



Diciembre 2004
Enero 2005

Unir, Compartir, Distribuir: la Búsqueda de Información en la Red. El caso Google



Número actual

Números

Editorial

Sitios de Interés

Libros

Ediciones Especiales



Carr. Lago de
Guadalupe Km. 3.5,
Atizapán de Zaragoza
Estado de México.

Tels. (52)(55) 58645613
Fax. (52)(55) 58645613

Por Antonio Gómez
Número 42

Resumen

La revolución tecnológica de las comunicaciones se basa en la posibilidad de unir, compartir y distribuir la información de todos los usuarios de la Red a nivel mundial y en tiempo real. Por este motivo, si en Internet está todo, lo importante ya no es saber, sino saber encontrar lo más relevante, la posesión de la información no es ya la cuestión central, sino el acceso, saber encontrar la información que necesitamos a la mayor brevedad posible y con la mayor concreción posible. En este sentido el uso y el mercado de los buscadores se presentan como uno de los mercados más prometedores y uno de los temas más polémicos para un futuro próximo.

La búsqueda de información con el uso de motores de búsqueda implica tres cuestiones fundamentales y muy debatidas en los actuales estudios sobre Internet: la cuestión de la libertad de expresión, la cuestión de la formación del conocimiento y la cuestión de la amenaza de la privacidad.

Buscadores

El mercado de las búsquedas en Internet se perfila como uno de los más disputados en un futuro próximo. Yahoo!, MSN¹ y Google parecen ser los tres gigantes que compiten en este mercado. Sus ataques circulan entorno a sus buscadores y sus servicios de correo². La próxima pelea parece que será el mercado de la información local³.

Tras varios años de competencias entre los distintos tipos de buscadores, la búsqueda a través de algoritmos del tipo de Google está imponiéndose a la indexación tradicional de páginas como Yahoo!, al ser mucho más eficaz. Por otro lado la publicidad encuentra en las búsquedas un campo de cultivo muy interesante ya que al hacer una búsqueda un internauta expone muy claramente sus intenciones lo que significa que se le puede ofrecer información comercial completamente ajustada a lo que está buscando, por este motivo la publicidad en este servicio es la que más está creciendo en la Red.

Cuando los matemáticos creadores de Google, Segey Brin y Lawrence Page, estaban preparando su proyecto de investigación hace seis años, y crearon su herramienta de búsqueda estudiando las relaciones combinatorias entre sitios web probablemente no pensaron que Google sería actualmente el buscador más utilizado de la Red, que va camino de convertirse en herramienta de consulta universal. Su primer servicio, el buscador, nació hace seis años. Hoy utiliza miles de ordenadores PC (se dice que hoy tienen más de 100.000), con lo cual consiguen varias ventajas: una red descentralizada, que les da más agilidad a las búsquedas, en casos de fallos nunca afecta a todo el sistema y además es más barato. Con ellos sirven diariamente más de 150 millones de consultas en menos de 0,2 segundos. Tras la salida a bolsa el pasado mes de agosto de Google, sus creadores se han convertido en dos de los jóvenes más ricos el mundo.

El objetivo inicial del buscador perfecto ha traído consigo muchos usos interesantes, y no todos estaban pensados en un principio. Para mantener su posición de motor de búsqueda más popular, Google va incorporando nuevas habilidades, como el listado de imágenes⁴, tiendas, noticias o citas, fruto de la creatividad de sus laboratorios⁵. Allí nació Google News⁶, que le confiere a computadoras la misión de presentarnos las noticias más importantes del momento. Google News es un desarrollo del mismo software del buscador, que analiza el número de links a un documento o noticia y la credibilidad del sitio web del que procede. Esta 'primera plana', que tiene a toda la industria informativa desencajada, selecciona información extraída de más de 4.000 sitios de fuentes informativas, las agrupa por secciones – Negocios, Ciencia y Tecnología, Deportes – y las ofrece a los internautas actualizadas cada quince minutos; y además sorprende por su pertinencia. Todo esto lo hace de forma automática, sólo con un software, y sin un solo periodista. Es decir, Google aprovecha el trabajo combinado de cientos de miles de periodistas y de miles de directores de periódico.

Frente a este nuevo servicio muchos se preguntan si el periodismo de la era de Internet será una acumulación sin alma de conexiones electrónicas entre negocios mediáticos globales o si continuará requiriendo un agudo criterio editorial, reporteros que estén al día, columnistas provocativos y una dosis de personalidad.

Otro producto es Froogle⁷, que ofrece información y fotografías de todo lo que se vende. Lanzando a toda prisa, justo antes de las Navidades de 2002, tiende aún a dar demasiadas respuestas, lo que confunde a los compradores. En otra línea el Viewer⁸, apodado Google TV, permite que el consumidor vea desfilan, sin *clickear*, las páginas que corresponden a las respuestas que busca. Pero aun no está a pleno rendimiento porque hay problemas con los diferentes estándares de los vídeos y con la velocidad de descarga; aunque estos problemas técnicos se podrían resolver, la cuestión parece estar sobre todo en los problemas de derechos de autor. La oferta del motor de búsqueda se completa con las Citas de Google⁹; que permite conocer lo que se dice sobre de un sitio. Es una extensión de la tecnología de base que no sólo se ocupa de los vínculos que llevan a una página sino de lo que dicen sobre ella.

Frente a esta avalancha algunos editores de enciclopedias defienden la información contrastada, comprobada y seleccionada de las enciclopedias frente a los problemas de fiabilidad de la información de la Red. Otros no ven posibilidades de competencia con una base de datos tan grande como la que ofrece Internet y que Google rastrea. Pero frente a la red, como medio abierto, y las enciclopedias, como medios cerrados, consideran que el abierto puede ayudar al cerrado pero al revés no. en contra alguno piensan que la batalla del papel contra el soporte informático está perdida y que el futuro probablemente pase por enciclopedias en red permanentemente actualizadas y contrastadas, donde compraremos en lugar de cuarenta tomos una clave de acceso.

El secreto del sistema de algoritmos para la búsqueda de datos que rigen a Google está guardado con enorme celo, pero se sabe que PageRank, nombre de la tecnología, establece el valor de un sitio en base al número de enlaces que apuntan a su dirección. Es decir que mide la popularidad de cada web en comparación con otras del mismo contenido y enlazan con ella situando las mejores en los primeros lugares. De este modo, un enlace que permite ir de la página A a la página B es interpretado como un voto de A a favor de B, además PageRank pondera los resultados al dar una importancia mayor a los votos de sitios que reciben más tráfico, y no indexa todas las páginas de un sitio.

Con este sistema se favorece a las páginas más visitadas haciéndolas más visitadas aún, con lo que la posición monopolística de Google contribuye a que toda la estructura de la

web sea más rígida. Para romper esta rigidez y mejorar su posición, algunas empresas utilizan, entre otros recursos, link farms, páginas llenas de enlaces que apuntan a su sitio. Cuando detecta la maniobra, Google baja la posición de dicho sitio¹⁰. La revolución tecnológica consiste en sus posibilidades no troncales de acceso a la información, en forma de red, de todos con todos, al margen de una vía principal, y las fórmulas de indexación de los motores de búsqueda pueden estar favoreciendo la formación de vías principales de flujo de información (Mattelart, 1998). Otro recurso es pagar por un espacio publicitario en la parte alta de Google.

Libertad de expresión y visibilidad (publicidad)

La industria de los buscadores en Internet es un negocio creciente y a medida que suben las cifras de los ingresos se agudizan los movimientos empresariales para tomar posiciones en este mercado.

El negocio más próspero es el de la inclusión de enlaces pagados en los resultados que se ofrecen durante una búsqueda. En algunos casos, estos enlaces de pago se diferencian claramente del resto, pero algunos buscadores cobran por situarlos en la cabecera de la lista de respuestas suministradas sin advertir al internauta de que se ha cobrado por su aparición en lugar preeminente.

Numerosas compañías pagan a los motores de búsqueda para que sus sitios aparezcan bien ubicados cuando un internauta hace una consulta relacionada, de lejos o de cerca, con su área de actividad¹¹. Constituye una buena práctica comercial para ambas partes. Sin embargo, es grave que el público no esté debidamente informado. Este tipo de acuerdos limita las opciones del consumidor y puede llegar a alterar nuestra relación con el conocimiento.

Los motores de búsqueda tienen dos maneras de favorecer a quienes pagan. La primera es el *posicionamiento pagado*, que consiste simplemente en colocar al inicio de la lista de los resultados de búsqueda a todos aquellos que se hayan suscrito al servicio. De esta manera cuando realizamos una búsqueda aparecen tanto sitios relacionados con el tema, como sitios de empresas que han pagado por estar ahí.

La segunda manera de favorecer a los pagadores, más compleja, consiste en indexar con mayor frecuencia ciertos sitios y, por consiguiente, en tener a los usuarios mejor informados de sus desarrollos más recientes; esto se llama *inclusión pagada*.

La mayoría de los usuarios no tienen conocimiento de que los motores de búsqueda más populares cobran tarifas especiales a cambio de colocar algunos sitios de manera más visible que otros¹². La falta de claridad de los motores de búsqueda es grave para el consumidor quien puede considerar comprar en un sitio convencido de que la marca ha convencido a la mayoría de los usuarios, cuando realmente le ha pagado al buscador más que sus competidores por figurar en buena posición.

Google parece respetar – hasta ahora- de manera escrupulosa las fronteras entre información y publicidad. Utiliza para su publicidad el sistema Adwords, que aparte de cobrar al anunciante sólo cuando alguien pincha el anuncio, también relega a los anuncios, e incluso los retira, si no interesan al internauta. La colocación de un anuncio depende por tanto de que lo enlace mucha gente, pero también de que lo paguen. De esta forma la clasificación del anuncio corre al 50% a cargo de los anunciantes y de los usuarios, de manera que si un sitio está muy arriba en la lista no es sólo porque han pagado mucho, sino porque, teóricamente, a la gente le interesa. En Mayo de 2003 Google compró Applied Semantics, una empresa dedicada a analizar el contenido de las webs para incrustar la publicidad más idónea. Sus programas discernen entre si el usuario, cuando busca sobre un término ambiguo, por ejemplo Ibiza, está interesado en

coches o en un viaje a la isla para colocar publicidad automovilística o de ocio.

El éxito de Google, cuyo negocio no es sólo la publicidad adosada a los listados de búsqueda (aunque algunas fuentes aseguran que representa el 95% de sus ingresos¹³) sino la venta de su tecnología a terceros, ha animado nuevas aventuras en este mercado. El mismo software que busca documentos en la Red puede rastrear información en la intranet de una empresa. Y a Google no le faltan clientes dispuestos a pagar entre 28.000 y 250.000 euros por estas herramientas de hecho Google es rentable desde 2001.

La información no siempre es poder (conocimiento)

Siguiendo una interpretación conducta humana, cuando navegamos por Internet en muchas ocasiones seleccionamos lo que nos parece más relevante entre la ingente masa de información que recibimos, lo que nos lleva a cometer errores. A esto se añade que Internet no es una fuente de información, es un canal por donde fluye una enorme cantidad de información cuya fiabilidad depende de las fuentes de las que procede. No siempre lo que está disponible en Internet es bueno, es decir, significativo. Una de las claves del éxito del cambio tecnológico en la enseñanza es la producción, gestión y acceso a los contenidos de calidad.

El acceso a toda la información no sustituye la competencia previa para saber qué información pedir y qué uso hacer de ella. El acceso directo a no suprime la jerarquía del saber y de los conocimientos. Hay algo de fanfarronada en creer que uno se puede instruir sólo con tener acceso a las redes (Wolton, 2000:97).

Pero muchos estudiantes piensan que si no encuentran la información que buscan en Internet de manera inmediata, es que no existe. Se olvidan así de otras fuentes impresas como los libros, las revistas, bibliotecas y de los propios profesores que conforman la fuente de referencia de información y conocimiento más importante. Incluso los propios estudiantes no valoran lo que la Red les puede realmente proporcionar, fácilmente olvidan la información y documentación que esconde la misma Internet y que nos muestra en 'primer lugar' el buscador de turno. El volumen de información que puede ser accesible en la llamada 'Internet visible' es ínfimo respecto a la zona 'invisible'.

Que Internet ofrece una gran libertad de expresión es teóricamente cierto, una vez publicada la información está disponible para los millones de usuarios, entre otros muchos miles de millones de documentos, lo que en la práctica equivale a que la posibilidad, de que esa información se vea está muy próxima a cero. La parte visible de la WWW está concentrada en unos cuantos nodos hacia los que apuntan multitud de enlaces, lo que queda fuera de ellos es prácticamente invisible, ya que los buscadores utilizados para localizar información en Internet están sesgados hacia los grandes nodos, cuestión que deberían considerar los nuevos empresarios que fabrican memoria, no bienes (Rifkin, 2000: 196).

Si dos personas buscan el mismo dato en una enciclopedia es imposible que lleguen a soluciones distintas. El peligro de Google, lo que falla, es la fiabilidad de las fuentes. A esto se añade que desde que somos conscientes de que todo el mundo busca en Google, existe la tentación de meter información falsa. Aun así, Google es una herramienta fundamental, cuya misión utópica original es organizar la información mundial y hacerla universalmente accesible y útil. Google es la diversidad y la diferencia; pone a disposición de los usuarios la mayor diversidad de información de la historia de la humanidad.

Pero ninguno de estos artilugios resuelve el problema que mortifica a todos los internautas: la lista interminable de

respuestas, entre las cuales es difícil escoger. Los internautas son impacientes y van directamente a al primer resultado. La casi omnipotencia de Google preocupa a pequeños empresarios cuyas entradas dependen de la posición obtenida en la clasificación de este sitio de referencia. Ser excluido de las primeras respuestas equivale a una condena, por la sencilla razón de que la mayoría de los usuarios no pasan de los primeros veinte resultados. La mayoría nunca va más allá de la primera página, y si no encuentran lo que buscan cambian la pregunta. Conscientes del problema, algunos de los nuevos motores de búsqueda clasifican los resultados en categorías¹⁴ u ofrecen una representación gráfica para facilitar la selección¹⁵. Aún así, hoy por hoy, parece que las tecnologías de categorización y visualización no están listas para satisfacer las necesidades de los usuarios. Pero a medida que los usuarios se vuelvan más sofisticados y que la información accesible (e indexada por Google) vaya aumentando, este enfoque podrá resultar insuficiente.

Amenaza de privacidad personal

Otro problema que plantean los motores de búsqueda es el de las denuncias de monitorización a través de las búsquedas, que invaden nuestra privacidad. Parece que nos encaminamos hacia una época en la que la libertad y privacidad estarán restringidas, donde la privacidad quedará como un invento de la revolución industrial. Hoy, la generación *puntocom* sólo quiere estar conectada. Inconscientemente definen la libertad como inclusión, mientras las generaciones anteriores definirían la libertad como exclusividad (privacidad y autonomía). La privacidad se está perdiendo, porque no podemos estar conectados constantemente y al mismo tiempo mantener nuestra privacidad. Parece que la privacidad puede convertirse en una reliquia en el mundo en red y que lo que necesitamos sean nuevos regímenes reguladores.

Google tiene respuesta para todo, y se acuerda de cada pregunta que hacemos, lo cual empieza a preocupar a los defensores de la privacidad¹⁶. La barra de herramientas de Google transmite datos sobre todos los sitios que visitamos. La cookie conserva los datos casi sin límite de tiempo, y aguanta hasta el 2038. Además del cookie insertado en nuestra computadora para reconocernos en cada visita, Google identifica nuestra dirección IP, la cual prácticamente revela el lugar desde el cual nos conectamos y puede ser utilizado para desplegar publicidad a medida.

Los miles de servidores de Google le permiten guardar rastro de nuestras preguntas, de la hora y día en la cual fueron formuladas y de los sitios visitados consecuentemente. Ésta, que podríamos denominar tecnología de minería de datos, junto a la extracción de información de bases de datos, amenazan con su uso generalizado a la privacidad personal. Sabiendo por ejemplo lo que preguntamos a Google alguien podría dar una descripción bastante precisa de lo que estábamos pensando¹⁷, lo que puede permitir elaborar un perfil preciso de nuestros usos y costumbres. Según los creadores el objetivo es memorizar los gustos de los usuarios para que obtengan lo que ellos quieren, pero que los consumidores son cientos de millones y las páginas miles de millones lo cual hoy por hoy supone demasiadas variables.

Pero la amenaza de Google a la intimidad no cesa en su buscador; parece ser que su nuevo servicio de correo que todavía no ha visto la luz ya tiene problemas. La estrategia de Google para rentabilizar su correo es la misma que la utilizada con su buscador; analizar palabras y venderlas al mejor postor. Esta estrategia no les ha causado problemas en el buscador, pero analizar los contenidos de los correos electrónicos de sus usuarios e insertar publicidad relacionada con ellos ya ha provocado quejas de las asociaciones defensoras de los derechos de la intimidad. La empresa se escuda en que el análisis será automático y no realizado por personas, y que sólo se insertará publicidad en los mensajes recibidos y no en los enviados, y no descartan que sea un servicio opcional que los usuarios puedan deshabilitar.

Conclusiones

La tecnología no es ni buena, ni mala, ni neutra; refleja la condición humana. La tecnología es poder y el poder nunca es neutral. Cuando uno lo ejerce siempre hay ganadores y perdedores. La neutralidad de la ciencia y de la tecnología es un gran mito que se ha creado, probablemente, por la comunidad de ingenieros.

Internet aún se encuentra en una etapa de desarrollo bastante inicial y todavía no se adapta fácilmente a los usos sociales. Sigue un proceso lento y gradual en su adaptación a las necesidades de los usuarios, y es cada vez más grande y potente en su camino hacia la madurez. Revelará muchos nuevos conceptos sociales no incorporados en su tecnología que, por su condición de transformadora, no la debemos entender limitada a unos fines determinados. Pero la naturaleza de estos impactos dependerá mucho más de las elecciones sociales y políticas que de la evolución tecnológica.

En la definición de cualquier tecnología, a partir de un determinado grado de desarrollo, que podemos establecer en el momento en el que tenemos tecnologías para problemas que no existen, cesa la innovación y cuentan más los factores sociales y culturales que los aspectos puramente tecnológicos.

Internet no sólo depende del espíritu emprendedor para difundirse en el conjunto de la sociedad, sino que también está ligada a sus orígenes en la comunidad científica y académica, donde se originaron los criterios de búsqueda de la excelencia, de apertura a la crítica de los colegas y de comunicación abierta del trabajo de investigación (Castells, 2001: 52-53).

A medida que nuestro uso de las tecnologías de la información evoluciona, estas prácticas utilizadas por los motores de búsqueda, son un riesgo que podría alterar nuestra relación con el conocimiento. De hecho, estamos pasando de una era en la cual el acceso a la información, siendo limitado, lo importante era el saber –el tener en mente la mayor información –, a una época en la que la información accesible siendo casi ilimitada, lo que cuenta es saber encontrar la más relevante. Sobre Google sobrevuelan sospechas de la posible entrega de datos a una agencia del gobierno de Estados Unidos o sobre la eliminación del sitio de un grupo fundamentalista cristiano anti-aborto o de material pro nazi en google.fr y Google.de, sobrevuelan al buscador. Aunque los creadores alegan tener por norma respetar la legislación de cada país y no mostrar pornografía, e insisten en que ellos no censuran, existen algunas dudas sobre una empresa cuyo poderío le confiere una responsabilidad que va más allá de sus imperativos comerciales. En la era del conocimiento, las técnicas de indexación y de visión están adquiriendo cada vez más un papel de utilidad pública, y ninguna alteración en su funcionamiento puede ser considerada inocente.

Notas:

¹ El buscador de MSN podría ir incluido en la próxima versión de Windows, lo que le está creando problemas a Microsoft con las autoridades antimonopolio de EEUU y la UE.

² El correo de Google y el buscador de MSN de próxima aparición.

³ Google ya ha lanzado Google Local en EEUU que no sólo ofrece información sobre objetos o sitios, sino que también incluye mapas para acceder a los sitios.

⁴ Ver <<http://images.google.com>>

⁵ Ver <<http://labs.google.com>>

⁶ Ver <<http://news.google.com>>

⁷ Ver <<http://froogle.google.com>>

⁸ Ver <<http://labs.google.com/gviewer.html>>

⁹ Ver <<http://labs.google.com/cgi-bin/webquotes>>

¹⁰ Para ver el listado de empresas implicadas que han tomado acciones judiciales <<http://www.chillingeffects.org/search.cgi?search=google>>

¹¹ Actualmente ha habido ya querellas interpuestas en esta sentido. Ver la querella de la organización Commercial Alert, en <<http://www.commercialalert.org/index.php>> y la carta de apoyo de la Federal Trade Commission, en <<http://www.commercialalert.org/PDFs/ftcresponse.pdf>>

¹² Ver encuesta de la organización en defensa del consumidor Consumer Report, en

<http://www.consumerwebwatch.org/news/1_TOC.htm>
13 Ver EL PAÍS, sábado 21 de agosto de 2004, p.56.
14 Ver <<http://vivisimo.com>>
15 Ver <<http://kartoo.com>>
16 Ver <<http://www.privacyinternational.org>>
17 Ver al respecto el seguimiento de <<http://www.google-watch.org>>

Referencias:

HABERMAS, J. (1992): *Ciencia y técnica como "ideología"*, Madrid, Tecnos, 1984.
MATTELART, A. (1998): *La mundialización de la comunicación*, Barcelona Paidós.
MANSELL, R. (ed.) (2003): *La revolución de la comunicación. Modelos de interacción social y técnica*, Madrid, Alianza Editorial.
RIFKIN, J. (2000): *La era del acceso: la revolución de la Nueva Economía*. Barcelona, Paidós.
ROJO, P.A. (2001): *Las nuevas tecnologías de la información en la Unión Europea*, Murcia, Universidad Católica San Antonio de Murcia.
WOLTON, D. (2000): *Internet, ¿y después qué?*, Barcelona, Gedisa.
CASTELLS, M. (2001) *La galaxia internet*. Barcelona, Plaza & Janés Editores, S.A.

Antonio Gómez Aguilar

*Grupo de Investigación en Comunicación y Cultura, Universidad de Sevilla,
España*