

Lenguaje de Marcas



1. Introducción

El reconocimiento de las características de lenguajes de marcas es un tema **fundamental en el mundo de la informática y la programación web**. Los lenguajes de marcas desempeñan un papel crucial en la creación y representación de contenido en la web y en otros contextos. Veremos qué son los lenguajes de marcas, sus características principales y su importancia en el desarrollo web y la gestión de datos.

Este módulo implica una aproximación a los lenguajes de marcas, con especial incidencia a **XML** (acrónimo de la expresión inglesa eXtensible Markup Language, que podríamos traducir como 'lenguaje de marcas extensible'), y también en otras tecnologías íntimamente relacionadas con XML.

Necesariamente, pues, la unidad “Lenguajes de marcas” proporciona una introducción a los diferentes lenguajes de marcas, ofrece una clasificación y muestra cómo debe ser la estructura de un documento XML bien formado.

En definitiva, **XML es un lenguaje de etiquetas** (como el **HTML**), desarrollado por el World Wide Web Consortium (W3C). Sin embargo, a diferencia de otros lenguajes de marcas, el **XML** tiene una potencialidad que le **permite definir la gramática de otros lenguajes específicos**. Por tanto, el **XML no es realmente un lenguaje** en particular, **sino una manera de definir lenguajes para diferentes necesidades**.

Hay que decir que el XML es una tecnología relativamente sencilla que tiene a su alrededor otras que la complementan y la hacen notablemente más extensa, además de proporcionarle unas posibilidades mucho mayores. Actualmente tiene un papel muy importante, ya que permite la compatibilidad entre sistemas, haciendo posible compartir información de una manera segura, fiable y sencilla.

Los lenguajes de marcas están en pleno auge. Apoyándose en el gran éxito que han representado las páginas web han ido apareciendo nuevas tecnologías basadas en lenguajes de marcas que permiten a los usuarios ya los programas conseguir resultados con los que antes sólo podía soñarse.

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).

Gramáticas y DTD's (Document Type Definition)

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.

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`

Definiciones

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.