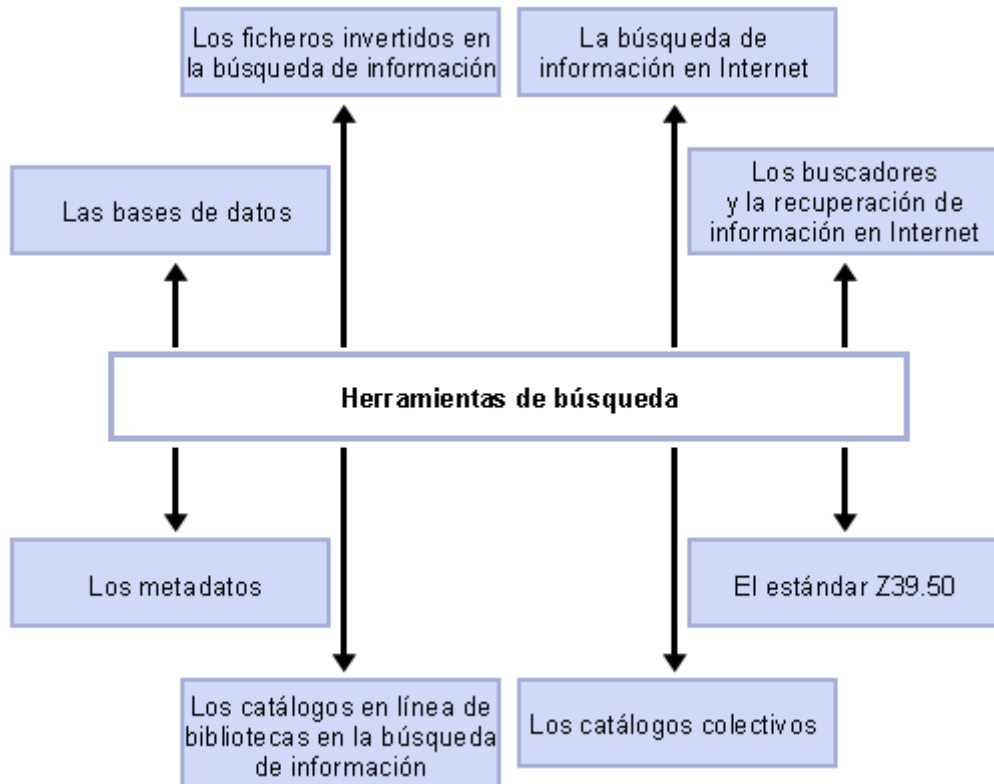


Herramientas de búsqueda



Las bases de datos

En este apartado, dedicado a las herramientas de búsqueda, queremos hacer especial incidencia en las bases de datos, ya que son herramientas indispensables en la búsqueda de información. La consulta de una base de datos especializada nos permite localizar información de cualquier tipo: últimos estudios publicados en todo el mundo sobre un tema concreto, la sentencia sobre un caso judicial de actualidad, datos de facturación de empresas de la competencia, toda la legislación relacionada con una materia concreta, los componentes químicos de un producto farmacéutico, etc.

Las bases de datos son para los documentalistas la fuente de información por excelencia.

Hasta que no apareció la red Internet, y con ella el gran volumen de páginas web accesibles, la forma más fácil y ágil -y a menudo la única- de consultar información especializada era mediante las bases de datos. Actualmente, según el tipo de información que estemos buscando, podemos utilizar un buscador en Internet o ir directamente a una base de datos especializada. También podemos hacer una combinación de los dos tipos de fuentes de información.



Muchas bases de datos especializadas son accesibles también por Internet. Debemos tener presente, sin embargo, que cuando hacemos una búsqueda, el buscador no rastrea el contenido de las bases de datos. En todo caso, el resultado obtenido puede incluir una base de datos como fuente útil para localizar información sobre la materia en cuestión.



Un usuario nos pide información sobre Bali. Vemos que el tema de la búsqueda es muy conocido. Probablemente, en Internet encontraremos muchas páginas web sobre esta isla de Indonesia. Nuestra búsqueda, pues, se iniciaría en Internet.

Un usuario está interesado en conocer los últimos estudios que se han publicado sobre una enfermedad muy poco conocida por el gran público. En este caso, nuestra búsqueda tendrá que hacerse directamente en una base de datos especializada en medicina, como *Medline*.

A menudo, cuando hablamos de bases de datos, pensamos en bases de datos, gratuitas o comerciales, que las bibliotecas y los centros de documentación suscriben para su posterior consulta. Sin embargo, el concepto de base de datos es más amplio.

¿A qué nos referimos cuando hablamos de bases de datos?

Una base de datos no es más que una colección organizada de datos constituida por los propios datos y una serie de índices asociados que permiten que estos datos sean recuperables.

Hace falta, sin embargo, diferenciar entre lo que es propiamente una base de datos y el programa informático que hace posible explotarla. Estos sistemas informáticos se llaman *sistemas de gestión de bases de datos* (SGBD) y hacen posible introducir los datos, modificarlos, eliminarlos, recuperarlos y también editarlos por pantalla e imprimirlos.

Más adelante estudiaremos qué opciones y técnicas utilizan los SGBD para la recuperación de la información.

Todos estaremos de acuerdo en que una de las tareas que desarrolla una biblioteca o un centro de documentación consiste en la búsqueda y recopilación de información. A partir de los datos recopilados, estos centros confeccionarán sus propias bases de datos internas, como podrían ser las siguientes:

- Base de datos de su propio fondo (catálogo).
- Base de datos en la que vacían los artículos de las publicaciones periódicas a las que están suscritos.
- Base de datos para el control de sus adquisiciones.
- Base de datos de sus usuarios.
- Base de datos de las preguntas y respuestas más frecuentes recibidas en el centro, etc.

Todas, pues, corresponderían a bases de datos internas de la biblioteca.

Evidentemente, si la biblioteca decide que la consulta de sus bases de datos internas sea accesible a otras bibliotecas o al público en general -de forma gratuita o mediante previa suscripción-, estas bases de datos pasarán a ser bases de datos externas para la biblioteca que las utilice como fuente de información.

Así pues, las bases de datos externas son herramientas que utiliza la biblioteca para localizar información que no corresponde a su propio fondo; es decir, son externas a la propia biblioteca. Estas bases de datos pueden ser gratuitas o bien de pago. El acceso a ellas tiene lugar por medio de servicios de búsqueda en línea, en CD-ROM y, como ya hemos comentado, cada vez más por Internet.

En el caso concreto de Internet, podemos encontrar tanto bases de datos especializadas gratuitas normalmente, este tipo de bases de datos son mantenidas por organismos oficiales y organizaciones sin ánimo de lucro-, como las denominadas *bases de datos comerciales de pago*, a las que es necesario suscribirse para poder consultarlas.

Algunas de las bases de datos comerciales permiten la consulta gratuita de información referencial. Sin embargo, si se quiere obtener el texto completo o los datos con valor añadido se debe ser miembro de la institución propietaria de la base de datos, como sucedería en el caso de la base de datos Guiame de Esade (<http://www.guiame.net/>), o bien pagar por el texto completo, como ocurre en la base de datos de convenios colectivos de Lex-Nova (<http://www.lexnova.es>).

Otra modalidad sería la que encontramos en bases de datos como ERIC o Medline, que pueden consultarse gratuitamente por Internet, a pesar de que también disponen de versión de pago que ofrece más prestaciones que la versión gratuita.



Podéis consultar el listado de bases de datos comerciales existentes en el mercado en FACSNET: Directory of Database Service (<http://www.facsnet.org>), página web publicada por Poynter Institute for Media Studies.



En Estados Unidos podéis localizar una relación de bases de datos especializadas y gratuitas en Internet Public Library (IPL) Reference Center (<http://www.ipl.org/ref>); se trata de una biblioteca virtual que sirve como punto de inicio para localizar trabajos de referencia, guías temáticas de recursos y bases de datos especializadas.

Tipos de bases de datos

Las bases de datos, según la información que almacenen, pueden clasificarse en dos grandes grupos: referenciales y factuales.

Referenciales	Factuales
<ul style="list-style-type: none">■ Bibliográficas■ Catálogos de bibliotecas■ Referenciales	<ul style="list-style-type: none">■ Numéricas■ Textuales■ Textuales-numéricas■ Multimedia

a. Bases de datos referenciales

Dirigen al usuario hacia otra fuente que contiene la información primaria, como por ejemplo un documento o un organismo. Dentro de esta tipología podemos hablar básicamente de tres grupos:

- **Bases de datos bibliográficas:** incluyen cita o referencias bibliográficas y, en muchos casos, también un resumen. Informan al usuario de lo que se ha publicado sobre una materia en concreto, y también indican en qué tipo de fuente está ubicado.
- **Catálogos de bibliotecas:** incluyen el fondo propio de una biblioteca o de una red de bibliotecas. Estas bases de datos incluyen monografías, títulos de publicaciones periódicas y otros tipos de documentos que tiene la biblioteca. En este sentido, los catálogos son bases de datos bibliográficas, sin embargo, dado que su orientación es diferente de las otras bases de datos bibliográficas -informan sobre el propio fondo de la biblioteca-, es importante considerarlas como una categoría diferente.
- **Bases de datos referenciales:** son directorios que ofrecen datos para identificar organismos, instituciones o personas, como por ejemplo nombres, direcciones, información sobre la actividad que realizan, etc.

b. Bases de datos factuales (*source databases*)

Contienen la información original en formato electrónico. Cuando consultamos una base de datos factual, obtenemos la información directamente y no tenemos que consultar ninguna fuente de información complementaria. La información puede estar disponible, por ejemplo, en formato HTML, PDF y también en formato de impresión. Según su contenido, podemos destacar las siguientes bases de datos:

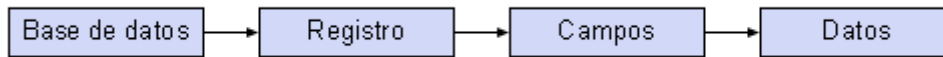
- **Numéricas:** contienen datos numéricos como por ejemplo estadísticas.
- **Textuales:** contienen textos completos de documentos.
- **Textuales-numéricas:** contienen datos textuales y también numéricos. Por ejemplo, el informe anual de una empresa.
- **Multimedia:** contienen documentos en los cuales podemos encontrar no sólo texto, sino también sonido, imagen fija e imagen en movimiento.

Estructuración de las bases de datos: los índices y la búsqueda de información

La estructura de las bases de datos

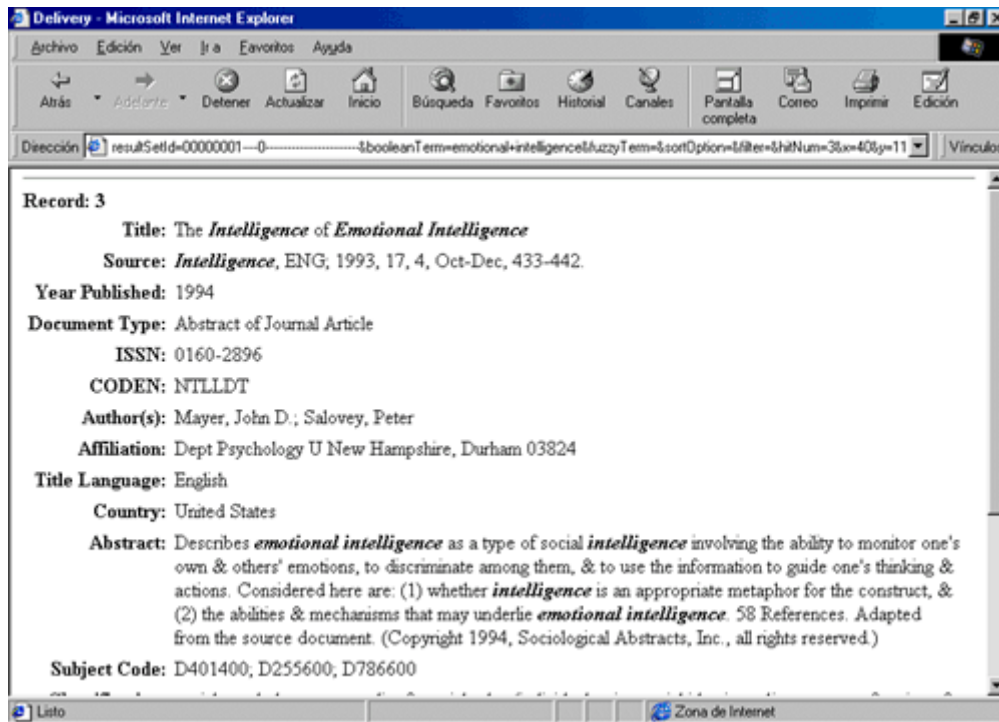
La unidad básica de una base de datos es el registro. Cada registro de una base de datos contiene toda la información relativa a un documento. Por ejemplo, cada registro de una base de datos de patentes contiene toda la información sobre una patente en particular. Del mismo modo, en una base de datos de productos químicos cada registro corresponde a la identificación de un producto concreto.

Por otra parte, los registros están divididos en una serie de campos. Cada campo contiene un elemento de información del registro (es decir, contiene datos). Por ejemplo, en una base de datos de empresas, la información que se facilita sobre cada empresa (nombre de la empresa, dirección, datos de facturación, etc.) se presenta en un campo concreto del registro.

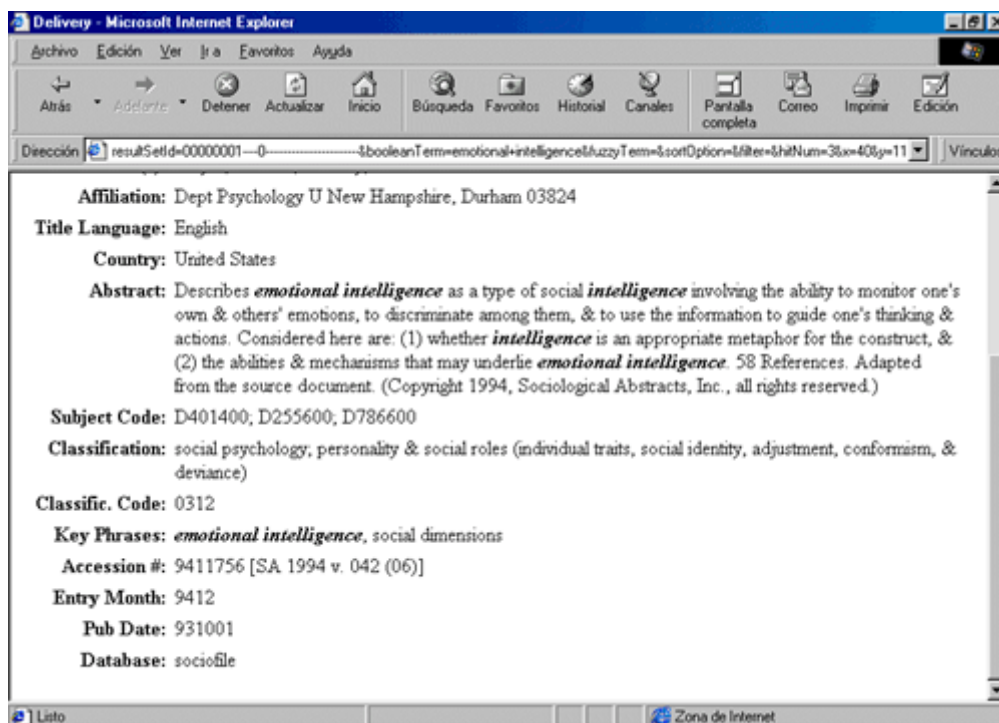


Los campos de los registros de una base de datos son su elemento más diferenciador; estos campos serán unos u otros según el tipo de información que la base de datos almacene.

Es muy probable que una base de datos bibliográfica incorpore los campos siguientes:

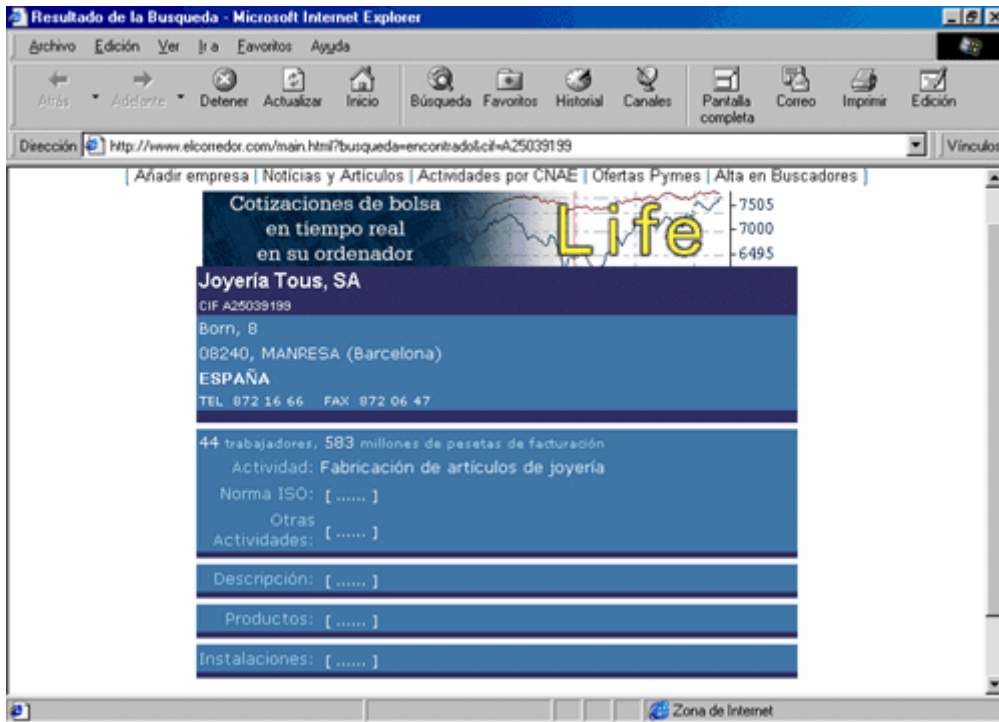


Registro de la base de datos *Sociofile* de Ebsco.



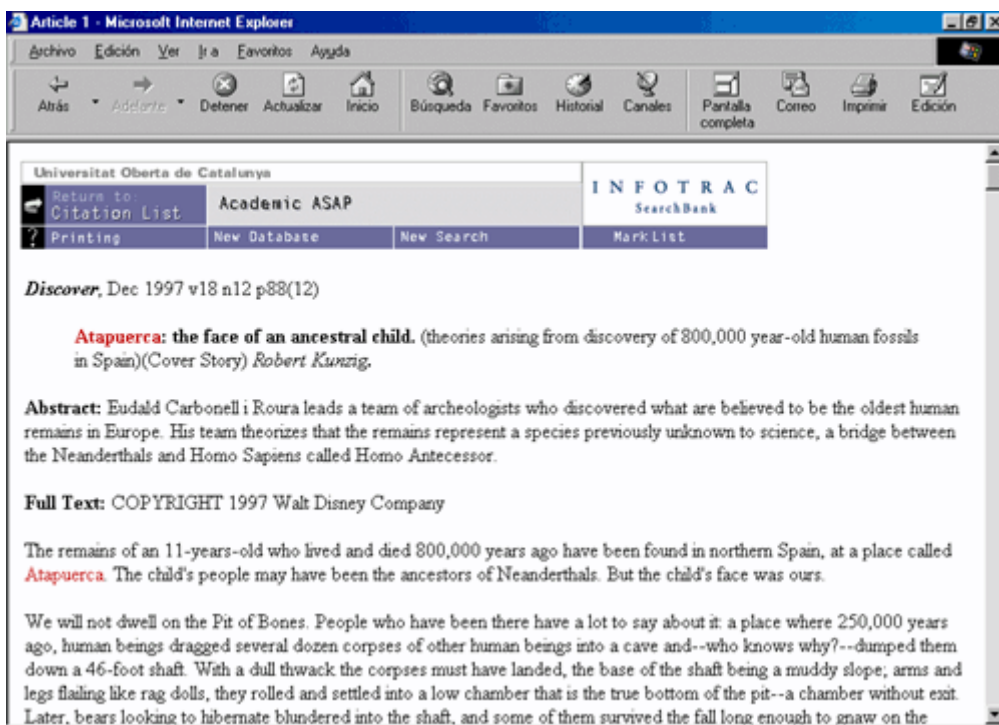
Registro de la base de datos *Sociofile* de Ebsco.

Una base de datos de información de empresas tendrá una estructura de campos diferente:



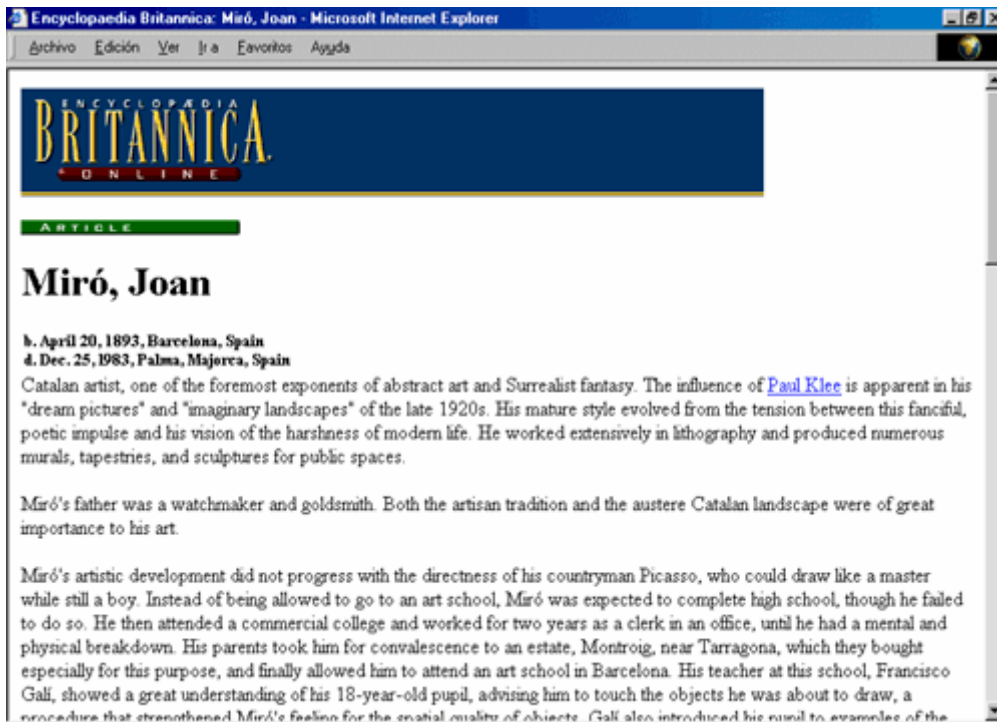
Registro de la base de datos *Buscador de empresas españolas* <http://www.elcorredor.com>

Igualmente, una base de datos a texto completo tendrá algún campo diferente:



Registro de la base de datos *Academic ASAP* de Information Access.

En el caso de una enciclopedia en línea, la base de datos tendrá pocos campos:



Registro de la enciclopedia *Britannica Online*.

Los índices y sus efectos en la búsqueda de información

En la obra *Information Seeking in the Online Age: Principles and Practice*, A. Large, L. A. Tedd y R.J. Hartley consideran que:



"... cuando ejecutamos una búsqueda en una base de datos, esperar diez segundos para obtener una respuesta es un tiempo excesivo, incluso cuando estamos interrogando una base de datos como quizá Medline, que contiene más de ocho millones de registros."

A. Large y otros (1999)

Evidentemente, este tiempo de espera está condicionado por el *hardware*, el *software* y las comunicaciones que se utilicen, pero también tiene un papel muy importante la estructuración e indización de los ficheros.

Un índice pretende proporcionar acceso a los documentos o a la información que éstos contienen. Los índices mantienen un orden preclasificado de uno o más campos, lo cual reduce el tiempo de respuesta de la base de datos.

El productor de la base de datos es quien decide cómo se debe procesar cada campo para generar el índice, de manera que el sistema generará automáticamente los índices correspondientes a cada nuevo registro.

Cada campo puede indexarse de varias maneras. Las siguientes son las más habituales:

■ **Por palabra:**

Este tipo de indización suele utilizarse en los campos que contienen datos sobre el contenido del documento, que son básicamente el título, los descriptores y el resumen. Se trata, pues, de campos en los que cada palabra tiene un significado útil en la recuperación; por lo tanto, el sistema selecciona todas las palabras de este campo excepto las conocidas como *palabras vacías de significado*. Este tipo de palabras corresponden a aquellas que aparecen frecuentemente en un texto, pero que no tienen ningún valor en el momento de recuperar información (*y, sin embargo, de, a*).

Una palabra se identifica como una secuencia alfanumérica de caracteres separados por un espacio o por el final del campo. Por lo tanto, cuando buscamos por un campo que ha sido indizado por *palabra* podremos buscar cualquiera de sus términos, excepto las palabras vacías, independientemente de la posición dentro del campo en que éstos se encuentren.

■ **Por frase:**

Este tipo de indización se utiliza, normalmente, en campos que están sujetos a algún tipo de normalización, como pueden ser los campos de autor y materia. La indización se hace por frase, no por palabras individuales.

Una frase se identifica habitualmente por un punto y coma seguido de un espacio o bien por el final del campo. Si se decidiera indexar un campo de título por frase, deberíamos conocer el inicio exacto del título que estamos buscando para poder recuperarlo.

■ **Por palabra y frase:**

Para algunos tipos de campos, resulta de mucha utilidad el empleo de las dos formas de indización. Normalmente, este tipo de indización se utiliza para campos sometidos a control de vocabulario (descriptores, materias, etc.) en el caso de una base de datos bibliográfica, o bien para el campo del nombre de un organismo en una base de datos de referencia.

■ **Sin ningún tipo de indización:**

Hay campos que contienen información sin ningún valor en un ámbito de búsqueda y que, por lo tanto, no es necesario indexar. Sería, por ejemplo, el caso del campo de paginación en una base de datos bibliográfica.

Los mismos autores nos ilustran los diferentes tipos de indización que pueden establecerse en una base de datos bibliográfica.

Por ejemplo, el caso siguiente corresponde a un registro de una base de datos bibliográfica con los datos introducidos correspondientes a un artículo de revista. Vemos que el registro está dividido en 11 campos: *an* (número de registro), *ave* (autores), *aa* (afiliación de los autores, lugar de trabajo), *ti* (título), *so* (revista fuente), *vo* (volumen), *is* (número), *pg* (páginas), *yr* (año), *ct* (términos controlados) y *ab* (resumen).

```
an: 1245
ave: Jones, Jean; Wilkinson, Michael
aa: Department of Electronic Communication Studies; University of Widdgetown
ti: Developments in medical information on the Internet and its use by nurses
so: Journal of Medical Internet Applications
vo: 6
is: 3
pg: 129-137
yr: 1998
ct: Information seeking; medical informatics; World Wide Web
ab: A study of the growth of medical information on the Web and its effect on the information seeking behaviour of nurses, particularly in community health centres.
```

Fuente: A. Large y otros (1999)

En el caso siguiente, los autores han generado un índice alfabético en el cual detallan entre paréntesis a qué campo corresponde el término de indización, y en el cual vemos también qué campos se han indizado por palabra o por frase:

- Campos indizados por palabras: *ab, an, ti, yr*.
- Campos indizados por frase: *aa, ave, so*.

- Campos indizados por palabra y por frase: *ct*.
- Campos no indizados: *vo*, *is*, *pg*.

1245(**an**)
1998(**yr**)
Behaviour(**ab**)
Centres(**ab**)
Community(**ab**)
Department of Electronic Communication Studies(**aa**)
Developments(**ti**)
Effect(**ab**)
Growth(**ab**)
Health(**ab**)
Informatics(**ct**)
Information(**ti,ab,ct**)
Information seeking (**ct**)
Internet(**ti**)
Jones, Jean(**ave**)
Journal of Medical Internet Applications(**so**)
Medical(**ti,ab,ct**)
Medical informatics(**ct**)
Nurses(**ti,ab**)
Particularly(**ab**)
Seeking(**ab,ct**)
Study(**ab**)
University of Widgettown(**aa**)
Use(**ti**)
Web(**ab,ct**)
Wide(**ct**)
Wilkinson, Michael(**ave**)
World(**ct**)
World Wide Web(**ct**)

Fuente: A. Large y otros (1999)

Para disminuir el tiempo de respuesta, muchas bases de datos crean unos ficheros invertidos que agilizan la búsqueda.

Los ficheros invertidos en la búsqueda de información

Un fichero invertido puede estar compuesto de dos o tres ficheros índice. Si continuamos con el discurso de A. Large, L.A. Tedd y R.J. Hartley, vemos que muchas bases de datos crean tres ficheros que corresponden a:

- **Fichero índice** (*index file*)

En este fichero el sistema selecciona cada una de las palabras con significado de todos aquellos campos del registro que se ha decidido que había que indexar, y las ordena alfabéticamente. Seleccionará palabras o frases según el tipo de indización establecida para cada campo concreto.

A medida que se vayan introduciendo registros en la base de datos, este fichero índice se irá actualizando automáticamente.

Si nos fijamos en la tabla siguiente, podemos ver que un fichero índice puede introducir tres tipos de informaciones:

- **El término** (*term*).
- **Número de ocurrencias** (*number of postings*): número de registros en los que aparece el término en la base de datos.
- **Localización del fichero** (*postings file address*): corresponde a un número de identificación propio de cada término dentro de la base de datos.

Inverted file structure: Index file

Term	No. of postings	Postings file address
...		
Browsers	27	1233
Centres	23	3245
Community	6	6100
Developments	14	1456
Doctors	29	2378
Effect	234	65
Email	145	6740
Engine	26	2354
Engineers	35	4356
Growth	153	3297
Health	635	8234
Informatics	23	3298
Information	347	2682
Internet	2345	5478
Learning	156	6390
Medical	89	7789
Nurses	21	8256

Fuente: A. Large y otros (1999)

Si cuando entramos un nuevo registro, el sistema detecta que uno de los términos no está incluido en este fichero índice, lo añadirá y en la información correspondiente al número de ocurrencias pondrá un 1 y dará un número de localización único para el nuevo término. En el caso de que el término aparezca en el fichero índice, el sistema únicamente tendrá que incrementar en 1 el número de ocurrencias.

■ **Fichero de localización** (*postings file*)

Contiene un número de identificación para cada término del fichero índice. Establece un vínculo de este número con los números de registro de la base de datos que incluyen el término.

Address	Record numbers
1233	88,91,427...
1456	18,98,1245...
2378	5,6,7,8,1455...
2682	23,45,561,678,... 1245
3448	666,783,1011
5478	1,2,3,4,7,8,9,23,45,56,67...1245
6100	456,789,987,1034,1123,1245
6390	45,47,269,345,589
6740	2,3,5,18,34,56,269,345,589...
7789	36,47,59,63,269...
8256	1,3,578,892,1245...

Fuente: A. Large y otros (1999)

■ **Fichero de impresión (*print file*)**

Este fichero está formado por los registros de la base de datos ordenados por número de registro.

Record N.º 1244...

Record N.º 1245

Jones, Jean; Wilkinson, Michael

Department of Electronic Communication Studies; University of Widdgettown

Developments in medical information on the Internet and its use by nurses

Journal of Medical Internet Applications 6 3 129-137 1998

Controlled terms: information seeking; medical informatics; World Wide Web;

A study of the growth of medical information on the Web and its effect on the information seeking behaviour of nurses, particularly in community health centres.

Fuente: A. Large y otros (1999)

Podemos ver, pues, que el fichero de localización establece el vínculo entre el fichero índice y el fichero de impresión.

El funcionamiento de estos ficheros puede llegar a ser bastante complejo según el tipo de búsqueda que hagamos. Sin embargo, lo ejemplificaremos con una búsqueda sencilla:



Cuando introducimos un término para realizar una consulta, por ejemplo, *nurses*, correspondiente a la imagen del *index file*, el sistema testea el fichero índice para ver si este término está incluido en su secuencia alfabética. Si no es así, el usuario recibe un mensaje del sistema en el que se indica que no hay ningún registro en la base de datos relacionado con *nurses*.

Si el término *nurses* está incluido en el fichero índice, el sistema informa al usuario del número de ocurrencias; es decir, del número de registros que contienen información relativa a *nurses*, en este caso 21; para hacer esta función, consultaría el fichero de localización (*postings file*).

Teniendo en cuenta que el resultado de la búsqueda es razonable (21 documentos), el usuario puede decidir ver estos registros por pantalla. En este caso, el sistema encuentra los números de los registros (*accession numbers*) en la localización 8256 del fichero de ocurrencias, y encuentra estos registros en el fichero de impresión; por lo tanto, podrá visualizar la información de los registros 1, 3, 578, 892 y 1245.

La búsqueda de información en Internet

A estas alturas ya todos sabemos que Internet es una red de ordenadores interrelacionados o una red de redes. Históricamente, la red Internet era en esencia una red de ámbito académico, pero el uso por parte de las empresas está creciendo y en estos momentos ya no se trata sólo de una red de elite para la comunicación entre centros de investigación, sino que también es accesible y llega a ser una vía de comunicación muy importante para organismos, instituciones, escuelas, empresas, bibliotecas y centros de documentación de todo el mundo.

Las bibliotecas y los centros de documentación están haciendo un doble uso de la red Internet. Según nos detallan Ernest Ackermann y Karen Harman, estos usos son:



"1. Para dar acceso a sus servicios y recursos. Este uso puede concretarse en:

- Información general o básica sobre la biblioteca, relativa, por ejemplo, a su horario, el directorio del personal, su ubicación física, su política de préstamo, etc.
- Acceso a sus servicios:
 - Consulta de su catálogo y otros catálogos de bibliotecas.
 - Acceso a bases de datos suscritas por la propia biblioteca.
 - Consulta de revistas electrónicas.
 - Presentación de sus servicios, visitas "virtuales" por la biblioteca, etc.
- Nuevas formas de comunicación a distancia con sus usuarios, mediante, por ejemplo:
 - Formularios para hacer consultas, peticiones de préstamo, peticiones de búsquedas bibliográficas, etc.
 - Desideratas para sugerir la adquisición de un libro, y facilidades para pedir documentos en préstamo o hacer reservas de documentos.
- Recopilación de enlaces de interés para la comunidad a la que sirven.
 - El personal de la biblioteca puede identificar aquellas páginas web que considera interesantes para la comunidad a la que da servicio y confeccionar páginas temáticas con estos recursos.

2. Como fuente de referencia para localizar información."

E. Ackermann y K. Harman (1999)

Como ya hemos comentado, cuando hablamos de Internet hacemos referencia a una fuente de fuentes en la cual la información crece a diario. Por lo tanto, localizar aquella información o aquella página web que necesitamos puede convertirse en una tarea bastante pesada.

La mayoría de los autores coincide en afirmar que la llegada del web a principios de la década de 1990 ha sido la responsable de este crecimiento exponencial de la información en Internet. Según Oliver Andrieu:



"... a mediados de 1996 el ritmo de creación de páginas web era de 2.000 por día como media. De hecho, hoy por hoy, la mayoría de información accesible en Internet se encuentra en este formato."

O. Andrieu (1997)

Según el mismo autor, cuando hablamos de la WWW (World Wide Web) nos estamos refiriendo simplemente a una colección de información accesible por medio de Internet. La información contenida en un web se conoce como página web, y normalmente está en formato multimedia, lo cual permite combinar texto, imagen, sonido y vídeo. Una página web puede incorporar enlaces a otros recursos y también a otras informaciones contenidas en la misma página. De hecho, una página web puede incluir información sobre cualquier área del conocimiento (arte, educación, humanidades, política, ciencias, información gubernamental, derecho, etc.); esta información puede ir desde informes técnicos hasta novedades cinematográficas, pasando por estudios de mercado, informes gubernamentales, publicidad, páginas personales, etc.



Como documentalistas, debemos tener en cuenta que Internet es una fuente más, pero no la única; por lo tanto, no tenemos que subestimar las fuentes tradicionales que hasta ahora hemos utilizado (bases de datos comerciales, catálogos de bibliotecas, contacto con otras bibliotecas o centros de documentación especializados, etc.)

En el contexto de la resolución de una consulta, tenemos que elegir siempre aquella fuente de información que pueda darnos la respuesta de la manera más rápida y económica posible.

Como hemos comentado, llegar a extraer sólo la información que necesitamos de esta gran colección de información no es un trabajo fácil. Desgraciadamente, aunque en algunos ámbitos (bibliotecarios y no bibliotecarios) puede llegar a hablarse de Internet como la biblioteca mundial de la era digital, no existe ningún sistema de clasificación común que nos organice toda la información disponible. Contamos, sin embargo, con unos servicios o métodos de búsqueda, también conocidos como *buscadores*, que nos facilitan este trabajo.

Los buscadores y la recuperación de información en Internet

Cuando hablamos de buscadores, estamos hablando básicamente de dos métodos de búsqueda:

- Los directorios
- Los motores de búsqueda

Los directorios

Los directorios presentan una selección de recursos web organizados siguiendo una estructura o clasificación jerárquica de materias que va de categorías más amplias a categorías más específicas.

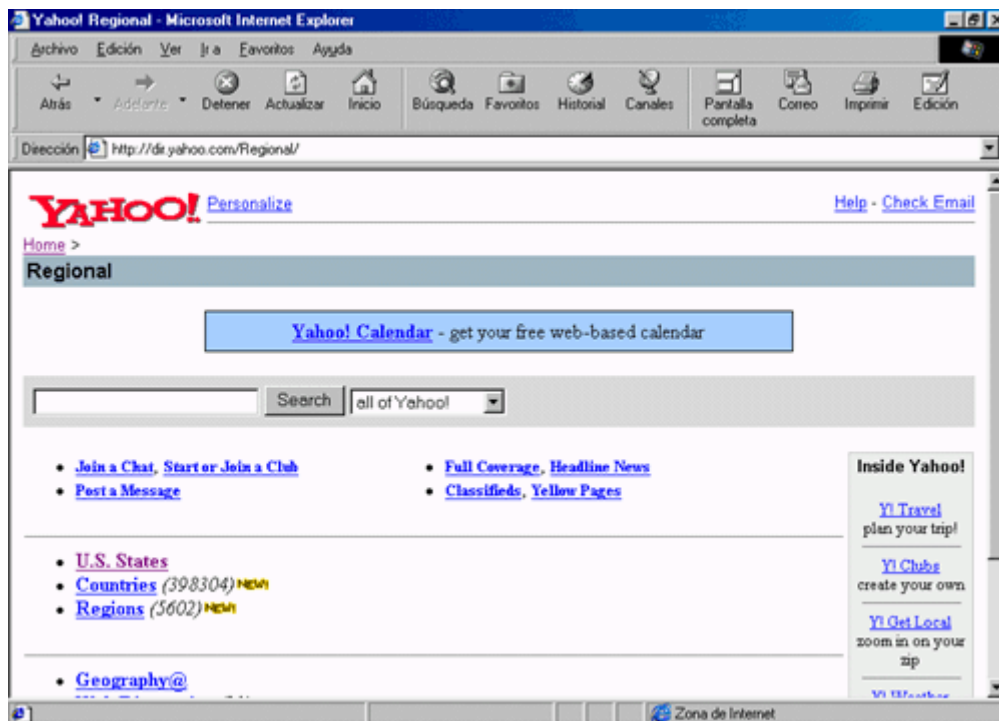
La búsqueda se lleva a cabo mediante la técnica del *browsing*. Esta técnica se inicia a partir de una materia o categoría general, para ir bajando a materias más específicas hasta llegar a la más precisa. Esto es posible porque estamos en un entorno hipertexto como el de World Wide Web, que nos permite navegar de una materia a una parte más específica de esta primera materia de una forma muy ágil y fácil. Cada directorio utiliza sus propias categorías y sus propios niveles de jerarquías.

Respecto a la cobertura, los directorios no pretenden incorporar todas las páginas web accesibles en Internet. De hecho, los directorios son una pequeña colección de recursos, sobre todo si los comparamos con los motores de búsqueda.

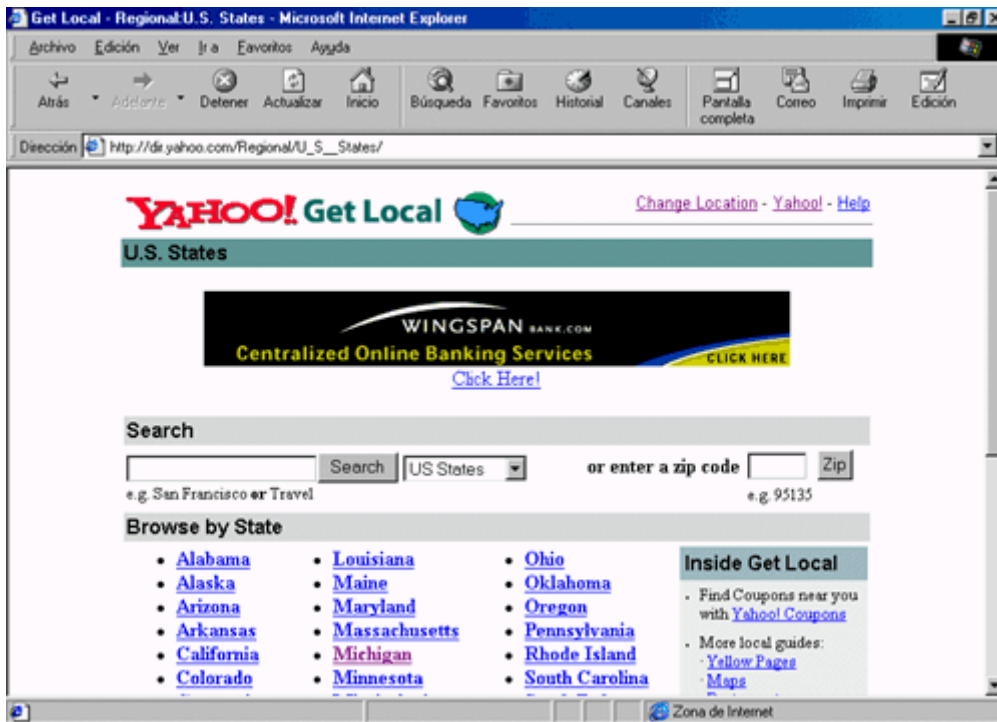
Se recomienda el uso de directorios cuando estamos buscando información que pertenece a una categoría clara (imaginad, por ejemplo, que estamos buscando organismos del ámbito educativo) o también cuando queremos localizar información general sobre un tema.



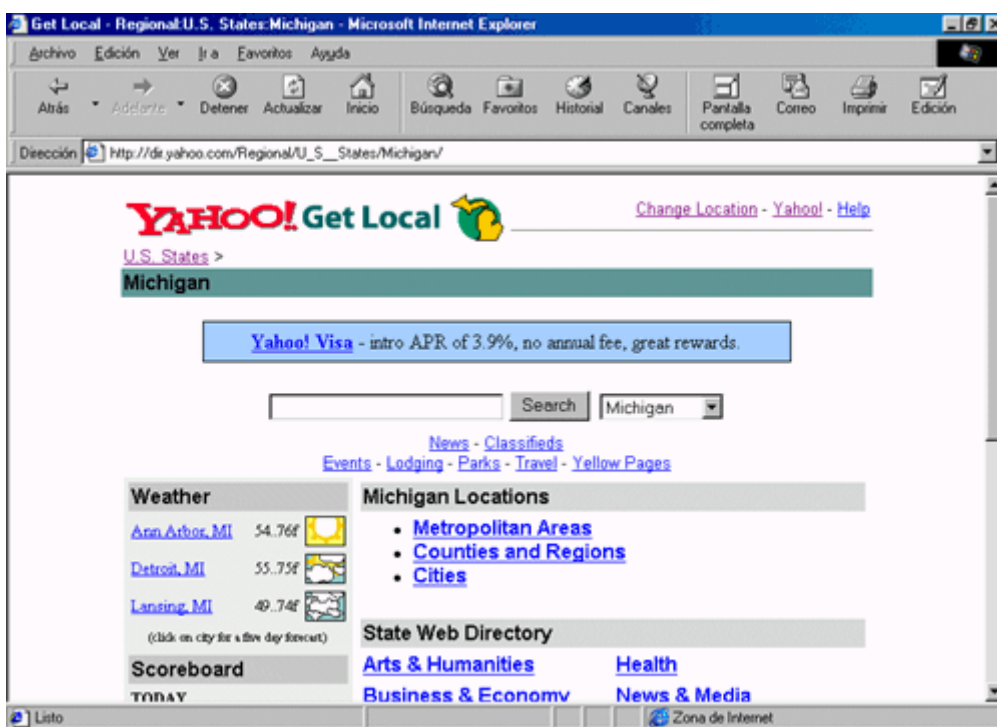
Un usuario nos pide información sobre el Estado de Michigan, de Estados Unidos. Seguramente, si utilizamos un buscador que tiene como método de búsqueda un robot lógico, localizaremos páginas web en las que aparezca la palabra *Michigan*, pero todas no harán necesariamente referencia al Estado de Michigan. En este caso, por lo tanto, es mucho más útil usar un directorio como, por ejemplo, Yahoo.



Un usuario nos pide información sobre el Estado de Michigan de Estados Unidos. Seguramente, si utilizamos un buscador que tiene como método de búsqueda un robot lógico, localizaremos páginas web en las que aparezca la palabra *Michigan*, pero todas no harán necesariamente referencia al Estado de Michigan. En este caso, por lo tanto, es mucho más útil usar un directorio como, por ejemplo, Yahoo.



Un usuario nos pide información sobre el Estado de Michigan, de Estados Unidos. Seguramente, si utilizamos un buscador que tiene como método de búsqueda un robot lógico, localizaremos páginas web en las que aparezca la palabra *Michigan*, pero no todas harán necesariamente referencia al Estado de Michigan. En este caso, por lo tanto, es mucho más útil usar un directorio como, por ejemplo, Yahoo.



En el momento de elegir un buscador que utilice como método de búsqueda un directorio, debemos tener presentes los inconvenientes que mostramos a continuación:

- La estructura jerárquica se crea de manera arbitraria y, por lo tanto, debemos intuir a qué categoría se ha asociado un recurso concreto.
- La actualización no se hace con frecuencia regular.
- La evaluación y selección de los recursos puede haberse realizado sin utilizar criterios lo suficientemente establecidos.

Algunos de los directorios más conocidos son:

- Galaxy (<http://galaxy.einet.net>).
- El Índice (<http://www.elindice.com/>).
- Infoseek (<http://www.infoseek.com>).
- Linkaweb ([http:// http://www.cercat.com/lincaweb/](http://http://www.cercat.com/lincaweb/)).
- LookSmart (<http://www.looksmart.com>).
- Lycos TOP 5% (<http://point.lycos.com/categories>).
- Magellan Web Reviews (<http://magellan.excite.com/>).
- Ozú (<http://www.advernet.es/ozu/>).
- WebCrawler Channels (<http://www.superinternetdeals.com/creditcard.html>).
- Yahoo (<http://www.yahoo.com>).

Hay buscadores que se conocen como *virtual libraries* o bibliotecas virtuales. De hecho, se trata también de directorios especializados, en los cuales los recursos han sido seleccionados y son mantenidos por bibliotecarios con un criterio muy riguroso y profesional.

La gran diferencia que existe entre los directorios y las bibliotecas virtuales es que las bibliotecas virtuales contienen menos recursos, pero éstos han sido seleccionados con más rigor.

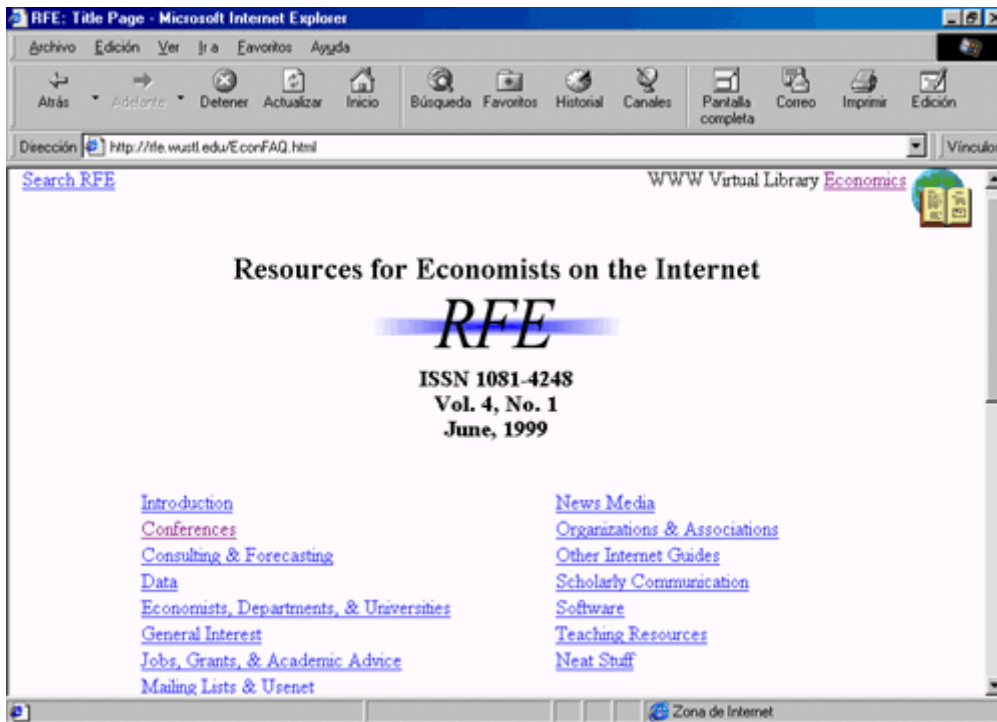
Dentro de estos directorios denominados *virtual libraries*, habitualmente podemos encontrar tres grandes categorías: *subject guides*, *reference works* y *specialized databases*.

- **Subject guides:** son recursos web que incluyen una selección de otros recursos relacionados con una temática concreta y con enlaces a otras páginas web.



RFE Resources for Economists on the Internet <http://www.rfe.org>

Versión hipertexto de la guía de Bill Goffe de recursos disponibles en Internet sobre economía. Esta guía, financiada por la American Economic Association, incluye más de 900 recursos de interés tanto en un ámbito académico como para el ejercicio de la profesión.

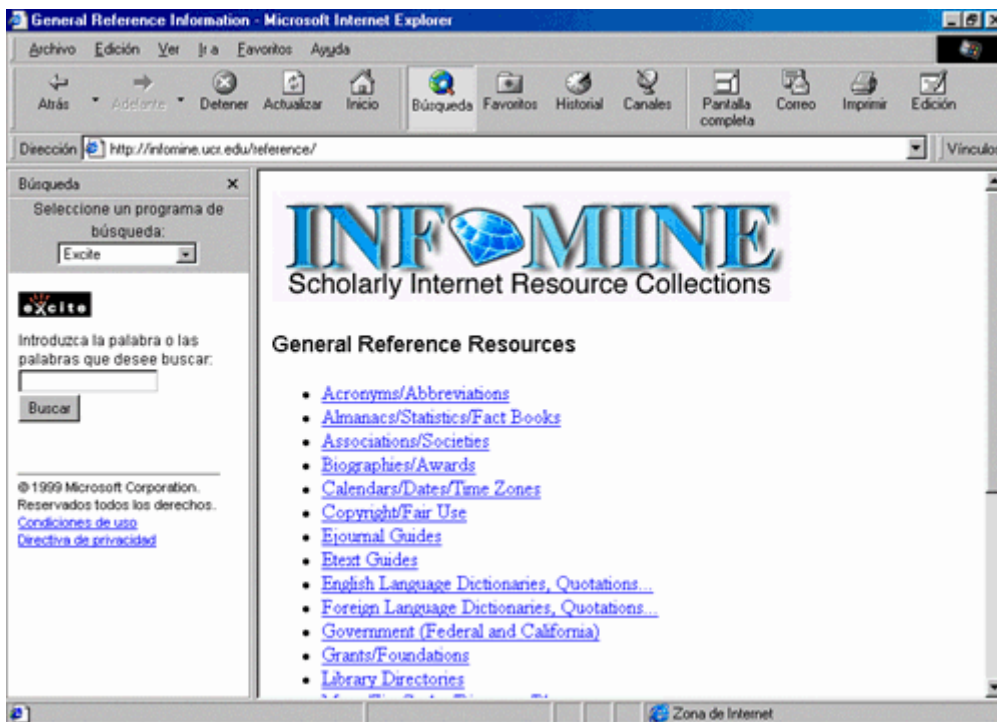


- **Reference works:** son recopilaciones de recursos a texto completo, como diccionarios, enciclopedias, etc.



Infomine. General Reference Resources

<http://infomine.ucr.edu/reference/>



- **Bases de datos especializadas** o *specialized databases*: como ya estudiaremos más adelante, son índices que permiten la búsqueda de diferentes tipos de materiales, como por ejemplo artículos de revistas, datos financieros, etc.

Las bibliotecas virtuales más conocidas son:

- The Argos Clearinghouse (<http://www.clearinghouse.net>).
- BUBL Link (<http://bubl.ac.uk>).
- Infomine (<http://infomine.ucr.edu/reference/>).
- Internet Public Library (<http://www.ipl.org>).
- Librarians' Index to the Internet (<http://lii.org/>).
- World Wide Web Virtual Library (<http://www.vlib.org>).

Los motores de búsqueda

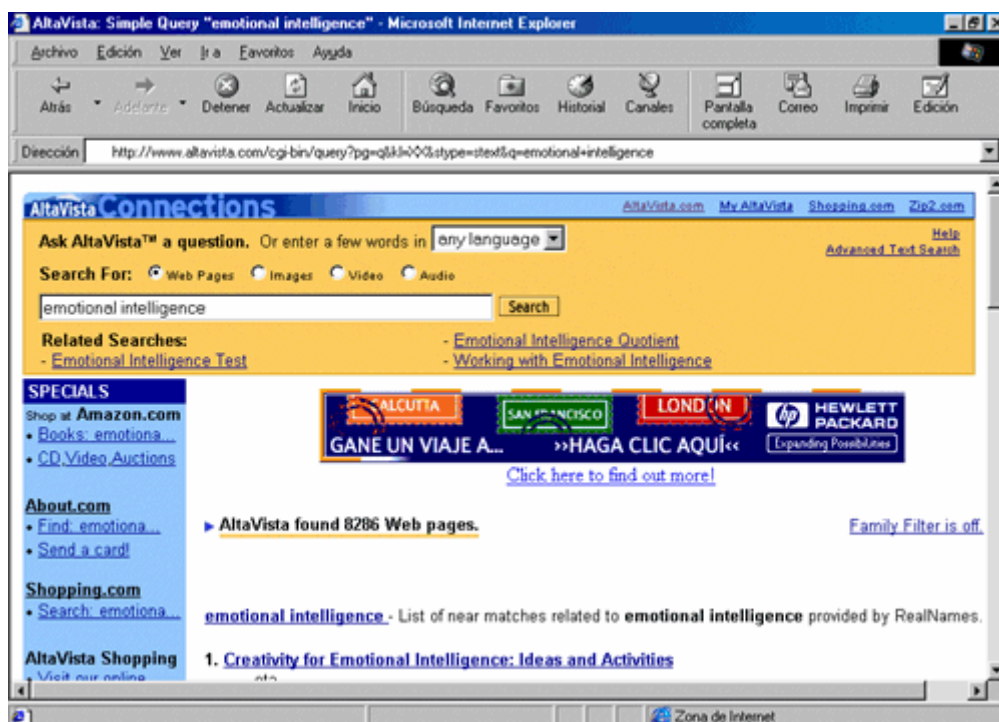
Los motores de búsqueda (en inglés, *search engines*) permiten la búsqueda de webs por medio de descriptores. Son herramientas que utilizan programas de ordenadores, denominados *spiders* o robots lógicos, que rastrean y recopilan automáticamente información disponible en Internet. Después, incorporan estos recursos a una base de datos que podemos consultar mediante los buscadores.



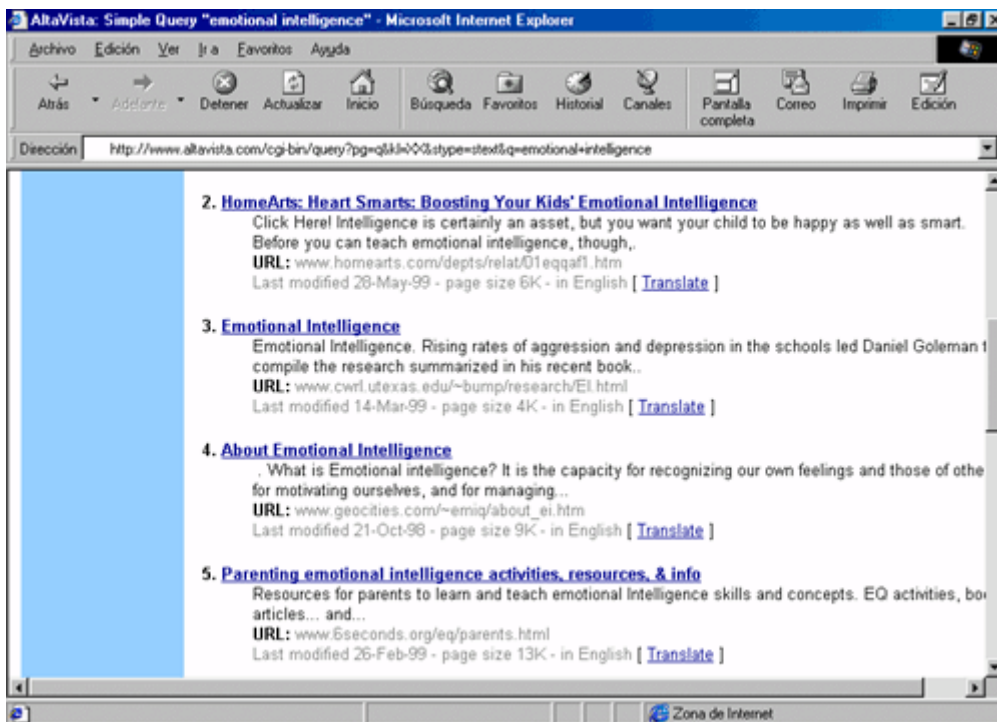
El rápido incremento de páginas web en Internet hizo inviable la indexación manual de la información contenida en páginas web, lo cual favoreció la creación de los robots lógicos o *spiders*.



Necesitamos localizar información general sobre inteligencia emocional. En este caso, la búsqueda en un directorio no será la opción más apropiada, ya que se trata de un concepto muy específico que no sabemos en qué categoría y subcategoría del directorio puede haber sido clasificado. Por lo tanto, seleccionaremos un buscador que utilice un motor de búsqueda como, por ejemplo, Altavista.



Necesitamos localizar información general sobre inteligencia emocional. En este caso, la búsqueda en un directorio no será la opción más apropiada, ya que se trata de un concepto muy específico. Por lo tanto, seleccionaremos un buscador que utilice un motor de búsqueda como, por ejemplo, Altavista.



Las principales ventajas que suponen los motores de búsqueda son las siguientes:

- Se actualizan frecuentemente.



Mientras que las bases de datos de los motores de búsqueda son creadas de manera automática por programas de ordenadores, los directorios son creados y mantenidos por personas.

-
- Permiten el acceso a grandes colecciones de información.
 - Facilitan un resultado exhaustivo.

Se recomienda la utilización de este tipo de buscadores cuando buscamos por un concepto o una frase específica. También se recomienda la utilización de más de uno de ellos, ya que el resultado obtenido seguramente será distinto.

Los buscadores más importantes que utilizan motores de búsqueda son los siguientes:

- Altavista (<http://www.altavista.digital.com>).
- Excite (<http://www.excite.com>).
- HotBot (<http://www.hotbot.com>).
- Infoseek (<http://www.infoseek.com>).
- Lycos (<http://www.lycos.com>).
- Northern Light (<http://www.northernlight.com>).
- Olé (<http://www.ole.es>).
- VilaWeb (<http://vilaweb.com/>).
- WebCrawler (<http://www.webcrawler.com>).

Los metabuscadores también son muy útiles y nos permiten buscar en diferentes motores de búsqueda al mismo tiempo, en lugar de tener que ir interrogando los distintos buscadores por separado.

Las ventajas de utilizar este tipo de herramientas son obvias: ahorramos un tiempo considerable y los resultados son exactamente los mismos que si ejecutáramos la búsqueda en cada uno de los motores de búsqueda por separado.

Sin embargo, también tienen las desventajas siguientes:

- No todos los motores de búsqueda se interrogan con la misma sintaxis.
- No hay control de duplicados: un mismo recurso aparecerá duplicado el mismo número de veces que se encuentre indizado en motores de búsqueda.

Algunos de los metabuscadores más conocidos son los siguientes:

- All4one (<http://all4one.com/>)

Busca en AltaVista, Lycos, HotBot y Excite.

- MetaCrawler Searching (www.metacrawler.com)

Realiza la consulta simultánea en los buscadores siguientes: About.com, AltaVista, Excite, GoTo.com, Infoseek, LookSmart, Lycos, Thunderstone, Webcrawler y Yahoo.

Si queréis entender cuál es el sistema que utilizan los robots lógicos para actualizar la información de sus bases de datos, leed el texto de Oliver Andrieu de la obra *Cómo buscar y encontrar en Internet*.



O. Andrieu (1997)

El tipo de búsqueda efectuado más habitualmente por los internautas es sin duda la búsqueda por introducción de palabras claves. En este sentido, desde hace unos meses se ha creado un gran número de servicios. Su principio es simple: el servicio indexa previamente un cierto número de páginas web de todo el planeta. Esta indexación se efectúa por texto completo, y todas las palabras de todas las páginas referenciadas se convierten en entradas del índice, potencialmente objeto de búsqueda. El usuario introducirá a continuación una o más palabras claves, enlazadas entre sí por operadores booleanos Y o bien O y el motor buscará las ocurrencias de esas palabras en las páginas fuente. Se establece a continuación una clasificación de las páginas más pertinentes y se presenta finalmente una lista de enlaces hacia dichas páginas.

Los motores de búsqueda son más o menos potentes, bien por el número de páginas indexadas por adelantado, bien por la posibilidad de enlazar las palabras clave por medio de operadores booleanos distintos, o bien por un cierto número de opciones que especificaremos en las páginas siguientes.

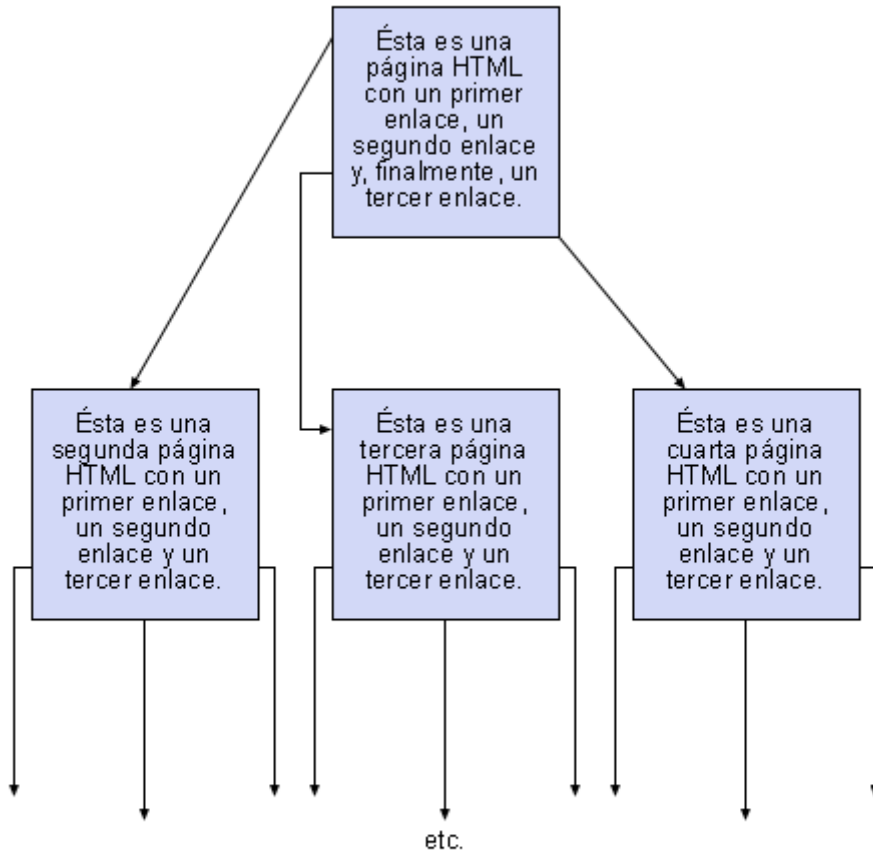
Indexación de las páginas

La indexación previa de las páginas puede hacerse de dos formas distintas:

Por declaración voluntaria del responsable de la sede web remota que indica al motor de búsqueda (en forma de un formulario en la web) la existencia de su servicio. El motor indexa entonces todas las páginas del servicio así referenciado.

Por medio de robots que recorren de modo automático la web y sus páginas de información. Estos robots lógicos parten de un cierto número de páginas básicas, y persiguen todos los enlaces del hipertexto en cada una de las páginas encontradas. Cada página identificada se indexa seguidamente por texto completo. Este sistema evita la declaración previa de un servicio, desde el momento en que un enlace de hipertexto apunta hacia él en cualquier parte del mundo.

¿Cómo funcionan los robots lógicos?



El robot lógico parte de una o más páginas definidas previamente y recorre todos sus enlaces de hipertexto. Hace lo mismo con todas las páginas encontradas a medida que progresa por la web mundial. Así, recorre en unos doce días la mayor parte de las páginas HTML disponibles en Internet.

Los robots lógicos recorren actualmente todo el universo (o casi) de la web en unos doce días.

Límites de los robots lógicos

Es importante comprender correctamente el concepto de funcionamiento de estos servicios. Cada indexación de página provoca una "captura" de las informaciones presentes en ellas y un almacenamiento del archivo obtenido en los discos duros del motor de búsqueda, en modo local. Cuando un usuario introduce una palabra clave, el sistema no se conecta en tiempo real a todos los servidores del planeta -esto sería totalmente irrealizable-, sino que efectúa sus búsquedas en el interior de las páginas que anteriormente ha almacenado en sus peregrinaciones por la web, con la indexación correspondiente.

Puede ocurrir, pues, que una información actualizada diariamente esté presente en una sede web, pero no se encuentre por medio del motor de búsqueda utilizado, ya que éste habrá capturado sus informaciones la semana anterior. Concretamente, ilustremos este hecho con un ejemplo: un periódico español pone en línea sus informaciones en la web diariamente. Supongamos que un motor de búsqueda, por medio de un robot lógico, tarda diez días en recorrer la web. Si las páginas del periódico español son indexadas por el robot el día 1 de enero, no volverá a actualizar su base de conocimientos hasta el 11 de enero, es decir, diez días más tarde. El motor de búsqueda sólo conservará en la memoria las informaciones del día 1 de enero almacenadas en su disco duro, mientras que, en la propia sede, estas informaciones se actualizan diariamente.

Sería necesario, para tener una imagen más precisa de las informaciones realmente presentes en la web en un momento dado, que los robots recorrieran la web en un día, o incluso menos. Es de suponer que esto llegará algún día, pero no es el caso por el momento.

Otra limitación de estos servicios es que indexan sólo la información encontrada en las páginas HTML presentes en la web. No tienen en cuenta la información contenida en las bases de datos específicas de los organismos presentes en la web, accesibles mediante formularios. Veamos un nuevo ejemplo: *Le Monde* pone sus archivos en línea desde hace algunos meses mediante el sistema de pago electrónico Globe online:



Service d'archives du Monde

Chercher les articles qui contiennent les mots suivants:

PARTOUT ou

Uniquement dans les titres

après le

04/ 1995

et avant le

04/ 1995

Dans le thème suivant:

Signataire:

Mots clés:

Valider

Réinitialiser la recherche

Cuando el usuario busca la información en los archivos del periódico, introduce una palabra clave en un cuadro de edición del formulario, realiza si lo desea otras elecciones (fecha de publicación, etc.) y obtiene, una vez validada la operación, una pantalla de representación de los artículos encontrados, junto con su precio (en este caso una búsqueda sobre el término *Sarajevo*):

Résultats de la recherche

53 articles ont été trouvés. [Page suivante]. [Nouvelle Recherche]

N°	Date	Page	Titre	Thème	Catégorie	Prix
1	29/04/1995	5	A trois jours de l'expiration de la trêve, Sarajevo se prépare à de nouvelles heures noires	International		5FF
2	28/04/1995	R09	La Yougoslavie pour l'exemple	Livres		5FF
3	28/04/1995	18	L'alibi des hypocrites	Horizons	bulletin	5FF
4	27/04/1995	14	Pâques sanglantes à Sarajevo	Horizons	opinion	5FF
5	27/04/1995	14	Bosnie - Nations unies: des leçons pour l'avenir	Horizons	opinion	5FF
6	27/04/1995	1	Le retour impérial de la Russie	La Une		5FF
7	26/04/1995	3	BOSNIE: deux civils à Sarajevo	International		5FF
8	26/04/1995	3	Radovan Karadžić devant la justice internationale	International		5FF
9	25/04/1995	12	Le Serbes ferment l'aéroport de Sarajevo aux médiateurs occidentaux	International		5FF
10	25/04/1995	3	Nouvelles consignes de sécurité pour les "casques bleus"	International		5FF

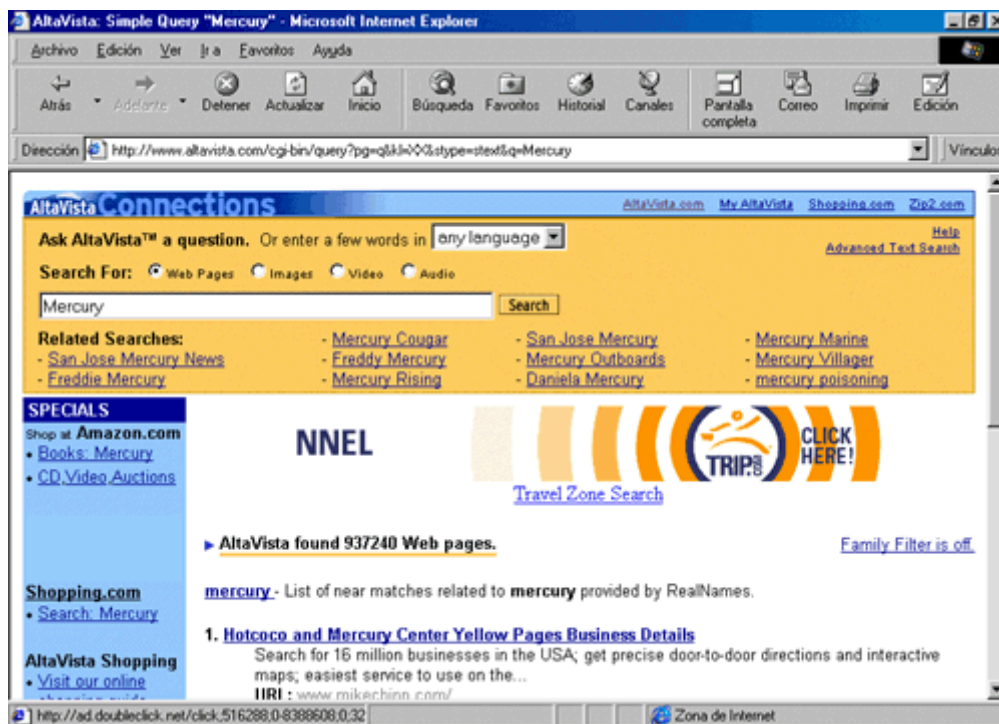
El usuario no tiene más que pulsar sobre el artículo deseado y pagarlo con su monedero virtual Globe Online para que aparezca en su pantalla. Pero las informaciones así tratadas son dinámicas, a partir de una consulta basada en criterios de las palabras claves elegidas, en la base de datos interna de los archivos del periódico *Le Monde*. Estas informaciones no se encuentran en forma de páginas HTML "en bruto" en la web y por tanto no se referenciarán en ninguno de los motores de búsqueda.

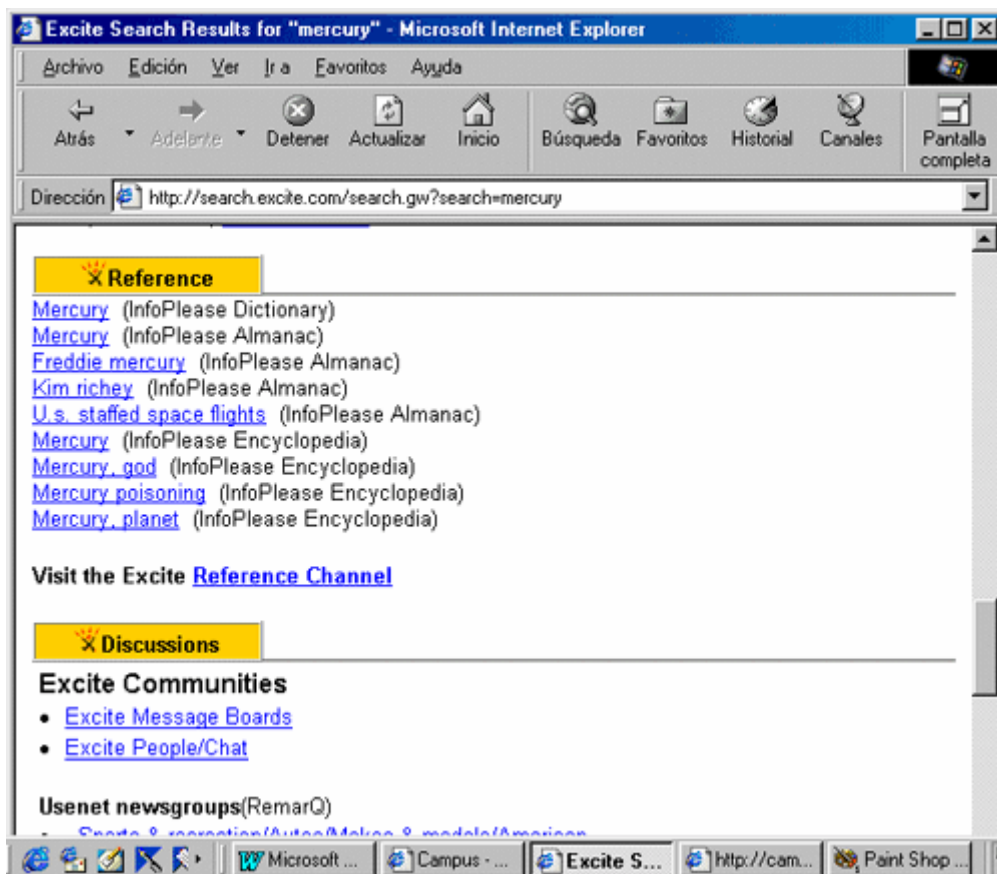
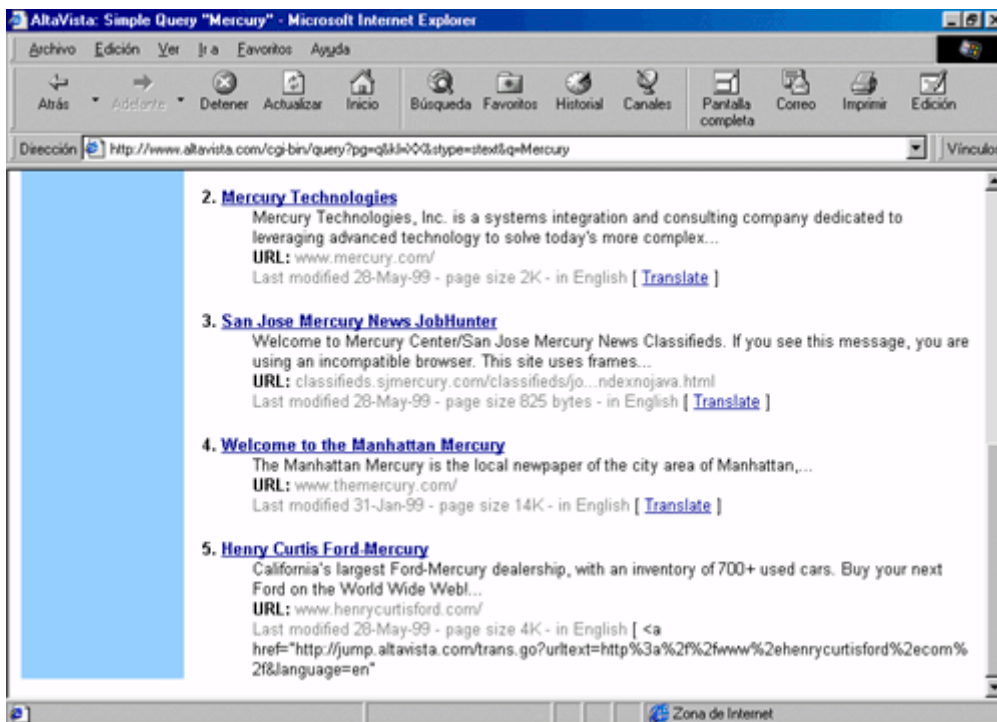
Una última limitación: el creador de una página tiene la posibilidad de indicar en su código HTML (o en el servidor donde se encuentra) su deseo de que no sea indexada por los robots lógicos. Esta restricción resulta sin embargo a menudo anecdótica, porque no se comprende por qué una página publicada en la web no debe referenciarse en los motores de búsqueda, salvo en el caso de páginas de uso privado o en el marco de servicios en fase de creación.

Debemos tener presente que la mayoría de los principales buscadores permiten la búsqueda de información tanto por un directorio como por un motor de búsqueda, aunque generalmente dan más importancia a la búsqueda por medio de uno de estos métodos. El directorio que forma parte de un motor de búsqueda es, normalmente, un subconjunto de la base de datos entera, y los recursos listados en el directorio están por lo general evaluados, disponen de un resumen y se han ordenado siguiendo un criterio de ponderación. Igualmente, algunos directorios incorporan unas opciones de búsqueda que suelen ser muy sencillas.

Ya hemos comentado en varias ocasiones que el gran crecimiento de la información disponible en Internet provoca que el hecho de recuperar sólo la que necesitamos sea un proceso cada vez más difícil. Y hemos estudiado también que los motores de búsqueda llegan a ser herramientas útiles en la recuperación de información, pero que, al mismo tiempo, presentan importantes inconvenientes, como por ejemplo que el resultado obtenido no es siempre el deseado.

Hoy por hoy, el hecho de que los motores de búsqueda rastreen el texto completo de los documentos hace que el resultado de las búsquedas tenga, muy a menudo, un número de documentos inalcanzable y, por otro lado, que no todos sean pertinentes. Por ejemplo, si buscamos en un buscador el término *Mercury*, obtendremos un resultado como el del ejemplo.





Vemos, pues, que recuperamos páginas sobre el planeta, sobre el mercurio como elemento químico, sobre el dios griego Mercurio e incluso sobre Freddy Mercury.

De toda esta cantidad de información, ¿cómo podemos recuperar una imagen del planeta que no tenga ningún tipo de texto asociado?

Ante esta situación, e igual que ocurre en las bibliotecas tradicionales, se ve clara la necesidad de establecer algún tipo de estandarización que nos ayude a organizar la información que hay en Internet y, por lo tanto, a localizar sólo documentos pertinentes. De hecho, aplicar los métodos tradicionales que se utilizan en las bibliotecas (es decir, que profesionales de la documentación sean los responsables de catalogar e indexar todas las páginas web o los recursos accesibles en Internet) es, obviamente, una tarea imposible. Sin embargo, a pesar de este gran impedimento, continúa siendo un trabajo imprescindible y en este contexto entran en juego los metadatos.



Fuentes de información en Internet

Hasta aquí hemos visto cómo podemos localizar información en Internet. También hemos comentado que este medio es una fuente de fuentes. Sin embargo, ¿qué tipologías de fuentes de información podemos encontrar en Internet? En este punto, estudiaremos cuáles son las que nos propone Jennifer Rowley en el libro *The electronic library*.

En primer lugar, observemos los tipos de servicios más generalizados de Internet:

- **E-mail:** permite a los usuarios enviarse mensajes o ficheros entre ellos.
- **New:** comunica a los usuarios la información disponible.
- **ftp** (file transfer protocol): permite a los usuarios acceder y localizar ficheros en ordenadores remotos.
- **Conexión a servidores remotos:** permite a los usuarios el acceso a ordenadores remotos.

Dentro de este último grupo, desglosamos la información y los servicios relacionados con la conexión en servidores remotos, que son útiles para los profesionales de la documentación y que llegan a ser fuentes de información importantes:

- **Listerserv y grupos de discusión** sobre una gran variedad de materias. Los participantes tienen la oportunidad de intercambiar información actual y establecer un diálogo.
- **Bases de datos temáticas**, concretamente elaboradas por instituciones académicas. Va aumentando el número de instituciones, especialmente académicas y de investigación, que confeccionan bases de datos especializadas en su ámbito o materia y que las hacen accesibles a través de Internet. Un caso sería, por ejemplo, el de la Universidad Complutense de Madrid, que da acceso gratuito a sus bases de datos de vaciado de revistas COMPLUDOC, o la National Library of Medicine de Estados Unidos, que permite la consulta de su base de datos *Medline*.
- **Información sobre comunidades locales concretas.** Cada vez con mayor frecuencia las comunidades están dando acceso a información local; por ejemplo, información turística, información sobre el tiempo, datos demográficos, etc.
- **Información gubernamental.** Instituciones gubernamentales, tanto en el ámbito estatal como local, están utilizando Internet para difundir sus actividades, así como para dar a los ciudadanos otra vía de acceso a información de interés público.
- **Catálogos de bibliotecas.** Va aumentando el número de bibliotecas que están haciendo accesible su catálogo mediante Internet. Resulta especialmente interesante el acceso a los catálogos colectivos; este tipo de acceso nos permite consultar en una sola búsqueda los catálogos de todas las bibliotecas que forman parte, por ejemplo, de la Red de Bibliotecas Universitarias Españolas REBIUN (<http://www.crue.org/cgi-bin/rebiun>).
- **Fuentes comerciales.** Las bases de datos comerciales -no gratuitas- también son accesibles mediante Internet. Cada vez más, todas las grandes empresas suministradoras de bases de datos están haciéndolas accesibles por web: Dialog, SilverPlatter, etc.
- **Diarios electrónicos.** Algunos diarios permiten consultar en Internet toda la información que publican en papel o parte de ella.
- **Compras y otras transacciones comerciales.** Ya hay un número significativo de empresas que permiten a sus clientes efectuar las compras desde casa. En este sentido, es necesario destacar las librerías y editoriales que ya nos permiten comprar libros y otros materiales mediante Internet.
- **Obtención de documentos:** muchas bibliotecas y servicios de obtención de documentos, como podría ser el caso de las bases de datos del CSIC, permiten la búsqueda de los documentos y su envío por Internet.

Los metadatos

Todos los autores coinciden en afirmar que los metadatos son datos sobre datos. Y desde el punto de vista de la documentación, podemos decir que corresponden a un estándar para describir documentos electrónicos tanto en lo que respecta a la forma (catalogación) como al contenido (indización).

A diferencia de los documentos típicos de las bibliotecas, este estándar de metadatos tiene que ir más allá, ya que

debe permitir la identificación no sólo de documentos textuales, sino también de documentos no-textuales, como podrían ser imágenes, gráficos, sonido o cualquier formato electrónico. En este punto los metadatos se diferencian de otros estándares de bibliotecas, pues tienen que permitir que los motores de búsqueda también puedan localizar documentos sin ningún texto asociado, como por ejemplo una imagen de una pieza de museo, un vídeo sobre violencia, etc.

Así pues, podemos afirmar que los metadatos son muy importantes en la búsqueda de información, ya que su finalidad coincide plenamente con la que nos proponemos cuando elaboramos una estrategia de búsqueda: que coincidan totalmente los conceptos de búsqueda con los campos y las palabras que hemos utilizado para describir los documentos, en este caso electrónicos.

Los metadatos surgen como método de estandarización tanto de la estructura como del contenido de los documentos electrónicos, mediante la catalogación y la indización de la información.

¿Para qué sirven los metadatos?

- En primer lugar, actúan como un sustituto de todo el recurso o documento electrónico. Permiten identificar el documento lo suficiente como para que el usuario pueda entender el contenido, así como el propósito, la fuente y, si es necesario, las condiciones de su utilización.
- Como complemento a esta función descriptiva, y para que sean realmente útiles, es necesario que establezcan una estructura y una terminología estándar. Serán útiles para constituir un único campo que englobe conceptos similares; por ejemplo, *creador*, *autor*, *escultor* y *compositor*. Y también deben permitir crear formas alternativas a las distintas entradas de los campos, de manera que resuelvan los ya conocidos y clásicos problemas de la recuperación de la información: la sinonimia, la polisemia y la ambigüedad.

Evidentemente, para que los metadatos sean una realidad práctica hay que plantearse y tener claro quién gestiona sus formularios.

¿Quién se encargará de rellenar los formularios de los metadatos?

Ya hemos comentado que, hoy por hoy, es completamente inviable que toda la información que se encuentra en la red Internet sea catalogada por profesionales de la información. No obstante, este colectivo está dirigiendo parte de sus esfuerzos hacia la creación de unos estándares de metadatos que faciliten su descripción a los propios autores de las páginas web.

Podemos decir que la mayoría de los creadores de recursos electrónicos no es consciente de la importancia que la aplicación de metadatos tiene con relación a la recuperación de la información; en todo caso, el argumento de la recuperación no tendrá el suficiente peso como para motivarlos en el momento de rellenar los formularios. Sin embargo, probablemente el hecho de que los motores de búsqueda los utilicen cuando deben seleccionar los recursos, y también en el momento de establecer el orden de los documentos más pertinentes, podrá ser el argumento que motive a los propios creadores para introducir los datos correspondientes.

En este sentido, los profesionales somos conscientes de la necesidad de crear unos estándares que faciliten la introducción de datos y que, al mismo tiempo, esto se haga de manera uniforme. Este punto será el que, si se cumple, hará que los metadatos sean útiles en la búsqueda de información.

Por eso se está trabajando con el objetivo de establecer un estándar común -el más conocido es el Dublin Core- y también se están elaborando muchos proyectos al respecto. Por ejemplo, la International Standards Organization (ISO) ha creado recientemente un grupo de trabajo sobre metadatos que se responsabilizará de crear las especificaciones de los estándares y de gestionar los metadatos.



Actividades

- Para profundizar más en el concepto de metadatos, leed el artículo obligatorio siguiente:

Milstead, J.; Feldman, S. (1999) "METADATA - Cataloging by any other name". *Online* (n.º 23, i1, enero, pág. 25).

- Para conocer con detalle los proyectos y estándares que se están desarrollando relacionados con los metadatos, leed el artículo siguiente:

Milstead, J.; Feldman, S. (1999) "METADATA - Projects & Standards". *Online* (n.º 23, i1, enero, pág. 32).

- Para conocer con detalle los metadatos de Dublin Core, os recomendamos que leáis el artículo siguiente:

San Segundo, R. (1999) "Organización del conocimiento en Internet: metadatos bibliotecarios Dublin Core". En: *VI Jornadas Españolas de Documentación "Fesabid 98"* (pág. 14). Valencia.

Los catálogos en línea de bibliotecas en la búsqueda de información

En su nivel más básico, un catálogo en línea -en inglés OPAC (Online Public Access Catalog)- corresponde a la versión automatizada de los tradicionales catálogos manuales de las bibliotecas.

La aparición de los primeros catálogos en línea tuvo lugar a finales de la década de los setenta, y su desarrollo se produjo durante los primeros años de la década de los ochenta. De hecho, la aparición de los catálogos en línea representó una gran revolución en el mundo de las bibliotecas, ya que les permitió abrirse a sus usuarios, reemplazando los catálogos manuales -con las limitaciones que comportaban- por catálogos automatizados.

El diseño de los catálogos en línea se ha hecho siempre pensando en los usuarios, que son los clientes finales de esta herramienta. Por eso el desarrollo de los OPAC se ha centrado en hacerlos cada vez más amigables, es decir, fáciles de consultar y con interfaces atractivas.

¿Cómo ha tenido lugar la evolución de los catálogos en línea con relación a la búsqueda?

Las primeras versiones de los catálogos en línea de finales de los años setenta permitían a los usuarios básicamente las mismas opciones de búsqueda que los catálogos manuales; es decir, búsqueda por autor, título, clasificación y, en algunos casos, por materias.

Estos campos siempre se indizaban como frase, lo cual obligaba a conocer el inicio de las entradas de cada campo.

Posteriormente, durante los años ochenta, algunos propietarios de programas introdujeron opciones de búsqueda en los catálogos en línea similares a las técnicas que se habían desarrollado para los servicios de búsqueda en línea como, por ejemplo, Dialog. Estos OPAC ya permitían la búsqueda por palabra en campos como por ejemplo título, materias y autores, y también la utilización de los operadores booleanos que permiten buscar por más de un término al mismo tiempo.

En los noventa apareció la interfaz Windows. La mayoría de los catálogos en línea incorporaron esta interfaz gráfica, lo cual hizo mejorar considerablemente la amigabilidad de estos sistemas hacia los usuarios. Ya a finales de los noventa, la aparición de Internet también influyó en los OPAC, pues se creó una nueva necesidad para las bibliotecas: hacer accesible la consulta de su catálogo independientemente del lugar donde estuviese el usuario. De este modo, investigadores y estudiantes de cualquier lugar del mundo pueden acceder a los catálogos de las bibliotecas que los hayan hecho consultables por Internet.

Los dos tipos de entornos que en la actualidad permiten el acceso remoto a los catálogos en línea de bibliotecas son el acceso telnet y el acceso web, aunque se está imponiendo el entorno web, porque es mucho más amigable que el entorno telnet.

Del mismo modo, cada vez más los catálogos en línea de bibliotecas están pasando de ser únicamente una base de datos bibliográfica a incluir también el texto completo de los documentos: libros enteros, legislación, artículos de revista, etc. En definitiva, documentos electrónicos. En estos momentos, un usuario puede buscar en un catálogo de biblioteca la *Constitución Española de 1978* y además de visualizar información bibliográfica sobre este documento, podrá conocer también si está disponible para el préstamo y dónde está ubicado físicamente. Asimismo, y lo que tiene más importancia, le será posible obtenerlo entero en el mismo instante.



Biblioteca de la UOC - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

Atrás Adelante Detener Actualizar Inicio Búsqueda Favoritos Historial Correo Imprimir

Dirección <http://xina.uoc.es:443/be/inici.html> Ir a Vínculos

Biblioteca [Página principal](#) [Catálogo UOC](#) [Biblioteca digital](#) [Servicios](#)

Lunes, 30 de julio

TÍTULOS

Anterior Siguiente Buscar Ayuda

Documentos	Escoja línea
[-]	Constitución
[-]	Constitución 1812
[-]	Constitución 1834
[-]	Constitución 1837
[-]	Constitución 1845
[-]	Constitución 1869
[-]	Constitución 1876
[-]	Constitución 1931
[-]	Constitución 1978
[-]	Constitución (1978). Català
1	La Constitución como norma y el Tribunal Constitucional / Eduardo García de Enterría
1	Constitución de 1837 [Document electrònic]
1	Constitución de 1845 [Document electrònic]
1	Constitución de 1869 [Document electrònic]

UOC

Listo Internet

Imagen del catálogo de la Biblioteca de la UOC.

Biblioteca de la UOC - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda





Atrás Adelante Detener Actualizar Inicio Búsqueda Favoritos Historial Correo Imprimir

Dirección <http://xina.uoc.es:443/be/inici.html> Ir a Vínculos »

Biblioteca [Página principal](#) [Catálogo UOC](#) [Biblioteca digital](#) [Servicios](#)

Lunes, 30 de julio

Información bibliográfica

 Copias
  Reservas
  Buscar
  Ayuda

AUTOR	Espanya
TIT.UNIFORM.	Constitución (1978)
TÍTULO	Constitución y Tribunal Constitucional ; edición preparada por Enrique Linde Paniagua
EDICIÓN	16ª ed.
PUBLICADO	Madrid : Civitas, 2000
DESC.FÍSICA	487 p.
COLECCIÓN	Biblioteca de legislación ; 10
MATERIA	España. Tribunal
MATERIA	España -- Constitución --
MATERIA	España -- Derecho
AUTOR SEC.	Linde Paniagua, Enrique
ISBN	84-470-1459-2
CDU	342(460)(094)

UOC

Listo Internet

Imagen del catálogo de la Biblioteca de la UOC.

Constitución de 1978

- [Preámbulo](#)
- [Título preliminar](#)
- [Título I. De los Derechos y Deberes Fundamentales:](#)
 - [Capítulo I. De los españoles y los extranjeros](#)
 - [Capítulo II. Derechos y libertades](#)
 - [Capítulo III. De los principios rectores de la política social y económica](#)
 - [Capítulo IV. De las garantías de las libertades y derechos fundamentales](#)
 - [Capítulo V. De la suspensión de los derechos y libertades](#)
- [Título II. De la Corona](#)
- [Título III. De las Cortes](#)
 - [Capítulo. I De las Cámaras](#)
 - [Capítulo II. De la elaboración de las leyes](#)
 - [Capítulo III. De los Tratados Internacionales](#)
- [Título IV. Del Gobierno y de la Administración](#)
- [Título V. De las Relaciones entre el Gobierno y las Cortes Generales](#)
- [Título VI. Del Poder Judicial](#)
- [Título VII. Economía y Hacienda](#)
- [Título VIII. De la Organización Territorial del Estado](#)
 - [Capítulo I. Principios Generales](#)
 - [Capítulo II. De la Administración Local](#)
 - [Capítulo III. De las Comunidades Autónomas](#)
- [Título IX. Del Tribunal Constitucional](#)
- [Título X. De la Reforma Constitucional](#)
- [Disposiciones Adicionales](#)
- [Disposiciones Transitorias](#)
- [Disposición Derogatoria](#)
- [Disposición Final](#)

Imagen del catálogo de la Biblioteca de la UOC.

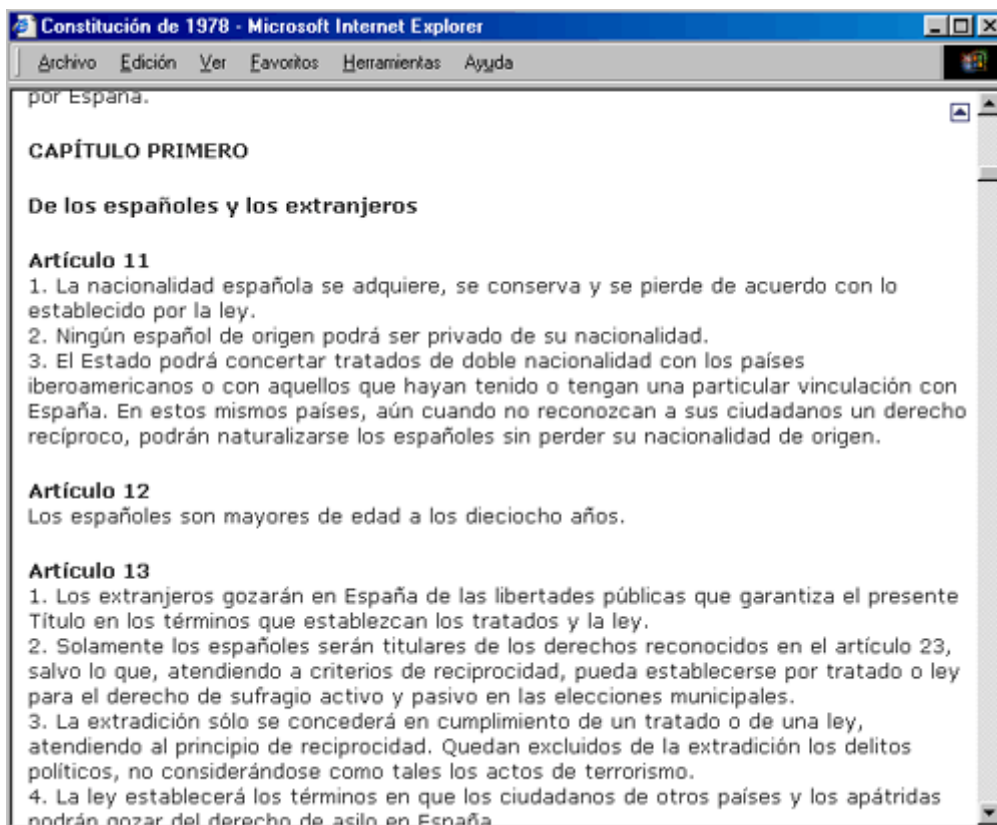


Imagen del catálogo de la Biblioteca de la UOC.

Otras mejoras que se están produciendo en los catálogos en línea son, por ejemplo, la posibilidad que se ofrece a los usuarios de hacer reservas de documentos o, incluso, pedir documentos en préstamo, como ocurre en el caso de la Biblioteca de la UOC.

Algunas bibliotecas están haciendo que sus catálogos en línea no sólo den información bibliográfica, sino que, previo escaneo, los usuarios puedan visualizar los sumarios de los documentos, resúmenes del contenido y, como ya hemos comentado, el texto completo .



-
- Para localizar qué bibliotecas tienen accesible su catálogo vía Telnet, consultad: WebCATS: Library Catalogues on the World Wide Web.
<http://www.libdex.com>
 - Para localizar qué bibliotecas tienen accesible su catálogo en Web, consultad: HYTELNET
(<http://www.einet.net/hytelnet/START.TXT.html>)

Si buscar el catálogo de una biblioteca desde cualquier lugar ya representa un gran beneficio, el hecho de poder consultar más de un catálogo al mismo tiempo todavía puede serlo más.

Los catálogos colectivos

Los catálogos colectivos permiten la consulta simultánea del fondo de más de una biblioteca o más de un centro de documentación.

Por lo tanto, ejecutando una sola estrategia de búsqueda, estaremos consultando el fondo de varias bibliotecas al mismo tiempo.

Según Manuel Carrión:



"Los catálogos colectivos tienen como finalidad el acceso al documento. Son un medio y no un fin, sirven en tanto en cuanto están integrados en sistemas de cooperación interbibliotecaria o en sistemas oficiales capaces de proporcionar el acceso al documento y se justifican para materiales de no fácil acceso. Su justificación primera no es el control bibliográfico, aunque pueden ser un resultado de éste.

[...] Los catálogos colectivos tienen su máxima justificación en dos campos:

el primero es el de la bibliografía científica actual entre bibliotecas científicas (universitarias, especiales, nacional). El segundo campo lo constituyen los fondos más o menos raros por distintas causas, entre las que resalta la antigüedad..."

M. Carrión (1993)

Así pues, los catálogos colectivos están normalmente relacionados con redes de bibliotecas que ofrecen (entre otros servicios) precisamente la creación de un único catálogo con el fondo de todas las bibliotecas que lo conforman. Lluís Anglada, en su artículo *Cooperación bibliotecaria en España*, dice:



En algunos casos, el nacimiento de los catálogos colectivos y otras experiencias cooperativas se produce por el uso del mismo programa de automatización de sus catálogos, como podría ser la red RUEDO (<http://buo.uniovi.es/Buo-Ruedo.html>), configurada por bibliotecas universitarias que utilizan el programa Dobis-Libis y que es posible consultar por Internet.



"El término red, sin embargo, se ha usado y se usa con significados diferentes. En sentido estricto, y a pesar de reconocer que el término se utiliza de forma un poco indiscriminada, red se aplica a aquellas actividades cooperativas en las que intervienen diferentes bibliotecas y que se sustentan en algún tipo de conexión mediante la red informática."

Por lo tanto, la consulta del catálogo colectivo es uno de los servicios que ofrece una red de bibliotecas. Este catálogo es el resultado de la catalogación compartida entre las bibliotecas que lo conforman.

Otro servicio que se ve favorecido por la aparición de estos tipos de catálogos es el préstamo interbibliotecario. Una vez que hemos obtenido la relación de documentos resultantes de una búsqueda de un catálogo colectivo, el sistema nos informará de las bibliotecas que tienen estos documentos, y es muy probable que nos facilite algún tipo de formulario para hacer la petición de préstamo interbibliotecario.

El estándar Z39.50

En el momento de consultar catálogos en línea accesibles por Internet, tanto si se trata del catálogo de una sola biblioteca como del catálogo de una red de bibliotecas, se nos plantea una gran dificultad: cada biblioteca o red de bibliotecas utiliza un programa de automatización concreto que, muy a menudo, no coincide con el que el usuario conoce. Por lo tanto, consultar el catálogo en línea de las bibliotecas implica que tengamos que formarnos previamente en su programa de interrogación, y también familiarizarnos con su interfaz de consulta.

Lo mismo sucede con la consulta de bases de datos. Cada proveedor de bases de datos utiliza su programa para interrogarlas; por lo tanto, las interfaces y las características de consulta varían de una base de datos a otra, según quién sea el proveedor. Así pues, nos encontramos con el hecho de que en las bibliotecas es necesario que nuestros usuarios aprendan tantos programas de consulta de bases de datos como suscripciones tengamos con distribuidores diferentes.

El estándar Z39.50 se ha ideado para solucionar estos problemas.

Aunque aquí no entraremos con mucho de detalle en este estándar, sí que queremos hacer una pequeña introducción sobre qué es y, a un nivel muy básico, ya que su aplicación representa un progreso importante para la búsqueda remota.

¿Qué es el estándar Z39.50?

Algunas definiciones posibles del estándar Z39.50 son las siguientes:

Según Clifford A. Lynck:



"Z39.50 -más concretamente Information Retrieval System (Z39.50); Application Service Definition and Protocol Specification, ANSI/NISO Z39.50-1995'- es un estándar que especifica una estructura de datos y unas normas de intercambio que permitan a un ordenador cliente (también denominado *origen*) buscar en bases de datos ubicadas en un ordenador servidor (también denominado *target* en inglés) y recuperar registros que han resultado de una búsqueda."

C.A. Lynck (1997)

Según Frances G. Livingston y Martin Rosen:



"El estándar de búsqueda de información Z39.50 permite a usuarios individuales utilizar las opciones de búsqueda y navegación de su sistema habitual para acceder a información almacenada en otros sistemas."

F.G. Livingston y M. Rosen (1998)

El estándar Z39.50 fue desarrollado inicialmente por ANSI, American National Standards Institute, y ha sido aprobado como estándar internacional por la ISO, International Standards Organization.

Actualmente lo está manteniendo la Library of Congress.

¿Cómo actúa el estándar Z39.50 en la búsqueda de información?

Un usuario inicia una búsqueda y formula la estrategia utilizando la interfaz del sistema local, que en este caso actúa como un ordenador cliente. Este ordenador convierte la estrategia de búsqueda en un formulario Z39.50 y transmite esta información ya codificada al servidor remoto donde está ubicada la base de datos que queremos consultar. El ordenador correspondiente a este servidor remoto vuelve a convertir la estrategia de búsqueda en un formulario legible para su propio sistema, y ejecuta la búsqueda.

Los resultados son retornados al ordenador cliente también con un formato Z39.50, y este ordenador lo convierte al lenguaje propio de su programa.

El estándar Z39.50 nos permite también consultar bases de datos y catálogos de bibliotecas independientemente del idioma en el que esté el catálogo en línea. Por lo tanto, mediante el estándar Z39.50 se pueden consultar, por ejemplo, catálogos de bibliotecas rusas, alemanas, chinas, etc. aunque no conozcamos estos idiomas.



Actividades

- Para más información sobre el estándar Z39.50, consultad la página web *Library of Congress Maintenance Agency page for International Standard Z39.50*. <http://lcweb.loc.gov/z3950/agency/>.
- Para conocer con más detalle la relación entre Z39.50 y las páginas web, leed la ponencia siguiente:

López de Sosoaga Torija, A. (1998). "OPAC-Web-Z39.50: ¿Redundantes o complementarios?: la realidad es multilingüe". En: *VI Jornadas Españolas de Documentación "Fesabid 98"* (pág.24).Valencia. http://www.florida-uni.es/~fesabid98/Comunicaciones/a_lopez.htm .

Actividades

1. Haced una búsqueda del término *Atapuerca* en dos metabuscadores diferentes.

Comparad qué buscadores interroga cada uno de ellos, cómo los interrogan y también cómo visualizan los resultados.

2. Para ver cómo afecta el tipo de indización de un campo al resultado de una búsqueda, localizad artículos relacionados con un concepto como, por ejemplo, *dirección de personal* en la base de datos ISOC del CSIC (accesible desde la Biblioteca de la UOC). Haced la búsqueda en la opción *Búsqueda en campos básicos*. Buscad en primer lugar por palabra y, a continuación, por frase. Evaluad el resultado.
3. Haced una búsqueda en el catálogo de la Library of Congress por web y por Telnet (<http://lcweb.loc.gov/catalog/>) y, a continuación, realizad la misma búsqueda en su servidor Z39.50 / (<http://lcweb.loc.gov/z3950/gateway.html>).

Nota: si no tenéis instalada la aplicación Telnet, podéis conseguir una copia y las instrucciones de instalación en la Biblioteca de la UOC (botón *Información / Formación - Manuales / Instrucciones de consulta en catálogos VTLS por Telnet*).

4. Describid y analizad los buscadores siguientes:
 - The Argus Clearinghouse (<http://www.clearinghouse.net>)
 - BUBL Link (<http://bubl.ac.uk>)

Nota: podéis utilizar como guía los criterios aplicados por Ángeles Maldonado y Elena Fernández en su estudio "Evaluación de los principales 'buscadores' desde un punto de vista documental: recogida, análisis y recuperación de los recursos" ([http://www.florida-uni.es/~fesabid98/Comunicaciones/a_maldonado/A_Maldonado .htm](http://www.florida-uni.es/~fesabid98/Comunicaciones/a_maldonado/A_Maldonado.htm)).

Ejercicios de autoevaluación

1. Estableced las diferencias más importantes que hay entre buscadores que utilizan como método de búsqueda un directorio y los que utilizan un motor de búsqueda.
2. Explicad qué método o métodos utilizan los motores de búsqueda para seleccionar e indexar las páginas web que incluyen.
3. ¿Qué son los metadatos?
4. Determinad cuáles son las dos funciones más importantes de los metadatos.
5. Comentad muy brevemente qué es el Dublin Core.
6. Elaborad un esquema en el que queden reflejadas las diferentes tipologías de bases de datos comentadas en este apartado, y explicad brevemente sus características principales.
7. ¿En qué consiste la indización por palabra? ¿Y la indización por frase?
8. Un investigador está interesado en conocer el fondo de bibliotecas médicas de Estados Unidos. ¿Dónde podría localizar esta información? Elaborad un documento con una selección de bibliotecas americanas especializadas en medicina que tengan accesible su catálogo para web.
9. Elaborad un pequeño esquema en el que se refleje la evolución de los catálogos en línea de bibliotecas.
10. Comentad en qué aspectos el estándar Z39.50 nos resulta de utilidad en la búsqueda de información.

Bibliografía

Ackermann, E.; Harman, K. (1999). *The information specialist's guide to searching and researching on the Internet and the World Wide Web*. Wilsonville [Or.] : ABF Content, cop.

Andrieu, O. (1997). *Cómo buscar y encontrar en Internet* (pág. 35-40). Barcelona: Ediciones Gestión 2000.

Anglada, LI. (1995) "Cooperació bibliotecària a Espanya". *Item* (núm. 16, enero-junio, pág. 51-67).

Bradley, Ph. (1999). *The advanced internet searcher's handbook* (pág. 232). Londres: Library Association.

Carrión Gútiérrez, M. (1993). "Presente y futuro del catálogo en la biblioteca". En: *Manual de bibliotecas* (pág. 289-319). Madrid: Fundación Germán Sánchez Ruizpérez; Madrid: Pirámide.

Large, A.; Tedd, A.L.; Hartley, R.J. (1999). *Information Seeking in the Online Age: Principles and Practice*. London: Bowker-Saur, cop.

Livingston, F.G.; Rosen, M. (1998). "Consortial use of Z39.50: the Kentuckiana Metrouniversity Linking Project". En: *Computers in Libraries* (v. 18, abril, pág. 16).

López de Sosoaga Torija, A. (1998). "OPAC-Web-Z39.50: ¿Redundantes o complementarios?: la realidad es multilingüe". En: *VI Jornadas Españolas de Documentación "Fesabid 98"* (pág. 24). Valencia.

http://www.florida-uni.es/~fesabid98/Comunicaciones/a_lopez.htm (consulta: 23 de julio de 1999).

López Yepes, J. (coord.) (1996). *Manual de información y documentación*. Madrid: Pirámide.

Lynck, C.A. (1997). "The Z3950 Information Retrieval Standard. Part I: A strategic view of its past, present and futur". En: *D-Lib Magazine* (abril)

<http://www.dlib.org/dlib/april97/04lynch.html> (consulta: 25 de julio de 1999).

Maldonado Martínez, A.; Fernández Sánchez, E. (1998). "Evaluación de los principales 'buscadores' desde un punto de vista documental: recogida, análisis y recuperación de recursos de información". En: *VI Jornadas Españolas de Documentación "Fesabid 98"* (pág. 24). Valencia.

http://www.florida-uni.es/~fesabid98/Comunicaciones/a_maldonado/A_Maldonado.htm (consulta: 23 de julio de 1999).

Milstead, J.; Feldman, S. (1999). "METADATA - Cataloging by any other name". *Online* (v. 23, i1, enero, pág. 25).

Milstead, J.; Feldman, S. (1999). "METADATA Projects & Standards". *Online* (v. 23, i1, enero, pág. 32).

Montes Hernández, A. (1999). "Posibilidades de consulta en los buscadores". *El profesional de la información* (v. 8, núm. 3, marzo, pág. 8-14).

Moscoso, P. (1996). "Bases de datos documentales: estructura y organización". En: *Manual de información y documentación* (pág. 393-403). Madrid: Pirámide.

Moscoso, P.; Ortiz-Repiso, V.; Rios, Y. (1998). "Telnet versus web en el acceso a la información catalográfica: estudio comparativo del caso español". En: *VI Jornadas Españolas de Documentación "Fesabid 98"* (pág. 5). Valencia.

http://www.florida-uni.es/~fesabid98/Comunicaciones/p_moscoso.htm (consulta: 23 de julio de 1999).

Rowley, J.E. (1998). *The electronic library* (pág. 396). Londres: Library Association.

San Segundo, R. (1998). "Organización del conocimiento en Internet: metadatos bibliotecarios Dublin Core". En: *VI Jornadas Españolas de Documentación "Fesabid 98"* (pág. 14). Valencia.

http://www.florida-uni.es/~fesabid98/Comunicaciones/r_sansegundo.htm (consulta: 23 de julio de 1999).