

T1.2. Teoría: Introducción

Historia

El pasado

Los lenguajes de marcas son bastante antiguos aunque solo se han popularizado con la llegada de Internet. De hecho, los comienzos de los lenguajes de marcas se pueden situar en el lenguaje SGML (Standard Generalized Markup Language).

Dicho lenguaje en realidad nunca llegó al gran público, aunque sigue existiendo y usándose en entornos muy específicos. Lo más usado y conocido hoy día es HTML (HyperText Markup Language).

El presente

El lenguaje HTML ha ido cambiando mucho a lo largo del tiempo. La versión más actual es la conocida como HTML5 y es la que se contará en estos apuntes. El fundamento de Internet y todas las tecnologías asociadas se basa en estándares abiertos e independientes de la tecnología.

El organismo que regula estos estándares sin ninguna contrapartida a cambio es el World Wide Web Consortium (W3C). Todos sus estándares (que ellos llaman «Technical Reports») pueden descargarse gratuitamente en su web.

El futuro

El futuro de todas estas tecnologías sigue estando en sistemas abiertos, que facilitan la competencia y por tanto benefician al usuario.

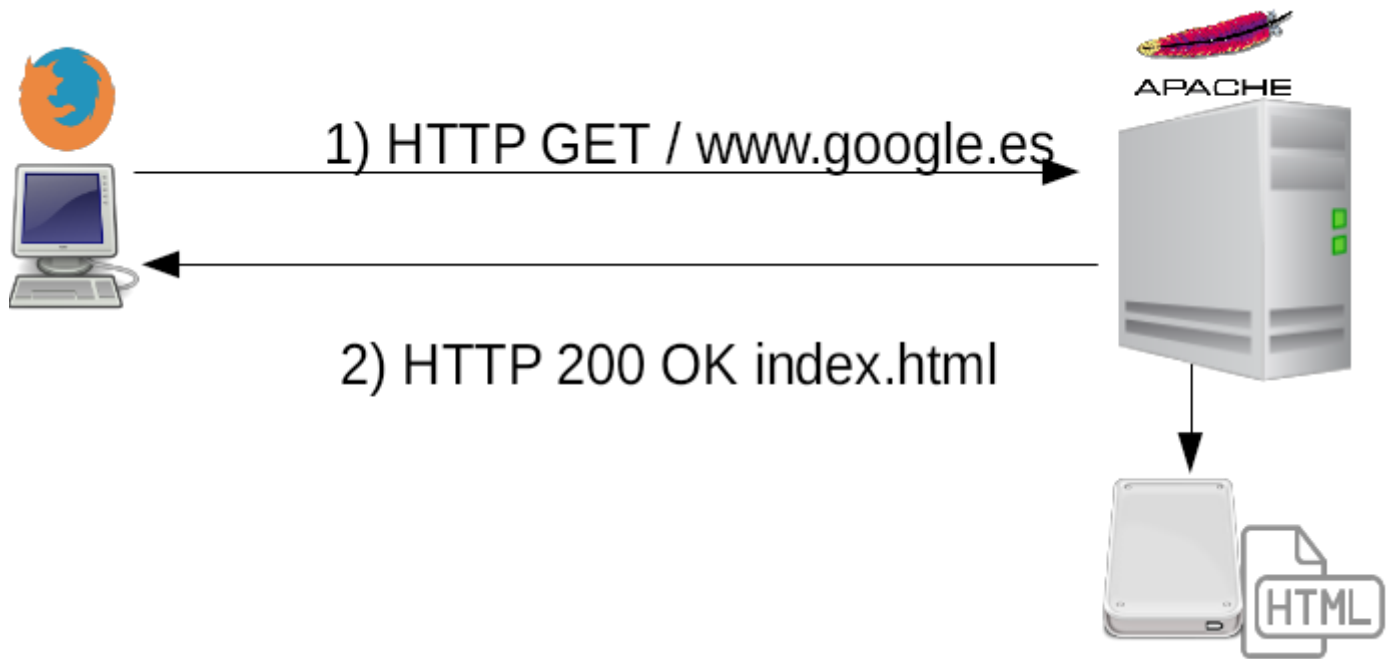
La web ha tenido un crecimiento exponencial debido a la existencias de las plataformas móviles, que han multiplicado los accesos a las páginas web (y que también han dificultado el diseño web, hablaremos más de ello en el tema sobre CSS).

Servidores web

Este apartado intenta ilustrar muy por encima como funciona un servidor web. Un servidor web es un programa que “entiende” los protocolos HTTP y HTTPS y que atiende peticiones de navegadores sirviendo páginas a medida que se van solicitando.

Tener un servidor no implica obligatoriamente un nombre de dominio. Es decir cuando en el navegador escribimos «https://www.goggle.es» contactamos con un programa que se está ejecutando en un ordenador al que se le ha dado el nombre «www.google.es» y le pedimos establecer una conexión para intercambiar información usando el protocolo HTTP.

Así, cuando un servidor recibe una petición examina sus directorios para ver si tiene ese archivo en el directorio que le han dicho, si es así, el fichero se transmite al que hizo la petición. Si no existe, el navegador devuelve un código 404 (recurso no encontrado) y nos mostrará la página web asociada a ese código 404. Si todo va bien nuestro navegador recibe un código 200



Intercambio de mensajes entre cliente web (o navegador) y servidor web.

Un servidor web puede ser privado o alquilado, existiendo grandes diferencias entre la forma de gestionar ambos.

- Uno privado, tiene la ventaja de ofrecer un control absoluto. No siempre es fácil mantener un ordenador doméstico conectado 24x7. En especial, en España, las conexiones ADSL no suelen ofrecer las capacidades necesarias para un sitio web de tamaño mediano.
- Los alquilados suelen conllevar un mayor precio cuando se necesita mas espacio para alojar nuestros ficheros. Ofrecen muchas garantías como por ejemplo anchos de banda muy aceptables a precios bastante competitivos y sobre todo que permiten al diseñador web liberarse de la gestión de los recursos de red.

Como configurar un servidor en casa

1. Se necesita un servidor web instalado en un equipo (Apache del paquete XAMPP)
2. Apuntar la IP del equipo en el que se instala Apache.
3. Entrar en el router y **abrir el puerto 80** indicando que se debe reenviar el tráfico a la IP que se apuntó.
4. Se debe averiguar la IP pública del router (por ejemplo podemos visitar [WhatsMylp](#) ,
5. La IP que aparece es la que se puede dar a clientes o amigos para que naveguen por nuestro sitio web.
6. (Optativo) se puede alquilar un nombre de dominio y solicitar que ese nombre (loquesea.com) sea redirigido a nuestra IP pública.

Los nombres de dominio

El servicio de nombres de dominio es un sistema que convierte de direcciones tipo [www.loquesea.com](#) a direcciones IP. Esto permite que sea más fácil recordar direcciones de páginas. Sin embargo, DNS es un sistema muy complejo que funciona de forma distribuida entre distintos países.

Los nombres de dominio se resuelven de final a principio. La última parte se llama TLD o Top Level Domain o dominio de primer nivel.. Estos dominios son administrados por países ocupándose cada uno de ellos de los nombres o marcas que hay dentro de dichos países.

Los sistemas de gestión de información

Definamos primero algunos términos:

- Sistema: conjunto de elementos interrelacionados que colaboran en la consecución de un objetivo.
- Gestión: conjunto de operaciones que resultan de relevancia para una persona o empresa.
- Información: conjunto de datos que resultan de utilidad a las funciones de la empresa.

Un SGI no tiene por qué estar informatizado.

Son programas que se pueden adaptar las necesidades de la empresa y que a veces necesitan importar (o exportar) datos e información.

El uso de un formato unificado (normalmente basado en marcas) facilita enormemente los trasvases de datos.

Tipos de lenguajes

Aparte de los lenguajes de marcas típicos, existen otros de uso bastante común. Siendo el caso más conocido LaTeX.

Programas como LaTeX tienen la desventaja de obligar a aprender “marcas”. Sin embargo, los algoritmos de colocación del texto suelen ser más sofisticados que otros programas. Los programas de redacción de documentos más utilizados suelen ser los del tipo MS-Word o LibreOffice que son del tipo WYSIWYG (What You See Is What You Get).

Dentro de los tipos de formatos, RTF es un mecanismo público que permite describir el aspecto de documentos. Al ser público, muchos procesadores pueden implementarlo si necesidad de pagar “royalties”.

Adobe lidera la especificación de documentos PDF (Portable Document Format). Antes de PDF, existía un lenguaje abierto denominado PostScript. Por otro lado RTF o LaTeX, son lenguajes descriptivos, mientras que PostScript es un lenguaje completo.

Organismos de regulación

1. La International Standards Organization emite estándares para la documentación.
2. El W3C (o World Wide Web Consortium) emite “technical recommendations” (o TR's) que los interesados en la web pueden seguir para garantizar la interoperabilidad. Su web es muy útil.

Gramáticas y DTD's

Las gramáticas HTML indican el conjunto de reglas para determinar lo que se acepta o no se acepta. Si se elige una gramática (como por ejemplo la de HTML5) es muy recomendable respetar las reglas de esa gramática. A modo de ejemplo, esto es una página válida.

```
<!DOCTYPE html>
```

```

<html>
  <head>
    <title>Page Title</title>
  </head>
  <body>
    Mi primera página
  </body>
</html>

```

y esto no lo es.

```

<!DOCTYPE html>
  <title>Esto es el título de la página
<body>

```

A lo largo del tiempo ha habido diversas versiones de HTML (con sus correspondientes gramáticas) y tales documentos deben llevar en la cabecera algo que diga a qué estándar se ciñen.

Las tres últimas familias de estándares han sido

- HTML4: muy permisivo, lo que dificulta a los navegadores el procesar el HTML dando lugar a que fuera bastante difícil para ellos el mostrar correctamente y de igual forma todos los HTML
- XHTML: es HTML con las estrictas reglas que impuso XML. Esto simplificó el desarrollo de navegadores y se avanzó en facilidad para mostrar páginas en distintos navegadores.
- HTML5: es una nueva revisión de XHTML en el que se han incluido nuevas posibilidades como etiquetas <audio> y <video> así como posibilidad de hacer muchas cosas desde JavaScript.

Un ejemplo de DTD, sería esto:

```

<!ELEMENT lista_de_personas (persona*)>
<!ELEMENT persona (nombre, fechanacimiento?, sexo?, numeroseguridadsocial?)>
<!ELEMENT nombre (#PCDATA) >
<!ELEMENT fechanacimiento (#PCDATA) >
<!ELEMENT sexo (#PCDATA) >
<!ELEMENT numeroseguridadsocial (#PCDATA)>

```

Y un ejemplo de archivo aceptado por esa DTD sería este

```

<lista_de_personas>
  <persona>
    <nombre> Pepe Pérez </nombre>
    <sexo> Varón </sexo>
    <numeroseguridadsocial>555</numeroseguridadsocial>
  </persona>
  <persona>
    <nombre> Angela Lopez </nombre>
    <fechanacimiento>13-2-1995</fechanacimiento>
    <sexo> Mujer </sexo>
    <numeroseguridadsocial>355</numeroseguridadsocial>
  </persona>
</lista_de_personas>

```

XML Schemas

Los XML Schemas surgen para mejorar las faltas de precisión que tenían las DTD. Sin embargo, la mejora en la precisión de la definición ha implicado que escribir XML Schemas sea mucho más complicado.

Un ejemplo de XML Schema (tomado de Wikipedia):

```
<?xml version="1.0" encoding="UTF-8"?>
  <xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
    <xsd:element name="Libro">
      <xsd:complexType>
        <xsd:sequence>
          <xsd:element name="Título"
            type="xsd:string"/>
          <xsd:element name="Autores"
            type="xsd:string"
            maxOccurs="10"/>
          <xsd:element name="Editorial"
            type="xsd:string"/>
        </xsd:sequence>
        <xsd:attribute name="precio"
          type="xsd:double"/>
      </xsd:complexType>
    </xsd:element>
  </xsd:schema>
```

Definiciones

Etiqueta: Es una secuencia de texto encerrada entre < y >

Elemento: Es todo lo que va entre una cierta etiqueta de apertura y cierre. En el ejemplo siguiente si nos hablan de la etiqueta libro se refieren simplemente a la etiqueta entre los paréntesis angulares.

Si hay que procesar el elemento libro esto significa procesar los sub-elementos o “elementos hijo”.

Atributo: Es un texto junto a la etiqueta que amplía información sobre la misma. En el ejemplo anterior podemos ver un atributo precio en la etiqueta titulo

Árbol del documento: Todo documento XML y HTML5 puede representarse como un árbol que se puede recorrer desde distintos lenguajes.

Este árbol a veces se llama el árbol DOM o simplemente el DOM (Document Object Model).

Relaciones de parentesco: En un árbol DOM, los distintos elementos (o nodos). Se dice que un nodo es hijo de otro si aparece más abajo en el árbol DOM.

Se dice que dos nodos son hermanos si están en el mismo nivel del árbol DOM. Se dice que un nodo es padre de otro si está en un nivel más arriba en el árbol DOM.