

Acceso y recuperación de la Información

Concepto de Internet



Aplicaciones básicas de Internet



Recuperación de la información y bases de datos



Opciones de búsqueda



Lenguajes de interrogación y operadores



Programas informáticos de asistencia a la traducción



Conceptos de Internet

“(...) el ciberespacio; pero cualquiera que conecta ahora, tiene que saber que es un universo entero. Y cada año hay más cosas ahí dentro (...)”

William Gibson, *Conde Cero* (1986)

Tecnologías y acceso a la información

La *World Wide Web* (WWW)

presenta la información

da acceso a los diferentes recursos



A través de documentos hipertextuales
(páginas web)



formados por un entrelazamiento de **texto**, **imágenes**, a veces **sonido** e **hiperenlaces** a otros documentos u otras webs

Conceptos

<http://tramullas.com/documatica/>

- ◉ **Teledocumentación:** consiste en el acceso y la consulta de bases de datos existentes en una localización remota, lejana al usuario, mediante la utilización de un ordenador conectado a una red de telecomunicaciones.
- ◉ La **Documentación Automática/Documática** es la disciplina que se ocupa de la investigación y aplicación de las Tecnologías de la Información en todos los ámbitos de las Ciencias de la Documentación, teniendo como material de trabajo la información generada y requerida durante el proceso documental que se trate, en el marco de los sistemas de información, en un contexto que integra los medios automáticos, el ser humano y las interacciones entre ambos.
- ◉ La **telemática o teleinformática** es la disciplina que surge de la confluencia entre la tecnología de comunicaciones y la Informática. Esta confluencia se produjo en ambas direcciones: los sistemas informáticos se conectaron mediante redes telefónicas y éstas mejoraron mediante dispositivos informáticos

¿Qué es Internet?

El fenómeno
social,
cultural,
sociológico y
comercial

de Internet,
red de redes



tiene su origen en la década del 60 (Siglo XX)



se relaciona con un proyecto de
defensa financiado por el gobierno
de Estados Unidos

El Departamento de Defensa de Estados Unidos decidió promover la **investigación** y desarrollo de una **estructura computacional nacional defensiva**, que respondiera a los requerimientos de funcionamiento bajo circunstancias extremas y que incluso ante un eventual ataque nuclear que destruyera parte del sistema, el resto de éste pudiera continuar activo. Se descubrió que para lograr este objetivo (mantener la comunicación en circunstancias especiales), se debía establecer una malla o red interconectada donde todos los participantes estuvieran comunicados entre sí.

¿Qué es Internet?

Luego la conexión se extendió a computadores pertenecientes a entidades universitarias, que participaron activamente en este y en otros proyectos del Departamento de Defensa para intercambiar mensajes e información a través de esta red.

Desde ese momento el crecimiento de Internet ha sido espectacular, especialmente en la década del 90, época en que la red se convirtió en una herramienta fundamental de comunicación, información e integración, que permite a los usuarios ahorrar tiempo y dinero, además de tener a su alcance todos los productos y servicios que requieran sin fronteras de espacio o tiempo.

Conceptos básicos sobre los recursos web

- ◉ **World Wide Web (WWW):** sistema de información distribuido, basado en hipertexto [...]. La información puede ser de cualquier formato –texto, gráfico, audio, imagen fija o en movimiento- y es fácilmente accesible a los usuarios mediante los programas navegadores
- ◉ **Hipertexto:** sistema de organización y acceso a la información basado en la asociación de documentos, que utiliza para ello un sistema informático, y que sigue la metáfora de la red de elementos
- ◉ **Sitio o Sede Web:** Localización en la WWW, identificada por una URL, que almacena páginas web para facilitar su acceso y uso. En este concepto se integraría el de *Portal*, del que hablaremos más adelante.

Conceptos básicos sobre los recursos web

- ⦿ **Página Web:** Una de las páginas de un documento hipertextual de un sitio web
- ⦿ **Página principal o inicial:** Primera página de una sede web
- ⦿ **URL (Uniform Resource Locator = Localizador Uniforme de Recursos):** Sistema unificado de identificación de recursos en la red. Este tipo de direcciones permite identificar objetos WWW, Gopher, FTP, News... Ejemplos de URL son: <http://www.uji.es>. En el ámbito de la aplicación Web una URL es la dirección telemática de una web o de una sección de una web.

Conceptos básicos sobre los recursos web

- ◉ **Directorio:** Los índices o directorios constituyen una opción de búsqueda basada en palabras clave. Son una forma organizada de navegar por la información, ya que clasifican los recursos incluidos en su base de datos para facilitar el acceso a los mismos. El directorio puede (y suele) ser selectivo en la elección de servidores que incluye en la base de datos. Eg. Yahoo, LookSmart, (atención a las opciones de búsqueda avanzada que ofrecen).

Conceptos básicos sobre los recursos web

- ◉ Por otro lado, la implantación y el desarrollo del marketing *on-line* está promoviendo una serie de conceptos que, en cierta medida, vienen a solaparse con las categorías que acabamos de exponer. Nos referimos al fenómeno del «**portal**».
 - Un «portal» es un sitio web cuyo objetivo es ofrecer al usuario, de forma fácil e integrada, el acceso a una serie de recursos y de servicios, entre los que suelen encontrarse buscadores, foros, compra electrónica, etc.

Conceptos básicos sobre los recursos web

- ◉ Ej. *Google* es, desde un punto de vista formal, un sitio web, identificado por su URL (<http://www.google.com>), que dispone de varias páginas web (Búsqueda Avanzada, Preferencias, etc.). Consta de una página principal y es un recurso electrónico o digital porque requiere del ordenador para ser manipulado, además de conexión a Internet. Por lo que respecta a su contenido es un buscador, esto es, un robot que permite localizar y recuperar recursos electrónicos disponibles en la Red.



Aplicaciones básicas de Internet: servicios de aplicación básicos

Conexión remota (telnet): Protocolo que permite la conexión de un terminal a un ordenador remoto, que actúa como servidor o *host*.



Es decir, gracias a Telnet es posible acceder a otro ordenador y ejecutar alguna aplicación en dicho servidor, siempre y cuando se disponga de los servicios necesarios o se acceda a un servidor de servicio libre

Eg: acceso a bases de datos desde el servidor de la biblioteca

Aplicaciones básicas de Internet: servicios de aplicación básicos

- ⦿ **FTP: transmisión de ficheros** (*File Transfer Protocol*) es un sistema de transmisión de ficheros que hace que enviar y recibir archivos de gran tamaño de un lugar a otro de Internet sea rápido y más cómodo que mediante correo electrónico o la Web.
- ⦿ Los servidores FTP pueden ser
 - **privados** (con acceso restringido),
 - pero por lo general son de **acceso público**, lo que se denomina **FTP anónimo** (sin necesidad de clave ni contraseña).

Aplicaciones básicas de Internet: servicios de aplicación básicos

WWW (World Wide Web): Servicio de información electrónica que permite a los usuarios conectados a Internet a través del protocolo http acceder, distribuir y gestionar información de todo tipo. Está formada por la conexión hipertextual de multitud de hipervínculos (hiperenlaces) que se codifican mediante lenguajes de marcado, como HTML, XML, etc.

Aplicaciones básicas de Internet: comunicación electrónica

- **Correo electrónico:** servicio de mensajería que permite el intercambio de mensajes y de documentos adjuntos entre ordenadores a través de una red, tanto local como Internet.
- Teniendo una cuenta de correo electrónico (e-mail) se puede formar parte de las llamadas «**listas de correo**» o «**listas de distribución**». Ej: infotrad@uji.es

Aplicaciones básicas de Internet: comunicación electrónica

- ⦿ **Grupos de noticias (Newsgroup):** Grupo de personas que, aprovechando la infraestructura de USENET (red de acceso público de INTERNET que proporciona acceso a los grupos de noticias), intercambian información y discuten sobre un tema particular.
 - Las aportaciones quedan archivadas de forma jerarquizada en cadenas de respuestas, mostrando con claridad los debates que se van produciendo.
- ⦿ **Discusiones interactivas (*chats: conversational hypertext access technology*):** Aplicación que a través de Internet, con un programa específico, permite tener contacto directo, y en tiempo real, con otros usuarios del programa que en ese momento estén conectados a la red.

Internet en traducción

VENTAJAS

- Simplifica la búsqueda documental del proceso traductor
- Posibilita el teletrabajo, modalidad laboral que consiste en trabajar a distancia al disponer de una infraestructura telemática apropiada (ver el concepto de *Workstation* al final de esta presentación)
- Puede servir como medio publicitario para darse a conocer como traductor

(Palomares, 1999: 180-181)

PALOMARES PERRAUT, Rocío (1999) «Evaluación de recursos de comunicación para el traductor en Internet». En: Pinto, María y Cordón, José Antonio (eds.) *Técnicas documentales aplicadas a la traducción*. Madrid, Síntesis, pp. 179-193.

INCONVENIENTES

- La propia estructuración de la red hace que no sea eficaz
- El tiempo de espera que necesita la conexión y transferencia de información, sobre todo en horas punta cuando mucha gente está conectada y los servidores funcionan más lentamente
- Dudas acerca de la fiabilidad de la información que circula en Internet

Recuperación de la información (RI)

La **recuperación de información** es el conjunto de **tareas** mediante las cuales el usuario **localiza** y **accede** a los **recursos** de información que son pertinentes para la **resolución del problema planteado**.

Concepto de Sistema de recuperación de la información

Proceso donde se accede a una información previamente almacenada (y estructurada antes del almacenamiento), mediante herramientas informáticas que permiten establecer ecuaciones de búsqueda específicas.

Se compone de:

Documentos estructurados

Es necesario establecer un proceso donde se establezcan herramientas de indización y control terminológico

Bases de datos

Donde estén almacenados los documentos. Definir lenguajes de interrogación y operadores que soportará la base de datos y, establecer qué tipo de ecuaciones serán permitidas.

BASES DE DATOS

Las bases de datos constituyen la tecnología más avanzada para tratar grandes volúmenes de información



facilitan en gran medida las **tareas propias del proceso documental** y, en particular, las **tareas de almacenamiento y recuperación de la información.**

BASES DE DATOS

El término «base de datos» suele utilizarse indistintamente tanto para referirse:

- al programa que gestiona la información, denominado **Sistema de Gestión de Bases de Datos (SGBD)** y que permite crear, modificar y mantener un volumen de información
- al **conjunto de información que comparte características comunes**

Son dos cosas conceptualmente diferentes y que no deben confundirse: la primera acepción se refiere a la aplicación informática que gestiona y manipula una colección de datos. Ej.: Access de Microsoft o FileMaker de Claris.

Base de datos documental bibliográfica

conjunto de informaciones relativas a una materia determinada del conocimiento, registradas en soporte legible por ordenador y articuladas en unidades lógicas denominadas registros

La naturaleza de la información va a determinar:

- la organización de los datos que se van a almacenar y procesar.
- las formas en las que éstos se van a poder recuperar.

Bases de datos documentales: Tipología

Según la **naturaleza de la información** contenida en las bases de datos:

- ⦿ **Bases de datos referenciales:** contienen datos que hacen referencia a información primaria.
- ⦿ **Bases de datos factuales o fuente:** contienen los datos originales o información primaria directamente utilizable. Entre ellas, se encuentran:
 - Textuales: contienen los documentos originales o completos, tal y como han sido publicados.
 - Numéricas: contienen datos estadísticos, censos, encuestas...
 - Alfanuméricas: contienen datos de personas, objetos...
 - Icónicas: contienen imágenes, fotografías, carteles...
 - Multimedia: películas, vídeos...

Bases de datos documentales: Tipología

Dependiendo del tipo de **cobertura temática** pueden ser:

- ***Bases de datos multidisciplinares***: la documentación almacenada abarca distintas disciplinas científicas.
- ***Bases de datos especializadas***: la documentación almacenada abarca sólo una disciplina.

Según la **cobertura geográfica**:

- ***Internacionales***: almacenan la documentación publicada en cualquier lugar del mundo .Ej.: LISA.
- ***Nacionales***: almacenan la documentación publicada sólo en un ámbito nacional. Ej.: IME.

BASES DE DATOS

Con independencia de la clasificación que se acaba de exponer es necesario indicar que los sistemas de gestión de bases de datos son las aplicaciones más idóneas para estructurar la información, sea del tipo que sea.

Es decir: cualquier tipo documental, ya se trate de una enciclopedia, un directorio o guía, un anuario, un informe... puede estar estructurado en forma de base de datos, lo que origina una amplia gama de bases de datos.

Estructura interna de una base de datos

- Las bases de datos se componen de diferentes **registros** con su número de identificación correspondiente.



Cada uno se corresponde con una **referencia de un documento original**, que puede ser

- publicación independiente: monografía, tesis, informe...
- parte con autonomía de contenido dentro de otra publicación
- artículo incluido en una revista, ponencia o comunicación en las actas de una reunión científica...

La información contenida en cada registro bibliográfico se estructura en diferentes *campos* para facilitar su control sistemático y su recuperación individualizada. Algunos campos reflejarán la descripción formal del documento y otros van destinados a reflejar su contenido temático.

Estructura interna de una base de datos

- En una base de datos donde se incorporan referencias de artículos de revista se encuentran campos como los siguientes:
 - **Autor o autores** del documento original.
 - **Título del artículo.**
 - **Título de la revista** en la que aparece.
 - **Año de publicación.**
 - **Datos fuente:** volumen, número, páginas.
 - **Idioma** en el que está escrito el documento.
 - **Clasificación** que describe el tema central tratado en el documento.
 - **Palabras clave** que caracterizan cada uno de los temas tratados con suficiente relevancia en el artículo.
 - **Resumen:** texto breve que sintetiza el contenido del documento.

Estructura interna de una base de datos (ejemplo)

Base de datos ISOC: Registro de artículo de revista:

Núm. registro: 411476

Autores: Camacho Mene, María Teresa

Título del artículo: La inmigración en la comunidad de Murcia: Una experiencia de acogida

Lugar de trabajo: Cáritas Murcia, España

ISSN de la revista: 0417-8106

Título Revista: DOCUMENTACION SOCIAL

Datos fuente: 2000, (121): 295-307

Tipo documento: AR (artículo de revista)

Lengua: ES

Clasificación: 751407;751602

Localización: ISOC

Descriptor: Inmigrantes; Política social; Acción social; Integración social; Centros de acogida; Acogimiento familiar; Necesidades básicas; Formación profesional; Protección de menores

Identificadores: Cáritas (nombres propios de personas, obras, o instituciones)

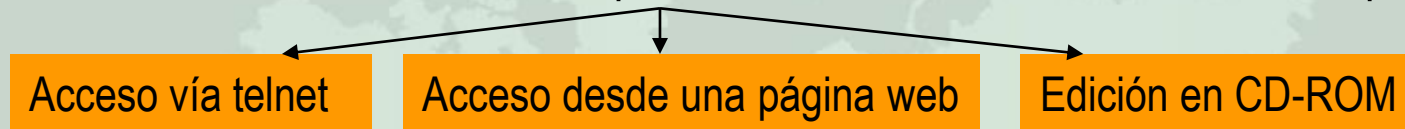
Topónimos: Región de Murcia (nombres propios de lugar)

Resumen: El fenómeno de la inmigración responde a causas estructurales, muchas de las cuales están motivadas por la injusticia, que hay que buscar tanto en la situación de los países de origen como de los de acogida. La inmigración no es un fenómeno nuevo en Europa, ni en España y ni siquiera en Murcia. La novedad está en que la Región ha pasado de ser emisora a receptora de emigrantes, y además en mayor número que la del resto de España, lo que supondría unos 45.000 inmigrantes..(A.)

(resumen cortado para elaborar este ejemplo)

La recuperación de información en bases de datos documentales

Un mismo sistema de información puede ser accesible desde distintos soportes



En el acceso tradicional en línea (telnet) la consulta puede realizarse a través

Formato de menús



el usuario va seleccionando sucesivas opciones

Lenguajes de recuperación



formados por comandos y operadores

La recuperación de información en bases de datos documentales

En la consulta en páginas web y en la mayoría de los CD-ROM



el usuario debe completar **formularios** diseñados en entorno Windows

Algunos sistemas permiten que la consulta pueda realizarse en varias bases de datos simultáneamente.

Esta opción es posible a través de algunos distribuidores de bases de datos y en determinadas ediciones en CD-ROM.



Opciones de búsqueda

1. Búsqueda directa



Se teclea directamente una o varias palabras en el espacio reservado para ello por el sistema de interrogación en la base de datos

Pueden distinguirse a su vez dos tipos dentro de esta modalidad

Interrogación en texto libre

Interrogación en campos individuales

Opciones de búsqueda

A. Interrogación en texto libre:

- El usuario realiza su consulta sin tener en cuenta la estructura de campos de los registros de la base de datos.
- Generalmente, el sistema busca el término o los términos solicitados en la totalidad de los campos de la base.

Búsqueda en campos básicos: (Título, descriptores, identificadores, topónimos, resumen)

alguna palabra ▾

microbiología

B. Interrogación en campos individuales:

El usuario realiza su consulta directamente sobre el campo o campos que ha seleccionado previamente: autor, revista, descriptores, etc...

Búsqueda por campos: Autores ▾

la frase ▾

Ruiz Hernández

Opciones de búsqueda

2. Búsqueda a través de índices



El usuario visualiza un diccionario o índice alfabético de las entradas de todos los campos o de uno concreto (índice de autores, descriptores,...) y selecciona las más adecuadas a su búsqueda.

El proceso puede parecer más lento pero es más seguro, ya que la búsqueda se adapta mejor al contenido real de los registros de una base de datos.

Índice por palabras

Índice por frase

Opciones de búsqueda

A. Índices por palabras

Incluyen solamente unitérminos (conceptos expresados en una sola palabra).

Se elaboran, por ejemplo, con todas las palabras extraídas del título o del resumen de los documentos.

Para reducir el tamaño de estos índices se establecen listas de “palabras vacías” (preposiciones, artículos,...).

Mostrar índice **del campo:** Título - Índice
comenzando desde: microbi

Documentos	Términos
1	<input type="checkbox"/> MICROBENTOS
1	<input type="checkbox"/> MICROBIALES
86	<input type="checkbox"/> MICROBIANA
22	<input type="checkbox"/> MICROBIANAS
32	<input type="checkbox"/> MICROBIANO
32	<input type="checkbox"/> MICROBIANOS
3	<input type="checkbox"/> MICROBICIDA

Ejemplo extraído de la consulta en la base de datos ICYT

Opciones de búsqueda

B. Índices por frase:

Incluyen entradas compuestas por varias palabras.

Por ejemplo: apellidos y nombre del autor, descriptores o encabezamientos de materias...

Mostrar índice **del campo:** Autores (frase) - Índice ▼
comenzando desde:
alvarez

Documentos	Términos
1	<input type="checkbox"/> <u>ALVARES, F.</u>
2	<input type="checkbox"/> <u>ALVAREZ ACEVEDO Y ALONSO, M.</u>
2	<input type="checkbox"/> <u>ALVAREZ AGUAYO, F.</u>
2	<input type="checkbox"/> <u>ALVAREZ AGUILAR, J.</u>
3	<input type="checkbox"/> <u>ALVAREZ AHEDO, J.</u>
3	<input type="checkbox"/> <u>ALVAREZ ALBA, J.</u>
3	<input type="checkbox"/> <u>ALVAREZ ALCON, M.</u>

Opciones de búsqueda

3. Búsqueda jerarquizada

A partir de un concepto genérico se pueden localizar, no sólo los registros en los que aparece dicho término, sino además todos aquellos en los que figure algún concepto más específico de su campo semántico.

Sólo será posible en aquellos sistemas de información que cuenten con campos codificados o que utilicen lenguajes estructurados jerárquicamente en la clasificación o en los descriptores

4. Búsqueda a través de códigos

En determinados campos de algunas bases de datos la interrogación se realiza a través de códigos numéricos o alfanuméricos. Pueden ser campos codificados la clasificación, el idioma, la tipología documental...

Búsquedas en bases de datos

Pasos recomendados:

- ⦿ *Seleccionar la base de datos* más adecuada a la consulta.
- ⦿ *Familiarizarse* con las características de la base de datos y del sistema de recuperación.
- ⦿ *Seleccionar los conceptos de búsqueda* que representen de forma precisa el tema de interés.
- ⦿ *Determinar la estrategia de búsqueda*: en qué campos puede encontrarse de forma más pertinente la información buscada y cómo combinar los diferentes conceptos, con operadores booleanos o de adyacencia.
- ⦿ *Realizar la búsqueda y visualizar los resultados*, valorando la adecuación de los mismos y la eficacia de la estrategia utilizada.
- ⦿ Si el resultado es poco satisfactorio o existen dudas sobre su idoneidad, se debe *repetir la búsqueda* ampliando o limitando la estrategia utilizada.
- ⦿ *Seleccionar los registros adecuados* a la necesidad de información.

Búsquedas en bases de datos

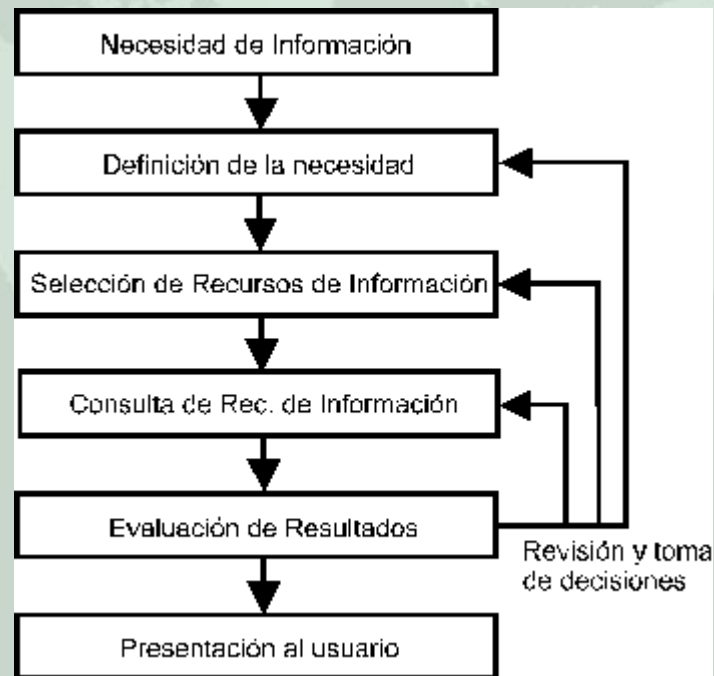
Se recomienda tener en cuenta:

- Si el sistema permite la consulta de una única base de datos o de varias simultáneamente.
- Cuál es su cobertura temática.
- Si se trata de un catálogo o de una base de datos documental con descriptores y/o con resúmenes.
- En el caso de que haya descriptores o encabezamientos de materias, si éstos proceden o no de un léxico controlado accesible.
- Cuál es la estructura de campos y en cuáles se realiza la búsqueda por defecto cuando no se selecciona uno concreto.
- Cuando se teclean dos términos se debe saber si el sistema interpreta que debe buscar registros en los que aparezcan ambos a la vez (intersección) o bien cualquiera de los dos (suma booleana).

Proceso de recuperación de la información

1. Definición de las necesidades informativas del usuario.
2. Selección y ordenación de las fuentes a utilizar.
3. Traslación de las necesidades del usuario al lenguaje documental propio de la fuente a utilizar en cada caso.
4. Traducción de la expresión de lenguaje documental al lenguaje de interrogación propio de cada sistema.
5. Ejecución de las expresiones del lenguaje de interrogación obtenidas.
6. Consulta de las respuestas obtenidas.
7. Replanteamiento, si procede, de las expresiones utilizadas.
8. Selección y obtención de los documentos.
9. Transmisión del resultado, preparado adecuadamente, al usuario.

Proceso de recuperación de la información



Lenguajes de interrogación

- Se puede definir un lenguaje de interrogación como un conjunto de órdenes, operadores y estructuras que, organizados según unas normas lógicas, permiten la consulta de fuentes y recursos de información electrónica.
- El resultado de la combinación de estos elementos, siguiendo las normas establecidas, es una expresión, a la que se identifica con el nombre «ecuación», capaz de interrogar el contenido de la fuente de información.

Herramientas de búsqueda

- ⦿ Precisar los **aspectos temáticos implicados** (y la posibilidad de expresarlos, mediante sinónimos, etc.)
- ⦿ Construir una **estrategia** y relacionar de forma clara los diferentes términos utilizados
- ⦿ Tener **criterios** para realizar búsquedas en Internet
- ⦿ No es fuente de información todo lo que consta en Internet
- ⦿ Búsquedas en Internet
 - Tener en cuenta la autoría de la información
 - Cuidado con la caducidad de la información
 - No olvidar la existencia de la Internet profunda
 - Ejemplo: [INTERNET INVISIBLE](http://www.internetinvisible.com)
<http://www.internetinvisible.com>

Herramientas de búsqueda

Buscadores:

Almacenan las páginas con determinadas características (metadatos) y posteriormente tras utilizar unas palabras clave emiten un listado de las más relevantes.

Generales:

Google <http://www.google.com>
Alltheweb <http://www.alltheweb.com>
Altavista <http://www.altavista.com>
Excite <http://www.excite.com>
Infoseek <http://www.infoseek.com>
Lycos <http://www.lycos.com>
Webcrawler <http://webcrawler.com>
Hotbot <http://www.hotbot.com>

Directorios y motores especializados

Humbul <http://www.humbul.ac.uk>
Librarian Index to de Internet <http://lii.org>
Internet Public Library <http://www.ipl.org>
Scirus <http://www.scirus.com>
Search4Science
<http://www.search4science.com>

Herramientas de búsqueda

Metabuscadores:

- son buscadores.
- al introducir los conceptos de búsqueda hace el barrido en distintas bases de datos.
- la amplitud de resultados es mayor.

Vivísimo <http://www.vivisimo.com>

Dogpile <http://www.dogpile.com>

Kartoo <http://www.kartoo.com>

Qbsearch <http://www.qbsearch.com>

Metacrawler <http://www.metacrawler.com>

Agentes inteligentes:

- permiten localizar información de forma automática.
- necesitan que se defina un perfil de búsqueda y dónde debe lanzarla (bases de datos, sitios web...)
- automáticamente va presentando un informe sobre la nueva información que va surgiendo.

BookWhere <http://www.bookwhere.com>

BullEye Pro <http://www.intelliseek.com>

WebSeeker 5

<http://www.bluesquirrel.com>

WebFerret <http://www.ferretsoft.com>

Metodología de búsqueda

- Definir de manera clara lo que se busca, acotando tanto como sea posible.
- Definir en forma de palabras claves nuestra demanda de información.
- Conocer los distintos tipos de documentos que existen para poder decidir, en función del conocimiento que inicialmente poseamos del tema, qué documentos debemos consultar: obras de referencia, fuentes primarias...
- Conocer los distintos tipos de organizaciones (bibliotecas, hemerotecas, filmotecas...) a los que podemos acudir en busca del material documental que nos es necesario.
- Definir el idioma de consulta.
- Diseñar la estrategia de búsqueda a seguir.
- En caso de navegar por Internet conviene seguir una estrategia sistemática:
 - 1º herramientas de búsqueda por índice temático del tipo Yahoo!
 - 2º herramientas de búsqueda por contenidos, del tipo Altavista.
 - 3º como complemento se pueden utilizar uno o varios metabuscadores.

Operadores para búsquedas en Internet

1. Operadores lógicos booleanos	
AND (“Y”)	<p>Indica que han de aparecer en el mismo documento todos los términos que especificamos restringiendo, de este modo la búsqueda.</p> <p>ej.: <i>política AND europa (política + europa)</i></p>
OR (“O”)	<p>Indica que han de aparecer en el documento indistintamente uno u otro de los términos que especificamos, ampliando la búsqueda.</p> <p>ej.: <i>europa AND (política OR gobierno)</i></p>
NOT (“No”)	<p>Indica que sólo se recuperarán aquellos documentos en los que figure el primer término especificado pero nunca el segundo; esto es, excluye al segundo término.</p> <p>ej.: <i>europa NOT francia (europa-francia)</i></p>

Operadores para búsquedas en Internet

La combinación de diferentes operadores en una misma búsqueda puede exigir la necesidad de utilizar paréntesis para marcar claramente el orden del proceso de búsqueda. Ejemplo, si se desea buscar documentos sobre artes plásticas, escultura o pintura en Italia no sería correcta la siguiente fórmula:

Italia AND Artes plásticas OR Escultura OR Pintura

En este caso la intersección con el conjunto “Italia” afectaría solamente a “artes plásticas”. Por el contrario, se debería utilizar un paréntesis para conseguir que la intersección se realice con los tres términos siguientes y no sólo con el primero:

Italia AND (Artes plásticas OR Escultura OR Pintura)

Operadores para búsquedas en Internet

2. Operadores de proximidad o sintácticos

ADYACENCIA (ADJ)	<p>Recupera los términos introducidos en la misma posición y en el orden establecido, sin posibilidad de existir otros términos entre medias. Una variante es el empleo de comillas, pues sirven igualmente para buscar una frase exacta.</p> <p><i>“política agraria europea”</i></p>
PROXIMIDAD (NEAR)	<p>Recupera todos los documentos en los que el término introducido en segundo lugar está próximo al primer término pero no necesariamente en el orden marcado, es decir, puede estar entre los 80 caracteres anteriores o posteriores al término introducido en primer lugar.</p>
FOLLOWED BY	<p>Recupera todos los documentos en los que el término introducido en segundo lugar se encuentra respecto al primer término, expresados en el mismo orden de la pregunta, separados por un máximo de ochenta caracteres.</p>

Operadores sintácticos, de proximidad o de adyacencia

Se desea localizar registros que traten sobre el escritor 'Juan Madrid' la expresión 'Juan and Madrid' dará como resultado todas las referencias en las que aparecen ambos términos pero no necesariamente relativos a este autor.



Solución: combinar conceptos relacionados entre sí por criterios sintácticos

Ej.: se selecciona el operador de *búsqueda por frase* y se teclea la expresión 'Juan Madrid' el sistema seleccionará sólo aquellos documentos en los que estas dos palabras aparezcan juntas y en este orden.

Operadores para búsquedas en Internet

2. Operadores de truncamiento (grafías)

ASTERISCO	<p>Para buscar las diferentes formas de una palabra se agregará un asterisco (*) que sustituye a cualquier cadena de caracteres.</p> <p>Ej: <i>document* recuperaremos: document-ación, document-ar, document-al, document-ales, documenta-lista...</i></p>
INTERROGACIÓN	<p>Sustituye sólo a un carácter, normalmente, al final del término, aunque también puede colocarse en medio de una palabra.</p> <p>Ej.: <i>polític? recuperaremos: politic-a, politic-o</i> <i>ca?a recuperaremos: casa, caza, cala...</i></p>
PUNTO	<p>Se utiliza (especialmente en Lycos) para establecer los límites exactos de un término.</p> <p>Ej.: <i>documental. No recuperamos términos de la misma raíz, como documentales, documentalista...</i></p>

Truncado

Los sistemas de recuperación permiten, mediante los signos de truncado (*, ? o \$, según el programa), realizar una búsqueda conjunta de todas las palabras de un índice que comparten un mismo comienzo o raíz.

Ejemplo, con la expresión: 'Econom*' se recuperan todos los documentos en los que figure cualquier término de esta raíz: economía, económico, económica...

Tipos de truncado:

- por la derecha (sufijos)
- por la izquierda (prefijos)
- truncado limitado a uno o varios caracteres dentro de una palabra.

Recomendaciones de búsqueda por Internet

- 1- Efectuar las búsquedas escribiendo los términos preferentemente en minúsculas y sin acentos.
- 2- Vigilar los efectos negativos de la sinonimia y la polisemia.



Workstation (estación de trabajo)

- ⦿ Este término se refiere al ordenador para uso personal
- ⦿ En el ámbito de la traducción, se refiere al equipo con el que cuenta todo/a traductor/a para desarrollar profesionalmente su trabajo.
- ⦿ Es básico saber que quien traduce ha de contar con un equipamiento informático adecuado:
 - estar conectado a Internet
 - disponer de materiales de consulta organizados
 - programas de edición de textos
 - correo electrónico...

Programas informáticos de asistencia a la traducción

- Breve introducción de estos recursos que forman parte de los contenidos de la materia de *Informática Aplicada a la Traducción*
- Estos programas abarcan desde la **ayuda terminológica** hasta la **traducción automática**

Programas informáticos de asistencia a la traducción

- Destacan las memorias de traducción, también conocidas como «**CAT-TOOLS**»



Son programas informáticos diseñados para ayudar en la traducción
Suelen utilizarse en combinación con un procesador de texto



Constan de una **base de datos** que contiene segmentos de texto (palabras sueltas o frases) en el idioma de partida y la traducción de los mismos en una o más lenguas de llegada

El traductor profesional va creando y almacenando la traducción con la ayuda del programa a medida que desarrolla su trabajo. Podrá utilizarla en el mismo documento (siempre que se presente una coincidencia con un segmento ya almacenado) o en trabajos sucesivos.

Programas informáticos de asistencia a la traducción

Para reutilizar los segmentos ya traducidos y proponerlos de nuevo



- algunas memorias de traducción buscan sólo correspondencias literales, es decir, sólo pueden recuperar coincidencias exactas de las frases
- otras utilizan algoritmos para encontrar cadenas de texto similares (*fuzzy matches*) en las que marcan las diferencias

El traductor debe traducir de forma manual los segmentos para los que no se encuentre coincidencia, segmentos que se añaden a la base de datos para utilizarse en futuras traducciones.

Programas informáticos de asistencia a la traducción

Las memorias de traducción funcionan de forma óptima con **textos muy repetitivos**, como los manuales técnicos.

También son útiles para realizar revisiones, correcciones y pequeñas modificaciones. Utilizar una memoria de traducción de forma reiterada en los textos adecuados durante algún tiempo puede ahorrar mucho trabajo a un traductor.

Las memorias de traducción no son demasiado productivas para textos literarios o creativos por el hecho de que hay muy poca repetición en el lenguaje empleado, aunque su utilización puede ayudar de todos modos al traductor, pues la utilización de estos programas ayuda a mantener de forma automática el formato del texto (negrita, tipos de fuente y de estilos utilizados, etc).

