HTTP (HyperText Transfer Protocol)

ÍNDICE

1. Introducción

HTTP es el «**Protocolo de Transferencia de Hipertexto**» (*HyperText Transfer Protocol*) fue creado en 1990 en el CERN (*Conseil Européen pour la Recherche Nucléaire* o *Consejo Europeo para la Investigación Nuclear*), como un medio para compartir los datos científicos a nivel internacional, instantáneamente y a bajo costo.

Es el método más común de intercambio de información en la *World Wide Web* o **WWW**, mediante el cual se transfieren páginas desde un ordenador servidor a uno cliente.

Este protocolo está descrito en la RFC 1945 (1996), que ha sido ampliado y modificado por otras RFC.

Aún siguen apareciendo más adaptaciones que resuelven las nuevas necesidades de este protocolo, como la versión segura de HTTP denominada **HTTPS**, que puede utilizar diversos métodos de cifrado (siempre que sea entendido tanto por el servidor como por el cliente).

Aunque la forma de intercambio de información, sobre todo en cantidad, más utilizada en Internet es mediante HTTP, y la red más conocida es la WWW, Internet es mucho más, tanto en cantidad de protocolos soportados como en volumen de información.

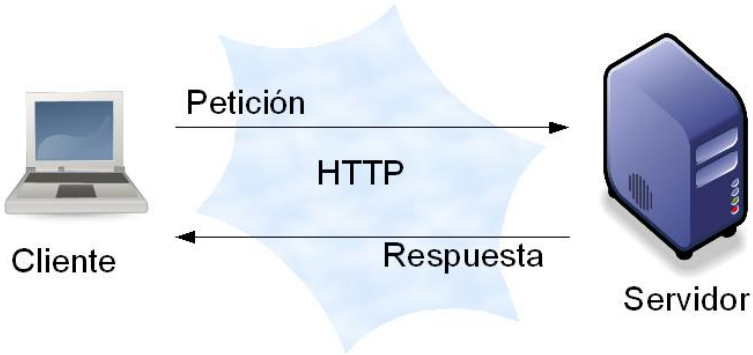
WWW (*World Wide Web*), también llamada «la Web», es una agrupación de miles de páginas electrónicas basadas en textos y en todo tipo de información multimedia, conectadas entre sí, a través de hiperenlaces, con la finalidad de recuperar información de forma sencilla, sin que sea necesario conocer previamente su ubicación exacta.

2. Modelo cliente-servidor de la Web

La Web sigue un modelo cliente-servidor en una red TCP/IP, ya sea local o interconectada con las demás redes a través de Internet. La petición de las páginas en formato HTML se realiza desde un cliente a un servidor. Ese servidor puede, o no, pertenecer a la misma red que el cliente.

Si el servidor no está en la red del cliente, dicho cliente debe poder alcanzar el servidor mediante los dispositivos y mecanismos necesarios que interconecten su red a la del servidor.

El cliente es el programa navegador con el que el usuario interacciona y realiza la petición a un servidor Web de envío de páginas de información. Por ello se denomina *cliente HTTP* o *navegador Web*.

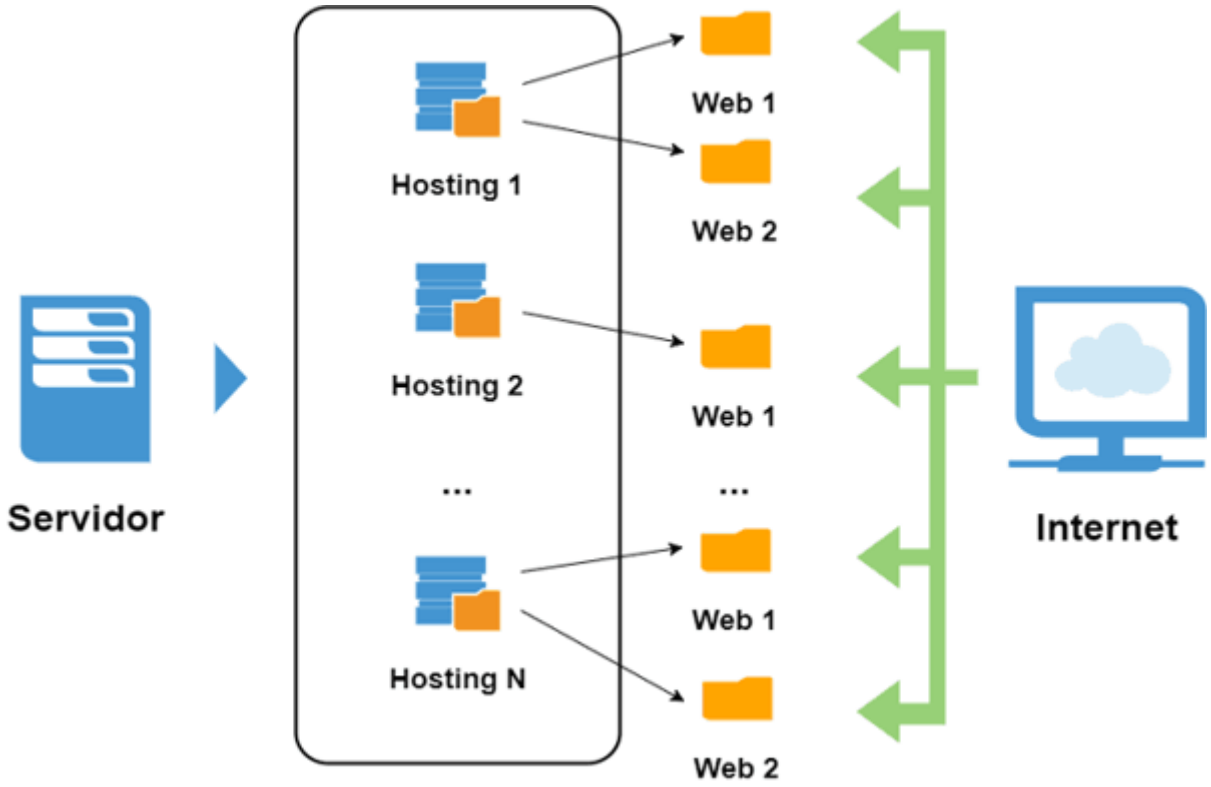


Las páginas que recibe el cliente HTTP son documentos de texto plano, estructurado por etiquetas en lenguaje HTML y que el programa interpretará para poder mostrarlo al usuario con el formato adecuado.

Si lo que recibe no es texto, sino un objeto multimedia (vídeo, sonido, etc.) no reconocido por el cliente HTTP, éste deberá activar una aplicación externa capaz de gestionarlo.

Las páginas que solicita el cliente se almacenan en directorios físicos en el servidor web que reciben el nombre de **sitio web** en sistemas operativos Windows y **hosts virtuales** en sistemas operativos Linux.

Un sitio web o host virtual no es más que una carpeta que contiene las páginas a mostrar y que está configurada por el servidor web para permitir acceder a ella de forma remota mediante un navegador utilizando un nombre de dominio concreto.



3. Funcionamiento

Desde el punto de vista del servidor, su función es atender las peticiones de páginas y otros documentos procedentes de los programas clientes HTTP y enviárselas. El protocolo HTTP es un protocolo sin estado, es decir, no recuerda ningún suceso de conexiones anteriores a la actual.

El esquema de funcionamiento es el siguiente:

- Si el servidor encuentra el documento HTML solicitado por el cliente, lo envía.
- Si no existe dicho documento, se envía una página de error.

En ambos casos, y por ser un protocolo sin estado, al final se libera la conexión.

3.1. Mantenimiento de estado. Cookies

Para resolver las situaciones de "falta de memoria" del protocolo, además de recordar la información de la sesión actual (cuya información se pierde, por ejemplo, al cerrar el navegador), se utilizan las cookies (*galletas*).

Las **cookies** son archivos de texto que se intercambian entre el cliente y el servidor, de modo que, la próxima vez que se solicite un intercambio de información entre estos mismos puntos, se tendrá en cuenta la información de dichos archivos. Su utilización fue propuesta y utilizada por la compañía Netscape y su uso se ha oficializado con el RFC 2109.

3.2. Otras cuestiones de interés

El funcionamiento del protocolo contempla además el acceso por contraseña, caché de documentos, soporte de programas externos y más recientemente, el manejo de transacciones seguras que no deben ser visibles por otros usuarios de la red (por ejemplo, transacciones entre entidades bancarias y sus clientes).

4. Transferencia de páginas Web

La petición de una página web se realiza escribiendo en el navegador su dirección. La estructura de dicha dirección es:

```
protocolo://dirección:puerto/recurso
```

Protocolo	Dirección IP o nombre de dominio	Puerto	Recurso
http	www.iesvillablanca.edu	80	index.html
http://www.iesvillablanca.edu:80/index.html			

4.1. Comportamiento por defecto

Existen una serie de valores o parámetros que con los que se trabaja por defecto. Por ejemplo, el puerto con el que trabaja el servicio HTTP, por defecto, es el 80, es decir, en la anterior URL podríamos haber omitido su indicación.

El cliente HTTP asume por defecto el protocolo (http), así como el puerto (80). En caso de no indicar ningún fichero a servir, asume por defecto que debe servir los catalogados en su configuración como iniciales (index.html, index.htm, index.php, default.html, etcétera).

En caso de no existir ninguna de las páginas iniciales anteriores, se cargará la página de información de error.

4.2. Caracteres codificados

Los caracteres no disponibles en el conjunto de caracteres URL, o los caracteres y secuencias de caracteres que puedan causar confusión, están representados en las direcciones URL con un código. Un carácter codificado aparece como un signo de porcentaje (%) seguido de un número, por lo general en base hexadecimal. Cuando una

cadena contiene caracteres no válidos en el conjunto de URL, como una frase de búsqueda, tiene que ser transformada en una URL y los caracteres no válidos pueden ser codificados. Los programas como los navegadores, suelen hacer esto de forma automática.

Actividad. Investiga sobre los caracteres codificados existentes.

4.3. Flujo de comunicación

Una vez que el usuario especifica en el cliente HTTP la dirección de la página que desea consultar, ocurre lo siguiente:

1) El cliente codifica la información de la URL, segmentando las distintas partes (protocolo de acceso, IP o nombre de dominio del servidor, puerto, etcétera).

2) El cliente establece una conexión con el servidor Web y solicita la página y/o el objeto deseados.

3) El servidor envía dicha página u objeto (en ausencia de éstos, envía una página de error) y el cliente inicia la tarea de interpretación de los códigos HTML.

4) Se cierra la conexión.

5. Tipos MIME

Los **tipos MIME** (*Multipart Internet Mail Extension*) son una forma abierta y extensible de representar el contenido de los datos. También reciben el nombre de IMT (*Internet Media Types*).

Inicialmente se utilizaron para extender las características del correo electrónico mediante los documentos adjuntos, pero en la actualidad se ha generalizado su utilización por los contenidos Web.

El registro de los tipo MIME es controlado por IANA (*Internet Asigned Numbers Authority*) según la RFC 2048. Su función es, entre otras, evitar que dos tipos de contenidos distintos tengan el mismo nombre.

MIME adjunta un fichero de cabecera a los documentos en el que se indica el tipo de contenidos para que el servidor Web y el navegador puedan manejar y mostrar los datos correctamente.

Un MIME está formados por tipos y subtipos con el formato general: `tipo/subtipo`.

Ejemplos:

- `text/html`: define todos los ficheros de texto que contienen código HTML.
- `video/mpeg`: define todos los ficheros de vídeo almacenados en formato mpeg.
- `image/*`: define todos los ficheros de imagen almacenados en cualquier formato (gif, jpeg, bmp, …).

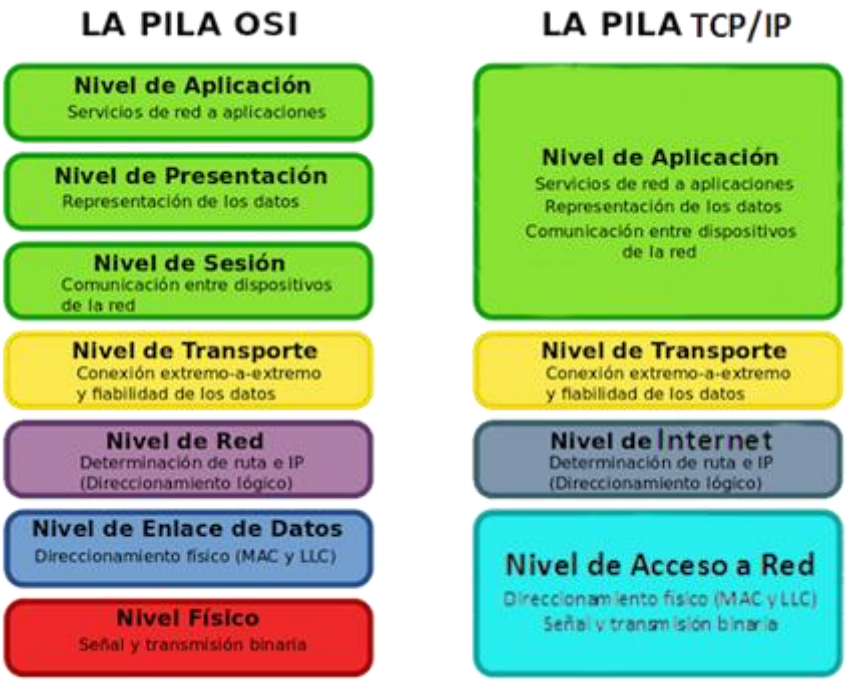
Actividad. Busca todos los tipos MIME que existen, para qué sirve cada uno y piensa de qué forma pueden incidir en tu labor como profesional IT.

6. Seguridad

6.1. Protocolo seguro SSL

El protocolo de encriptación SSL (*Secured Sockets Layer* o *Capa de Conexión Segura*) es utilizado en transacciones seguras vía web entre un cliente y en servidor. Utiliza el puerto 443.

Es un sistema diseñado y propuesto por *Netscape Communications Corporation* y se encuentra en la pila OSI en el nivel de sesión.



Los servicios proporcionados por SSL son autenticación, integridad y confidencialidad, habilitando la posibilidad de utilizar firmas digitales, algoritmos de criptografía, etc.

6.2. Protocolo seguro HTTPS

Uno de los usos comunes de SSL es el de establecer una comunicación web segura entre un navegador Web (cliente) y un servidor web. Es aquí donde se usa HTTPS, que es básicamente HTTP sobre SSL, con un esquema de invocación por medio de URL.

La utilización de HTTPS no invalida la opción de utilizar HTTP, ya que los navegadores, de forma estándar, avisan del paso de uno a otro protocolo.

6.3. Certificados de servidor

Las páginas web seguras son accesibles mediante el protocolo HTTPS sobre SSL, en vez de utilizar el habitual HTTP.

Los certificados de servidor son necesarios para poder definir páginas web seguras en un servidor determinado y son emitidos por entidades certificadoras.

Una entidad certificadora o servidor de certificados es una institución que emite certificados y se compromete a certificar como auténticos los certificados emitidos por otras entidades certificadoras subordinadas, en las cuales ha expresado de forma explícita su confianza.

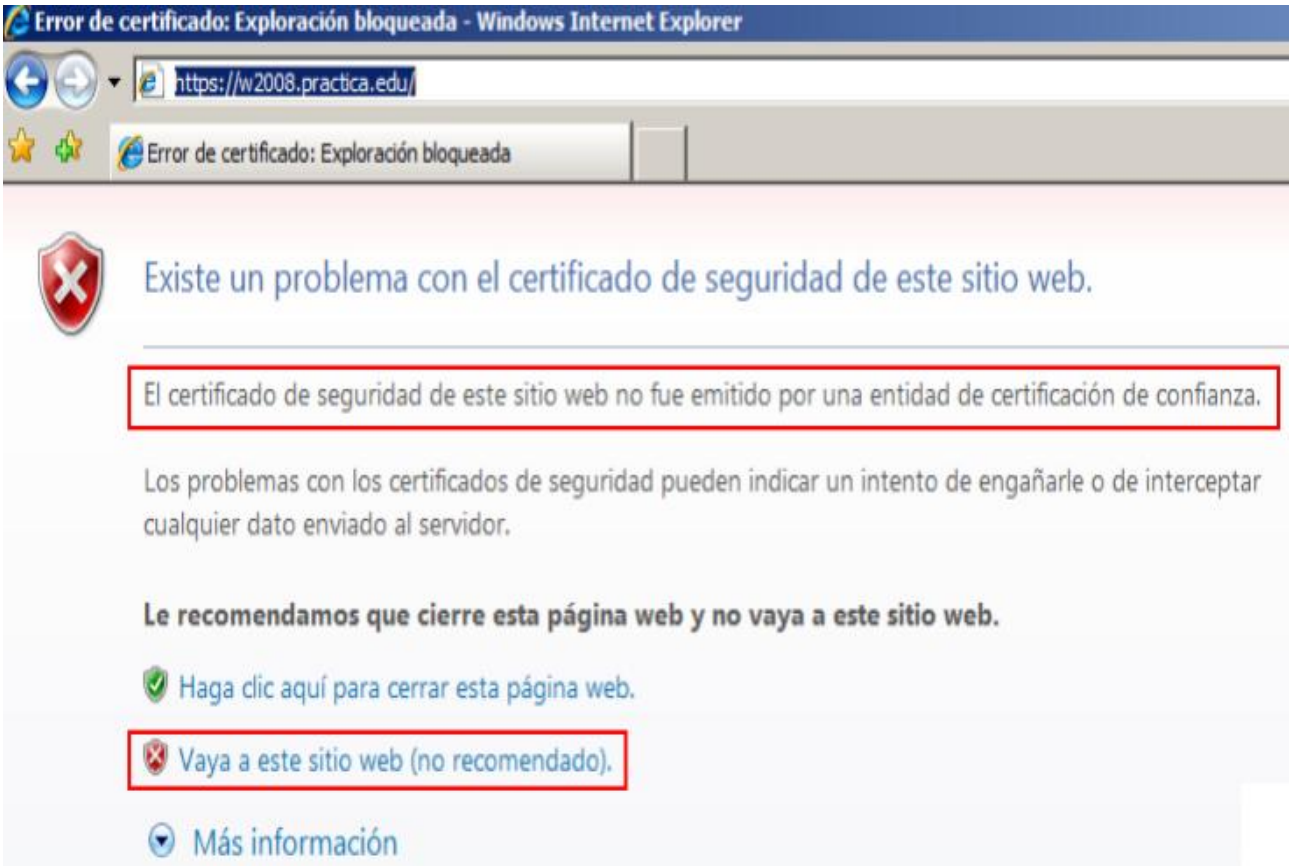
Debemos utilizar una página web segura (y por tanto debe accederse vía HTTPS) cuando los datos que van a enviarse a través de Internet por medio de ella son críticos en cuanto a su seguridad, como páginas donde debemos introducir nuestra contraseña, el número de nuestra tarjeta de crédito, etc.

Cuando un usuario desea navegar por una página segura, el servidor al cual está accediendo le envía un certificado de servidor, para indicarle que toda la información que va a ser transferida no podrá ser interceptada ni leída por terceros.

La única referencia que tendrá el usuario de estar en una página Web segura, será el nombre del protocolo y, en algunos casos, un pequeño icono específico de seguridad que se le mostrará en el navegador.

Si la entidad que presenta el correspondiente certificado al equipo del usuario está reconocida como una entidad de confianza en el navegador del equipo del cliente, se mostrará directamente la página.

Si en el navegador del equipo cliente no se tiene definida de forma explícita confianza en la entidad certificadora que presenta dicho certificado, se mostrará una ventana en la cual el usuario podrá dar su aprobación o no al proceso de envío de datos a ese servidor.



Podemos utilizar la entidad certificadora de un servidor Windows para certificar páginas web seguras en el servidor web IIS mediante el protocolo SSL y accediendo por HTTPS. El programa que me permite emitir y gestionar los certificados para IIS estará integrado en Active Directory.

También podemos utilizar el programa *OpenSSL* basado en software libre para la creación de certificados.

En cualquier caso, nuestra entidad certificadora no será válida en Internet y únicamente tendrá vigencia en el ámbito de nuestra red local, pero es suficiente para el fin pedagógico que pretendemos.

7. Servidor web

El servidor web atiende las peticiones recibidas de los navegadores (cliente web) y sirve dichas peticiones con la mayor eficiencia y seguridad.

La ubicación lógica de un servidor web es Internet, pero también se utilizan dentro de una Intranet o red local.

El servidor web, inicialmente, solamente suministraba páginas web estáticas, pero hoy en día permiten la ejecución de pequeños programas en diferentes lenguajes (PHP, Servlets, CGI, JavaScript, etc.) que proporcionan dinamismo en las páginas.

En la página web *http://news.netcraft.com* se publica información relativa a los servidores web más utilizados en Internet, siendo los más utilizados: IIS, Apache y nginx.

7.1. Apache

Es un servidor basado en software libre (proyecto GNU). Su facilidad de configuración, sus prestaciones, la posibilidad de instalarse sobre diversas plataformas y el software libre y gratuito, lo han convertido en líder en Internet.

7.2. IIS (Internet Information Services)

Fue desarrollado por Microsoft para sus servidores con sistemas Windows (NT/2000/2003/2008), aunque también funciona bajo Windows XP/Vista/W7/W10.

En la actualidad es soporte de un gran número de servidores de la Web, pero superado por Apache.

7.3. nginx

Es un servidor web ligero de software libre y código abierto. Originariamente fue diseñado para satisfacer las necesidades del sitio web de Rambler (potente motor de búsqueda ruso con más de 500 millones de peticiones al día).

En la actualidad, ha sufrido un ascenso que le permite destacar como el servidor web más utilizado en Internet (Diciembre 2019) y entre sus clientes más importantes está Facebook.

8. Navegadores web

El navegador web es quien realiza la labor de cliente HTTP, convirtiéndose en el punto con más responsabilidad en el intercambio de información de la Web.

Los navegadores Web más utilizados son:

- Mozilla Firefox

- Iceweasel

- Google Chrome

- Opera

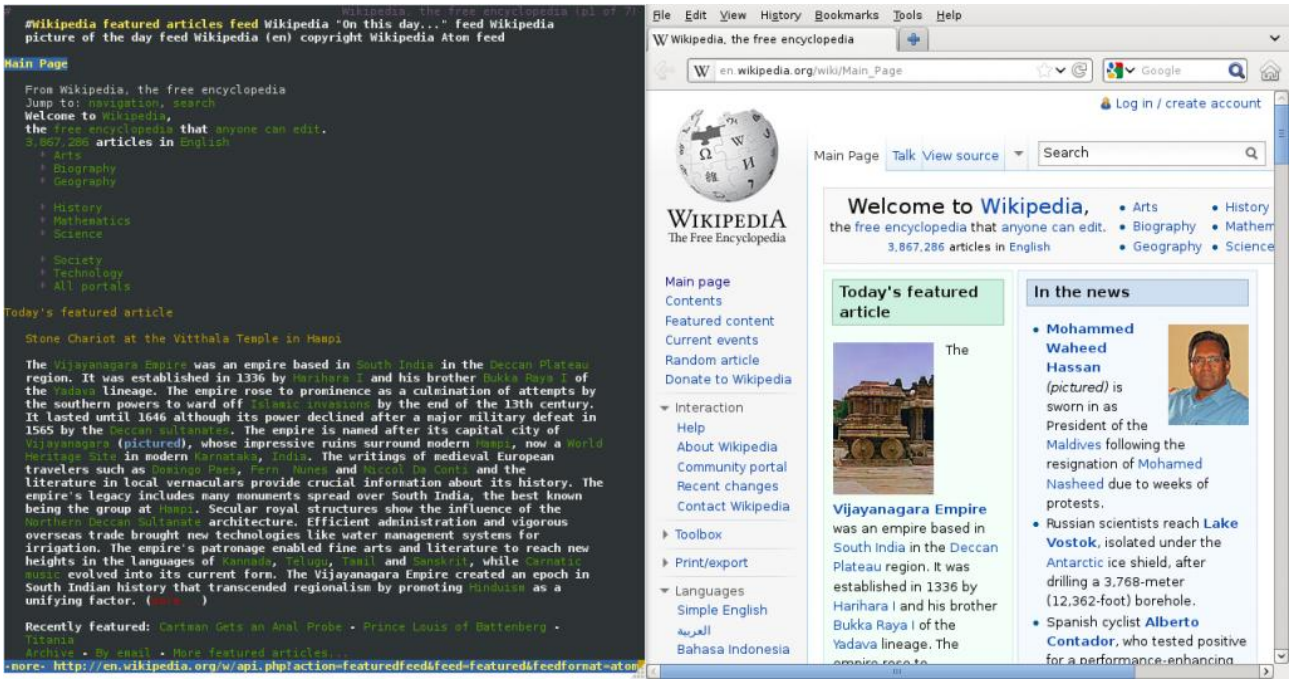
- Internet Explorer / Microsoft Edge

- Safari

Uno de los aspectos más importantes es el hecho de que un cliente HTTP esté disponible para diversos sistemas operativos, de modo que el cambio o alternancia de sistema del usuario no suponga un cambio de funcionamiento. Navegadores como *Mozilla Firefox* o *Google Chrome* cumplen esta condición.

La ausencia de entorno gráfico no supone impedimento para no poder utilizar clientes HTTP, pasando a utilizar navegadores en modo texto.

Los primeros clientes fueron de este tipo al no existir entornos gráficos. Al no realizar tanto intercambio de información (únicamente texto) la navegación Web con este tipo de clientes o navegadores tiene un gran rendimiento, aunque hoy en día no se utilizan. Uno de los principales navegadores modo texto es Lynx.



Como interfaz de usuario, un navegador puede ser configurado de modo que su apariencia sea la más acorde con las preferencias del responsable de su manejo.

Todos los navegadores poseen modos de configuración, donde podemos indicar parámetros de seguridad como aceptación o no de cookies, contraseñas guardadas en caché, historial de sitios visitados, versión de SSL admitida, tamaño de letra, etc.

Como regla general, los navegadores gráficos suelen poseer un menú llamado Opciones o Preferencias, desde donde podemos configurar estos valores.