



4ª Jornada sobre la Biblioteca Digital Universitaria JBDU 2006

“La Web y las bibliotecas digitales en el ámbito académico: acceso, interacción, visibilidad”

“INTERNET INVISIBLE” SU IMPORTANCIA PARA LAS BIBLIOTECAS UNIVERSITARIAS

Adrián Néstor Curti

1. Resumen

En este trabajo se describe a la *Internet*, en general y a la *Internet Invisible* en particular; consignando las principales características cualitativas y cuantitativas de cada una y reseñando la génesis del concepto de ésta última. Asimismo se hace una reseña histórica de los buscadores y de la incidencia de éstos a la hora de recuperar información; brindándose algunos consejos de búsqueda. Se establecen tipologías de buscadores, su estructura y su funcionamiento. Se clasifica a la Web invisible según sus contenidos para poder identificar cuáles son los sitios accesibles y útiles a las bibliotecas universitarias. Se espera que el presente documento se utilice como punto de partida para la investigación y exploración de esa parte de internet que no “ven” los buscadores y que, sin embargo, resultan de vital importancia para las bibliotecas académicas.

1.1 Internet. La WWW.

La web (WWW) no es sinónimo de Internet sino, apenas, uno de sus servicios y con 12 mil millones de páginas web y cerca de dos millones de páginas creadas o actualizadas diariamente, encontrar información relevante en Internet es una tarea cada vez más dificultosa (<http://www.sociedadelainformacion.com/20011103/invisible/internetprofundo.htm>).

Existen dos grupos de contenidos en la web. El primero es la web visible o superficial y se corresponde con la versión tradicional de la web. Se trata de las páginas estáticas, disponibles de forma pública y abierta y es, en rigor, la parte más pequeña de la web conocida.

El segundo grupo está conformado por la web invisible o profunda, que consiste en bases de datos y sitios dinámicos desconocidos por la mayoría de quienes navegan, aunque la información existente en estas direcciones es entre 400 y 550 veces mayor que la información disponible en la web superficial. (www.absysnet.com, 2005)

A fines del año 2000 la web superficial tenía alrededor de 2.500 millones de documentos que, cuatro años antes, contaba con menos de la mitad, 1.000 millones. Su tasa de crecimiento era en esos años de 7,3 millones de páginas por día (0,1 terabytes {Un terabyte (tb) equivale a 1.000 millones de bytes}). El tamaño promedio de las páginas oscilaba entre los 10 y los 20 kb. Así, el tamaño total de la web de superficie oscilaba entre los 25 y los 50 terabytes de información, de los cuales correspondían a información textual entre 10 y 20. (Piscitelli, 2005)

Por otro lado, con respecto a la web profunda, si se toma como base de medida todas las bases de datos conectadas, las páginas dinámicas (foros, weblogs, catálogos, sitios de remates) y las intranets de las organizaciones, hay cerca de 550.000 millones de documentos web, con una página promedio de 14 kb. Esto suma la impresionante cantidad de 7.500 terabytes.

Dos de los sitios profundos más grandes del mundo son las bases de datos del National Climatic Data Center y de la NASA. Entre las dos, contienen un total de 585 terabytes de información, prácticamente el 7,8% de la web profunda. Solo 60 de los sitios web más grandes almacenan 750 terabytes de información, lo que supone el 10% de la web profunda. (Piscitelli, 2005)

1.2 Internet Invisible

“La búsqueda de información en la red se ha convertido, con la proliferación de buscadores, en un ritual que alcanza a casi cualquier ámbito de nuestra vida cotidiana. Viajar, comprar, participar activamente en foros de discusión, mantenerse informado de la actualidad, resolver cuestiones laborales, satisfacer puras curiosidades y un largo etcétera que cada usuario configura según sus intereses personales. Todas estas búsquedas obtienen su satisfacción como resultado del potencial indizador de los motores de búsqueda que constantemente rastrean las páginas del universo de la red. Pero éste sólo es el universo VISIBLE de Internet. Hay otro, cuyas últimas estimaciones calculan en 500 veces más grande que el total de la información indizada por los buscadores, que permanece INVISIBLE” (internetinvisible.com, 2005)

En 1994, la Dra. Jill Ellsworth, especializada en el estudio de la internet, acuñó el término “internet invisible” para referirse a la información que no podían encontrar los buscadores más comunes (Ellsworth, 1995). En el año 2000, un estudio elaborado por Michael Bergman para la consultora BrightPlanet (Bergman, 2001), confirmaba la existencia de una “red profunda” de aproximadamente unos 7500 terabytes (7500 billones de bytes) frente a los 19 de la web visible accesible por medio de los buscadores convencionales. Aunque hubieron estudios posteriores en los cuales las cifras eran menores: entre 2 y 50 veces mayor (Sherman y Price, 2001). En todo caso, el valor de la información en la internet invisible es muy importante y merece su estudio.

Esta internet invisible está compuesta principalmente por bases de datos y documentos no indexados por los buscadores. Esto no significa que sea imposible acceder a estos documentos. Conocer el funcionamiento de los motores de búsqueda es fundamental para comprender el porqué de la Internet Invisible, ya que la supuesta invisibilidad de la información deriva de la tecnología y política de los buscadores para indexar los sitios porque cada buscador posee una gran base de datos y es allí donde busca ante una consulta y no en toda la web.

1.2.1 Clasificación

Sherman y Price (2001) categorizaron cuatro tipos de invisibilidad: la web opaca (*opaque web*), la web privada (*private web*), la web propietaria (*proprietary web*) y la web realmente invisible (*truly invisible web*)

a) La web opaca

Páginas que pueden ser indexadas, pero no son incluidas en los buscadores. Los motivos para que los buscadores “decidan” no incluirlas pueden ser:

- Profundidad de exploración: los sitios tienen “profundidad”. La home page o página principal es el primer nivel; ahí llegan todos los buscadores. De ahí se linkean a páginas internas del sitio, ése sería el segundo nivel, al que no llegan los Directorios y algunos motores de búsqueda. Esas páginas, a su vez, enlazan con páginas “más” internas, que no estaban en la home page. A este nivel llegan muy pocos buscadores. Cuanto más profundo sea el nivel, menos buscadores lo indexarán.
- Frecuencia de exploración: un sitio puede cambiar todos los días, pero muchos de los robots de los buscadores que exploran los sitios los visitan una vez por mes o menos por una cuestión de costos. Todos los cambios entre una visita y otra no figuran en los buscadores.
- Supera el número máximo de resultados: cada buscador define qué cantidad de páginas de un sitio mostrará. Si un sitio tiene más páginas que las que el buscador incluye, las restantes quedarán sin indexar.
- Errores de exploración: puede haber un problema en el sitio o en el robot del buscador (o en la compatibilidad entre ambos) que impida que una página (o hasta un sitio completo) sea incluida en la base de datos del buscador.

b) La web privada

Sitios completos o páginas individuales que son técnicamente indexables por el buscador, pero que fueron excluidos por el dueño del sitio. Se trata de información no pública (bases de datos de una empresa para exclusivo uso interno, discos virtuales, intranets, etc.). Eligieron la web como medio por todas sus ventajas, pero no tienen ningún interés en que la información que allí almacenan se haga pública. En esos casos se dice que es un *sitio cerrado*.

Técnicas de protección: Alguien puede no mandar su sitio a un buscador para ser indizado y el buscador puede indizarlo igual. Por eso, los sitios se valen de diferentes técnicas de protección para que ni los buscadores, ni los usuarios no registrados puedan acceder a información protegida. Entre las técnicas de protección de la información, las siguientes son las tres más frecuentes:

- Password: el sitio está cerrado y sólo puede consultarlo aquél que tenga nombre de usuario y contraseña.
- Robots.txt: es un archivo de texto que elabora cada sitio que le indica a cada buscador que páginas debe indexar y cuáles no.
- Noindex metatag: similar al anterior, pero es un comando que se pone en cada página que no se quiere indexar. Si el buscador la detecta, no la incorpora en su base de datos.

c) La web propietaria

Sitios accesibles solo para aquellos que se registran: muchos foros de discusión permiten el acceso sólo a los usuarios registrados. En esos casos se dice que es un *sitio semiabierto*. Como los robots de los buscadores no pueden (ni quieren) “llenar un formulario de registración” para acceder a la información, no es incorporada en los buscadores.

En buena parte de la web propietaria los contenidos se cierran con una finalidad económica: hay que pagar una tarifa por tiempo (día, año) o por contenido (artículo, ejemplar de un diario) para acceder a la información.

Tanto en la web propietaria como en la privada los sitios protegidos no son indexados por los buscadores, pero en la web propietaria se puede acceder a esta información por otras vías (registrándonos, pagando), mientras que en la privada no.

d) La web *realmente* invisible

Este contenido no puede ser indexado por los buscadores por razones técnicas. Los documentos pueden estar en un formato que los robots no reconozcan (música, videos) o por páginas generadas dinámicamente (la página se autodiseña, no hay ningun diseñador humano de por medio, se genera sola). Ese tipo de páginas (foros de discusión, sitios de remates, catálogos, diccionarios, etc) no son tenidas en cuenta por los buscadores.

2. Buscadores. Reseña histórica

Los buscadores de grandes colecciones de datos son anteriores a la web por muchas décadas. Antes de la llegada de la World Wide Web existían catálogos informatizados en muchas bibliotecas. También los profesionales de la información recurrían a servicios de búsqueda como Dialog o LexisNexis, que todavía siguen existiendo. Buscar datos a través de estos

servicios no era una tarea sencilla, muchos eran pagos, muy pocos permitían el acceso a distancia y eran contados los que ofrecían información en español.

Con la aparición de los BBS (sigla de Bulletin Board System, que permitía acceder rudimentariamente a datos remotos usando un modem) el acceso a documentos desde una computadora hogareña se simplificó bastante. Las bocas de producción de documentos electrónicos se multiplicó, así como la cantidad de usuarios con computadora que intercambiaban información. Sin embargo, cada BBS actuaba de manera autónoma (década del '80 y los primeros '90.)

Y entonces llegó la web. Para bien y para mal. Para bien, porque Internet conectó a todas las máquinas que producían (y recababan) información, porque abarató y democratizó la producción y recolección de información y porque la cantidad de información creció (y sigue creciendo) en proporciones antes nunca imaginadas. Pero también para mal, porque ahora la información no estaba toda en el mismo formato, porque no necesariamente era verdadera, porque podía estar desactualizada, o con errores.

En ese contexto aparecieron los buscadores de Internet, para tratar de ponerle un poco de sentido a todo ese inabarcable océano de información online. Los buscadores evolucionaron rápidamente, tratando de ayudar, cada vez mejor, a organizar los miles de millones de documentos que se producen.

2.1 Tipología. Estructura

A grandes rasgos, hay dos tipos de buscadores en Internet: los directorios y los motores de búsqueda.

Los directorios son buscadores organizados a partir de una jerarquía temática (taxonomía). El más conocido de los directorios es Yahoo!. Se puede navegar un directorio adentrándose en sus categorías y subcategorías o ingresando una palabra clave que mostrará las distintas categorías y sitios en los que esa palabra aparece. Hay directorios generalistas -como Yahoo!- y especializados -como Ariadna-, buscador de recursos periodísticos (<http://www.periodismo.com/buscador/>). Aunque la mayoría de los buscadores generalistas imitan la taxonomía de Yahoo! no hay un estándar en este sentido. Tampoco hay ningún tipo de homogeneidad o criterio común entre los buscadores especializados. El directorio muestra los resultados de su búsqueda basándose únicamente en el título y la descripción del sitio. Además, los directorios incluyen sitios completos, no páginas o secciones dentro de un sitio. Otra característica de los directorios es que las páginas son revisadas por seres humanos e incorporadas sólo si cumplen con los criterios de pertinencia de ese buscador. Esto hace que la cantidad de sitios de los Directorios sea pequeña en comparación con la totalidad de sitios existentes.

El otro tipo de buscadores son los motores de búsqueda. El más conocido actualmente es Google (www.google.com). Los motores de búsqueda no tienen una taxonomía, se puede

acceder a los resultados sólo a partir de una palabra clave. Todo el proceso de indizado de páginas es automático, no hay seres humanos revisando sitio por sitio. A diferencia de los directorios, los motores de búsqueda buscan en toda la página de un sitio web, no se limitan al título y la descripción. Si bien no indizan todas las páginas de un sitio, tampoco se limitan a indizar una sola página. La cantidad de páginas indizadas es enorme: al momento de escribir esto Google tenía indizadas más de 8.000.000.000 páginas web. (<http://www.google.com.ar/intl/es/corporate/>, 2006)

Los motores de búsqueda tienen robots buscadores que exploran los sitios web y los incorporan a sus bases de datos. A esta acción se la llama indizar o indexar (la Real Academia acepta los dos términos).

En la actualidad muchos buscadores combinan la potencia de los motores de búsqueda con la lógica de los directorios. Si Yahoo! no encuentra resultados en su directorio, muestra los resultados de su motor de búsqueda. Pero también los motores de búsqueda recurren a los directorios: para quién necesite resultados más ordenados, Google usa los datos del directorio Dmoz, también conocido como Open Directory. (<http://dmoz.org/>)

2.2 Cómo funcionan

Robots

En una escala microscópica los robots del siglo XXI son invisibles e inmateriales. Estos robots se dedican a hacer algo clave: indexar las páginas que se visitan.. (Piscitelli, 2005).

Hoy la situación es muy diferente de hace cuatro o cinco años atrás: en ese momento, la fe en los robots buscadores, hacía suponer que si buscadores como *Altavista* o *Hotbot* no encontraban aquello que se buscaba, era sencillamente porque tal información no existía en la red.

Sin embargo, se empezó a dar importancia a la calidad por encima de la cantidad. Esto determinó que era preferible indexar sitios de calidad, antes que apilar meramente la mayor cantidad de sitios posibles, acudiendo a los buscadores. El universo web estaba lleno de páginas que no valía la pena visitar nunca.

En ese momento interesaba el área del aprendizaje robótico y se construyó un robot llamado *Inquirus*, capaz de interrogar a otros robots acerca de la existencia de documentos que cumplieran con cierta estructura de búsqueda; este robot podía arrojar un beneficio secundario más valioso que el buscado originalmente estimando el tamaño real de la red, un número que en ese momento nadie conocía a ciencia cierta. Entre los resultados que cosechó el *Inquirus* aplicado al buscador Hotbot, fue descubrir, en 1997, que la web contaba con alrededor de 320 millones de documentos (por lo menos el doble de lo que se creía entonces) .Y no solo eso: Hotbot se preciaba de ser el más exitoso y exigente de los robots en aquella época pero, de pronto, se vio devaluado al descubrirse que solo indexaba el 34% de toda la web. Como premio de consuelo, pudo jactarse de que a los otros robots les iba aún peor: Altavista solo cubría un 28%, y otros buscadores —como Lycos, que pronto caería en manos de Terra y Telefónica— apenas cubrían un 2% de la red.

En febrero de 1999, cuando se repitió el mismo ejercicio, los investigadores encontraron que la red había crecido (tenía 800 millones de documentos) pero la capacidad de indexación de sitios a manos de los robots había empeorado.

Un excelente buscador de la época, Northern Light, ocupó entonces la *pole position* cubriendo el 16% de la web, pero Altavista había bajado al 15% y Hotbot reseñaba apenas el 11% de las páginas existentes. Mientras tanto Google, que era un benjamín entre los pesos pesados,

apenas veía en ese entonces un 7,8% de las páginas estimadas. En junio de 2001 Google cubrió por primera vez 1.000 millones de documentos, seguido de cerca por Alltheweb. Hoy, Google está cerca de alcanzar los 9.000 millones de documentos.

Por muy impresionante que sea la capacidad de indexación de los motores, la web crece infinitamente más rápido que la posibilidad de éstos de analizarla. Además, existe la web profunda que es por lo menos 550 veces más grande que la que los robots alcanzan, por lo que la asimetría entre lo visible y lo existente se ahonda muchísimo más.

En el año 2000, 6 de cada 10 páginas no habían sido visitadas nunca. Hoy los resultados deben alcanzar cifras de entre 8 y 9 de cada 10.

Los buscadores en su rol de “gatekeeper”

El noción de gatekeeping (cuidado de la puerta o del acceso) investiga la manera irregular en que las informaciones circulan y se encuentran sometidas a instancias que las demoran o “traban” en algún punto de la cadena comunicacional, y la fluidez con que circulan luego aquellas que consiguen pasar la barrera. Estos lugares de demora o nudos que actúan como barrera y filtro en la circulación de la información serían los gatekeepers o porteros (“arquero” en el fútbol). El concepto de “gatekeeper” fue introducido por el psicólogo Kurt Lewin en 1947 trabajando en dinámica de grupos, y observó que la información circulaba de una manera muy irregular, ya que por momentos podía interrumpirse debido a los nudos o fluir de manera muy amplia después de superarlos.

Hay que imaginar al buscador como a un gatekeeper: dentro del universo de todas las páginas de la web, el buscador tiene el poder de orientar en el camino hacia la búsqueda de la información.

Ya se sabe que los buscadores no indexan todas las páginas de la web. Ahí ya hay una primera selección. El gatekeeper-buscador deja afuera de sus resultados una enorme cantidad de contenido. La segunda selección está en la relevancia: el buscador define que determinadas páginas son más importantes que otras. Y que este criterio, aunque automatizado, es “subjetivo” (valga la paradoja) se comprueba fácilmente comparando los resultados de los distintos buscadores.

¿Cómo desafiar estos criterios? El segundo criterio es más fácil de burlar: utilizando diversos buscadores y directorios se puede llegar a una “intersubjetividad de resultados”. Metabuscadore como Vivisimo (<http://vivisimo.com/>) le muestran al usuario las diferentes posiciones en que cada página figura en los distintos buscadores.

Desafiar la lógica del buscador en cuanto a los contenidos que deja afuera de sus resultados, lleva a internarse en la llamada Internet Invisible.

2.3 Sugerencias

Cómo buscar

El primer paso en una búsqueda es saber qué es lo que se busca. No necesariamente hay que saberlo con precisión. Puede interesar, puntualmente, encontrar la bandera de Rusia o, más vagamente, hallar legislación sobre jubilación privada en América latina.

Luego que se pudo conceptualizar lo que se va a buscar, se sabrá hacia dónde ir. Si la búsqueda es más específica, se empezará con un motor de búsqueda que conduzca hacia el sitio que se necesita. Si es más general, sería conveniente empezar por un directorio que agrupe a todos los sitios comunes al tema que se investiga.

Si lo que se necesita es local o regional, habrá que restringir las búsquedas a esos países o regiones o, mejor, consultar buscadores de la zona en cuestión.

En el mismo sentido, si la temática es específica, habría que partir de un buscador generalista, buscar allí un directorio temático y realizar en el directorio temático una segunda búsqueda, más acotada.

Cada buscador tiene sus reglas (sintaxis) y por eso es recomendable leer la documentación y páginas de ayuda para entender bien sus opciones de búsqueda.

La mayoría de los buscadores aceptan los operadores booleanos. Estos son AND, OR y AND NOT (esto último en algunos buscadores, funciona al poner solamente NOT o colocando el signo -).

Por defecto, la mayoría de los buscadores funciona con el operador AND o +. Esto quiere decir que poner en un buscador las palabras *mapa argentina* o mapa AND argentina o mapa + argentina es equivalente. Esta búsqueda traerá sólo los documentos que contengan ambas palabras.

Colocando la palabra OR, mostrará los documentos que contengan al menos una de esas palabras. Por ejemplo, si se pone argentina OR uruguay mostrará las páginas que contengan las palabras argentina, las páginas que contengan la palabra uruguay (y en consecuencia también las que contengan ambas palabras). El operador OR es también útil si no se sabe como se escribe una palabra (volswagen OR volkswagen): traerá documentos que contengan al menos una de las palabras.

El operador NOT o el signo - excluye palabras de la página de resultados. “Cindy Crawford” -sex -porn -adult -xxx -nude traerá documentos que que mencionen a la supermodelo, pero no contenido pornográfico. A esto se llama “filtrar” o “refinar” una búsqueda. También puede usarse si, por ejemplo, se quiere información solamente sobre Windows 95 (habrá que poner windows + 95 -98 -3.1).

Las comillas (como en el ejemplo anterior) sirven para denotar una frase exacta: términos que uno sabe que deben ir juntos como el título de un libro o de una película o de una

canción o de un juego. Hay que tener la certeza que se escribe de ese modo, porque si no, ignorará el pedido.

Los operadores AND y OR deben escribirse con mayúsculas.

Todos estos operadores pueden ser muy útiles usándose en forma combinada. Así, por ejemplo, si se quiere encontrar todas las páginas en las que aparezca mencionado Estados Unidos en su forma castellana, excluyendo la grafía en inglés, se podría poner: **“Estados Unidos” OR EE.UU. OR EEUU -“United States” -USA.**

3. Internet invisible, bibliotecas y recuperación de la información. Conclusiones

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 y, estos productos o servicios, se encasillan dentro de lo que Sherman y Price (2001) denominaron *Web Propietaria* y *Web realmente invisible*.

Atendiendo a criterios documentales, se podría agrupar dentro de la Internet invisible los siguientes sitios:

- Catálogos de bibliotecas
- Bases de datos de bibliotecas (propiedades 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 indizables, 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.

Además:

- Foros de profesionales (p.e. ABGRA)
- Blogs de asociaciones profesionales (p.e. <http://blogs.ya.com/abprosario/>)

Otros sitios con mucho contenido “invisible” son los organismos públicos de gobierno, ONGs, etc., que en sus sitios web disponen de consultas a sus bases de datos.

Por eso, una buena forma de localizar la información contenida en la Internet invisible, será a través de las secciones de referencia de las bibliotecas virtuales y de sus catálogos en línea. Así como las bases de datos, catálogos y archivos de instituciones educativas, gubernamentales, sociales, etc.

Asimismo, con el auge de la edición de publicaciones de acceso abierto (open access), internet se convirtió en una herramienta esencial para mejorar o brindar un servicio de calidad en la bibliotecas universitarias o especializadas, ya que estas publicaciones de libre acceso y a texto completo cubren muchas áreas del conocimiento.

Bibliografía

- Bender, Cristina [et.al]. Clasificación de páginas *web* como posprocesamiento a la recuperación de la información. En: Víctor M. Castel, Comp. (2005) *Desarrollo, implementación y utilización de modelos para el procesamiento automático de textos*. Mendoza : Facultad de Filosofía y Letras, UNCuyo. Cap. 1.
- Deco, Claudia [et.al]. Expansión de consultas utilizando recursos lingüísticos para mejorar la recuperación de información en la web. En: Víctor M. Castel, Comp. (2005) *Desarrollo, implementación y utilización de modelos para el procesamiento automático de textos*. Mendoza : Facultad de Filosofía y Letras, UNCuyo. Cap. 4.
- Fainholc, Beatriz (2004). *Lectura crítica en internet : Análisis y utilización de los recursos tecnológicos en educación*. Rosario : Homo Sapiens.
- Piscitelli, Alejandro (2005). *Internet, la imprenta del siglo xxi*. Barcelona : Gedisa.

Buscadores y directorios

- Gigablast Inc. **Gigablast** [en línea] <<http://www.gigablast.com/>> [Consulta 28 febrero 2006]
- Periodismo.com. **Ariadna** [en línea] <<http://www.periodismo.com/buscador/>> [Consulta 10 marzo 2006]
- Google. **Google** [en línea] <<http://www.google.com/>> [Consulta 28 febrero 2006]
- Grub Buscador [en línea] <<http://www.grub.org/>> [Consulta: 06 agosto 2005]
- IAC Search & Media. **Ask.com** [en línea] <<http://www.ask.com/>> [Consulta 28 febrero 2006]
- IAC Search & Media. **Excite** [en línea] <<http://www.excite.com/>> [Consulta 28 febrero 2006]
- LookSmart, Ltd. **Wisnut** [en línea] <<http://www.wisnut.com/>> [Consulta 28 febrero 2006]
- Lycos, Inc. **Hotbot** [en línea] <<http://www.hotbot.com/>> [Consulta 28 febrero 2006]

- Lycos, Inc. **Lycos search** [en línea] <<http://www.lycos.com/>> [Consulta 28 febrero 2006]
- Microsoft. **MSN** [en línea] <<http://www.msn.com>> [Consulta 28 febrero 2006]
- Overture Services, Inc. **Alltheweb**, find it all [en línea] <<http://www.alltheweb.com/>> [Consulta 28 febrero 2006]
- Overture Services, Inc. **Altavista** [en línea] <<http://www.altavista.com/>> [Consulta 28 febrero 2006]
- The New York Times Company. **About** [en línea] <<http://www.about.com/>> [Consulta 28 febrero 2006]
- Walt Disney Internet Group (WDIG). **Go.com** [en línea] <<http://go.com/>> [Consulta 28 febrero 2006]
- WebFile.com. **Webfile** [en línea] <<http://www.webfile.com/>> [Consulta 28 febrero 2006]
- Yahoo! Inc. **Yahoo!** [en línea] <<http://www.yahoo.com/>> [Consulta 28 febrero 2006]

Metabuscadores

- Copernic Technologies, Inc. **Copernic** [en línea] <<http://www.copernic.com/>> [Consulta 28 febrero 2006]
- Digital Tsunami, Inc. **Quickfindit** [en línea] <<http://www.quickfindit.com/>> [Consulta 28 febrero 2006]
- Ez2find.com. **ez2finf** [en línea] <<http://ez2find.com/>> [Consulta 28 febrero 2006]
- InfoSpace, Inc. **Dogpile** [en línea] <<http://www.dogpile.com/>> [Consulta 28 febrero 2006]
- InfoSpace, Inc. **Metacrawler** [en línea] <<http://www.metacrawler.com/>> [Consulta 28 febrero 2006]
- Intelliseek, Inc. **ProFusion** [en línea] <<http://www.profusion.com/index.htm>> [Consulta 28 febrero 2006]
- Mamma, Inc. **Mamma** [en línea] <<http://www.mamma.com/>> [Consulta 28 febrero 2006]
- Surfboard BV. **Ixquick** [en línea] <<http://www.ixquick.com/>> [Consulta 28 febrero 2006]
- Vivísimo, Inc. **Vivísimo** [en línea] <<http://vivisimo.com/>> [Consulta 28 febrero 2006]
- Netscape Communications Corporation. **DMOZ** Open Directory Project [en línea] <<http://dmoz.org>> [Consulta 10 marzo 2006]

Buscadores de buscadores

- Multibuscador.com [en línea] <<http://dir.multibuscador.com/>> [Consulta 28 febrero 2006]
- Buscopio [en línea] <<http://www.buscopio.net/esp/>> [Consulta 28 febrero 2006]

Sitios

- @bsysnet.com (2005). Internet Invisible [en línea] <www.absysnet.com> [Consulta 5 marzo 2006]
- Bergman, Michael K (2001). *The Deep Web: Surfacing Hidden Value*. *Bright Planet* [en línea] <<http://www.brightplanet.com/technology/deepweb.asp>> [Consulta: 6 agosto 2005]

- Codina, Lluís. Internet invisible y web semántica: ¿el futuro de los sistemas de información en línea? [en línea] <<http://www.fti.uab.es/tradumatica/revista/num2/articles/>> [Consulta 10 marzo 2006]
- Ellsworth, Jill and Ellsworth, Matthew V (1995). Marketing on the Internet : Multimedia Strategies for the World Wide Web. New York: John Wiley & Sons [Citado por: Moreno Jiménez, Pilar María. Estrategias y mecanismos de búsqueda en la web invisible. Disponible en: http://biblio.colmex.mx/recelec/web_invisible.htm]
- *Internetinvisible.com* [en línea] <<http://www.internetinvisible.com/>> [Consulta: 06 agosto 2005]
- Invisible-web.net. www.invisible-web.net : *The invisible web directory* [en línea] <<http://www.invisible-web.net/>> [Consulta: 06 agosto 2005]
- Merlo Vega, José Antonio. 50 aplicaciones bibliotecarias de internet. [en línea] <<http://exlibris.usal.es/merlo/escritos/pdf/aplica.pdf>> [Consulta 10 marzo 2006]
- Merlo Vega, José Antonio. El servicio de internet en la biblioteca pública [en línea] <<http://exlibris.usal.es/merlo/escritos/pdf/oviedo.pdf>> [Consulta 10 marzo 2006]
- Merlo Vega, José Antonio. Las bibliotecas como clientes y servidoras de información web [en línea] <<http://exlibris.usal.es/merlo/escritos/cliser.htm>>
- Moreno Jiménez, Pilar María (2005). *Estrategias y mecanismos de búsqueda en la web invisible* [en línea] <http://biblio.colmex.mx/recelec/web_invisible.htm> [Consulta: 06 agosto 2005]
- Price, Gary; comp. *Direct Search* [en línea] <<http://www.freepint.com/gary/direct.htm>> [Consulta: 06 agosto 2005]
- Ruiz Felipe, José Angel. Recuperar información de la internet profunda [en línea] <<http://www.sociedadelainformacion.com/20011103/invisible/internetprofundo.htm>> [Consulta 09 marzo 2006]
- Sherman, Chris and Price, Gary (2001). *The invisible Web*. CyberAge Books [Citado por: Moreno Jiménez, Pilar María. *Estrategias y mecanismos de búsqueda en la web invisible*. Disponible en: http://biblio.colmex.mx/recelec/web_invisible.htm]
- UC Berkeley Library, Teaching Library Internet Workshops. *Invisible Web: What it is, Why it exists, How to find it, and Its inherent ambiguity* [en línea] <<http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/InvisibleWeb.html>> [Consulta 06 agosto 2005]
- White, Gary; McGuigan, Glenn S. Unearthing the Invisible Web: Business Resources Search Engines Don't Find [en línea] <http://www.ala.org/ala/rusa/rusaourassoc/rusasections/brass/brassprotools/brasspres/invisibleweb.htm#invisible>> [Consulta 5 marzo 2006]