

C.2. Motores de búsqueda para usos académicos: ¿cambio de paradigma?

Por Lluís Codina

Codina, Lluís. "Motores de búsqueda para usos académicos: ¿cambio de paradigma?". En: *Anuario ThinkEPI*, 2007, pp. 98-100.



"Scirus declara indizar más de 250 millones de sitios vinculados con la ciencia, mientras que Bubl o Sosig apenas tienen unas pocas decenas de miles"

"Cuando los motores de búsqueda ordenan sus resultados aplicando el análisis de enlaces (o análisis de citas, que es lo mismo), la navegación parece quedar obsoleta"

EN LOS ÚLTIMOS AÑOS, estamos asistiendo a una transformación en las capacidades y prestaciones de los motores de búsqueda que podría significar, en poco tiempo, si es que no lo está significando ya, un cambio de paradigma en el mundo de la búsqueda y recuperación de la información científica y técnica, y por tanto, en el mundo de la publicación científica.

Hasta ahora, hablar de motores y ciencia parecía una antinomia, una contradicción lógica: o hablamos de motores de búsqueda, o hablamos de publicación científica, pero no de ambas cosas.

Sin embargo, algunos desarrollos de diversos buscadores ponen en cuestión no solamente la idea anterior sino otras ideas más o menos arraigadas en el mundo de la Documentación acerca de los procesos de búsqueda y recuperación de información.

La propuesta clásica

Para buscar contenidos de calidad, filtrados y fiables, la propuesta clásica sugería limitar las búsquedas de información a tres tipos de servicios de información, a saber:

–directorios como *Yahoo!* o *Dmoz*, siempre limitando la búsqueda a sus respectivas secciones de ciencia, por supuesto;

–*information gateways* o portales como *Bubl* o *Sosig* que son, en realidad una clase de directorios especializados en ciencia;

–bases de datos como *ISI Web of knowledge* o *Science direct*.

<http://www.yahoo.com/>

<http://www.dmoz.org/>

<http://www.bubl.ac.uk/>

<http://www.sosig.ac.uk/>

<http://www.isiwebofknowledge.com/>

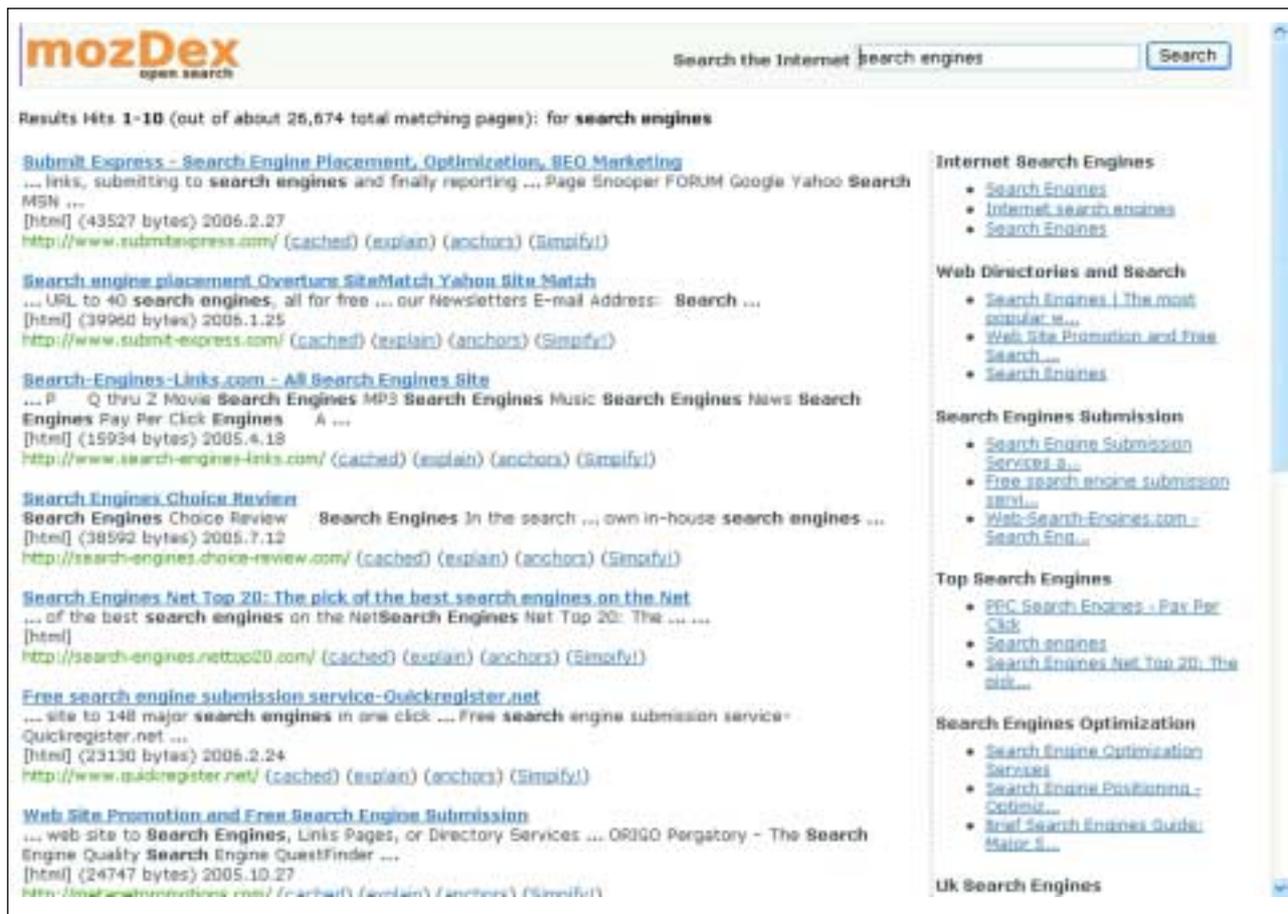
<http://www.sciencedirect.com/>

El problema es que casi nadie parece utilizar actualmente directorios como *Yahoo* o *Dmoz*.

Un experimento rápido puede aclarar muchas cosas: acceda el lector al sitio *Way-Back machine* de *Internet archive* y compare la posición relativa que el directorio ha ido ocupando en *Yahoo* a lo largo de los últimos años en comparación con sus inicios. Comprobará que ha pasado de ser el centro a prácticamente desaparecer.

<http://www.archive.org/web/web.php>

Por su parte, los portales o *information gateways* parecen haber entrado en una situación de estancamiento, cuando no de crisis, debido a la siguiente contradicción: filtrar contenidos de la web mediante análisis intelectual impide llegar a siquiera una fracción significativa de las páginas o sitios web sobre ciencia. Por ejemplo, un motor de búsqueda sobre el que volveremos después (*Scirus*) declara indizar más de 250 millones de



sitios vinculados con la ciencia (sitios .edu o .ac.uk, etc.), mientras que *Bubl* o *Sosig* apenas indizan unas pocas decenas de miles: o sea, que tenemos aquí un desfase de varios órdenes de magnitud entre lo que realmente ofrece la Web y lo que los servicios que funcionan *a mano* pueden asimilar. Esto sin contar que *Scirus* indiza y da acceso a páginas individuales, mientras que los sistemas anteriores se limitan al sitio como un todo.

El nuevo paradigma

Además (o precisamente por ello), desarrollos recientes de los motores de búsqueda y la dinámica de la Web han puesto aún más en cuestión no solamente el supuesto conflicto de ideas entre motores y ciencia, sino la propuesta clásica anterior.

¿A qué desarrollos nos referimos? Sin ánimo de exhaustividad, creemos que el cambio de paradigma queda muy bien representado por los siguientes casos:

1. Google Scholar
<http://scholar.google.es/>

2. Google Books
<http://books.google.com/>

3. Scirus
<http://www.scirus.com/>

4. Amazon Search Inside
<http://www.amazon.com>

5. A9
<http://a9.com>

6. Mozdex
<http://www.mozdex.com/>

Dada la necesaria limitación de esta nota, y sin perjuicio de desarrollar el tema en trabajos posteriores, nos gustaría destacar lo siguiente:



Google Scholar y *Google Books* presentan un escenario en el cual un motor de búsqueda llega donde nunca han llegado ni los directorios ni las bases de datos. Entre ambas opciones el usuario del mundo profesional o de la ciencia puede consultar un índice que está formado por el contenido a texto completo de páginas y sitios web científicos, publicaciones científicas impresas y libros impresos, también a texto completo.

Scirus indiza el contenido de más de 250 millones de servidores científicos del tipo .edu, .ac.uk, etc., sumando a sus resultados diversos bancos de tesis internacionales, también a texto completo (julio 2006).

Amazon Search Inside representa una alternativa a *Google* por sí, por alguna razón, deseáramos *una segunda opinión*, es decir, consultar otra base de datos gigante de libros indizados a texto completo.

A9 y *Mozdex* aportan novedades susceptibles de tener su respectivo impacto en el uso de motores de búsqueda con fines académicos. *A9* (en realidad, un desarrollo de *Amazon* que utiliza el índice de *Google*) aporta la idea de los perfiles e intereses personales de sus usuarios. En *A9* el usuario (si lo desea) no es anónimo, sino que tiene un perfil concreto del que el sistema aprende para refinar sus búsquedas, aparte de ofrecer de manera rutinaria el contenido de libros impresos como parte de las respuestas. Finalmente, *Mozdex* es un caso peculiar de *¿open search?*, es decir, de un motor de búsqueda que explica, para cada página o

sitio web de su página de resultados el porqué de la ordenación (ranking) presentada al usuario.

Consecuencias

Al menos de modo provisional, pueden apuntarse algunas conclusiones: en primer lugar, la contradicción ciencia vs. motores está siendo (adecuadamente) destruida. En segundo lugar, tal vez habrá que revisar algunas ideas clásicas de la disciplina conocida como *information seeking and retrieval*, que afirmaba el lugar destacado de la navegación junto con la interrogación en el acceso a la información. Revisión que debería llevarse a cabo al menos cuando el contexto es la web y los motores de búsqueda pueden aplicar análisis de enlaces. La evidencia actual indica que, cuando los motores de búsqueda ordenan sus resultados aplicando el análisis de enlaces (o análisis de citas, que es lo mismo), la navegación parece quedar obsoleta o, al menos, habrá que buscar el motivo por el cual ya nadie parece interesado en utilizar directorios como *Yahoo* o *Dmoz* salvo para practicar posicionamiento web, pero esto es otra historia como diría el bueno de Kipling.

Lluís Codina

lluís.codina@upf.edu

<http://www.lluiscodina.com/>