

## Tema 1 - Reconocimiento de las características de los lenguajes de marcas

### 1. Lenguajes de Marcas. Definición y clasificación.

En los lenguajes de marcas se combina la información que contiene el documento con las marcas que indican el comienzo y el final de algo. Dependiendo del lenguaje de marcas utilizado se utilizarán unas u otras marcas. Nosotros a lo largo del curso vamos a utilizar lenguajes de marcas tales como *HTML*, *XML*, *XSL*, ... La característica común a todos ellos es el uso de etiquetas que marcan el inicio y el final de algo. Dichas etiquetas, por ejemplo, en el caso de *HTML* no serán visibles si abrimos el documento en un navegador, las etiquetas serán interpretadas por el navegador y no nos serán mostradas.

Veamos un ejemplo de *XML*:

```
<alumno id="1">
  <nombre>Luisa</nombre>
  <ap1>Martínez</ap1>
  <ap2>Pérez</ap2>
  <curso>IDAW</curso>
  <asignaturas>
    <asignatura>LMSGI</asignatura>
    <asignatura>Programación</asignatura>
    <asignatura>Bases de Datos</asignatura>
    <asignatura>Entornos de Desarrollo</asignatura>
    <asignatura>Sistemas Informáticos</asignatura>
    <asignatura>FOL</asignatura>
  </asignaturas>
</alumno>
```

Vemos que hay una etiqueta de inicio de la información del alumno `<alumno>` y toda la información de dicho alumno está contenida entre dicha etiqueta y la de cierre `</alumno>` y lo mismo ocurre para cada uno de los "trozos" de información relativos al elemento alumno, `<nombre>...</nombre>`, etc.

### 2. Distintos lenguajes de marcas

Existen principalmente tres tipos de lenguajes de marcas:

- **Orientados a la presentación**, utilizan marcas que pueden ser binarias o etiquetas. Por ejemplo en el LibreOffice Writer cuando marcamos un trozo de texto en negrita, se sitúa una marca binaria (invisible para el usuario) delante y detrás de dicho texto, indicando aquí empieza texto en negrita y aquí acaba. Las primeras versiones de *HTML* también eran orientadas a la presentación, siguiendo con el mismo ejemplo en *HTML*, `<b>Hola</b>`, hacía que el texto "Hola" apareciese en negrita. *HTML* ha dejado de ser orientado a la presentación en la última versión, la versión 5, ya no se pueden usar etiquetas de presentación, dejándose eso para las hojas de estilo (ficheros css).

- **Procedurales**, son "parecidos" a las instrucciones que se ejecutan en los lenguajes de programación, siguiendo con el ejemplo del punto anterior, en LaTeX `\textbf{Hola}`, haría que la palabra "Hola" apareciese en negrita.

- **Descriptivos**, se utilizan marcas para indicar la estructura del documento, utilizando una etiqueta de inicio y otra de fin para indicar donde empieza y acaba una sección. Por ejemplo, en *HTML* la etiqueta `<body>...</body>`, indica el comienzo de la sección donde se incluirá todo el contenido de nuestra página web. Otro ejemplo, `<table>....</table>` abarca el contenido de una tabla de datos.

### 3. Historia de los lenguajes de marcas (HTML)

El lenguaje de marcas que revolucionó Internet es *HTML*. HTML es el lenguaje de marcas con el que se escriben las páginas web. En 1989 existían dos técnicas que permitían vincular documentos electrónicos, por un lado los hipervínculos o enlaces (hyperlinks o links) y por otro lado un poderoso lenguaje de etiquetas denominado SGML. Por entonces, Tim Berners-Lee, quien trabajaba en el Centro Europeo de Investigaciones Nucleares CERN da a conocer a la prensa que estaba trabajando en un sistema que iba a permitir acceder a ficheros en línea que funcionaba sobre redes de computadoras o máquinas electrónicas basadas en el protocolo TCP/IP. Inicialmente fue desarrollado para que se pudiera compartir fácilmente información entre científicos de distintas universidades e institutos de investigación de todo el mundo.

A principios de 1990, define por fin el HTML como un subconjunto del conocido SGML y crea algo más valioso incluso, el World Wide Web.

Tim Berners-Lee creó el proyecto World Wide Web (Tejido o Telaraña Mundial), así como un sistema que facilitaba la lectura de información, mediante un programa de navegación. Sería el primer navegador web, llamado WorldWideWeb, y desarrollado durante la segunda mitad del año 1990; siendo tiempo después rebautizado como Nexus, para evitar confusiones por su nombre que era igual al de la tecnología que representaba.

HTML fue evolucionando pasando por distintas versiones añadiéndole capacidades, como por ejemplo, facilidades para crear tablas, hacer que el texto fluyese alrededor de las figuras y mostrar elementos matemáticos complejos. Aparece los navegadores Netscape e Internet Explorer que mantendrían una larga pugna por ser el más puntero, añadiendo etiquetas para producir efectos que estaban fuera del estándar.

En 1997, HTML 4.0 se publicó como una recomendación del W3C. HTML 4.0 adoptó muchos elementos específicos desarrollados inicialmente por los navegadores citados, pero al mismo tiempo comenzó a limpiar el HTML señalando algunos de ellos como «desaprobados» (deprecated, en inglés).

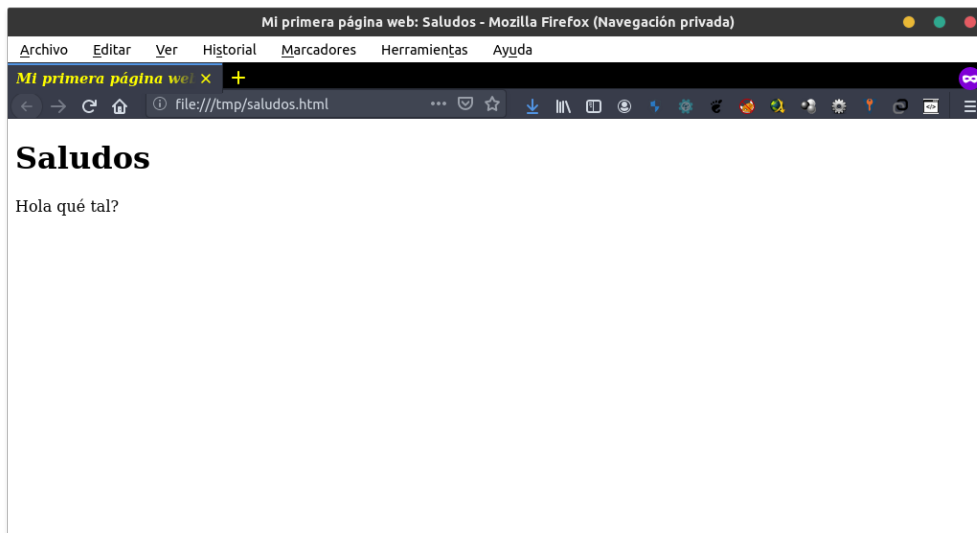
En 2006, el W3C se interesó en el desarrollo de HTML5, y en 2007 se unió al grupo de trabajo del WHATWG para unificar el proyecto. HTML 5 es la versión más reciente. En ella se toma la decisión de separar definitivamente el contenido (ficheros *.html*) de la presentación (ficheros *.css*). En la actualidad el "W3C" se ha convertido en el organismo de referencia para las tecnologías *HTML*, *XML*, *CSS*, *XSL*, etc.

**Actividad 1.1.** Utilizando un editor de *texto plano* copia y guarda el siguiente fichero con el nombre de *saludos.html*:

```
<!DOCTYPE html>
<html lang="es">
  <head>
    <meta charset="UTF-8"/>
    <title>Mi primera página web: saludos</title>
  </head>
  <body>
    <h1>Saludos</h1>
    <p>Hola qué tal?</p>
  </body>
</html>
```

Una vez hayas guardado el fichero ábrelo con el navegador (con el explorador de ficheros, vete a la carpeta donde lo hayas guardado, haz clic con el botón derecho del ratón, selecciona abrir con, y ábrelo con *Firefox*, *Chrome*, ...)

Notarás que en el navegador no se ven las etiquetas, sino que se "*interpretan*" y el resultado será algo parecido a esto:



En el siguiente tema se profundizará en el lenguaje de marcas *HTML* (*HyperText Markup Language*) y en las hojas de estilo en cascada *CSS* (*Cascade Style Sheets*).

#### 4. Etiquetas. Elementos. Atributos.

**Etiqueta.** Una etiqueta comienza con un carácter "<" y acaba con un carácter ">". La mayoría de etiquetas tienen etiqueta de apertura y cierre. En el ejemplo anterior, `<h1>...</h1>`, son etiquetas, la primera es la apertura y la segunda, idéntica a la primera sólo que el nombre va precedido de un carácter "/", la de cierre.

**Elementos.** Elemento es todo lo que esté contenido entre una etiqueta de apertura y una etiqueta de cierre. Siguiendo con el ejemplo anterior `<h1>Saludos</h1>`, es un elemento. El elemento `h1` es un título de primer nivel (lo veremos con mayor detalle en el siguiente tema). En este caso estamos diciendo que **Saludos** es un título principal del documento.

**Atributos.** Un atributo es una palabra clave seguida de un carácter "=", seguida de un valor entrecomillado, las comillas pueden ser dobles o simples. En el ejemplo anterior, el elemento **html** contiene un atributo `lang="es"`. La etiqueta `<html lang="es">`, marca el inicio del contenido de nuestro documento, el cual será representado dentro de la pantalla del navegador. En este caso el atributo `lang` nos está indicado que el contenido de nuestra página web está en español. Podemos inferir que un atributo suele usarse para establecer el valor de una propiedad de un elemento.

Otro ejemplo, en este caso de *XML*

```
<?xml version="1.0" encoding="UTF-8"?>
<countries>
  <country name="Afghanistan" population="22664136" area="647500">
    <language percentage="11">Turkic</language>
    <language percentage="35">Pashtu</language>
    <language percentage="50">Afghan Persian</language>
  </country>
```

```

<country name="Albania" population="3249136" area="28750"/>
<country name="Algeria" population="29183032" area="2381740">
  <city>
    <name>Algiers</name>
    <population>1507241</population>
  </city>
</country>
<country name="American Samoa" population="59566" area="199"/>
<country name="Andorra" population="72766" area="450"/>
<country name="Angola" population="10342899" area="1246700"/>
<country name="Anguilla" population="10424" area="91">
  <language percentage="100">English</language>
</country>
</countries>

```

En el ejemplo anterior hay un elemento `<countries>....</countries>`, que contiene toda la información del documento, a este elemento se le llama elemento raíz (lo veremos en temas posteriores), dentro de él hay una serie de elementos `<country>...</country>`, que a su vez contienen información de un país determinado, por ejemplo el primer país que aparece vemos que es *Afghanistan*, el elemento correspondiente a dicho "country", contiene tres atributos que son **name**, **population** y **area**, y dentro de dicho elemento hay a su vez tres elementos `<language>...</language>`, que a su vez tiene un atributo **percentage**. En temas posteriores veremos ésto con más profundidad.

**Actividad 1.2.** Partiendo del ejemplo XML anterior, piensa en cómo almacenarías una agenda de teléfonos de tus amigos, donde la información relevante sería el **número** de teléfono, el **nombre** y la **dirección** de tu amigo.

## 5. Definición de la estructura de los documentos

Asociado a un documento escrito en un determinado lenguaje de marcas normalmente habrá otro documento que especifica **qué** estructura debe tener dicho documento. En el ejemplo anterior de **countries** esto se referiría a qué estructura debe llevar un **country**, que **atributos** están permitidos en dicho elemento, **qué** elementos están permitidos dentro de un **country**, en **qué** orden deben de aparecer, **qué** atributos pueden tener éstos últimos, y así sucesivamente. Para ello veremos en temas posteriores dos tipos de documentos para definir esta estructura, **DTD (Document Type Definition - Definición de Tipo de Documento)** y **XSD (Xml Schema Definition - Definición de Esquema XML)**. Veremos que ambos sirven para definir **qué** elementos, **qué** atributos, en **qué** orden pueden aparecer dentro de un documento escrito en un determinado lenguaje de marcas, **cuántas** veces pueden aparecer y **qué** valores pueden tomar.