web (WWW) index services: Infoseek, Lycos, Webcrawler and WWWorm [texto en línea] Junio 1997. Disponible en: <URL: http://www.winona.msus.edu/library/webind.htm >. [Acceso 10 de Diciembre 2001].
- 13. Notess G. On the net: Tips for evaluating web databases. Database [publicación en línea], 1998:21(2). Disponible en: <URL:http://www.onlineinc.com/database/DB1998/net4.html>. [Acceso 25 de Diciembre 2001].
- 14. Wighton D. Searching FAQs [texto en línea] Julio 1997. Disponible en: <URL: http://www.cln.org/searching faqs.html>. [Acceso 25 de Diciembre 2001].
- 15. Barlow L. The Spider's Apprentice: A helpful guide to search engines [texto en línea], March 2, 2001. Disponible en: <URL:http://www.monash.com/spidap.html>. [Acceso 3 de Enero 2002].
- 16. Campbell K. Tips on popular search engines [texto en línea], March 1997. Disponible en: <URL:http://www.hamline.edu/library/bush/handouts/slahandout.html>. [Acceso 20 de Diciembre 2001].
- 17. Cohen L. Quick reference guide to search engine syntax [texto en línea], May 2, 2001. Disponible en: <URL:http://library.albany.edu/internet/syntax.html>. [Acceso 23 de Diciembre 2001].
- 18. Introduction to search engines [texto en línea]. Disponible en: <URL: http://www.kcpl.lib.mo.us-search-srchengines.htm>. [Acceso 3 de Enero 2002].
- 19. Notess G. On the fly search engine analysis. Online [publicación en línea], 1999:23(5). Disponible en: <URL:http://www.onlineinc.com/onlinemag/OL1999/net9.html>. [Acceso 25 de Diciembre 2001].
- 20. Price G. Web search engines FAQS: Questions, answers, and issues [texto en línea], 1999. Disponible en: <URL: http://www.infotoday.com/searcher/oct01/price.htm>. [Acceso 10 de Diciembre 2001].
- 21. Salpeter J. In Search of the perfect search engine [texto en línea], Febrero 2001. Disponible en: <URL: http://techlerning.com/db area/archives/TL/200102/picksofmonth.html >. [Acceso 25 de Diciembre 2001].
- 22. Vernon H, Srivastava J. Precision among World Wide Web search services (search engines): Altavista, Excite, Hotbot, Infoseek, Lycos [texto en línea] Junio 1997. Disponible en: <URL: http://www.winona.msus.edu/library/webind2/webind2.htm>. [Acceso 10 de Diciembre 2001].
- 23. Gresham K. Surfing with a Purpose: Process and strategy put to the test on the Internet. Educom Review [publicación en línea], 1998:33(5). Disponible en:

- <URL:http://www.educause.edu/ir/library/html/erm9851.html >. [Acceso 3 de Enero 2002].
- 24. Lawrence S, Lee C. Tips for searching the Web [texto en línea]. Disponible en: <URL: http://www.neci.nj.nec.com-homepages-lawrence-searchtips>. [Acceso 25 de Diciembre 2001].
- 25. Boolean Searching on the Internet [texto en línea]. Disponible en: <URL:http://library.albany.edu/internet/boolean.html>. [Acceso 3 de Enero 2002].
- 26. Corredor J. Buscadores en Internet, guía para sobrevivir [texto en línea]. Disponible en: <URL: http://www.baquia.com/com/20001212/art00003>. [Acceso 9 de Noviembre 2001].
- 27. Sullivan D. How search engines work [texto en línea]. Disponible en: <URL: http://www.searchenginewatch.com/webmasters/rank.html>. [Acceso 10 de Diciembre 2001].
- 28. Chakrabarti S. Hypersearching the Web [texto en línea] June 1999. Disponible en: <URL: http://www.sciam.com/1999/0699issue/0699raghavan.html >. [Acceso 3 de Enero 2002].
- 29. Buscadores: Qué son y cómo funcionan. [texto en línea]. Disponible en: <URL:http://personal.redestb.es/necora/buscador/1.htm>. [Acceso 12 de Noviembre 2001].
- 30. Cómo buscar inteligentemente en la web? [texto en línea]. Disponible en: <URL: http://www.monografias.com>. [Acceso 9 de Noviembre 2001].
- 31. Janah M. The new world of search engines: Pressure to make money means clutter, confusion for consumers [texto en línea], 1999. Disponible en: <URL: http://www.mercurycenter.com/svtech/news/indepth/docs/search081699.htm>. [Acceso 25 de Diciembre 2001].
- 32. Rakitin J. Search and ye shall find? [texto en línea], 1999. Disponible en: <URL: http://www.nwfusion.com/news/1020search.html >. [Acceso 25 de Diciembre 2001].
- 33. Sherman C. Search engine help: documentation and resources on the web. On line [publicación en línea], 1998:22(6). Disponible en:

 <URL:http://www.onlineinc.com/onlinemag/OL1998/sherman11.html>. [Acceso 25 de Diciembre 2001].
- 34. Sherman C. Super searchers' search secrets [texto en línea] January 1, 1999. Disponible en: <URL: http:// websearch.miningco.com/internet/websearch/library/weekly/ag010899.htm>. [Acceso 3 de Enero 2002].
- 35. Sullivan D. Search engine watch [sitio en Internet]. Disponible en: <URL: http://searchenginewatch.com>. [Acceso 10 de Diciembre 2001].
- 36. Tanaka J. The perfect search. Newsweek [periódico en línea], 27 de Septiembre de 1999. Disponible en: <URL: http://www.newsweek.com/nw-srv/focus/ty/foty01131.htm>. [Acceso 3 de Enero 2002].
- 37. Winship I. World Wide Web searching tools, an evaluation [texto en línea] 1995. Disponible en: <URL: http://bubl.ac.uk/journals/lis/oz/vine/n09995.winship.htm>. [Acceso 25 de Diciembre 2001].

Recibido: 4 de enero del 2003 Aprobado: 24 de febrero del 2003

Ania Torres Pombert

Centro Nacional Coordinador de Ensayos Clínicos

Calle 200 esq. 21, Atabey, Playa. Ciudad de La Habana. Cuba

Correo electrónico: <u>atorresp73@yahoo.com</u>

<u>1 Licenciada en Información Científico-Técnica y Bibliotecología. Jefa del Grupo de Información Científica y Documentación. Departamento Técnico-Organizativo. Centro Nacional Coordinador de Ensayos Clínicos.</u>

© 2004 2000, Editorial Ciencias Médicas

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

