

Págs.:

Concepto e historia de Internet.....	2
Funciones y servicios de Internet.....	4
Concepto de WWW.....	5
–Pagina Web.....	6
–Sitio Web.....	7
–Hospedaje Web.....	8
Correo Electrónico (E–Mail).....	9
¿Qué es el Outlook Express.....	11
Concepto de Netscape Messenger.....	12
Concepto de Telnet.....	12
Concepto de Netscape.....	13
Concepto de Protocolo TCP/IP.....	13
Concepto de FTP (Transferencia de ficheros).....	13
–FTP Anónimo.....	14
–Ficheros comprimidos.....	15
–Una sesion FTP con su cuenta de usuario.....	16
–Inicio de sesion con FTP.....	16
–Comandos de ayuda FTP.....	17
Motores de búsqueda.....	17
Los elementos que componen un motor de búsqueda.....	22
–Buscador sin Robot.....	24
–Buscador con Robot.....	25
Funcionamiento de los motores de búsqueda.....	25
Algunos motores de búsqueda y sus diferencias.....	28

–Yahoo.....	28
–Altavista.....	31
–Lycos.....	33
–Infoseek.....	35
–Webcrawler.....	37
Bibliografía.....	40

INTERNET: Concepto e Historia:

Se suele definir a **Internet** como la "**red de redes**". En esencia es una definición cierta, pero un poco parca para el profano en comunicaciones. Ciertamente se trata de ordenadores –millones de ellos– que están ínter conexionados entre sí, mayoritariamente mediante cables. Es una red descentralizada, donde unas redes convergen con otras de diferente tamaño y de diferente tecnología. Redes dispersas geográficamente se ven unidas gracias a **Internet**.

Lo que hace que todos estos ordenadores puedan comunicarse entre sí, es un juego de protocolos (programas de comunicación) común a todos ellos. Se trata de los **TCP/IP**, desarrollados en 1974 por **Robert Kahn**.

El **Internet**, fue concebido por la agencia de nombre **ARPA** (Advanced Research Projects Agency) del gobierno de los Estados Unidos en el año de 1969 y se le conocía inicialmente como **ARPANET**. El propósito original fue crear una red que permitiera a los investigadores en un Campus poder comunicarse a través de los sistemas de cómputo con investigadores en otras Universidades.

Robert Kahn dio el empuje tecnológico que necesitaba **ARPANET**, la semilla de **Internet**, que había comenzado en 1969 como un proyecto patrocinado con fondos del Ministerio de Defensa de los EEUU.

La intención del proyecto **ARPANET** era asegurar el flujo de información entre lugares distantes en caso de conflicto bélico, aunque muchos ordenadores fueran destruidos entre los puntos del emisor y el receptor, la información debía conseguir llegar a su destino, reconociendo el mejor camino para ello.

En 1983 se unieron a **ARPANET** otras redes dando el primer paso a la construcción de **Internet** tal y como hoy la conocemos. Su utilidad en entornos científicos fue el principal argumento para que la red fuera creciendo. Un científico de Los Angeles podía enviar información para que la procesara un superordenador en Nueva York y este le devolviera los resultados en un tiempo récord.

A estas primeras redes, se fueron sumando otras muchas hasta llegar a cubrir hoy en día a casi todos los países del mundo.

La herramienta principal de **Internet** ha sido el **correo electrónico**, aunque otras muchas se le sumaron rápidamente, la transferencia de ficheros (**FTP**), la charla en tiempo real (**IRC**), las charlas diferidas a modo de tablón de anuncios (**News**), el **TELNET** que permite manejar un ordenador desde otro distante en la red, el **Gopher** donde poder publicar y leer documentos y finalmente el **WWW** (World Wide Web), más conocido como Web, que ha sido el detonador final para la expansión tan impresionante que ha tenido la red desde hace unos años.

Hoy en día, el **Internet** es un medio de comunicación público, cooperativo y autosuficiente en términos económicos, accesible a cientos de millones de gentes en el mundo entero. Físicamente, el **Internet** usa parte

del total de recursos actualmente existentes en las redes de telecomunicaciones. Técnicamente, lo que distingue al **Internet** es el uso del protocolo de comunicación llamado **TCP/IP** (Transmission Control Protocol/Internet Protocol).

En sus primeras versiones solo era posible incluir texto, aunque con posterioridad avanzó hacia el que conocemos hoy en día, un lugar donde poder interactuar con otro ordenador que nos puede mostrar imágenes, vídeo, sonido, etc., y además con un simple movimiento de ratón nos ofrece navegar hacia otra página o hacia otro ordenador.

Para muchos usuarios del **Internet**, el **correo electrónico (e-mail)** ha reemplazado prácticamente al servicio postal para breves mensajes por escrito. El **correo electrónico** es la aplicación de mayor uso en la red. También se pueden realizar conversaciones "en vivo" con otros usuarios en otras localidades usando el **IRC** (Internet Relay Chat). Más recientemente, el software y hardware para telefonía en **Internet** permite conversaciones de voz en línea.

Hoy en día el servicio mas popular de **Internet** es el **correo electrónico**, o el **e-mail** esta es la aplicación mas común de **Internet** ya que es uno de los sistemas de comunicación mas importante, tanto por su rapidez como su eficacia. Al estar todos los ordenadores conectados a la misma red, es posible contactar con cualquiera de ellos.

FUNCIONES Y SERVICIOS DE INTERNET:

Internet es una red extensa construida por una cantidad de redes menores de ámbito mundial conectadas entre si. A través de **Internet** se puede acceder a servicios como transferencias de archivos, correo electrónico, noticias, investigaciones y juegos, entre otros.

Una de las características importantes de **Internet** es que se trata de un medio de comunicación interactivo.

Esta red que conecta a muchos redes informáticas de todo el mundo, se utiliza como canal de comunicación para mensajes electrónicos, pero también contiene enormes cantidades de información muy útil, almacenada por personas, centros educativos y de investigación, así como organizaciones comerciales.

A cada ordenador se le asigna una dirección, única en la red. Este servicio lo suministra también los servidores cuando nos damos de alta en una cuenta. Existen programas para gestionar la recepción y envío de mensajes e incluso los principales navegadores incluyen este programa.

Las posibilidades que ofrece **Internet** se denominan servicios. Cada servicio es una manera de sacarle provecho a la Red independiente de las demás. Una persona podría especializarse en el manejo de sólo uno de estos servicios sin necesidad de saber nada de los otros.

Sin embargo, es conveniente conocer todo lo que puede ofrecer Internet, para poder trabajar con lo que más nos interese. Hoy en día, los servicios más usados en Internet son: **Correo Electrónico, World Wide Web, FTP, Grupos de Noticias, IRC y Servicios de Telefonía.**

CONCEPTO DE WWW (WORDL WIDE WEB):

World Wide Web, o simplemente **Web**, es el universo de información accesible a través de Internet, una fuente inagotable del conocimiento humano.

El **WWW** fue un proyecto iniciado desde el **CERN** (Laboratorio Europeo de Física de Partículas) y su principal objetivo era relacionar diferentes documentos mediante la técnica del hipertexto.

El componente más usado en el **Internet** es definitivamente el **Web**. Su característica sobresaliente es el texto remarcado, un método para referencias cruzadas instantáneas. En la mayoría de los Sitios **Web**, ciertas palabras aparecen en texto de otro color diferente al resto del documento. Por lo general, este texto es subrayado. Al seleccionar una palabra o frase, uno es transferido al sitio o página relacionado a esa frase.

En algunas ocasiones hay botones, imágenes, o porciones de imágenes que pueden activarse mediante un clic. Si Usted mueve el apuntador sobre el contenido del documento y el apuntador cambia a un símbolo con una mano, eso indica que Usted puede realizar un clic para ser transferido a otro sitio.

Usando el **Web**, se tiene acceso a millones de páginas de información. La exploración en el **Web** se realiza por medio de un software especial denominado **Browser** o **Explorador**. La apariencia de un Sitio **Web** puede variar ligeramente dependiendo del explorador que use. Así mismo, las versiones más recientes disponen de una funcionalidad mucho mayor tal como animación, realidad virtual, sonido y música.

Sin duda, el **WWW** (World Wide Web) ha sido el detonante para la expansión de **Internet** más allá del ámbito académico y científico. Sus páginas plagadas de atractivos gráficos, la característica hipertexto que contiene y la nueva visión que ofrece como medio de comunicación, hace que el **Web** tenga los ingredientes para convertirse en la semilla del medio de comunicación del futuro.

Hasta hoy, vivimos en una dictadura de los medios de comunicación.. Pocos medios publican, frente a los millones de receptores. Con el **Web** todo se invierte, cualquier persona puede publicar con el mínimo coste, con igualdad de condiciones (en teoría) con cualquier multinacional de la comunicación.

Es lógico pensar, que el **Web** siga aumentando los contenidos para llegar a cubrir cualquier necesidad que a un usuario se le ocurra: información, compras, entretenimiento, formación, etc.

La evolución es imparable y ya hay quien se aventura en predecir, que cuando se produzcan las condiciones tecnológicas necesarias, el Web absorberá a la televisión. Aunque todo esto está por ver.

El Web debemos entenderlo como dos partes diferenciadas, la parte del ordenador que contiene la dirección que visitamos (**servidor**) y nuestro ordenador (**cliente**) que interpreta la información con un programa específico, llamado navegador (**browser en inglés**).

PAGINA WEB:

Una página de **Internet** o **página Web** es un documento electrónico que contiene información específica de un tema en particular y que es almacenado en algún sistema de cómputo que se encuentre conectado a la red mundial de información denominada Internet, de tal forma que este documento pueda ser consultado por cualesquier persona que se conecte a esta red mundial de comunicaciones y que cuente con los permisos apropiados para hacerlo.

Una **página Web** es la unidad básica del **World Wide Web**.

Una **página Web** tiene la característica peculiar de que el texto se combina con imágenes para hacer que el documento sea dinámico y permita que se puedan ejecutar diferentes acciones, una tras otra, a través de la selección de texto remarcado o de las imágenes, acción que nos puede conducir a otra sección dentro del documento, abrir otra página **Web**, iniciar un mensaje de **correo electrónico** o transportarnos a otro Sitio **Web** totalmente distinto a través de sus hipervínculos.

Estos documentos pueden ser elaborados por los gobiernos, instituciones educativas, instituciones públicas o privadas, empresas o cualquier otro tipo de asociación, y por las propias personas en lo individual.

Las **páginas Web** están construidas con un lenguaje de marcas bastante sencillo. Las marcas indican al ordenador cliente donde debe colocar los objetos (texto, imágenes, etc.) en el interior de la página y también le indica las direcciones de los enlaces que contiene. Así, con solo ligeros movimientos del ratón podemos desplazarnos por una página o bien saltar a otra distinta que se halle relacionada en la primera.

Esta característica de hipertexto en la red, ha dado lugar al término "**telaraña mundial**".

SITIO WEB :

Es un conjunto de archivos electrónicos y **páginas Web** referentes a un tema en particular, que incluye una página inicial de bienvenida, generalmente denominada **home page**, con un nombre de dominio y dirección en **Internet** específicos.

Empleados por las instituciones públicas y privadas, organizaciones e individuos para comunicarse con el mundo entero. En el caso particular de las empresas, este mensaje tiene que ver con la oferta de sus bienes y servicios a través de **Internet**, y en general para eficientar sus funciones de mercadotecnia.

Su Sitio **Web** no necesariamente debe localizarse en el sistema de cómputo de su negocio. Los documentos que integran el Sitio **Web** pueden ubicarse en un equipo en otra localidad, inclusive en otro país. El único requisito es que el equipo en el que residan los documentos esté conectado a la red mundial de **Internet**. Este equipo de cómputo o **Servidor Web**, como se le denomina técnicamente, puede contener más de un sitio **Web** y atender concurrentemente a los visitantes de cada uno de los diferentes sitios .

Al igual que los edificios, oficinas y casas, los Sitios **Web** requieren de una dirección particular para que los usuarios puedan acceder a la información contenida en ellos.

Estas direcciones, o **URLS** (por sus siglas en inglés **Uniform Resource Locator**), aparecen cotidianamente en todos los medios de comunicación como son prensa escrita, radio, televisión, revistas, publicaciones técnicas y en el propio **Internet** a través de los **motores de búsqueda** (por su denominación en inglés **search engines**). Los nombres de estos sitios **Web** obedecen a un sistema mundial de nomenclatura y están regidos por el **ICANN** (Internet Corporation for Assigned Names and Numbers).

Los Sitios **Web** pueden ser de diversos géneros, destacando los sitios de negocios, servicio, comercio electrónico en línea, imagen corporativa, entretenimiento y sitios informativos.

PORTAL :

Portal es un término, sinónimo de puente, para referirse a un Sitio **Web** que sirve o pretende servir como un sitio principal de partida para las gentes que se conectan al **World Wide Web**. Son sitios que los usuarios tienden a visitar como sitios ancla.

Los **portales** tienen gran reconocimiento en **Internet** por el poder de influencia que tienen sobre grandes comunidades.

La idea es emplear estos portales para localizar la información y los sitios que nos interesan y de ahí comenzar nuestra actividad en **Internet**. Un Sitio **Web** no recibe el rango de portal por tratarse de un sitio robusto, importante o por contener información relevante. Un **portal** es mas bien una plataforma de despegue para la navegación en el **Web**.

HOSPEDAJE WEB :

Es el servicio de almacenamiento, acceso y mantenimiento de los archivos que integran un Sitio **Web**. Más

importante que el espacio en disco provisto para estos archivos, es el acceso rápido al **Internet** lo que adquiere mayor relevancia. Una empresa que pretenda hospedar su Sitio **Web** en sus propias instalaciones, requiera invertir una fuerte cantidad en recursos de equipos, sistemas y medios de comunicación generalmente caros. Los servicios de **Hospedaje Web** permiten a las compañías compartir el costo de una conexión rápida a **Internet**.

Cuando un Sitio **Web** llega a ser muy robusto, es muy probable que el servidor **web** en el que se encuentren instalados los archivos electrónicos que lo integran, sea dedicado única y exclusivamente a atender a este sitio. Este servicio se conoce como **Hospedaje Web** dedicado. En este caso, el equipo de cómputo que funciona como **servidor web** puede ser propiedad de la empresa que publica el Sitio **Web** o del propio proveedor de **Internet**.

CORREO ELECTRÓNICO (E- MAIL):

El **correo electrónico** (email, electronic mail) es el intercambio de mensajes almacenados en computadora por medio de las telecomunicaciones. Los mensajes de **correo electrónico** se codifican por lo general en formato de texto **ASCII** (American Standard Code for Information Interchange). Sin embargo, se pueden también enviar archivos en otros formatos, tales como imágenes gráficas y archivos de sonidos, los cuales son transferidos como archivos anexos en formato binario.

El **correo electrónico** representa una de las primeras aplicaciones del **Internet** y sigue siendo la de mayor uso. Un alto porcentaje del tráfico total en el **Internet** se debe al **correo electrónico**. El **correo electrónico** puede ser también intercambiado entre usuarios de proveedores privados de servicios en línea, tales como **America On Line** y **Compuserve**, y aquellos conectados a través de otros medios diferentes al **Internet**, redes públicas y privadas.

El **correo electrónico** es la principal herramienta en **Internet** y es una de sus principales aplicaciones.

Podemos escribir mensajes a cualquier parte del mundo, recibiéndose en segundos con total seguridad, si un mensaje no encuentra a su destinatario, es devuelto inmediatamente a su origen indicando las razones de su devolución.

El **correo** puede ser enviado tanto a individuos en lo particular como a listas de distribución. Una lista de distribución puede ser administrada usando software especializado (**e-reflector**). Algunas listas de **correo** permiten a uno suscribirse con tan sólo enviar una solicitud a su administrador. Una lista de correo que se administra automáticamente se le nombra servidor de listas (**list server**).

El **correo electrónico** es uno de los protocolos incluidos dentro del **TCP/IP (Transport Control Protocol/Internet Protocol)**, protocolo que gobierna la operación del **Internet**. Un protocolo comúnmente empleado para enviar correo electrónico es el **SMTP (Simple Mail Transfer Protocol)** y un protocolo muy popular para recibirlo es el **POP3 (Post Office Protocol 3)**. Tanto **Netscape** como **Microsoft** incluyen un software básico para administración de correo junto con sus programas para navegación en el **Web**.

Para entender el **correo electrónico** (email en inglés) se suele hacer un símil con el correo convencional: un emisor, un transporte y un buzón de llegada. En **Internet** ocurre de manera idéntica, solo que los buzones son espacios destinados en los ordenadores conectados en la red, espacios identificados bajo una dirección electrónica, con el formato:

nombre@organizacion.dominio

Así tenemos que en **Internet** una dirección electrónica corresponde con un único usuario. Además de texto, actualmente se puede enviar mediante el correo electrónico todo tipo de información digital: fotografías,

sonidos, documentos, etc.

Una de las principales aplicaciones derivadas del **correo electrónico** son las "**news**".

Para entender mejor este concepto, mejor volver a utilizar un símil: imaginemos un tablero de anuncios donde toda persona interesada en comprar o vender productos cuelgan su anuncio o comentario de manera ordenada.

Las "**news**" no son más que un conjunto muy amplio de tableros de anuncios. Las "**news**" suele ser un buen lugar para conocer a personas interesadas, bien por su profesión, bien por afición, en temas que nos interesen. Como en cada grupo de "**news**" suelen haber preguntas que se repiten con mucha asiduidad, se crearon las **FAQ (Frecuent Asked Question)**. Preguntas Formuladas Frecuentemente.

Las **FAQs** son archivos con las preguntas y respuestas más habituales, es por donde deberíamos empezar, leyendo la **FAQ** correspondiente a un grupo nos haríamos una idea de los temas que se tratan en un lugar determinado y probablemente nos respondería a unas cuantas dudas que tendríamos.

Otra gran aplicación del **correo electrónico** son las **listas de distribución**. Se trata de una suscripción a un debate con un grupo de personas con interés común por un tema determinado. Es similar a las **news** aunque tiene una peculiaridad.

Todos los mensajes son enviados a un ordenador que actúa como servidor de la lista, y este se encarga de distribuir a todos sus subscriptores una copia del mensaje enviado. De esta manera, solo debemos enviar un primer mensaje al ordenador servidor para que todos los mensajes generados en la lista nos lleguen directamente a nuestro buzón de **correo electrónico**.

:Qué es Outlook Express:

Outlook Express es uno de los mejores programas para administrar correo electrónico, muy utilizado en el **Internet**, es de **Microsoft** y, casi siempre, viene conjuntamente con el navegador **Microsoft Internet Explorer**.

Una particularidad de **Outlook Express** es que puede enviar mensajes que contengan código **HTML**, es decir mensajes con formatos muy avanzados, otros administradores de **e-mail**, como **Eudora Light**, solo pueden enviar texto puro en el mensaje.

Otra particularidad de **Outlook Express** es la capacidad de administrar más de una casilla de correo a la vez, aquí llamadas "**cuentas**".

Concepto De netscape messenger:

Netscape Messenger es una aplicación que permite recibir y enviar correo electrónico (**e-mail**) desde un PC con conexión a **Internet**. Además incluye una serie de utilidades que complementan y mejoran la herramienta:

Agenda de Direcciones: **Netscape Messenger** permite mantener una agenda de personas guardando de cada una de ellas su nombre, apellidos, dirección de mail, teléfono, entre otras muchas cosas.

Buscador de direcciones mail en directorios LDAP: podemos buscar direcciones electrónicas de personas que tengan cuenta en el servidor de la Universidad (P.A.S., profesores y alumnos).

Servicio de News (Noticias): **Netscape Messenger** permite configurar y leer todos los grupos de noticias que ud. desee.

Este programa se distribuye incluido en el paquete **Netscape Communicator**, que incluye **Netscape Navigator**, **Netscape Composer**, **Netscape Messenger**, etc. Si necesita mas información sobre este paquete o quiere descargar el programa de instalación desde Internet puede visitar la página **web** de **Netscape**:

CONCEPTO DE TELNET:

TELNET Tele Network. Tele Red. Conexión a un **Host** en la que el ordenador cliente emula un terminal de manera que se configura como terminal virtual del ordenador servidor.

Telnet (sesiones con ordenadores remotos): podemos utilizar otros ordenadores para ejecutar aplicaciones en ellos, aprovechando de esta forma la capacidad de procesamiento y herramientas que éstos posean.

CONCEPTO DE NETSCAPE:

Netscape: popular navegador de páginas **HTML** (World Wide Web) creado por la empresa **Netscape Inc.** que empezó a cotizar en bolsa de Nueva York a principios de 1995. Actualmente es el mejor **browser** existente.

CONCEPTO DE PROTOCOLO TCP/IP: (Transport Control Protocol/Internet Protocol)

Protocolo: conjunto de directrices que regulan las comunicaciones entre ordenadores. Existen protocolos para diversas tareas: transferencia de archivos (en cualquier sentido), verificación de errores, control de flujo, etc.

CONCEPTO DE FTP (TRANSFERENCIA DE FICHEROS)

Las redes de ordenadores se crearon para el intercambio de información y la compartición de recursos. Aunque, utilizando el **correo electrónico** es posible el intercambio de información, hay ocasiones en las que necesitamos mover ficheros por la red. Los nuevos estándares de **correo** nos permiten realizar esta operación (**ficheros adjuntos**), pero existe un servicio especial en **Internet** dedicado a la transferencia de ficheros entre dos ordenadores .

El **Protocolo de Transferencia de Ficheros (FTP)** le permite conectarse a un ordenador y transferir ficheros, de tipo texto y de tipo binario, entre su máquina (local) y la máquina a la que se conecta (máquina remota).

Para utilizar este servicio necesita un cliente **FTP**, que es el que establece la conexión con el programa servidor **FTP** situado en la máquina remota. Si ya dispone de uno y su ordenador se encuentra configurado, lo único que le queda por hacer es indicarle al cliente la máquina remota a la que quiere conectarse. Cuando se conecte a este tipo de máquinas necesitará permiso para poder acceder a sus ficheros. Es decir, el servidor **FTP** le solicitará un nombre de usuario (**login**) y una contraseña (**password**).

Supongamos que usted tiene una cuenta de usuario en una máquina **Unix**, donde tiene todos sus ficheros. En un momento determinado le interesa transferir los ficheros de su cuenta a su PC, ya que le es más cómodo trabajar con este tipo de máquina. Utilizando un cliente **FTP** puede conectarse a su cuenta, identificarse (**login** y **password**), moverse por el árbol de directorios de su cuenta y transferir cualquier fichero a su PC. Esta sería una sesión de trabajo normal con un cliente **FTP**.

Existe una gran cantidad de clientes **FTP**, en distintos entornos **Unix**, **Ms-Dos**, **Windows 3.x**, **Windows 95**, y la mayoría de ellos en entorno gráfico que mejora en mucho a los primeros clientes que se desarrollaron, los clientes basados en texto. Un ejemplo de este tipo de clientes es el que puede encontrar en la máquina en la que posee cuenta de usuario.

FTP ANÓNIMO

Los servidores de **FTP anónimo** ofrecen sus ficheros libremente a todos los usuarios. El usuario que desee un fichero situado en este tipo de máquinas sólo tiene que identificarse con el **login**:

anonymous

de ahí el nombre de **FTP anónimo**. Este nombre de usuario le permite acceder a la máquina con cualquier **password** que indique (**guest**, su dirección de **correo**, etc...), aunque algunos servidores le obligan a especificar su dirección de **correo electrónico**.

Los nuevos clientes de **FTP** le facilitan aún más esta tarea, ya que se conectan y ellos mismos se presentan a la máquina remota indicando el usuario '**anonymous**' cuando lo detectan.

Un ejemplo de este tipo de servidores está en **RedIris**. Si queremos obtener un fichero (binario) de una imagen llamado **rediris.gif** en el directorio **/pub/imagenes** a una cuenta de un usuario:

1. Nos conectamos al servidor **ftp.rediris.es**, el servidor nos pedirá nuestro nombre de usuario.
2. Cuando solicite el **username** se escribe: User: **anonymous**, a continuación nos pedirá nuestro **password**.
3. En **password**: se teclea la dirección de **correo-e** del propio usuario, por ejemplo, Password: **mperez@ujaen.es** (aunque sólo basta con **mperez@**).

En algunas ocasiones, nosotros podemos colaborar en este tipo de servidores incorporando ficheros a él. Para ello se suele utilizar un directorio especial denominado, "**incomming**".

FICHEROS COMPRIMIDOS

Para mejorar el rendimiento de la red los ficheros que podemos descargar de la red, éstos se suelen comprimir. Un fichero comprimido es un fichero que ha reducido su tamaño mediante un medio de compresión que le aplica un programa compresor. Por lo tanto, para poder utilizar ese programa necesitaremos el descompresor correspondiente. Los formatos más usuales en los que nos podemos encontrar estos programas son:

Compresor	Descompresor	Extensión	Sistema Operativo
compress	uncompress	.Z	Unix
gzip	gunzip	.gZ	Unix
pkzip	pkunzip	.zip	DOS
arj	arj	.arj	DOS

Algunos programas son capaces de comprimir y generar ficheros con extensión **.EXE**, eso significa que no necesitaremos un descompresor para obtener sus ficheros, ya que este se encuentra incorporado en el fichero. Estos ficheros son auto-extractables, aunque como desventaja, un fichero **.EXE** no puede ser usado en una máquina **Unix**.

UNA SESIÓN FTP CON SU CUENTA DE USUARIO

Para iniciar una sesión **FTP** con su cuenta de usuario necesita:

1. Un cliente **FTP**

- Cliente **FTP** basado en texto de las máquinas **UNIX**
- Un cliente **FTP** gráfico, encontrará uno en el grupo **PATHWORKS**

desde un navegador de la **WWW**

2. Conocer el nombre de la máquina o la dirección **IP** a la que se desea conectar
3. Su nombre de usuario.
4. Su clave de usuario.

```
Connected to jabalcuz.ujaen.es.  
220 jabalcuz.ujaen.es FTP server (Digital UNIX Version 5.60) ready.  
User (jabalcuz.ujaen.es:(none)): 
```

Inicio de sesión con FTP:

Los clientes **FTP**, basados en texto, tienen una serie de comandos que nos permiten desplazarnos por los directorios de la máquina local y remota.

```
ftp> help  
Commands may be abbreviated. Commands are:  
  
!           delete      literal  
?           debug       ls  
append      dir         mdelete  
ascii       disconnect  mdir  
bell        get         mget  
binary      glob        mkdir  
bye         hash        mls  
cd          help        mput  
close       lcd         open  
ftp> 
```

Comandos de ayuda de FTP:

Estos comandos son implementados mediante botones en los clientes **FTP** gráficos facilitando mucho su utilización.

Muchos ordenadores reservan una parte de su disco para ofrecer y distribuir sus ficheros públicamente a todos los usuarios que los deseen. Este nuevo servicio que surge a partir del **FTP** es lo que se denomina **FTP anónimo**.

MOTORES DE BÚSQUEDA:

Internet es una red descentralizada que le confiere una imagen un tanto anárquica. Por ello, al principiante le asaltan dudas en el momento de buscar información.

Un **motor de búsqueda** es una herramienta **Web** que localiza, de forma rápida, información existente en **Internet**.

Está formado por tres elementos bien diferenciados: **un interfaz, un robot y una base de datos**.

El éxito de las búsquedas en un motor específico estará determinado por el volumen y el contenido de la base de datos del motor, la velocidad de búsqueda, la facilidad del uso, entre otros factores.

Su funcionamiento a nivel de usuario es muy sencillo, al introducir una o varias palabras claves en la herramienta, da como resultado un listado con gran semejanza a lo introducido. El uso de operadores lógicos simplifica notoriamente la búsqueda de la información, prácticamente todas las herramientas de búsqueda admiten la utilización de lógica matemática (**OR, AND, NOT**,etc.).

Existen otros dos términos que se usan como sinónimos de motores de búsqueda: **índices y directorios**.

Podríamos diferenciar cuatro niveles de búsqueda en **Internet**:

Página actual

Lugar actual

En el Web

En Internet

Podemos buscar palabras dentro de la **página** que estamos visitando actualmente mediante el uso del comando apropiado en nuestro navegador, normalmente situado en los iconos de la barra de herramientas o bien dentro del menú **'Edición'**.

Muchos lugares, disponen de servicios de búsqueda simples, donde poder encontrar la información de manera rápida en el interior de aquel servidor. Normalmente al acceder a la página de búsqueda, aparece un pequeño cuadro de texto donde introducir las palabras a buscar junto a un botón para ejecutar la búsqueda.

En el **Web** existen muchos lugares para hacer búsquedas, son sitios especializados y pueden realizarse simples búsquedas o bien búsquedas complejas utilizando argumentos boleados.

Básicamente hay dos tipos de buscadores: los buscadores tradicionales y los buscadores **robot** o de **araña**.

Los **buscadores tradicionales** son básicamente, bases de datos de direcciones que alguien ha dado de alta. Es decir, lugares donde están almacenadas las direcciones de las páginas indexadas junto a una pequeña descripción del contenido del lugar.

Son muy útiles para buscar recursos grandes, lugares importantes, pero son muy poco eficaces para encontrar un documento determinado, puesto que la búsqueda se realiza únicamente dentro de la descripción que acompaña a la dirección. Tienen la ventaja de ofrecer las direcciones clasificadas jerárquicamente por temas y de realizar búsquedas con resultados más ajustados.

Los **buscadores araña** son buscadores más especializados. Las «**arañas**» son programas específicos que navegan por la red, acceden a un servidor y recogen todas las páginas **web** siguiendo todos los enlaces. Cada una de las páginas del servidor se indexa y se incluye en la base de datos siguiendo una metodología.

Por lo tanto, este tipo de buscadores contienen mucha más información donde buscar, por lo que sí queremos buscar en documentos algo escondidos, será el mejor recurso donde podamos acudir.

Tienen la desventaja de producir muchos resultados y por lo tanto es más difícil depurar la búsqueda.

La búsqueda de información dentro de la red nunca fue una cuestión sencilla de realizar. Sin embargo, la

existencia de herramientas de búsqueda tales como los **robots, agents, webcrawlers, webwanderers, rastreadores, spiders, search engines, índices, directorios**, entre otros, son términos que en ocasiones son utilizados como sinónimos.

En concepto los **motores de búsquedas** son **sites de Internet** que funcionan gratuitamente permitiendo a los navegantes encontrar páginas con información útil.

Su funcionamiento a nivel de usuario es muy sencillo, al introducir una o varias *palabras claves* dentro de esta herramienta (**search engines**), basta con solo presionar un **enter** o hacer **click** en **buscar** (según sea el buscador) y es entonces cuando se formula la búsqueda, finalmente arroja como resultado un listado ordenado según la semejanza que posea con la palabra introducida.

Los **motores de búsqueda** poseen en su gran mayoría un índice temático que tiene como función primordial brindar una orientación de forma amigable dentro de la red mundial.

Los buscadores según su estructura se dividen principalmente en dos diferenciados tipos:

Ø **Clasificadores ó buscadores de categorías** (comúnmente llamados Índice). Por ejemplo Yahoo.

Ø **Buscadores de contenido** (ó de palabras por contexto). Un ejemplo sería Altavista de Digital.

Los principales buscadores suelen incluir publicidad y, cada vez más, una gran cantidad de recursos de valor añadido como pueden ser: la búsqueda de direcciones de **correo**, búsqueda de artículos de **news, chats**, acceso a sitios de venta a través de **Internet**, acceso a información meteorológica actualizada, enlaces a prensa electrónica, enlaces a otros buscadores, servicios de traducción automática de páginas, servicios de localización de imágenes o ficheros audio, y acceso a las últimas noticias de actualidad entre otros servicios que brinda.

La localización y recuperación de información en la red puede llegar a representar un grave problema, debido a la inmensa magnitud de información, se dificulta el acceso a los datos que realmente importan, debido a la desorganización que posee la información dentro de la red y los resultados que arrojan los buscadores pueden ser datos de buena calidad (fiable y actualizado) ó datos de escasa a nula calidad.

Los **motores de búsqueda** basan la recuperación en el uso de palabras clave y en la ordenación de los resultados de búsqueda por relevancia.

Existen aquellos que solo buscan en los sitios **web** de quienes se han suscripto a esas bases de datos.

La mayor parte de motores usan por defecto el operador **O (OR)** para enlazar los términos de búsqueda, en caso de que no se haya especificado ningún operador, esto genera mucho ruido.

Para evitar este problema conviene precisar la frase de búsqueda a través de las posibilidades que nos ofrece el lenguaje de consulta: **comillas, operadores boléanos, paréntesis, ponderación de términos de búsqueda, operadores de adyacencia y proximidad, mayúsculas y acentos**. Estas posibilidades suelen variar de un buscador a otro, por lo que se hace imprescindible consultar las pantallas de ayuda del buscador.

Los Buscadores se podrían clasificar en dos diferenciados grupos, **directorios ó índices y motores de recuperación**.

Los **Directorios o Índices** en su formato poseen un listado de recursos organizados por categorías temáticas, las cuales están organizadas jerárquicamente en forma de árbol, de esta manera se pueden acceder a la información pasando de lo general a lo particular. La mayoría de los índices permiten el acceso a estos

recursos a través de dos sistemas: navegación a través de la estructura de las categorías temáticas, y búsquedas por palabras claves sobre el conjunto de referencia del buscador.

Los **índices** son más precisos pero menos exhaustivos que los **motores de búsqueda**. El resultante es menor.

Sus ventajas, mayor precisión, menor ruido, y el poder contar en algunos casos con comentarios y valoraciones de los recursos referenciados realizados por especialistas en el tema.(Citado de Manual teórico y práctico Internet F.U.C.E.).

Motores de búsqueda son ni más ni menos, base de datos generadas como resultado de la indexación a texto completo ó parte de los documentos recogidos previamente de **Internet**. Estos suelen recoger documentos **web** en formato **HTML, NEWS**, entre otros. Esta tarea de recoger información es realizada por un programa denominado **Robot**.

Los **motores** devuelven gran volumen de páginas referenciadas, problema que es salvado en gran parte gracias a la ordenación por relevancia de los resultados de la búsqueda. Existen tantos **motores de búsquedas** en la red y tan variados que si aplicáramos una misma consulta en distintos buscadores veríamos que en su gran mayoría los casos resultantes serían dispares.

Los puntos a considerar a la hora de valorar la calidad de un buscador son: n° de documentos **Internet** que almacena en su base de datos, la flexibilidad y calidad del lenguaje de consulta, la pertinencia de los resultados (ruido y silencio), los servicios de valor añadido que incorporan, la periodicidad de actualización de la base de datos, la velocidad en la recuperación y las dificultades de conexión.

Los **motores** son más exhaustivos que los índices en cuanto al volumen de páginas referenciadas (varias decenas de millones frente a unos pocos cientos de miles), pero son mucho menos precisos que los índices, al no ser su contenido objeto de indización humana.

Los elementos que componen a un Motor de Búsqueda son:

1) Interfaz.

Es la **pagina web** a la que accede el usuario. En ella se establece los tipos de búsquedas:

Con formularios: se presenta una pagina con formularios en los que se introducen las palabras claves de búsqueda.

Con directorios: estructura la información jerárquicamente por materias. Para encontrar la información hay que recorrer el árbol desde los temas más generales hasta los más específicos.

2) Base de datos textual.

Índice de palabras frases y datos asociados con la dirección de **páginas web(URL)**, ficheros, programas, etc. La información se da de alta y de baja, pero siempre dejando una referencia para que después quien busque un tema relacionado pueda encontrar la dirección y un pequeño resumen de lo que contiene.

3) Robot.

Es un programa que atravesía una estructura de hipertexto recuperando ese enlace y todos los enlaces que estén referenciados allí.

Las funciones que lleva a cabo son las siguientes:

Análisis estadístico : Miden el crecimiento de la Web, número de servidores conectados, etc.

Mantenimiento de la estructura hipertextual de la World Wide Web : Verificando la corrección de los enlaces entre documentos y eliminando o guardando información de los denominados "**enlaces muertos**"(**dead links**), es decir, **páginas Web** que ya han desaparecido.

Duplicación de directorios Ftp (Mirrors): Incrementando su utilidad a un número mayor de usuarios.

Creación automática de bases de datos textuales a partir de los documentos **html** distribuidos por los distintos servidores. Debido a que cada **robot** está programado para buscar en la Red de diferente forma, la información almacenada en cada base de datos puede ser diferente.

Los **robots** pueden ser usados para varios propósitos:

Indexar, Validar HTML, Validar Links, Monitorear que hay de nuevo, Generar imágenes.

Estos son usualmente llamados **Web Wanderers**, **Web Crawlers**, ó **Spiders** suelen compara sus movimiento entre los sitios como si fuesen **virus**, un **robot** simplemente visita sitios y extrae los enlaces que están incluidos dentro de estos.

Arañas (spiders): Es un **robot**, pero otorga resultados más fríos.

Gusanos (worms): Es lo mismo que un **robot**, aunque técnicamente un **gusano** es una replica de un programa, a diferencia de un **robot** que es un programa original.

Orugas (web crawlers): Es lo mismo que un **robot**, aunque es un tipo más específico.

Hormigas (web ants): Cooperativa de **robots**.

Vagabundos (wanderers): **Robot** encargados de la medición del crecimiento de la **Web**.

Knowbots: permite evaluar el impacto de las distintas aportaciones que engrosan las distintas áreas de conocimiento presente en la red.

Agente autónomo: es un programa que viaja entre los sitios **Web**, decidiendo por ellos mismos cuando son movidos a otros lugares y también deciden que hacer.

Agente inteligente: son programas que ayudan al usuario por ejemplo: a elegir productos, llenar formularios o a encontrar determinada cosa.

Agente de usuario: es un nombre técnico para programas que ejecutan tareas para un usuario en la red.

Ejemplo: **Netscape Navigator, Microsoft Internet Explorer**, etc.

Buscador sin Robot

Ø Las direcciones añadidas, se ubican en secciones dentro de una estructura de árbol. Los contenidos en muchos casos, son analizados y procesados por personas que visitarán la dirección añadida, determinando si éste cumple con los requisitos necesarios para ser dados de alta y si los datos introducidos son correctos.

Ø Hay menos resultados totales debido al menor contenido de sus bases de datos. Estos son mejores, más fiables y presentan menos enlaces erróneos o poco efectivos.

Ø Los resultados aparecen por orden alfabético en la mayoría de los casos y ordenados por categorías temáticas.

Ø Son ideales para hallar páginas sencillas con temas comunes, pues las búsquedas se facilitan.

Buscador con Robot

Ø Los contenidos son indexados por medio de un **robot**, **araña** o **gusano**.

Ø No imprescindible dar de alta a un sitio **Web** para figurar en él. Es aconsejable pero no imprescindible, pues la mayoría de los **robots** buscan por la **Web** por ellos mismos, indexando todo lo que hallan a su paso. Aun así, el contenido de sus bases de datos no suele superar el 5% del total de la red.

Ø Hay más resultados totales, estos son menos fiables y presentan más enlaces erróneos o poco efectivos.

Ø Los resultados aparecen por orden de prioridad, dependiendo de las características del **robot**. Pueden tomar las palabras clave del título, descripción o contenido.

Ø Las altas pueden llegar a demorar varios meses en algunos casos.

Ø Son ideales para hallar temas intrincados o prohibidos en los directorios de búsqueda, o para encontrar temas poco comunes.

Funcionamiento de los Motores de Búsqueda:

Los **Motores de Búsqueda** realizan un trabajo doble, bien diferenciado:

El primero, hace referencia al **robot** y a su forma de actualizar la base de datos. Periódicamente, y tomando como partida una **URL** inicial, el robot recupera un fichero en formato **html** que transfiere al sistema local, en donde procede a su indización, guardando en su base de datos tanto los términos

indexados, como la localización **URL** completa del documento. Una vez indexado el documento, el **robot** identifica los enlaces hipertextuales que contiene el mismo y de forma recursiva procede a recuperar los documentos referenciados en esos enlaces, creando un índice, obteniendo nuevos enlaces, etc.

En el segundo, el usuario acude a ellos como si de una hoja **Web** más se tratara. Con él puede seleccionar el tema de interés de un directorio, llenar los campos de búsqueda con las palabras claves relacionadas con dicho tema, seleccionar el tipo de búsqueda (lógica booleana, truncamientos, etc.) y lanzar la pregunta.

El **motor** revisa su base de datos en busca de términos que coincidan con los de la pregunta y en breves instantes devuelve una lista con las direcciones de las hojas **Web** que más se ajusten a esa pregunta. Dependiendo de la complejidad del **motor** y del algoritmo de búsqueda empleado, tendremos una lista con las direcciones que mejor se adaptan al perfil de búsqueda. Después de comprobar las direcciones interesantes, sólo hay que pulsar sobre ellas para conectarse y ver si hemos acertado.

Los programas comúnmente denominados "**Robots**" o "**Web-Crawlers**" pueden estar escritos en varios lenguajes (**Perl**, **C** .etc.) pero su funcionamiento básico depende del **protocolo HTTP (Hyper Text Transfer Protocol)**. El **HTTP**: Cada vez que solicitamos una página en **Internet**, nuestro navegador (además de convertir el "nombre del sitio" a un nodo **IP**, envía información que es denominada "**HEADERS**", esta información es interpretada por el servidor de páginas.

Los "**HEADERS**" le informan al servidor de páginas una gran cantidad de cosas :

Ø Que navegador esta solicitando la información? (Esto tiene implicaciones para enviar formatos de **HTML**, e inclusive es la forma en que se distingue cuando la solicitud proviene de un aparato inalámbrico).

Ø El navegador aceptará "cookies"?

Ø El navegador puede desplegar fotografías.

Los "**HEADERS**" que comúnmente envía un Navegador son los siguientes:

GET /index.html HTTP/1.0 User-Agent:

Mozilla/4.51 Accept: */*

Lo anterior indica que el cliente (**Navegador** o "Web-crawler") esta solicitando la página principal (**index.html**) del respectivo sitio, con la versión **1.0 HTTP** , **User-Agent** indica que el cliente es : **Mozilla version 4.51** (**Mozilla** es **Netscape Navigator**), y **Accept: */*** indica que esta dispuesto a aceptar todo tipo de información (**GIF's, JPEG's, Java, etc.**).

Una vez analizados estos "**HEADERS**" por el servidor de paginas, éste envía sus respectivos "**HEADERS**" con la información pertinente. Estos **HEADERS** también son de suma importancia ya que le indican al **Navegador** o "Web-Crawler", como esta siendo enviada la información. Los "**HEADERS**" que típicamente envía un Servidor de Paginas son los siguientes:

HTTP/1.0 200 OK Last-Modified: Fri, 08 Jul 2001 08:57:12 GMT

Content-Type: text/html; charset=iso-8859-1 **MIME-Version: 1.0** **Date: Sat, 09 Jul 2001 12:16:41 GMT** **Server: AOLserver** **Content-Length: 4807**

Lo anterior indica al cliente (**Navegador** o "Web-crawler"):

La requisición fue favorable: **200 OK** (esto pudo ver sido otro código como : **404 Not Found** , o **403 Forbidden** ,etc).

La última fecha de modificación del documento: **Last-Modified: Fri, 08 Jul 2001 08:57:12 GMT** .

El tipo y características del documento: **Content-Type: text/html; charset=iso-8859-1** **MIME-Version:1.0.**

La fecha de la solicitud: **Date: Sat, 09 Jul 2001 12:16:41 GMT**.

El servidor de paginas que esta enviando la información Server: **AOLserver** .

El tamaño del documento en **bytes** : **Content-Length: 4807**.

Gráficamente:

Ejemplo citado de pagina de Internet (

Algunos Motores de Búsqueda y sus diferencias: (yAHOO, ALTAVISTA, LYCOS, INFOSEEK, WEBCRAWLER).

Yahoo: www.yahoo.com

Aparecen sitios clasificados por temas, referencias y enlaces a lugares ya existentes.

Como buscar en Yahoo: puede seleccionar una categoría de las que aparecen en el listado inicial de **Yahoo**.

Allí se encuentran por ejemplo: artes, humanidades, negocios y economía, computadoras e **Internet**, entre otros. Dentro de cada categoría aparece un listado de subtemas que permite hacer una búsqueda más específica.

Si sabe lo que quiere buscar, pero no sabe dónde hallarlo, puede utilizar la ventana de búsqueda (**search**) para especificar una o varias palabras clave, **yahoo** no establece diferencias entre mayúsculas y minúsculas, el idioma predeterminado es el inglés. **Yahoo** realiza una búsqueda a través de las cuatro áreas de su base de datos: categorías, **Web sites**, eventos y charlas y artículos de noticias mas recientes.

Si no encuentra coincidentes, el sistema realiza una búsqueda general en la red a través de **Alta Vista**.

Para resultados de búsquedas menos extensos se utiliza un menú de opciones (**options**), allí se puede especificar la base de datos sobre la que se realiza la búsqueda, la antigüedad de los documentos que se debe hallar y el número deseado de resultados por página, a través de un formato que cada usuario puede modificar. También brinda la posibilidad de seleccionar el método de búsqueda, según las siguientes posibilidades:

Inteligente por defecto, que entrega los documentos y categorías que incluyen todas o algunas de las palabras clave.

Coincidencia exacta con una frase: donde las palabras clave deben encontrarse en el documento exactamente como se escribieron en la ventana de búsqueda (en la misma secuencia).

Coincidencia de todas las palabras: como si se utilizará el conector "y".

Coincidencia con alguna palabra: como si se utilizara el conector "o".

Nombre de una persona: cuando no se trata de lugares sino de gente.

Sintaxis en la ventana de búsqueda: La búsqueda también se puede restringir utilizando algunos filtros (**operadores**) al escribir las palabras claves dentro de la ventana:

Comillas dobles (" "): tiene el mismo efecto que utilizar la opción de búsqueda de frases completas. Le indica a **Yahoo** que las palabras clave forman parte de una frase.

Signo más (+): antes de una palabra este indica que debe aparecer obligatoriamente en los resultados de la búsqueda.

Signo menos (-): Especifica las palabras que no deben aparecer en los resultados.

t:: busca las palabras solamente en los títulos de los documentos.

u:: busca las palabras solamente en las direcciones (**URL**) de los lugares de **Internet**.

Asterisco (*): Se utiliza como comodín **Col*** arrojará como resultado un listado de lugares que incluyan palabra que comiencen por "**Col**". Estos operadores se pueden combinar, pero se debe conservar un orden lógico, de la siguiente manera: **+, -, t:, u:, " "** y finalmente *****.

Un lugar personal: **My yahoo** este servicio esta diseñado para que cada persona tenga la posibilidad de crear

su propio lugar y seleccionar la información que quiere que aparezca en él.

Servicios adicionales: Aparte de los servicios de búsqueda tradicionales **Yahoo** ofrece, entre otros:

Yahooligans For Kids: índice con temas especiales para jóvenes y niños que incluye: computadoras, juegos y servicios en línea, alrededor del mundo, sopa de artes, entretenimiento, escuela, entre otras categorías.

YahooInternet life: un recorrido al estilo "Magazin" por las actividades diarias en **Internet**.

Yahoo Shop: encontrarás camisetas, gorras, libros, teclados y demás artículos con el sello de **Yahoo**.

Today's events and chats: recorrido por las actividades del día que se desarrollan en **Internet** (clasificadas por temas),

Weekly picks: página de enlaces a lugares que pueden ser buenos para los usuarios, según **Yahoo**.

National Yahoos y Yahoo Metros: versiones de **Yahoo** preparadas para países y ciudades específicas.

Enlaces. Estos enlaces se implementan haciendo del gráfico un mapa sensitivo, que de acuerdo con la posición donde se haga **click** responde como una dirección **Internet (URL)** diferente.

News: conduce al cibernetico a un listado con los enlaces nuevos recibidos por **Yahoo** en la última semana, organizados por categorías.

Cool: lleva a una selección de sitios divertidos, extraordinarios o de contenido útil para el posible visitante.

Random: escoge al azar un enlace de la base de datos de **Yahoo** para que el visitante recorra. Es para quienes quieren ir "a la deriva" al navegar por el **Web**.

Headlines: provee noticias de última hora, además de algunas en la categoría de negocios, deportes, internacional y política de EE.UU.

Yahoo Info: como su nombre lo dice, brinda información específica de **yahoo**, tal como el "**help**".

Add URL: permite mediante un sencillo formulario inscribir un sitio Web en el directorio **Yahoo**.

Altavista. www.altavista.digital.com

La máquina más poderosa que existe en el momento. No solo busca en los títulos, direcciones o descripciones de un sitio enlazado; mas bien busca en todas las palabras de un texto o hipertexto y de acuerdo con ello genera la respuesta query (pregunta mediante palabras llave).

Altavista no depende de que un sitio **Web** se registre en ella, pues tiene "**robots**" que permanentemente están recorriendo todo el **Web** para indexarlo a su base de datos. Utiliza cualquier lengua ya que no distingue palabras de diferentes idiomas.

A diferencia de un directorio, **Altavista** no tiene una estructura jerárquica para hacer búsquedas temáticas. Su configuración se basa en dos sistemas, uno para la indexación y base de datos con 10 procesadores y otro para recoger todo el **Web** basado en tecnología **DEC 300**, este último es considerado por parte de **DEC** como la **araña** más rápida de la red. **Altavista** tiene dos modos de búsqueda directa: **sencillo** y **avanzado**.

El modo avanzado tiene varias diferencias del sencillo y no lo incluye. En el pantallazo inicial de **Altavista** es

clara la ancha zona para escribir las llaves de búsqueda al lado izquierdo del botón "**Submit**" (que al hacerle **click** efectúa la búsqueda). En el renglón superior se pregunta, con dos **pop menús**, si la búsqueda es en el **Web** o en el **Usenet** (**grupos de interés o NewsGroups**) y si los resultados se mostraran de forma estándar, compacta o detallada.

La forma estándar muestra los **URL** de respuesta conformando tres renglones por hit al incluir el texto de arranque de la página encontrada. La forma compacta solo muestra enlaces.

Parámetros para búsqueda sencilla: considera palabras y frases donde los signos de puntuación se consideran como separadores de palabras.

Utiliza las comillas para buscar específicamente lo ingresado. Para obligar a que una palabra o frase deba estar en los resultados, es necesario anteponerle un signo (+=, y para obligar a que una palabra o frase no este, es obligatorio anteponerle un signo menos (-)).

Existe además la posibilidad de usar la "**notación estrella(*)**", para poder denotar raíces de palabras con diferentes sufijos, prefijos o infijos.

Una característica adicional estriba en la puntuación y mayúsculas, no es necesario escribir una palabra en mayúscula para encontrarla, se puede buscar sin tildes de igual forma las encontrara con o sin tildes.

Altavista tiene un sistema de "ranking" para ordenar los resultados de un query. En la búsqueda sencilla el orden es así:

- Primero los documentos que tengan las palabras llave o frases en las primeras palabras de un documento.
- Luego los que tengan repetida más de una vez una palabra o frase.

Las restricciones dependen mas bien de las cláusulas del lenguaje **HTML**.

Title:"EL TIEMPO" encuentra las páginas de tengan la frase "**EL TIEMPO**" en su título.

url:indice.html encuentra las páginas que tengan las palabras índice o **html** en su **URL**. Es equivalente a **URL "índice html"**

Parámetros para búsqueda avanzada. La búsqueda avanzada se diferencia de la sencilla en lo siguiente:

- No maneja los símbolos (+) y (-).
- No supone el "**o lógico**" de default. No se pueden escribir palabras contiguas.
- Requiere del uso de operadores lógicos y paréntesis.
- Permite cambiar los criterios de ranking y los rangos de fechas de búsqueda.
- Los operadores son: **and**, **or**, **not** y **near**; que además son validos en mayúsculas o con equivalentes (**and=&**, **or=|**, **near=~**).
- La procedencia de operadores es la usual en lógica, nombrando primero al de mayor procedencia: **not near and or**. Para acceder al query avanzado es necesario hacer **click** en la zona de "**advanced**". Tiene la opción de navegar sólo con textos para así acelerar la navegación por el sitio.

Lycos: www.lycos.com

Es una herramienta de navegación por el ciberespacio que proporciona contenido editorial propio y un catálogo de 60 millones de lugares en Internet. Su estructura es bastante similar a la del motor **Yahoo** : una ventana de búsqueda a través de palabras clave, un índice por temas y una sección de recomendados.

Tiene algunas diferencias en la sintaxis empleada para especificar la búsqueda y en la manera de llegar al índice temático, pues este no aparece en la pantalla inicial, como en **Yahoo**, sino que es necesario hacer clic sobre un botón de enlace (**sites by subject**) para llegar a él. Como ventaja, ofrece la posibilidad de realizar búsquedas de sonidos y gráficas únicamente.

Permite navegar por la red utilizando tres mecanismos fundamentales : a través de una búsqueda por temas específicos, mediante una guía de categorías interesantes, o por una lista de lugares recomendados.

Para determinar la relevancia de un documento dentro de una rutina de búsqueda, **Lycos** utiliza agentes llamados **arañas**, que encuentran las páginas y crean sumarios que contienen su título encabezamientos, las 100 palabras que más aparecen, las primeras 20 líneas del documento, el tamaño en **bytes** y el número de letras, **Lycos** elimina palabras como el. **a**, **y**, **o**, que no tienen valor dentro de las búsquedas.

Coincidencia de resultados: Cada usuario escoge lo que quiere ver.

El botón **Custom Search**, lleva a un menú de opciones donde se pueden cambiar el conector entre las palabras clave y la manera como **Lycos** entrega los resultados. En cuanto a los conectores se puede utilizar "**or**" , "**and**". El menú también tiene opciones para seleccionar el número de términos coincidentes que deben aparecer (de dos a siete).

En cuanto a las coincidencias de los resultados con las palabras clave de la búsqueda, se tiene las siguientes opciones : **loose match** (no es la más confiable), **fair match** (se acerca más a lo que el usuario quiere), **good match** (las probabilidades de que todos los documentos los satisfagan son del 50 %), **close match** (este tal vez sea la opción) y **strong match** (coincidencia total de los términos de la búsqueda con los resultados).

El usuario también puede elegir el número de documentos por página que quiere ver una vez se realiza la búsqueda : 10, 20, 30 ó 40. Finalmente, cada documento puede tener una descripción detallada (**detailed results**), una breve (**standard results**) o una resumida (**summary results**).

Sintaxis de las búsquedas:

Signo más (-) : especifica las palabras que no deben aparecer en los resultados.

El punto (.): se utiliza para búsquedas de palabras exactas.

El signo pesos (\$) : se utiliza como comodín al igual que el (*) en el motor **Yahoo**.

Servicio de Lycos:

Super pages : lugar interactivo que contiene búsquedas en **Internet**, guía para compradores, páginas amarillas y avisos clasificados.

Top News : noticias del mundo, deportes, negocios, estado del tiempo, gente.

Stock Find : información de las bolsas de valores, ganancias, resultados financieros.

Sites by subject : índice temático de **Lycos**. Como en **Yahoo**, hay categorías y subcategorías.

Pictures and sounds : imágenes y sonidos, nada de texto.

Point review y Top 5% sites : **Lycos** selecciona lo mejor de **Internet** para usted.

Infoseek: www.infoseek.com.

Info: apócope de información. **Seek**: buscar, anhelar, solicitar.

Su estructura general se parece mucho a la de los sistemas (**Yahoo**, **Lycos** y **Webcrawler**), pero tiene algunas diferencias circunstanciales en la sintaxis de las búsquedas y en la presentación de su página inicial. Contiene una ventana de búsqueda para que el usuario escriba las palabras que desea encontrar y un listado de temas generales.

La diferencia de **Infoseek** radica en que, si el usuario lo desea, puede esconder el índice temático y realizar las búsquedas solamente a partir de palabras clave. Se supone que de esta manera se obtienen resultados más precisos, pues la herramienta se vuelve más funcional.

Infoseek tiene un nombre para estas dos modalidades: **Utrasmart**, para la que incluye el listado de temas, y **Ultraseek**, para la herramienta que emplea solo palabras clave. A pesar de que los ingresos al sistema se realizan siempre por página de **Utrasmart** (la que tiene ventana de búsqueda e índice de temas).

Ejemplos de búsqueda con claves: Estas son algunas de las variantes a diferencia de los otros motores existentes.

Palabras con inicial en mayúsculas : Si utiliza mayúsculas en palabras seguidas, **Infoseek** las entiende como un solo nombre propio o un título.

Uso de comas: Si se quiere buscar varios títulos o nombres propios, sepárelos con comas.

Comillas dobles o guiones (" o -) : Se utilizan para especificar palabras que deben aparecer seguidas en los resultados.

Signo más (+) : Obliga a que esta aparezca dentro de los resultados de la búsqueda.

Signo menos (-) : Obliga a que esta no aparezca dentro de los resultados de la búsqueda.

Barra vertical (|) : Es el equivalente a buscar un subtema dentro los temas generales del índice.

Búsquedas específicas: Si quiere especificar dónde realizar una búsqueda, puede utilizar las siguientes claves:

link: infoseek.cpm

site: sun.com

url: science

title: "The New York Times"

Existen dos mecanismos para **Infoseek** que son **Ultrasmart** y **Ultraseek**

Ultrasmart : El más completo esta diseñado para satisfacer las necesidades de la mayoría de los usuarios: es especialmente útil para aquellas personas que tienen una idea general de lo que quieren encontrar pero que no saben a que sitio específico de la red mundial acudir. Adicionalmente, ofrece 2 características : una búsqueda de temas relacionados, en inglés **related topics** (diferentes áreas en el directorio principal del motor que tengan términos relevantes) y otra de noticias relacionadas **related news** que ofrece información novedosa sobre los temas de interés del usuario.

De igual manera tienen un menú que ofrece diferentes servicios como centros de noticias (tecnología, deportes, entretenimiento), información inteligente (contiene referencias de bolsa y planos de calles) y páginas amarillas (para buscar personas, negocios y direcciones electrónicas).

Ultraseek : Esta diseñado para usuarios que saben lo que quieren y que no desean mucho material de desecho en sus resultados.

El sistema busca ser una herramienta rápida, segura, actualizada y muy inteligente. Ofrece un menú que muestra una versión resumida de las operaciones de noticias, información inteligente y páginas amarillas, así como cuatro posibilidades específicas de búsqueda. Estas determinan la coincidencia de las palabras clave con una parte específica de los resultados : **Imageseek** (se refiere a las imágenes), **Site search** (lugares que contengan dicha información), **Link search** (enlaces a páginas con términos coincidentes), **URL search** (coincidencia de palabras en la dirección del lugar) y **Title search** (que los términos aparezcan en el título de la página).

Los mismos resultados se obtienen utilizando las claves que antes mencionamos en búsquedas específicas.

Webcrawler: www.webcrawler.com.

Webcrawler comenzó como un proyecto de investigación, más adelante se trasladó a **America Online**, donde continua su operación como una herramienta de búsqueda en **Internet** de libre acceso, cuyo objetivo es proveer un servicio de alta calidad, rápido y gratuito. La página recibe un promedio de dos millones de usuarios por semana, lo que en términos de búsquedas representa mas de tres millones al día.

Webcrawler ofrece las características de búsqueda básicas de **Lycos** y **Yahoo** (una ventana para especificar palabras clave y un índice temático), además de una sección especial con servicios adicionales.

El sistema tiene tres funciones diferentes :

- Construye índices para los documentos que encuentra en la red mundial y los pone a disposición de los usuarios.
- Actúa como un agente que busca documentos de interés para los navegantes del ciberespacio, a partir del conocimiento acumulado en sus índices.
- Es un espacio de prueba para experimentar mecanismos de búsqueda en Internet.

Así mismo **Webcrawler** está compuesto por tres partes esenciales : el motor de búsqueda propiamente dicho, una base de datos y agentes que recuperan los documentos. **Webcrawler** le permite determinar la manera como los resultados de su búsqueda se presentan en la pantalla, a través de dos cajas de dialogo.

En la primera de ellas, **Show**, se pueden seleccionar entre títulos y sumarios. Los primeros entregan solamente el nombre de las páginas encontradas, mientras que los segundos presentan, además un resumen de su

contenido, las direcciones, un puntaje que califique su nivel de satisfacción con respecto a los términos de la búsqueda y la posibilidad de ver páginas similares para cada resultado.

La otra caja, **For**, determina el número de resultados por página, que puede ser de 10, 25 o 100. Si lo prefiere, puede utilizar el enlace **options**, ubicado debajo de la ventana de búsqueda, que ofrece las siguientes alternativas adicionales:

Mostrar la calidad de las páginas encontradas de manera numérica (con porcentajes) o a través de un ícono.

Ver los resultados en la misma ventana activa del programa de navegación, en una nueva para cada resultado o en una ventana diferente para todos los resultados.

Servicios de webcrawler: Al hacer clic sobre el botón **Special**, se tiene acceso a los siguientes servicios :

Direcciones en un mapa : Si necesita ayuda para encontrar al sitio al que va o solamente quiere saber donde se encuentra, utilice el servicio **Find and Address**. Presenta un mapa interactivo donde se puede ubicar cualquier dirección dentro de EEUU.

Ruleta Wed : Una colección de sitios que el sistema selecciona de forma aleatoria para usted.

Sintaxis de las búsquedas: Al utilizar los operadores que aparecen en la siguiente tabla se puede obtener resultados mas precisos. Los operadores son **AND**, **OR**, **NOT**, **NEAR/x** (determina qué tan cerca debe aparecer una palabra de la otra, sin importar el orden. **ADJ** (significa que las palabras deben aparecer en el texto, una cerca de la otra y en el mismo orden especificado). "..." (Especifica frases de búsquedas exactas). (...) Sirve para simplificar la creación de búsquedas complejas. Se puede usar en combinación con otros operadores.