



UD 1 - Introducción a las aplicaciones web

INTRODUCCIÓN AL FUNCIONAMIENTO DE LA WEB

1. Funcionamiento de las aplicaciones web	2
1.1.Introducción	2
1.2. Funcionamiento de la Web	6
2. Aplicaciones web.....	9
2.1. Webmail.....	9
Funcionamiento del correo.....	9
Protocolos de correo.....	11
Protocolo de transporte SMTP.	11
Protocolos de acceso a correo (POP e IMAP).....	12
2.2. Navegadores	13
La función principal de un navegador	13
Componentes principales del navegador.....	14
Estándares web	15
Protocolos.....	15
HTTPS (Hyper Text Transfer Protocol Secure)	15
SSL (Security Socket Layer)	17
FTPS (File Transfer Protocol Secure) /SFTP (Secure Shell File Transfer Protocol)	18
Contraseñas	19
Cookies.....	19
Añadiendo funcionalidad a los navegadores	20
2.3. Buscadores	20
Buscador Google	21
Sintaxis especiales.....	22
2.4. Herramientas cooperativas.....	23
Grupos de noticias	23
Protocolo NNTP (Network News Transport Protocol)	23
Clientes de noticias.....	24
Grupos Unset públicos contra los privados.....	24
Ofimática WEB	24
CMS-Gestores de contenido.....	25
3. Almacenamiento web	25
4. Gestión de servicios web multimedia.....	25
Audio digital en la web	25
Imagen digital en la web	26
Video digital en la web	26

1. Funcionamiento de las aplicaciones web

1.1.Introducción

Hoy en día Internet —también conocida como «la red»— supone un excelente medio para obtener información de los más variados temas a cualquier hora del día y sin necesidad de moverse de casa.

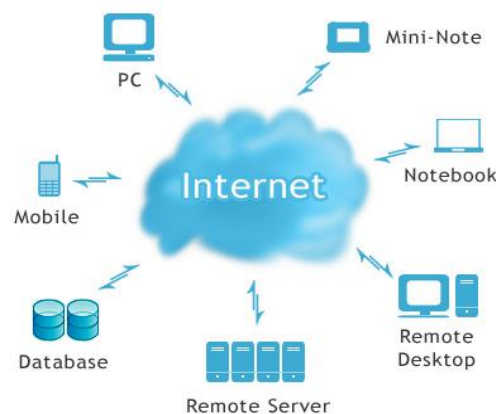
Así, se pueden conocer las noticias de última hora, ver el tráiler del próximo estreno de cine, visitar lugares lejanos, reservar los billetes de avión para las vacaciones, contactar con personas de todo el mundo o comprar cualquier cosa que a uno se le pueda ocurrir.

Esto se debe a que la información disponible en Internet es casi ilimitada, y aumenta día a día. Las empresas vieron en la red primero un medio para anunciarse y, actualmente, una vía para atender a sus clientes y ofrecer sus productos.

Muchas instituciones públicas y privadas la utilizan para dar a conocer sus actividades y publicar datos de interés general o de un tema específico. Y los particulares disponen de un número creciente de servicios accesibles a través de Internet.

Para que toda esta montaña de información sea realmente útil es imprescindible que el acceso a ella sea simple e intuitivo, de forma que cualquier persona pueda encontrar y utilizar la información que desea con tan sólo unos conocimientos básicos.

Esto es hoy posible gracias a la Web. En este capítulo se verán los fundamentos y el modo de trabajo de la Web, mientras que en los siguientes capítulos se aprenderá a diseñar documentos propios y a incorporarlos a esta gran telaraña mundial.



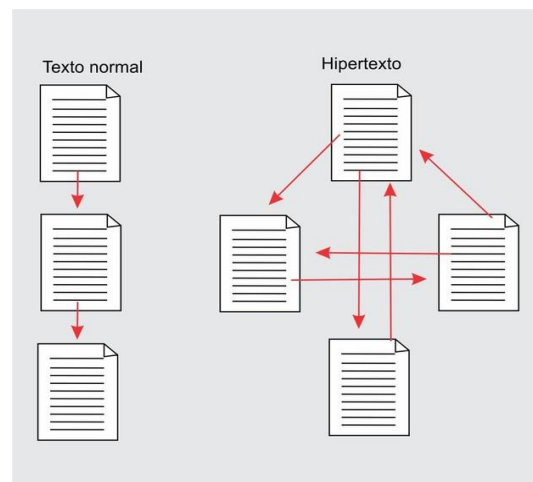
Físicamente, **Internet (Interconnected Networks)** es una red compuesta por miles de redes independientes pertenecientes a instituciones públicas, centros de investigación, empresas privadas y otras organizaciones.

Estas **redes comparten unas normas** que aseguran la comunicación entre ellas:son los denominados protocolos Internet. Un protocolo es un conjunto de normas que permite el intercambio de información entre máquinas de diversos tipos conectadas entre sí.

Existen **diversos protocolos** en función del tipo de comunicación que se establece entre las máquinas: FTP (para la transferencia de ficheros), POP y SMTP (para el envío y recepción de correo electrónico), TELNET (para la conexión con terminales remotos), y HTTP (para el acceso a servidores de información). El protocolo HTTP (HyperText Transfer Protocol) es la base de la World Wide Web o telaraña mundial, abreviada WWW y más conocida como Web.

La Web es un sistema que facilita el **acceso a los distintos recursos** que ofrece Internet a través de una interfaz común basada en el hipertexto. La Web se ha hecho enormemente popular con la aparición de unos programas dotados de interfaces gráficas amigables que permiten a cualquier persona acceder a los documentos de la red: son los conocidos **navegadores**.

El **hipertexto** es un texto con enlaces. Los documentos de hipertexto pueden contener enlaces, que son referencias a otras partes del documento o a otros documentos. De esta forma, el documento no tiene porque ser leído secuencialmente, sino que en ciertos puntos se puede saltar a otra parte del documento y desde allí volver al punto original, o bien saltar a otro punto distinto.





El término **hipertexto** fue acuñado por Ted Nelson en 1965, aunque sus orígenes se remontan a 1945. El hipertexto supone una ampliación del concepto habitual de texto al permitir que una serie de documentos enlazados unos con otros y posiblemente ubicados en lugares remotos aparezcan formando una unidad.

Las referencias entre las partes de un documento de hipertexto se establecen mediante anclas y enlaces: un anda es un fragmento de información dentro de un documento al que se asocia un enlace. Este fragmento puede ser una palabra, frase o párrafo completo. Un enlace es una referencia, o puntero, a otro fragmento de información.

El enlace debe contener toda la información necesaria para acceder al fragmento enlazado: su nombre, ubicación y mecanismo de acceso (protocolo).

La idea de los enlaces está ya presente en la mayoría de los documentos que se utilizan a diario: índices, tablas de contenidos, pies de página, referencias bibliográficas. La diferencia con los documentos de hipertexto basados en computadora es que los saltos a estos enlaces se realizan automáticamente, y funcionan de la misma manera tanto si se refieren al propio documento como a documentos externos. De hecho, el lector no apreciará la diferencia.

El hipertexto da un gran salto con el desarrollo de internet, ya que un documento puede estar físicamente distribuido en distintas máquinas conectadas entre sí. Esta es la idea que da origen a la Web.



La **Web** fue creada en 1989 por **Tim Berners-Lee** (foto) en el CERN (Ginebra). Su objetivo inicial era compartir información entre grupos de investigadores de gran tamaño localizados en diversos lugares del mundo, utilizando para ello el concepto de hipertexto.

Como resultado se crearon unos protocolos y especificaciones que han sido adoptados universalmente e incorporados en Internet,

gracias a aportaciones posteriores como el desarrollo por la NCSA (National Center for Supercomputing Applications) de una popular interfaz, el MOSAIC. En la página web del World Wide Web Consortium o W3C (www.w3.org) puede conocerse la visión primitiva del creador sobre la Web y algunas charlas sobre la misma impartidas por directivos de la institución.

Por Web se pueden entender tres cosas distintas:

- El proyecto inicial del CERN,
- El conjunto de protocolos desarrollados en el proyecto o
- El espacio de información formado por todos los servidores interconectados (el denominado «hiperespacio»); cuando se habla de la Web habitualmente se hace referencia a esto último.

Las ideas básicas de la Web son tres:

1. La **desaparición de la idea de un servidor único** de información mantenido por un equipo de profesionales dedicado. Cualquiera puede crear sus propios documentos y referenciar en ellos a otros documentos. Así pues, no existe una autoridad central en la Web.
2. La especificación de un mecanismo para localizar de forma unívoca documentos distribuidos geográficamente. Cada documento tiene una dirección llamada **Uniform Resource Locator, URL**. Esta dirección indica no sólo la localización del documento, sino también el mecanismo para acceder a él (es decir, el protocolo, que en el caso de una página web es el HTTP).
3. La existencia de una **interfaz de usuario uniforme** que esconde los detalles de los formatos y protocolos utilizados para la transferencia de la información, simplificando por tanto el acceso a la misma.

Los programas que utilizan estas interfaces son denominados examinadores (browsers), visualizadores o, más comunmente, navegadores.

El elemento básico de la Web es la **página web**, un documento que contiene enlaces de hipertexto a otros documentos relacionados que pueden estar localizados en la propia máquina o en máquinas remotas. De esta forma, siguiendo los enlaces se puede navegar por la red visitando páginas ubicadas en distintos lugares como si formaran una unidad.

El protocolo de la Web especifica un formato para escribir los documentos que forman parte de ella: es el Hypertext Markup Language (HTML). Sin embargo, también es posible acceder a documentos de otros formatos a través de la interfaz web sin necesidad de reescribirlos, gracias a las funcionalidades incorporadas por los navegadores.

1.2. Funcionamiento de la Web

La Web funciona siguiendo el denominado **modelo cliente-servidor**, habitual en las aplicaciones que funcionan en una red. Existe un servidor, que es quien presta el servicio, y un cliente, que es quien lo recibe.

El **cliente web** es un programa con el que el usuario interacciona para solicitar a un servidor web el envío de páginas de información. Estas páginas se transfieren mediante el protocolo HTTP.

Las **páginas** que se reciben son documentos de texto codificados en lenguaje **HTML**. El cliente web debe interpretar estos documentos para mostrárselos al usuario en el formato adecuado. Además, cuando lo que se recibe no es un documento de texto, sino un objeto multimedia (vídeo, sonido, etc.) no reconocido por el cliente web, éste debe activar una aplicación externa capaz de gestionarlo.

Entre los clientes web (también conocidos como visualizadores o navegadores) más usuales están el Netscape Navigator y el Microsoft Internet Explorer. La mayoría de ellos soportan también **otros protocolos**, como el FTP (File Transfer Protocol), para la transferencia de ficheros, y el SMTP (Single Mail Transfer Protocol), para el envío y la recepción de correo electrónico.

El **servidor web** es un programa que está permanentemente escuchando las peticiones de conexión de los clientes mediante el protocolo HTTP.

El servidor funciona de la siguiente manera: si encuentra en su sistema de ficheros el documento HTML solicitado por el cliente, lo envía y cierra la conexión; en caso contrario, envía un código de error que cierra la conexión.

El servidor web también se ocupa de controlar los aspectos de seguridad, comprobando si el usuario tiene acceso a los documentos.

El **proceso completo**, desde que el usuario solicita una página hasta que el cliente web se la muestra con el formato adecuado, es el siguiente:

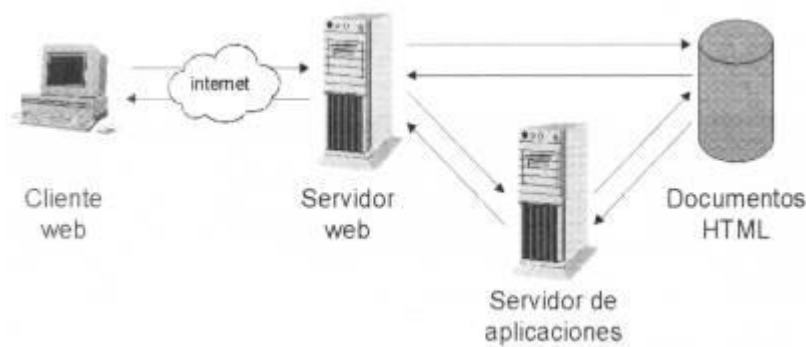
1. El usuario especifica en el cliente web la dirección (URL) de la página que desea consultar.
2. Una vez resuelta la dirección IP a partir del URL (DNS) el cliente establece la conexión con el servidor web.
3. El cliente solicita la página deseada dentro de todo el repositorio del servidor.

4. El servidor busca la página que ha sido solicitada en su sistema de ficheros, Si la encuentra, la envía al cliente; en caso contrario, devuelve un código de error.
5. El cliente interpreta los códigos HTML y muestra la página al usuario.
6. Se cierra la conexión.

Es muy probable que en el paso 4 la página solicitada no exista como HTML, sino que se cree en el momento de su petición. En estos casos el servidor web cede el control al denominado servidor de aplicaciones que es quien se encarga de construir la página HTML a partir del código de la misma (PHP, JAVA, etc), y quizás con consultas a una base de datos (MySQL, SQLServer, etc...).

Una vez creada la página ya en HTML la pasa al servidor web, que a su vez la envía al cliente, donde el navegador visualiza el texto estático html como si éste realmente hubiera existido previamente a la llamada.

El esquema general de la transferencia de páginas web es pues el que se muestra en la Figura siguiente.



Ejemplo:

Código PHP en el servidor	Código HTML enviado al cliente (el cliente nunca ve el código php):
<pre> <html> <body> La fecha del día de hoy es: <?php echo(date("d/m/Y")); ?> </body> </html> </pre>	<pre> <html> <body> La fecha del día de hoy es: 01/09/2010 </body> </html> </pre>

Y en el navegador se vería simplemente:

La fecha del día de hoy es:01/09/2010

La conexión siempre se libera al terminar la transmisión de la página.

Volviendo sobre el modelo de las aplicaciones cliente-servidor, se puede decir que se sigue el denominado modelo de transacciones: el cliente realiza una petición, el servidor la atiende e inmediatamente se cierra la comunicación.

Las transacciones son, pues, independientes, y no se mantiene una memoria entre las sucesivas peticiones.

Esto es un grave inconveniente para muchas aplicaciones en las cuales esta memoria es fundamental, como por ejemplo las operaciones de compra a través de la Web. No obstante, existen mecanismos para resolver el problema, como el uso de cookies o de bases de datos, y que se comentarán más adelante.

La petición desde el cliente (en el caso de la web) se realiza con cierta sintaxis y haciendo uso de la URL (Uniform Resource Locator), en español Localizador Uniforme de recursos. La sintaxis, para la web, sería:

esquema://anfitrión:puerto/ruta?parámetro=valor#enlace

Donde:

- **esquema** : En el caso que nos ocupa HTTP, es en la mayor parte de los casos http, aunque a veces cuando se utiliza una conexión segura (criptografía) se sustituye por https.
- **anfitrión** : También se llama host, nombre de dominio o servidor (en nuestro primer ejemplo sería www.bowwww.bollssaammaaddrriid.d.eess).
- **puerto**: Es el número de puerto por el que se establece la comunicación. Generalmente es el puerto 80, y por eso suele omitirse.
- **ruta**: Indica generalmente un nombre de archivo (con cualquiera de las extensiones jsp, php o asp según el lenguaje) , precedido por nombres de directorios que establecen cierta ruta o path específicas dentro del servidor.
- **?parámetro=valor**: Indica el paso de un parámetro llamado parámetro cuyo valor es valor que acompaña a la petición.

2. Aplicaciones web

2.1. Webmail

El correo electrónico o email es el servicio más utilizado de Internet, junto con la Web.

Hay dos formas básicas de comunicarse usando Internet, el correo electrónico y la mensajería instantánea (WhatsApp, Skype,...).

El correo por Internet, igual que el correo normal, debe ser privado, por esto todos los programas para leer el correo disponen de protección mediante contraseña. Aunque esto no asegura que alguien pueda interceptar el correo mientras viaja a través de la red.

Para evitar esto, se pueden utilizar sistemas de cifrado, como el PGP, que codifican el mensaje antes de enviarlo por la red y lo descodifican al llegar al destino. También disponemos de las ifrmas y certificados electrónicos para garantizar la autenticidad y confidencialidad de los correos.

Funcionamiento del correo.

El correo electrónico o e-mail es una forma de enviar mensajes entre ordenadores conectados a través de Internet.

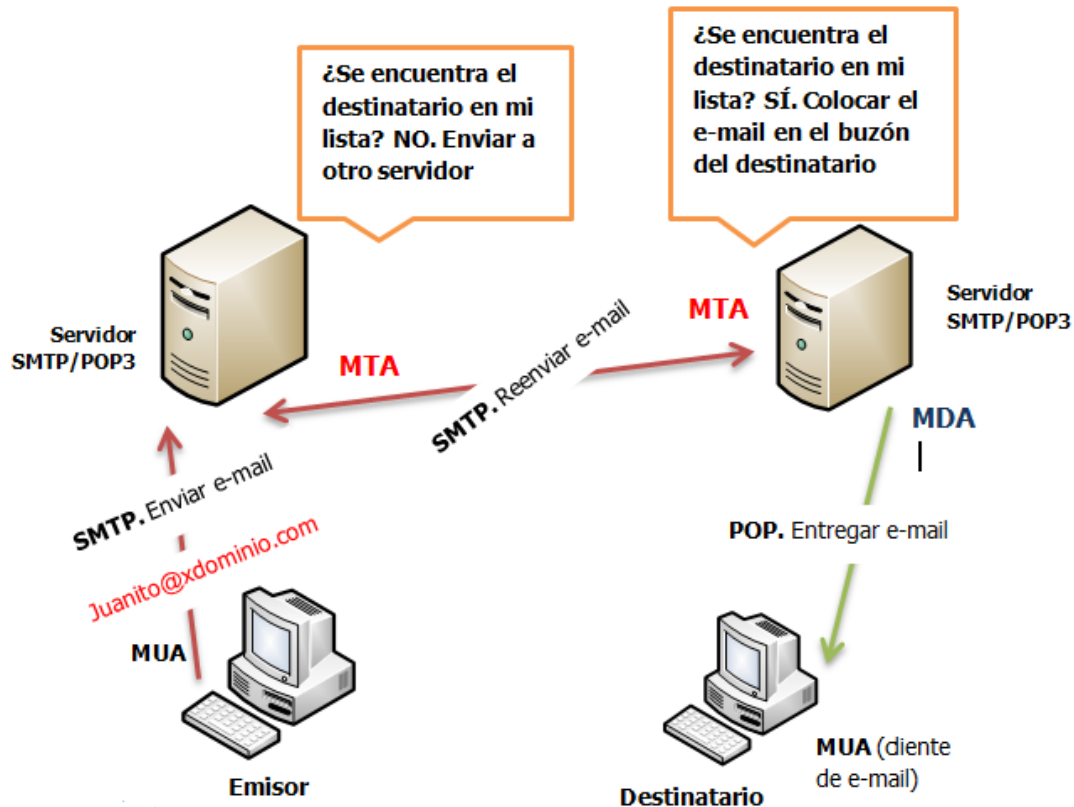
Como la mayoría de los servicios de Internet el correo se basa en la **arquitectura cliente/servidor**. El servicio e-mail utiliza formatos de mensajes y cadenas de comando definidas por el protocolo SMTP

Al escribir un mensaje de correo electrónico, generalmente utilizamos una aplicación denominada **Agente de usuario de correo (MUA)** o comúnmente conocido como **cliente de e-mail**. MUA permite realizar dos procesos:

1. Enviar los mensajes al **buzón del cliente ubicado en el Servidor de correo**;
2. Ubicar y acceder a los mensajes recibidos también desde el **buzón del cliente** .

El **servidor e-mail** ejecuta también 2 procesos individuales:

1. Agente de transferencia de correo (**MTA, Mail Transfer Agent**). Se utiliza para enviar correos electrónicos, es el encargado del intercambio de e-mail entre servidores.
2. Agente de entrega de correo (**MDA, Mail Delivery Agent**). Recibe de MTA un e-mail y realiza el envío real al buzón del destinatario. El MDA también puede resolver temas de entrega final, como análisis de virus, correo no deseado filtrado y manejo de acuses de recibo.



Hay dos tipos de clientes de correo:

- **Webmail o correo web o cliente de correo web** es un cliente de correo electrónico, que proporciona una interface web, a la que accedemos empleando el navegador, y que nos permite acceder al correo electrónico.

Como su nombre indica, consiste en utilizar el correo desde un navegador Web. Una de las ventajas del webmail, es que desde cualquier ordenador que tenga conexión a Internet podemos leer y enviar nuestro correo sin tener que configurar nada.

Cada usuario tiene asociada una dirección de correo electrónico, con formato `nombre@dominio_servidor_correo`, que identifica al usuario en ese servidor de correo. Esta dirección es la utilizada para enviar y recibir correos electrónicos.

- **Cliente de correo local o programa de correo.** Por ejemplo el Outlook, Windows Live Mail, el Thunderbird de Mozilla, Eudora, etc. Estos son programas específicos para trabajar con el correo y que tenemos que instalar en nuestro equipo. La primera vez que se utilizan hay que configurarlos con los datos de la cuenta y servidor de correo. Por lo tanto sólo es práctico utilizarlos en dispositivos que vayamos a utilizar con cierta frecuencia, como el portátil, el equipo personal o en el trabajo.

Pueden manejar varias cuentas a la vez sin importar quién nos haya proporcionado la cuenta. Tienen muchas más opciones que el webmail.

El cliente de correo permite enviar, recibir, leer, escribir y borrar los correos almacenados en el servidor remoto. También proporciona servicios comunes como la gestión de una libreta de contactos (direcciones de correo de otros usuarios), administración de carpetas para gestión de correos, etc. Otros servicios menos comunes como aplicación de filtros, exportación de correos a archivos externos, copias de seguridad de los buzones de correo, etc.

Cada dirección de correo es única para todo el mundo, no pueden existir dos direcciones de correo iguales.

Protocolos de correo.

Los protocolos de correo definen como se realiza la comunicación entre el cliente y el servidor a través de la red. Los más usados son POP e IMAP. Los servicios de Webmail, como Gmail, soportan ambos protocolos. Las cuentas que utilizan el protocolo POP permiten que sea un único usuario el que se conecte, descargue todos los mensajes del servidor a su equipo, y los gestione directamente en el cliente. Puede trabajar con ellos de modo offline.

Con el protocolo IMAP, en cambio, se trabaja únicamente con los correos cuando se está conectado. Está más pensado para trabajar en equipo, empresas o en cooperación, ya que permite que varios usuarios trabajen sobre la misma cuenta de correo a la vez. Por lo tanto, los usuarios no suelen gestionar los contenidos directamente, sino que simplemente los consultan.

Protocolo de transporte SMTP.

La entrega de correo desde una aplicación cliente a un servidor, y desde un servidor origen al servidor destino es manejada por el Simple Mail Transfer Protocol (SMTP)

El objetivo principal del Protocolo simple de transferencia de correo, SMTP, es transferir correo entre servidores de correo. Sin embargo es crítico para los clientes de correo. Para poder enviar correo, el cliente envía el mensaje a un servidor de correo saliente, el cual contacta con el servidor de correo de destino para la entrega. Por esta razón, es necesario especificar un servidor SMTP cuando se esté configurando un cliente de correo. Siempre y cuando no utilicemos correo basado en Web (Webmail).

Un punto importante sobre el protocolo SMTP es que no requiere autenticación. Esto permite que cualquiera en la Internet pueda enviar correo a cualquiera otra persona o a grandes grupos de personas. Esta característica de SMTP es lo que hace posible el correo basura o spam. Los servidores SMTP

modernos intentan minimizar este comportamiento permitiendo que sólo los hosts (PC's) conocidos accedan al servidor SMTP. Los servidores que no ponen tales restricciones son llamados servidores open relay.

Protocolos de acceso a correo (POP e IMAP)

A continuación vamos a ver los dos protocolos más importantes usados por las aplicaciones de correo cliente para recuperar correo desde los servidores de correo: el Post Office Protocol (POP) y el Internet Message Access Protocol (IMAP).

A diferencia de SMTP, estos protocolos requieren autenticación de los clientes usando un nombre de usuario y una contraseña. Por defecto, las contraseñas para ambos protocolos son pasadas a través de la red de forma encriptada.

Protocolo POP.

Cuando se utiliza el protocolo POP, los mensajes de correo son descargados a través de las aplicaciones de correo cliente. Por defecto, la mayoría de los clientes de correo POP son configurados automáticamente para borrar el mensaje en el servidor de correo después que éste ha sido transferido exitosamente, sin embargo esta configuración se puede cambiar.

Para añadir seguridad, es posible utilizar la encriptación Secure Socket Layer (SSL) para la autenticación del cliente y las sesiones de transferencias de datos.

Protocolo IMAP.

Cuando se utiliza un servidor de correo IMAP, los mensajes de correo se mantienen en el servidor donde los usuarios pueden leer y borrarlos. IMAP también permite a las aplicaciones cliente crear, renombrar o borrar directorios en el servidor para organizar y almacenar correo.

IMAP lo utilizan principalmente los usuarios que acceden a su correo desde varias máquinas. El protocolo es conveniente también para usuarios que se estén conectando al servidor de correo a través de una conexión lenta, porque sólo la información de la cabecera del correo es descargada para los mensajes, hasta que son abiertos, ahorrando de esta forma ancho de banda. El usuario también tiene la habilidad de eliminar mensajes sin verlos o descargarlos

Para seguridad adicional, es posible utilizar la encriptación SSL para la autenticación de clientes y para las sesiones de transferencia de datos. También están disponibles otros clientes y servidores de correo IMAP gratuitos así como también comerciales, muchos de los cuales extienden el protocolo IMAP y proporcionan funcionalidades adicionales.

2.2. Navegadores

Un navegador o navegador web (del inglés, web browser) es un **cliente web**, esto es, un programa que permite gestionar recursos web alojados ya sea en un servidor Web dentro de la World Wide Web o en un servidor Web local.

El navegador, en el caso de que el recurso sea una página Web, interpreta el código HTML en el que está escrita y lo presenta en pantalla permitiendo al usuario interactuar con su contenido, que puede incluir recursos multimedia incrustados además de los hipervínculos que enlazan una porción de texto o una imagen a otro documento, permitiendo navegar hacia otros lugares de la red.

En el caso de que el recurso web no esté escrito en HTML, el navegador ejecuta la aplicación correspondiente para su visualización o facilita su descarga.

La función principal de un navegador

Las funciones de un navegador son:

1. **La principal, solicitar** al servidor los **recursos web** que elija el usuario y **mostrarlos** en una ventana. El recurso suele ser un documento HTML, pero también puede ser un archivo PDF, una imagen o un objeto de otro tipo.

El usuario especifica la ubicación del recurso mediante el uso de una **URL (siglas de Uniform Resource Locator, identificador uniforme de recurso)**.

2. El **seguimiento de enlaces** de una página a otra, ubicada en cualquier computadora conectada a la Internet, se llama **navegación**, de donde se origina el nombre navegador (aplicado tanto para el programa como para la persona que lo utiliza, a la cual también se le llama cibernauta).
3. Los navegadores también **ofrecen otras características o utilidades** al usuario para asegurar la navegación, incorporar nuevas funciones o permitir la personalización de la navegación.

Componentes principales del navegador

A continuación se especifican los componentes principales de un navegador.

1. **Interfaz de usuario:** incluye la barra de direcciones, el botón de avance/retroceso, el menú de marcadores, y las utilidades para el usuario etc. (en general, todas las partes visibles del navegador, excepto la ventana principal donde se muestra la página solicitada).

En los últimos años, entre otras, se han añadido **funciones de seguridad** para garantizar una navegación segura. Aunque no todos los navegadores las incorporan, la mayoría ya permiten la gestión de contraseñas, de firmas o certificados de seguridad.

2. **Motor de búsqueda:** coordina las acciones entre la interfaz y el motor de renderización.
3. **Motor de renderización:** es responsable de "renderizar", es decir, mostrar el contenido solicitado. Por ejemplo, si el contenido solicitado es HTML, será el responsable de analizar el código HTML y CSS y de mostrar el contenido analizado en la pantalla.

De forma predeterminada, el motor de renderización puede mostrar imágenes y documentos HTML y XML. Puede mostrar otros tipos mediante el uso de complementos (o extensiones); por ejemplo, puede mostrar documentos PDF mediante un complemento capaz de leer archivos PDF. Sin embargo, en este capítulo nos centraremos en su uso principal: mostrar imágenes y código HTML con formato definido con CSS.

4. **Red:** es responsable de las llamadas de red, como las solicitudes HTTP. Tiene una interfaz independiente de la plataforma y realiza implementaciones en segundo plano para cada plataforma.
5. **Servidor de la interfaz:** permite presentar widgets básicos, como ventanas y cuadros combinados. Muestra una interfaz genérica que no es específica de ninguna plataforma. Utiliza métodos de la interfaz de usuario del sistema operativo en segundo plano (transparente al usuario).
6. **Intérprete de JavaScript:** permite analizar y ejecutar el código JavaScript.
7. **Almacenamiento de datos:** es una capa de persistencia. El navegador necesita guardar todo tipo de datos en el disco duro (por ejemplo, las cookies). La nueva especificación de HTML (HTML5) define el concepto de "base de datos web", que consiste en una completa (aunque ligera) base de datos del navegador.

Estándares web

Los **estándares web** son un conjunto de recomendaciones dadas por el World Wide Web Consortium (W3C) y otras organizaciones internacionales acerca de cómo crear e interpretar documentos basados en el Web. Son un conjunto de tecnologías orientadas a brindar beneficios a la mayor cantidad de usuarios, asegurando la vigencia de todo documento publicado en el Web.

Estos estándares incluyen especificaciones para CSS y HTML, y su objetivo es crear una Web que trabaje mejor para todos, con sitios accesibles a más personas y que funcionen en cualquier dispositivo de acceso a Internet.

Durante años, los navegadores cumplían solo una parte de las especificaciones y desarrollaban sus propias extensiones. Esto provocó graves problemas de compatibilidad para los creadores de contenido web. En la actualidad, la mayoría de los navegadores cumplen las especificaciones en mayor o menor grado.

Protocolos.

HTTPS (*Hyper Text Transfer Protocol Secure*)

El protocolo de Transferencia de Hiper-Texto (HTTPS) es **la versión segura de el http (Hyper Text Transfer Protocol)** . La diferencia es que HTTPS permite realizar transacciones de forma más segura ya que cifra el intercambio de información con los servidores web.



Ya que evitar cotillas en una comunicación es algo que se puede catalogar como muy difícil o imposible, la solución ha consistido en hacer que **nuestro amigo mirón no pueda comprender lo que mira a través de ese agujerito** que ha hecho en la tubería, y que tampoco pueda meter en la tubería mensajes haciéndose pasar por nuestro banco (servidor) o por nosotros mismos (cliente).

1. Antes de empezar esa conexión segura, se establece una comunicación entre el cliente y el servidor en la que se acuerdan los detalles necesarios. **Esta fase se denomina acuerdo**, o en inglés handshake (apretón de manos), y es donde todo se complica un poco.

El handshake entre cliente y servidor puede variar dependiendo de los modos de autenticación (verificación de la identidad del remitente de un mensaje) y cifrado que se vayan a utilizar durante la conexión segura.

2. Los modos de autenticación se deciden en función del certificado que se elige utilizar. Una vez hecho el acuerdo, se **codifica la sesión con un certificado digital**. De este modo, el usuario tiene ciertas garantías de que la información que envíe desde dicha página no podrá ser interceptada y utilizada por terceros.

Este protocolo HTTPS pretende lograr dos objetivos, en función del modo de autenticación utilizado:

- **Verificar nuestra identidad:** Si un mensaje lo firmamos con nuestro certificado digital éste nos identifica ya que sólo nosotros podemos utilizarlo, y además permite verificar la integridad del mismo.
- **Evitar mirones:** Si utilizamos un certificado con claves asimétricas, no sólo garantizamos la identidad del emisor y la integridad del mensaje, sino que podemos garantizar también que sólo el destinatario podrá interpretar nuestro mensaje.

En los navegadores comunes, como Firefox o Explorer, cuando estamos empleando un protocolo https **podemos ver el icono de un candado, que aparece en la barra principal de nuestro navegador**. Además, en la barra de direcciones podremos ver que “http://” será sustituido por “https://”.

Estos certificados de seguridad son conocidos como SSL. Cuando estos están instalados en la página Web veremos el candado del que hablábamos anteriormente. Por otro lado, si están instalados Certificados de Validación Extendida (utilizan claves asimétricas), los usuarios, además del candado, podremos ver que la barra de URL's del navegador toma un fondo verdoso. (siempre que estemos utilizando las últimas versiones de Firefox, Explorer o Opera).

SSL (Security Socket Layer)

El **SSL (Security Socket Layer)** y el **TLS (Transport Layer Security)**¹ son los protocolos de seguridad de uso común que establecen un canal seguro entre dos ordenadores conectados a través de Internet o de una red interna.

Técnicamente, el protocolo SSL es un método transparente para **establecer una sesión segura** que requiere una mínima intervención por parte del usuario final. Por ejemplo, el navegador **alerta al usuario de la presencia de un certificado SSL** cuando se muestra un candado o cuando la barra de dirección aparece en verde cuando se trata de un certificado EV SSL con validación ampliada.



¿Qué es un certificado SSL?

SSL es un protocolo para cuyo uso es necesario contar con un certificado SSL. Un certificado SSL es un pequeño archivo de datos que vincula digitalmente una clave de encriptación o claves con los datos de su empresa. Generalmente estos datos son:

- **Su nombre de dominio, nombre de servidor y nombre de *host***
- **El nombre y la ubicación de su empresa**
- **En determinados casos, la información de contacto de su empresa**

Para poder **iniciar sesiones SSL con navegadores**, las organizaciones deben instalar el certificado SSL en sus servidores web. Dependiendo del certificado SSL solicitado, la organización deberá superar determinados niveles de inspección.

Tras la instalación, al intentar acceder a una de sus páginas web el certificado solicitará al servidor que establezca una conexión segura con el navegador, lo que protegerá todo el tráfico entre el servidor web y el navegador.

¹TLS (Transport Layer Security) es el sucesor del protocolo SSL. Versiones de TLS tienen un equivalente en SSL, por ejemplo TLS 1.2 corresponde a SSL 3.3; de ahí que aún sea común que se refiera al protocolo TLS como SSL y que en un contexto informal se utilicen estos términos de forma intercambiable

Para **visualizar un certificado SSL**, basta hacer clic en el candado y seleccionar “Visualizar certificado”. Cada navegador mostrará el certificado de manera ligeramente diferente, pero siempre incorporará la misma información.

FTPS (File Transfer Protocol Secure) /SFTP (Secure Shell File Transfer Protocol)

En una conexión normal por FTP, el usuario y contraseña se envían a través de internet en **texto plano**, sin ningún tipo de seguridad. Aunque es poco probable, esta información podría ser interceptada por el camino.

Los protocolos FTP fueron avanzando, con alternativas como SFTP o FTPS

FTPS sobre SSL/TLS

FTPS utiliza SSL/TLS para cifrar la conexión. El uso más común de FTP y SSL es:

- *AUTH TLS* o FTPS Explícito, nombrado por el comando emitido para indicar que la seguridad TLS es obligatoria. Este es el método preferido de acuerdo al RFC que define FTP sobre TLS. El cliente se conecta al puerto 21 del servidor y comienza una sesión FTP sin cifrar de manera tradicional, pero pide que la seguridad TLS sea usada y realiza la negociación apropiada antes de enviar cualquier dato sensible.
- *FTPS Implícito* es un estilo antiguo, pero todavía ampliamente implementado en el cual el cliente se conecta a un puerto distinto (como por ejemplo el 990), y se realiza una negociación SSL antes de que se envíe cualquier comando FTP.

El cliente FTP se conecta al servidor de forma normal, negocia el nivel de seguridad, y una vez comprobado que puede enviar **datos encriptados utilizando certificados SSL**, envía de esta forma los datos sensibles.

Secure Shell FTP (SFTP)

El Protocolo de transferencia de archivos SSH (SFTP) es un protocolo de red que utiliza Secure Shell (SSH) para transferir archivos.

SSH utiliza **criptografía de clave pública** para autenticar el sistema remoto y permitir al mismo autenticar al usuario, si es necesario.

A **diferencia del FTP estándar**, cifra tanto mandatos como datos, evitando que las contraseñas y la información confidencial se transmitan en texto claro a través de la red. **La totalidad de la sesión de inicio de sesión, incluida la transmisión de la contraseña, está cifrada**, haciendo mucho más difícil que una persona ajena pueda ver y capturar contraseñas. Al cifrar todo el tráfico, SFTP elimina de forma efectiva las escuchas no autorizadas, el secuestro de conexiones y otros ataques a nivel de red.

SFTP no es simplemente FTP ejecutado sobre SSH; SFTP es un nuevo protocolo. Es funcionalmente similar a FTP pero, debido a que utiliza un protocolo

diferente, no se puede utilizar un cliente FTP estándar para comunicarse con un servidor SFTP, ni tampoco puede conectarse a un servidor FTP con un cliente que sólo soporte SFTP.

Contraseñas

Los navegadores nos permiten almacenar las **contraseñas** personales que introducimos en diferentes sitios webs. Estas contraseñas se almacenan con un nivel de seguridad bajo en nuestro ordenador de forma que podrían ser descubiertas por cualquier persona que tuviese acceso a él.

Es necesario comprobar que nuestra versión para el navegador posee un gestor de contraseñas que permite, como mínimo, utilizar una contraseña maestra para codificarlas. Este gestor también debe aportar algún mecanismo para recuperar fácilmente estas contraseñas almacenadas.

Cookies

Una **cookie** es un archivo de texto que se almacena en el disco duro del visitante de una página web a través de su navegador, a petición del servidor de la página con cierta información de su visita a esa página. Esta información puede ser luego recuperada por el servidor en posteriores visitas.

Al ser el protocolo HTTP incapaz de mantener información por sí mismo, para que se pueda conservar información entre una página vista y otra (como login de usuario, preferencias de colores, etc), ésta debe ser almacenada, ya sea en la URL de la página, en el propio servidor, o en una cookie en el ordenador del visitante.

Otro aspecto importante es el hecho de que se establece una conexión independiente para cada documento u objeto que se transmite. Es decir, si una página contiene, por ejemplo, cuatro imágenes, entonces se establecen cinco conexiones independientes: una para la página propiamente dicha y otras cuatro para las imágenes.

De esta forma, los usos más frecuentes de las cookies son:

- Llevar el **control de usuarios**: cuando un usuario introduce su nombre de usuario se almacena una cookie para que no tenga que estar introduciéndolas para cada página del servidor. Sin embargo una cookie **no identifica a una persona, sino a una combinación de computador-navegador-usuario**. Ojo: las contraseñas que guarda el navegador para recordarlas la próxima vez que nos conectemos a esa web no se almacenan en cookies.
- Conseguir información sobre los **hábitos de navegación** del usuario, e intentos de spyware, por parte de agencias de publicidad y otros. Esto

puede causar problemas de privacidad y es una de las razones por la que las cookies tienen sus detractores.

Añadiendo funcionalidad a los navegadores

¿Qué es un complemento? Los complementos son pequeños elementos software que añaden nuevas características o funcionalidades a tu navegador.

¿Cuáles son los diferentes tipos de complementos?

- **Las extensiones** añaden nuevas funciones o modifican alguna funcionalidad ya existente. Hay extensiones que te permiten bloquear publicidad, descargar videos de sitios web, mejorar la integración con las redes sociales y añadir características que se ven en otras aplicaciones.
- **Los temas** cambian el aspecto, normalmente incluyendo iconos, colores, cuadros de diálogo y otros estilos visuales.
- **Personas** son temas ligeros que utilizan imágenes de fondo para personalizar las barras de herramientas.
- **Los buscadores** añaden más opciones a la lista desplegable del cuadro de búsqueda. Estos buscadores te permiten buscar rápidamente cualquier página web.
- **Los diccionarios y los paquetes de idiomas** permiten utilizar idiomas adicionales en Firefox.
- **Los plugins** ayudan a visualizar o entender los diferentes contenidos multimedia, como Adobe Flash o Apple Quicktime.

2.3. Buscadores

Un **buscador** es una página web en la que se ofrece consultar una base de datos en la cual se relacionan direcciones de páginas web con su contenido. Su uso facilita enormemente la obtención de un listado de páginas web que contienen información sobre el tema que nos interesa.

El origen de los buscadores se remonta a abril de 1994, año en el que una pareja de universitarios norteamericanos (David Filo y Jerry Yang) decidieron crear una página web en la que se ofreciera un directorio de páginas interesantes clasificadas por temas, pensando siempre en las necesidades de información que podrían tener sus compañeros de estudios. Había nacido **Yahoo!**. El éxito de esta página fué tan grande que una empresa decidió comprarla y convertirla en el portal que hoy conocemos. Además del buscador, hoy Yahoo! ofrece muchos más servicios.

Existen varios tipos de buscadores, según la forma de obtener las direcciones que almacenan en su base de datos, pero todos ellos tienen en común que permiten una consulta en la que el buscador nos devuelve una lista de direcciones de páginas web relacionadas con el tema consultado.

Los buscadores se pueden clasificar en tres tipos:

1. **Índices de búsqueda.** En los índices de búsqueda, la base de datos con direcciones la construye un **equipo humano**. Es decir, un grupo de personas va rastreando la red en busca de páginas y las **clasifican por categorías** ó temas y subcategorías en función de su contenido. La consulta de un índice se realiza a través de categorías. Yahoo, inicialmente buscador de índices, combina ahora ambos métodos, índices y motor de búsqueda.
2. **Motor de búsqueda.** El rastreo de la web lo hace un programa, llamado *araña* ó motor (de ahí viene el nombre del tipo de buscador). Este programa va visitando las páginas y, a la vez, creando una base de datos en la que **relaciona la dirección de la página con las 100 primeras palabras que aparecen en ella**. El acceso a esta base de datos se hace por palabras clave: la página del buscador me ofrece un espacio para que yo escriba la ó las palabras relacionadas con el tema que me interesa, y como resultado me devuelve directamente un listado de páginas que contienen esas palabras clave. Google
3. Los **metabuscadors** son páginas web en las que se nos ofrece una búsqueda sin que haya una base de datos propia detrás: utilizan las bases de varios buscadores ajenos para ofrecernos los resultados. Un ejemplo de metabuscador es Metacrawler.

Buscador Google

Google, como motor de búsqueda de texto completo, incluye en su índice páginas Web completas, en lugar de incluir sólo los títulos y las descripciones. Mediante los comandos adicionales, denominados sintaxis especiales, nos permite realizar búsquedas en partes determinadas de páginas Web o tipos específicos de información. Esto resulta práctico cuando se está tratando con dos mil millones de páginas y necesitamos aprovechar cualquier oportunidad para restringir esos criterios.

- **AND:** Esta opción está establecida como predeterminada en Google, siempre que introducimos las palabras busca a lo largo de todo el contenido.
- **"varias palabras":** Si introducimos las comillas dobles, conseguimos que busque en las páginas que contengan esas palabras juntas.
- **OR:** Puede mostrarnos páginas que contengan una palabra u otra.
- **-palabra / -"varias palabras":** Usaremos el signo de "menos" cuando queremos que esa palabra no aparezca en los resultados.
- **+palabra / +"varias palabras":** Usaremos el signo de "más" para

asegurarnos de que la palabra o palabras aparecen en los resultados (google a veces prescinde de ciertas palabras para buscar)

También es posible combinar estas opciones, pero hemos de tener en cuenta que Google solo reconoce un máximo de 10 palabras.

Sintaxis especiales

- `intitle:palabras` - Devuelve las páginas que contengan en su título ALGUNAS de las palabras especificadas.
- `allintitle:palabras` - Devuelve las páginas que contengan TODAS las palabras en el mismo título. No se usa muy a menudo ya que no se combina muy bien con algunas de las otras sintaxis.
- `inurl:palabras` - Devuelve las páginas donde en su URL aparecen ALGUNAS de las palabras.
- `allinurl:palabras` - Devuelve las páginas donde en su URL están TODAS de las palabras. Tampoco os aconsejo mucho su uso, igual que `allintitle`, no se combina muy bien con otras sintaxis.
- `intext:palabras` - Busca sólo en el cuerpo del texto que contenga ALGUNAS palabras, ignorando el texto de los vínculos, URL y títulos.
- `inanchor:palabras` - Busca entre los textos que describen los vínculos.
- `site:sitio` o `site:dominio` - Permite restringir esos resultados bien a un sitio Web bien a un dominio de nivel superior. Esta sintaxis se utiliza mucho en el posicionamiento ya que permite ver el número de páginas almacenadas en los buscadores.
- `link:dirección` o `link:dirección/página` - Devuelve una lista de las páginas que tienen vínculos al URL especificado. No hemos de olvidar las “www” pero sí podemos ahorrarnos el “http://” ya que Google parece que lo ignora. últimamente parece que han cambiado algo y no sé si es que son mas rigurosos o han eliminado parte de la funcionalidad.
- `cache:dirección` - Encuentra una copia de la página que Google muestra en el índice incluso si esa página ya no está disponible en su URL original, o si ha cambiado su contenido completamente desde entonces. Os recuerdo que Google tiene las páginas almacenadas y cuando buscamos lo hacemos en su caché.
- `related:dirección` - Encuentra páginas relacionadas con la página que especifiquemos, pero no todas las páginas están relacionadas con otras. En mi ejemplo, obtendríamos como resultado una variedad de motores de búsqueda.
- `info:dirección` - Debajo los resultados nos proporciona vínculos que ofrecen más información sobre un determinado URL, como: la caché de Google, lista de páginas similares, tengan enlaces, aparezcan los términos... Hay que tener en cuenta que esta información depende de si Google tiene en su índice dicho URL o no. Si no lo tiene, la información será, obviamente, más limitada.
- `type:tipo` (tipo puede ser pdf, doc, ppt, etc...). Devuelve solo documentos en el formato especificado.

2.4. Herramientas cooperativas.

Uno de los usos más importantes de las webs es el trabajo cooperativo, la capacidad para compartir información cooperativamente. Por ejemplo: realizar un documento entre varios usuarios de un proyecto al mismo tiempo y que otros puedan visualizar dicho documento.

Grupos de noticias

Los grupos de Usenet se desarrollaron por primera vez en 1979 en Duke University como una manera para que los estudiantes de Duke se comunicaran entre sí a través de los diferentes campus. Como otras universidades establecieron sus propios servidores de Usenet, expandieron los servicios. Los grupos de Usenet fueron los precursores a los foros y los blogs basados en la web, ofreciendo debate en línea sobre miles de temas.

Los usuarios leen y envían mensajes textuales a distintos tableros distribuidos entre servidores con la posibilidad de enviar y contestar a los mensajes.

Protocolo NNTP (Network News Transport Protocol)

El NNTP es un protocolo inicialmente creado para la lectura y publicación de artículos de noticias en Usenet

El funcionamiento del NNTP es muy sencillo: consta de un servidor en el que están almacenadas las noticias y a él se conectan los clientes a través de la red.

La conexión entre cliente y servidor se hace de forma interactiva consiguiendo así un número de artículos duplicados muy bajo. Esto supone una gran ventaja respecto de servicios de noticias anteriores, en los que la tecnología por lotes era su principal aliada.

Esta conexión se realiza sobre el protocolo TCP. El puerto 119 está reservado para el NNTP. Sin embargo cuando los clientes se conectan al servidor de noticias mediante SSL se utiliza el puerto 563.

Cada artículo de noticias almacenado en el servidor está referenciado por el nombre de la máquina del cliente que ha publicado dicho artículo. Esta referencia queda presente en un campo de la cabecera llamado NNTP-Posting-Host.

Clientes de noticias

Al igual que otros servidores de noticias, los clientes podrán listar los grupos de noticias disponibles y descargar, enviar y publicar artículos mediante órdenes específicas del NNTP.

Los programas cliente para leer y escribir a grupos de noticias, generalmente integrados con un programa cliente de correo electrónico o en navegadores. Los mensajes suelen ser temáticos y el tráfico es enorme, por lo que solo aparecen los mensajes más recientes. Algunos grupos de noticias son moderados.

Los grupos de noticias tienen un sistema de nombres jerárquicos. Los grupos se dividen en varias categorías, tales como "rec" para "recreación", luego se subdividen en, a veces, muchas veces, para reducir el foco de cada grupo a un tema muy específico. Los períodos se utilizan para mostrar la jerarquía de nombres de grupos de noticias como "rec.arts.movies" y "alt.comp.mail.qmail."

Grupo de noticias RedIris:

http://www.rediris.es/netnews/infonews/usenet_info.es.html

Grupos Unset públicos contra los privados

Mientras que la mayoría de grupos de noticias son grupos públicos, los administradores de servidores pueden crear grupos que sólo están disponibles en su servidor Usenet local. Muchas universidades, por ejemplo, establecen grupos de discusión para estudiantes y profesores para que no se propaguen fuera de su intranet local.

Hay 9 jerarquías principales, cada una dedicada a discusiones sobre un tipo de temas:

- comp.*: Temas relacionados con las computadoras.
- news.*: Discusión del propio Usenet.
- sci.*: Temas científicos.
- humanities.*: Discusión de humanidades (como literatura o filosofía).
- rec.*: Discusión de actividades recreativas (como juegos y aficiones).
- soc.*: Socialización y discusión de temas sociales.
- talk.*: Temas polémicos, como religión y política.
- misc.*: Miscelánea (todo lo que no entre en las restantes jerarquías).
- Alt.*: Salió como alternativa a talk, pero es usada por usuarios P2P.

Los grupos de noticias públicos cada vez se utilizan menos, en favor de otros servicios web como las Web's académicas, blogs, redes sociales, etc.

Pero todavía podemos probar alguna versión gratuita de exploradores específicos para Usenet como Mimo de <http://es.giganews.com/>

Ofimática WEB

Las últimas tendencias se basan en el traslado de paquetes ofimáticos desde un sistema nativo a un servidor remoto basado en Software como Servicio (SaaS) y

la Computación en nube (cloud computing) dos conceptos que permiten disponer de la información en cualquier momento, en cualquier lugar, utilizando simplemente un navegador web.

Estudiaremos con más detalle estas herramientas en la UD6 del módulo.

CMS-Gestores de contenido.

Los gestores de contenido son aplicaciones web que permiten la creación de páginas web de forma intuitiva, sin necesidad de poseer conocimientos técnicos específicos, lo que ha facilitado que personas sin conocimientos de informática, sistemas o programación puedan realizar y mantener un sitio web.

Existen diferentes tipos de gestores de contenido, cada uno de los cuales es más adecuado para el desarrollo de sitios web de un tipo determinado. Los más cooperativos son las wikis y los blogs.

Estudiaremos con más detalle estas herramientas en las UD3 y 4 del módulo.

3. Almacenamiento web

En ocasiones necesitamos almacenar datos en algún lugar seguro y que podamos recuperarlos desde cualquier lugar. Los servicios de almacenamiento web ofrecen discos duros virtuales en Internet a los que se accede mediante un navegador. También se les conocen como servicios de almacenamiento y distribución en línea.

Los parámetros que definen un servicio de almacenamiento vienen condicionados por el tipo de servicio contratado, ya que normalmente, todos los sitio de almacenamiento, ofrecen opciones gratuitas y de pago, en función de las cuales se modifican las condiciones de uso del servicio.

Estudiaremos con más detalle estas herramientas en las UD5 del módulo.

4. Gestión de servicios web multimedia

El término multimedia define la capacidad de un ordenador de mostrar o reproducir conjuntamente audio, animación, imagen, vídeo, etc., haciendo para ello uso de software específico llamado aplicaciones multimedia.

Con la aparición de tecnologías de acceso dedicado se ha ampliado el ancho de banda consiguiendo conexiones más potentes, lo que ha llevado a una auténtica revolución multimedia en la web.

Audio digital en la web

Las mayores innovaciones en este campo han venido dadas por lo que se ha denominado streaming, la transmisión de audio digital por la red en tiempo real bajo demanda, tanto en eventos en diferido como en directo.

En 2004 apareció el concepto de podcasting, que consiste en la distribución de archivos multimedia (normalmente audio o vídeo, que puede incluir texto

como subtítulos y notas) mediante un sistema de redifusión (RSS) que permita suscribirse y usar un programa que lo descarga para que el usuario lo escuche en el momento que quiera. No es necesario estar suscrito para descargarlos. Así se pueden reproducir directamente desde la página que los distribuye o bien descargarlos y reproducirlos cuando se desee.

Imagen digital en la web

A raíz de una comunidad de juegos on-line surgió Flickr, que en 2004 creó este servicio específico para compartir imágenes bajo licencia Creative Common. Dado su éxito, Yahoo! La adquirió en 2005.

Como era de esperar, empresas como Google o Microsoft, con Picasa Web y Live, están tratando de ganar usuarios a las redes sociales como Facebook o Tuenti, que tanto éxito han tenido justamente por compartir recursos multimedia entre sus usuarios.

También se puede crear un servicio con software libre y servidores web propios dentro de tu propio dominio, y añadir los complementos necesarios para hacer más dinámica la galería de imágenes. Dos de las más populares son **Gallery** y **Coppermine**, gratuitas y que funcionan sobre AMP.

Video digital en la web

En el campo de los videos digitales almacenados en la web, destacamos los primeros pootales que surgieron a principios de 2005, prácticamente a la par: Google Videos y Youtube.

El primero, Google Videos, no tuvo demasiado éxito y Google compró a finales de 2006 Youtube abandonando más adelante su proyecto inicial.

En la actualidad Youtube, líder indiscutible en vídeo por internet, es la plataforma perfecta para llegar al máximo público en el menor tiempo. Además ha abierto y mejorado sus API, lo que ha permitido crear mashups.

Todos los videos en Youtube disponen de un enlace externo o un código HTML con la etiqueta object, bien para ser enlazados, bien para ser incrustados en páginas web externas. Así mismo, permite que los usuarios opinen sobre los videos, y que empresas e instituciones creen canales con sus videos, a los que los usuarios pueden suscribirse.

En cuanto a la edición de video desde la web, existe una versión específica de Adobe Premier, llamada Express y disponible en el portal Photobucket.com. Basta con cinco pasos para montar un vídeo a partir de grabaciones o fotografías, añadir textos o etiquetas, marcos, música y transiciones entre los elementos que lo componen.