

La World Wide Web

Gregorio Robles Martínez

PID_00150271



Universitat Oberta
de Catalunya

www.uoc.edu



Los textos e imágenes publicados en esta obra están sujetos –excepto que se indique lo contrario– a una licencia de Reconocimiento-NoComercial-SinObraDerivada (BY-NC-ND) v.3.0 España de Creative Commons. Podéis copiarlos, distribuirlos y transmitirlos públicamente siempre que citéis el autor y la fuente (FUOC. Fundació para la Universitat Oberta de Catalunya), no hagáis de ellos un uso comercial y ni obra derivada. La licencia completa se puede consultar en <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.es>

Índice

| | |
|--|----|
| Introducción | 5 |
| Objetivos | 6 |
| 1. Historia | 7 |
| 1.1. La web, el servicio estrella de Internet | 7 |
| 2. HTTP: cliente y servidor | 9 |
| 2.1. El servidor | 9 |
| 2.2. El cliente | 12 |
| 3. El protocolo HTTP | 16 |
| 3.1. Un ejemplo de comunicación con HTTP | 18 |
| 4. El lenguaje HTML. Nociones básicas | 20 |
| 4.1. Un ejemplo de página web | 22 |
| 5. La web dinámica o interactiva y aplicaciones web | 26 |
| 5.1. Buscadores | 28 |
| 6. Web 2.0 | 31 |
| 6.1. Páginas wiki | 32 |
| 6.2. Blogs | 35 |
| 6.3. Suites ofimáticas en línea | 39 |
| 6.4. Otros sistemas web 2.0 | 40 |
| 7. La web multimedia | 42 |
| 7.1. Imágenes | 42 |
| 7.2. Audio | 44 |
| 7.3. Vídeo digital y la compresión de datos | 46 |
| 7.4. Sistemas de compartición de imágenes, audios y vídeos | 49 |
| Resumen | 50 |
| Actividades | 51 |
| Ejercicios de autoevaluación | 51 |
| Solucionario | 52 |
| Glosario | 53 |

Bibliografía..... 56

Introducción

La **World Wide Web** (o WWW, comúnmente conocida como **web**) es el servicio más utilizado en Internet, hasta el punto de que, aun siendo sólo uno de los servicios existentes, popularmente se confunde todo Internet con él.

En este módulo vamos a adentrarnos en todo lo relacionado con la web. Empezaremos con una perspectiva histórica, desde su concepción hasta su popularización. A continuación, conoceremos el modelo arquitectónico en el que se basa, entendiéndose aquí la arquitectura desde un punto de vista informático. De esta manera, hablaremos de los componentes que necesitamos para que la web funcione:

- un cliente que pedirá información,
- un servidor que la proveerá,
- un protocolo de comunicaciones (HTTP) para poder intercambiar esta información, y
- un lenguaje estándar (HTML) para enriquecer el texto y convertirlo en hipertexto.

Pero la web moderna tiene mucho más, y parte de todo ello se tratará también en este módulo. En particular, hablaremos de la web dinámica o interactiva, las aplicaciones web y la web 2.0. Finalmente, veremos cómo se han integrado en la web elementos multimedia como las imágenes, el audio y el vídeo.

Objetivos

Tras el estudio de este módulo, el estudiante deberá haber adquirido las siguientes competencias:

- 1.** Conocer la historia de la web. Se estudiará quién creó la web y cómo ha ido evolucionando desde sus comienzos.
- 2.** Tener nociones de cómo funciona la web. Se verá qué elementos intervienen en este servicio y cómo se relacionan entre ellos.
- 3.** Comprender los conceptos básicos del lenguaje HTML. El lenguaje HTML es el lenguaje en el que están escritas las páginas web. El estudiante entenderá los mecanismos básicos de éste y tendrá las nociones suficientes para utilizar sus elementos en páginas web sencillas y en otros lugares donde también pueden ser de utilidad (por ejemplo, en foros).
- 4.** Conocer los fundamentos de la web dinámica y de la web 2.0. La web ha evolucionado desde sus comienzos estáticos e interactúa en la actualidad con el usuario de una manera más sencilla y rica. Esto abre un mundo de posibilidades, como se verá en el caso de la web 2.0.
- 5.** Entender los elementos multimedia y cómo se integran en la web. La web no es sólo texto; también hay imágenes, audio y vídeo. Veremos cómo se componen estos elementos multimedia y cómo se incluyen en las páginas web.

1. Historia

Aunque los orígenes de Internet se pueden identificar en la década de los sesenta del siglo pasado, la web es un servicio más bien reciente.

Un **servicio de Internet** resuelve una necesidad de un usuario por medio de Internet.

En los servicios de Internet intervienen varios ordenadores, por lo que se sostienen sobre protocolos y estándares. Normalmente cada servicio tiene su propio protocolo para transmitir la información durante el intercambio de datos.

Aparte del servicio web, existen otros servicios de Internet ampliamente utilizados:

- Correo electrónico.
- Mensajería instantánea.
- Acceso remoto.
- Compartición P2P.
- Juegos multi-usuario.
- *Streaming* de vídeo.
- Telefonía por Internet.
- Vídeo-conferencia.
- Computación masivamente distribuida.

Gran parte del éxito de la web reside en su facilidad de uso, lo que ha hecho que muchos de los servicios de Internet mencionados ofrezcan una interfaz web. Por lo tanto, hacemos uso de ese servicio dentro de un entorno que resulta familiar para el usuario: su navegador web.

Ejemplo de servicio de Internet a través del navegador web

El servicio de correo electrónico es un ejemplo de servicio de Internet al que se accede mediante un navegador web. Hoy en día existen múltiples plataformas que ofrecen el servicio de correo electrónico vía web, como pueden ser Gmail, Yahoo! mail, Hotmail o el correo de la UOC. Los usuarios pueden abrirse una cuenta y gestionar sus buzones y mensajes de correo electrónico navegando por páginas web.

1.1. La web, el servicio estrella de Internet

Como ya hemos comentado, el servicio web no es uno de los primeros que existían en Internet. El servicio web es muy reciente, pero ha evolucionado muy rápidamente.

Ved también

Recordad que los protocolos de comunicaciones se explicaron en el módulo "Aspectos tecnológicos de las redes e Internet".

La historia de la web tiene su comienzo en el año 1989, cuando Tim Berners-Lee (conocido popularmente como TBL) propone un nuevo proyecto para facilitar el intercambio de información científica.

Berners-Lee, físico de formación, trabajaba en la Organización Europea para la Investigación Nuclear¹ (CERN, siglas en francés) en Ginebra. Se dio cuenta de que en la Internet de entonces ya existía una gran cantidad de información, pero que estaba vagamente organizada y estructurada. Así, la información estaba disponible y en gran cantidad, pero era difícil acceder a ella, ya que se encontraba en texto plano en múltiples documentos, sin referencias entre sí.

Por esta razón, Berners-Lee hizo uso del concepto **hipertexto** para lanzar un nuevo proyecto. La idea de hipertexto, conocida desde los años sesenta, se basa en que la información se puede organizar de manera dinámica mediante conexiones y enlaces –que reciben el nombre de **hipervínculos**– entre los documentos. Incluso existen documentos que simplemente sirven de índice, o sea, que son documentos con un listado de documentos enlazados. Cuando un usuario selecciona un hipervínculo dentro de un hipertexto, el programa le mostrará el documento al que hacía referencia ese hipervínculo. Había nacido la web, un tejido en el que los documentos se enlazan unos con otros.

Para llevar a cabo su idea, Berners-Lee ideó y creó lo siguiente:

- un **protocolo de transmisión**, que bautizó como protocolo de transmisión de hipertexto² (HTTP). Este protocolo sirve para indicar cómo se ha de pedir un documento de hipertexto y cómo se ha de enviar el documento.
- un **lenguaje de marcado** para enriquecer el texto plano y convertirlo en hipertexto, llamado *HyperText Markup Language* (HTML). La razón por la que se creó el HTML es que se necesitaba añadir información a los documentos de texto plano para que incluyeran los hipervínculos y otros elementos que veremos más adelante.
- un **servidor** en el que alojar las páginas de hipertexto (que llamó CERN HTTPd).
- un **cliente**, conocido como **navegador**, para acceder a las páginas alojadas en el servidor (y que curiosamente recibió el nombre de World Wide Web).

Hacia finales de 1990, con la ayuda de Robert Cailliau, un joven estudiante del CERN consiguió la primera comunicación con éxito entre el cliente y el servidor. Debido a su facilidad de uso, en poco tiempo el servicio WWW se convirtió en popular, lo que desplazó o integró otros servicios existentes en Internet con anterioridad. En 1994 Berners-Lee creó el World Wide Web Consortium (W3C) para regular las tecnologías relacionadas con la web (HTTP, HTML, etc.).

⁽¹⁾Conocido en inglés como *European Organization for Nuclear Research*.

Texto plano

El texto plano es texto sin formato.

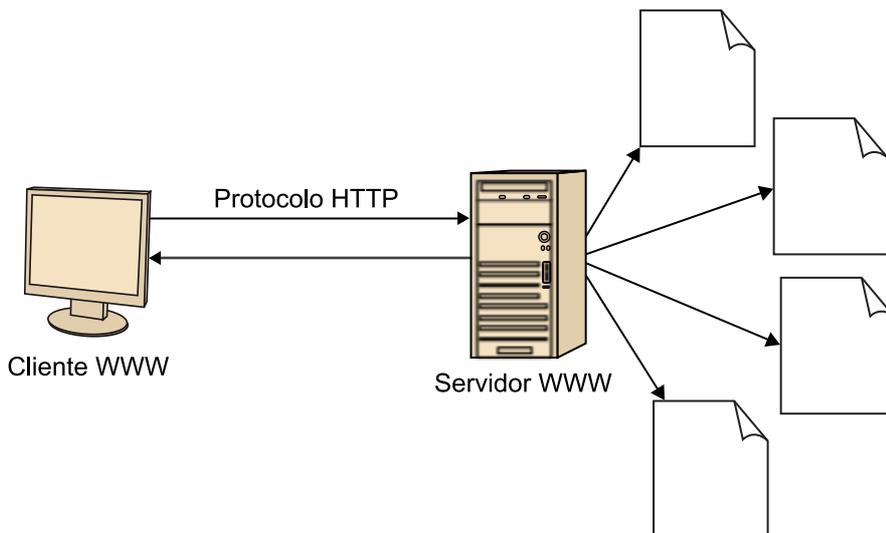
⁽²⁾En inglés *HyperText Transfer Protocol*.

2. HTTP: cliente y servidor

La web se basa en el **paradigma cliente-servidor**, una de las arquitecturas que podemos encontrar comúnmente en Internet. En este paradigma, los actores que intervienen no están al mismo nivel ni tienen los mismos requisitos.

Tal y como muestra la figura siguiente, tenemos un nuevo modo de acceder a documentos que se encuentran en otro ordenador (al que llamamos **máquina servidora** o simplemente el **servidor**), y que son transferidos mediante un protocolo específico llamado HTTP. La máquina servidora tiene un software, que se denomina software **servidor web**, cuya tarea se limita a traducir las direcciones que recibe por HTTP en la ruta en local de los archivos para enviarlos al **cliente web** (un navegador) que los ha pedido.

Arquitectura cliente-servidor de un servicio WWW básico



2.1. El servidor

El servidor contiene **páginas web**, documentos de texto enriquecidos con hipervínculos y está atendiendo constantemente peticiones por parte de clientes. De esta manera, su labor reside en estar continuamente escuchando peticiones provenientes de clientes, interpretándolas y atendiéndolas, posiblemente respondiendo con la página web (u otros elementos que haya pedido). De este modo, *ad infinitum*. Uno de los requisitos esenciales de un servidor es que esté siempre encendido, porque de lo contrario no podrá atender convenientemente las peticiones que le llegan. Otro requisito es que posea una dirección IP fija, ya que de alguna manera los clientes han de conocer a quién dirigirse de antemano. El servidor web será un programa en el nivel de aplicación.

Ved también

La definición del paradigma cliente-servidor se puede consultar en el subapartado 1.1 del módulo "Aspectos tecnológicos de las redes e Internet".

Ved también

Recordad que los niveles OSI se explicaron en el subapartado 1.3 del módulo "Aspectos tecnológicos de las redes e Internet".

Si el número de peticiones que un servidor tiene que atender es muy elevado, entonces será necesario contar con una máquina más potente (o sea, con mayor capacidad de procesamiento, mayor tamaño de memoria y mayor capacidad de almacenamiento en disco duro, por ejemplo). Por esta razón, aunque cualquier máquina pudiera actuar de servidor, cuando se espera mucho tráfico suelen ser generalmente máquinas bastante potentes (y caras).

Pero es muy probable que si la web que aloja es muy frecuentada, no baste con una sola máquina. Entonces entra en escena el concepto granjas de servidores³, que permiten escalar un servicio ofrecido a muchas máquinas. Las granjas de servidores actúan de manera transparente al cliente; esto es, el cliente realizará la petición como si sólo hubiera un servidor y no es consciente de cómo se trata su petición, si por una única máquina o por varias.

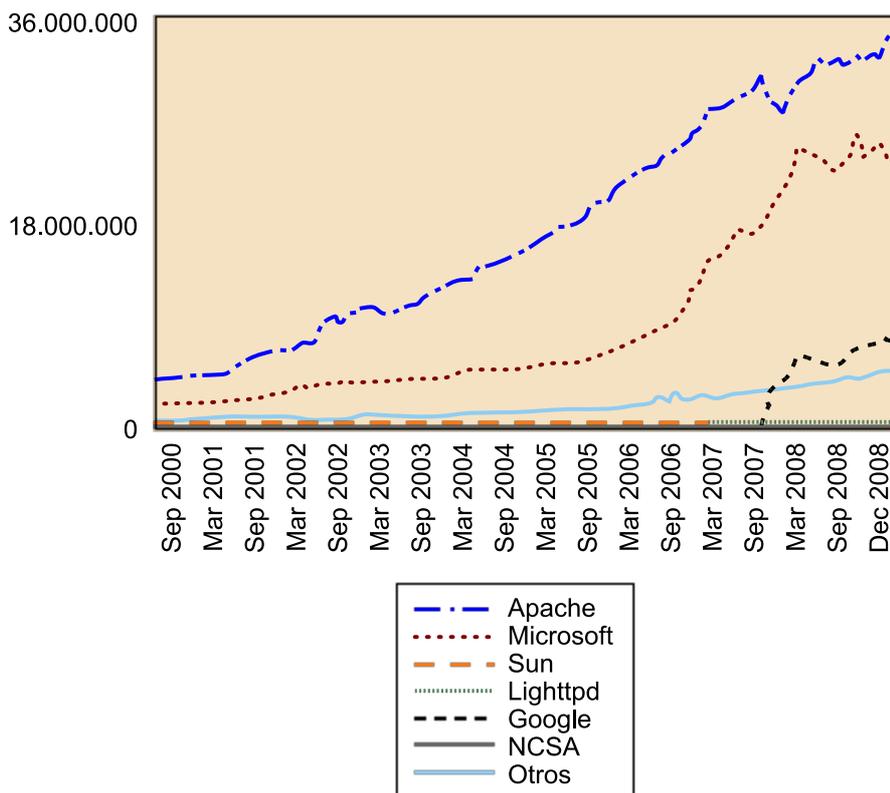
⁽³⁾En inglés, *server farms*.

Ha de notarse que servidor es una palabra polisémica y que se conoce como servidor web tanto a la máquina (física) que alberga las páginas web, como al propio software (lógico) que procesa y responde a las peticiones de los clientes.

Como se ha comentado con anterioridad, cualquier máquina física puede actuar de servidor y se pueden adquirir máquinas especialmente dotadas para servicios muy frecuentados.

En cuanto al servidor lógico, existen múltiples programas de ordenador en el mercado que pueden realizar sus funciones, aunque hay unos cuantos que históricamente han sido muy utilizados y que creemos conveniente conocer.

Crecimiento en términos absolutos del número de servidores web



(c) Netcraft LTD. Fuente: Netcraft LTD

Existen dos programas que son mayoritariamente utilizados: el servidor web Apache, un programa de software libre, y el Microsoft Internet Information Server (también conocido como *IIS*), una aplicación privativa creada por el gigante tecnológico Microsoft. A gran distancia de estos dos, se pueden encontrar otros software de servidor, como uno creado por la multinacional SUN, e incluso el incipiente crecimiento de servidores basados en el software de Google en los últimos tiempos.

Pero desde el punto de vista histórico, resulta más interesante observar la gráfica en la siguiente figura con la cuota de mercado que cada servidor ha tenido desde que se toman referencias para la encuesta de NetCraft (en agosto de 1995). Se puede observar que mientras que desde el año 2000 los resultados no varían mucho –el servidor web Apache y Microsoft Information Server son los servidores más utilizados–, esto no era así al principio. En 1994, el servidor web más utilizado era el NCSA, un software creado por una compañía llamada Netscape. Esta compañía, además de tener en catálogo este servidor web, ofrecía un navegador, conocido como Netscape Navigator. Mientras que el navegador podía descargarse de manera gratuita, había que pagar por el software de servidor web. Se trataba de una estrategia encaminada en ampliar el mercado de personas que navegaban por Internet (ofreciendo su navegador gratis), pero cobrando a los que ofrecían información y servicios (vendiendo el servidor).

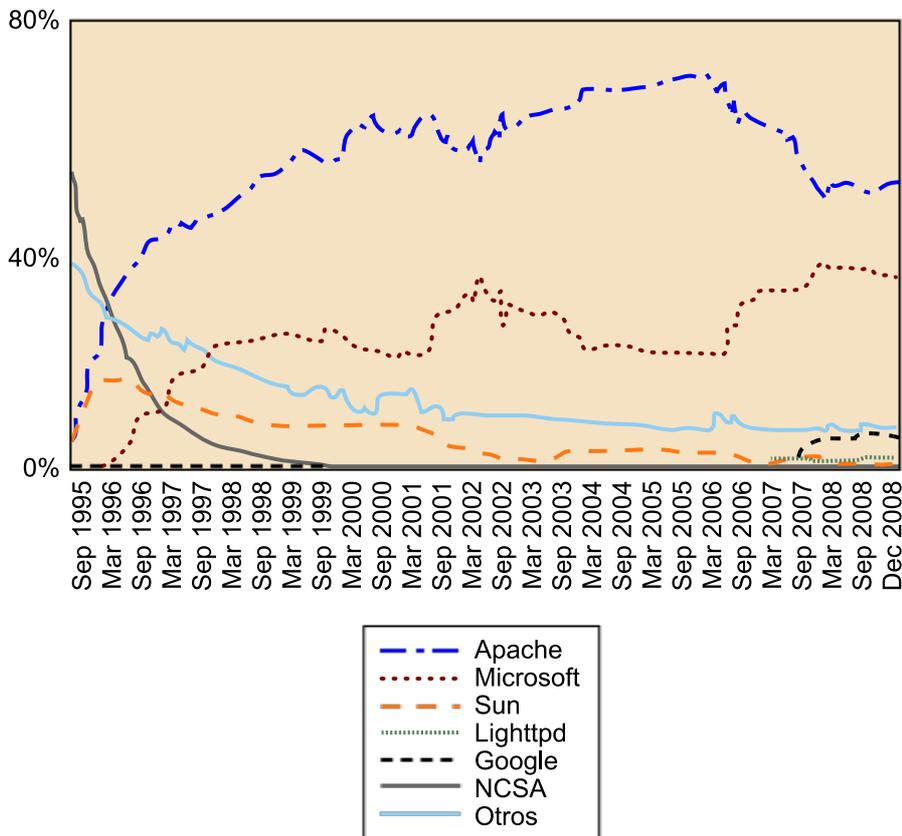
NetCraft

NetCraft es una empresa que se dedica a sondear el software que se utiliza como servidor web y que realiza un informe mensual publicado en la web.

Ved también

El software libre se describirá en el módulo "La cultura libre".

Cuota de mercado de los servidores web más utilizados en todos los dominios de Internet



(c) Netcraft LTD. Fuente: Netcraft LTD

La figura anterior muestra cómo Microsoft apostó de manera tardía por la web, ya que su servidor empezó a ser importante bien entrado el año 1997, cuando Apache y SUN ya le habían tomado la delantera.

2.2. El cliente

Por un lado, hemos visto que tenemos el servidor web, que requiere de una serie de condiciones. Por otro, está el **cliente web**, conocido generalmente como el **navegador**. El cliente web es un programa que se comunica con el servidor, para lo que utiliza una serie de reglas que veremos más adelante. Una de las características más importantes del modelo cliente-servidor es que en todo momento la comunicación sólo existe entre un cliente y un servidor, no habiendo nunca comunicación entre clientes. El cliente web es un programa en el nivel de aplicación y su acometido se divide en dos partes:

- Realizar una **petición**, por ejemplo, una página web alojada en el servidor web.
- Interpretar la **respuesta del servidor**. Esto significa que el cliente web no sólo ha de conocer las reglas para comunicarse con el servidor web (el protocolo HTTP), sino que también deberá saber qué hacer con los contenidos que éste le envía (generalmente utilizando el estándar HTML). Gran

parte de la complejidad de los clientes web actuales reside precisamente en mostrar los contenidos de manera conveniente en pantalla.

A diferencia del servidor, no es necesario que los clientes estén continuamente conectados a Internet y pueden tener IPs diferentes (dinámicas), ya que el servidor responde a la dirección de petición.

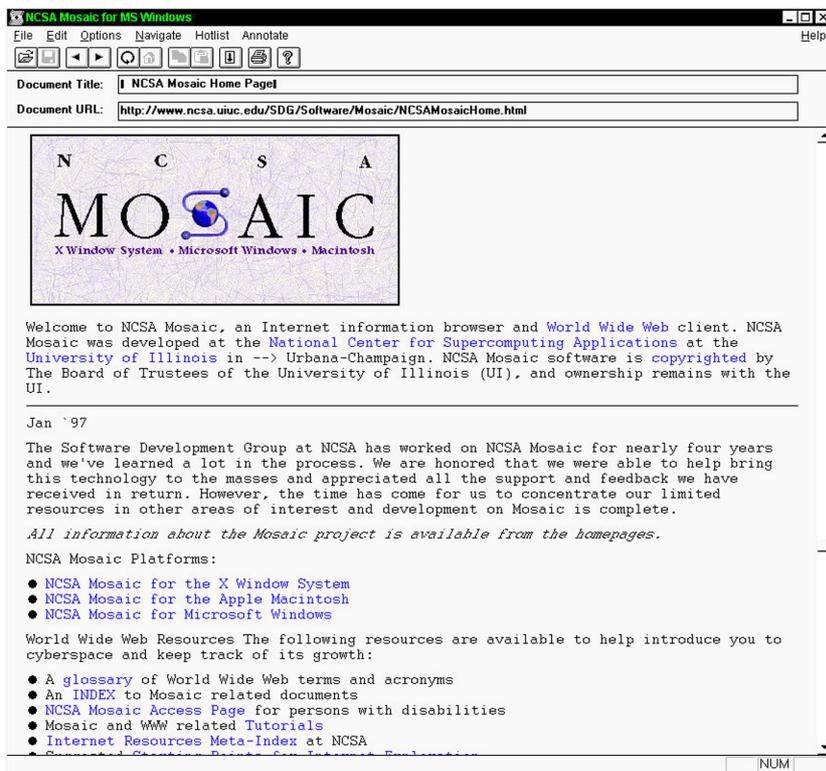
Operadoras de comunicaciones

Las operadoras de comunicaciones (ISP) tienen un conjunto limitado de direcciones IP que asignan dinámicamente cuando uno de sus clientes se conecta a Internet. Por lo tanto, cada vez que un usuario se conecta a Internet cambia su dirección IP.

Ejemplo de IP dinámica

Si pedimos una página web desde nuestro portátil en la oficina, nos la enviará a la IP actual (desde la que hemos realizado la petición). Una vez en casa, si pedimos de nuevo la página, el servidor nos volverá a enviar la información a nuestro portátil, que tendrá una dirección IP diferente.

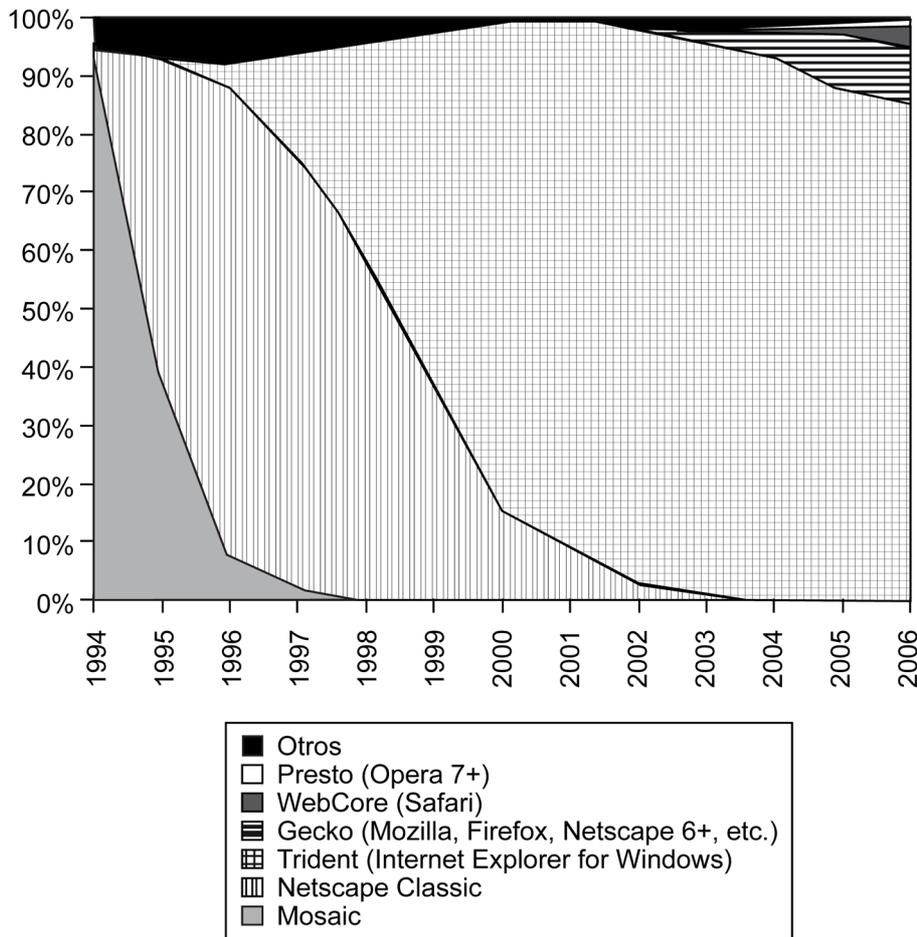
Como ya hemos comentado con anterioridad, el primer cliente de la World Wide Web fue creado por Berners-Lee en el CERN. Sin embargo, el primer navegador realmente popular fue Mosaic, concebido en el National Center for Supercomputing Applications (NCSA) en 1992. Los navegadores actuales, aun siendo más elaborados y completos, tienen gran parecido visual con Mosaic. Tal fue el éxito de Mosaic que sus creadores montaron una *spin-off* llamada Netscape que, como hemos visto con anterioridad, también vendía servidores.



Captura de pantalla del navegador web Mosaic, en su versión 6beta a finales de 1993. Fuente: National Center for Supercomputing Applications/University of Illinois Board of Trustees.

Mientras que el mercado de servidores ha estado fragmentado entre varios competidores, el de navegadores web siempre ha tenido un navegador que ha sido utilizado por la mayoría de los internautas. Hasta 1995 el cliente más utilizado era Mosaic, aunque Netscape Navigator tomó el testigo de manera rápida; al fin y al cabo, eran los continuadores de Mosaic.

Cuota de mercado histórica de los navegadores web más populares desde 1994



Fuente: Wikipedia. Licencia: Dominio Público

Entre 1997 y 1999 se puede observar un período que es conocido como la guerra de los navegadores. Para entonces, el público ya se había acostumbrado a que los navegadores web fueran gratuitos –siguiendo el modelo de negocio de Netscape–, por lo que las compañías que competían en este mercado lo hacían por otras razones. El principal motivo era que la congruencia de todos los servicios de Internet en la web convertía este servicio en un mercado muy apetitoso. Quien consiguiera poner su navegador en el mayor número de ordenadores tendría una posición muy interesante, ya que controlaría la puerta de entrada a la web. Microsoft, aunque tardíamente, se dio cuenta de esto y lanzó un software cliente web llamado Internet Explorer.

La **guerra** se recrudeció debido principalmente a dos movimientos:

- En primer lugar, Microsoft utilizó como plataforma de lanzamiento de su navegador su sistema operativo Windows. Esto significaba que muchos

Ved también

Recordad que el modelo de negocio de Netscape se ha tratado en el subapartado 2.1 de este módulo.

usuarios que compraban un ordenador y se instalaban el sistema operativo –o les venía preinstalado– ya tenían un navegador disponible. Por esta razón, se llevó a Microsoft a juicio por abuso de posición dominante. Cuando los jueces fallaron, Netscape había desaparecido –fue adquirido por AOL– y el desarrollo de su navegador se paralizó. Curiosamente, en 1997, en vista de que su cuota de mercado menguaba drásticamente, Netscape decidió liberar su código, ofreciéndolo bajo una licencia de software libre. De esta manera nació el proyecto Mozilla y de ahí el navegador Firefox (Gecko), que aparece en los últimos años como una alternativa a Microsoft Explorer.

- En segundo lugar, Netscape y Microsoft se lanzaron en una carrera por innovar para ofrecer mejores páginas web, más allá del estándar HTML promulgado por el W3C, el organismo encargado de estandarizar HTML. Lo que en un principio puede resultar una situación beneficiosa de innovación, no lo fue tanto, ya que se daba el caso de que los diseñadores web creaban páginas web con información que sólo podía ser vista correctamente por uno de los navegadores. Aunque las dos compañías fueron muy proclives a hacer este tipo de modificaciones propias, en cuanto Microsoft ganó una cuota de mercado importante acentuó estas prácticas. De esta manera, se aprovechaba del llamado efecto red. Un diseñador web tenía muchos incentivos para crear páginas web que llegaran al máximo número de usuarios. Al mismo tiempo, los usuarios utilizaban el navegador web con el que pudieran visualizar correctamente más páginas web. Por el efecto red, el navegador utilizado fue Internet Explorer, ya que la mayoría de los usuarios lo utilizaban.

El efecto red

El **efecto red** se da cuando un bien o servicio tiene mayor valor cuanto más gente lo utilice. En otras palabras, el número de personas que han adoptado antes un servicio es un factor del valor disponible para el próximo adoptante. Un ejemplo clásico es el servicio telefónico: si poca gente tiene teléfono, es menos interesante tener uno que si todo el mundo tiene uno.

3. El protocolo HTTP

Una vez conocidos los lados cliente y servidor, vamos a conocer cómo se intercambia información. Para ello, se utiliza el protocolo HTTP creado por Tim Berners-Lee.

El protocolo utilizado en el servicio web tiene como nombre **protocolo de transferencia de hipertexto**⁴(HTTP). Las primeras versiones del protocolo fueron diseñadas por Tim Berners-Lee, pero después ha pasado a ser coordinado por el World Wide Web Consortium (W3C) y el Internet Engineering Task Force (IETF). El protocolo viene especificado por un RFC. El RFC de la versión actual de HTTP, la versión 1.1, es el RFC 2616 y fue publicado en junio de 1999.

El protocolo HTTP es un protocolo basado en texto, por lo que puede ser entendido fácilmente por humanos. Existen dos tipos de mensajes HTTP:

- Los mensajes de **petición**, realizados por el cliente al servidor.
- Los mensajes de **respuesta**, enviados del servidor al cliente.

Un mensaje de petición contiene los siguientes elementos:

- La línea con la petición, en la que se especifica un **método** y un **objeto**. El método es la acción que se debe realizar, mientras que el objeto es, como su propio nombre indica, el objeto de la acción. Existen varios métodos de peticiones diferentes como se muestra en la tabla siguiente. Entre los métodos HTTP podemos distinguir métodos seguros e inseguros. Los métodos seguros reciben ese nombre porque no cambian el estado del servidor y sólo realizan peticiones de información. Son HEAD, GET, OPTIONS, CONNECT y TRACE. El resto de métodos, POST, PUT y DELETE, sí cambian el estado del servidor o realizan acciones de envío de información al servidor. Evidentemente, con estos últimos métodos hay que tener más cuidado en el servidor, ya que pueden poner en riesgo su seguridad.
- Ninguna, una o varias **cabeceras**. Las cabeceras permiten enviar información adicional sobre la petición, como el navegador utilizado o el lenguaje preferente. Esta información puede ser útil para el servidor. Por ejemplo, si el servidor ofrece los contenidos en varios idiomas, puede ver cómo hemos configurado nuestro cliente a partir de las cabeceras y mandar la información en el idioma seleccionado.
- Una línea en blanco.
- El cuerpo del mensaje opcional. Aquí irán los contenidos que enviamos al servidor. Generalmente, una petición de página web no tiene conteni-

Ved también

Recordad que los protocolos de comunicaciones se explicaron en el módulo "Aspectos tecnológicos de las redes e Internet".

⁽⁴⁾En inglés *Hypertext Transfer Protocol*.

Los RFC

Recordad que los RFC (*Request For Comments*) son acuerdos de la IETF con relación a un protocolo o estándar.

dos. Pero esto no es el caso cuando enviamos un formulario o subimos un archivo.

Métodos HTTP más importantes utilizados para realizar peticiones a un servidor web

| Método HTTP | Uso |
|-------------|--|
| GET | Pide al servidor un recurso (página web, imagen, archivo de sonido, etc.). |
| POST | Envía datos al servidor para ser procesados. |
| HEAD | Pide un recurso (página web, imagen, archivo de audio, etc.), pero sin que se le envíe el cuerpo de la respuesta, de manera que sólo recibirá las cabeceras. |
| PUT | Manda el recurso (página web, imagen, archivo de sonido, etc.) dado por la dirección al servidor. |
| DELETE | Borra el recurso. |
| OPTIONS | Pide al servidor los métodos que pueden ser utilizados con el recurso. |
| TRACE | Método utilizado para realizar pruebas. |
| CONNECT | Es un método reservado para realizar túneles en proxies. |

Los mensajes de respuesta enviados por el servidor tienen la misma estructura que los de petición, aunque difieren en la primera línea. Así, en vez de enviar un método y un objeto, el servidor devuelve un código numérico indicando el **estado de la petición**. Al código numérico se le añade una breve descripción textual sobre su significado. En la siguiente tabla se muestran algunos de los códigos de respuesta HTTP más comunes, su descripción y su significado.

Códigos HTTP de respuesta (y su descripción) más utilizados

| Código de estado | Descripción | Significado |
|------------------|-------------------|--|
| 200 | OK | La petición ha sido realizada con éxito. |
| 301 | Moved permanently | El recurso pedido tiene una nueva dirección permanente. |
| 302 | Found | El recurso pedido tiene una nueva dirección temporal. |
| 403 | Forbidden | El servidor ha entendido la petición, pero no la ejecutará. |
| 404 | Not Found | El servidor no ha encontrado nada por esa dirección. |
| 500 | Server Error | El servidor ha encontrado un error interno que le impide procesar la petición. |

Una de las ventajas del protocolo HTTP es que al ser texto plano, podemos leerlo y entenderlo con facilidad. Por desgracia, no sólo el cliente y el servidor lo podrán hacer, sino que cualquiera en un *router* intermedio de Internet podría hacer lo mismo sin muchos problemas. Para ciertos usos, la facilidad de lectura se convierte en una desventaja. Podríamos estar interesados en que

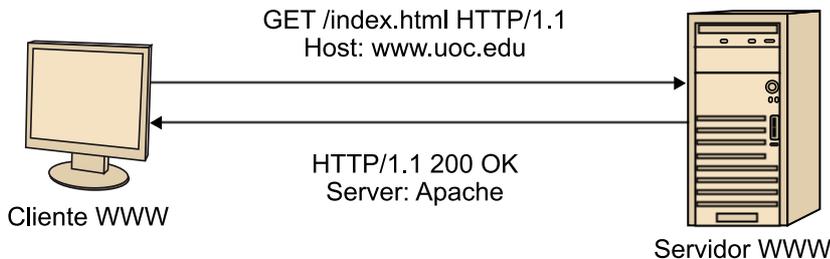
nadie leyese nuestras peticiones o las respuestas del servidor, como al realizar una transferencia bancaria, pagar con nuestra tarjeta de crédito o mirar nuestros datos fiscales en la web del ministerio correspondiente.

Para solucionar este problema, se propuso el **protocolo HTTPS** (HTTP seguro), estandarizado en el RFC 2818, que es una combinación del protocolo HTTP con técnicas de cifrado. Por ello, cuando accedemos a la página en línea de nuestro banco, lo hacemos pidiendo una página cuya dirección web empieza por "https" y no "http". Mediante este mecanismo, no sólo se asegura la confidencialidad de la transmisión, sino que también se evita la suplantación.

3.1. Un ejemplo de comunicación con HTTP

Una petición HTTP de la página principal de la página de la UOC podría ser tal y como se muestra de manera gráfica (y resumida) en la figura siguiente.

Ejemplo de intercambio de mensajes HTTP entre un cliente un servidor



La petición completa que envía el cliente al servidor web de la UOC es la siguiente:

```
GET /index.html HTTP/1.1
Host: www.uoc.edu
User-Agent: Mozilla/5.0 (X11; Linux i686; en-US; rv:1.9.0.5)
```

Se puede observar que en la petición aparece información sobre el cliente que estamos utilizando (en el caso del ejemplo se trata del navegador Mozilla Firefox 1.9.0.5, que utiliza el motor Mozilla 5.0), el sistema operativo (basado en Linux), el entorno gráfico (en este caso, X11, un interfaz de usuario gráfico para Linux) e incluso el idioma (en-US) en el que tenemos configurado el navegador.

La respuesta del servidor HTTP de la UOC a nuestra petición anterior sería la siguiente:

```
HTTP/1.1 200 OK
Server: Apache
Last Modified: Fri, 14 Dec 2007 10:16:44 GMT
Content-type: text/html
Content-length: 1371
```

Ved también

Recordad que en el subapartado 4.1 del módulo "Seguridad en la sociedad de la información" se habla de identidad digital.

Las direcciones web

Las direcciones web son también conocidas como URL o *Uniform Resource Locator*, localizador uniforme de recurso.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"  
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">  
<html xmlns="http://www.w3.org/1999/xhtml" lang="ca">  
<head>  
[...]
```

Se puede observar que en la primera línea se especifica tanto el protocolo (y su versión) como el código de respuesta (200 OK). A continuación, en líneas sucesivas y hasta una línea en blanco, recibimos información mediante cabeceras. Se puede comprobar qué software de servidor se está utilizando (Apache), la fecha de la última modificación del documento, el tipo de contenido que se está enviando (en este caso, una página de texto en lenguaje HTML) y, finalmente, la longitud de la página en bytes (1.371 bytes). Luego viene el contenido, que es una página web en lenguaje HTML. Por su excesiva longitud, en la figura se muestra sólo las cabeceras de la página web.

4. El lenguaje HTML. Nociones básicas

Como ya se ha comentado con anterioridad, uno de los aspectos fundamentales de la popularidad de la web es el concepto de hipertexto. El texto plano es una buena manera de contener información, pero adolece de elementos estructurales y de apariencia. Hasta cierto punto, nos gustaría añadir además información sobre el texto en sí y referencias a otros textos. En este apartado veremos cómo conseguirlo con el lenguaje de marcado HTML.

Berners-Lee pensó que la manera de conseguir esa riqueza adicional en texto plano es añadiéndole **etiquetas**. Si queremos que un texto aparezca en negrita, debemos crear un **elemento HTML** del siguiente modo:

Esta es una frase con `texto en negrita`.

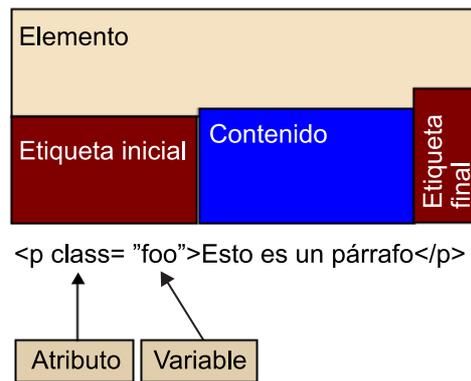
Nótese que lo que estamos añadiendo es información al texto plano para que el navegador la interprete de manera conveniente. Y lo más interesante es que esta información que estamos añadiendo, ¡la estamos añadiendo también en texto plano! Los elementos HTML vienen dados por etiquetas de principio y de final. El hecho de que las etiquetas se encuentren entre signos de menor y mayor es precisamente para que el navegador sepa distinguir entre la información que queremos representar y la información sobre cómo la queremos representar.

Las etiquetas y la manera de utilizarlas vienen especificadas en un estándar (en realidad, varios estándares, porque ha habido varias versiones) y han sido publicados como un estándar del W3C. Este lenguaje recibe el nombre **lenguaje de marcado de hipertexto**⁵(HTML).

⁽⁵⁾En inglés, *Hypertext Markup Language*.

En general, los **elementos HTML** se forman de la siguiente manera:

- Una etiqueta marcando el principio de un elemento.
- Opcionalmente, se aceptan **atributos** con sus valores (variables) dentro de la etiqueta.
- El contenido del elemento (que puede ser texto y/u otros elementos).
- Una etiqueta final.



Estructura general de un elemento en HTML. Fuente: Wikipedia.
Licencia: Dominio Público.

Se da el caso de elementos que son vacíos, por lo que no precisan de contenido. El ejemplo más conocido es el del salto de línea (`
`). En tales casos, la etiqueta final no es necesaria. En la siguiente tabla se pueden ver algunas de las etiquetas HTML más utilizadas.

Etiquetas HTML más frecuentes

| Etiqueta | Significado/uso |
|------------------------------|---|
| <code>html</code> | Inicio de una página HTML. |
| <code>title</code> | Título de la página HTML. |
| <code>head</code> | Cabecera de la página HTML. Aquí se incluye metainformación sobre la página, como el título, autoría, palabras clave, etc. |
| <code>meta</code> | Metainformación sobre la página, generalmente incluida en la cabecera. |
| <code>body</code> | Cuerpo de la página. Es donde viene el contenido de ésta. |
| <code>p</code> | Etiqueta de párrafo. |
| <code>br</code> | Etiqueta de salto de línea. |
| <code>b</code> | Etiqueta de texto en negrita. |
| <code>i</code> | Etiqueta de texto en itálica. |
| <code>hr</code> | Línea horizontal. |
| <code>a</code> | Etiqueta de enlace. Cuando se trata de un enlace a otra página, requiere al menos un atributo, <code>href</code> , que indica la dirección web de destino. |
| <code>center</code> | Etiqueta para centrar contenido. |
| <code>img</code> | Etiqueta de imagen. Requiere al menos dos atributos: <code>src</code> para indicar la dirección de la imagen y <code>alt</code> para mostrar texto alternativo en el caso de que no se pueda mostrar la imagen. |
| <code>table</code> | Etiqueta de tabla. |
| <code>td</code> | Etiqueta de elemento de una tabla. |
| <code>tr</code> | Etiqueta de fila de una tabla. |
| <code>hn</code> | Etiqueta de encabezado. El n da el nivel de encabezamiento y va desde 1 (más grande) a 6 (más pequeño). |

4.1. Un ejemplo de página web

A continuación, se muestra un ejemplo del código de una sencilla página web (creada a partir de las etiquetas que se muestran en la tabla del apartado 4). Se ha utilizado el espaciado horizontal para que se pueda determinar visualmente qué elementos son contenidos en otros.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN"
"http://www.w3.org/TR/1998/REC-html40-19980424/loose.dtd">

<html>
  <head>
    <title>Página de ejemplo: Fundamentos Tecnológicos</title>
    <meta http-equiv="Content-Type" content="text/html; charset=utf-8">
  </head>

  <body>
    <h2>Bienvenidos a mi primera página web</h2>

    <p>Ésta es la <b>primera página web</b> para la que leo el código.
      Y lo hago porque estoy cursando un máster en la
      <a href="http://www.uoc.edu">Universitat Oberta de Catalunya</a>.
    </p>

    <hr>

    <p>A continuación, vamos a introducir como imagen el logo de la W3C, el
      organismo encargado de estandarizar el HTML. Para ello, vamos a hacer
      referencia al logo que podemos encontrar en su página principal
      (tomando su dirección, o URL). Aquí va:</p>

    <hr>

    <p>Pero podríamos ir más allá y combinar varias etiquetas, por ejemplo,
      para conseguir que la imagen fuera a la vez un enlace y estuviera
      centrada. ¿Cómo? Véase a continuación:</p>

    <center>
      <a href="http://www.w3c.org">
        
      </a>
    </center>

    <hr>
```

```
<p>Vamos a probar a introducir una <i>tabla</i> en HTML.</p>

<table border="1">
  <tr>
    <td>Elemento 1.1</td>
    <td>Elemento 1.2</td>
    <td>Elemento 1.3</td>
  </tr>
  <tr>
    <td>Elemento 2.1</td>
    <td>Elemento 2.2</td>
    <td>Elemento 2.3</td>
  </tr>
</table>

</body>

</html>
```

La primera línea indica el tipo de documento del que se trata: un documento que sigue el estándar "HTML versión 4.0 Transitional"; asimismo, se adjunta la dirección de la especificación técnica del estándar.

El resto del documento empieza y termina con una etiqueta `html` que indica que se trata de un documento HTML y que contienen el resto de elementos. Así, si `html` es la etiqueta raíz (o de primer nivel), podemos observar que en el segundo nivel contamos con dos etiquetas: `head` y `body`.

En la cabecera, dentro de `head`, podemos encontrar información sobre la página web en sí; en nuestro ejemplo, simplemente hemos indicado el título de la página web utilizando la etiqueta `title`. Además, mediante una etiqueta `meta`, se informa de que el tipo de contenido es una página de texto, en particular HTML, y que el tipo de codificación de caracteres es UTF-8. Si quisiéramos se podría añadir mucha más información, como viene indicado en el estándar HTML: autor de la página web, palabras clave, etc.

En el cuerpo, entre las etiquetas `body`, encontramos el contenido de la página web. En nuestro caso, empezamos con un texto que queremos que sea más grande de lo habitual, por lo que utilizamos la etiqueta `h2`. Recordemos que si quisiéramos que este texto fuera todavía más grande, deberíamos utilizar en su lugar la etiqueta `h1`.

Justo a continuación viene el primer párrafo (que se puede identificar fácilmente al estar entre etiquetas `p`). Este primer párrafo contiene otras etiquetas: primera página web aparecerá en negrita (**bold**) al encontrarse etiquetado como `b`, mientras que `Universitat Oberta de Catalunya` es un enlace

a la página principal de la UOC. Los enlaces vienen dados por la etiqueta `a` y requieren de un parámetro adicional, en este caso `href`, para saber cuál es la dirección de destino.

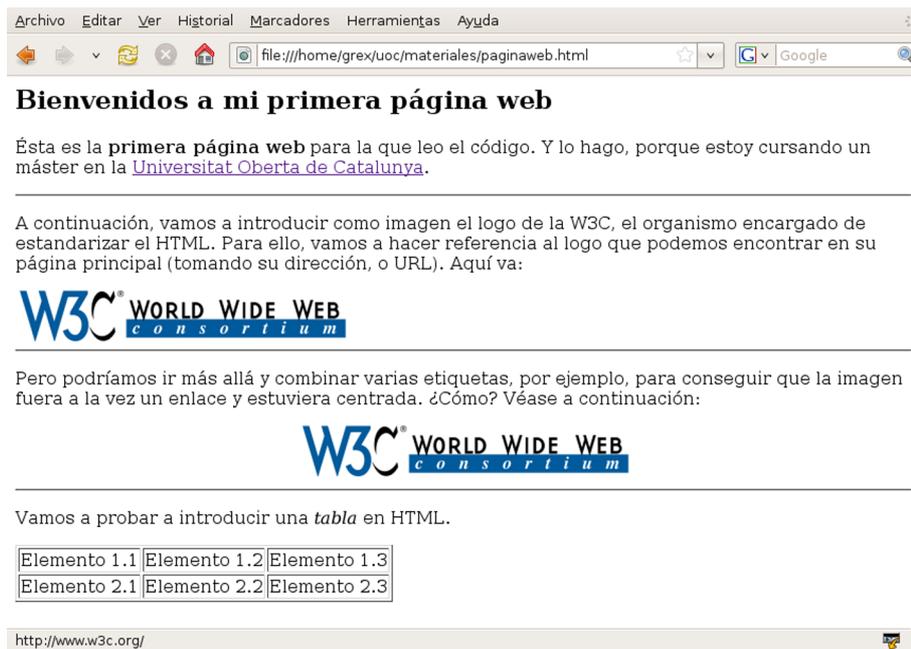
Algo parecido ocurre en el caso de querer insertar una imagen, tal y como hacemos en el siguiente párrafo. Para incluir imágenes, la etiqueta que debemos utilizar es `img` y el número de parámetros necesarios son dos: la dirección de la imagen (`http://www.w3.org/Icons/w3c_main`) y un texto alternativo (Logo W3C), utilizando el atributo `alt`. Este texto aparecerá en caso de que no se encuentre la imagen, de que el navegador no sea un navegador gráfico o incluso de que sea un navegador para invidentes (en cuyo caso se leerá este texto alternativo por los altavoces).

Se pueden encadenar varias etiquetas para conseguir el efecto agregado de todas ellas. Así, el siguiente párrafo contiene una imagen centrada, que es a su vez un enlace a la página web del W3C. Ya conocemos todos los elementos de este enlace, salvo el atributo `border`, que está puesto a 0. Si no se especifica este atributo, a las imágenes se les añade por defecto un borde de color azul para que visualmente se pueda reconocer que se trata de una imagen que enlaza otra página.

Finalmente, se muestra cómo incluir una tabla en HTML. La etiqueta `table` es la que indica que vamos a construir una tabla, y le pasamos como parámetro `border` con valor a 1 para que se puedan ver los bordes de la tabla. La tabla tiene dos filas (dos etiquetas `tr`) de tres columnas (por eso hay tres etiquetas `td` en cada fila).

Para crear esta página web, sólo es necesario copiar y pegar el texto en un editor de texto (como el "bloc de notas" o el "Wordpad") y guardarlo en el disco con el nombre `miprimeraweb.html`. Haciendo un doble clic sobre el fichero se abrirá la página web en el navegador. En la siguiente figura, se muestra cómo es interpretado por un navegador web.

Visualización con el navegador de la página web que hemos creado



En el momento de la captura, el puntero del ratón (que no aparece en la captura) estaba situado sobre la segunda imagen. Esta imagen, centrada, era además un enlace a la página principal del W3C. Por ello, al pie de la ventana del navegador, se puede ver la dirección a la que apunta: <http://www.w3c.org>. Si hacemos clic con el ratón, iremos a parar a esa página.

El lenguaje HTML es un estándar que tiene varias versiones. Una vez que hemos terminado de hacer nuestra página web, podemos ir al servicio de validación web del W3C, que nos indicará si hemos hecho un uso correcto del estándar. En caso contrario, nos dirá los errores que hemos cometido, por ejemplo, si falta algún atributo o si no hemos cerrado alguna etiqueta.

Servicio de validación del W3C

El servicio de validación de la W3C se puede encontrar en <http://validator.w3.org/>.

5. La web dinámica o interactiva y aplicaciones web

Hasta ahora, y así fue en los primeros años de existencia del servicio web, hemos considerado que un servidor web ofrece una serie de documentos que los navegadores pueden pedir mediante el protocolo HTTP. Estos documentos son unos documentos de texto enriquecidos con etiquetas HTML que hacen no sólo que podamos darle estructura y mostrarlos visualmente de una manera muy rica, sino que también permitían estar enlazados entre sí. Esto nos permitía pasar fácilmente de uno a otro, y se ha venido a llamar "navegar" por la web.

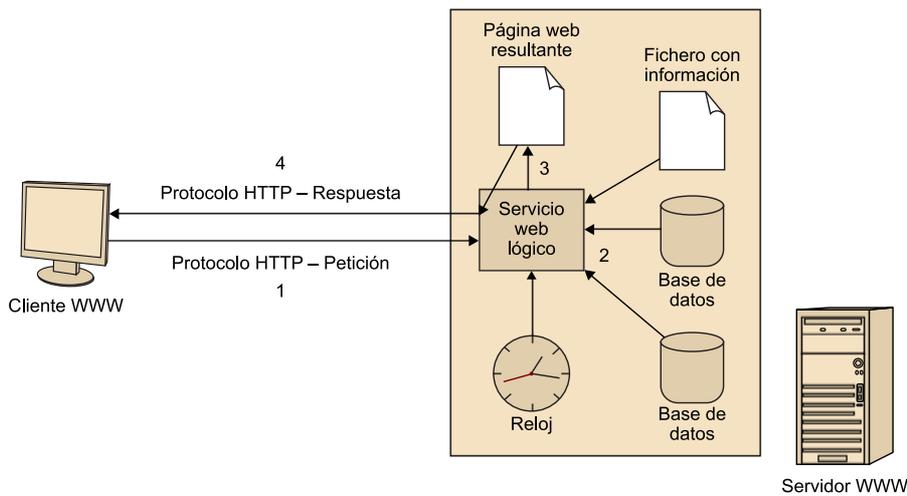
Sin embargo, pronto muchos vieron que este sistema, aun siendo muy simple y de gran provecho, tenía sus limitaciones. Una de las limitaciones más importantes era que la interactividad se reducía precisamente al hecho de navegar por páginas en los servidores. En esta situación, la información ya se encontraba de antemano en el servidor y no se podía modificar, añadir o corregir de manera sencilla. Y otra limitación es que esta información es estática, que no interactúa con el usuario. En otras palabras, en la web que hemos presentado hasta ahora los servidores son los contenedores de información y los clientes los que la piden. El usuario era un mero consumidor de información. Por otro lado, esta información no se genera con la petición del cliente, sino que debía existir con anterioridad en el servidor.

En este apartado estudiaremos cómo la web evolucionó de un marco estático, como el descrito, a otro dinámico, en el que el servidor interactúa de manera mucho más estrecha con el cliente. Ofreceremos una perspectiva de cómo se consigue esta interacción individualizada con el usuario y veremos que de esta tendencia nacieron las aplicaciones web.

El caso es que pronto muchos se dieron cuenta que el software servidor podría hacer más cosas que simplemente identificar el documento que se ha pedido y enviarlo al cliente. Así, por ejemplo, se podría crear ese documento dinámicamente o "al vuelo" y enviarlo al cliente. O se podría construir una página web a partir de datos almacenados en una o varias bases de datos. O muchas otras posibilidades, ya que el software de servidor puede realizar una gran cantidad de tareas. Esto da una enorme versatilidad a la web.

La siguiente figura intenta mostrar esto de manera comprensible.

Esquema de un servicio web que ofrece contenidos generados dinámicamente de varias fuentes de información



Por un lado, a la izquierda, tenemos al cliente, que es el que realiza peticiones; en este caso, una página web dinámica. En el otro lado, a la derecha, se encuentra el servidor web. Al recibir la petición HTTP del cliente (paso 1), el servidor web lógico (que es un programa que se ejecuta dentro del servidor físico y que hemos representado por un rectángulo) toma información posiblemente de varias fuentes de información (paso 2). Así, se podría dar el caso de que esas fuentes de información fueran una o más bases de datos, archivos en el disco duro e incluso información propia de ejecución (como, por ejemplo, la hora). Esta información se combina con etiquetas HTML de la manera que se desee para crear una nueva página web (paso 3). Finalmente, la página web es enviada de vuelta por parte del servidor web lógico mediante un mensaje de respuesta HTTP (paso 4). El aspecto crucial de todo este proceso es que la página web resultante estará codificada en el lenguaje HTML que entienden los navegadores, y que no se encuentra almacenada como tal en el sistema de archivos del servidor. Mediante esta simple idea, hemos creado un modo de obtener contenidos dinámicos que interactúen con el cliente. Un mundo nuevo de posibilidades se abre ante nuestro navegador.

Precisamente, el hecho de que la web sea considerada como la **puerta de entrada a Internet** y que muchos confundan Internet con la web es debido a esta versatilidad. Al fin y al cabo, si el servidor web se conecta al servidor de correo electrónico y ofrece sus contenidos al usuario mediante páginas web, lo que estamos haciendo es acceder a un servicio de Internet clásico como el correo electrónico mediante la web. El usuario no tiene ni siquiera noción de que está utilizando otro servicio; para él sigue siendo la web. Si a esto le añadimos que "usar" la web es muy sencillo, entenderemos que la **convergencia hacia la web** es un hecho natural, lo que hace caer en el olvido el resto de servicios de Internet.

Las **aplicaciones web** son aquellas aplicaciones a las que los usuarios pueden acceder mediante el navegador web.

Hemos añadido tanta complejidad al programa servidor web que éste se puede asemejar a una aplicación como las que tenemos instaladas en nuestro propio disco duro. Con la salvedad, evidentemente, de que una aplicación web se encuentra almacenada y se ejecuta en la máquina servidor. La manera de interactuar con la aplicación web es mediante páginas HTML que visualizaremos en nuestro navegador. Esto constituye una gran ventaja, ya que el interfaz del navegador y el concepto de hiperenlaces es conocido por el usuario. Además, los creadores de una aplicación web se pueden olvidar de los problemas de instalación en los ordenadores de sus usuarios. No es necesario detectar qué sistema operativo utiliza el usuario en su ordenador personal, ya que con que éste tenga un navegador es suficiente.

Ejemplos de aplicaciones web

Existen muchas aplicaciones web, la mayoría muy populares. Entre las aplicaciones web más utilizadas podemos contar con:

- Servicios de correo-web (o *webmail*): ofrecen acceso a nuestro buzón de correo electrónico mediante el navegador. Los servicios modernos de correo-web ofrecen gran capacidad de almacenamiento, libreta de direcciones, filtros frente al correo no deseado (SPAM), etc.
- Periódicos en línea: las páginas web de los periódicos modernos en línea rara vez existen como tal en el disco duro de un servidor. Generalmente se crean "al vuelo" incluyendo cabeceras, la noticia o bloques laterales (noticias relacionadas, noticias más leídas, etc.).
- Centrales de reserva de viajes: permiten acceder a las grandes bases de datos de compañías aéreas y promotores de viajes, de manera que se pueda organizar un viaje desde el ordenador de casa.

Más ejemplos de aplicaciones web lo constituyen los buscadores, las páginas wiki, los blogs, los sitios de redes sociales, etc.

5.1. Buscadores

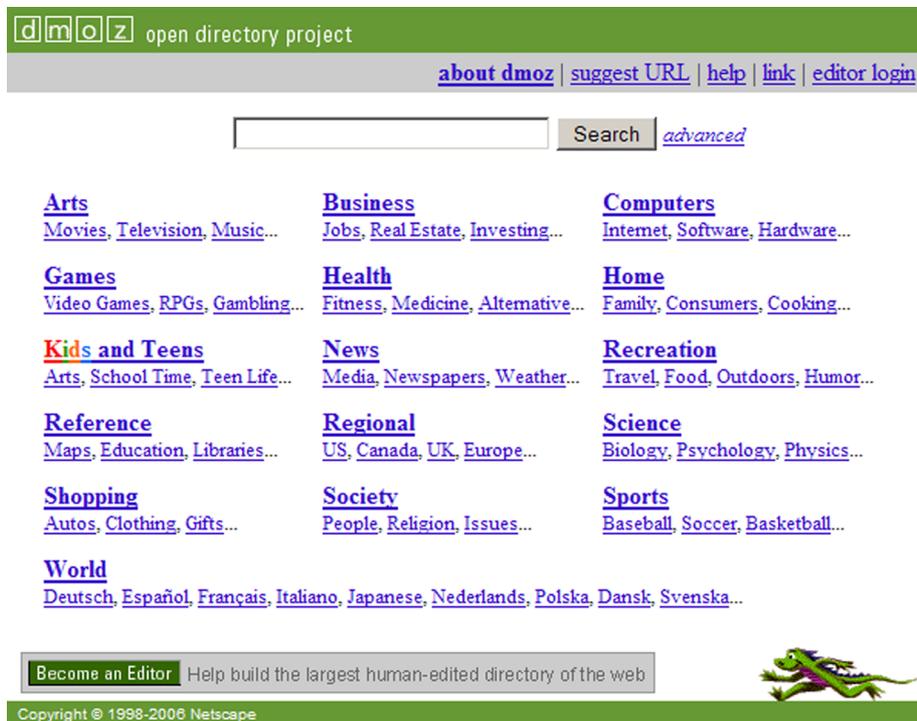
Las aplicaciones web de busca de páginas web son un servicio indispensable en la actualidad, pero no ha sido siempre así. En este subapartado, además de la evolución histórica de los buscadores, introduciremos brevemente cómo funcionan.

Los **buscadores** son aplicaciones web que nos permiten encontrar información en Internet de manera rápida. Mediante la introducción de una o varias palabras clave, nos devuelve una página índice que enlaza páginas web con información relacionada.

Con el auge de la web y toda la información que había en ella, también surgió el problema de cómo encontrar la información deseada. Los primeros esfuerzos, allá por 1993, se dedicaron a crear páginas web que enlazaban con otras páginas web, generalmente de un mismo tema.

Así, surgieron los primeros **índices temáticos**, sistemas de búsqueda por temas como Lycos (en 1994) o Yahoo! (en 1995). Estos sitios tenían una jerarquía de temas por los que se podía navegar y que contenían páginas web o subtemas. Los índices temáticos contaban con personas que se encargaban de identificar, clasificar y listar las páginas web según temas o categorías. Se trataba de una tarea laboriosa y con sus limitaciones, ya que era realizada por humanos, aunque –precisamente por ser revisada– eran de gran calidad. Los índices temáticos tenían, además de su índice navegable por temas o categorías, un formulario que permitía realizar búsquedas dentro sus propias páginas. Si buscabas "fútbol", te llevaba a la página del índice temático sobre fútbol con enlaces sobre el tema; todo lo que no estaba incluido en el índice no era accesible.

En la actualidad, los índices temáticos han pasado a un segundo plano por cuestiones tecnológicas, ya que se han conseguido formas automáticas y eficientes de clasificar y mostrar la información. Aun así, existe un proyecto llamado **Open Directory Project** (ODP o DMOZ de directory.mozilla.org; en castellano se podría traducir como Proyecto de Directorio Abierto) que ofrece este servicio con la colaboración de una comunidad de editores voluntarios.



[d](#) [m](#) [o](#) [z](#) open directory project

[about dmoz](#) | [suggest URL](#) | [help](#) | [link](#) | [editor login](#)

[advanced](#)

Arts
[Movies](#), [Television](#), [Music](#)...

Business
[Jobs](#), [Real Estate](#), [Investing](#)...

Computers
[Internet](#), [Software](#), [Hardware](#)...

Games
[Video Games](#), [RPGs](#), [Gambling](#)...

Health
[Fitness](#), [Medicine](#), [Alternative](#)...

Home
[Family](#), [Consumers](#), [Cooking](#)...

Kids and Teens
[Arts](#), [School Time](#), [Teen Life](#)...

News
[Media](#), [Newspapers](#), [Weather](#)...

Recreation
[Travel](#), [Food](#), [Outdoors](#), [Humor](#)...

Reference
[Maps](#), [Education](#), [Libraries](#)...

Regional
[US](#), [Canada](#), [UK](#), [Europe](#)...

Science
[Biology](#), [Psychology](#), [Physics](#)...

Shopping
[Autos](#), [Clothing](#), [Gifts](#)...

Society
[People](#), [Religion](#), [Issues](#)...

Sports
[Baseball](#), [Soccer](#), [Basketball](#)...

World
[Deutsch](#), [Español](#), [Français](#), [Italiano](#), [Japanese](#), [Nederlands](#), [Polska](#), [Dansk](#), [Svenska](#)...

Help build the largest human-edited directory of the web

Copyright © 1998-2006 Netscape



La segunda generación de sitios de búsquedas, que datan de finales de los años noventa del siglo XX, viene dada por la aparición de los **motores de búsquedas**.

Los motores son programas que trabajan en tres etapas:

- La primera etapa consiste en la descarga de páginas web. Para ello, los motores de búsqueda se sirven de programas que se denominan **arañas**, se descargan las páginas web y siguen los enlaces que encuentran en esas direcciones para seguir descargando otras páginas. Así, una página aparecerá en un motor de búsqueda si, y sólo si, es enlazada por otra página web a la que el motor de búsqueda tenga acceso. Aunque existen páginas web que son visitadas por las arañas con mayor frecuencia, por ejemplo los sitios de noticias o algunos foros, en general, durante un mes se recorren todas las páginas web. Es por ello que si subimos una página web al servidor web no aparece automáticamente en un motor de búsqueda; debemos esperar a que una araña visite nuestra página web para que aparezca. También ha de notarse que, al descargarse la página web, el motor de búsqueda está guardando una copia en sus discos duros. Algunos servicios de motor de búsqueda ofrecen la posibilidad de ver esa copia (conocida como la copia *caché*).
- El segundo paso del proceso consiste en la indexación de la página web. Para ello, el motor de búsqueda utiliza una serie de algoritmos para extraer las **palabras clave**. Las palabras que aparecen en el título de la página suelen tener mayor peso que las que aparecen en el cuerpo. Asimismo, los motores de búsqueda almacenan metainformación sobre la página que pudiera ser de interés (como intentar la identificación del idioma en el que está escrita la página web o si ésta tiene una licencia de uso específica).

El algoritmo de Google

Google tiene un algoritmo llamado PageRank que ofrece una valoración de una página a partir del número de páginas que la enlazan y del PageRank de estas páginas. Así nuestra página personal tendrá mayor PageRank si nos enlaza a muchas páginas con un PageRank alto. El razonamiento que podemos encontrar en este algoritmo es que la importancia de la propia página web va en función de la importancia de las páginas web que la enlazan.

- El último paso tiene como objetivo mostrar la información de las bases de datos, de manera que sea útil para el usuario. Parte de esta tarea consiste en optimizar el tiempo de búsqueda para que la espera no sea muy larga.

Ejemplos de motores de búsqueda

El más conocido de los motores de búsqueda es Google, aunque Yahoo! Search, AllTheWeb o Bing, por ejemplo, funcionan de una manera similar.

6. Web 2.0

Como se ha podido comprobar de la lectura de los últimos apartados, la web ha ido evolucionando desde un repositorio de información estática a principios de los años noventa hasta un servicio interactivo con el usuario, dinámico en su esencia y altamente cambiante. El nuevo escenario se ha venido a denominar **web 2.0**. Es importante entender que aunque el concepto de web 2.0 esté fuertemente vinculado con tecnologías, la web 2.0 en sí no es una tecnología, sino un modo de entender y usar la web. Por lo tanto, a pesar de que el nombre parece indicar que se trata de una nueva versión de la web, esto no es así, ya que no existe ningún RFC o documento similar que lo estandarice ni lo describa. La web 2.0 es, en realidad, una manera nueva de entender la web, mucho más centrada en el usuario final.

Este apartado pretende arrojar un poco de luz sobre todo esto.



Compendio de conceptos y fundamentos asociados con la idea de web 2.0. Fuente: Wikipedia. Licencia: Creative Commons Attribution ShareAlike 2.0.

Ejemplos de webs 2.0

La red social Facebook, el archivo fotográfico Flickr, el servicio de blogs de Blogger, la cartografía de Google Maps, el correo electrónico etiquetado de Gmail, la opinión de los clientes acerca de los libros de Amazon, la realimentación generadora de confianza por parte de vendedores y compradores de eBay o la compartición de marcadores con páginas interesantes de Del.icio.us, son ejemplos de webs 2.0.

Dentro de los conceptos importantes que se atribuyen a la web 2.0 están los siguientes: plataforma, web de lectura y escritura y controlado por los usuarios.

La web 2.0 está formada por una serie de **plataformas tecnológicas**, que no dejan de ser aplicaciones web complejas. Al HTML original se le han añadido tecnologías que permiten introducir dinamicidad en el navegador y peticiones

asíncronas de nuevos datos. Recuérdese que antes, para obtener nuevos datos, había que realizar una nueva petición HTTP al servidor; ahora, existen tecnologías que permiten descargarse nueva información sin la petición explícita del usuario.

En los inicios de la web, la información que existía en el servidor no podía ser modificada, ahora, con la web 2.0 esto no es así. Un ejemplo de ello son los wikis, que permiten precisamente cambiar las propias páginas web, o los blogs, que suelen ofrecer la posibilidad de comentar las historias.

Finalmente, lo importante es el usuario final y no los contenidos. Esto se debe a que lo realmente difícil es captar nuevos usuarios a los sitios de Internet (el efecto red del que ya hemos hablado con anterioridad), mientras que la información interesante o útil se ha visto que es más fácil de crear.

Ejemplo de efecto red

Un ejemplo paradigmático del efecto red viene dado por las redes sociales, en las que la Red gana en valor con cada nuevo usuario que se apunta; al fin y al cabo, uno no se apuntaría a una red social que no utilizara nadie de sus conocidos.

6.1. Páginas wiki

En este subapartado introducimos el wiki, un ejemplo de aplicación web desarrollada dentro de la web 2.0. En el esquema clásico de páginas web almacenadas en el disco duro del servidor, el cliente recibía contenidos estáticos. Mediante las páginas wiki, el cliente puede editar la propia página, modificándola y almacenándola de nuevo en el servidor.

Una **wiki** es una página web que permite ser editada mediante el navegador web.

La historia de las páginas wiki se remonta a 1995, cuando Ward Cunningham produjo la primera versión de un servidor WikiWiki.

Las ventajas de este sistema de edición "al vuelo" son evidentes: cualquiera que visita la página puede editarla, y permitir así la elaboración, la corrección y la mejora en común. Además de permitir la modificación de la página, los wikis permiten –salvo que esta funcionalidad esté inhabilitada o el usuario carezca de permisos para llevarla a cabo– la creación de nuevas páginas wiki. Por lo tanto, existe un mecanismo para que dentro de la página wiki podamos crear un enlace a una nueva página wiki.

WikiWiki

El nombre WikiWiki viene de que en la lengua de las islas Hawaii *wiki* significa rápido.

Las páginas wiki cuentan con lo que se ha venido a denominar **vistas**, que no es otra cosa que varias representaciones de la propia página wiki o de información relacionada. El usuario puede elegir entre las diferentes vistas, utilizando normalmente las pestañas que existen en la parte superior de la página.



Página de la Wikipedia sobre el wiki. Fuente: Wikipedia: Licencia: GFDL y Creative Commons Attribution ShareAlike 3.0.

Las vistas más comunes son:

- La **vista tradicional** es la del artículo, en el que se muestra el contenido de la página.
- La **vista discusión** está pensada para la edición de wikis cooperativos, en particular de enciclopedias. Es un espacio en el que los que aportan contenidos pueden discutir y comentar la página wiki, por ejemplo, sobre la veracidad del contenido, si sigue la guía de estilo, etc.
- En la **vista de código fuente** se puede editar la página wiki. Se trata de un simple formulario en el que podemos introducir el texto que deseemos. Al pulsar el botón de envío, pasaremos a la primera vista y podremos ver de manera inmediata los cambios que hemos realizado.
- Finalmente, contamos con la **vista de historia**. Mediante la historia podemos trazar los cambios que ha sufrido la página, incluyendo información sobre quién los ha realizado, cuándo fueron realizados y qué líneas fueron cambiadas. Además de poder ver cómo ha ido evolucionando la página wiki, permite deshacer cambios y volver a versiones anteriores de la página wiki.

Como se ha comentado con anterioridad, en la vista de edición de una página wiki podemos modificar la propia página. En los inicios, y todavía en algunos wikis, el código de las páginas wiki se escribía en HTML. Sin embargo, pronto se alzaron muchas voces en contra del uso del HTML y a favor, de una manera más sencilla de editar. Esto es debido a que el estándar HTML, aunque amplio, a veces es farragoso y ciertamente complejo para usuarios no técnicos. A su vez, el HTML permite algunas cosas que van en contra de la consistencia visual de la página wiki y, aunque de manera moderada, también introduce algunos riesgos a la seguridad.

En consecuencia, se ideó una sintaxis específica para los wikis, más sencilla de utilizar. El problema es que cada aplicación web de wikis utiliza sus propias convenciones ligeramente diferentes, aunque en los últimos tiempos podemos encontrar algunos esfuerzos por unificar las convenciones, de manera que las reglas sintácticas sean iguales. En la tabla siguiente se pueden ver algunas de las convenciones utilizadas en Wikimedia, la aplicación web utilizada por la Wikipedia. Como se puede observar, aun así existen algunos códigos que son HTML "puro".

Etiquetas wiki más frecuentes

| Estilo | Significado/uso |
|--|-------------------|
| <code>'''negrita'''</code> | negrita |
| <code>''itálica o cursiva''</code> | itálica o cursiva |
| <code><u>subrayado</u></code> | subrayado |
| <code><spanstyle="color:green">verde</code> | verde |
| <code><center>Centrado</center></code> | Centrado |
| <code>[[Enlace]]</code> | Enlace |
| <code>[[Enlace Texto visible]]</code> | Enlace |
| <code>== Sección 1 ==</code> | Sección |
| <code>=== Sección 1.1 ===</code> | Subsección |
| <code>==== Sección 1.1.1 ====</code> | Apartado |
| <code>[[Imagen:Isaac.Asimov02.jpg Isaac Asimov.]]</code> | Imagen |

Fuente: Wikipedia

El wiki más conocido es la enciclopedia **Wikipedia**, iniciada en el 2001 con el objetivo de crear una obra de consulta cooperativa en múltiples idiomas y con contenidos libres.

La Wikipedia funciona como un proyecto autogestionado y tiene una serie de normas consensuadas por la comunidad.

Guía sobre la edición de páginas wiki

Una guía detallada sobre la edición de páginas wiki se puede encontrar en la ayuda correspondiente de la Wikipedia (http://es.wikipedia.org/wiki/Ayuda:Cómo_se_edita_una_página)

Ved también

Veremos lo que son contenidos libres en el módulo "Cultura libre".

Al ser un wiki, cualquiera puede editar en la Wikipedia pasando a ser un **wikipedista**; de hecho, salvo excepciones, lo puede hacer de manera anónima o siendo usuario (o sea, creándose una cuenta). Existe una serie de usuarios, conocidos como **bibliotecarios**, que tienen permisos especiales para borrar artículos o bloquear páginas, pero no tienen mayor poder decisorio que los wikipedistas y deben seguir las normas adoptadas por la comunidad. Asimismo, existen **bots** (diminutivo de robot), programas que realizan tareas tediosas como la creación de páginas de manera automática a partir de datos del censo, de manera que todo pueblo incluido en el censo tenga una página en la Wikipedia.

Se ha debatido mucho sobre la calidad de los contenidos de la Wikipedia, en especial en comparación con las enciclopedias tradicionales elaboradas por especialistas en las materias que se presentan. Algunos estudios han demostrado que la calidad de la Wikipedia en términos generales es similar o mejor que los de la Enciclopedia Británica. Sin embargo, la posibilidad de que cualquiera pueda editar una página de la Wikipedia tiene también sus aspectos negativos. Uno de ellos es la existencia de **vandalismo**, que es como se llama cuando alguien introduce modificaciones incorrectas, inapropiadas u ofensivas. Si el vandalismo se sucede repetidamente, los bibliotecarios tienen la potestad de bloquear la página wiki, evitando que sea modificada.

6.2. Blogs

Un **blog** es un sitio web donde se incluyen **historias** o **entradas**⁶, artículos o textos que pueden incluir imágenes y otros elementos multimedia, en orden cronológicamente inverso.

⁽⁶⁾En inglés, *posts*.

Blog

Hay personas que utilizan el término bitácora en castellano, pero el anglicismo se utiliza hoy en día ampliamente.

El concepto de blog viene de **weblog**, donde *log* significa diario en inglés. El carácter de diario es el que ha dado una cierta informalidad a los contenidos publicados en los blogs. En general los blogs ofrecen la posibilidad de comentar esos contenidos, de manera que los lectores de las historias pueden interactuar con el autor (o autores). Por esta razón, los blogs han tenido amplia aceptación más allá del ámbito personal, siendo utilizados profusamente en periodismo y por instituciones y empresas para comunicarse con sus lectores y clientes. Aunque generalmente los blogs suelen incluir textos, hay blogs de fotos (los **fotoblogs**) o de vídeos (los **videoblogs**).

A día de hoy cualquiera puede abrirse un blog en Internet y llevar su diario personal allí.

Ejemplos de alojamientos blog

Hay muchos sitios que ofrecen alojar blogs de manera gratuita en Internet, como por ejemplo:

- Blogger
- WordPress

- Bitacoras.com
- LiveJournal

Aunque no hay normas sobre cómo escribir en un blog (al fin y al cabo, se supone que se trata de un espacio personal donde uno escribe sobre lo que quiere y cómo quiere), sí hay una serie de recomendaciones que uno debería seguir si quiere tener un blog interesante. Así, es importante darle pausas al lector, utilizando convenientemente párrafos y espacios, y usar listados y elementos gráficos. Un texto compacto y complejo, con párrafos grandes "invita" al lector a no seguir leyendo. La claridad y la simplicidad son reglas básicas de un buen blog.

La longitud de las historias en los blogs ha de ser necesariamente corta, salvo casos especiales como manuales por ejemplo. El tiempo de lectura de la pantalla al que está habituado el lector es muy breve, especialmente en un entorno como el de la web donde hay múltiples enlaces para navegar por páginas sin fin. Por eso es importante ir a la esencia rápidamente. Enlazar otras páginas web es una buena práctica para ser concisos y a la vez originales. No es buena práctica copiar párrafos de otra página web, incluyéndolos en nuestro blog. Lo recomendable es crear un hipervínculo a esa página, para que el lector si así lo desea la visite, y mostrar nuestra postura o comentario en nuestro blog.

La mayoría de los blogs ofrecen la posibilidad de utilizar el lenguaje HTML a la hora de elaborar historias o comentarios, permitiendo utilizar al menos un subconjunto de las etiquetas del estándar. Finalmente, aún siguiendo los anteriores consejos, el mayor hándicap de los blogs –como diarios que son–, es que requieren actualizarse con cierta frecuencia para ser interesantes.

Generalmente la página que vemos del blog es la unión de los contenidos creados por el autor, a los que se les da forma para que tengan un aspecto visual atractivo y práctico. Véase el caso de la figura siguiente, que muestra la página principal de un conocido blog.

Página principal de un conocido blog sobre tecnología y sistemas de información



Fuente: <http://www.enriquedans.com>. Licencia: Creative Commons Reconocimiento 2.5 España.

En aras de una mejor visualización, los títulos de las historias se encuentran en otro tamaño y color, incluyen una imagen con el nombre del blog en la cabecera de la página, e información adicional en el pie, entre otras muchas cosas. En un lateral de la página, podemos ver que la mayoría de los blogs cuentan con alguna manera de navegar por historias pasadas, de buscar en los contenidos del mismo o de enlaces a otros blogs que el autor considera interesante.

Uno de los aspectos más interesantes que ofrecen los blogs es la posibilidad de **exportar sus contenidos**. Todos los elementos decorativos son eliminados a la hora de exportar los contenidos de un blog. Recordemos que los contenidos no dejan de ser el texto (generalmente en HTML) escrito por autor para escribir una historia y algunos metadatos, como pudiera ser la dirección de la historia, el autor, etc.

Para facilitar el intercambio de historias, se toman los contenidos y se ofrecen de manera pública siguiendo un formato estandarizado. Hay varios de estos formatos, como **RSS** o **Atom**. La mayoría de los blogs ofrecen un enlace a una página con sus contenidos en estos formato.

En la anterior figura, el enlace está arriba a la derecha como "RSS" junto a un icono cuadrado de color naranja (el logotipo de RSS). Si el lector lo visita, podrá observar que se trata de una página de texto plano con etiquetas similares, pero ligeramente diferentes, a las que conocemos del HTML, como puede verse en la figura siguiente.

RSS del blog de Enrique Dans sobre tecnologías y sistemas de información

```

Archivo  Editar  Ver  Ayuda

<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" media="screen" href="/-d/styles/rss2spanishfull.xsl"?><?xml-stylesheet type="text/xsl" media="print" href="/-d/styles/rss2spanishprint.xsl"?>
<channel>
  <title>El Blog de Enrique Dans</title>
  <link>http://www.enriquedans.com</link>
  <description>Investigación y opinión acerca de los Sistemas y Tecnologías de Información</description>
  <pubDate>Tue, 07 Apr 2009 07:41:38 +0000</pubDate>
  <generator>http://wordpress.org/?v=2.7.1</generator>
  <language>en</language>
  <sy:updatePeriod>hourly</sy:updatePeriod>
  <sy:updateFrequency>1</sy:updateFrequency>
  <creativeCommons:license>http://creativecommons.org/licenses/by/2.0/</creativeCommons:license>
  <title>Los nuevos influyentes, de Paul Gillin</title>
  <link>http://feedproxy.google.com/~r/ELBlogDeEnriqueDans/~3/iTY6yEjy6PQ/los-nuevos-influyentes-de-paul-gillin.html#comments</link>
  <comments>http://www.enriquedans.com/2009/04/los-nuevos-influyentes-de-paul-gillin.html#comments</comments>
  <pubDate>Tue, 07 Apr 2009 07:41:38 +0000</pubDate>
  <dc:creator>Enrique Dans</dc:creator>
  <guid isPermaLink="false">http://www.enriquedans.com/?p=6325</guid>
  <description><![CDATA[&#8220;Los nuevos influyentes&#8221; es la edición española de &#8220;Los nuevos influyentes - LID Editorial" href="http://www.lideditorial.com/2009/04/los-nuevos-influyentes-de-paul-gillin.html" height="1" width="1"/></description>
  <content:encoded><![CDATA[<p><a title="Los nuevos influyentes - LID Editorial" href="http://www.lideditorial.com/2009/04/los-nuevos-influyentes-de-paul-gillin.html">El libro fue un éxito en su edición original: era uno de los primeros que se aventuraba a entrar en la experiencia de escribir el prólogo fue muy interesante, porque el libro produce en la mayoría de los bloques de texto. La historia de los nuevos influyentes es un brutal ejercicio de democracia: para cada uno de los nuevos influyentes se ha escrito un capítulo. EL libro está disponible <a title="Los nuevos influyentes - LID Editorial" href="http://www.lideditorial.com/2009/04/los-nuevos-influyentes-de-paul-gillin.html">http://www.lideditorial.com/2009/04/los-nuevos-influyentes-de-paul-gillin.html</a>.</p></content:encoded>
  <fwv:commentRss>http://www.enriquedans.com/2009/04/los-nuevos-influyentes-de-paul-gillin.html#comments</fwv:commentRss>
  <feedburner:origLink>http://www.enriquedans.com/2009/04/los-nuevos-influyentes-de-paul-gillin.html</feedburner:origLink>
</channel>
<item>
  <title>Dominios .tel, curva de difusión y valor añadido</title>
  <link>http://feedproxy.google.com/~r/ELBlogDeEnriqueDans/~3/Ref7p1nqSks/dominios-tel-curva-difusion-y-valor-anadido.html#comments</link>
  <pubDate>Tue, 07 Apr 2009 07:41:38 +0000</pubDate>
  <dc:creator>Enrique Dans</dc:creator>
  <guid isPermaLink="false">http://www.enriquedans.com/?p=6325</guid>
  <description><![CDATA[&#8220;Dominios .tel, curva de difusión y valor añadido" href="http://www.enriquedans.com/2009/04/dominios-tel-curva-difusion-y-valor-anadido.html" height="1" width="1"/></description>
  <content:encoded><![CDATA[<p>El artículo trata sobre el dominio .tel, la curva de difusión y el valor añadido. El artículo está disponible en el blog de Enrique Dans.</p></content:encoded>
</item>

```

RSS es, por tanto, otro lenguaje de marcado. Su filosofía es similar a la del HTML, aunque su finalidad es diferente. Con los conocimientos que ya tenemos sobre HTML, podemos ver que no es muy difícil entender el RSS de la figura anterior. Se pueden identificar fácilmente etiquetas para el título (<title>) del blog, para la dirección del mismo (<link>), el lenguaje en el que está escrito (<language>), la licencia (<creativecommons:license>), así como los títulos (<title>) y contenidos de las historias (<content:encoded>).

Actualmente, algunos navegadores web ofrecen la funcionalidad de gestionar los RSS de los blogs que un usuario quiera leer con frecuencia. Además, existen páginas que también aglutinan los contenidos de varios blogs. Esto es muy útil si se quiere tener en un sitio centralizado blogs que compartan temática, como blogs de empleados de una empresa, de compañeros de clase o de un proyecto de software libre. El funcionamiento de estos sitios, conocidos como **planetas de blogs**, es el siguiente: un software toma los contenidos de los blogs del planeta exportados en RSS o Atom e integra las historias de todos los blogs como si fueran de un único blog, mostrándolas en orden cronológicamente inverso. La unión de la informalidad de escribir en un blog con la posibilidad de compartir tus experiencias y conocimientos con otros bloggers provoca que se creen comunidades virtuales muy interesantes, y muchos proyectos y empresas fomentan su uso porque potencian la difusión de información, que es difícil que se propague por los mecanismos habituales.

Ved también

Se hablará de las licencias Creative Commons en el módulo "Cultura libre".

6.3. Suites ofimáticas en línea

Las **suites ofimáticas en línea** están compuestas por los programas más comunes de las suite ofimáticas, principalmente un procesador de textos, una hoja de cálculo, un gestor de presentaciones, pero accesibles y operables mediante un navegador web.

El hecho de que la aplicación se ejecute en el servidor es muy interesante, ya que no sólo supone un ahorro al no tener que comprar el software, sino que también permite que varios usuarios editen un documento a la vez. Asimismo, el documento queda almacenado en el servidor y no en los discos duros locales de los usuarios. Esto tiene la ventaja de que se puede acceder a aquél desde cualquier ordenador con el sencillo requisito: que éste se halle conectado a Internet y tenga un navegador.

Pero también hay críticas a estos servicios. Los más escépticos argumentan que su uso es inseguro: al hecho de que se pueda perder la conexión o de que el servidor se pueda desconectar, debemos añadir que si estamos realizando una tarea crítica –y un informe para nuestra empresa lo puede ser–, lo estamos realizando con infraestructura ajena, con todos los problemas que esto puede traer (privacidad). Por otro lado, está la velocidad: ejecutar aplicaciones en el ordenador local suele ser mucho más rápido que hacerlo en remoto.

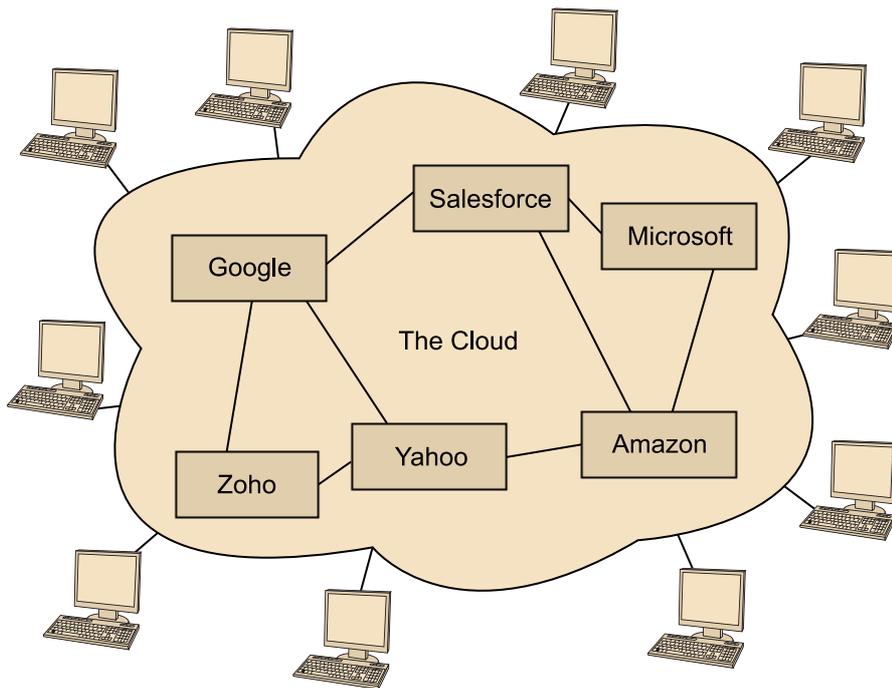
Las suites ofimáticas en línea son el ejemplo más popular de **la nube**⁷. El nombre se debe a que en los diagramas de redes de ordenadores se utiliza la metáfora de la nube para Internet, como se puede comprobar en la siguiente figura. La idea detrás de la nube es que se pueden utilizar servicios externos accediendo a ellos mediante un navegador web. Y dentro de estos servicios también podemos encontrar tareas que generalmente realizamos en nuestros ordenadores en local, como escribir un documento. En el caso de la nube, la infraestructura de estos servicios, su creación, puesta a punto y mantenimiento son llevadas a cabo por otros y no por nosotros mismos (o por nuestra empresa), por lo que el ahorro de costes puede ser importante.

Google Docs

El servicio más conocido de suite ofimática en línea es **Google Docs**. El servicio **Google Docs** se puede encontrar en <http://docs.google.com>.

⁽⁷⁾En inglés, *the cloud* o *cloud computing*.

Una perspectiva de la nube



Fuente: Wikipedia. Licencia: Creative Commons Attribution ShareAlike 3.0 Unported.

6.4. Otros sistemas web 2.0

Además de las aplicaciones explicadas en los anteriores subapartado, existen otros ejemplos de sitios web 2.0. En este subapartado referenciamos los más relevantes. Notad que todas estas aplicaciones no son mutuamente excluyentes: puede haber webs que las incluyan todas, cada una para una finalidad distinta.

- Los **sistemas de gestión de contenidos** son aplicaciones web utilizadas para crear, editar, buscar y publicar textos y multimedia en Internet. Estos sistemas funcionan con una base de datos que almacena los contenidos y datos asociados a éstos. Un contenido podría ser una noticia, mientras que los datos asociados pueden ser comentarios a la noticia, información sobre el autor, la fecha, etc. Los sistemas de gestión de contenidos tienen funcionalidad para gestionar los contenidos de manera sencilla. Esto se debe a que, entre otras cosas, permiten editar los contenidos de manera fácil mediante páginas web, ofrecen la posibilidad de establecer un flujo de trabajo (indicando que el contenido es un borrador o definitivo) e incluyen políticas para tener varios roles (autor, revisor, etc.). La diferencia entre los sistemas de gestión de contenidos y los wikis radica en que mientras los primeros están pensados como un lugar para ofrecer información en la web, los segundos están abiertos a que cualquiera pueda modificarlos de manera sencilla. Así, los sistemas de gestión de contenidos son interesantes cuando se quiere ofrecer mucha información y en constante cambio (por ejemplo, un periódico o la web de un ministerio), mientras los wikis tienen como objetivo la creación de información de manera cooperativa.

- Los **sistemas de docencia electrónica** permiten crear y gestionar cursos en línea. Estos sistemas ofrecen funcionalidades como calendario, foros, entrega de trabajos, gestión de alumnos, gestión del currículum, etc. También pueden integrar otros elementos como los que se han presentado hasta ahora, como los wikis y los blogs. Desde un punto de vista formal, los sistemas de docencia electrónica se pueden considerar como sistemas de gestión de contenidos específicamente creados para entornos docentes.
- Los **servicios de redes sociales** permiten a los usuarios apuntarse, indicar quiénes son sus "amigos" (o contactos) y compartir con ellos mensajes, fotos, música, noticias, juegos y comentarios.
- A caballo entre las redes sociales y los blogs, podemos encontrar el servicio **Twitter**. Twitter permite compartir con tus contactos mensajes cortos de hasta 140 caracteres. Los usuarios generalmente lo utilizan para comentar sus experiencias y pensamientos y compartirlos con sus seguidores, que es el nombre que reciben aquéllos que se apuntan como contacto. Debido a lo limitado del espacio de cada mensaje y del carácter marcadamente personal e informal que suelen tener, se habla de **micro-blogging**.

Ejemplo de sistema de docencia electrónica

El Campus Virtual de la Universitat Oberta de Catalunya es un sistema de docencia electrónica, y uno de los más populares es **Moodle**, que tiene una licencia de software libre.

Ejemplo de servicios de redes sociales

Los servicios de redes sociales más populares actualmente son **Facebook** y **LinkedIn** –una red social orientada a contactos profesionales–, aunque hay muchas más.

7. La web multimedia

Hasta el momento, hemos visto que la web ha evolucionado mucho y que ha cambiado sensiblemente desde los tiempos en los que la información que recibía un cliente era un documento estático en un servidor. En el apartado dedicado al lenguaje HTML, hemos podido comprobar que se pueden incluir otro tipo de elementos que no sea texto en las páginas, en particular imágenes. Del uso diario de la web sabemos que existe otro tipo de elementos que también se pueden encontrar, como pueden ser vídeos o audio. En este apartado, vamos a conocer un poco más a fondo estos elementos.

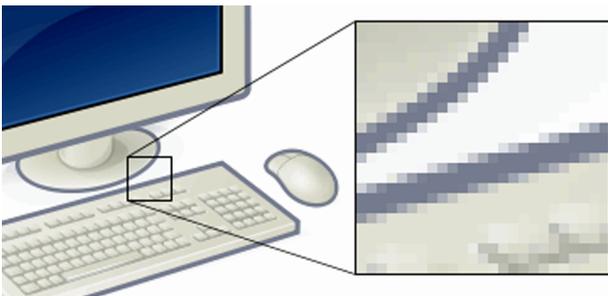
Multimedia viene de la unión de *multum* (muchos) y *medium* (medios) y denota el uso de distintos medios para la información: audio, gráficos, vídeo, animación, etc.

7.1. Imágenes

Sabemos cómo se representa la información, en especial, la información en formato texto. Podríamos resumir que el método consiste en que cada una de las letras tenga una secuencia asociada de unos y ceros, o sea bits. De esta modo tenemos una codificación de caracteres, como el ASCII.

La cuestión que vamos a abordar ahora es cómo representar la información gráfica. Tomemos una imagen y supongamos que está compuesta de minúsculas partículas de colores únicos. La unión de todas las partículas nos dará la imagen. Cada partícula es lo que conoceremos como **píxel** y representará un color. La siguiente figura pretende clarificar esta idea: la imagen global está compuesta por píxeles de forma cuadrada, como se puede observar al aumentar una parte de la imagen. Cada píxel vendrá dado por un color. Desde lejos, la capacidad integradora del ojo humano provoca que la imagen no parezca estar compuesta por píxeles.

Imagen con una parte aumentada



En la parte aumentada se pueden ver los píxeles de manera cuadrada que componen la imagen. Fuente: Wikipedia. Licencia: GFDL, y Creative Commons by-sa-2.5.

Ved también

La representación de la información y la codificación ASCII se explican en el módulo "Aspectos tecnológicos de los sistemas informáticos".

Pixel

Píxel es la contracción de *picture element*, elemento del dibujo.

Supongamos, para simplificar, que estamos tratando una imagen en blanco y negro. Entonces, podríamos convenir que un píxel tendría un valor de 1 cuando fuera negro (N), y de 0 cuando fuera blanco (B). Cada píxel ocuparía un bit y la secuencia 101010 correspondería a NBNBNB. Si quisiéramos tener más colores, necesitaríamos darle a cada píxel más bits. Esto es conocido como la **profundidad de color**. La regla general es que con n bits podemos obtener 2^n colores. En el caso de blanco y negro, n es igual a 1, ya que solo tenemos dos colores.

Con esta regla, podríamos codificar todos los colores que existen con un código único, como hacíamos con las letras. El problema es que el número de colores es infinito, por lo que no tendríamos bits suficientes. Por suerte, el ojo humano no es capaz de distinguir entre todos los colores. Llega un momento en el que un pequeño cambio en la tonalidad de un color es inapreciable para el ojo humano. Se conoce como **color verdadero** la codificación con 24 bits con la que podemos representar 16.777.216 colores.

Para terminar con la representación del color, queda una cuestión más. Todos los colores pueden representarse por una mezcla de tres colores, conocidos como **colores primarios**, que son el rojo, el verde y el azul. A partir de los tres colores primarios, podemos, por ejemplo, obtener el amarillo (sumando rojo puro y verde puro), el cian (sumando verde puro y azul puro) y el magenta (sumando rojo puro y azul puro). La suma de los tres colores puros nos daría el blanco.

Si en vez de tomar los colores puros elegimos tonalidades, por ejemplo un rojo más débil, y lo sumamos con el verde, el amarillo que obtendremos será más débil que el amarillo puro. Con este mecanismo se puede obtener cualquier color a partir de los tres colores primarios y sus tonalidades.

El color de un píxel se puede dividir en tres partes de 8 bits, una parte por cada color primario. Con 8 bits podemos representar 256 tonalidades (recuérdese que $2^8 = 256$). En otras palabras, tenemos 256 tonalidades de rojo, 256 tonalidades de verde y 256 tonalidades de azul, que podemos mezclar entre sí. Si quisiéramos obtener el azul puro, lo que deberíamos es tomar la tonalidad más pura del azul y nada del rojo ni del verde. Para el cian, necesitaríamos las tonalidades más puras de verde y azul, pero nada de rojo. Sin embargo, si quisiéramos un cian algo más claro, entonces habría que añadir algo de rojo.

Para obtener la imagen completa, tendremos una secuencia de píxeles, tantos como vienen dados por la **resolución**. La resolución es una matriz de A columnas por B filas de píxeles. Como las imágenes suelen ser rectangulares, la resolución se suele indicar en los siguientes términos $A \times B$, como 800×600 . En otras palabras, la imagen consta de 800 píxeles de ancho por 600 píxeles de alto. En total, tendríamos 480.000 píxeles en la imagen, que es el resultado de

multiplicar el ancho por el alto. En las cámaras digitales la resolución viene dada en megapíxeles, que no es otra cosa que millones de píxeles. Así, una cámara de 5 megapíxeles puede crear fotos de hasta 2.560x1.920.

Finalmente, el **tamaño de la imagen** será el resultado de multiplicar la resolución por la profundidad de color:

$$\text{Tamaño imagen} = \text{Resolución} \times \text{Profundidad de color}$$

donde la profundidad de color nos da el tamaño en bits de cada píxel, mientras que la resolución indica el número de píxeles de nuestra imagen. Así, una foto de una cámara de 5 megapíxeles a color verdadero ocupa $2.560 \times 1.920 \times 24$ bits = 117.964.800 bits o unos 14 megabytes.

Sin embargo, las fotos en el ordenador no suelen ocupar tanto. En el subapartado 7.3, al tratar de vídeo, hablaremos de una técnica conocida como **compresión**, que es la responsable de que una imagen ocupe menos de lo que hemos contado aquí. En la siguiente tabla se presentan algunos de los formatos de imágenes más habituales, indicando sus características más importantes.

Formatos de imagen

| Formato | Características más importantes |
|------------|---|
| Raw | Imagen "cruda" sin compresión. |
| BMP | Formato de imagen de los iconos de Microsoft Windows. Admite compresión sin pérdidas. |
| TIFF | Formato popular en escáneres (puede tener compresión con pérdidas). |
| GIF | Formato muy popular en la web. Es de 8 bits (256 colores máximo) y permite crear animaciones. |
| PNG | Formato diseñado para reemplazar el formato GIF. Los colores pueden codificarse con 8 bits (256 colores), 24 bits (4 millones de colores) y 48 bits (unos 256 billones de colores). |
| JPG o JPEG | Utilizado para fotografías e imágenes con muchos colores. Es un formato comprimido con pérdida de calidad ajustable. |

7.2. Audio

El **audio digital** es la codificación en digital de una señal eléctrica obtenida a partir del sonido.

Los humanos sólo tenemos capacidad auditiva en un rango específico, que va desde 20 hasta 20.000 Hz. Para obtener audio digital a partir de una señal eléctrica es necesario realizar dos pasos: el muestreo y la cuantificación.

El **muestreo** consiste en tomar muestras de una señal eléctrica con regularidad, tal y como se puede ver en la figura siguiente. Así, cada período de tiempo T , tomamos una muestra de la señal $S(t)$ (básicamente, nos quedamos con el

El sonido

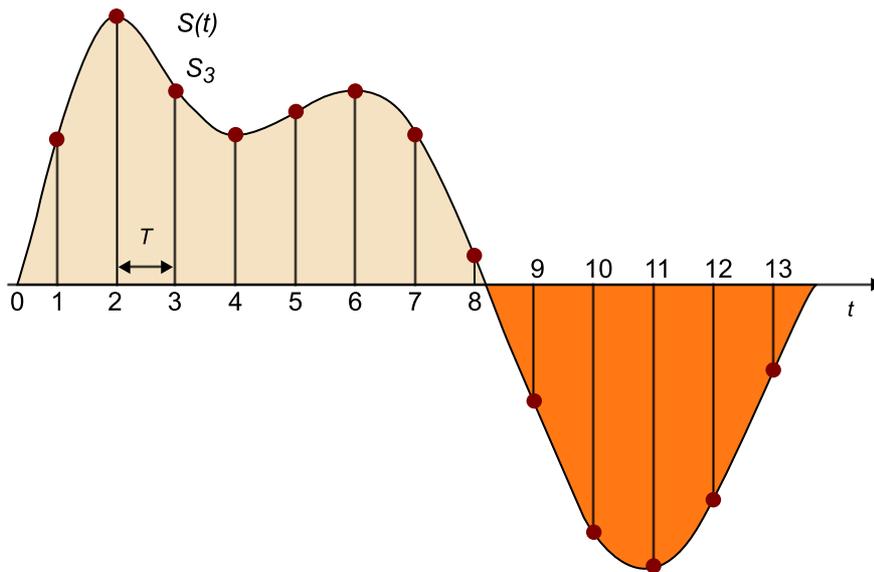
El sonido es en realidad una onda de presión.

Hercio

Un hercio, o 1 Hz es equivalente a una variación de la onda de presión por segundo.

valor de la señal $S(t)$ en ese instante de tiempo). Por ejemplo, en $t = 3$, el valor de la señal es S_3 . La regularidad necesaria viene dada por la condición de no perder información, ya que en caso contrario la codificación es de baja calidad. Matemáticamente, se ha demostrado que es suficiente con muestrear con una frecuencia el doble de rápida que la frecuencia más alta de la señal eléctrica. Esto quiere decir que si los humanos escuchamos hasta 20.000 Hz, deberíamos muestrear como poco a 40.000 Hz. Por ello, la frecuencia de muestreo de los CD de audio es de 44.100 Hz.

Muestreo de una señal de audio



Fuente: Wikipedia. Licencia: Dominio Público.

El segundo paso es la **cuantificación**. Esta consiste en asignar un valor concreto al valor de la onda que se ha muestreado. Cuantos más bits usemos para representar este valor, más precisión habrá y, por lo tanto, más calidad. Supongamos que una señal de 20.000 Hz se cuantifica con dos bits, el 0 y el 1. Para los valores menores de 10.000 Hz se asigna un 0 y para los valores mayores de 10.000 Hz, un 1. Así, se podrían cuantificar todos los valores obtenidos mediante el muestreo del paso anterior. El problema es que con tan pocos bits la cuantificación sería muy mala, porque estaríamos dando un 0 tanto a un valor de unos pocos hercios como a otro de 9.999 Hz

La solución consiste en dedicar más bits a la cuantificación. ¿Cuántos? Pues depende de la calidad que deseemos obtener. Se ha comprobado que con 16 bits (con lo que obtenemos 2^{16} , 65.536 valores de cuantificación diferentes) se puede obtener una calidad más que suficiente para una persona común. La cuantificación de 16 bits se utiliza en los CD de audio tradicionales, que almacenan el audio en formato digital. Al utilizar 16 bits, tenemos suficientes valores para muestrear los 20.000 Hz.

El **tamaño de archivo de audio digital** se calcula multiplicando el tiempo que dura por la frecuencia de muestreo por la cuantificación:

$$\text{Tamaño archivo audio} = \text{Duración} \times \text{muestras/segundo} \times \text{bits/muestra}$$

Mientras que la frecuencia de muestreo nos da el número de muestras por segundo, la cuantificación nos da el número de bits que utilizamos para cada una de las muestras. Con la multiplicación de estos dos factores, se obtiene el número de bits que ocupa un segundo de audio. Los CD de audio se suelen grabar en estéreo (o sea, que tenemos dos canales que muestrear y codificar), por lo que 1 segundo de audio ocupa 1.411.200 bits o 176.400 bytes.

Con un sencillo cálculo se puede comprobar que en un CD se puede grabar unos 70 minutos de música. Una sencilla operación matemática, que se deja como ejercicio, demuestra que el audio ocupa algo más de 700 megabytes. Precisamente lo que se podría grabar si se utilizara ese mismo CD para grabar datos.

La siguiente tabla presenta los formatos de archivos de audio más comunes y sus características más importantes. El lector ha de tener en cuenta que si graba canciones de audio en un CD de audio caben hasta un máximo de 80 minutos de canciones, pero que si las graba en formato MP3 le entran muchísimas más (como diez veces más). Esto también se debe a la compresión, que se explica en el siguiente subapartado. Además del popular formato MP3, existe otro formato de compresión de audio ampliamente utilizado: OGG. La principal diferencia entre uno y otro radica en que utilizan técnicas de compresión diferentes, siendo las del formato OGG un estándar libre y abierto (que no es el caso del MP3).

Formatos de audio

| Formato | Características más importantes |
|---------|---|
| WAV | Formato de audio sin compresión. |
| MP3 | Formato de audio con compresión con pérdidas. |
| OGG | Formato (libre) de audio con compresión con pérdidas. |

7.3. Vídeo digital y la compresión de datos

El **vídeo digital** viene dado por una sucesión de imágenes junto a una secuencia de audio. El número de imágenes (denominadas **frames** en la jerga) necesario para conseguir la ilusión de una imagen en movimiento debe ser, según algunos estudios, mayor de 15. En el cine, la velocidad es de 24 *frames* por segundo (fps), mientras que en la televisión –dependiendo del estándar utilizado– va de 25 a 30 fps.

El tamaño de una imagen de vídeo vendrá dado, por lo tanto, por el tamaño de cada uno de los *frames*, por el número de frames por segundo, por el tamaño del audio por segundo y todo ello multiplicado por la duración del vídeo.

$$\text{Tamaño archivo vídeo} = (\text{frames/segundo} \times \text{bits / frames} + \text{bits audio/segundo}) \times \text{duración del vídeo}$$

Si tenemos un frame de 1.280x720 píxeles, 25 fps, en color verdadero, y el audio es de una calidad como la del CD-audio (16 bits, 44.100 Hz de muestreo y estéreo). Un segundo de vídeo ocuparía más de 66 megabytes, ya que 1 segundo de vídeo ocupa debido al audio 0,17 megabytes, y debido a las imágenes, casi 66 megabytes. En un DVD de capa simple, cuya capacidad es de 4,7 gigabytes, apenas cabrían 74 segundos de vídeo.

Para reducir el volumen de datos aparece la **compresión de datos**. Para entender la compresión, se debe tener en cuenta que la información puede ser de tres tipos:

- **Redundante.** Información que se repite o que se puede inferir aunque faltara. Sea una imagen monocolor o de múltiples colores, el número de píxeles (resolución) y el número de bits para cada píxel (profundidad de color) no cambian. Sin embargo, al utilizar técnicas de compresión podríamos indicar de alguna forma que la imagen tiene píxeles repetidos y reducir tamaño.

Ejemplo de información redundante

El ejemplo más clásico de información redundante es una imagen que tiene todos los píxeles en blanco. Sin técnicas de compresión, esta imagen ocuparía lo mismo que cualquier otra imagen, ya que como hemos visto el tamaño de una imagen depende de su resolución y de la profundidad de color.

- **Irrelevante.** Se trata de información que no podemos apreciar.

Ejemplo de información irrelevante

Un ejemplo típico de información irrelevante se da en el audio, donde hemos visto que se suele muestrear para captar frecuencias de hasta 20 kHz. Y aunque existen personas con el oído tan fino como para escuchar variaciones a 20 kHz, lo usual es que la mayoría no llegue más allá de los 16 kHz. Si se eliminaran estas frecuencias, irrelevantes en general, el archivo ocuparía menos.

- **Básica:** Es la información que ni es redundante ni irrelevante.

Dependiendo del tipo de información que se utiliza, existen dos tipos de compresión: compresión sin pérdidas y compresión con pérdidas.

- La compresión **sin pérdidas** es aquella para la que podemos recuperar toda la información original a partir del archivo comprimido. Para realizar este tipo de compresión, se trabaja y elimina la información redundante.
- Por otro lado, la compresión **con pérdidas** es aquella para la que no podemos recuperar la información original a partir del archivo comprimido. Dentro de la compresión con pérdidas, debemos tener en cuenta que –aun habiendo pérdidas– puede que el receptor no se dé cuenta de éstas. Es lo que se conoce como **compresión con pérdidas, pero subjetivamente**

Ejemplo de compresión sin pérdidas

En el formato *zip* se utiliza la compresión sin pérdidas. De hecho, imaginemos un texto que, tras haber sido comprimido, tuviera pérdidas; sería ilegible.

sin pérdidas: estaremos eliminando la información redundante y la irrelevante.

- Si la compresión fuera con pérdidas y se eliminara también parte de la información básica, hablaríamos de **compresión subjetivamente con pérdidas**. Puede parecer algo raro un escenario en el que se pierda información relevante, pero esto es algo muy común, particularmente en el vídeo digital, debido a su gran tamaño. La razón detrás de una compresión tan fuerte es que muchas veces preferimos ver algo, aunque sea sin mucha nitidez pero de manera inmediata, a tener que esperar un rato para verlo con mayor calidad. Un vistazo a los vídeos del portal de compartición de vídeos YouTube convencerá a los más escépticos: la imagen de los vídeos no será la mejor, pero se valora más el poder verlos "al vuelo".

Ejemplo de compresión con pérdidas

El caso del audio con la eliminación de frecuencias por encima de 16 kHz es un ejemplo de compresión con pérdidas.

Dentro de la compresión de vídeo, se pueden diferenciar dos formas de comprimir:

- La primera es **compresión en el mismo frame**. Estas técnicas se asemejarían a la compresión utilizada en una imagen, aunque suelen ser más fuertes (entiéndase, con mayores pérdidas), ya que una imagen que sólo vamos a ver una fracción de segundo no necesita contener muchos detalles.
- La segunda es **compresión entre frames**. La razón de esta compresión es que un *frame* y el siguiente suelen tener mucho en común. Por ello, en vez de enviar dos frames diferentes, utilizando las técnicas de compresión en un mismo *frame*, se podría enviar un primer *frame* de mayor calidad y sucesivos *frames* con las diferencias con el primero. A este tipo de técnicas de compresión se las conoce también como compresión diferencial.

En la siguiente tabla se pueden ver algunos formatos de vídeo. Todos ellos son formatos con compresión con pérdidas, porque sino los vídeos ocuparían un gran volumen de datos.

Formatos de vídeo digital

| Formato | Características más importantes |
|-----------|---|
| DivX | Formato de vídeo con compresión con pérdidas. Los ficheros a menudo tienen la extensión AVI |
| MPEG | Formato de vídeo con compresión con pérdidas |
| QuickTime | Formato de vídeo con compresión con pérdidas |
| FLV | Formato de vídeo en flash con compresión con pérdidas |

7.4. Sistemas de compartición de imágenes, audios y vídeos

Uno de los hitos más importantes de la web 2.0 ha sido la aparición de sitios en los que los usuarios pueden compartir sus imágenes, sus canciones y sus vídeos.

Dentro del mundo de las imágenes –especialmente fotografías–, el lugar más popular actualmente es **Flickr**.

Flickr va más allá de ser un repositorio de fotografías con funcionalidades típicas de un sistema de gestión de contenido (añadir comentarios, puntuar las fotos, etc.). También se puede añadir información acerca de las fotografías (esta información sobre la información se denomina **metainformación**) mediante el uso de **etiquetas**⁸. De este modo, se pueden realizar búsquedas por palabras clave en Flickr que mostrarán fotografías con etiquetas relacionadas. Las etiquetas pueden incluir información del lugar, la fecha y la licencia de uso (en particular, en Flickr se puede indicar si la fotografía tiene una licencia Creative Commons).

En lo que se refiere al audio, existen múltiples proyectos de radio en Internet que emiten "en vivo". En la jerga informática, esto se conoce como emitir mediante **streaming**: no se descarga un archivo para ser escuchado posteriormente, sino que se emiten datos que son reproducidos mientras éstos llegan. Existen múltiples radios en Internet que ofrecen un servicio similar a la radio convencional. Pero fuera de las radios, una comunidad interesante es **Jamendo**, un portal en el que los músicos se pueden dar a conocer ofreciendo su música bajo las condiciones de las licencias Creative Commons.

Finalmente, están los sitios de compartición de vídeo. Aunque hay muchos, el más conocido es **YouTube**. Estos sitios permiten subir vídeos en una serie de formatos (MPEG, AVI, etc.) que los transforman para poder ser vistos a baja calidad, alta calidad, incluso para móviles. Generalmente, la transformación consiste en bajar la resolución (o sea, hacer los frames más pequeños) y aplicar técnicas de compresión con pérdidas. La gran ventaja es que un usuario sin una gran conexión de Internet es capaz de ver los vídeos "al vuelo" sin tener que esperar mucho. Asimismo, los vídeos se pueden integrar de manera sencilla, mediante la inserción de una serie de etiquetas HTML que se presentan en un cuadro, en otras páginas web y en blogs. Por contra, en YouTube la inclusión de vídeos con contenidos de terceros ha provocado muchos problemas con los derechos de autor.

⁽⁸⁾En inglés, *tags*.

Ved también

Trataremos las licencias Creative Commons en el módulo "Cultura libre".

Resumen

En este módulo hemos introducido la web, el servicio de Internet más conocido y usado. Para ello, se ha ofrecido una perspectiva histórica de su evolución, así como de su arquitectura cliente-servidor. También se han mostrado brevemente los protocolos más importantes sobre los que se asienta la web, básicamente HTTP –para la petición y respuesta de páginas web entre el cliente y el servidor– y el estándar HTML, mediante el cual se elaboran las páginas web.

Tras ver las limitaciones de las primeras aproximaciones de la web, se han presentado algunas nociones sobre la web dinámica e interactiva y cuáles son los fundamentos de las aplicaciones web. De entre las aplicaciones web, hemos visto en detalle varios tipos: los buscadores, los wikis y los blogs. De esta manera, el módulo finaliza presentando el concepto de web 2.0, mucho más centrado en el usuario, considerando interactividad y participación.

Actividades

1. Cread un blog en alguno de los sitios indicados en el texto. Probad a editar una entrada de presentación en HTML, incluyendo imágenes y estilos de caracter (negrita, itálica, subrayado, etc.).
2. En los últimos tiempos hemos podido presenciar una amplia polémica entre los blogs y el periodismo tradicional. Recopilad argumentos a favor y en contra de ambas posturas.
3. Crearos una opinión sobre la siguiente afirmación: "la web 2.0 no aporta nada nuevo sobre lo que ya existía", y defended vuestra postura.
4. Discutid la aportación y riesgos (de funcionamiento, legales, etc.) para una empresa de utilizar aplicaciones web en vez de aplicaciones instaladas en sus ordenadores.
5. Cread una página wiki en WikiOle. Intentad introducir código HTML y código estructurado. Comprobad la diferencia entre el código estructurado de WikiOle y el utilizado en Wikimedia, que es el presentado en este texto.
6. Buscad en DMoz por páginas sobre los Juegos Olímpicos de Pekín 2008 sin utilizar el formulario de búsquedas (o sea, navegando por la jerarquía de conceptos).
7. Utilizad la búsqueda avanzada de Google para encontrar páginas web que hablen de la sección de baloncesto Fútbol Club Barcelona, pero no de la de fútbol. Utilizad para ello elementos de búsqueda avanzada, tales como los descritos en la Ayuda de Google.
8. Cread una página web personal en HTML utilizando un editor de textos. La página ha de incluir vuestro nombre como título, una foto y unos dos párrafos de currículum. Una vez creada la página, aprobadla con el servicio de validación del W3C.

Ejercicios de autoevaluación

1. Nos hemos comprado un marco electrónico de los que nos permiten mostrar varias fotos digitales en nuestra mesilla de noche. Las dimensiones de las fotos son de 2.000x1.500 píxeles. En las instrucciones leemos que si pasamos fotos en formato BMP (con color verdadero) por el puerto USB, el software integrado en el marco automáticamente las comprime en JPEG y pasan a ocupar una décima parte de su tamaño original. ¿Cuántas fotos caben si la capacidad de la memoria incluida en el marco es de 120 megabytes?

2. Queremos crear un sintetizador de sonido sencillo para utilizarlo en las maquetas de nuestro grupo. El sintetizador será capaz de producir las siete notas musicales (do, re, mi, fa, sol, la, si) en 8 octavas (o sea, que podrá crear ocho veces las siete notas). ¿Cuántos bits necesitáis para codificar las notas? Si cada segundo de nuestro sintetizador puede emitir hasta 1.000 notas diferentes, ¿qué tamaño en kilobytes tendría el fichero de una canción de 20 segundos? Un amigo nos comenta que podríamos guardar esos mismos datos en un fichero de menor tamaño sin que se pierda calidad. ¿Es eso cierto? ¿Se os ocurre alguna manera de hacerlo? ¿Cómo?

3. En nuestro viaje de interrail pasamos por la bella ciudad de Praga, y para inmortalizar la visita nos hemos llevado nuestra flamante cámara que hace fotos de hasta 10 megapíxeles (con color verdadero) y que cuenta con una memoria de hasta 1GB. En un principio, pensamos en hacer fotos de 10 megapíxeles y no utilizar compresión a la hora de almacenar nuestras fotos. ¿Cuántas fotos podríamos hacer de esta manera?

Vemos que el número de fotos que se pueden hacer de esta manera es muy pequeño para una ciudad tan bonita como Praga. Uno de nuestros compañeros de viaje nos comenta que podemos hacer dos cosas:

- Utilizar compresión al almacenar las fotos. Miramos las instrucciones y vemos que nuestra cámara permite compresión de hasta un factor de 10 creando imágenes en formato JPEG.
- Configurar la cámara para que el número de megapíxeles de las fotos sea menor. Vemos que para obtener un número de fotos igual que con la otra solución, sería necesario bajarlo a 1 megapixel.

¿Cuál de las dos soluciones es mejor? ¿Por qué?

Solucionario

Ejercicios de autoevaluación

1. Para calcular cuánto puede ocupar una foto como máximo, debemos calcular primero el número de píxeles, luego multiplicarlo por el tamaño que ocupa cada píxel (a partir de la profundidad de color) y finalmente aplicarle el factor de compresión.

$2.000 \times 1.500 = 3.000.000$ de píxeles

$3.000.000$ de píxeles \times 4 bytes/píxel (color verdadero) = $12.000.000$ bytes = 12 megabytes (hemos tomado para eso la aproximación de que un millón de bytes es un megabyte).

Si aplicamos el factor de compresión, cada foto puede llegar a ocupar 1,2 megabytes.

Si tenemos 120 megabytes de capacidad en el marco, podríamos tener hasta 100 fotos en él como mínimo.

2. Tenemos siete notas (do, re, mi, fa, sol, la, si) y ocho octavas, lo que en total nos da $7 \times 8 = 56$ notas que queremos codificar. Para codificar 56 notas de manera unívoca, necesitamos 6 bits (lo que nos da la posibilidad de codificar hasta 64 cosas diferentes, lo que es suficiente para nosotros).

6 bits/nota \times 1.000 notas/segundo \times 20 segundos = 120.000 bits.

Entendemos 1 kilobyte como, aproximadamente, 1.000 bytes. En ese caso, el tamaño de 20 segundos de canción es de alrededor de 15 kbytes.

Lo que nos comenta nuestro amigo sobre compresión es, por lo tanto, cierto. Una manera de hacerlo es tomar la primera nota (6 bits) y, a partir de ahí, en vez de utilizar la codificación total de cada nota (que sería de 6 bits), utilizar una codificación relativa a la nota anterior de menos bits (por ejemplo, 4).

3. Primero debemos calcular cuánto ocupa una foto. Tenemos el número de píxeles que hemos de multiplicar por el tamaño que ocupa cada píxel (a partir de la profundidad de color), así que las cuentas son:

$10.000.000$ píxeles \times 4 bytes (color verdadero) = $40.000.000$ bytes = 40 megabytes (para ello hemos considerado la aproximación de que un millón de bytes es un megabyte).

Si nuestra memoria es de 1GB (aproximadamente 1.000 megabytes), el número de fotos que podemos hacer es 25.

La primera solución es claramente mejor, porque al comprimir se aplican algoritmos que permiten almacenar más información en menos espacio.

Glosario

agregador de blogs *m* Véase planeta de blogs.

Apache *m* Es el servidor web más popular, con una cuota de mercado superior al 60% desde hace años. Véase también Apache Software Foundation.

Apache Software Foundation *m* Fundación que se encarga de velar por el desarrollo y la promoción del servidor web Apache y de otros proyectos generalmente relacionados con tecnologías web. Más información en su página web.

aplicación web *f* Aplicaciones (programas software) a los que el usuario accede mediante un cliente web (navegador).

Araña *m* Programa que rastrea y almacena la web siguiendo los hipervínculos en las páginas web para su posterior procesamiento.

atom *m* Formato estructurado que se emplea para exportar los contenidos de blogs, entre otros. Véase también RSS.

bitácora *f* Véase blog.

blog *m* Sitio web en el que se insertan historias o entradas (en inglés *posts*), artículos o textos que pueden incluir imágenes y otros elementos multimedia, en orden cronológicamente inverso.

buscadores *m pl* Aplicaciones web que nos permiten encontrar información en Internet de manera rápida. Mediante la introducción de una o varias palabras clave, nos devuelve una página índice que enlaza páginas web con información relacionada.

cliente *m y f* Programa software que accede a un servicio ofrecido por un servidor. En el caso de la web, el cliente web es el navegador.

color verdadero *m* Dícese de aquella profundidad de color de al menos 24 bits. Con el color verdadero, se pueden representar al menos 16.777.216 colores.

compresión *f* Las secuencias de audio y vídeo ocupan mucho espacio, por lo que se utilizan técnicas de compresión que permiten disminuir el tamaño. De este modo se facilita su intercambio (el tiempo de descarga de un vídeo comprimido en formato DivX es menor que sin comprimir, y un vídeo que antes requería mucho espacio ahora cabe en un CD-ROM). DivX y MP3 comprimen con pérdidas, lo que significa que el archivo comprimido no tiene la calidad del original. Sin embargo, las pérdidas son –en muchas ocasiones y dependiendo del factor de pérdidas introducido– tolerables (incluso indetectables) para el ser humano. Existen formatos de compresión sin pérdidas, como el usado en los archivos zip, pero su utilidad es menor para audio y vídeo.

cuantificación *f* Proceso por el que se le asigna un valor binario a una muestra (véase muestreo).

DivX *f* Tecnología de compresión de vídeo basada en el estándar de compresión MPEG-4.

fotoblog *m* Blog de fotos.

Google Docs *m* Es una suite ofimática en línea de la empresa Google.

hipertexto *m* Documento de texto con enlaces (véase hipervínculos) a otros documentos.

hipervínculo *m* Conexión o enlace entre documentos.

HTML *m* Acrónimo de *HyperText Markup Language* (lenguaje de marcado de hiperTexto). Es el lenguaje en el que están escritas las páginas web.

HTTP *m* Protocolo de comunicaciones utilizado para intercambiar documentos de hipertexto entre dos ordenadores, generalmente un servidor y una máquina cliente. Acrónimo de *Hypertext Transfer Protocol*.

índice temático *m* Sitio web en el que las páginas web se encuentran ordenadas de manera jerárquica por temas.

internet *f* conjunto global de redes de ordenadores interconectados mediante el protocolo Internet Protocol (IP).

la nube *f* Paradigma de computación en el que los servicios se proveen por medio de Internet, sin que el usuario tenga percepción, conocimiento ni control de la infraestructura tecnológica de la que hace uso. Ejemplos de *cloud computing* son Gmail o Google Docs.

lenguaje de marcado *m* Lenguaje que permite marcar texto, de manera que se pueda añadir información adicional al contenido. Véase, por ejemplo, HTML.

modelo cliente-servidor *m* Arquitectura de computación distribuida, en la que una parte ofrece un servicio (el servidor) que es demandado por la otra (el cliente). La World Wide Web es un ejemplo de arquitectura cliente-servidor.

motor de búsqueda *m* Aplicación web que permite buscar por palabras clave en una base de datos de página web previamente almacenada por arañas.

Mozilla *m* Proyecto iniciado por la compañía Netscape a finales de la década de los noventa tras liberar su navegador Netscape Navigator. El programa más conocido de Mozilla es el navegador Firefox. Más información en su página web.

MP3 *m* Formato de compresión para audio.

muestreo *m* Toma de muestras de una señal eléctrica con regularidad, para que posteriormente le sea asignada a cada muestra un valor (véase cuantificación).

Netscape *m* es la compañía que a mediados de la década de los noventa comercializaba el popular navegador Netscape Navigator. Cuando ya daba por perdida la *guerra de los navegadores* con el Internet Explorer de Microsoft, publicó el código fuente de su navegador bajo una licencia de software libre y fundó el proyecto Mozilla con el objetivo de remontar el vuelo. Fue adquirida posteriormente por AOL. Hoy el proyecto Mozilla es totalmente independiente de Netscape.

palabras clave *f pl* Palabras más importantes de un texto.

página web *f* Documento cuyo contenido está en formato HTML.

planeta de blogs *f* Programa software que toma contenidos de varios blogs (generalmente disponibles como RSS o Atom) y los muestra de manera integrada en una única página web como si se tratara de una nueva bitácora.

profundidad de color *f* Número de bits por cada píxel. Cuanto mayor sea la profundidad de color, más calidad tiene una imagen.

resolución *f* Número de píxeles de una imagen. Cuanta más resolución, más calidad tiene una imagen.

RSS *f* (de *Really Simple Syndication*, sindicación realmente simple). Formato estructurado que se emplea para exportar los contenidos de blogs, entre otros. Véase también Atom.

servidor *m* Puede referirse tanto al componente físico (hardware) como al componente lógico (software) que ofrece un servicio, como páginas web. Véase también cliente y Apache.

Streaming (en español: flujo) *m* Se trata de un modo de transmisión de datos entre cliente y servidor un poco peculiar (en relación con lo que es común en Internet). Fue ideado principalmente para la transmisión de datos multimedia. El flujo de datos ha de ser constante, para evitar interrupciones incómodas en audio y vídeo, y se ha de mantener la ordenación original de los datos, porque en caso contrario se dificulta la comprensión de audio y vídeo. El streaming suele utilizarse junto con técnicas de compresión.

suite ofimática en línea *f* Composición de los programas más comunes de las suite ofimáticas, principalmente un procesador de textos, una hoja de cálculo y un gestor de presentaciones, pero accesibles y operables mediante un navegador web.

The Cloud *m* véase la nube.

videoblog *m* Blog de vídeos.

web *f* véase World Wide Web.

web 2.0 *f* Páginas y aplicaciones web que facilitan compartir información y la colaboración en Internet. Su diseño está más orientado hacia el usuario que hacia los datos, como sucede en la web "tradicional".

Wiki *f* Aplicación web que permite que la propia página web sea editada con el propio navegador web.

Wikipedia *f* Enciclopedia en línea totalmente libre que utiliza como infraestructura informática un wiki.

World Wide Web *f* Servicio de tipo cliente-servidor basado en el protocolo de comunicación Hypertext Transfer Protocol (HTTP), que sirve para intercambiar información entre ordenadores conectados a una red, generalmente Internet.

Bibliografía

Cobo Romaní, C.; Pardo Kuklinski, H. (2007). *Planeta Web 2.0. Inteligencia colectiva o medios fast food*. ISBN 978-84-934995-8-7.

Eguíluz Pérez, J. (2008). *Introducción a XHTML*.

Herrera, A. M.ª; Villabona, J. (2008). *15 años de Internet*. Madrid: Grupo Buho. ISBN: 9788493625986.

O'Reilly, T. (2005). *What is Web 2.0?*

Rojas, O.; Alonso, J.; Antúnez, J. L.; Orihuela, J. L.; Varela, J. (2005). Madrid: Editorial ESIC. *Blogs. La conversación en Internet que está revolucionando medios, empresas y ciudadanos*. ISBN: 84-7356-427-8.

Tannenbaum, A. S. (2003). *Redes De Computadoras* (4.ª ed.). Madrid: Pearson. ISBN: 9702601622.