

## SISTEMAS DISTRIBUIDOS BASADOS EN LA WORLD WIDE WEB

Sistemas informáticos I



## 2.2 WEB HIPERTEXTO

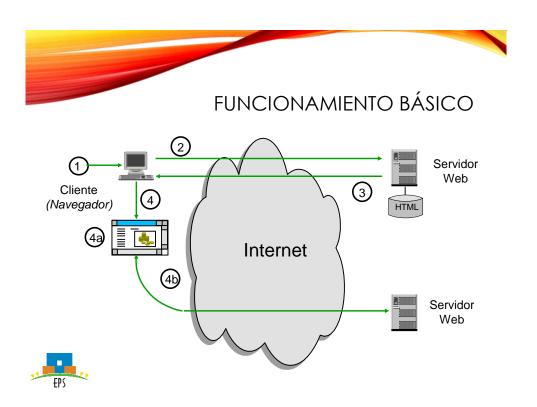
Sistemas distribuidos basados en la World Wide Web





- Es el modelo **original** y más sencillo de la Web
- Documentos distribuidos en servidores en Internet
  - Identificados por una URI
- Pueden contener otros objetos o documentos incrustados
- Enlaces desde cualquier documento a cualquier otro en la red
- El documento se define atendiendo a los elementos que contiene y no a la forma de presentarlo
  - Tendencia actual hacia una web semántica → metadatos y ontologías
  - Qué vs. Cómo
- Clientes estándar para visualizar documentos







- 1. El usuario solicita un recurso mediante su URL en un navegador (cliente)
- 2. El navegador genera una petición HTTP y la envía al servidor Web
- 3. El servidor Web recibe la petición y envía el recurso solicitado
  - Una sesión TCP por cada solicitud
  - · Los documentos se codifican utilizando HTML
- 4. El cliente interpreta y muestra el documento recibido
  - Puede tener asociados nuevos elementos en su interior
  - Nueva petición HTTP a los servidores que los contienen para su recuperación
  - Originalmente nueva sesión TCP, el protocolo HTTP actual permite reutilizar sesiones





- Protocolo de comunicaciones entre cliente y servidor Web
- Transporte: Conexión TCP sobre el puerto 80 (por defecto, 443 HTTPS)
- Intercambio de mensajes ASCII entre ambos:
  - Cliente realiza una petición

GET /hypertext/www/TheProject.html HTTP/1.0

• El servidor responde con un mensaje MIME (Multipurpose Internet Mail Extensions):

HTTP:/1.0 200 Document follows
MIME-Version: 1.0
Server:CERN/3.0
Content-Type: text/html
Content-Length: 8247

<HEAD><TITLE>The World Wide Web Consortium (W3C)</TITLE></HEAD>
<BODY>
<H1><IMB ALIGN=MIDDLE ALT="W3C" SRC="icons/WWW(w3c\_96x67.gif>
The World Wide Web Consortium</H1><P>...





- Formato de las peticiones: [método] URI [protocolo]
- URI: Identifica el objeto sobre el que aplicar el método
- Protocolo: El del mensaje enviado: HTTP/1.0, HTTP/1.1
- Método: Acción a realizar:
  - GET: Lectura del objeto
  - HEAD: Lectura de la cabecera del objeto
  - PUT: Almacenamiento del objeto
  - POST: Añadir al objeto la información enviada
  - DELETE: Eliminar el objeto
  - LINK: Conectar el objeto con otro determinado
  - UNLINK: Desconectar el objeto de otro determinado







- Estándar ISO-8879-1/1986 para la definición de texto electrónico independiente de dispositivos, sistemas y aplicaciones
- Proviene de GML (IBM, 70's)
- Metalenguaje para definir lenguajes de diseño descriptivos
- Proporciona un medio de codificar mensajes cuyo destino sea el intercambio directo entre sistemas o aplicaciones
- · Múltiples lenguajes definidos mediante SGML
  - Hypertext Markup Language, HTML
- Evolución de SGML: Extended Markup Language, XML





- SGML tiene cinco características principales:
  - Permite crear lenguajes de codificación descriptivos
    - Códigos incluidos en el propio texto definen los elementos que lo componen
  - Define una estructura de documento jerárquica, con elementos y componentes interconectados
  - Proporciona una especificación formal completa del documento
    - Contenida en la Document Type Definition, DTD
  - No tiene un conjunto implícito de convenciones de señalización. Soporta, por tanto, un conjunto flexible de juegos de etiquetas
    - SGML es un metalenguaje. Definición de etiquetas para cada lenguaje en su DTD
  - · Los documentos generados con él son legibles por personas





```
<!-- Parameter Entities -->
<!ENTITY % head.misc "SCRIPT | STYLE | META | LINK" -- repeatable head elements -->
<!ENTITY % heading "H1 | H2 | H3 | H4 | H5 | H6">
<!ENTITY % list "UL | OL | DIR | MENU">
<!ENTITY % preformatted "PRE">
....
<!ELEMENT FONT -- (%inline)* -- local change to font -->
<!ATTLIST FONT</p>
size CDATA #IMPLIED -- [+] nn e.g. size="+1", size=4 --
color CDATA #IMPLIED -- #RRGGBB in hex, e.g. red: "#FF0000" --
face CDATA #IMPLIED -- comma separated list of font names --
>
```





- Aplicación del estándar SGML. <a href="https://www.w3.org/MarkUp/">https://www.w3.org/MarkUp/</a>
- Lenguaje de definición de formato de documentos
  - Define los elementos que pueden aparecer dentro del documento
  - Se especifican mediante etiquetas (Tags) de comienzo y final de documento
     Texto ASCII encerrado entre < y >
  - Sólo se define el tipo de elemento y <u>no la forma de representarlo</u>
- <u>Lenguaje estándar</u> utilizado en la Web para el intercambio de documentos hipermedia que incluyen:
  - Texto
  - Imágenes (Fotos, Vídeo)
  - Audio
  - Vínculos (Links)
- Estándar especificado por el World Wide Web Consortium, W3C





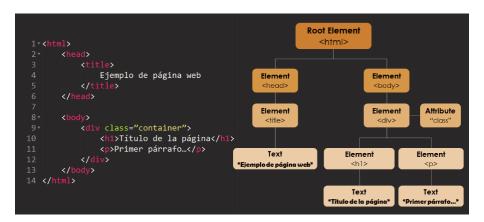
- · Identificación SGML
  - Permite identificar la DTD adecuada para procesarlo

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN">

- · Delimitación del documento HTML
  - Identificado con la etiqueta <HTML>
- · Cabecera. Información asociada al documento
  - Identificado por la etiqueta <HEAD>
  - Contiene información descriptiva del documento y definiciones de formato para los elementos del resto del mismo
- · Cuerpo del documento
  - Identificado con la etiqueta <BODY>
  - Contiene los elementos de presentación del documento









https://www.w3schools.com/html



- Etiquetas/Elementos
  - No vacías:
    - La mayoría de elementos comienzan por un tag, luego contenido, finalizando por un tag de terminación

<TAG> contenido </TAG>

- Vacías: por ejemplo <LI>, <BR>, etc.
- Atributos:
  - Específicos de cada etiqueta
  - Texto ASCII entre comillas (doble o simple) dentro de la etiqueta de comienzo
- Existen multitud de etiquetas que no vamos a ver en detalle, aquí solo vamos a tratar algunos (muy pocos) ejemplos representativos. Consultar el manual de referencia de el W3C
- HTML es insensible a Mayúsculas/Minúsculas:
  - <TAG> es equivalente a <tag> o <TaG>
- Espacios en blanco, tabulaciones y retornos de carros no son significativos

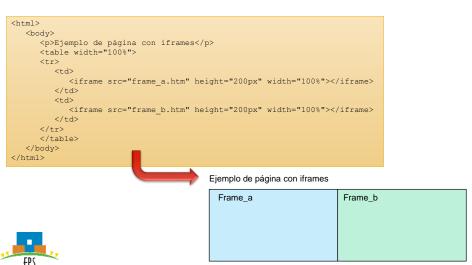




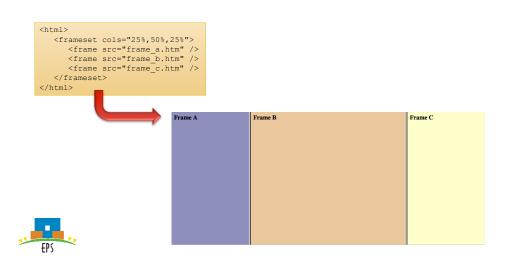
- Con los frames (iframes en HTML5) varios documentos HTML pueden ser mostrados en la misma ventaja del navegador
- Cada frame es independiente de los otros
- Han caído en desuso: seguridad y AJAX





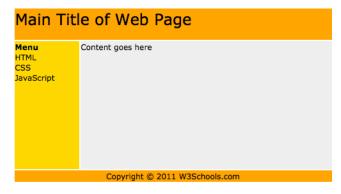








- Dos formas de organizar dónde va cada cosa:
  - Tablas
  - Etiqueta <div>
    - CSS





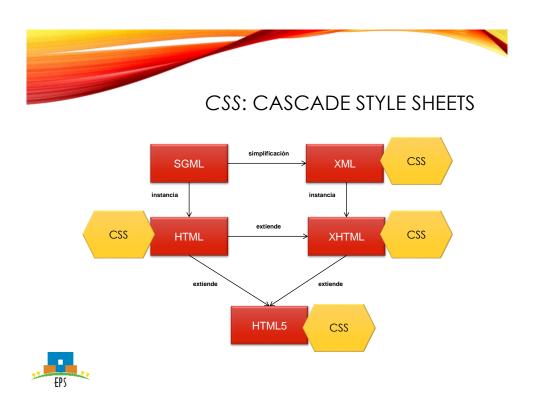


- El único disponible en las primeras versiones de HTML
- Solía combinarse con el uso de frames, actualmente es más habitual el uso de AJAX
- · Desventajas:
  - Obliga a crear ficheros muy grandes y complejos
  - Obliga a descargar información de presentación con cada página que se descarga
  - El rediseño de sitios y contenido existente es más laboriosos (y costoso)
  - Es difícil (y costoso) mantener consistencia visual en un sitio
  - Las páginas basadas en tablas son menos accesibles a usuarios con necesidades especiales y a los que usan teléfonos móviles, PDAs o tabletas para acceder











- Una regla fundamental del diseño de software es que funciones distintas deben ser llevadas a cabo por entidades distintas
  - · Separation of concerns
- HTML originalmente fue diseñado para definir contenido de documentos...
   <h1>This is a heading</h1>

This is a paragraph.

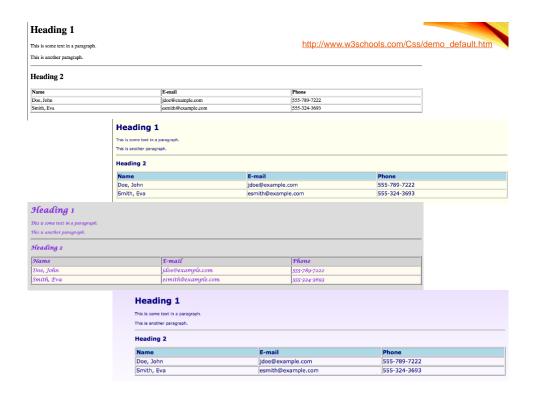
- ... pero no para definir formato, la forma en que se presentan los contenidos
  - Cuando se agregaron etiquetas como <font> (a partir de HTML 3.2) la cosa empezó a complicarse
  - Se mezcla el contenido (el qué) con el formato (el cómo)!!!





- Usado para describir la semántica de presentación de un documento escrito con un lenguaje de marcado
  - El uso más común es darle estilo a páginas escritas en HTML y XHTML
- El objetivo es separar el contenido del documento (HTML o similar) de cómo se presenta (layout, colores, fuentes, etc.)
- · Ventajas:
  - Forma rápida y sencilla de cambiar el aspecto del documento
  - · Ofrecer accesibilidad
  - · Más control y flexibilidad de presentación
  - · Compartir formato entre múltiples páginas
  - Distinta presentación para distinto tipo de dispositivo (ordenador, móvil, impreso...) o entre distinto tipo de usuarios de a aplicación (p.ej., usuarios de distintas compañías)







- Un estilo CSS es una combinación de propiedades predefinidas a las que se asigna un valor (propiedad:valor). Cada declaración separada por ";"
- Estilos inline (mala práctica, solo en casos muy específicos): aplican directamente sobre un elemento del documento

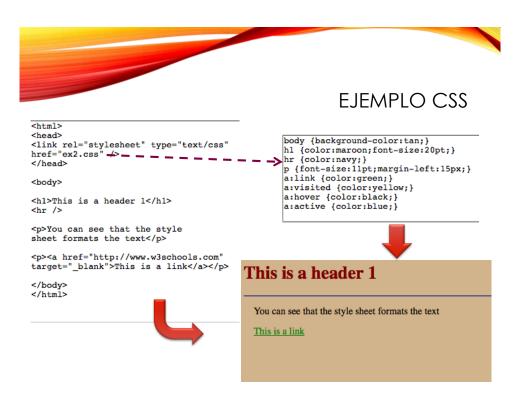
This is a paragraph.

- Lo más habitual es definir hojas de estilos compuestas por reglas de aplicación
- Las reglas constan de un selector y las declaraciones del estilo entre llaves



- · Se pueden definir:
  - · Hojas de estilo internas: en el propio documento como contenido del tag <style>
  - Hojas de estilo externas: en ficheros externos vinculados al documento





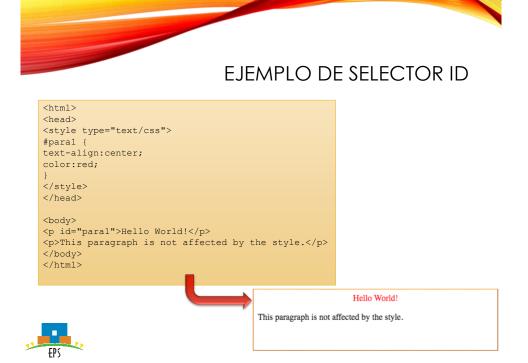


- · Selector de elemento
  - Usado para especificar el estilo de los elementos de un tipo determinado
  - Se define con el nombre del elemento
- - Usado para especificar el estilo de un elemento con un id determinado
    Usa el atributo id de los elementos HTML y se define con "#"
- · Selector class
  - Permite especificar un estilo para todos los elementos de una determinada "clase"
  - Utiliza el atributo class de HTML y se define con "."
- Combinación de selectores:
  - Reglas de selección en función de la relación entre selectores:

    - Espacio: selector de descendientes
       : selector de hijos
       : selector de hermanos
       : selector de hermanos

https://www.w3schools.com/css/css\_combinators.asp







```
<html>
<head>
<style type="text/css">
.center {
    text-align:center;
}
</style>
</head>

<body>
<h1 class="center">Center-aligned heading</h1>
Center-aligned paragraph.
</body>
</html>
```

# **Center-aligned heading**

Center-aligned paragraph.





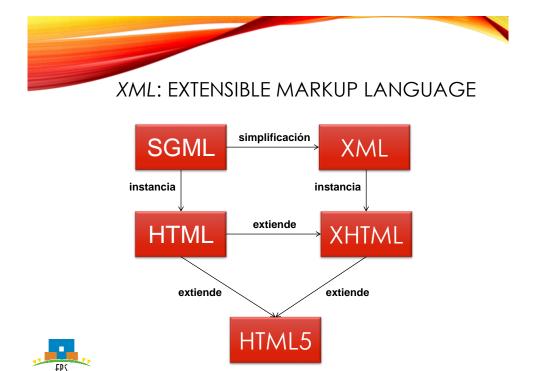
- ... lo que por fin explica el nombre
- Sobre un elemento determinado pueden aplicar distintos estilos definidos en distintos "sitios"
- Se impone un orden de precedencia que determina qué selector prevalece según tres criterios (de mayor a menor):
  - Importancia (!important)
  - Especificidad:
    - Estilo inline
    - Selector id
    - Selector class
    - · Selector de elemento
  - Orden





- Hacer una página web personal estática, recomendablemente añadiendo hojas de estilo CSS, que al menos incluya los siguientes elementos:
  - Un enlace
  - Una imagen (no es necesario que sea de una altísima calidad)
  - Una tabla
- Entregar de forma individual a través de Moodle







- Especificación del W3C (<u>www.w3.org/XML</u>) para <u>crear lenguajes descriptivos</u> de propósito específico
  - Fecha de primera especificación: 1998
    - https://www.w3.org/TR/1998/REC-xml-19980210
  - https://www.w3schools.com/xml/xml\_whatis.asp
    - XML is a software- and hardware-independent tool for storing and transporting data
    - · XML Does Not DO Anything
- Procede de SGML reduciendo su complejidad
- Se puede ver como una generalización de HTML para modelar estructuras de datos de cara al procesamiento automático de información
  - Lenguaje de meta-marcado
  - Diseñado para describir estructuras de datos
  - Utilizado para realizar intercambios de datos entre sistemas
  - Etiquetas, elementos y atributos con estructura de árbol
- En los últimos años hay una tendencia a sustituir XML por JSON





## Root





- El documento XML es de texto, normalmente Unicode
- Cada documento XML debe tener un único elemento raíz
- Todo elemento no vacío debe tener etiqueta de comienzo (<elemento>) y final (</elemento>)
- Todo elemento que no lleve contenido puede ser marcado con una etiqueta de elemento vacío (<elemento/>)
- Todos los valores de atributos deben ir entre comillas dobles
- Los elementos se pueden anidar, pero no solapar con otros
- Los nombres de los elementos son sensibles a mayúsculas y minúsculas
- Permite el uso de distintos espacios de nombres tomados de distintos vocabularios en el mismo documento (namespace:nombre)





- Document Type Definition
- · No es obligatorio
- · Especifican un sublenguaje de XML
  - Lista de etiquetas permitidas
  - Etiquetas y atributos permitidos dentro de cada etiqueta
- Pueden ser muy complejos
  - Para SVG (Scalable Vector Graphics) más de 1000 líneas
  - Para XHTML 1.0 más de 1500 líneas
  - Para DocBook más de 11000 líneas
- Alternativa: XML Schema (versión 1.0, mayo 2001)

https://www.w3schools.com/xml/xml\_dtd\_examples.asp



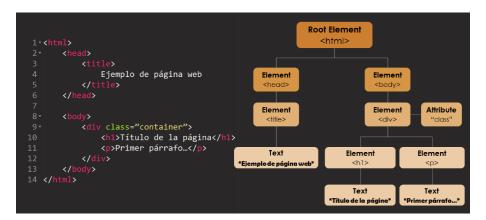


### SAX: Simple API for XML

- · Orientado a eventos
- El programador proporciona métodos callback que son invocados por el parser a medida que lee el documento XML
- Es rápido y eficiente, pero difícil de usar
  - Cada nodo es tratado independientemente de lo que había antes o después
  - No se puede "volver atrás" en el procesado
- · DOM: Document Object Model
  - Los documentos se estructuran como árboles
  - También sirve para otros lenguajes como HTML
  - Estándar del W3C
  - XML DOM:
    - Modelo de objetos estándar para XML
    - Una API estándar para XML, aunque no todos los parser DOM siguen el estándar
    - Independiente de la plataforma y del lenguaje de programación











- Todo en DOM son nodos:
  - El documento completo es un document node
  - Cada elemento XML es un element node
  - · El texto en elementos XML son text nodes
  - · Cada atributo es un attribute node
  - · Los comentarios son comment nodes
- La API DOM define una serie de propiedades y métodos que debe ofrecer cualquier implementación de DOM para acceder y manipular nodos en las distintas dimensiones del árbol DOM
  - x.nodeName: el nombre de x
  - x.nodeValue: el valor de x
  - x.parentNode: el nodo padre de x
  - x.childNodes: los nodos hijos de x
  - x.attributes: los nodos atributos de x



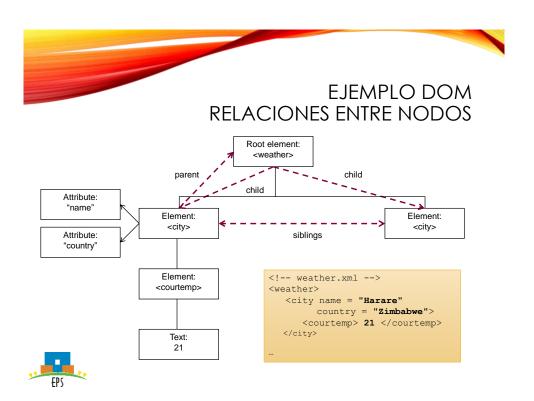


- XPATH es un mecanismo para definir expresiones para navegar por las distintas dimensiones del árbol DOM
  - En cierta medida análogo a la definición de rutas en un sistema de archivos
- Recomendación del W3C independiente del lenguaje de programación
  - · Hay implementaciones en múltiples lenguajes
  - En particular, es uno de los principales elementos de XSLT
- Distingue entre 7 tipos de nodos:
  - Element
  - Attribute
  - Text
  - Comment Document

  - Namespace Processing-instruction
- · Relaciones entre nodos:
  - Parent
  - Children
  - Siblings
  - Ancestors
  - Descendants



https://www.w3schools.com/xml/xml\_xpath.asp





- /bookstore/book[1]
- /bookstore/book[last()]
- /bookstore/book[last()-1]
- /bookstore/book[position()<3]
- //title[@lang]
- //title[@lang='en']
- /bookstore/book[price>35.00]
- /bookstore/book[price>35.00]/title





- · CSS dice cómo mostrar un documento HTML
  - HTML tiene un conjunto predefinido de etiquetas
  - Decirle a un navegador cómo mostrar un elemento en una fuente o color especial es sencillo
- XML en cambio permite definir cualquier etiqueta
  - ¿Qué significa TABLE?
  - ¿Cómo debe mostrarlo el cliente?







- XSL le dice al navegador cómo se debe mostrar el documento XML
- Una plantilla XSL es un documento XML





```
EPS -
```

### datos.xml

### dtd-datos.dtd

```
<!DOCTYPE helloworld
[
<!ELEMENT helloworld (person)>
<!ELEMENT person (name)>
<!ELEMENT name (#PCDATA)>
]>
```



- Según el W3C:
  - JSON is a syntax for storing and exchanging data
  - JSON is text, written with JavaScript object notation
- Muy popular: a día de hoy es seguramente el formato más utilizado para intercambiar datos entre aplicaciones web
- Es legible por personas e interpretable por algoritmos
- Almacena los datos siguiendo la notación de objetos y arrays JS





- Un objeto JS es un conjunto desordenado de pares nombre/valor entre llaves ({})
  - El nombre es una cadena de caracteres que identifica el atributo
  - El valor puede ser un objeto, un array, un número, cadena de caracteres, true, false o null

```
{ "artistname" : "Pink Floyd" }

{ "artistname" : "Pink Floyd",
   "formed" : "1964"
}

{ "artistname" : "Pink Floyd",
   "formed" : "1964",
   "origin" : "Cambridge, United Kingdom"
}
```





- Un array es una colección ordenada de valores
- Permite construir listas de valores











https://www.w3schools.com/js/js\_json\_xml.asp



- En principio, XML y JSON sirven para lo mismo
- Sus diferencias son:
  - JSON no utiliza tags
  - JSON es más corto
  - JSON es más rápido de leer y escribir
  - JSON utiliza arrays
  - XML debe ser parseado por un parser (tradicionalmente DOM), mientras que JSON puede serlo por el interprete JavaScript como objetos y arrays JS "listos para utilizar"
  - · Los mecanismos de validación XML
    - https://json-schema.org/
    - Actualmente en versión draft









