

Aproximación general a la evaluación de la recuperación de información mediante motores de búsqueda en Internet

Francisco Javier Martínez Méndez

Grupo de Tecnologías de la Información. Universidad de Murcia

0.1. Resumen

El aumento del número de motores de búsqueda en la web ha venido acompañado de una importante producción de literatura científica relacionada con la evaluación de estos sistemas, que podemos calificar como dispersa, esporádica y poco uniforme. La dispersión deriva de la ocasional presencia de autores que posteriormente abandonan esta línea de trabajo; la ausencia de periodicidad se encuentra muy vinculada a la característica anterior e incide negativamente en la calidad de estas evaluaciones; y por último, la escasa uniformidad propicia poca normalización en la determinación de los parámetros más apropiados para llevar a cabo estos procesos de evaluación. Es por ello que nos parece interesante estudiar el estado de la cuestión, identificando esas posibles carencias y pergeñando una serie de líneas de actuación que favorezcan la superación de las carencias detectadas en este campo. (Autor)

Palabras clave: World Wide Web. Motores de búsqueda. Sistemas de recuperación de información. Evaluación.

0.2. Abstract

The growing number of web search engines has resulted in an important production of scientific literature related with the evaluation of these systems, which can be defined as disperse, sporadic and very heterogeneous. This dispersion derives from the occasional presence of these authors in this line of research, which they use to leave out before; the absence of periodicity is highly associated with the previous feature and affects negatively to the quality of this these evaluations; finally, the heterogeneous methodology implies a small level of normalization in establishing the set of appropriate for deploying these processes of evaluation. So, it seems important studying the state of the art with the goal of detecting the possible failures and designing proposals to solve them. (Author)

Keywords: World Wide Web. Search engines. Information retrieval systems. Evaluation

1. Introducción

De forma paralela al crecimiento de la información depositada en diversas fuentes dentro de la red Internet, ha aumentado también a un considerable ritmo el número de los motores de búsqueda que aseguran poseer la mayor cantidad de recursos debidamente indexados y accesibles a través de ellos. Al poco tiempo de su aparición y de la popularización de su uso en el seno de los usuarios de la red de redes, éstos se han repetido muchas veces, sin duda alguna, la interrogante: “¿Cuál es el mejor de estos ingenios?”. En una primera instancia, seguramente debido a razones de familiaridad con su uso, los usuarios avezados de la red suelen tener una respuesta preparada para la anterior cuestión, aunque, si se detuvieran a pensar un momento, ¿cuáles son las bases que sustentarían su respuesta?

Con la idea de conferir una mayor certidumbre a esta respuesta ha proliferado la producción científica (especialmente anglosajona) en este campo. Los resultados que arrojan estos estudios ofrecen resultados dispares y dispersos. La disparidad, lógica por otra parte, proviene de la obtención de resultados distintos entre la mayor parte de los estudios, aunque siempre se pueden identificar algunas conclusiones comunes en todos ellos. La dispersión surge cuando el conjunto de motores evaluados es distinto, lo que dificulta la extracción de conclusiones válidas y cuando estos estudios no suelen repetirse con el tiempo, lo que conlleva a que los mismos pueden ofrecer conclusiones válidas sobre una situación determinada, pero resultan incapaces de percibir la evolución de esta situación y por tanto, incapaces de recoger la evolución de esta tecnología, algo verdaderamente imperdonable considerando los continuos avances que se vienen produciendo en este campo.

Profundizando en la disparidad de los estudios realizados, también se detectan grandes diferencias relacionadas con el método o con el alcance de estos estudios. Podemos encontrar una considerable cantidad de trabajos cuya evaluación se limita a los aspectos explícitos (también se podrían denominar como “formales” o “testimoniales”) del motor de búsqueda: amigabilidad de la interfaz, velocidad de respuesta, formatos de presentación, documentación existente, ayuda del sistema —aspectos todos importantes, sin duda alguna, pero que no recogen, en modo alguno, la efectividad de las operaciones de recuperación implementadas. En muchos casos, estos estudios se limitan a una descripción de las características de los motores de búsqueda sin ninguna aportación valorativa. Otro segundo grupo de evaluaciones sí se centran en el estudio de esa efectividad, aunque se limitan preferentemente a estudiar los parámetros de exhaustividad y precisión de la recuperación de información, parámetros que por sí solos tampoco pueden arrojar una serie de conclusiones concluyentes sobre las prestaciones de estos motores de búsqueda y que además cuentan con el problema de unas altas dosis de subjetividad subyacentes— aunque, evidentemente sí arrojan algunos resultados mucho más concluyentes que las evaluaciones formales anteriores.

Por último, existen algunos estudios más profundos que suman otra serie de conceptos a los indicados anteriormente y que, incluso llegan a contemplar el estudio de los algoritmos empleados por estos motores para el alineamiento (o *ranking*) de los documentos a la hora de recuperarla información.

Con este trabajo, pretendemos ofrecer una visión general de cómo se encuentra la investigación en este campo, identificando los parámetros generales que se han venido manejando y comentar su funcionalidad. En este trabajo, por tanto, no pretendemos llevar a cabo ninguna evaluación, sino sólo seleccionar los instrumentos de medida que pudieran ser empleados por otros investigadores y usuarios de la red Internet en general en el desarrollo de sus trabajos de evaluación de los distintos motores de búsqueda existentes y de los que se desarrollarán.

2. Estudios explícitos

Los primeros estudios datan de 1995, aproximadamente un año después de la aparición de los primeros motores de búsqueda. En un estudio que se está convirtiendo en un clásico en este campo, Chu y Rosenthal (Chu, 1996), de la Universidad de Long Island en Nueva York, recopilan parte de ellos y presentan sus conclusiones en la Conferencia Anual de ASIS de 1996. Los estudios recogidos por estos autores se encuentran basados, en la mayoría de los casos, en las características formales del propio ingenio de búsqueda y en las descripciones técnicas que proporciona el sistema. En su trabajo citan a Courtois, Baer y Stark (Courtois, 1995), quienes destacan la potencialidad del motor Webcrawler en todo lo relacionado con su flexibilidad a la hora de plantear las ecuaciones de búsqueda y su rápida respuesta; resaltando además su interfaz, la cual consideran muy adecuada para usuarios poco iniciados. Basándose también en la idea de la flexibilidad de la interfaz, Scoville (1996) apuesta por los motores Excite, Infoseek y Lycos. Chu y Rosenthal citan el trabajo de Kimmel (1996) quien califica a Lycos como el mejor basándose en la documentación aportada por varios motores y en sus características externas. Similar es el estudio de Davis (1996), quien toma en consideración el tamaño de índice del motor y las posibilidades de recuperación de información, decantándose por Alta Vista, Hot Bot e Infoseek, sobre un total de siete motores evaluados.

Lebedev (1997) lleva a cabo en 1996 un sencillo experimento donde introduce ocho palabras clave, relacionadas con la Química y con la Física, en varios motores de búsqueda y suma el total de ocurrencias de las ocho búsquedas. Para él, el número de documentos es el indicador más primario, pero indicador al fin y al cabo, de la calidad de los distintos motores de búsqueda. A continuación sugiere un segundo criterio, el de la incorporación de nuevas páginas al motor. A este factor lo denomina “dinámica de los motores de búsqueda”, concluyendo que, bajo su punto de vista, Alta Vista es el mejor motor de búsqueda, seguido muy de

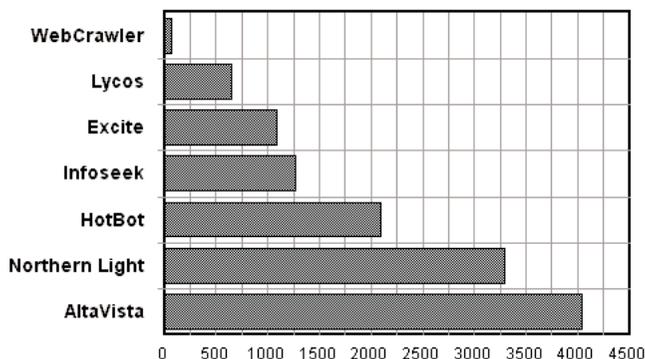


Figura 1. Aciertos totales en quince búsquedas el 29 de agosto de 1998 (Notess, 1999).

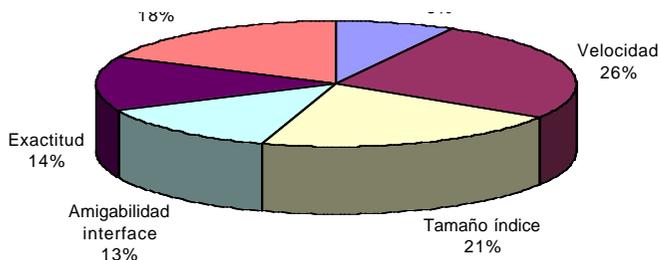


Tabla 1. Evolución del ranking de buscadores (Notess, 1999).

cerca por HotBot. Llama la atención en este experimento el pésimo resultado de Yahoo, Galaxy y Webcrawler.

Notess (1999), antes de recopilar sus estudios en la web Searchengine Showdown llevó a cabo diversas evaluaciones similares a las anteriores. La realizada en agosto de 1998 ofrecía el resultado de la figura 1. En ese mismo estudio, Notess nos indica que había realizado estudios similares a lo largo de diversos períodos de tiempo y que los resultados habían ido variando de forma sustancial, tal como podemos ver en la tabla 1. Similar línea de trabajo prosigue el mismo autor en la web Searchengine Showdown (Notess, 2000a), donde calcula, entre otras estadísticas, los documentos recuperados en distintos motores tras realizarles una serie de veinticinco cuestiones. El resultado obtenido se recoge en la tabla 2.

Otro criterio explícito, más empleado que lo que podríamos imaginar, lo encontramos en la popularidad o “audiencia” del motor de búsqueda, como es el

Motor analizado	Hits
All the Web	4929
Google	4720
Northern Light	3768
IWon Advanced Search	3608
AltaVista	2904

Tabla 2. Aciertos de los distintos motores de búsqueda en octubre de 2000 (Notess, 2000a)

caso del trabajo de Sullivan (2000) titulado “Media Metrix Search Engine Ratings”, donde clasifica los distintos motores según un estudio de audiencia sobre los accesos que más de cien mil usuarios de Internet realizan a los distintos motores de búsqueda, mostrándonos como motor más popular a Yahoo (61.9%) seguido por Microsoft Network (52.3%) y American Online (44.1%). Si bien este estudio viene avalado por su continua interactividad con los usuarios, presenta un planteamiento polémico al basar todas sus conclusiones en la “audiencia” y también introduce distorsión en sus resultados, en tanto que para muchos autores, ninguno de estos tres sistemas destacados es verdaderamente un motor de búsqueda, sino que se tratan más bien de directorios electrónicos de uso muy popular gracias a la amplísima cantidad de recursos que tienen recopilados.

En otro estudio de 1995, Winsip (2000) estudia cuatro motores de búsqueda: World Wide Web Worm, Webcrawler, Lycos y Harvest; junto a dos directorios: Yahoo y Galaxy, centrando su análisis en la porción de página indexada, en la interfaz del sistema, en las capacidades de búsqueda, en los formatos de presentación de los documentos y en el número de documentos recuperados, destacando en primer lugar a Lycos, ligeramente por encima de Harvest.

En 1996, en la Universidad de Sunderland, Stobart y Kerridge (Stobart, 1996) desarrollaron un estudio sobre un conjunto de cuatrocientos usuarios de Internet (profesores e investigadores universitarios en su mayor parte), que cumplimentaron un cuestionario. La primera de las preguntas consistía en indicar cuáles eran los motores de búsqueda que más empleaban, obteniéndose como respuesta Alta Vista, Yahoo y Lycos (por ese orden) de forma preferente. En segundo lugar, se preguntaba a quienes aseguraban usar más de un motor de búsqueda cuál era al que acudían primero, siendo Alta Vista el preferido con una amplísima diferencia. En tercer lugar, se preguntaban las posibles causas de esa fidelidad a un motor frente a otros y la respuesta la encontramos en el gráfico de la figura 2.

Westera (2000) realiza varios análisis de las prestaciones de los distintos motores de búsqueda. Uno de ellos se centra en las capacidades de su interfaz a

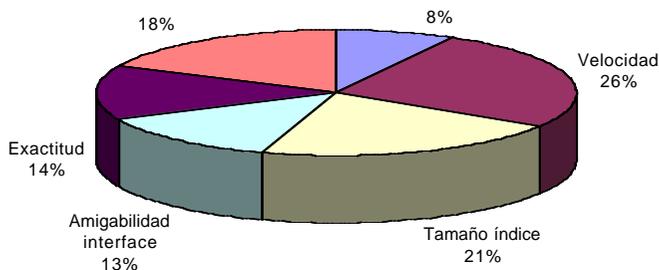


Figura 2. Causas de la fidelidad a un motor de búsqueda

la hora de realizar las operaciones de recuperación de información. Una particularidad interesante de este estudio de Westera es que se actualiza constantemente desde 1996. Divide las capacidades de búsqueda en dos grupos, básicas y especiales. En el primer grupo destacan Alta Vista y Google; en el segundo, Alta Vista y Hot Bot.

Esta amplia disparidad de resultados ofrecidos por esta pequeña muestra de estudios recopilados, viene a mostrar claramente la tendencia de este tipo de análisis hacia la divergencia en sus conclusiones. Gordon y Pathak (1999) consideran que la evaluación procedente de la observación de las características más técnicas de los motores de búsqueda no deja de ser una mera percepción testimonial. En su trabajo, ellos recogen también las conclusiones de algunos de estos estudios pero afirman que “aunque los tests testimoniales pueden proporcionarnos datos útiles para la toma de decisiones sobre qué motor de búsqueda emplear en un determinado momento, únicamente nos pueden proporcionar información indirecta sobre cuál es el más efectivo en la recuperación de información”, por lo cual proponen proceder a otro tipo de estudios más relacionados con esa efectividad.

Sin abandonar aún este grupo de estudios, nos encontramos el desarrollado por Arents (1997) en 1995, donde selecciona distintos motores de búsqueda y los clasifica según una escala de cinco ítems: mejor, muy bueno, bueno, útil. Arents basa sus apreciaciones en la facilidad de uso y en la efectividad, entendida ésta como la cantidad, precisión y legibilidad de los documentos recuperados. Esta clasificación se ha venido actualizando desde entonces y actualmente presenta a los motores Northern Light, Google, Alta Vista y Excite como los mejores.

Chu y Rosenthal comentan también que C|net, empresa especializada en evaluar productos en línea, realizó un estudio comparativo de diecinueve motores de búsqueda considerando el acierto en sus búsquedas, la facilidad de su uso y la cantidad de opciones avanzadas que proporcionaba cada ingenio. En este estudio se destaca a Alta Vista como el mejor de todos.

Zorn, Emanoil y Marshall (1996) llevan a cabo un estudio parecido sobre las capacidades de búsqueda de varios motores, sobre la extensión de sus índices y sobre su capacidad de acierto; concluyendo que resulta difícil clasificar a alguno de los motores como el mejor pero que Alta Vista, Infoseek y Open Text aúnan suficientes méritos para destacarse sobre Lycos, el cuarto motor analizado.

Resulta significativo el hecho de que, mientras el acierto o precisión en la búsqueda no se contemplaba en los estudios, la disparidad de resultados era muy grande, y que cuando comienzan a considerarse este criterio (Stonbart y Kerridge, Arents y C|net, por ejemplo), se empiezan a extraer conclusiones más cercanas unas de otras, debido, sin duda alguna, a que se comienzan a introducir en los estudios una serie de parámetros más objetivos. Esta breve presentación de diversos estudios realizados nos sirve para mostrar el alto grado de diversidad de opiniones detectado y las distintas consideraciones evaluadas por los distintos investigadores. En la tabla 3 se construye una clasificación —a modo de ranking— de las elecciones de estos autores siguiendo un procedimiento compuesto de dos pasos: 1) se asigna un coeficiente de 10 puntos al primero de los motores destacados por el autor; 6 puntos al segundo y 3 puntos a cualquiera de los demás que resalte sobre el grupo analizado. 2) posteriormente, se suman los puntos asignados a cada uno de ellos y se presentan en la tabla 4 a modo de clasificación resumen de los estudios analizados.

Motor	Puntuación	Estudios analizados
Alta Vista	63	7
Lycos	26	4
Google	18	3
Yahoo	16	2
Infoseek	15	3
Hot Bot	15	3
Excite	13	3
Northern Light	13	2
Webcrawler	10	1
All the Web	10	1
Microsoft Network	6	1
American Online	3	1
Harvest	3	1
Open Text	3	1

Tabla 3. Ranking de motores a partir del metanálisis de la literatura

3. Estudios experimentales

Con esta denominación vamos a recoger el conjunto de estudios que han sometido a los distintos motores de búsqueda evaluados a una serie de tests o experimentaciones que van más allá de la meras consideración de las características externas de cada motor. Chu y Rosenthal (1996) comentaban que las diferencias de opinión de los estudios anteriores residía en la ausencia de una metodología clara para la evaluación y se preguntaban: “si podíamos llegar a pensar en desarrollar esa metodología que facilite unos resultados que ayuden al usuario a discernir qué herramienta es la más apropiada para sus necesidades específicas”. En su propuesta, los autores retoman las típicas medidas de evaluación de los sistemas de recuperación de información propuestas por Lancaster y Fayen en 1973 (Lancaster y Fayen, 1973), donde establecen seis criterios:

1. Cobertura
2. Exhaustividad
3. Precisión
4. Tiempo de respuesta
5. Esfuerzo del usuario
6. Formato de presentación de los documentos

Chu y Rosenthal afirman que estos criterios siguen siendo plenamente válidos casi tres décadas después de enunciarse, aunque, por ejemplo, no se puede comparar el esfuerzo del usuario de principios de los setenta para llevar a cabo

Criterio	Nº de veces
Prestaciones en la recuperación de información	6
Amigabilidad de la interfaz	4
Precisión	3
Número de ocurrencias	3
Formato de presentación	2
Documentación	1
Frecuencia de incorporación de nuevos documentos	1
Audiencia	1
Porción de página indexada	1
Legibilidad de los documentos	1
Tamaño del índice	1

Tabla 4. Características valoradas en los distintos estudios de evaluación

una recuperación de información en línea con el esfuerzo de cualquiera de nosotros en la actualidad cuando pretendemos realizar una búsqueda en la web, no admitiendo mucha comparación tampoco los formatos de presentación. Ellos plantean entonces un método de evaluación inspirado en los criterios anteriores y adaptado al contexto del World Wide Web, donde se evalúan cuatro criterios:

1. *Composición de los índices*: Se estudia la cobertura, la frecuencia de actualización y la porción de página indexada.
2. *Capacidades de búsqueda*: Se analizan las prestaciones que los distintos motores poseen a la hora de recuperar información —operadores booleanos, búsquedas por frase literal, truncamiento y refinamiento de búsquedas.
3. *Ejecución de la recuperación de información*: Se emplean tres parámetros, la precisión, la exhaustividad y el tiempo de respuesta, procedentes todos de las propuestas clásicas de Lancaster y Fayen.
4. *Esfuerzo del usuario*: Se valora la documentación y la interfaz del sistema.

Como podemos observar, en la propuesta de Chu y Rosenthal se produce un salto cualitativo frente a las evaluaciones explícitas, aprovechando de aquéllas los parámetros más fiables e incorporando una serie de parámetros que permiten una mejor ponderación, parámetros que, aunque continúan afectados por la subjetividad, sí aportan una información más determinante a la hora de evaluar a los motores de búsqueda analizados. Verdaderamente, el único aspecto negativo de este estudio es que sólo se aplica a tres motores —Alta Vista, Lycos y Excite—, por lo que nos hace falta recurrir a otras experimentaciones para hacernos una idea más completa sobre el estado de la cuestión. El primer criterio no encontró diferencias sustanciales entre los motores. El segundo arrojó unos mejores valores para Alta Vista (en torno al 78% sobre una media del 59%). El tercero se decanta claramente hacia Alta Vista con la detección de fallos imperdonables en Lycos (ausencia de posibilidad de “búsquedas laterales”). Por fin, en el cuarto criterio, Lycos ofrece mejores prestaciones seguido de Alta Vista.

Gordon y Pathak (1999) desarrollan un estudio mucho más amplio que el desarrollado por Chu y Rosenthal, ya que analizan siete motores de búsqueda —Hot Bot, Magellan, Infoseek, Excite, Lycos, Alta Vista y Open Text— junto a un directorio —Yahoo. Estos autores reflexionan sobre la excesiva dependencia que las evaluaciones tradicionales de los sistemas de recuperación de información poseen del concepto de la efectividad medida en torno a los parámetros de precisión y exhaustividad. De hecho indican que:

[...] uno de los planteamientos que más asaltan de forma general a todos los estudios de este tema y a los expertos usuarios de los sistemas de recuperación de información es conseguir que sus estrategias de búsqueda encuentren el término medio más adecuado entre estos dos parámetros. Una exhaustividad muy grande llevará, sin

duda alguna, implícita un amplio número de documentos recuperados lo que implicará, por regla general, que los niveles de precisión no serían aceptables, ya que el total de documentos recuperado sería muy amplio y gran parte de los mismos no satisfarían en lo más mínimo nuestras necesidades de información”.

Estos autores reflexionan también sobre el concepto de “documento relevante”, el cual no parece estar excesivamente claro. Sobre este aspecto ya nos indicaba Blair (1990):

[...] no está claro en un primer término, se nos presenta como un concepto afectado de una gran dosis de subjetividad que puede ser explicado de múltiples maneras por distintas personas y, por tanto, dentro del contexto de una búsqueda en un sistema de recuperación de información este concepto puede ser precisado de muchas maneras distintas por todos aquellos que realicen búsquedas en el mismo. En un segundo lugar, no debe sorprendernos que un usuario afirme que unos determinados documentos son relevantes a sus necesidades de información y que, en cambio, no sea capaz de precisarnos con exactitud qué significa ser relevante para él. Esto no quiere decir que el concepto carezca de importancia, sino que la realización de un juicio de relevancia viene a formar parte de ese amplio conjunto de tareas cotidianas que llevamos a cabo los seres humanos (procesos cognitivos, por tanto), pero que, generalmente, no podemos encontrar las palabras adecuadas para proceder a su descripción”.

Otro aspecto muy importante a considerar es la propia naturaleza intrínseca de los motores de búsqueda, especialmente en todo lo relacionado con las fuentes de información a las que acceden. Estos ingenios indexan páginas (bien por su cuenta o bien a sugerencia de sus usuarios) de temática muy heterogénea y sin establecer restricciones muy amplias sobre el contenido y el formato de esos documentos, excepto si su contenido es presuntamente delictivo. Esto interesa en tanto que no podremos extrapolar los tests desarrollados sobre los sistemas de recuperación de información; sistemas cuyas bases de conocimiento se encuentran mucho más estructuradas y cuya naturaleza informativa es mucho más homogénea. Efectivamente, la web constituye un contexto mucho más abierto, tanto en estructura de los documentos como en la naturaleza de los contenidos.

Por último, Gordon y Pathak nos recuerdan que, a la hora de expresar la necesidad de información del usuario en los términos empleados por el motor de búsqueda, cada uno de ellos introduce una serie de limitaciones; y que, además, hay que considerar que cada motor de búsqueda emplea un algoritmo distinto, que la base de conocimientos gestionada por cada uno de ellos es diferente y que, incluso si varios motores de búsqueda llevaran a cabo la misma operación de recuperación de información sobre el mismo conjunto de documentos, el resultado será diferente, pues cada motor de búsqueda usa un algoritmo de alineamiento distinto.

Por todo lo anterior, Gordon y Pathak afirman que hace falta considerar nuevos parámetros de evaluación de la recuperación de información que complementen a los tradicionales. En concreto, estos autores proponen un total de siete:

1. Hay que tener en cuenta que los experimentos que puedan llevarse a cabo sobre estos motores de búsqueda, suelen realizarse por medio de usuarios experimentados en la recuperación de información, quienes, en virtud de su conocimiento de los sistemas que van a emplear, pueden introducir algunas modificaciones sobre las necesidades de información originales, e incluso pueden llegar a favorecer a un motor frente a otro en virtud de la costumbre de uso del mismo. Este sesgo no debe obviarse en nuestras consideraciones, teniendo en cuenta además, que la amplia diversidad producida de la “frescura” derivada de las interrogaciones que pueden llevar a cabo la multitud de usuarios de estos sistemas quedaría muy atenuada al restringir los ensayos a un grupo de expertos.
2. Si en una recuperación de información una persona busca documentos que alguien más ha encontrado previamente, la necesidad de información de esa persona debe ser capturada en su contexto inmediato en la medida de lo posible, sin que incidan búsquedas anteriores.
3. Un número suficientemente grande de búsquedas debe ser diseñado para producir evaluaciones satisfactorias. Es decir, van a encontrarse inducidas al éxito de antemano.
4. Los estudios realizados con detenimiento han de incluir más y mejores motores de búsqueda.
5. La efectividad de los motores de búsqueda debe ser analizada explotando las características especiales de cada motor de búsqueda.
6. Los juicios de relevancia deben realizarse por los individuos que precisan la información. De otra manera, si los experimentadores deciden por su cuenta qué páginas web son relevantes o no, se introducirían muchas disfunciones debidas a la familiaridad con la materia o al desconocimiento exacto de las necesidades de información del usuario, de las motivaciones que provocan la necesidad de información y de otra serie de detalles más o menos subjetivos que escapan a la percepción del localizador de información. No se puede dejar de enfatizar suficientemente la importancia de los juicios de relevancia realizados por quienes verdaderamente necesitan la información. En realidad, propiamente hablando los documentos recuperados no son relevantes o dejan de serlo; no se trata de una decisión binaria, pues los contenidos de los documentos pueden coincidir en mayor o menor parte con las necesidades de información. Lo que sí podemos determinar es si son o no relevantes para una determinada persona. Desde un punto de vista pragmático, el mismo documento puede significar varias cosas para personas diferentes; los juicios de relevancia pueden sólo realizar evaluaciones semánticas o incluso sintácti-

cas de documentos o preguntas. Pero estos juicios fallan al involucrar a los usuarios particulares, y también al identificar por qué el usuario realmente encuentra un documento particularmente relevante. Gordon y Pathak bromean un poco al respecto, parafraseando de forma ligeramente modificada la vieja frase: “la relevancia reside en los detalles”.

7. Finalmente, redundando en los criterios anteriores, los experimentos bien dirigidos resultan necesarios para obtener unas medidas significativas de la eficacia de las búsquedas. Esto exige: 1) seguir un diseño experimental apropiado —por ejemplo, convirtiendo en aleatorio el orden en el que los documentos se presentan a los evaluadores para superar los problemas vinculados al orden de presentación—; 2) adoptar las medidas tradicionalmente aceptadas de la recuperación de información (como la precisión y la exhaustividad), para permitir que los resultados puedan evaluarse dentro de un contexto familiar; y 3) emplear tests estadísticos para medir las diferencias detectadas entre los distintos motores de búsqueda.

En su experimento, Gordon y Pathak contaron con la ayuda de treinta y seis profesores familiarizados con el uso de los distintos motores de búsqueda evaluados. Estos profesores analizaron las distintas necesidades de información sobre Economía y Nuevas Tecnologías que les transmitieron diversos usuarios por medio de diversos cuestionarios y efectuaron sus búsquedas en todos los sistemas, indicando el éxito o fracaso de los documentos recuperados entre las veinte primeras respuestas que proporcionaba el motor de búsqueda. El estudio posterior de Gordon y Pathak se centra en estudiar el comportamiento de los parámetros de efectividad en cada cuarteto de cinco documentos recuperados y ofrece un importante número de conclusiones, entre las que destacamos:

1. La efectividad absoluta en la recuperación de información es baja, alrededor del diez por ciento en los primeros diez documentos recuperados.
2. Se detectan diferencias estadísticas considerables entre la precisión de los motores de búsqueda analizados. Alta Vista, Open Text y Lycos resultan los mejores, y Yahoo ocupa el último lugar. Otra cuestión importante a destacar es la amplia distancia entre el primero, alrededor del 40%, y el segundo, sobre el 20%.
3. En el caso de la exhaustividad, no se detectan diferencias sustanciales entre los motores.
4. La efectividad de las búsquedas aparece más fuertemente vinculada al algoritmo de localización que a los tipos de búsqueda permitidos
5. Aunque los motores siguen en su funcionamiento los principios generales de la recuperación de información, su funcionamiento varía, ya que es clara la tendencia a recuperar documentos irrelevantes o lejanamente

relevantes, hecho difícilmente aceptable en el caso de un sistema de recuperación de información tradicional.

Gordon y Pathak también nos aportan otras conclusiones no menos importantes, una de ellas se refiere a los tamaños notablemente diferentes de los índices de los motores de búsqueda (algunos son diez veces más grandes). Además, aunque alguno en su publicidad suela decir lo contrario, ninguno pretende indexar toda la Web. Los motores también difieren en la actualización periódica de los datos, en la posibilidad de añadir los usuarios páginas por su cuenta, en el plazo de tiempo que se tarda en incorporar una nueva página indexada tras tener noticia de su existencia y en el seguimiento de la disponibilidad de los enlaces.

Un tercer estudio a destacar dentro de este grupo es el realizado por Leighton y Srivastava en 1999, en el que compararon la precisión de los primeros veinte resultados devueltos por cinco motores de búsqueda: Alta Vista, Excite, Hot Bot, Infoseek y Lycos a quince preguntas. En este caso, los autores diseñaron una función de evaluación que confería un peso específico a la capacidad de los motores de búsqueda de colocar documentos relevantes entre los primeros veinte mostrados al usuario. Es decir, una función que mide la precisión y el acierto en la tarea de mostrar los documentos relevantes antes que los que no lo son. Además, en este estudio se penaliza la existencia de enlaces inactivos, de manera que aquellas páginas que no hayan sido actualizadas hace bastante tiempo influyen menos en los resultados. Destacaron en precisión Alta Vista, Excite e Infoseek (por este orden), mientras que en su capacidad de alineamiento (o ranking) se apreciaron pocas diferencias aunque cambió el orden (Infoseek, Excite y Alta Vista).

En esta misma línea encontramos un completo estudio de Ming (2000), en el que se contempla el análisis de diversos parámetros tales como la precisión, el tiempo de respuesta, la interfaz y el número de aciertos. Además, introduce un factor que denomina “sensibilidad”, factor sobre el cual incide la calidad de los enlaces devueltos por el motor de búsqueda. Su estudio sólo afecta a Yahoo, Alta Vista y Lycos, y ofrece como resultados más destacados el hecho de no encontrar diferencias significativas en la precisión entre los tres ingenios evaluados (en los primeros diez documentos), su preferencia por la interfaz y el tiempo de respuesta de Yahoo (aunque reconoce que es una opinión subjetiva), la superioridad de Yahoo en el parámetro “sensibilidad”.

Westera (2000) también ha estudiado la precisión de varios motores de búsqueda, destacando a Alta Vista, Lycos e Infoseek en su estudio, donde también ha introducido juicios sobre la calidad de los enlaces suministrados y sobre el crecimiento de los índices. Wishard (2000) lleva a cabo un estudio de la precisión de distintos motores a los que pregunta cuestiones relacionadas con la Geología y no encuentra diferencias significativas entre todos ellos.

Ljosland (2000a), realiza otro interesante y completo estudio para calcular la precisión en veinte motores de búsqueda sobre los que realiza diez búsquedas simples utilizando términos de poco uso (él los denomina “raros”, como es el caso de, por ejemplo, la palabra “haliography” o la palabra “peleidou”). En sus resultados indica que los motores que más documentos recuperaron fueron Fast, AskJeeves y Northern Light. En cobertura destaca a AskJeeves, Fast, Excite y Northern Light. En precisión Euroseek y WebCrawler obtienen el máximo valor (100%) y Lycos obtiene un promedio del 95%. Por último, en lo relacionado con la exhaustividad, el orden es AskJeeves, Fast y Excite.

Este mismo autor lleva a cabo posteriormente otro estudio comparando únicamente tres motores (Ljosland 2000b) —Alta Vista, Google y AlltheWeb—, siendo entonces el primero un motor de consolidada posición y reconocido prestigio y los otros dos nuevos proyectos que buscaban hacerse un hueco dentro de este amplio conjunto. Ljosland introduce una variante sobre estudios anteriores al considerar la posibilidad de que un documento sea “relevante parcialmente”, en lugar del binomio tradicionalmente empleado de “relevante/no relevante”. El experimento arrojó como resultado que, cuando no se consideró la relevancia parcial, se obtiene un valor de 0,4 para Alta Vista, 0,7 para Google y de 0,4 para AllTheWeb, cuando se consideró la relevancia parcial suben un poco todos los valores anteriores: 0,5 para Alta Vista, 0,9 para Google y 0,5 para AllTheWeb. Por tanto, Google se presentó como el mejor motor de los tres analizados.

Por último, Nicholson (2000) en un reciente estudio publicado en JASIS, evalúa la precisión de distintos motores y estudia el mantenimiento de esa preci-

Motor	Puntuación	Estudios
Alta Vista	49	7
Excite	28	5
Infoseek	22	4
Lycos	21	4
Google	10	1
Yahoo	10	1
Webcrawler	10	1
Euroseek	10	1
Fast	6	1
Open Text	6	1
All the Web	3	1

Tabla 5. Clasificación de los motores de búsqueda apartir de los estudios analizados

sión a lo largo del tiempo durante diez semanas. En sus conclusiones destaca a Excite, Infoseek y Alta Vista sobre Lycos.

En la tabla 5, se presenta —a modo de resumen— una clasificación de los motores de búsqueda a partir de los estudios experimentales analizados. Se ha obtenido asignando un coeficiente de 10 puntos al primero de los motores destacados por el autor; 6 puntos al segundo de los destacados y 3 puntos a cualquiera de los demás que resalte sobre el grupo analizado; y sumando posteriormente los puntos asignados a cada uno de ellos.

Este tipo de estudios, mucho más rigurosos que los incluidos en el apartado anterior, nos proporcionan una interesante cantidad de datos, pero, al mismo tiempo, resultan muy laboriosos, por lo que se realizan en menor número y presentan, por tanto, un riesgo alto de obsolescencia al evolucionar con tanta rapidez los motores de búsqueda en la web. También presentan el inconveniente de abarcar, por lo general, pocos motores de búsqueda.

4. Estudios divulgativos

Dentro de este tercer grupo, queremos recoger el amplio conjunto de evaluaciones realizadas por revistas divulgativas especializadas en Informática y que poseen una gran aceptación por parte de los usuarios de Internet. Si bien estas evaluaciones pueden no poseer el rigor científico de las anteriores, creemos que poseen dos factores importantes a tener en cuenta: en primer lugar, el contacto directo con los usuarios, con las ventajas inherentes que se desprenden de esa cercanía; y, en segundo lugar, la actualidad de sus estudios, que suelen repetirse cada cierto tiempo y permiten recoger en ellos la presencia de nuevos motores que se vienen desarrollando —en este punto no debemos olvidar que gran cantidad de los estudios anteriores carecen de evaluaciones de Google, uno de los ingenios actualmente más empleados y mejor considerados por los usuarios. A este tipo de estudios lo denominamos como “divulgativos” en tanto que no vienen avalados por una elaborada experimentación científica previa, aunque no por ello su valor va a ser menor que cualquiera de los estudios englobados en los dos apartados anteriores.

Con anterioridad hemos mencionado a C|net (2000a), empresa que viene estableciendo rankings entre los distintos motores. En su estudio de julio de 1999 establecía el siguiente orden de preferencias: Hot Bot, Alta Vista, Excite, Infoseek y Lycos. En cambio, en junio de 2000 modifica sustancialmente sus resultados al establecer el siguiente orden (C|net, 2000b): Google, Yahoo, Microsoft Network, Alta Vista, Lycos y Netscape Netcenter.

En la web del Harlingen College del Instituto Técnico de Texas (Texas State Technical College, 2000) se nos ofrece una serie de sugerencias sobre el motor más apropiado a elegir dependiendo del tipo de búsqueda que deseemos realizar, aun-

que se decanta por Alta Vista como el mejor para labores de investigación inicial gracias a su alto número de documentos indexados. En la misma línea encontramos a Sherman (2000), quien mantiene en websearch.about.com una serie de interesantes documentos relacionados con los motores de búsqueda. Entre ellos destaca la guía *How to choose the best general-purpose search tool* donde nos indica qué motor es, según sus criterios y estudios previos, el mejor para distintas situaciones de búsqueda. En la tabla 6 recogemos un resumen de la misma

5. Otros factores a tener en cuenta

5.1 Porcentaje de enlaces inactivos.

Uno de los factores más importantes a tener en cuenta para evaluar los distintos motores de búsqueda es la proporción de enlaces inactivos que proporcionan al usuario. En el apartado segundo hemos recogido algunos considerandos sobre este tema por parte de algunos autores, pero resulta difícil encontrar estudios que los introduzcan, situación curiosa más si cabe cuando los porcentajes resultan bastante llamativos en algunos casos. Por ejemplo, Alta Vista llega al 13,7% en el estudio de la web Searchengine Showdown de febrero de 2000

Si deseamos encontrar...	Motor recomendado
Una amplia revisión sobre una materia determinada	Looksmart, que posee más de 24000 categorías para elegir
Una materia determinada usando una frase corta de dos o tres palabras	Infoseek, que reconoce frases y no ignora las palabras más comunes, que otros motores se olvidan
La respuesta a una pregunta simple en idioma inglés	AskJeeves, que posee respuestas directas para más de 7 millones de preguntas
Un tema empleando palabras raras o desconocidas	AltaVista, su tamaño le avala para esta tarea
Los sitios más populares en Internet	HotBot
Páginas en un idioma determinado	Alta Vista o HotBot
Páginas publicadas en una fecha determinada	Alta Vista o HotBot
Un conjunto de documentos y luego refinar la búsqueda	Infoseek
Documentos sobre ciudades, compañías o equipos deportivos	Excite
Sitios web que han pagado para aparecer en lugares preferentes de las búsquedas	GoTo.com
Un directorio de enlaces de calidad	About.com

Tabla 6. Utilidad de los motores de búsqueda para diferentes tareas

(Notess, 200a), Excite alcanza el 8,7% y Northern Light el 5,7%, porcentajes quizás, excesivamente altos. Otro aspecto importante que no debemos olvidar es la percepción negativa que un usuario de un sitio web puede alcanzar si en el mismo proliferan los enlaces negativos.

5.2 Frecuencia de actualización del índice

Muy vinculado a este factor está la frecuencia de actualización del índice de cada motor de recuperación, valor que tiende a ser grande a medida que crece el tamaño del mismo. Ese aumento también puede ser un factor a considerar, aunque no tan determinante como los anteriores, y el mismo se refleja muy claramente en la figura 3 que hemos tomado de la web anteriormente citada y que corresponde a un estudio de noviembre de 2000 (Notess, 2000b). Este estudio presenta a Northern Light como el motor de ritmo de crecimiento más consistente y a Google como el de crecimiento más explosivo, recogiendo también comentarios relacionados con un lento aumento del tamaño de Alta Vista.

5.3 Solapamiento y referencia única

Otros factores considerados en algunos análisis de este tipo son el solapamiento entre los motores —coincidencias de documentos recogidos por los distintos motores— y la referencia única —es decir, cuándo un documento aparece recogido exclusivamente por un motor de búsqueda.

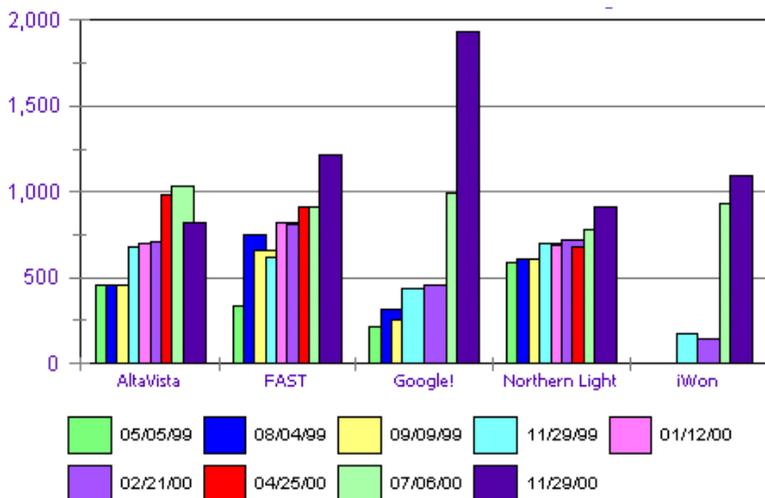


Figura 3. Evolución de las bases de datos de algunos motores de búsqueda (Notess, 2000b)

Ambos factores vienen gozando de una cierta consideración aunque se nos presenta difícil extrapolar algún tipo de conclusión de ambos. En el primer caso, porque un alto grado de solapamiento no implica necesariamente nada negativo sino más bien que un amplio conjunto de fuentes de información forman parte de los índices de los distintos motores de búsqueda. En el segundo caso, el hecho de que un recurso sólo aparezca recogido por un solo motor puede deberse a que el procedimiento de recopilación de recursos del motor es muy exhaustivo y alcanza muchas fuentes de información a las que otros no han llegado, o bien puede deberse a que un usuario de Internet haya confeccionado una página y sólo la dé de alta en un motor determinado, en cuyo caso dicha página puede estar durante algún tiempo recogida sólo en ese motor. Sin embargo, no creemos que ello confiera ningún valor añadido al motor de búsqueda. Tampoco podemos olvidar la posibilidad de que dicha exclusividad pueda ser producto de una casualidad.

Tomando datos de Searchengine Showdown, el 76% de los recursos pueden recuperarse al menos en tres motores de búsqueda. En el mismo estudio se nos proporciona una tabla con el número de referencias únicas localizadas en cada motor de búsqueda (Notess, c), que reproducimos en la tabla 7.

De dicha tabla podemos deducir que la relación mayor tamaño de índice - mejores resultados en factor referencia única debe comprobarse más veces, en tanto que Excite obtiene en este estudio mejores valores que Alta Vista, Northern Light y Google, poseyendo, en cambio, un índice más pequeño.

6. Conclusiones

Como principal conclusión de este estudio recopilatorio aportamos la certeza de que no existe una metodología de análisis de los motores de búsqueda en la web que se imponga claramente sobre las demás. Es decir, no se ha llegado a establecer un método de análisis estándar de estos ingenios, ya que los estudios realizados presentan sustanciales diferencias en su método y, por tanto, sus con-

Motor	% Referencias únicas
Fast	27
Excite	23
Alta Vista	18
Lycos	13
Northern Light	8
Google	7
Microsoft Network	4

Tabla 7. Porcentaje de referencias únicas en varios motores de búsqueda (Notess2000d)

clusiones también resultan dispares. También es importante recordar que en algunos de los trabajos analizados ni siquiera se lleva a cabo un proceso de análisis mínimamente riguroso, sino que las recomendaciones de esos trabajos se basan en criterios excesivamente subjetivos. Es por ello que, siendo rigurosos, no podríamos contestar a la respuesta planteada en la introducción de este trabajo con total certeza.

En segundo lugar, cabe resaltar la dificultad de encontrar estudios amplios y detallados, con base en experimentaciones científicas en este campo, de hecho, la mayoría de los trabajos (algunos muy citados, por cierto), simplemente analizan tres motores de búsqueda, con lo cual el análisis no llega a representar exactamente la realidad.

En esta misma línea, también llama la atención el hecho de que, en la mayor parte de los casos, los autores analizados no repitan sus experimentaciones al cabo del tiempo, sino que suelen abandonar esta línea de trabajo, no contribuyendo, por tanto, a profundizar en ella. Muy relacionado con esta presencia esporádica de investigadores está el problema de la obsolescencia de los informes (especialmente de aquellos publicados en revistas científicas), los cuales difícilmente recogen las novedades que se vienen produciendo de forma paulatina y dejan fuera de su análisis a algunos motores de búsqueda en pleno auge.

Por último, concluimos que se hace preciso desarrollar metodologías de análisis de la calidad de estos motores de búsqueda, rigurosas y alejadas de cualquier tipo de subjetividad, métodos; pero que, por otra parte, tampoco resulten excesivamente complicados y permitan la actualización periódica de los estudios con mucha frecuencia y que, al mismo tiempo abarquen al mayor número de motores posible, incorporando las novedades que se vienen produciendo con gran profusión en este campo.

7. Bibliografía

- Arents, H. (2000). A rated selection of Internet search tools [En línea]. Gent: University 1997. URL: <<http://www.lib.rug.ac.be/internet/search.html>>. Consulta: 29 diciembre 2000.
- Blair, D.C. (1990). Language and representation in information retrieval. Amsterdam [etc.]: Elsevier Science Publishers, 1990. ISBN 0-444-88437-8
- C|Net (2000a). CNET's ultimate guide to search: The top five search sites [En línea]. San Francisco: Cnet.com, 2000. URL: <<http://www.cnet.com/internet/0-3817-7-1922934.html?tag=st.int.3817-7-1922932.txt.3817-7-1922934>>. Consulta: 29 de diciembre de 2000.
- C|Net (2000b). Search engine CNET compares the top 5 engines [En línea]. San Francisco: Cnet.com, 2000. URL: <<http://www.cnet.com/internet/0-3817-7-276910.html?tag=st.cn.sr.inet.1>>. Consulta: 29 de diciembre de 2000.

- Courtois, M.P. ; Baer, W. M. ; Stark, M. (1995). Cool tools for searching the Web: A performance evaluation. // *Online*, 19 : 6 (1995) 14-32,.
- Chu, H. ; Rosenthal, M (1996). Search Engines for the World Wide Web: A Comparative Study and Evaluation Methodology [En línea]. // *Asis 1996 Annual Conference Proceedings*, 1996. URL: <<http://www.asis.org/annual-96/ElectronicProceedings/chu.html>>. Consulta: 28 de diciembre 2000.
- Davis, E. A. (1996). Comparison of Seven Search Engines [En línea]. Kent : University, 1996. URL: <<http://www.iwaynet.net/~lsci/Search/paper.htm>>. Consulta: 29 de diciembre de 2000.
- Gordon, M. ; Pathak, P. (1999). Finding information on the World Wide Web: the retrieval effectiveness of search engines. // *Information Processing and Management*. 35 (1999) 141-180
- Kimmel, S. (1996). Robot generated databases on the world wide web. // *Database*. 19 : 1 (1996) 41-49.
- Lancaster, F.W. ; Fayen, E. G. (1973). *Information Retrieval on-line*. Los Angeles, CA: Melville Publishing Co, 1973.
- Lebedev, A. (1997). Best search engines for finding scientific information in the Web. [En línea] Moscow : State University, 1997. URL: <<http://www.chem.msu.su/eng/comparison.html>>. Consulta: 10 de noviembre de 2000.
- Leighton, V. ; Srivastava, J. (1999). First 20 Precision among World Wide Web Search Services (Search Engines). // *Journal of the American Society for Information Science* 50 (July 19, 1999) 870-881.
- Ljosland, M. (2000a). Comparison between two and one well established Web search engines [En línea]. Trondheim: Dept. of Computer and Information Science, Norwegian University of Science and Technology, 2000. URL: <<http://www.idi.ntnu.no/~mildrid/paper/resultat.html>>. Consulta: 10 de noviembre de 2000.
- Ljosland, M. A. (2000b). comparison between twenty Web search engines on ten rare words [En línea]. Trondheim: Dept. of Computer and Information Science, Norwegian University of Science and Technology, 2000. URL: <<http://www.idi.ntnu.no/~mildrid/paper/Comp20.html>>. Consulta: 10 de noviembre de 2000.
- Ming, H. (2000). Comparison of Three Search Engines [En línea]. Toronto: University, 2000. URL: <<http://gypsy.rose.utoronto.ca/people/ming/report.html>>. Consulta: 29 de diciembre de 2000.
- Nicholson, S.(2000). Raising reliability of web search tool research through Replication and Chaos Theory. // *Journal of the American Society for Information Science* 51 : 8 (2000) 724-729.
- Notess, G. R. (1999). Comparing Internet search engines [En línea]. Charles Sturt University, 1999. URL: <<http://www.csu.edu.au/special/online99/proceedings99/103a.htm>>. Consulta: 28 diciembre 2000.
- Notess, G. R. (2000a). Search Engine Statistics: Relative Size Showdown [En línea]. Bozeman, MT: Notes.com, 2000. URL: <<http://www.searchengineshowdown.com/stats/size.shtml>>. Consulta: 28 diciembre 2000.

- Notess, G. R. (2000b). Search Engines Statistics: Database Change Over Time [En línea]. Bozeman, MT: Notes.com, 2000. URL: <<http://www.searchengineshowdown.com/stats/change.shtml>>. Consulta: 10 de noviembre de 2000.
- Notess, G. R.(2000c). Search Engines Statistics: Dead Links Report [En línea]. Bozeman, MT: Notes.com, 2000. URL: <<http://www.searchengineshowdown.com/stats/dead.shtml>>. Consulta: 14 de noviembre de 2000.
- Notess, G. R. (2000d). Search Engines Statistics: Unique Hits Report [En línea]. Bozeman, MT: Notes.com, 2000 URL: <<http://www.searchengineshowdown.com/stats/unique.shtml>>. Consulta: 10 de noviembre de 2000.
- Scoville, R. (1996). Special report: Find it on the net!. // PC World. 14 : 1 (1996). p. 125.
- Sherman, C (2000). How To Choose The Best General-purpose Search Tool [En línea]. New York: About.com Inc. 2000. URL: <http://websearch.about.com/internet/web-search/library/bl_981204a.htm?terms=choose+the+best>. Consulta: 29 de diciembre de 2000.
- Stobart, S. ; Kerridge, S. (1996). WWW search engine study [En línea]. Sunderland: University, 1996. URL: <<http://osiris.sunderland.ac.uk/sst/se/>>. Consulta: 10 de noviembre de 2000.
- Sullivan , D. (2000). Media Metrix Search Engine Ratings [En línea]. SearchEngineWatch.com: 2000 URL: <<http://searchenginewatch.com/reports/mediamatrix.html>>. Consulta: 29 diciembre 2000.
- Texas State Technical College (2000). Using Internet Search Engines [En línea]. Harlingen, Texas, 2000. URL: <<http://www.harlingen.tstc.edu/tstclibrary/metarsch.htm>>. Consulta: 3 de enero de 2001.
- Westera, G. (2000). Comparison of Search Engine User Interface Capabilities [En línea]. Curtin: University of Technology, 2000. URL: <<http://lisweb.curtin.edu.au/staff/gwpersonal/compare.html>>. Consulta: 1 de noviembre de 2000.
- Winsip, I. (1995). World Wide Web searching tools - an evaluation. // VINE. 99 (1995). p.49-54 También disponible en URL: <<http://bubl.ac.uk/journals/lis/oz/vine/n09995/winship.htm>>. Consulta: 5 de noviembre de 2000.
- Wishard, L. (1998). Precision Among Internet Search Engines: An Earth Sciences Case Study. [En línea]. Pennsylvania: State University, 1998. URL: <<http://www.library.ucsb.edu/istl/98-spring/article5.html>>. Consulta: 29 de diciembre de 2000]
- Zorn, P. ; Emanoil, M., Marshall, L. ; Panek M. (1996). Advanced Web searching: tricks of the trade. // Online. 20 : 3 (1996).