

Reseñas

Análisis comparativo de buscadores en internet

Por Ángeles Maldonado Martínez y Elena Fernández Sánchez

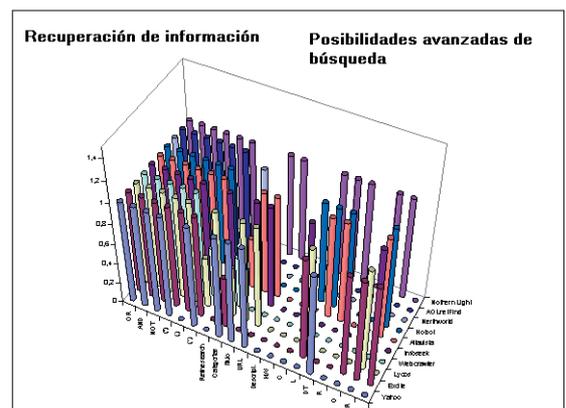
Resumen de la comunicación presentada por las autoras a la conferencia Online Information'99, Londres, 7-9 de diciembre de 1999.

1. Introducción

Internet es ya la mayor fuente de información existente y los índices temáticos y motores de búsqueda son las herramientas más importantes para seleccionarla. Unos y otros son bases de datos documentales que recogen, almacenan y gestionan documentos html.



Ángeles Maldonado y Elena Fernández, Cindoc



3.2. Recuperación de información

Respecto a las características de recuperación de información, en primer lugar se han tenido en cuenta las herramientas básicas que poseen todas las bases de datos documentales convencionales: operadores booleanos (Y, O y NO), truncado de palabras, uso de paréntesis y localización de términos compuestos. En una segunda fase se han evaluado las posibilidades de búsqueda avanzadas siguientes:

- buscar dentro de un conjunto,
- ordenar los registros temáticamente,
- realizar una búsqueda por campos,
- visualizar los índices de cada uno de los campos existentes, y finalmente,
- contar con herramientas de control de vocabulario.

Mención especial requiere la búsqueda por campos. En este trabajo se han considerado como necesarios: título, URL, descripción, palabras clave, país, idioma, fecha, tipo de fuente y tipo de organización responsable. Todos ellos están generalmente presentes en las bases de datos convencionales, salvo URL y tipo de organización, de especial interés para las páginas web.

En este apartado se ha establecido un sistema de puntuación para poder comparar los buscadores entre sí, que consiste en adjudicar un punto por variable operativa disponible, contemplándose sólo dos excepciones: 0,5 puntos en el caso de que el truncado sea obligatorio e igualmente 0,5 puntos para los buscadores que sólo clasifiquen una parte de las páginas existentes en su sistema. La máxima puntuación que puede ser alcanzada en total, herramientas básicas y posibilidades de búsqueda avanzada, es de 19 puntos.

Resultados

Prácticamente todos los buscadores tienen las herramientas de búsqueda básicas.

Sin embargo, en las posibilidades de **búsqueda avanzada** se detectan notables carencias. Sólo *Lycos* e *Infoseek* permiten realizar una búsqueda en un conjunto previo. Los únicos buscadores que tienen clasificadas temáticamente todas sus páginas son los índices temáticos *Yahoo!* y *Nerdworld*. Los **campos de búsqueda** con mayor presencia en los buscadores son: título, URL e idioma, apareciendo cada uno en cinco de ellos. *Northern Light* es el que más opciones de búsqueda por campos ofrece, ya que permite buscar por título, URL, país, idioma, fecha y tipo de organización; a continuación se sitúa *Altavista* que permite hacerlo por título, URL, idioma y fecha; *Yahoo!*, *Lycos*, *Infoseek* y *Hotbot* permiten interrogar en tres campos diferentes; *Excite* tan sólo ofrece hacerlo en dos; y finalmente *Webcrawler*, *Nerdworld* y *AOLNetfind* no contemplan la esa posibilidad por campos.

En cuanto a la presencia de **índices** únicamente se puede visualizar el de idioma en *Excite*, *Lycos*, *Altavista*, *Hotbot* y *Northern Light*, y el de país en *Infoseek*, *Northern Light* y *Hotbot*. Es de resaltar que *Hotbot*, *Infoseek* y *Altavista*, después de lanzar una búsqueda utilizando una palabra, son capaces de presentar aquellos términos compuestos que contienen dicha palabra, por si se quiere utilizar alguno de ellos.

Finalmente, con respecto al **vocabulario controlado** tan sólo *Excite* dice, pues no se ha podido comprobar, poder realizar una búsqueda conceptual; o sea, localizar páginas donde aparezca el concepto solicitado, aunque no esté representado por la palabra utilizada en la búsqueda.

Una vez comprobado el nivel de cumplimiento de las 19 variables seleccionadas, herramientas básicas y posibilidades de búsqueda avanzadas, la puntuación obtenida por cada uno de los buscadores estudiados puede observarse en el cuadro adjunto.

4. Conclusiones

1. No existe relación entre nivel de popularidad y capacidades de búsqueda. Por ejemplo, mientras que *Northern Light* es el que tiene menos número de páginas enlazadas, en cambio es el que ofrece mayores posibilidades en la búsqueda. Por el contrario, *Yahoo!*, el más enlazado de todos con diferencia, ocupa un puesto medio en la clasificación según las opciones de recuperación de información.

2. Desde el punto de vista de la recuperación de información, hay importantes diferencias entre buscadores. Compárense *Northern Light*, el de puntuación más alta con 13 puntos, y *Nerdworld*, el de puntuación más baja, que tan sólo obtiene 2 puntos.

3. También desde el punto de vista de la recuperación se puede decir que los buscadores ofrecen menos posibilidades que las bases de datos documentales convencionales. Esta afirmación se basa en los siguientes hechos:

a. Algunos buscadores no ofrecen todavía todas las herramientas básicas de búsqueda.

b. La búsqueda por campos no es muy frecuente entre los buscadores, no habiendo diferencias notables entre índices temáticos y motores de búsqueda, a pesar de que en principio los primeros están en mejores condiciones de ofrecer mayores posibilidades, pues desde sus formularios de alta se puede exigir sistemáticamente diversos datos sobre cada página incorporada.



¿Recibes ya EPI en casa?

Por sólo 9.900 pesetas + 4% IVA puedes tener todos los meses tu copia particular y leerla cómodamente cuando quieras.

c. Al no existir la recogida sistemática de datos sobre las páginas, la visualización de índices no es frecuente.

d. El control del vocabulario se puede decir que prácticamente no existe.

Ángeles Maldonado Martínez y Ele-

na Fernández Sánchez. Centro de Información y Documentación Científica (Cindoc). Csic.

Pinar, 25, 3ª planta. 28006 Madrid.

Tel.: +34-914 112 220 ext. 244;

fax: 915 645 069

urbasoc@cindoc.csic.es

elenaf@cindoc.csic.es

<http://www.cindoc.csic.es>