

Pablo Tachouet

Internet invisible: Un viaje a lo profundo de la Web

Copete La Web invisible es el término que describe toda la información disponible en Internet que no es recuperada por los buscadores convencionales. Por cada millón de páginas visibles hay otros 500 millones ocultas.

Texto

Generalmente es información almacenada y accesible mediante bases de datos. Si bien el 90% de las bases de datos están públicamente disponibles en Internet, los robots de los buscadores solamente pueden indexar su página de entrada (homepage) y son incapaces de ingresar dentro de las tablas e indexar cada registro, les falta la habilidad para interrogarlas, seleccionar sus opciones y teclear una consulta para extraer sus datos. En los capítulos siguientes se buscara desentrañar cuales son las características de este fenómeno, y los efectos que generan que estas páginas no sean indexadas por los buscadores más utilizados.

Ftes.

S/D(2008). [“¿Qué es la web invisible?”](#). En *Internetinvisible.com* Fecha de consulta: 01 de noviembre de 2008.

TS2

Los principales protagonistas de la Web invisible

Copete Los buscadores y las bibliotecas digitales juegan un rol fundamental en cuanto a la Internet invisible. A continuación se enumeran los más relevantes.

Texto

Google: Es el motor de búsqueda más utilizado en todo el mundo. Es el que más resultados brinda, más servicios adicionales tiene y, por ende, más popularidad alcanza. De las 61 mil millones de búsquedas llevadas a cabo en todo el mundo en agosto de 2007, 37 mil cien millones fueron realizadas mediante este buscador.

Quaero: Para contrarrestar la supremacía tecnológica de la influencia estadounidense, el sector público en Europa invirtió varios millones de euros en el desarrollo de este buscador. La idea es encabezada por dirigentes franceses, y apoyada por la Unión Europea.

Theseus: Buscador alemán respaldado por el gobierno de aquel país. Es financiado por la Unión Europea y por varias de las más importantes firmas alemanas en el área de la tecnología, como Siemens, SAP y Deutsche Thomson, las cuales colaborarán cada una en su área de mayor experiencia. Tiene como objetivo, al igual que Quaero, contrarrestar la influencia estadounidense.

Biblioteca Digital Europea: Busca promover el acceso de forma amplia y organizada al patrimonio cultural y científico de la Unión Europea en las redes informáticas mundiales. Nació en el año 2005, aglutinando los fondos de diferentes bibliotecas europeas en veinte lenguas. La misma fue fomentada por intelectuales y políticos de Europa.

Cuil: Es un motor de búsqueda de Internet lanzado el 28 de julio de 2008. Su objetivo es proveer al usuario de un resultado más extenso comparado con el de otros buscadores. La empresa declara tener un índice tres veces mayor que el de Google y diez veces mayor que el de Microsoft, con 124,426,951,803 páginas Web. Sus creadores son ex empleados de Google.

Completeplanet.com: Posee más de 70 mil bases de datos. Mediante el rastreo a través de CompletePlanet se puede ir a diferentes áreas temáticas, tales como la energía o la agricultura o la alimentación o la medicina, y encontrar los sitios en la web invisible que no son accesibles usando los motores de búsqueda convencionales.

Internetinvisible.com: recopila, describe y ofrece los enlaces a las bases de datos españolas gratuitas existentes en Internet. Ofrece, además de un formulario de búsqueda directa simple y avanzada, un directorio organizado en grupos temáticos, los cuales se subdividen a su vez en materias más específicas. Tras la consulta, se obtiene una ficha con la descripción del contenido de la base de datos relacionada, el enlace a la pantalla de búsqueda y la entidad o persona responsable de su creación. Existen enlaces a más de 2.400 bases de datos de acceso gratuito.

InvisibleWeb.com: es un directorio que contiene unas 10.000 bases de datos y archivos que contienen información que tradicionalmente no puede ser accedida por los buscadores tradicionales.

Yahoo: En sus inicios poseía un directorio de búsqueda, pero por falta de interés de los usuarios y por dificultades en su mantenimiento, fue dado de baja. En la actualidad Yahoo ofrece un motor de búsqueda. Produjo 8500 millones de búsquedas en agosto de 2007. Ubicándose en el segundo lugar de consultas de búsquedas mundiales.

Ftes.

Siri, Laura (2008). [“¿Buscar en la Web o “googlear”?”](#). En: *Blog de la Comisión 2, Taller de introducción a las TICs, Cátedra Beltrán* . 28 de agosto. Fecha de consulta: 01 de noviembre de 2008.

S/D (2005). [“Europa se lanza a destronar a Google”](#) . En *La Nación*. Sección: Política. 29 de diciembre. Fecha de consulta: 11 de noviembre 2008.

S/D (2008). [“Bruselas aprueba que Francia respalde con 99 millones el buscador Quaero”](#). En *CincoDías.com*. Sección: Empresas. 14 de marzo. Fecha de consulta: 13 de noviembre de 2008

Klew, Willy (2007) [“Theseus, el buscador europeo que competirá con Google”](#) . En *Neoteo* 21 de julio. Fecha de consulta: 12 de noviembre de 2008.

Arana, Patricio (2007). [“No se puede dejar a Google todo el manejo de la cultura”](#) . En *La Nación*. Sección: Cultura. 17 de enero. Fecha de consulta: 12 de noviembre 2008.

Falduto, Facundo. (2008) [“Cuil: ¿la alternativa a la supremacía de Google?”](#) . En *Perfil*. Sección: Tecnología. 29 de julio. Fecha de consulta: 11 de noviembre de 2008.

S/D (2008). [“Buscador Cuil, ¿una alternativa a Google?”](#) . En *Tecnocosas*. Sección: General. 14 de agosto. Fecha de consulta: 12 de noviembre de 2008.

TS3

Los diez conceptos fundamentales de la Web invisible

Copete A continuación se definen diez conceptos que ayudarán al lector a comprender el fenómeno en cuestión

Texto

Web invisible: Toda aquella información almacenada en bases de datos, material de archivo y herramientas interactivas tales como diccionarios o calculadoras, páginas dinámicas construidas con tecnologías Flash, ASP, PHP, etc. Estos recursos son embebidos dentro de miles de sitios web individuales y no son "visibles" para los motores de búsqueda tradicionales. Para acceder a todo ese incalculable acervo de información sólo podemos interrogar a las bases de datos directa e individualmente a través de sus propios formularios de búsqueda, puesto que las páginas indexables por los motores de búsqueda no dan cuenta de los recursos en ellas disponibles.

Web global (navegable): aquella Red de información libre y gratuita que es accesible teóricamente mediante la interconexión de computadoras. La forma de acceso se realiza mediante programas navegadores, Chats, mensajería o intercambio de protocolos (FTP, P2P).

Web oscura: Se define como los servidores o host que son totalmente inaccesibles desde nuestro ordenador. Según un estudio de la compañía Arbors Networks esta situación sucede en el 5% de los contenidos globales de la Red. La causa principal (78 % de los casos) se debe a zonas restringidas con fines de seguridad nacional y militar. El porcentaje restante, (22%) obedece a otros motivos: configuración incorrecta de routers, servicios de cortafuegos y protección, servidores inactivos y finalmente "secuestro" de servidores para utilización ilegal.

Arañas o robots: Las bases de datos de los motores de búsqueda son creadas por programas robot llamados spiders (arañas) que se arrastran por la web buscando el contenido de los motores de búsqueda. Estas arañas navegan la Web siguiendo los enlaces en las páginas web que ya están en la base de datos de su motor de búsqueda padre. Si no hay ningún enlace a una página, una araña no puede "verlo". Ellos no tienen la capacidad de mecanografiar o pensar en ninguna serie de caracteres. Ellos tampoco pueden examinar un grupo de opciones y escoger cual seleccionar.

Páginas dinámicas: Las páginas creadas como resultado de una búsqueda reciben el nombre de páginas "dinámicamente generadas". La respuesta a la búsqueda es colocada en una página web diseñada para llevar la respuesta y enviarla a la computadora. A menudo la página no se almacena en ninguna parte después, porque su contenido único (la respuesta a esa búsqueda específica) probablemente no es de utilidad para muchas otras personas. Es más fácil para la base de datos regenerar la página cuando sea necesario que guardarla.

Páginas estáticas: Lo opuesto a una página dinámica es una página estática. Las páginas estáticas residen en servidores, cada una identificada por un URL único, y esperando ser recuperada cuando su URL es invocado. Los spiders pueden encontrar una página estática si un enlace a ella aparece en cualquier otra página que ellos "conocen". Las páginas estáticas no son invisibles, aunque los motores de búsqueda puedan escoger omitirlas por razones de política

Motores de búsqueda: Indexan los documentos o recursos Web de forma automática por medio de robots (o spiders) y conforman una base de datos a la que posteriormente consultará el motor cuando se realiza una consulta.

Buscadores en el fondo: Determinados buscadores que sí indexan -y bucean- en parte de esa Web invisible. Denominados por algunos como "Submarinos", son bases de datos de bases de datos.

Directorios o Índices temáticos: se trata de listas de recursos organizados por categorías temáticas y estructuradas de forma jerárquica que permiten visualizar los contenidos ofreciendo una lista de enlaces a las páginas que aparecen referenciadas en el buscador.

Terabyte: Unidad de medida informática que se representa como "TB" y equivale a: 2 elevado al 40 bytes. En la práctica, un terabyte puede ser tanto 1.000.000.000.000 bytes (10 elevado al 12), como 1.099.511.627.776 bytes (1024 elevado a 4). Equivale a 1.024 veces un gigabyte. 1 terabyte = 1024 gigabytes = 1.048.576 megabytes = 1.073.741.824 kilobytes = 1.099.511.627.776 bytes.

Ftes.

Lamarca Lapuente, María Jesús (2008). ["La Web Invisible" y "Buscadores"](#). En: *Hipertexto: El nuevo concepto del documento en la cultura de la imagen*. 22 de julio. Fecha de consulta: 02 noviembre de 2008.

Barker, Joe(2003) [“Web Invisible”](#) . En: *Web Invisible*. Fecha de consulta: 02 de noviembre 2008

[Diccionario Informático Alegsa](#) Fecha de consulta: 02 de noviembre de 2008.

TS4

Internet Invisible y el derecho a la información

Copete En cuanto al fenómeno de la Web invisible no existe legislación o norma que la regule, ni a nivel local ni en el plano internacional. Lo mismo ocurre con Internet en sí; solo rigen regulaciones en cuanto al derecho de propiedad (copyright). Sin embargo, en referencia al derecho al acceso de la información, cabe citar algunos extractos de tratados internacionales y de la Constitución nacional que competen al tema aunque sea de manera indirecta.

Texto

La Declaración universal de los derechos humanos compuesta por las Naciones Unidas sostiene en su artículo 19 que:

“Todo individuo tiene derecho a la libertad de opinión y de expresión; este derecho incluye el de no ser molestado a causa de sus opiniones, el de investigar y recibir informaciones y opiniones, y el de difundirlas, sin limitación de fronteras, por cualquier medio de expresión.”

Por su parte el Pacto de San José de Costa Rica, redactado por la Organización de Estados Americanos, en su artículo 13 pregona que:

“1. Toda persona tiene derecho a la libertad de pensamiento y de expresión. Este derecho comprende la libertad de buscar, recibir y difundir informaciones e ideas de toda índole, sin consideración de fronteras, ya sea oralmente, por escrito o en forma impresa o artística, o por cualquier otro procedimiento de su elección.

2. El ejercicio del derecho previsto en el inciso precedente no puede estar sujeto a previa censura, sino a responsabilidades ulteriores, las que deben estar expresamente fijadas por la Ley y ser necesarias para asegurar: a) El respeto a los derecho o a la reputación de los demás, o b) La protección de la seguridad nacional, el orden público o a la salud o la moral pública.

3. No se puede restringir el derecho de expresión por vías o medios indirectos, tales como el abuso de controles oficiales o particulares de papel para periódicos, de frecuencias radioeléctricas o de enseres y aparatos usados en la difusión de información o por cualesquiera otros medios encaminados a impedir la comunicación y la circulación de ideas y opiniones.

4. Los espectáculos públicos pueden ser sometidos por la Ley a censura previa con el exclusivo objeto de regular el acceso a ellos para protección moral de la infancia y la adolescencia, sin perjuicio de lo establecido en el inciso 2.

5. Estará prohibida por la Ley toda propaganda en favor de la guerra y toda apología del odio nacional, racial o religioso que constituyan incitaciones a la violencia o cualquier otra acción ilegal similar contra cualquier persona o grupo de personas, por ningún motivo, inclusive los de raza, color, religión, idioma u origen nacional

En cuanto a la Constitución nacional, los artículos 14 y 42 expresan que:

“Art. 14.- Todos los habitantes de la Nación gozan de los siguientes derechos conforme a las leyes que reglamenten su ejercicio; a saber: de trabajar y ejercer toda industria lícita; de navegar y comerciar; de peticionar a las autoridades; de entrar, permanecer, transitar y salir del territorio argentino; de publicar sus ideas por la prensa sin censura previa; de usar y disponer de su propiedad; de asociarse con fines útiles; de profesar libremente su culto; de enseñar y aprender”

“Art. 42.- Los consumidores y usuarios de bienes y servicios tienen derecho, en la relación de consumo, a la protección de su salud, seguridad e intereses económicos; a una información adecuada y veraz; a la libertad de elección, y a condiciones de trato equitativo y digno. Las autoridades proveerán a la protección de esos derechos, a la educación para el consumo, a la defensa de la competencia contra toda forma de distorsión de los mercados, al control de los monopolios naturales y legales, al de la calidad y eficiencia de los servicios públicos, y a la constitución de asociaciones de consumidores y de usuarios. La legislación establecerá procedimientos eficaces para la prevención y solución de conflictos, y los marcos regulatorios de los servicios públicos de competencia nacional, previendo la necesaria participación de las asociaciones de consumidores y usuarios y de las provincias interesadas, en los organismos de control.”

El Manifiesto de la International Federation of Library Associations (IFLA) y la UNESCO sobre la biblioteca pública predica que:

“Todos los grupos de edad han de encontrar material adecuado a sus necesidades. Las colecciones y los servicios han de incluir todo tipo de soportes adecuados, tanto en modernas tecnologías como en materiales tradicionales. Son fundamentales su alta calidad y adecuación a las necesidades y condiciones locales. Los materiales deben reflejar las tendencias actuales y la evolución de la sociedad, así como la memoria del esfuerzo y la imaginación de la humanidad. Ni los fondos ni los servicios han de estar sujetos a forma alguna de censura ideológica, política o religiosa, ni a presiones comerciales”.

Ftes.

Organización de las Naciones Unidas (ONU)(1948) [“Declaración Universal de los Derechos Humanos”](#) 10 de diciembre. Fecha de consulta: 02 de noviembre de 2008.

Organización de los Estados Americanos (OEA) (1999) [“Pacto de San José de Costa Rica”](#)
. 7 de junio. Fecha de consulta: 03 de noviembre de 2008.

Artículos 14 y 42 de la Constitución Nacional (1994) [“Constitución de la Nación Argentina”](#) 22 de agosto. Fecha de consulta: 03 de noviembre de 2008.

International Federation of Library Associations and Institutions y UNESCO (1994) [“Manifiesto sobre la Biblioteca Pública”](#) noviembre 1994. Fecha de consulta: 04 de noviembre 2008.

TS5

La Web invisible y el acceso al conocimiento

Copete Internet invisible engloba toda aquella información que no es recuperada por los buscadores. A continuación se hará una descripción detallada de este fenómeno y del papel que Google desarrolla en torno al mismo. Se analizará la repercusión de dicho buscador en Argentina.

Texto

La Web invisible está compuesta por aquella información que no se puede recuperar ("ver") en los resultados de los motores de búsqueda. Según el estudio How much Information?, realizado por Peter Lyman y Hal R. Varian de la School of Information Management and Systems de la Universidad de California, Berkeley, “la cantidad de información de la Web navegable o visible es de 167 terabytes, mientras que la Web invisible es de 91.850 terabytes.” Se calcula que la información que permanece invisible en Internet es 500 veces más grande que aquella indexada por los motores de búsqueda. Los buscadores, a grandes rasgos, se dividen en:

Directorios o Índices temáticos

Los directorios son listas de recursos organizados por categorías temáticas que se estructuran jerárquicamente en un árbol de materias que permite visualizar los recursos descendiendo desde los temas más generales situados en las ramas superiores, a los temas más específicos situados en las ramas inferiores. Las categorías ofrecen una lista de enlaces a las páginas que aparecen referenciadas en el buscador. Cada enlace también ofrece una breve descripción de su contenido. Así pues, los directorios o índices se estructuran por temas o categorías principales que, a su vez, contienen otras subcategorías, y así sucesivamente hasta que al final se ofrecen enlaces directos a otras páginas o recursos de Internet.

Motores de búsqueda

La diferencia fundamental entre un índice y un motor de búsqueda es que mientras los índices mantienen su base de datos “manualmente”, utilizando para la inclusión de las direcciones a sus empleados o a los propios internautas que dan de alta sus páginas, los

motores de búsqueda emplean para ello un robot de búsqueda. Estos robots son programas que se dedican a recorrer la Web automáticamente recopilando e indexando todo el texto que encuentran, formando así enormes bases de datos en las que luego los internautas hacen sus búsquedas mediante la inclusión de palabras clave. Los robots recorren los distintos servidores de forma recursiva, a través de los enlaces que proporcionan las páginas que allí se encuentran, descendiendo a través de las distintas ramas de cada servidor.

La principal ventaja de los motores de búsqueda frente a los índices temáticos radica en la gran cantidad de información que recogen y la mayor actualización de sus bases de datos. Por el contrario, la ventaja de los Directorios frente a los motores reside en la mayor precisión en la búsqueda, aunque son menos exhaustivos que los motores de búsqueda, ya que se obtienen menos resultados.

Metabuscadores

Permiten buscar en varios buscadores al mismo tiempo. Los metabuscadores no almacenan las descripciones de páginas en sus bases de datos, sino que contienen los registros de los motores de búsqueda y la información sobre ellos, adecuando su expresión a los diferentes motores para remitir la consulta a los motores. Una vez que reciben la respuesta, la remiten al usuario no directamente, sino tras realizar un filtrado de los resultados. Este filtrado consiste en eliminar y depurar los enlaces repetidos y en ordenar los enlaces. Además, sólo aparecerá un número limitado de enlaces, los que se consideren más importantes. Los más repetidos ocuparán los primeros puestos ya que el metabuscador considerará que son los más relevantes por estar dados de alta en mayor número de buscadores. El problema es que, por lo general, no permiten realizar búsquedas tan avanzadas como en un motor de búsqueda, con lo que las consultas tienen que ser generales y no se puede buscar en profundidad. La gran mayoría de los metabuscadores utiliza el motor de búsqueda de Google.

Muchos datos quedan fuera de los buscadores, puesto que indexar informaciones contenidas en enormes bases de datos numéricas o textuales, exige gastar muchos recursos; resulta muy costoso para los buscadores almacenar en sus bases de datos este tipo de formatos. Por otro lado, los buscadores tampoco indexan muchos de los datos que se generan de forma dinámica en tiempo real, debido a que se convierten en obsoletos en un brevísimo lapso de tiempo y no merece gastar recursos en informaciones tan efímeras. A esto se une que muchas de estas bases de datos dinámicas han de rastrearse desde su propia ubicación o sitio Web, y con sus propias herramientas de búsqueda personalizadas, puesto que precisan contraseñas especiales para acceder a ellas. Los resultados de búsqueda de muchas de estas bases de datos son entregados en páginas Web que son sólo para una búsqueda. Ese tipo de páginas muy a menudo no se almacenan en ninguna parte: es más fácil y más barato generar dinámicamente la página respuesta para cada búsqueda que almacenar todas las posibles páginas que contienen todas las posibles respuestas para todas las posibles búsquedas que las personas puedan hacer en la base de datos.

Hay dos razones por las cuales un motor de búsqueda no contiene una página: barreras técnicas que prohíben el acceso, y decisiones “políticas” para excluirla. En cuanto a las barreras técnicas, si la única manera de acceder a determinada página Web requiere que se

mecanografié algo o se examine una página y se seleccione una combinación de opciones, los motores de búsqueda son incapaces de proceder. Las bases de datos de los motores de búsqueda son creadas por programas robot llamados spiders (arañas), que se arrastran por la Web buscando el contenido de los motores de búsqueda. Lo hacen siguiendo los enlaces en las páginas que ya están en la base de datos de su motor de búsqueda padre. Si no hay ningún enlace a una página, una araña no puede "verlo". Ellos no tienen la capacidad de mecanografiar o pensar en ninguna serie de caracteres, ni pueden examinar un grupo de opciones y escoger cual seleccionar.

Las páginas creadas como resultado de una búsqueda, son páginas dinámicamente generadas. La respuesta a la búsqueda es colocada en una página Web diseñada para llevar la respuesta y enviarla a la computadora. A menudo la página no se almacena en ninguna parte después, porque su contenido único (la respuesta a esa búsqueda específica) probablemente no es de utilidad para muchas otras personas. Es más fácil para la base de datos regenerar la página cuando sea necesario que guardarla.

Lo opuesto a una página dinámica es una página estática. Las páginas estáticas residen en servidores, cada una identificada por un URL único, y esperando ser recuperada cuando su URL es invocado. Los spiders pueden encontrar una página estática si un enlace a ella aparece en cualquier otra página que ellos "conocen". Ellos siguen los enlaces hacia ella y la recuperan. Las páginas estáticas no son invisibles, aunque los motores de búsqueda puedan escoger omitirlas por razones de política.

El contenido de muchas páginas Web es tanto "buscable" como "navegable" haciendo click en los enlaces. En la medida en que el contenido encontrado en la búsqueda se replica en páginas Web con enlaces en algún lugar, parte (o todo) el contenido puede ser encontrado en un motor de búsqueda general de la Web. Cualquier contenido no incluido en páginas estáticas enlazadas hacia algún otro sitio permanece invisible. Se debe buscar en una base de datos directamente para encontrarlas.

La incapacidad de los spiders de mecanografiar y pensar causa dos tipos de páginas Web Invisible: 1) el contenido de bases de datos especializadas con opción de búsqueda puede ser enteramente o parcialmente invisible o visible, dependiendo de cuánto está contenido en páginas estáticas con enlaces. Ejemplo: sitios con bases de datos con opción de búsqueda incluyen a la mayoría de los motores de búsqueda como Google o Northern Light o AltaVista. Los contenidos de todos los catálogos de bibliotecas on-line que no requieren una contraseña son también web invisible. Los resultados de las búsquedas se generan dinámicamente. A veces es posible retener esa URL en la parte superior del resultado de la búsqueda, y usarla para regenerar la página dinámicamente cuando se hace clic sobre ella. Sin embargo, las páginas de resultados no se almacenan en ninguna parte. 2) Requieren contraseña o login. Todos los sitios que requieren una contraseña o un login están cerrados para los spiders de los motores de búsqueda debido a que requieren mecanografiar algo que los spiders no pueden "saber". Los contenidos de estos sitios es muy poco probable que estén en cualquier motor de búsqueda general de la Web. Esto incluye todos los recursos que existen que requieren contraseña.

Los motores de búsqueda pueden escoger no incluir páginas porque el formato del documento sería buscado raras veces o sin éxito por los usuarios del motor de búsqueda. No hay ninguna razón técnica para que ellos deban excluirlas, sólo es una política tomada por muchas compañías de motores de búsqueda. Las bases de datos y spiders de los motores de búsqueda están optimizados para "leer" HTML, el lenguaje básico de la Web. Otros tipos de lenguajes de programación contienen códigos y requerimientos de formato que son incompatibles con HTML. Este lenguaje puede llevar enlaces a estas páginas, pero no el texto completo de su contenido en su formato especial. Las páginas con imágenes y sin texto son también a menudo omitidas porque, sin texto, no hay nada que utilizar como palabra clave para buscar esa imagen, entonces no son incluidas.

La información que permanece invisible es aquella que está almacenada en bases de datos, material de archivo y herramientas interactivas tales como diccionarios o calculadoras, páginas dinámicas construidas con tecnologías Flash, ASP, PHP, etc. Estos recursos son embebidos dentro de miles de sitios Web individuales y no son "visibles" para los motores de búsqueda tradicionales. Para acceder a todo ese incalculable acervo de información sólo podemos interrogar a las bases de datos directa e individualmente a través de sus propios formularios de búsqueda, puesto que las páginas indexables por los motores de búsqueda no dan cuenta de los recursos en ellas disponibles.

Se observa que nadie tiene acceso completo a todo Internet ya que no sólo existen áreas concretas de la red que son inaccesibles a la mayor parte de los internautas, sino también determinados contenidos que permanecen invisibles.

Ricardo Fornas Carrasco en "La cara oculta de Internet" establece 3 tipos distintos de Internet:

Internet global: Es aquella Red de información libre y gratuita que es accesible teóricamente mediante la interconexión de computadoras. La forma de acceso se realiza mediante programas navegadores, Chats, mensajería o intercambio de protocolos (FTP, P2P).

Internet invisible: Responde a todos aquellos contenidos de información que están disponibles en Internet pero que únicamente son accesibles a través de páginas generadas dinámicamente tras realizar una consulta en una base de datos. Esta particular naturaleza les hace inaccesibles a los procesos habituales de recuperación de la información que realizan buscadores, directorios y agentes de búsqueda. Pero podemos acceder a las mismas mediante nuestras habituales herramientas de navegación, correo, etcétera. La única condición es saber exactamente la dirección de acceso (URL o FTP).

Internet oscuro: Se define como los servidores o host que son totalmente inaccesibles desde nuestro ordenador. Según un estudio de la compañía Arbors Networks esta situación sucede en el 5% de los contenidos globales de la Red. La causa principal (78 % de los casos) se debe a zonas restringidas con fines de seguridad nacional y militar. El porcentaje restante, (22%) obedece a otros motivos: configuración incorrecta de routers, servicios de cortafuegos y protección, servidores inactivos y finalmente "secuestro" de servidores para utilización ilegal.

La Web invisible contiene un gran número de fuentes de información que no pueden buscarse porque su contenido no ha sido indexado, ni puede serlo, por los principales buscadores. Aun cuando se recupere un sitio que contenga una base de datos, es improbable que el buscador conduzca a la base de datos misma, puesto que requiere que se navegue por el sitio Web para encontrarla. Así pues, la Web invisible está constituida por toda esa información accesible vía Web, pero a la que no es posible llegar mediante una consulta a los buscadores tradicionales.

Google ahora tiene la habilidad de buscar el texto completo de muchos archivos PDF convirtiendo estos archivos a texto, y colocando el texto en HTML de manera que pueda trabajar como una página Web ordinaria en la base de datos de Google. En los resultados de la búsqueda aparece un enlace al documento original completo en PDF. Otros motores de búsqueda actualmente no proveen este servicio. Google funciona con un PageRank. Ampara una familia de algoritmos utilizados para asignar de forma numérica la relevancia de los documentos (o páginas web) indexados por un motor de búsqueda. El sistema PageRank es utilizado por Google para ayudarle a determinar la importancia o relevancia de una página.

En abril de 2008 Google comunicó que comenzó a explorar algunos formularios HTML para descubrir nuevas páginas y URLs que de otra manera no se podrían rastrear ni indexar. Cuando Google encuentra un elemento FORM (formulario) dentro de un sitio Web de calidad desarrollará una serie de consultas de prueba, insertando de manera automatizada palabras con cierto sentido en las cajas de texto y realizando selecciones en los menús desplegables radio buttons o check boxes. En caso de que la información que devuelva el formulario sea interesante y que aporte contenido no indexado todavía, esta se incluirá en el index de Google tal y como si fuera otra página web. Este experimento trata de mejorar la experiencia en las búsquedas y es parte de un amplio esfuerzo por parte de Google en incrementar la cobertura en la web. Sin embargo, sólo una pequeña parte de los sitios reciben este tratamiento. De la misma manera solo se procesarán formularios GET y se evitarán aquellos que solicitan información de carácter personal (contraseña, user, contactos, etc.) Los formularios HTML son la puerta de entrada a una enorme cantidad de datos situados más allá del alcance de los motores de búsqueda.

Google es el buscador más consultado en Internet. La expresión “Googlear” se ha convertido en nuestros días como el sinónimo hegemónico de la búsqueda en Internet. Pareciera que el “dominio” de Google sobre el común de los usuarios de la red no deja escapatorias. En cuanto al común me refiero a aquellas personas que utilizan la Web sin un conocimiento técnico y específico sobre la misma, sino como medio de consultas en general (ya sean académicas, turísticas, musicales, cinematográficas, etc.). Se infiere que tanto la falta de un conocimiento técnico, como la gran popularidad y publicidad que posee Google, hace que los usuarios de la Web, en su mayoría, elijan a este motor de búsqueda como opción principal. No hay que descartar tampoco que debido a las opciones que brinda sea el más utilizado. Sin embargo muchas páginas quedan fuera de su alcance, no logrando indexar sitios a los que quizás los usuarios por falta de conocimiento técnico nunca accederán. Según el experto en usabilidad Jakob Nielsen, de la consultora norteamericana Nielsen Norman Group, una cuarta parte de los usuarios no siempre consigue concretar acertadamente una búsqueda en Google. Y ese 24 por ciento no sólo está formado por

quienes no realizan las búsquedas correctamente, sino también por personas que ni siquiera logran encontrar la página misma del buscador, y terminan buscando en otro o bien abandonando la búsqueda.

En cuanto al acceso de la información, si bien no hay ninguna ley que regule el acceso a la misma en Internet, podría decirse que por más que la información que permanece en la Web invisible es accesible para los usuarios de Internet (si conocen la dirección exacta), la falta de conocimiento o el hecho de que “búsqueda en Internet” se ha convertido en buscar en Google para el común de los usuarios se ha vuelto una costumbre mundial, genera que aquellas páginas que no son indexadas por el buscador queden “vedadas” para la personas que las solicitan. Desde las estadísticas que se generan tanto mundialmente, como por regiones, pareciera que Google (al situarse como el buscador más utilizado) “administra” el conocimiento desde las páginas que indexa dejando sumergido aquello que no le es posible recuperar. El hecho de que haya páginas que no sean indexadas por motivos de seguridad nacional, pertenecientes a lo que se ha denominado como Internet oscura, se ajusta a lo expuesto en el Pacto de San José de Costa Rica (artículo 13 inciso 2) en cuanto a que el acceso a información vinculada con la seguridad nacional será restringido.

Según el investigador del CONICET, Adrián Curti, Internet Invisible puede considerarse como el mejor “recurso” para las bibliotecas académicas (universitarias, especializadas, científicas). Como se dijo más arriba, su principal contenido son las bases de datos en sus distintas formas. Las bibliotecas universitarias y de otras instituciones académicas son una de las principales proveedoras de productos de la Internet invisible. Atendiendo a criterios documentales, se podría agrupar dentro de la Internet invisible los siguientes sitios: Catálogos de bibliotecas, Bases de datos de bibliotecas (propiamente dichas), Catálogos de editoriales, Directorios de Publicaciones periódicas, Archivos, Repositorios de artículos de revistas, tesis y literatura gris, Directorios de bases de datos, Guías de sitios, Revistas electrónicas (en las que es necesario un registro previo y las que sólo se puede recuperar la información mediante búsquedas en su base de datos), Documentos en formatos no indexables, como archivos multimedia, Obras de referencia: enciclopedias, diccionarios, tesauros, etc.(en las que es necesario interrogar a la base de datos para acceder al contenido).

Por lo tanto, toda esta información al no ser indexada queda sumergida en las profundidades de la Web, motivo por el cual las personas que quieran acceder a ellas deberán tener el conocimiento de las páginas específicas, el conocimiento técnico como para utilizar otras herramientas de búsqueda, o quedar privadas de todas ellas.

Si bien la IFLA en su Manifiesto sobre la biblioteca pública rezaba por “incluir todo tipo de soportes adecuados, tanto en modernas tecnologías como en materiales tradicionales”, en la práctica quedan sepultadas a lo profundo de la Red.

Toda la información de la Web invisible no ha sido explorada hasta ahora. Sin embargo, existen en la red, determinados buscadores que sí indexan -y bucean- en parte de esa Web invisible. Denominados por algunos como “Submarinos”, son bases de datos de bases de datos. Pero no poseen demasiada difusión.

De los motores de búsqueda Google es el que más resultados brinda, más servicios adicionales tiene y, por ende, más popularidad alcanza. Según la medidora de tráfico Web Alexa, Google es la página Web más visitada de la Argentina. Nuestro país no escapa a la influencia mundial de la “gran G”.

Ftes.

Lamarca Lapuente, Maria Jesús (2008). [“La Web Invisible” y “Buscadores”](#). En: *Hipertexto: El nuevo concepto del documento en la cultura de la imagen*. 22 de julio. Fecha de consulta: 06 noviembre de 2008.

Barker, Joe(2003) [“Web Invisible”](#). En: *Web Invisible*. Fecha de consulta: 06 de noviembre 2008.

Curti, Adrián Néstor (2006) [“Internet invisible, su importancia para las bibliotecas universitarias”](#) En: *4ª Jornada sobre la Biblioteca Digital Universitaria* 19 de octubre. Fecha de consulta: 06 de noviembre 2008.

Madhavan, Jayant y Alon Halevy (2008) [“El rastreo a través de formularios HTML”](#) En: *Google Webmaster central blog* 11 de abril de 2008. Fecha de consulta 07 de noviembre de 2008.

[Alexa, compañía de información Web](#) Fecha de consulta: 10 de noviembre de 2008.

TS6

La influencia de Google en Latinoamérica

Copete Google es el sitio más visitado por los países más importantes de Latinoamérica. El idioma español se ha posicionado como el segundo más usado detrás del inglés en este buscador.

Texto

Google es uno de los portales de búsqueda más famosos del mundo para el cual Latinoamérica representa dos tercios de todas las búsquedas en español que se realizan a nivel mundial. Según Alexa en los principales países de Latinoamérica, entre los cuales figuran Brasil, Chile, Uruguay, Venezuela y México, Google lidera los rankings de las páginas más visitadas. Google emitió un comunicado en agosto de 2007, afirmando que el idioma español se había convertido en el segundo más usado para las búsquedas a través de Google en todo el mundo después del inglés y por delante del alemán, el japonés y el francés. De allí se desprende la gran influencia que posee el buscador en Latinoamérica.

Ftes.

Cepeda, Yuly (2008) [“América Latina: Mayor crecimiento mundial para Google”](#). En: *Minademas.com* blog colombiano de tecnología y ciencia. 10 de abril. Fecha de consulta: 10 de noviembre de 2008

S/D. (2008) [“Google habla español”](#). En: *CincoDías.com*, Sección: Empresas. 7 de agosto. Fecha de consulta: 10 de noviembre.

[Alexa, compañía de información Web](#) Fecha de consulta: 10 de noviembre de 2008.

TS7

La hegemonía de Google y sus principales adversarios

Copete Google es el sitio con más búsquedas realizadas en el mundo, dejando atrás por amplia diferencia a sus competidores directos. Desde Europa surgen proyectos para contrarrestar el dominio norteamericano de la Web.

Texto

Una encuesta llevada a cabo por la consultora Reuters en junio de 2004 arrojó que casi la mitad de todas las búsquedas que se realizan en Internet pasan por Google (48%). Un estudio realizado por Standard & Poor's señala que el 20% de los internautas usa las búsquedas de Yahoo, el 14% MSN y el 7% AOL.

En agosto de 2007, de las 61 mil millones de búsquedas llevadas a cabo en todo el mundo 37 mil cien millones fueron de Google. Yahoo produjo 8500 millones de búsquedas en ese período. El portal chino Baidu.com generó 3300 millones, y la MSN, de Microsoft, tuvo 2166 millones, según la consultora norteamericana comScore.

Según los datos recogidos por la medidora de audiencias Hitwise en agosto de 2008, Google captó el 70,77% de las búsquedas realizadas por los estadounidenses en el pasado mes de julio. El porcentaje alcanzado por Google es superior al 69,17% que la propia empresa había registrado en junio y al 64,35% obtenido en julio de 2007. Por el contrario, sus principales competidores no han dejado de perder cuota de mercado en el último año. Yahoo, ha visto como su cuota de mercado quedaba en el 18,65%. En junio, su posición era del 19,62% y en julio de 2007 superaba el 22,1%. También MSN de Microsoft ha sufrido una dura caída, al pasar de una cuota de mercado del 8,79% en julio de 2007 al 5,36% del pasado mes de julio.

Según un sondeo realizado por la consultora Netprovider en septiembre de 2007, Google es el buscador más utilizado en España con 77,93%, Yahoo solo alcanza un 12,62% Por cada búsqueda en Yahoo, se hacen 7 en Google.

Estas cifras demuestran la influencia y el poder de Google a nivel mundial. Sin embargo, para contrarrestar la supremacía tecnológica y cultural de la influencia estadounidense, el sector público en Europa está invirtiendo varios millones de euros en el desarrollo de un buscador denominado Quaero. El plan Quaero salió a la luz impulsado por los gobiernos de

Francia y Alemania, y especialmente del entonces presidente de la República gala, Jacques Chirac, y del ex canciller germano, Gerhard Schroeder, quienes comentaron que estaban dispuestos a aportar al programa entre 1.500 y 2.000 millones. Al proyecto también se sumaron empresas como France Télécom, Deutsche Telekom, el grupo audiovisual Bertelsmann, Siemens, el gigante del software SAP o la Agencia Europea para la Innovación Industrial. El objetivo final de los dos países pasaba por competir con los gigantes estadounidenses en el ámbito de las búsquedas MSN (Microsoft), Yahoo y la citada Google. Sin embargo, la llegada al poder de Angela Merkel en Alemania en septiembre de 2005 tuvo un impacto directo sobre Quaero, puesto que finalmente el país teutón se retiró del proyecto a principios de 2007 para desarrollar uno propio bautizado como Theseus. El proyecto está respaldado por varias de las más importantes firmas alemanas en el área de la tecnología, como Siemens, SAP y Deutsche Thomson, las cuales colaboran cada una en su área de mayor experiencia. Diversos medios comentaron entonces que habían surgido ciertas diferencias entre germanos y franceses sobre quien debía hacerse cargo de la dirección. En marzo de 2008 Quaero parece revivir. Avalado por un holding participado por 23 empresas, entre las que figuran France Télécom y el fabricante de televisores Thomson, y el propio Estado francés. El grupo trabaja en un buscador avanzado que incluya herramientas para la identificación de términos en Internet e integrar en índices imágenes, archivos audiovisuales y texto.

En cuanto a la hegemonía que ejerce Google con relación al acceso al conocimiento en la Web, el Presidente de la Biblioteca Nacional Francesa -Jean-Noël Jeanneney- ha manifestado que “no se le puede dejar a Google todo el manejo de la cultura”. Este comentario en cuanto a la digitalización de libros por parte del buscador mediante su sistema Google Search Books (creado en 2004). La crítica radica en por qué algunos libros son digitalizados y otros no (desde la perspectiva cultural norteamericana) mediante Google, otorgando muchas veces solo un punto de vista acerca de un tema. Además, los resultados de las búsquedas proporcionan mayoritariamente obras en inglés. También se critica el hecho de que los resultados de búsquedas están viciados de contenidos publicitarios.

Como contrapartida al Google Search Books surgió la Biblioteca Digital Europea. Busca promover el acceso de forma amplia y organizada al patrimonio cultural y científico de la Unión Europea en las redes informáticas mundiales. Nació en el año 2005, aglutinando los fondos de diferentes bibliotecas europeas en veinte lenguas. La misma fue fomentada por intelectuales y políticos de Europa, alentados por la adopción de la declaración universal de la UNESCO sobre la diversidad cultural. El objetivo es contrarrestar la posibilidad de que Google Search Books realice un monopolio cultural norteamericano, garantizando la mayor oferta y diversidad posible.

En julio de 2008, Cuil se presentó como el rival directo de Google. Un nuevo buscador que puede investigar unas 121,617,892,992 páginas, según prometen sus creadores (ex ingenieros de Google). El servicio asegura que puede indexar la red a una décima parte del costo que le lleva a Google, argumento principal para conseguir una inversión inicial de 33 millones de dólares. El método de búsqueda es novedoso: a diferencia de Google, que ofrece un ranking de resultados ordenados, entre otras cosas, por la "popularidad" (page rank) y la cantidad de sitios que enlazan a una página, Cuil propone centrarse en los

contenidos. El buscador analiza cada página en contexto, con palabras clave relacionadas entre sí, para devolver resultados ordenados de forma semántica, por tema, que sean más relevantes al usuario. Otra ventaja: el sitio promete que no recopila ni guarda información alguna sobre sus usuarios o sus búsquedas, algo que sí hace Google y por lo cual recibió muchas críticas.

Las búsquedas en español de Cuil son muy escasas lo que supone que han volcado sus recursos en el mercado anglosajón. En cuanto a su funcionamiento fue criticado por resultados de búsqueda laxos y porque el mismo día de su presentación el servidor colapsó.

Para explorar en la Web invisible existen los denominados “Submarinos”, la mayor parte de los recursos está en inglés y algunos tienen una interfaz complicada para los usuarios. Entre los más destacados figuran:

Internetinvisible.com (en español): recopila, describe y ofrece los enlaces a las bases de datos españolas gratuitas existentes en Internet. Ofrece, además de un formulario de búsqueda directa simple y avanzada, un directorio organizado en grupos temáticos, los cuales se subdividen a su vez en materias más específicas. Tras la consulta, se obtiene una ficha con la descripción del contenido de la base de datos relacionada, el enlace a la pantalla de búsqueda y la entidad o persona responsable de su creación. Existen enlaces a más de 2.400 bases de datos de acceso gratuito. Esta web pretende ser el directorio de referencia del contexto español (tanto lo producido en el estado español, en cualquiera de sus lenguas oficiales, como lo procedente de otros lugares geográficos de temática o habla hispana). También incluye recursos ajenos cuando constituyen un punto de referencia en su campo de aplicación.

Invisibleweb.com (en inglés): es un directorio que contiene unas 10.000 bases de datos y archivos que contienen información que tradicionalmente no puede ser accedida por los buscadores tradicionales.

Completeplanet.com: “The Deep Web Directory” contiene unas 70.000 bases de datos de la red profunda y otros recursos especializados.

Ftes.

S/D, (2004) [“Google sigue siendo el líder par alas búsquedas en Internet”](#). En: *Cinco días.com*. Sección: Empresas. 08 de junio. Fecha de consulta: 11 de noviembre 2008.

Siri, Laura (2008). [“¿Buscar en la Web o “googlear”?”](#). En: *Blog de la Comisión 2, Taller de introducción a las TICs, Cátedra Beltrán*. 28 de agosto. Fecha de consulta: 01 de noviembre de 2008.

S/D. (2008) [“Google arrincona a sus rivales en el negocio de las búsquedas en Internet”](#). En: *Cinco días.com*. Sección: Empresas. 20 de agosto. Fecha de consulta: 11 de noviembre 2008.

S/D. (2007) [“Google, con un 77.93%, es el buscador mas usado en España”](#). En: *Noticiascadadia*. Sección: Tecnología. 13 de Septiembre. Fecha de consulta: 11 de noviembre de 2008.

S/D (2005). [“Europa se lanza a destronar a Google”](#). En *La Nación*. Sección: Política. 29 de diciembre. Fecha de consulta: 11 de noviembre 2008.

S/D (2008). [“Bruselas aprueba que Francia respalde con 99 millones el buscador Quaero”](#). En *CincoDías.com*. Sección: Empresas. 14 de marzo. Fecha de consulta: 13 de noviembre de 2008

Klew, Willy (2007) [“Theseus, el buscador europeo que competirá con Google”](#). En *Neoteo* 21 de julio. Fecha de consulta: 12 de noviembre de 2008.

Arana, Patricio (2007). [“No se puede dejar a Google todo el manejo de la cultura”](#). En *La Nación*. Sección: Cultura. 17 de enero. Fecha de consulta: 12 de noviembre 2008.

Falduto, Facundo. (2008) [“Cuil: ¿la alternativa a la supremacía de Google?”](#). En *Perfil*. Sección: Tecnología. 29 de julio. Fecha de consulta: 11 de noviembre de 2008.

S/D (2008). [“Buscador Cuil, ¿una alternativa a Google?”](#). En *Tecnocosas*. Sección: General. 14 de agosto. Fecha de consulta: 12 de noviembre de 2008.

TS8

Los diez links más relevantes sobre la Web invisible

Copete A continuación se presentan diez links que permitirán al lector ampliar su conocimiento en cuanto al tema de estudio

Texto

Hipertexto: Tesis doctoral de María Jesús Lamarca Lapuente. Universidad Complutense de Madrid. Doctorado en Fundamentos, Metodología y Aplicaciones de las Tecnologías Documentales y Procesamiento de la Información. Esta tesis presenta información detallada sobre la Web invisible, tipos de buscadores, sus funcionamientos y otros temas relevantes al objeto de estudio.

Web Invisible: Documento de la Universidad de Berkeley, California. Allí se explica que comprende la Web invisible, cómo funcionan los buscadores, por qué no indexan ciertas páginas. Posee ejemplos de páginas no indexables por los motores de búsqueda.

Estrategias y mecanismos de búsqueda en la Web invisible: Trabajo publicado en el portal del Colegio de México. Posee una descripción de la Web invisible, sus características, y la forma de explorarla. Contiene un listado de recursos de búsqueda en Internet invisible.

[¿Buscar en la Web o “googlear”?](#): Nota periodística realizada por Laura Siri para la revista *Limen*. Posee una gran cantidad de estadísticas acerca de cuales son los buscadores mas utilizados mundialmente. Contiene un apartado donde se explica puntualmente el fenómeno de la Web invisible. Hacia el final de la nota se enumeran consejos de búsqueda para que el usuario pueda tener un mejor acceso a la información.

[La red invisible](#): Nota periodística publicada en el diario *La Gaceta*. Explica con términos claros y sencillos para el lector el fenómeno de la Web invisible.

[La generación Google cambia la forma de acceder al conocimiento](#): Nota periodística que refleja cómo están cambiando las formas de acceder al conocimiento en relación al pasado. La generación Google (aquellos nacidos desde 1993 en adelante) utiliza Internet como herramienta de búsqueda y conocimiento, a diferencia de las generaciones anteriores que utilizaban las bibliotecas.

[Internet invisible: su importancia para las bibliotecas universitarias](#): Diapositivas explicativas de la Internet invisible. Las mismas abordan: tamaño, clasificación, contenidos, ejemplos de sitios. Posee consejos para adquirir información de calidad en la Red.

[El libro de medio billón de páginas](#): Análisis de la Web invisible llevado a cabo por José Antonio Millán (lingüista y escritor español). Entrelaza el fenómeno estudiado con el dominio de la lengua inglesa por sobre las demás en la Web.

[Internetinvisible.com](#) : Es un directorio temático con más de 2.800 bases de datos. Posee una explicación acerca de qué es la Web invisible. Herramienta de habla hispana que permite “bucear” la Red invisible.

[99 Resources to Research and Mine the invisible Web](#): Listado de 99 herramientas que permiten explorar los contenidos de la Web invisible.

Ftes.

Las fuentes están linkeadas en el texto del TS

TS9

Entrevista con Mario Nicotra, DBA manager de Claro Argentina

Copete Mario Nicotra se desempeña como Jefe de grupo de DBA (database administration) en Claro Argentina. Es responsable de: la administración de las bases de datos, SAP basis, DW ETL, la administración de servidores de aplicaciones, la aplicación de cambios en el entorno de producción, y de la arquitectura de aplicaciones. Ingreso a la empresa en mayo de 1995 (cuando la razón social era CTI). Ese mismo años se recibió de Analista IT en la Universidad Blas Pascal.

Texto

¿Qué opina del fenómeno de la Web invisible?

La idea original fue la creación de una red que permitiera compartir información entre las distintas computadoras de una red corporativa. La Web invisible existe desde la creación misma de la Web. Es lógico que las compañías quieran compartir información sólo con sus empleados y evitar que sea difundida públicamente. Este tipo de redes corporativas se denominan "intranets", es decir, redes internas. Y es inútil guardar la información "dinámica" porque, justamente cuando la buscáramos, quizás haya cambiado.

¿Existe alguna regulación en cuanto al fenómeno de la Web invisible?

No existe ninguna regulación ni sobre la Web invisible ni sobre la Web en si, su uso, información a compartir, etc. Mientras no se viole el copyright de contenidos existentes, se tiene la libertad de publicar lo que se desee. Algunos gobiernos regulan o prohíben acceder a ciertos sitios, pero son regulaciones "regionales" (como en el caso de China, donde no se permiten acceder a sitios como youtube).

¿En algún momento los buscadores podrán indexar la totalidad de las páginas de la Web?

Es prácticamente imposible que se indexen la totalidad de las páginas Web. Por empezar, no es accesible (ni las corporaciones lo permitirían por seguridad y confidencialidad) acceder a información de sus intranets, ni indexar páginas dinámicas ya que varían constantemente. Independientemente de estos comportamientos, hay miles de millones de usuarios creando contenido, actualizando sus blogs, guardando fotos, videos, etc. Los robots automáticos que indexan este contenido pueden procesar millones de páginas por día, pero usualmente una nueva página tarda entre una semana y diez días en ser indexada.

¿Qué opina acerca de que los buscadores administran el conocimiento?

Esta práctica se está volviendo una de las más "peligrosas". Por definición, si una búsqueda mediante buscadores conocidos (Google, Yahoo, MSN Search) no arroja resultados positivos, se dice que no hay información sobre el tema. Eso es tan o más peligroso que creer que la información provista por diferentes sitios es real. Googlear se ha vuelto un verbo. Que el conocimiento sea administrado por una empresa privada siempre tiene sus riesgos, aunque se muestre transparente y ofrezca sus servicios "gratuitamente". Ninguna empresa privada vive del aire, ni Google, Yahoo o MSN Search son la excepción.

Google es el buscador más consultado ¿El "acceso al conocimiento" mediante la Web sólo se limita a Google?

Ésta es una de las afirmaciones más comunes. Google es una herramienta excelente y muy poderosa. Pero no todo el conocimiento pasa por Google. Uno de los proyectos más innovadores es el buscador Wikia (<http://search.wikia.com>), creado por el fundador e ideólogo de Wikipedia. Bajo el mismo concepto de Wikipedia, en Wikia los resultados de

las búsquedas los aportan, clasifican y editan los propios usuarios, descentralizando el conocimiento en los propios usuarios.

¿Cuales piensa que son los motivos de que Google sea el motor de búsqueda más utilizado?

La indexación de los sitios de Internet se hace mediante robots que recorren la Web millones de veces por día, almacenando esta información en un sistema de archivos creado especialmente por Google, lo que permite almacenar millones de terabytes de información por día. La lógica que permite a esos robots recorrer la Web y procesar la información fue la tesis de grado de Larry Page y Sergey Brin, los fundadores de Google. Su idea fue innovadora y revolucionaria, permitiendo llegar a Google a la compañía que es hoy. Independientemente de la forma en que Google procesa la información, uno de los conceptos que más agradaron a los usuarios fue la interfaz de usuario de Google. Los viejos sitios de búsqueda presentaban una interfaz cargada de enlaces, propagandas gráficas destellando y detalles periféricos que distraían y molestaban al usuario. Google presento una página inicial de búsqueda con una caja de texto y su logo; y los resultados de las búsquedas poseían propagandas a la derecha, sólo en texto y que no interferían en la lectura de los resultados. Gracias a esto, a los usuarios les resulta "visualmente agradable" utilizar Google para buscar información.

¿Existe alguna alternativa, que quizás por falta de difusión u otras cuestiones, supere a Google en los resultados de búsqueda?

Al día de hoy no existe una alternativa real que supere a Google en búsquedas de información. Para búsquedas de temas puntuales (documentación científica, de leyes, reglamentaciones, historia, etc.) existen buscadores especializados que superan a Google sólo en esos temas. Pero son casos aislados, que representan un porcentaje pequeño en la cantidad de búsquedas que se realizan a nivel general.

¿Cree que la falta de instrucción avanzada hace que la gente elija a Google o realmente es el buscador más completo?

Es un buscador muy bueno y muy completo. Pero se ha vuelto una costumbre para los usuarios. Adicionalmente, los navegadores más utilizados firman acuerdos con Google e incorporan una caja de búsqueda de Google a su interfaz, lo que hace que los usuarios realicen sus búsquedas utilizando esta facilidad sin preocuparse de buscar otras alternativas.

¿Conoce el buscador Quaero?

Lo he oído nombrar y leí algunos artículos. Siempre es bueno crear alternativas y evitar que se crea que toda la información debe pasar por un sólo lugar. Cuanta más competencia haya, mejores resultados y fuentes obtendremos los usuarios al realizar búsquedas.

¿Cuál cree que es la causa por la que los motores de búsqueda han desplazado a los directorios temáticos?

Los buscadores respetan el principio de la Web, cosa que los directorios temáticos no hacían. La Web es dinámica. Cambia y se actualiza todo el tiempo. Los directorios temáticos dependían de la intervención de operadores o usuarios externos que clasificaran y almacenaran la información. Con el volumen de información que se maneja hoy día, es prácticamente imposible procesar esa información manualmente.

Conclusiones

Existe una relación directamente proporcional entre el conocimiento técnico sobre la web y la libertad al acceso de la información en la misma. Solo accederán a la información que está disponible en la Red invisible aquellos que sepan como desenvolverse en ella, o tengan conocimiento de su existencia. Conocimiento que también se refiere a la multiplicidad de herramientas que están a disposición del usuario pero que no tienen una gran difusión.

La formula sería a mayor conocimiento mayor libertad. Confieso que en mi caso particular desconocía de la existencia de la Web invisible y las herramientas para desenvolverse en ella. Justamente por ese desconocimiento y como usuario de Google, quedaba “preso” de las ofertas del mismo en materia de búsquedas.

Desde la perspectiva de Diego De Charras ”La utilización de la red implica la posesión de un capital económico que garantice la infraestructura básica para poder acceder [...].Pero además, el acceso a la información no solo está limitado por los elementos descriptos sino que se le agrega el principal escollo que es que para poder navegar desde un ordenador es preciso poseer una capital cultural que permita, por un lado manejar una PC y, por otro, poder ordenar, procesar y seleccionar la información disponible.” Esta frase de De charras se ajusta perfectamente al tema estudiado. Justamente ese saber, ese capital cultural, que permite ordenar, procesar y seleccionar la información disponible en la Web, es el que permite un pleno y libre acceso a la información en la misma. Las personas que no posean ese capital cultural quedaran atadas a la “dominación” hegemónica de las ofertas de Google.

Bourdieu sostiene que “la distribución desigual del capital cultural reproduce y realimenta la estructura de dominación económica y la estratificación social.” Trazando un paralelismo al tema estudiado, puede afirmarse que aquellos que no posean el capital cultural quedarán bajo la “dominación” de Google, solo pudiendo acceder a las páginas que este recupere; mientras que aquellos que poseen el capital cultural podrán acceder “libremente” a la información que solicitan (mediante la utilización de otras herramientas). Quedando así marcados dos estratos, los que poseen el capital cultural y “libertad” de acceso a la información, y aquellos que no poseen el capital cultural y su acceso a la información en la Web es “administrado” por Google (mediante las paginas que este indexe).

En contra de los augurios de Castells que anunciaba una Sociedad Red, basada en el conocimiento, donde el tráfico de informaciones de una computadora a otra (a escalas mundiales) generaría una sociedad democrática e igualitaria, se observa una suerte de brecha informacional entre aquellos que poseen el capital cultural y quienes no lo tienen.

Si bien los artículos citados en el TS4 no hacen referencia explícita a la Web, impactan en ella de manera “indirecta”. La Red está conformada por hombres que pertenecen tanto al mundo como a nuestro país. Llegada esta etapa del trabajo puede afirmarse que no existe un libre y pleno acceso al derecho de la información, que consta en esos artículos, en la Web. Si bien en cuanto a la libertad de expresión la Red ha sido un aporte significativo, luego de realizar esta investigación no puede decirse lo mismo en términos de acceso informacional; debido a que la “exposición” de la misma no es equitativa en la Web

El page rank de Google refleja una forma típica de actuar de las Industrias Culturales. Google evalúa la calidad de los sitios desde la cantidad de visitas que reciben. Por lo tanto las cualidades se cuantifican. Una página adquiere valor por la cantidad de visitas que posee y no por su contenido. Y esto porque Google funciona en torno a la búsqueda de ganancias, mediante la publicidad, pudiendo llegar así a la mayor cantidad de personas posibles.

Las alternativas que desde Europa se le presentan a Google, enumeradas en el TS7 no pueden hacerle frente ya que no poseen la difusión mundial que tiene el buscador más utilizado del mundo. Google es tan poderoso debido a los millones de dólares que maneja en publicidad, además de poseer mejores herramientas que los motores de búsqueda competidores (Yahoo, MSN Live Search, etc.) La aparición de publicidad en el buscador de textos de Google se relaciona con la idea de Diego De Charras. Desde su perspectiva, los sitios comerciales existen gracias a las páginas Web gratuitas, debido a que para que pueda haber sitios comerciales es necesaria una cierta cantidad de usuarios, que difícilmente podría crecer a la velocidad que lo hace en el caso que los sitios fueran pagos. Muchos lugares gratuitos sirven para hacer publicidad de los que no lo son. Esa gratuidad es funcional al capital económico. Los sitios que se enfrentan a la hegemonía, tecnológica y cultural, norteamericana no logran insertarse en la lógica de la Industria Cultural. Ya que si bien reciben millones de dólares por parte de los Estados y empresas, al no explotar la publicidad de la forma en que lo hace Google no entran en la lógica de mercado y quedan relegados a segundo plano.

Google se presenta como el administrador del conocimiento de los usuarios comunes. Las herramientas de búsqueda en la Web invisible quedan relegadas a la profundidad debido a la falta de difusión, no sólo de las mismas sino también de la Web invisible.

Ftes.

Castells, Manuel (1999). *La era de la información*. Vol. 1. *La sociedad red*. México: Siglo XXI.

De Charras, Diego; (2006), *Redes, burbujas y promesas*. Prometeo, Buenos Aires.