

Conociendo mejor los buscadores

[Versió catalana]





ISIDRO F. AGUILLO 

Laboratorio de Cibermetría

CCHS - CSIC Madrid

isidro.aguillo@cchs.csic.es

Opciones

    [Imprimir](#) [Recomanar](#) [Citació](#) [Estadístiques](#) [Metadades](#) [<meta />](#)

El impacto de los motores de búsqueda sobre la Web, su estructura, percepción y sus usuarios ha sido enorme, posiblemente mucho más de lo que hasta la fecha se ha reconocido. Un breve recorrido histórico nos permite comprobar cómo la evolución de los buscadores ha modelado nuestra imagen de la Web y no siempre de forma correcta. En una profesión de intermediarios como la nuestra, el intermediario (motor) ha alcanzado tal protagonismo que según los vaivenes nos ha promovido (*Altavista* y la búsqueda booleana) o nos ha relegado a la cuneta (cualquiera puede buscar en *Google*).

A mediados de los noventa, los buscadores eran meras evoluciones de los sistemas de recuperación de bases de datos basados en frecuencia de aparición de términos. Sin embargo, lo que funcionaba en las bases bibliográficas parecía fallar en la Web, donde la relevancia y pertinencia de los resultados dejaba mucho que desear. Se oyeron voces acerca de la falta de control documental en la Red o del carácter caótico de los contenidos. Pocos sabían que en realidad el motor líder de la época, Lycos, truncaba por defecto y que ello generaba un enorme ruido que pocos achacaban al peculiar funcionamiento de este motor que requería el poco intuitivo añadido de un punto tras el término para hacer una búsqueda exacta.

En el 96 llega *Altavista*, que entre otras muchas novedades incorpora operadores booleanos y delimitadores remedando las posibilidades de los filtros por campos tan queridos por los documentalistas. Sin embargo, la presentación de resultados opta por un criterio, a menudo pasado por alto, que ofrece como texto referente de los contenidos las primeras líneas de cada página. En documentos muy largos, pertinentes a la búsqueda, ese supuesto "resumen" es a normalmente poco descriptivo y da lugar a una percepción generalizada errónea: *Altavista* devuelve mucha "basura".

Google cambia las reglas, el PageRank le da una gran ventaja sobre *Altavista* y le permite dejar de lado los operadores booleanos (que pasan a ser implícitos) en aras de la sencillez, con un interfaz limpio. Además, los textos de cada entrada ya no son las primeras líneas sino que se contextualizan, e incluyen frases que contengan los términos de búsqueda. Psicológicamente el usuario siente que los resultados son más pertinentes. Hay además algunas razones más, especialmente importantes para el documentalista, que justifican el éxito de *Google*: Es más rápido, en un momento en que todavía había limitaciones de velocidad, y es mucho más grande, lo que reduce nuestro mayor miedo, el silencio documental.

Pero el éxito de *Google* también impone restricciones a sus prestaciones. Hay millones de peticiones simultáneas y todos los usuarios deben ser atendidos en plazos relativamente breves, por lo que el tiempo de cómputo individual está restringido.

Google necesitará tiempo y dinero (salida a bolsa) para mejorar su búsqueda avanzada y ofrecer recuperaciones por rangos. Los ingenieros informáticos de *Google* son excelentes y los clústers de PCs en los Data Center suman más de 500.000 ordenadores, pero el crecimiento de peticiones sigue obligando a ciertos compromisos prácticos:

Las cifras de resultados ofrecidas por *Google* son siempre aproximaciones, a veces con grandes diferencias dependiendo del Data Center que conteste. Posiblemente dichos resultados están sobrestimados pues la búsqueda de frase no es exacta al igual que el uso de múltiples operadores combinados. Otras prestaciones avanzadas están en lista de espera dada la cantidad de recursos informáticos que requieren: clasificación de resultados (*racimos*), búsqueda multilingüe o

truncamientos.

En los últimos dos años una de las mejoras poco conocidas de *Google* sí que ha tenido un impacto significativo en la percepción de la calidad de los resultados. Se ha incorporado un sistema de localización (semifiltro lingüístico), que es relativamente simple, y que va ligado a la elección del idioma del interfaz por parte del usuario. Dicha elección condiciona los resultados presentados que son ahora más "próximos" geográfica e idiomáticamente. Sin embargo, es una opción que pasa desapercibida para la gran mayoría del público y los profesionales.

Google se ha decantado estos últimos años por diversificar su oferta de servicios, ofreciendo buscadores de imágenes y video, académicos o geográficos, y haciendo un gran esfuerzo por integrarlos. Esto puede tener consecuencias a medio plazo pues el interfaz se complicará notablemente y la fusión de contenidos puede generar más ruido. Sin embargo la idea de incorporar imágenes parece funcionar bien en Exalead (volcados) y es una de las bazas de *Cuil*, un nuevo motor diseñado por ex miembros de *Google*.

Es difícil predecir el futuro, pero si *Google* desea mejorar la calidad de sus respuestas, necesariamente necesitará recopilar mucha más información adicional. Un primer paso, a la imagen de lo que ya hace tiempo recoge *Yahoo*, es asociar peticiones con enlaces visitados. No es seguro que además se asocie algún tipo de *cookie* o identificación por IP pero este conocimiento extra acaba perfilando al usuario y mejorando con la experiencia. Pero es evidente que *Google* quiere ir más allá y tiene los medios necesarios para cruzar información personal de todas las fuentes que ya controla. Esto ha generado un fuerte debate sobre la privacidad y la protección de los datos personales, pero como el *boom* de las redes sociales muestran, esta parece ser una preocupación minoritaria.

Una alternativa a tan sombrío panorama pasa por la asunción de algunas responsabilidades por los intermediarios humanos (por ejemplo, documentalistas y bibliotecarios), quizá oponiendo la web semántica a la web social, ganado así la batalla de la intertextualidad. Pero como bien sabemos ni la web semántica ha tenido éxito todavía, ni necesariamente el ingente trabajo necesario para producirla estará en nuestras manos.