

Introducción a creación páginas web

CURSO 2019-2020

Índice de contenidos:

INTERNET	2
PROTOCOLOS DE INTERNET	2
Familia de protocolos de internet.....	2
PROTOCOLO TCP/IP.....	3
Dirección IP Y Dominio.....	4
Dirección IP	4
Nombre de dominio (DN) (Domain name)	4
DNS (Domain name system).....	5
Servidor DNS	5
World Wide Web.....	6
Protocolo HTTP y la URL	6
Protocolo HTTP	6
URL (Uniform Resource Locator)	7
HARDWARE, SOFTWARE y LENGUAJES EN LA WEB.....	8
Alojamiento web (Hardware).....	8
Servidor Web o servidor Http (Web server) (Software)	8
Cliente web (Hardware).....	8
Navegador (Web Browser) Software.....	9
Lenguajes de marcas (Etiquetas) (Markup Language).....	9
HTML (Hypertext Markup Language)	9
El Flujo de contenido	9
Versiones de HTML.....	10
La Web 2.0 y la Web Semántica	10

INFORMACIÓN IMPORTANTE:

A partir de este momento vamos a comenzar a programar. Para ello nos apoyaremos en una plataforma que nos permitirá un aprendizaje individualizado. De esta manera, todo el mundo podrá avanzar a su ritmo, empleando el tiempo que considere necesario tanto en clase como en casa. La idea no es avanzar y entregar, sino que estas prácticas entendiendo lo que hacemos, nos permitirán realizar una prueba en la que no deberemos estudiar, solo recurrir a las notas que iréis tomando en clase. Concretamente en este documento. Antes de nada, necesitamos una introducción teórica que reforzará nuestros conocimientos de cómo funciona la Internet y la WWW.

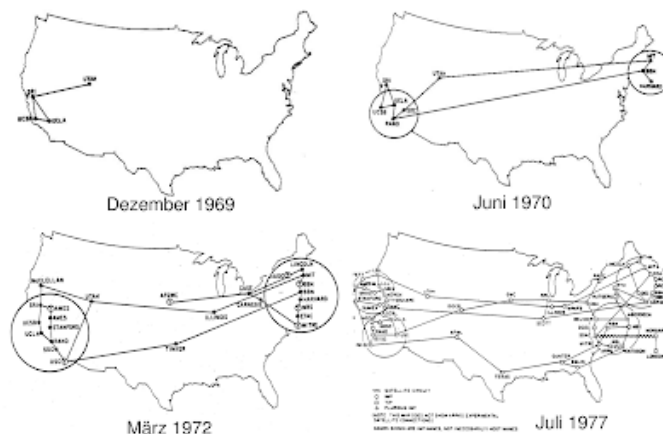
Conceptos básicos que debemos estudiar (examen)

INTERNET

Internet es el **conjunto descentralizado de redes** que utilizan protocolos que garantizan que las redes físicas que la componen, que son **heterogéneas y tienen localizaciones geográficas distintas**, formen una única red lógica de alcance mundial.

El origen de internet data de **1969** al establecerse la primera conexión conocida como **ARPANET (Advanced Research Projects Agency NETWORK)** entre tres redes físicas en ubicaciones geográficas distintas.

Dicho lo anterior, Internet es una red en la que podemos acceder a sub-redes, de ahí que la llamen la **red de redes**. A partir de la creación de Internet se comenzaron a desarrollar servicios y protocolos con distintos objetivos.



Evolución de La Internet en sus primeros años de vida.

PROTOCOLOS DE INTERNET

Familia de protocolos de internet

Se trata de un conjunto de protocolos de red en lo que se basa internet y permite la transmisión de datos entre computadoras.

Se **estructura en capas** en las cuales existen diferentes protocolos. A continuación, se especifican las capas y los protocolos más conocidos.

- **Capa de interfaz de red (Ethernet)(Concepto de conexión entre computadoras)**
 - El objetivo de los protocolos que existen en esta capa es que la información fluya en entre **máquinas conectadas directamente**. En este caso al contrario que el IP se trata de protocolos orientados a la conexión.
- **Capa de red internet (IP)(Concepto de no conexión entre computadoras)**
 - Es una capa en la que se sitúan los **protocolos para la conectividad y la selección de la mejor ruta** entre dos sistemas situados **en redes geográficamente distintas**.
 - El objetivo de esta capa es que los datos lleguen de extremo a extremo **sin que exista conexión directa**.

- Se puede decir que estos protocolos sirven a los de transporte.
- **Capa de transporte (TCP, DCCP, UDP)**
 - Se encarga de la **transferencia libre de errores de los datos entre emisor y receptor y mantienen el flujo de la red**. Es la base de toda jerarquía de protocolo. Alguno como el TCP está orientado a la **conexión confiable y otro como el UDP está orientado a una conexión no confiable**. Aquí se puede intuir cual usarán los programas tipo Torrent.
- **Capa de aplicación (SSH, FTP,SMTP,POP,DHCP,DNS,HTTP)**
 - Ofrece a las **aplicaciones acceder a los servicios de las demás capas** a través de sus propios protocolos (Los mencionados)
 - Estos protocolos también sirven a los de transporte.

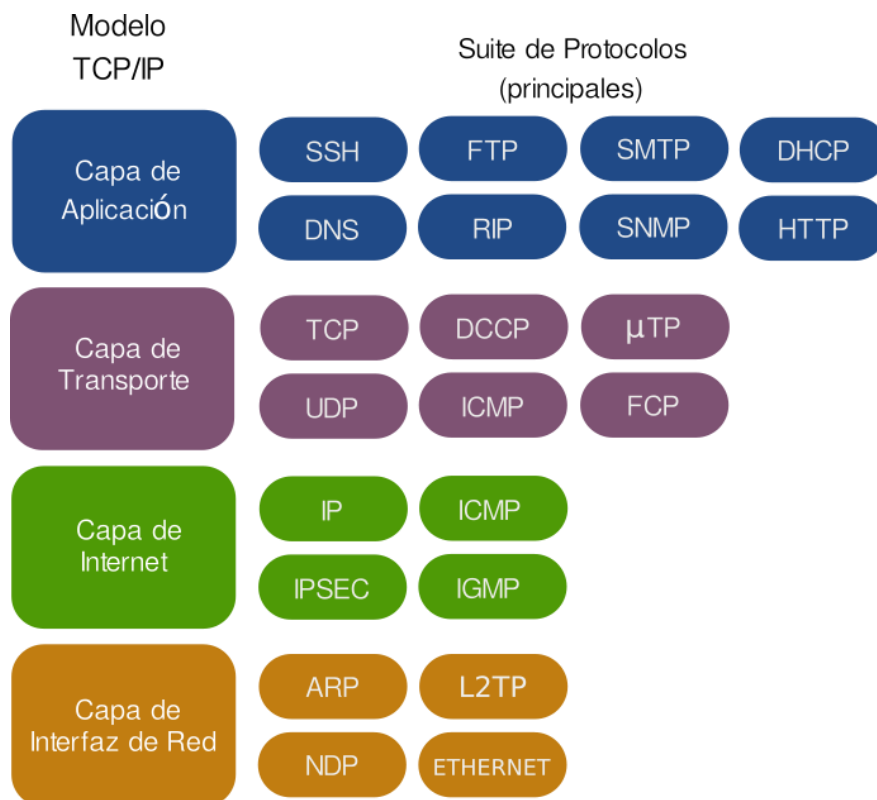


Gráfico de las distintas capas de la familia de protocolos usados en internet

PROTOCOLO TCP/IP

Generalmente se habla de los protocolos **TCP/IP** en referencia a **los dos más importantes** porque fueron los dos primeros en definirse y que son los más utilizados de la familia.

- **Protocolo TCP:** Se trata del protocolo de **control de la transmisión**.
 - Se asegura de que los datos sean transmitidos por completo y se diseñó para proporcionar un **flujo de bytes confiable de extremo a extremo a través de una interred no confiable**.
 - (Es interesante reflexionar sobre las fallas actuales de seguridad en protocolos SAMBA de intrared)
- **Protocolo IP:** Se trata de un **protocolo de comunicación de datos digitales** en la red de redes o Internet. Como aquellos que se encuentran en la capa de red (visto anteriormente), se trata de un protocolo **no orientado a la existencia de una conexión entre extremos**. Esto quiere decir que está diseñado para realizar una **comunicación entre dos puntos de una red sin que exista acuerdo previo para realizar dicha comunicación**. Actualmente estamos en la versión 6 **IPv6**

Dirección IP Y Dominio

Dirección IP

Identifica de manera lógica y jerárquica el elemento que está conectado a la red y que utilice el **IP (Internet Protocol)** o los **protocolos de (ETHERNET)**. No se debe confundir con la dirección **MAC** (Que es un número físico asignado a la tarjeta o dispositivo de red impuesto por el fabricante). Las **direcciones IP** tanto en intranet como en internet pueden ser del tipo: **IP fija/IP dinámica**. En las IP dinámicas debe existir un **servidor(DHCP)** que es el que **otorga las direcciones IP**.

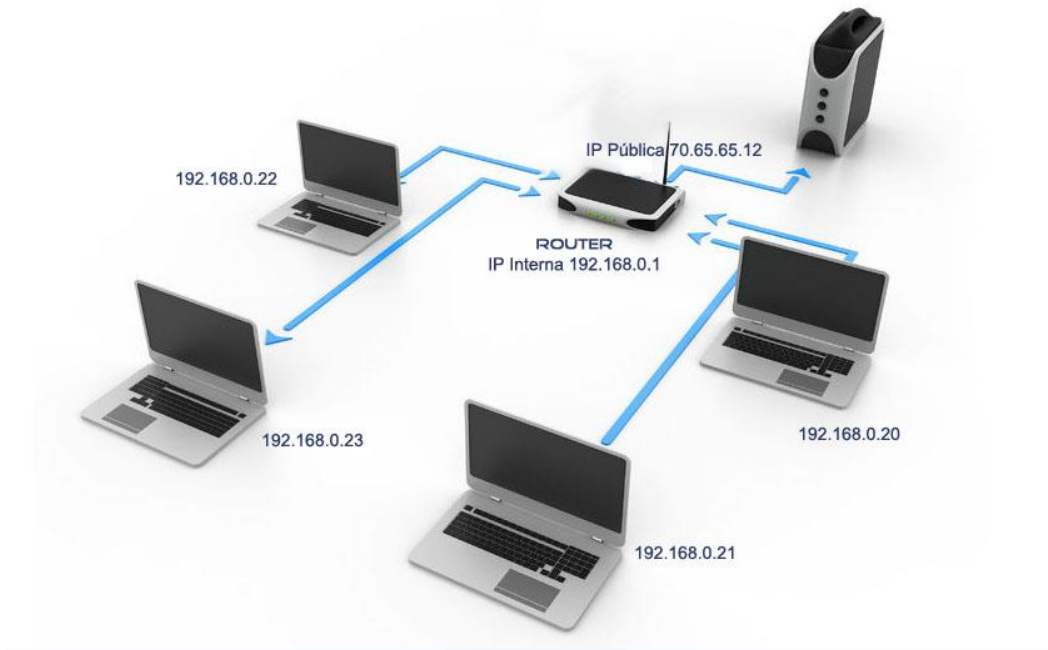


Imagen donde se Puede ver las direcciones IP (internas) y la dirección IP Pública. En una red con acceso a intranet con acceso a internet

Nombre de dominio (DN) (Domain name)

El **nombre de dominio** sirve para **asociar un nombre inteligible por los humanos a una dirección IP**. Para la **traducción** de nombres de **dominio** a **dirección IP** se **utiliza el DNS (Domain Name System)** del que hablaremos en el próximo apartado.

El siguiente ejemplo de lo que comúnmente es denominado un **dominio**: **www.microsoft.com**

Sin embargo, esto no es así. En realidad **www.microsoft.com** es la **dirección** de un **host** o **hostname**. Si descomponemos ese hostname:

www es un **subdominio** de **microsoft.com**

microsoft.com es el **nombre del dominio (Domain)**

com es un dominio de más alto nivel (**TLD**)

En cuanto al dominio de más alto nivel(TLD o Top Level Domain) nos encontramos con algunos como:

- gov - Government agencies
- edu - Educational institutions
- org - Organizations (nonprofit)
- mil - Military
- com - commercial business

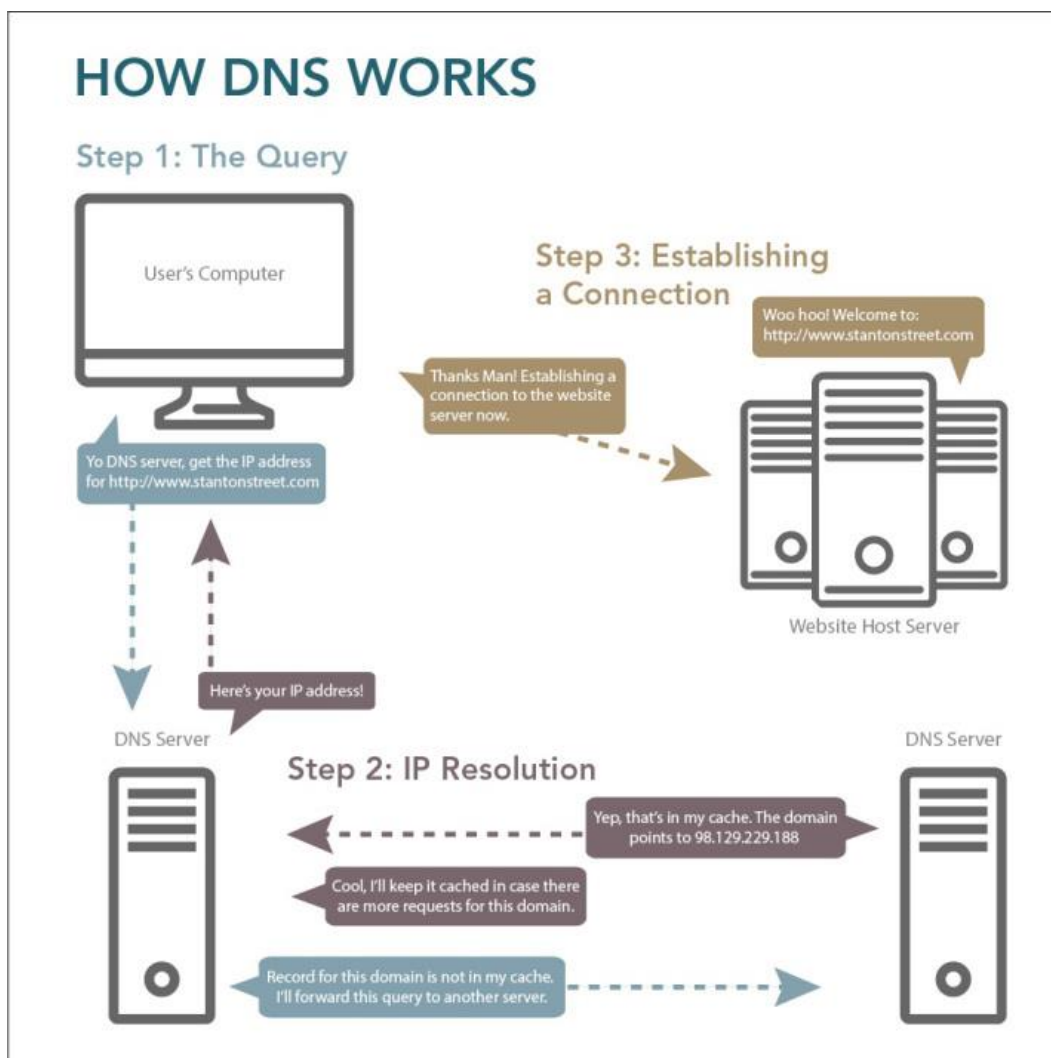
DNS (Domain name system)

Se trata el sistema de nombres de dominio (**Domain name system**). Se emplea para traducir nombres inteligibles para las personas (**Dominios**) en identificadores binarios asociados con los equipos conectados a la red (**Direcciones IP**) para lograr localizarlos y direccionar estos equipos.

Servidor DNS

Los servidores DNS realizan la traducción del **dominio** a la **dirección IP** cuando desde el cliente se pide dicha traducción. Un **servidor DNS** es un programa que se ejecuta en una máquina o host (como ocurre con los **servidores HTTP**). Cuando nuestro ordenador (cliente) quiere traducir una **dirección IP** para la que se ha introducido un **dominio**, este envía la petición a un **servidor DNS** predeterminado (Existen servidores DNS que deben evitarse, ya sea por saturación de clientes o por lentitud propia del hardware (host) donde están instalados. De lo contrario, pese a tener una buena conexión a internet y un hardware actualizado, nuestra experiencia de conexión puede ser lenta.

En muchas ocasiones, la tarea de traducción de dominios se lleva a cabo con varios servidores. Esto se ejemplifica en la figura inferior. Cuando el primer **servidor DNS** consultado no conoce la traducción del **dominio** a **dirección IP**, este se conecta a otro **servidor DNS** para obtener la respuesta. En todos los casos la traducción, es decir la **dirección IP**, vuelve al cliente para que pueda conectar con el equipo pretendido al introducir el **dominio correspondiente**.



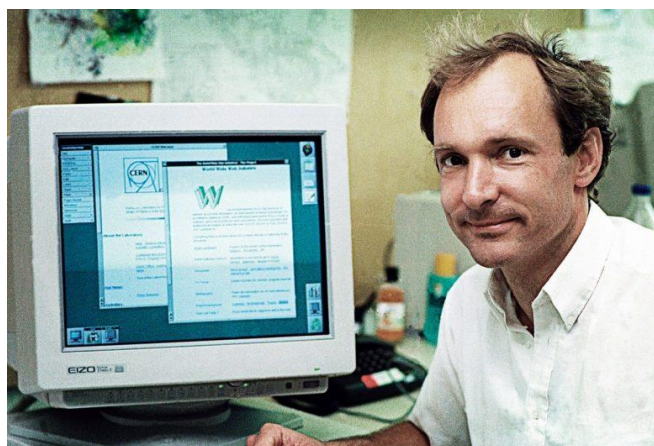
Esquema de funcionamiento de los servidores DNS cuando el navegador le pide que traduzca el dominio que hemos introducido en el navegador

IMPORTANTE: ESTO ES TODO LO QUE NOS INTERESA A NIVEL INTERNET. A PARTIR DE AQUÍ NOS CENTRAMOS EN LA WEB

World Wide Web

Unos años después de la creación de internet **1991** se desarrolló en el **CERN** por **Tim Berners-Lee** y otros investigadores, un servicio llamado **WWW (World Wide Web)**. Generalmente, el contacto del usuario con internet es a través de este servicio, las famosas páginas web. Debido a esto, en muchas ocasiones se confunde Internet con la WWW.

La World Wide Web no engloba a Internet, es tan solo el conjunto de protocolos concretos para la consulta de archivos de hipertexto o hipermedia. Estos archivos de hipertexto o de hipermedia incorporan un contenido propio junto con enlaces a otros archivos de hipertexto u contenidos audiovisuales localizados en otras redes. Esto se lleva a cabo a través de **vínculos** a esos otros archivos y contenidos, lo que genera una tela de araña de **información/contenidos**. Los archivos de hipertexto se leen con los **Navegadores (Browsers)** y al existir esta tela de araña de información descentralizada en la que nos podemos mover a través de los hiperenlaces y el acceso a nuevos dominios, decimos que **navegamos por la WEB**.



Tim Berners-Lee, desarrollador de la World Wide Web (No confundir con La Internet).

Primera página WEB creada en el CERN

<http://info.cern.ch/hypertext/WWW/TheProject.html>

- **www (World Wide Web)**
 - Se trata del sistema de distribución de documentos de hipertexto o hipermedia interconectadas y accesibles vía internet.
 - Se desarrolló en el **CERN** entre **1989 y 1990** por el inglés **Tim Berners-Lee**.
- **World Wide Web Consortium (W3C) (Se creó en 1994)**
 - Consorcio que se encarga de generar recomendaciones y estándares que aseguran el mantenimiento y crecimiento de la WWW.
 - Dirigido por el mismo **Tim Berners-Lee**

Protocolo HTTP y la URL

Protocolo HTTP

Cuando hablamos de capas dentro de la familia de protocolos de internet, existía una dedicada a las aplicaciones. Los protocolos contenidos en esa capa sirven para ayudar a las aplicaciones a realizar sus comunicaciones en internet (Usando los protocolos **TCP/IP**). En el caso del **protocolo HTTP**, **permite que los mensajes entre el navegador y el servidor HTTP se formateen y transmitan, y establece las acciones que los servidores web y los navegadores deben realizar como respuesta a estos mensajes y a otros comandos**. Por ejemplo, cuando incluimos una dirección web en nuestro navegador y apretamos la tecla intro, este envía una solicitud (**comando GET**) al servidor para que este envíe un archivo que por defecto es el **index.html (Archivo de Hipertexto)**.

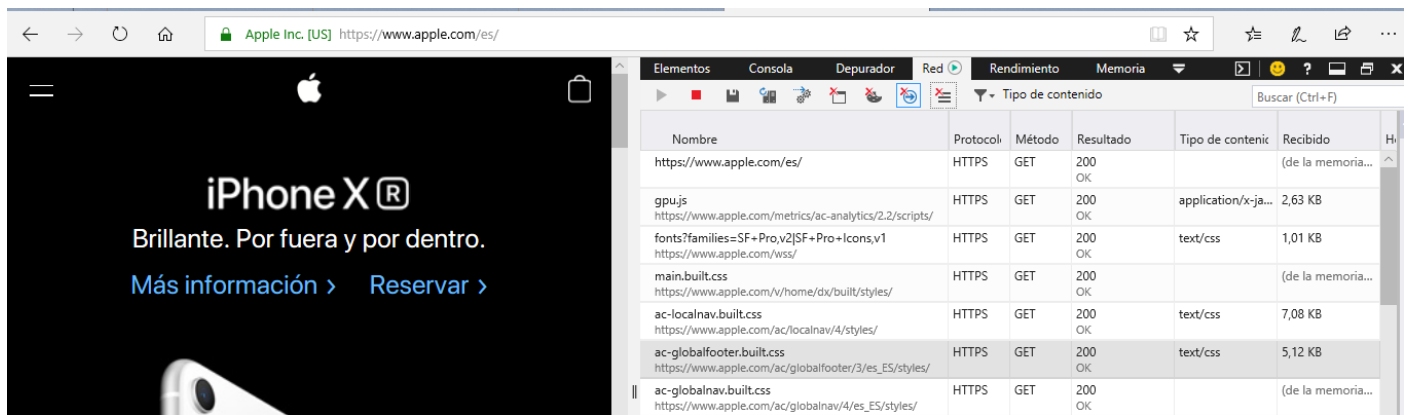


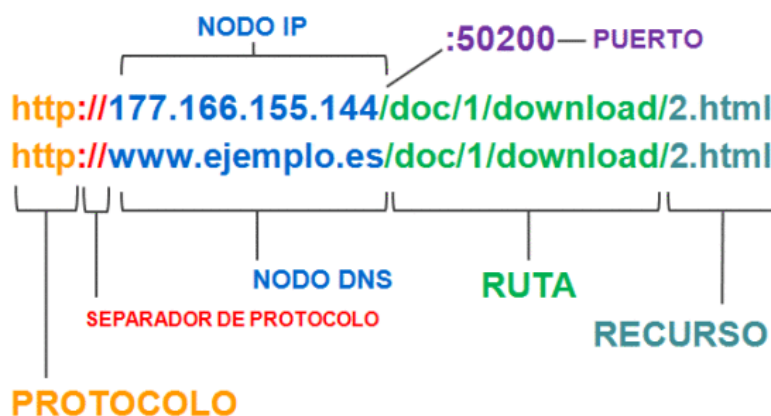
Imagen de las herramientas de desarrollador de EDGE. Se puede ver como interacciona el navegador con el servidor a través del protocolo.

URL (Uniform Resource Locator)

El **localizador uniforme de recursos (URL)**, se emplea para lo que su propio nombre indica, **localizar un recurso**. Entendemos por recurso lo siguiente:

- Un documento de **hipertexto** por ejemplo una página escrita en **html**
- Un archivo cualquiera (**imagen, archivo de audio, un pdf..**).
- También Una **carpeta de archivos**

Se trata de una **serie de caracteres con la cual se asigna una dirección única a cada uno de los recursos de información disponibles en Internet y combina los siguientes elementos:**



El esquema/Protocolo: indica el **protocolo de red** que se usa para recuperar la información del recurso identificado

- **http** (Recursos Hypertext Transfer protocol HTTP)
- **https** (Recursos Hypertext Transfer Protocol sobre Secure Sockets Layer (SSL). O lo que es lo mismo, la HTTP segura.
- **mailto** (Direcciones de correo electrónico)
- **ftp** (File transfer Protocol)
- **file** (Recursos disponibles en el sistema local)
- hay otros con los que no nos extendemos (**news, ldap, gopher, telnet, data**)

El Host o Hostname (nodo): Dirección de la computadora que proporciona la información (**Dirección IP**). Esta dirección sería la que nos lleva al **computador (host)** donde **se aloja el servidor (software)** con el que nos comunicamos para obtener y enviar información. Puede usarse la **dirección IP** o el **nombre de dominio** (que

tendrá que resolver el **servidor DNS** en la **dirección IP correspondiente** como habíamos visto). El dominio **www.apple.es** se identifica por el Dominio de mayor nivel **TLD “es”** y el subdominio **“www”**.

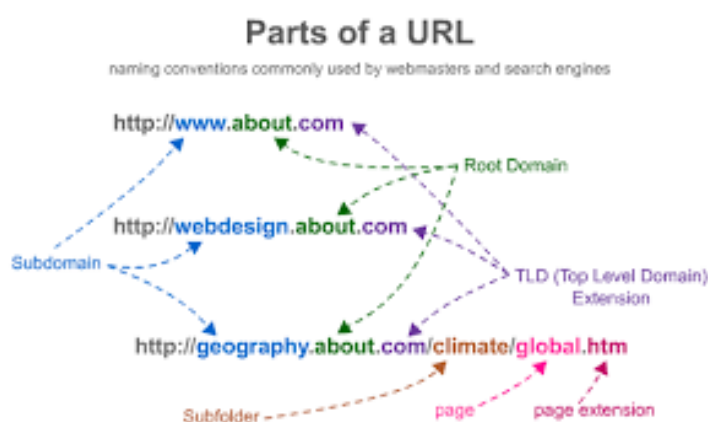
Ruta al directorio donde se encuentra el recurso dentro de la **computadora (host)** que **aloja al servidor**.

El nombre del recurso

Adicionalmente, puede incluir:

- Usuario y contraseña para el acceso al recurso
- El puerto a través del cual se accede

Otro ejemplo:



HARDWARE, SOFTWARE y LENGUAJES EN LA WEB

Alojamiento web (Hardware)

El **Web Hosting** es un servicio que ofrecen algunas empresas para el alojamiento web, ya sea de pago o gratuito (con publicidad).

<https://www.hostinger.es/> es uno de ellos por ejemplo. Dada la gratuidad y sencillez para subir nuestros archivos .html será el que usemos para subir nuestras páginas web.

Host a secas. Se trata de un **ordenador** donde está instalado el **servidor HTTP** y al cual apunta la **dirección IP** que corresponde al **hostname** correspondiente. Recordamos que el **hostname** (www.microsoft.com) incorpora el nombre de **subdominio** (www) y el **nombre de dominio** (Microsoft.com) que se traduce en una dirección reconocible por el Protocolo de Internet (IP) a la que se le solicitará la información.

Servidor Web o servidor Http (Web server) (Software)

Programa instalado en el **Host** (ordenador físico) que contiene la información de página Web.

Cuando nuestro navegador (**cliente**) ha localizado al **host** (donde se instala el servidor), se establece una conexión entre ambos empleando el protocolo HTTP. Una ejemplificación de esta conexión es la siguiente:

```
GET /index.html HTTP/1.1 HOST: www.host.com
```

El navegador **solicita al host**, con **hostname**: www.host.com y usando el **comando GET**, el recurso **index.html**

Cabe destacar que **en cualquier ordenador se puede instalar un servidor web**, convirtiendo este ordenador en un hosting o alojamiento web.

Cliente web (Hardware)

Se trata de **nuestro ordenador, teléfono móvil, tablet** o cualquier otro dispositivo donde instalemos un Software para la **obtención y renderizado de páginas WEB**, comúnmente llamado **Navegador Web (Web Browser)**

Navegador (Web Browser) Software

- Tiene por objetivo **integrar las comunicaciones con el servidor**.
 - Se comunica mediante **protocolo HTTP con el Servidor HTTP**
 - En su funcionamiento más básico, cuando introducimos una URL en él, este envía un requerimiento (**GET**) al servidor alojado en la **dirección IP** que corresponde al **dominio** introducido en la **URL**. El servidor envía de vuelta el archivo **index.html** (**siempre que no hayamos especificado otro**).
- Se encarga de la **seguridad**
 - **HTTPS** (Secure) Actualmente el protocolo HTTP se ha sustituido por este, garantizando la seguridad durante conexión y envío.
 - **CORS** (Protocolo de seguridad al ejecutarse el navegador)
- Se encarga de la **optimización de recursos**
 - **Compresión**
 - **Caché, etc...**
- Se encarga de **RENDERIZAR** en la pantalla del cliente (nuestro ordenador) los contenidos con la estructura y el formato definido en el archivo **.html**.

Lenguajes de marcas (Etiquetas) (Markup Language)

Se trata de la codificación de un documento con etiquetas que suponen una información adicional con diferentes objetivos. Principalmente son dos:

- Estructurar el texto
- Presentar el texto.

HTML (Hypertext Markup Language)

Se trata del lenguaje de programación de etiquetas (**creo que es más conveniente hablar de lenguaje de codificación mediante etiquetas**) que se empleó desde el comienzo de la WEB para realizar **páginas web**. Los documentos que se crean, archivos **que sólo contienen texto**, se codifican con la extensión **.html** y, como se había dicho antes, **son los que se requieren al servidor web** cuando introducimos el **hostname** junto a la **ruta** a la carpeta del disco duro **donde se aloja la página Web** o lo que es lo mismo, a través de una **URL**.

El texto que se incluye en un archivo **html** será en una parte el contenido (Por ejemplo, un texto de un artículo de blog y en otra, las etiquetas y atributos que permiten estructurar ese contenido. Si queremos que cuando se visualice nuestra web esta incorpore audio, imágenes vídeo u otros, deberemos hacer referencia a ese contenido. Esto es porque no se incluye en el archivo **.html** nada más que texto (nuevamente, la manera de indicar al navegador qué incluir y cómo incluir será a través de etiquetas y atributos de las mismas). En el siguiente ejemplo podemos ver cómo se introduce una imagen en una página web:

```

```

Nótese que, como se viene repitiendo hasta ahora, la inserción de una imagen corresponde a una línea de texto. La imagen se encontraría alojada en la ruta a donde apunta el **hostname** y el navegador que abra el archivo html se la pedirá al servidor usando de nuevo protocolos HTTP. En el caso expuesto, la ruta es la raíz de la carpeta donde se encuentra el archivo **index.html**, pero esta podría estar alojada en cualquier otro servidor de internet y tendría que ser referenciada con su URL. Esta se introducirá en el lugar de la página web donde se quiera colocar la imagen.

```

```

El Flujo de contenido

La programación de documentos HTML no dispone de funciones y variables, y tampoco se rige por un algoritmo (**secuencia de operaciones**). Frente a un lenguaje algorítmico, en este lenguaje (recordamos que es mejor llamarlo de

codificación) se genera un **flujo de contenido estructurado y formateado** a medida que se escribe y codifica con etiquetas (marcas). Estas etiquetas permiten variar el flujo en la pantalla, darle estilos o agregar los enlaces a los recursos (imágenes, vídeo, etc.) que se desea incluir en **ese flujo**.

El lenguaje HTML fue concebido para incorporar **etiquetas de estructura y formato**. Esto fue así durante las primeras versiones de este estándar. Al aumentar el tamaño y complejidad de las Web, tal y como las conocemos hoy, el lenguaje se convirtió en inoperativo, por lo que el lenguaje HTML se dejó para dar estructura y contenido, y se crearon alternativas para el formato. En la actualidad, el formato y apariencia (**estilos**) de una página se codifican con **CSS (Cascade Stylesheet)** dejando al Html para la estructura y la semántica (¿semántica?, en apartados posteriores entraremos en qué es la semántica).

En ese camino hacia la **Web 2.0** en la que el cliente comenzaba la interacción con la Web, surge la necesidad de emplear lenguajes de programación como apoyo a los anteriores (El lenguaje de programación, sigue un algoritmo y ejecuta unas acciones cuando el usuario interactúa con la página web. **Java, Flash, Javascript** son algunos de ellos. Los escasos recursos computacionales de los ordenadores de generaciones anteriores y la baja calidad de las redes afectaron significativamente al desarrollo y empleo de estos lenguajes. Hoy en día el más utilizado es **Javascript**.

En definitiva, en la última versión del HTML, **HTML5**, se relegan todas las tareas de formato a otro lenguaje (**CSS**) y se complementa con un lenguaje de programación (**Javascript**) para poder convertir la web en dinámica (interacción del usuario/cliente con el servidor)



Logos de los tres lenguajes que algunos autores engloban en el HTML5

Versiones de HTML

En estos apuntes se utiliza como referencia el HTML del W3C, simplemente para poder tener una referencia fija. Actualmente (octubre de 2018), la referencia es la última versión disponible, el HTML 5.2, publicada en diciembre de 2017.

Con respecto a la sintaxis, a partir del curso 2018/19 se utilizará la sintaxis HTML, pero sin omitir las etiquetas opcionales permitidas por la recomendación.

La Web 2.0 y la Web Semántica

Para finalizar esta introducción a la WWW, cabe destacar que actualmente nos encontramos en la llamada versión 2.0. Además, el consorcio **W3C** pretende la implantación de una nueva **Web semántica, la web interpretable por las máquinas**.

La **Web 2.0** se trata de una evolución la **Web 1.0** en la que nos encontrábamos con páginas web de contenido estático (solo consulta) y que, a medida que se produjo una mayor interacción del cliente con la Web, aparecieron nuevas aplicaciones para usuarios, comunicaciones y todo ello enfocado a las comunidades.

Según se ha definido la **Web 2.0** su funcionamiento requiere de los usuarios humanos, que mediante un terminal (cliente), realizan las tareas que todos cono

ceamos como buscar, encontrar y agregar información. Podemos limitar por tanto el concepto de **Web 2.0 a un producto dedicado exclusivamente a los humanos**.

Por otro lado, la **Web Semántica** cuyo **creador** ha sido de nuevo **Tim Berners-Lee**, tiene por objetivo que las páginas web también pueden ser manipuladas por los computadores. Se trata de un proyecto con el que cambiar la manera en la que se presentan los datos en una página Web, **incluyendo información que describa el contenido, su significado y la relación de este**. Gracias a la Web semántica serán los ordenadores los que realicen búsquedas, generen y combinen la información contenida en la Web sin un operador humano.