

La diversidad cultural en Google y los motores de búsqueda: una aproximación conceptual

Por Ernest Abadal y Lluís Codina

Resumen: Se analiza una preocupación expresada tanto de forma implícita como explícita en diversos ámbitos, según la cual los motores de búsqueda en general y Google en particular, podrían estar utilizando algún sistema para ocultar determinados contenidos culturales o bien para destacar aquellos que proceden del ámbito anglosajón en detrimento de los de otras culturas y lenguas. Dicho de otro modo, se considera, en una primera aproximación de modo exclusivamente conceptual, la idea de si el uso intensivo de los motores de búsqueda, tan característico de nuestro tiempo, resulta una amenaza para la diversidad cultural. Para ello, en primer lugar y con el fin de contextualizar este trabajo, se lleva a cabo una presentación de dos polémicas previas sobre el dilema proteccionismo o libre flujo de los productos culturales y a continuación se analizan las características más relevantes de los motores de búsqueda (centrándonos en Google en particular) con el fin de valorar si éstos constituyen una amenaza creíble a la diversidad cultural e idiomática de la Web.

Palabras clave: Google, Motores de búsqueda, Diversidad cultural.

Title: Cultural diversity in Google and other search engines: A theoretical approach

Abstract: In various contexts, a concern has been expressed, both implicitly and explicitly, that search engines in general, and Google in particular, may be using some system that obscures certain cultural content or emphasizes the Anglo-Saxon perspective, to the detriment of other cultures and languages. To put it in other terms, we present an exclusively theoretical approach to the idea that intensive use of search engines, which is so common now, represents a threat to cultural diversity. First, with the goal of putting our study in context, we present two previous debates on the dilemma of protectionism vs. the free flow of cultural products, and then analyse the most relevant characteristics of search engines (with a focus on Google) to determine whether they constitute a credible threat to cultural and linguistic diversity on the Web.

Keywords: Google, Search engines, Cultural diversity.

Abadal, Ernest; Codina, Lluís. “La diversidad cultural en Google y los motores de búsqueda: una aproximación conceptual”. En: *El profesional de la información*, 2008, marzo-abril, v. 17, n. 2, pp. 191-198.

DOI: 10.3145/epi.2008.mar.09



Ernest Abadal es profesor titular de la Facultad de Biblioteconomía y Documentación de la Universitat de Barcelona (UB); Adjunto al vicerrectorado de Política Docente para Sistemas de Información y Documentación (UB); y vicepresidente del Consorci de Biblioteques Universitàries de Catalunya (CBUC). Es autor de diversos artículos y monografías sobre recuperación de información y sobre publicaciones digitales.



Lluís Codina es profesor titular de universidad. Imparte docencia en los estudios de periodismo y en la Facultad de Comunicación Audiovisual de la Universidad Pompeu Fabra de Barcelona. Es el investigador principal del Grupo de Investigación DigiDoc de la misma universidad. Participa en el máster interuniversitario UB/UPF en gestión de contenidos digitales, en el programa de doctorado del Departamento de Periodismo y de Comunicación Audiovisual y es co-director del máster online de documentación digital.

Introducción

NADIE DUDA QUE LOS MOTORES DE BÚSQUEDA constituyen uno de los elementos fundamentales de la Web. De hecho, la búsqueda es, justo después del correo electrónico, la principal actividad de los internautas.

Desde el ámbito de la recuperación de la información se acostumbra a analizar sobre todo sus ca-

racterísticas tecnológicas. No obstante, los motores de búsqueda, en tanto que partes fundamentales de la Web entendida como un gran medio de comunicación y de difusión de contenidos, también se pueden considerar desde puntos de vista sociales, culturales o políticos. La cuestión es que causa inquietud que Google sea el buscador utilizado de forma casi exclusiva por los ciudadanos de todo el planeta. Además,

parece cada vez más necesario añadir estas dimensiones de análisis a

“Causa inquietud que Google sea el buscador utilizado de forma casi exclusiva por los ciudadanos de todo el planeta”

las “meramente” técnicas (pero sin ignorarlas ni dejarlas tampoco de lado).

Buen ejemplo de esta preocupación son los trabajos de **Lucas D. Introna** y **Helen Nissenbaum**, que exponen en sendos artículos conjuntos (2000a, 2000b) los fundamentos del posicionamiento web y de los algoritmos de relevancia para resaltar cómo las presiones e intereses del mercado condicionan las listas de resultados de los buscadores.

Según estos investigadores los motores de búsqueda indexan especialmente los ámbitos en los cuales existen intereses comerciales, en detrimento de aquel tipo de información o de público que no dispone de incentivos económicos. Esos autores realizan una aproximación desde el punto de vista político y sin concesiones; su objetivo es criticar y poner de manifiesto supuestos intereses ocultos. La tesis de fondo es que los buscadores (y la Web en general) tienen un carácter de bien público (*public good*) por lo cual no pueden estar guiados exclusivamente por intereses de mercado sino en base a principios de interés público. En este sentido, se refieren a diversas líneas de acción a llevar a cabo para que los motores de búsqueda estén abiertos a todos los contenidos y a todos los públicos. Se trata de unos textos de carácter teórico y con una toma de posición previa. Aunque se puedan compartir algunos de los presupuestos de fondo, es difícil estar de acuerdo con algunas de las explicaciones y justificaciones que ofrecen, ya que no se basan en datos empíricos contrastados.

Eszter Hargittai (2000), por su parte, destaca el papel de filtro de los buscadores y otros servicios web similares y los considera como unos nuevos “gatekeepers”, es decir, sistemas facilitadores (o denegadores) de acceso a la información. Toma como referencia los principales

conceptos utilizados por parte de la teoría de la comunicación y los aplica al entorno de la Web. Una de sus aportaciones se refiere a la distinción entre contenidos existentes (*available*) y accesibles (*accessible*), siendo estos últimos los que son fácilmente localizables por los usuarios gracias a los mecanismos de priorización de los motores de búsqueda. Podría considerarse una dicotomía comparable a la habitual en nuestro ámbito entre contenido indizable o web superficial y material no indizable o web invisible, pero con un componente un tanto “conspirativo”.

La desconfianza sobre la actuación de los motores de búsqueda no se limita al ámbito político y económico, sino que también se ha formulado desde el punto de vista cultural. Se trata de una cuestión recurrente que va saltando a la palestra por motivos diversos, ya se trate del anuncio de un buscador “europeo”¹, o de los proyectos de digitalización masiva de libros, y pone sobre la mesa la discusión sobre la necesidad de Europa de defenderse de un presunto sesgo cultural de los motores de búsqueda de origen norteamericano, y muy en particular de *Google*, dado su dominio del mercado global.

En este texto nos proponemos analizar esta idea y presentar algunas bases conceptuales que puedan ayudar a determinar su credibilidad, es decir, a establecer si los buscadores realmente constituyen una amenaza para la diversidad cultural como puede suceder con otros ámbitos de la comunicación social (por ejemplo, con el flujo de información en los noticieros televisivos, en la prensa escrita o con el cine en los países donde el predominio de la producción *hollywoodiense* puede llegar a eclipsar las producciones nacionales). Examinaremos, por tanto, si el sistema que utilizan los motores como *Google* para seleccionar y ordenar los resultados

de sus búsquedas incluye algún mecanismo que otorgue un trato privilegiado a los contenidos culturales norteamericanos o anglosajones en detrimento de los de otras culturas o lenguas.

Con el propósito de situar nuestro trabajo en un contexto más amplio, comenzaremos por revisar dos polémicas políticas y económicas que se han producido en el entorno de los medios de comunicación social y que tienen que ver con la necesidad de establecer medidas proteccionistas o no al flujo cultural y de información entre países con el fin de preservar la diversidad cultural. En concreto, se trata de las discusiones sobre el *nomis* y sobre la denominada “excepción cultural”.

A continuación se discutirán con detalle algunas características y el modo de funcionamiento del buscador de *Google* (hasta donde la empresa propietaria los ha hecho públicos y son conocidos por los analistas externos) para valorar si estos elementos presentan algún indicio o alguna evidencia de tener capacidad para afectar a la diversidad cultural e idiomática en este caso de los recursos presentes en la Web. La fundamentación de este segundo apartado se basará en la presentación de los sistemas de indización y en los mecanismos de ponderación de resultados (ranking) del buscador.

Producción y difusión cultural: ¿proteccionismo o libre flujo?

Está plenamente establecido que la información y la cultura no son un producto comercial como cualquier otro. Ciertamente, la difusión masiva y en una sola dirección de información y cultura provoca con facilidad casos de colonización y homogeneización culturales. Esta constatación ha generado al menos un par de polémicas internacionales de profundo calado que están lejos de haber sido resueltas. Nos referi-

mos a la discusión sobre el “nuevo orden mundial de la información y la comunicación” (*nomíc*) y a la defensa por parte de algunos países europeos y en especial de Francia de la denominada “excepción cultural”.

Las discusiones sobre el *nomíc* arrancaron a finales de 1970 y recogían las inquietudes de determinados países que, dado el sistema imperante de producción de noticias periodísticas y de otro tipo en aquel momento, nunca eran difusores de noticias. Lo que pretendían era no ser engullidos por el flujo procedente principalmente de los países anglosajones, y de los EUA más concretamente (con sus poderosas agencias de noticias y cadenas de televisión a la cabeza).

La *Unesco*, entonces dirigida por el senegalés **Amadou Mahtar M'Bow**, propició la realización de diversos estudios para intentar corregir estos desequilibrios. Era lo que se denominó *nomíc*, una propuesta presentada dentro del ámbito de esta organización que pretendía reducir el colonialismo informativo propiciado “de facto” por los países occidentales y favorecer la difusión de la cultura de los países del Tercer Mundo en los medios de comunicación. El político irlandés, y premio Nobel de la Paz (1974), **Sean MacBride** fue nombrado por la *Unesco* en 1977 presidente de una comisión encargada de analizar los problemas de la comunicación a nivel internacional. El denominado *Informe MacBride*, presentado en Belgrado en 1980, denunciaba un desequilibrio informativo entre norte y sur que podía llegar a afectar la diversidad cultural. Una de sus recomendaciones fue la creación de una agencia de noticias internacional independiente gestionada por la *Unesco* para contrapesar el poder de las grandes agencias comerciales.

Los EUA y la mayoría de países occidentales se opusieron frontalmente a lo que consideraban una forma de atentar contra el libre mer-

cado y la libre competencia. Como consecuencia de las discusiones los EUA y Gran Bretaña salieron de la *Unesco* en 1984 y 1985, respectivamente², organización a la que consideraron como un foro de propaganda anti-occidental. La crisis que sufrió este organismo dependiente de las *Naciones Unidas* fue profunda, y no sólo tuvo repercusiones a nivel institucional sino también económico (hay que recordar que los EUA sufragaban una cuarta parte del presupuesto total de la *Unesco*). La creación de algunos medios de comunicación avalados y financiados directamente por gobiernos (*Al Jezira* y otras cadenas árabes, o el canal *teleSUR* impulsada por Venezuela y otros gobiernos latinoamericanos, son un buen ejemplo) mostrarían que los motivos de los debates aún siguen vigentes.

La “excepción cultural”, por su parte, fue un término afortunado acuñado por los franceses en medio de las negociaciones finales del *GATT* (*Ronda Uruguay*)³ con el fin de preservar los productos audiovisuales (cine y televisión) de la liberalización comercial. Francia y los EUA oficiaron de líderes de una y otra postura, y el punto de discusión se centró en la industria audiovisual sobre la cual los americanos querían que se eliminara cualquier barrera arancelaria o proteccionismo legal, mientras que los franceses defendían justo lo contrario. La mayoría de países de la UE (excepto Gran Bretaña) estaban de acuerdo con los franceses en esta cuestión. El origen de la polémica hay que situarlo en 1989, cuando se aprobó una directiva europea conocida popularmente como “Televisión sin fronteras”, que recomendaba a los países europeos que reservasen la mayor parte del tiempo de emisión televisiva para producciones europeas. Su aprobación fue denunciada dentro del *GATT* como un atentado contra los principios que regían en este organismo y que habían sido también suscritos por la UE.

Las conocidas y polémicas cuotas de pantalla que afectan la exhibición cinematográfica en muchos países europeos (es obligado exhibir una determinada proporción de cine europeo en todas las salas) eran el caballo de batalla de los estadounidenses, que lo consideraban una medida proteccionista atentatoria contra los principios de liberalización del comercio mundial, es decir, contra la filosofía de la *Ronda Uruguay*. El argumento de los franceses se basaba en la especificidad de la industria de la cultura que no tiene el resto de sectores productivos. La discusión no está cerrada y de hecho muchos países europeos mantienen todavía las cláusulas de protección para defenderse de la penetración de productos culturales extranjeros (especialmente estadounidenses). Después de muchas controversias y negociaciones dentro del *GATT* se excluyeron temporalmente las “creaciones del espíritu” de las normas vigentes para otras mercancías.

Motores de búsqueda y contenidos culturales

Las dos polémicas anteriores están ampliamente documentadas en la bibliografía especializada. No pasa lo mismo cuando se traslada este debate a la web y más concretamente al papel que ejercen los motores de búsqueda. En este apartado vamos a intentar demostrar que la posible amenaza a la diversidad cultural de los buscadores estadounidenses contiene diferencias fundamentales con las dos polémicas antes descritas. Para ello, vamos a examinar el funcionamiento de tres elementos sustanciales de la difusión de información por parte de los medios de comunicación social y a comparar su aplicación en ellos y en el caso de los buscadores.

Estos son: el tipo de selección previa de la información (lo que los comunicólogos denominan *agenda setting* y que se refiere a qué con-

tenidos se incluyen); la saturación de canal (es decir, cuáles se pueden llegar a distribuir); y la priorización de la información (en qué posición de la lista se hacen aparecer los contenidos).

“La posible amenaza a la diversidad cultural de los buscadores contiene diferencias fundamentales con las polémicas del *nomie* y de la ‘excepción cultural’”

Es indudable que la combinación de estos tres factores puede amenazar la diversidad cultural, ya que unos determinados contenidos pueden no ser seleccionados para su difusión, o no disponer de tiempo o canal para ser difundidos o relegarse a ámbitos de bajo nivel de consulta. Ahora bien, ¿pueden aplicarse esos elementos de análisis al caso de los motores de búsqueda en general y de *Google* en particular? En lo que sigue, vamos a proceder a su examen más detallado.

1. Selección previa de la información

Los medios de comunicación social (tanto en su formato analógico como en el digital) deben seleccionar la información por diversos motivos. Incluso en un canal apa-

rentemente ilimitado como la web, los medios de comunicación seleccionan qué es noticia y qué no lo es. Se trata de la función denominada *agenda setting* según la cual, aunque los medios no dictan cómo debe pensar el ciudadano, sin duda establecen sobre qué corresponde pensar.

Por su parte, los motores de búsqueda no disponen en general de elementos para filtrar qué páginas incluyen o no en su base de datos en función de sus contenidos (por ejemplo, basándose en la calidad de la misma, en su solvencia, su idioma, etc.). Es decir, no se produce de forma intrínseca una exclusión de contenidos por motivos culturales o lingüísticos (y no existe evidencia de que *Google* se aparte de ello en este aspecto) ya que se limitan a indizar los documentos publicados en la web, a volcar en un índice (y a generar metadatos) de forma automática, con lo que los usuarios de cualquier rincón del mundo pueden encontrar recursos de su entorno cultural más próximo con el simple uso del idioma propio.

Las dudas sobre la cantidad de contenidos que los motores pueden indizar constituye uno de los principales argumentos de **Introna** (2000a) o **Hargittai** (2000, p. 19) para criticar la poca atención a la diversidad cultural y política por parte de los motores de búsqueda y otros portales que caracterizan la web. Para compensar esta idea

vamos a recurrir a un estudio de **L. Vaughan** y **M. Thelwall** (2004) en el que se analiza la cobertura geográfica de los motores de búsqueda con el objetivo de determinar si existe algún sesgo político o cultural por parte de los buscadores hacia los contenidos estadounidenses o de otros países.

El trabajo analiza la cobertura de las sedes web comerciales de cuatro países (EUA, China, Taiwán y Singapur) por parte de tres motores de búsqueda (*Google*, *Altavista*, *Alltheweb*) y en dos idiomas (inglés y chino). Los resultados muestran que las sedes web de los EUA están mejor representadas (83%) que las de China (61%), Taiwán (51%) o Singapur (49%) y que *Google* es el motor que incluye un mayor porcentaje de sedes web de cada uno de los países. Sin embargo, según los autores, los motivos que explican este diferente grado de inclusividad no se deben a causas políticas o sociales (es decir, a la existencia de sesgos premeditados por parte de los buscadores) si no que tienen una explicación técnica basada en que las sedes web de los EUA disponen de un mayor número de enlaces que las apuntan y, por tanto, tienen mayor visibilidad y facilidad para ser encontradas por parte de los motores de búsqueda. Se trataría, por tanto, de una variedad de lo que en el campo de la bibliometría se conoce como “efecto Mateo” (mencionado por primera vez por **Merton** en 1968, o “a quien más tiene

	Medios de comunicación social	Web
Selección previa	Limitación en función del interés (político, social, cultural, etc.) del medio.	Tan sólo existen algunos límites técnicos (localización de las páginas por parte del buscador).
Saturación del canal	Limitación en función de los horarios de emisión (radio y televisión) o del número de páginas (prensa escrita).	Sin limitación. Posibilidad de distribuir y acceder a todos los contenidos indexados.
Priorización de la información	Mejores horarios (radio y televisión) o primeras páginas (prensa escrita) seleccionados en función de criterios subjetivos.	Algoritmo de relevancia que ordena los resultados en función de criterios objetivos.

Tabla 1. Comparación entre medios de comunicación social clásicos y la web respecto a los tres factores analizados

más se le dará”). Tres años más tarde, **Vaughan** y **Zangh** (2007) ampliaron este mismo estudio a webs educativas, gubernamentales y de organizaciones, llegando a resultados similares.

La tesis de estos autores coincide con otros estudios que también señalan la importancia de los enlaces (“*links have economic power*”), no tanto por el hecho de aumentar el tráfico de visitas que proviene de ellos, sino porque permiten que una sede web sea incluida en los motores de búsqueda. Así pues, si adoptamos el principio que aconseja escoger de entre varias explicaciones la más simple (“navaja de Ockham” o “en igualdad de condiciones la solución más sencilla es probablemente la correcta”), todo parece indicar que las causas que explican la presencia de unas sedes web concretas en los motores de búsqueda son exclusivamente técnicas y no tienen una motivación política o cultural.

“Las causas que explican la presencia de unas determinadas sedes web en los motores de búsqueda son exclusivamente técnicas y no tienen una motivación política o cultural”

2. Saturación del canal

En los medios “clásicos”, es decir, en la prensa escrita y la radio y la televisión analógicas, existen unos canales a cuya saturación se llega con cierta facilidad. Por ejemplo, los diarios no pueden tener un número ilimitado de páginas en cada edición, y el espectro televisivo y radiofónico permite un número limitado de canales de televisión de emisoras de radio para cada zona geográfica.

En el mundo digital (es decir, el propio de los buscadores), en cambio, no hay una limitación previa sobre la cantidad de contenidos que puede soportar el canal. Se estima que puede haber en la web al menos varias decenas de miles de millones de páginas (se especula con unos 30 mil millones). Que un ciudadano de Manhattan o de Berlín publique un nuevo blog no va en detrimento de los ya preexistentes, y que la *Universidad de Barcelona* incorpore un nuevo repositorio de documentos académicos en su web no es perjudicial para lo que pueda hacer la *Universidad de México*, etc. A lo sumo, será más difícil hacer consultas efectivas en la web para ciertos temas durante un período de tiempo, pero todos los contenidos serán accesibles al mismo tiempo y por igual a todos los efectos prácticos para quienquiera que introduzca la url adecuada en el navegador (o formule la pregunta correcta en *Google*).

Durante un tiempo, la rumorología sobre *Google* especuló sobre las dificultades de construir un índice con más de diez mil millones de páginas. Puede que el límite existiera en algún momento y durante un cierto tiempo, pero *Google* pronto anunció contener en su índice información sobre 20 mil millones⁴. La cuestión es que no parecen existir fronteras tecnológicas prácticas para la cantidad de información que es capaz de albergar la web. Por su parte, el precio por gigabyte almacenado no deja de caer en picado desde hace años, así como el de las CPU y el del ancho de banda disponible en internet. Estos tres elementos combinados eliminan las barreras económicas al crecimiento del volumen de la información disponible en formato digital.

3. Priorización de la información

Los medios clásicos disponen (para bien y para mal) de diversas formas para, una vez aceptado un

contenido, privilegiarlo (o degradarlo): primera página de un diario, *prime time* en los informativos de televisión, acceso a un gran número de salas de cine para distribuir un film, etc.

El aspecto correspondiente a la priorización de los resultados en los motores de búsqueda que podría asimilarse con este apartado es el cálculo de relevancia. Como es sabido, *Google* no presenta los resultados de manera aleatoria ni cronológica, sino aplicando un cálculo de relevancia, mostrándolos en orden decreciente (ranking). Los componentes concretos del cálculo no solamente permanecen en secreto sino que además varían con frecuencia. No obstante, hay dos cosas que se conocen con seguridad: en primer lugar, se trata de un conjunto de algoritmos que tienen en cuenta características intrínsecas de las páginas web de tipo estadístico o posicional: frecuencia absoluta y relativa de la palabra clave, posición de la palabra en el cuerpo del documento, etc. En segundo lugar, los motores de búsqueda aplican un análisis de enlaces de tipo recursivo que otorga más peso a las páginas que reciben mayor número de enlaces de otras que a su vez son muy enlazadas.

En general, la combinación de ambos criterios (los intrínsecos de cada página con relación a cada pregunta, más el peso obtenido como resultado del análisis de enlaces) proporciona el índice de relevancia relativo de cada página para cada pregunta concreta. Obsérvese que, en este sentido, en un motor de búsqueda no se produce un orden previo de cada página, no hay por tanto lugares privilegiados, sino la selección y el orden de presentación se establecen a posteriori, como resultado de comparar el texto (en rigor, la cadena de caracteres) de cada pregunta concreta con las entradas del índice del motor de búsqueda.

Existe en *Google*, eso sí, un indicador, denominado *PageRank* que es estático, es decir, es propio de la página independientemente de la pregunta. Aunque a este respecto deben considerarse a su vez dos cosas: no es otorgado por un editor (como en la prensa escrita) sino por el conjunto de los creadores de páginas web mediante sus decisiones individuales cada vez que deciden a qué página enlazan (y a cuál no) desde las suyas propias. En segundo lugar, el *PageRank* final de cada página se computa de forma automática para todas las páginas del índice de *Google*, como no podría ser de otro modo cuando estamos hablando de una base de datos de 25 mil millones de páginas, cada una con su *PageRank* propio.

En cambio, tanto **Introna** como **Hargittai** coinciden de nuevo en dudar de que los intereses comerciales no prevalezcan de una forma u otra en las listas que elaboran los buscadores. **Introna** y **Nissenbaum** se refieren a la ausencia en la Web de determinados contenidos porque no tienen interés para el mercado. No obstante, no pueden ofrecer pruebas concretas de que exista sesgo por parte de los motores de búsqueda⁵. Por su parte, **Hargittai** también utiliza argumentos similares (**Hargittai**, 2000, p. 23): “el contenido producido por entidades con presupuestos suficientemente grandes puede optar a una situación prominente en portales muy visitados. Esto privilegia tal contenido sobre el de creadores más pequeños y menos dotados financieramente”.

De hecho, se llega a sugerir la creación de portales igualitarios, alternativos (se pone a *Dmoz* por ejemplo), que no estén sujetos a presiones comerciales. Los cambios que se han producido en el ecosistema web desde entonces (en especial la web 2.0) creemos que han desmentido estas teorías, ya que se ha podido comprobar cómo proyectos colaborativos y sin intereses comer-

ciales han llegado a las más altas cúspides (por ejemplo, *Wikipedia*).

Vistos los elementos anteriores, para que los buscadores pudieran representar una amenaza a la diversidad cultural en el mismo sentido que en otras ocasiones lo han sido determinados medios de comunicación social, sería necesario negar sus tres características básicas y que ya hemos discutido, a saber:

– Deberían discriminar contenidos en función de su origen geográfico o de alguna característica cultural de su contenido, pero hemos visto que no es así.

– El canal digital, internet en este caso, debería tener algún tipo de saturación según el cual si existen ya muchos contenidos de una cultura dominante, digamos la anglosajona, estos materiales taponarían el acceso a contenidos de otras culturas, pero de nuevo simplemente no sucede así.

– Por último, la priorización de los contenidos (es decir, el hecho de ocupar o no un lugar privilegiado en la lista de resultados) debería realizarse en función del origen del documento, de las características de su autor o de la fuente, de la lengua del mismo, de su pertenencia a un tipo de cultura u otra, etc., cosas de las que, no solamente ningún analista ha detectado hasta ahora el más mínimo indicio, sino que contradecirían el propio modelo de negocio de estas herramientas, basado en su capacidad de localización. Una priorización como la que estamos discutiendo privaría a los motores de búsqueda estadounidenses de la mayor parte de sus usuarios no anglosajones, y en cuanto a los de naturaleza local, serían una formidable competencia para *Google*, lo que (tal vez por desgracia) no sucede en absoluto.

La zona de sombras

Hasta ahora nos hemos referido a cuestiones de carácter teórico

que, sin embargo, parecen ser desmentidas por algunas ejemplificaciones prácticas. Recientemente se han invocado tres casos concretos para defender la teoría según la cual los motores de búsqueda (y *Google* en especial) realizan algún tipo de manipulación de los resultados que podría representar un sesgo en perjuicio de culturas no anglosajonas. Estos son: la censura de contenidos en China, la priorización de resultados en función del idioma y el proyecto *Google Books*.

Sobre el primer caso, se sabe que *Google* alteró el funcionamiento del algoritmo en su versión china para excluir los resultados que hacen referencia a asuntos “sensibles” del gobierno chino (Tiananmen, Falung Gong, etc.). Podemos señalar dos cosas sobre este tema: en primer lugar, esto supone una alteración consciente y reconocida (y por supuesto, del todo lamentable) del funcionamiento del algoritmo que no se ha detectado en ningún otro caso. En segundo lugar, se aplica igualmente de forma automática, es decir, el sistema simplemente no entrega resultados para determinadas palabras clave que el gobierno chino ha censurado. No estamos ante una priorización de resultados de un origen cultural sobre otro, sino que aquí la censura actúa de forma lamentablemente igualitaria. Además, aunque sea un triste consuelo, cabe señalar que un ciudadano chino que utilice la versión de *Google* en otra lengua, por ejemplo en inglés, no se verá afectado por esa censura (excepto que el aparato de Estado chino disponga de otras formas de controlar contenidos de la Web, por ejemplo, a través de los proveedores de acceso).

El segundo caso consiste en la denuncia de que las posibles búsquedas de contenidos sobre **Goethe**, **Baudelaire**, o **Cervantes** en *Google* sitúan en primer lugar los recursos que están escritos en inglés sin tener en cuenta que, en

la mayoría de los casos, las mejores páginas web para estos términos se encuentran en el idioma de los autores utilizados como ejemplo, es decir, en alemán, francés o español, respectivamente. En esta cuestión se tiene que tener en cuenta que existe una priorización de los resultados en función del idioma de la versión local de *Google* que se utiliza. Esto afecta a las distintas versiones, ya sea la alemana (.de), francesa (.fr), española (.es) o inglesa (.com).

Quiere decirse que se dará prioridad a la versión local en cada caso: si usamos *Google* en la versión francesa, priorizará resultados en francés, si usamos la versión española lo hará con aquellos en español. La explicación es clara: se entiende que cuando un usuario accede desde una versión nacional quiere disponer de la interfaz en su idioma y, además, de una cierta preferencia en los resultados que se encuentran en su idioma. En el caso de la versión internacional, que está en inglés, se presupone que el idioma del usuario que utiliza esta versión también es el inglés. No obstante, lo anterior solamente sucede con términos similares en más de una lengua. Por ejemplo, comprobaremos cambios en los resultados de diversas versiones locales si buscamos con la palabra clave “tv” que es igual para varias lenguas. En cambio, si la palabra clave se corresponde netamente con una lengua, por ejemplo, “Juegos Olímpicos” vs. “Olympic Games”, los resultados serán idénticos o muy similares en cualquiera de las versiones locales.

Google Books es el tercer caso que queremos comentar. Esta iniciativa trata directamente con contenidos (no con metadatos o índices inversos) y es por ello que puede conllevar un desequilibrio cultural a favor de las obras en lengua inglesa, ya que en este caso existe una labor de selección previa por parte de *Google* de las bibliotecas y editoriales. Hasta el momento, la colección se

está formando fundamentalmente a partir de acuerdos con editores y bibliotecas norteamericanos, aunque no faltan convenios europeos⁶, como se puede comprobar si se hacen búsquedas usando lenguas distintas del inglés.

“Se entiende que cuando un usuario accede desde una versión nacional quiere disponer de la interfaz en su idioma y de una cierta preferencia en los resultados que se encuentran en su idioma”

Lo cierto es que, en este caso, no estamos ante un rastreador que incluye contenidos de forma automática, sino que existe una selección ya que el número de objetos (libros impresos) a digitalizar e introducir en la colección no es ilimitado, sino que está condicionado, entre otros factores, por los costes del proceso de digitalización. No sabemos si en el futuro este desequilibrio se mantendrá, pero en estos momentos es patente.

En este sentido, es sabido que la *Comisión Europea*, con apoyo de los gobiernos de los diferentes países, planteó en marzo de 2005 el proyecto de *Biblioteca digital europea (European digital library–EDL)*⁷ que tiene por objetivo la preservación y difusión del patrimonio documental europeo. De esta forma se quiere asegurar una presencia relevante de los contenidos culturales del viejo continente, lo cual parece una actuación muy acertada. Más recientemente (abril 2007), la *Biblioteca Nacional de Francia* puso en funcionamiento el portal *Euro-peana*, una versión experimental para facilitar la consulta y navegación por los fondos bibliográficos de la *Biblioteca digital europea*.

<http://www.europeana.eu>

Conclusiones

Hasta ahora hemos examinado desde un punto de vista conceptual la credibilidad de la acusación de que el predominio de los buscadores estadounidenses en general y de *Google* en particular es una amenaza plausible para la diversidad cultural. Desde esta perspectiva, esta amenaza es difícil de fundamentar porque no hay nada en el funcionamiento de los motores de búsqueda que la justifique ni hasta la fecha nadie ha podido presentar evidencias en este sentido. Como estudiosos del fenómeno nuestra obligación consiste en descartar teorías no fundadas y sobre todo, en no aceptar conjeturas conspirativas, cuyo fundamento va directamente contra la filosofía de la ciencia.

En relación con la comparación con el mundo de la comunicación social y de los medios analógicos, cabe señalar que en el mundo digital los motores de búsqueda no actúan como los medios de comunicación tradicionales (televisión, prensa, cine, etc.). En primer lugar, en el medio digital no existe una limitación previa sobre la cantidad de contenidos que pueden ser publicados; por otra parte, los motores de búsqueda no son creadores de contenidos, sino que se limitan a proporcionar acceso a los creados por terceros. Si una lengua o zona geográfica es pobre en la creación de materiales, eso es exactamente lo que reflejarán las búsquedas comparadas. Una consulta formulada en una lengua con una amplia base demográfica ofrecerá más resultados que otra en una lengua con una baja demografía, simplemente porque hay muchas más personas creando contenidos en el primer caso.

Entendemos que es responsabilidad de los actores y agentes culturales de cada lengua y de cada nación promover la riqueza y la productividad culturales, pero esa producción cultural solamente

aflorará, naturalmente, en los buscadores cuando se usen palabras en esa lengua para realizar búsquedas. Esto es algo tan básico que al parecer nos pasa desapercibido. Precisamente, los motores de búsqueda son la única forma de privilegiar contenidos de culturas minoritarias.

“Si una lengua o zona geográfica es pobre en la creación de materiales, eso es exactamente lo que reflejarán las búsquedas”

Así pues, la cuestión que se ha puesto a debate con motivo del predominio del mercado por parte de motores de búsqueda de origen norteamericano en general y de *Google* en particular es totalmente diferente de las otras dos polémicas (*nomie* y excepción cultural) antes planteadas. La intervención estatal para estimular una industria europea propia en el campo de los motores de búsqueda está más que justificada⁸, pero debería buscar otros argumentos, como la necesidad de incrementar la competitividad de las naciones dentro de las prácticas internacionalmente aceptadas y sin provocar un proteccionismo innecesario ni alentar, aunque sea involuntariamente, teorías conspirativas, ya que sería una vía garantizada hacia el fracaso. Las políticas culturales de futuro tienen que sostenerse en ideas y datos contrastados.

Notas

1. El denominado proyecto *Quaero* (“preguntar”, en latín), anunciado por los gobiernos francés y alemán en 2005 y del que no se ha sabido nada más en concreto, a excepción del anuncio público de la retirada del gobierno alemán a finales de 2006.

2. En 1997 volvió a ingresar Gran Bretaña y en 2004, los EUA.

3. El *GATT* (*General Agreement on Tariffs and Trade*, Acuerdo General sobre Aranceles y Co-

mercio) era un organismo intergubernamental que pretendía, desde su fundación en 1948, favorecer el libre comercio entre los países reduciendo las trabas arancelarias. La última ronda de negociaciones, *Ronda de Uruguay*, acabó en diciembre de 1993 después de más de ocho años de conversaciones. El *GATT*, como tal, se disolvió y dio paso a la *Organización Mundial del Comercio* (*OMC*), que tiene por objetivo favorecer el comercio mundial.

4. La *Wikipedia* estima en unos 25 mil millones el número de páginas a las que da acceso el índice de *Google*.

http://en.wikipedia.org/wiki/Google_search

5. En un determinado pasaje lo reconocen abiertamente: “Although we don’t have empirical evidence (...)”. En: **Introna**, 2000a, p. 59.

6. En el caso de las bibliotecas hay que destacar los recientes acuerdos con la *Universidad Complutense* (2006) y la Biblioteca de Catalunya (2007) para digitalizar sus fondos de reserva.

7. Se puede consultar una explicación resumida de su alcance en la web.

<http://libraries.theeuropeanlibrary.org/press/edl-faq20060302.html>

8. Algo parecido se llevó a cabo en el caso de la industria aeronáutica, con la creación y el apoyo de distintos gobiernos europeos a la empresa *EADS* (productora de los famosos *Airbus*).

Bibliografía

Abadal, Ernest; Codina, Lluís. “¿Es Google una amenaza para la diversidad cultural?”. En: *Anuario ThinkEPI 2007*. Barcelona: El profesional de la información, 2006, pp. 94-96.

Ackermann, Ernest; Hartman, Karen. *The information specialist’s guide to searching and researching on the internet and the world wide web*. Wilsonville: ABF Content, 1998.

Altman, Alon; Tennenholtz, Moshe. “Ranking systems: the PageRank axioms”. En: *Proceedings of the 6th ACM Conference on electronic commerce*, 2005. New York: ACM Press, 2005, pp. 1-8.

Basch, Reba; Bates, Mary-Ellen. *Researching online for dummies*. Foster: IDG, 2000.

Belew, Richard K. *Finding out about: a cognitive perspective on search engine technology and the www*. Cambridge: Cambridge University Press, 2000.

Berry, Michael W.; Browne, Murray. *Understanding search engines: mathematical modeling and text retrieval*. Philadelphia: SIAM, 1999.

Bianchini, Monica; Gori, Marco; Scarselli, Franco. “Inside PageRank”. En: *ACM Transactions on internet technology*, 2005, February, v. 5, n. 1, pp. 92-128.

Brin, Sergey; Page, Lawrence. “The anatomy of a large-scale hypertextual web search engine”. En: *WWW/Computer networks*, 1998, v. 30, n. 1-7, pp. 107-117. <http://www-db.stanford.edu/~backrub/google.html>

Calishain, Tara; Dornfest, Rael. *Google: los mejores trucos*. Madrid: Anaya, 2004.

Chowdhury, G. Gobinda. *Introduction to modern information retrieval*. 2nd ed. London: Facet, 2004.

Hargittai, Eszter. “Open portals or closed gates: channeling content on the world wide web”. En: *Poetics*, 2000, v. 27, n. 4, pp. 233-254.

Hock, Randolph. *The extreme searcher’s guide to web search engines: a handbook for the serious searcher*. Medford: CyberAge Books, 1999.

Introna, Lucas D.; Nissenbaum, Helen. “Defining the web: the politics of search engines”. En: *Computer*, 2000a, pp. 54-62.

Introna, Lucas D.; Nissenbaum, Helen. “Shaping the web: why the politics of search engines matters”. En: *The information society*, 2000b, jul-sep, v. 16, n. 3, pp. 169-185.

MacBride, Sean. *Un solo mundo, voces múltiples: comunicación e información en nuestro tiempo: hacia un nuevo orden mundial más justo y eficaz de la información y la comunicación*. México: Fondo de Cultura Económica; París: Unesco, 1980.

Mattelart, Armand. “Excepción o especificidad cultural: los desafíos del GATT”. En: *Telos*, 1995, n. 42, pp. 15-27. http://www.campusred.net/tepos/anteriores/num_042/opi_perspectivas5.html

Mattelart, Armand. *Diversité culturelle et mondialisation*. Paris: la Découverte, 2005.

Qiu, Feng; Cho, Junghoo. “Automatic identification of user interest for personalized search”. En: *Proceedings of the 15th International conference on world wide web*, 2006. New York: ACM Press, 2006, pp. 727-736

Ramonet, Ignacio. *Internet, el mundo que llega*. Madrid: Alianza, 1998.

Thelwall, Mike. *Link analysis: an information science approach*. Amsterdam: Elsevier, 2004.

Vaughan, Liwen; Thelwall, Mike. “Search engine coverage bias: evidence and possible causes”. En: *Information processing and management*, 2004, n. 40, 2004, pp. 693-707.

Vaughan, Liwen; Zhang, Yanjun. “Equal representation by search engines? A comparison of websites across countries and domains”. En: *Journal of computer-mediated communication*, 2007, v. 12, n. 3, article 7. <http://jcmc.indiana.edu/vol12/issue3/vaughan.html>

Williams, Joseph, et al. *Bots and other internet beasts*. Indianapolis: Sams.net, 1996.

Ernest Abadal, Universitat de Barcelona.
abadal@ub.edu

Lluís Codina, Universitat Pompeu Fabra.
lluis.codina@upf.edu