

Estructura de una página web (01)

```
<!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" xml:lang="es" lang="es">
  <head>
    <title>Título</title>
    <meta http-equiv = "Content-Type"
      content = "text/html; charset=UTF-8"/>
  </head>
  <body>
    Contenido de la página web
  </body>
</html>
```

14

Estructura de una página web (02)

- **DOCTYPE:** permite establecer el documento DTD que se va a utilizar para validar la página. En este caso XHTML 1.0 Transaccional.
 - **html:** marca principio y final del documento.
 - xmlns: especifica el espacio de nombres al que pertenecen las etiquetas.
 - xml:lang: indica el lenguaje en que está escrito la página.
 - lang: también sirve para indicar el lenguaje de que la página. Es opcional pero recomendable para advertir a todos los navegadores.
- Códigos: es, es-ES (español España), en (inglés), en-US (inglés USA), fr (francés), etc.

15

Estructura de una página web (03)

- **head**: establece el bloque de cabecera donde se incluyen las propiedades de la página, título, estilos CSS, código Javascript y cualquier elemento que se incluya en la página.
 - title: título de la página.
 - meta: establece el tipo de codificación de texto que utiliza la página siendo el más recomendable UTF-8 (unicode de 8 bits).
- **body**: establece el bloque body que es el que incluye el contenido que se mostrará en la página web.

16

Estructura de una página web (04)

■ HTML 5

```
<!DOCTYPE html>
<html lang = "es">
  <head>
    <title>Título</title>
    <meta charset = "UTF-8"/>
  </head>
  <body>
    Contenido de la página web
  </body>
</html>
```

17

Estructura de una página web (05)

En HTML5 se tiende a simplificar:

- La etiqueta **DOCTYPE** únicamente indica que se trata de un documento HTML sin establecer el documento DTD asociado.
- La etiqueta **meta** para indicar el código de caracteres es mucho más simple.
- El lenguaje se indica utilizando el atributo **lang** únicamente.
- El resto de elementos es idéntico.

18

Estructura de una página web (06)

Atributos comunes

Son propiedades que se pueden establecer para cualquier elemento.

- **id**: es un atributo fundamental que permite identificar a los elementos.

Deben comenzar por letra y pueden contener letras y números pero no espacios en blanco.

Deben ser únicos, es decir dos elementos de una página no pueden tener el mismo id.

- **title**: permite poner un título al elemento.
- **lang**: indica el lenguaje del elemento y sólo es necesario cuando sea distinto al de la página.

19

Estructura de una página web (07)

Atributos comunes

- **xml-lang**: su significado es el mismo que lang pero es entendible por el software analizador de XML, sólo se utiliza en XHTML.
- **dir**: indica la dirección de lectura del texto contenido en el elemento:
 - **ltr** (left to right): de izquierda a derecha.
 - **rtl** (right to left): de derecha a izquierda.
- **class** y **style**: permiten especificar formato de tipo CSS al elemento.

20

Estructura de una página web (08)

Codificación de caracteres

El código utilizado tradicionalmente para la representación de caracteres en un ordenador es el ASCII (7bits). Este código es muy limitado ya que sólo permite representar 128 caracteres diferentes.

Los navegadores compatibles con HTML 4 y posteriores pueden utilizar códigos más extensos como ISO-8859-1 y Unicode (UTF-8, UTF-16) capaces de representar cualquier carácter nacional fuera de ASCII.

Sin embargo a la hora de publicar contenido en la web puede que nos encontremos con que algunos caracteres (fuera de ASCII) no se representan bien.

21

Estructura de una página web (09)

Codificación de caracteres

El problema surge cuando nuestra página es traducida en un servidor web que tiene configurado un charset diferente al nuestro, en cuyo caso las etiquetas meta de nuestro html son ignoradas y los caracteres fuera de ASCII pueden ser representados de forma incorrecta.

Una solución podría ser evitar cualquier carácter fuera de ASCII, pero eso limitaría mucho los caracteres que se pueden representar.

Otra solución es utilizar una notación especial para representar los caracteres que puedan suponer un problema: **entidades**.

22

Estructura de una página web (10)

Codificación de caracteres

Además las entidades sirven también para representar aquellos caracteres que podrían ser interpretados por el navegador.

Los caracteres que obligatoriamente se tienen que representar por medio de entidades son:

Carácter	Entidad
"	"
'	'
&	&
<	<
>	>

23

Estructura de una página web (11)

Codificación de caracteres

Sin embargo el auge de Unicode hace que la mejor solución sea optar por alguna de sus codificaciones UTF-8, UTF-16 que permiten utilizar los símbolos nacionales, y en caso de que la información no se muestre bien modificar la configuración de nuestro servidor web de manera que no haya un charset predefinido obligando así a utilizar la codificación que se establezca en las etiquetas meta de la página.

Por ejemplo en Apache deberíamos editar el archivo "**httpd.conf**" incluyeno la línea:

AddDefaultCharset off

24

Estructura de una página web (12)

Espacios en blanco

Cuando escribimos un texto en una página web hay que tener en cuenta que los tabuladores y saltos de línea se omiten, y si tenemos espacios en blanco consecutivos sólo el primero será tenido en cuenta.

```
<body>
  Contenido
  de la           página           web
</body>
```

Mostrará:

Contenido de la página web

25

Estructura de una página web (13)

Espacios en blanco

Para añadir espacios en blanco al texto y que esos espacios sean respetados por el navegador hemos de utilizar la entidad ** **.

```
<body>  
  Contenido de la&nbsp;&nbsp;&nbsp;&nbsp;&nbsp;página&nbsp;&nbsp;&nbsp;web  
</body>
```

Mostrará:

Contenido de la página web