



Lenguajes de marcas



Tema 1. Reconocimiento de las características de los lenguajes de marcas

1. Concepto de lenguajes de marcas y clasificación

Un lenguaje de marcas es un modo de codificar un documento donde, junto con el texto, se incorporan etiquetas, marcas o anotaciones con información adicional relativa a la estructura del texto o su formato de presentación. Permiten hacer explícita la estructura de un documento, su contenido semántico o cualquier otra información lingüística o extralingüística que se quiera hacer patente.

Las marcas pueden representarse con distintos caracteres, como en TeX que se usa la barra invertida para el inicio de una marca, pero las más usadas son las que se expresan con el signo mayor que y el menor que < y >.

No puede compararse un lenguaje de marcas con un lenguaje de programación como C, Java o Pascal, pero si permite combinar dentro de documento lenguajes como PHP o JavaScript, aportando funcionalidad, tal y como hace HTML.

La herramienta fundamental para visualizar documentos HTML, XML, XHTML, etc es el navegador.

```
html>
  <head>
    <title>hola mundo</title>
  </head>
  <body>
  </body>
/html>
```

Si abrimos este fichero en un navegador, no se verá marca alguna y solo se visualizará la frase “hola mundo”.

Aunque en la práctica en un mismo documento pueden combinarse varios tipos diferentes de lenguajes de marca, éstos pueden clasificarse de la siguiente manera:

- **De presentación:** Define el formato del texto. Es complicado de mantener y modificar a pesar de ser fácil de crear. Se usa, por ejemplo, para maquetación de documentos. (Lo vemos en Microsoft WORD activando marcas o caracteres ocultos)

Huelva y sus playas
Playa de Punta Umbría
Con una longitud de 4 kilómetros,
junto al Paraje Natural de las
Marismas del Odiel, en una zona
donde abundan los enebros....

- **De procedimientos:** Orientado también a la presentación pero, en este caso, el programa que representa el documento debe interpretar el código en el mismo orden en que aparece. Se pueden usar etiquetas para centrar un texto, cambiar el tamaño de la letra,... Por ejemplo son de este tipo los lenguajes de marcado TeX, LaTeX, Nroff...

```
\documentclass[a4paper]{article}
\begin{document}
\bfLARGE(Huelva y sus playas)
\bfLarge(Playa de Punta Umbría)
\bigskip
\rm(Con una longitud de 4 kilómetros, junto al Paraje Natural de las Marismas del Odiel,
en una zona donde abundan los enebros...)
\end{document}
```

- **Descriptivos o semánticos:** Describen las diferentes partes en que se estructura un documento pero sin especificar como deben representarse. Son flexibles en cuanto a que usan etiquetas sin especificar la manera de representarlas ni el orden. Las marcas aportan información sobre la estructura y una descripción del contenido. (SGML, XML, HTML, XHTML(es html cumpliendo las especificaciones de xml))

```
<playas>
  <provincia>Huelva</provincia>
  <población>Punta Umbría</población>
  <descripción>
    Con una longitud de 4 kilómetros, junto al Paraje Natural de las Marismas del Odiel,
    en una zona donde abundan los enebros...
  </descripción>
</playas>
```



Además, si atendemos a su ámbito de utilización podemos clasificarlos en:

- Documentación electrónica:
 - RTF: Formato de Texto Enriquecido. Fue desarrollado por Microsoft en 1987 para el intercambio de documentos de texto entre procesadores de texto.
 - TeX: Para creación de ecuaciones matemáticas complejas.
 - WikiText: permite la creación de páginas wiki en servidores preparados para soportar ese lenguaje.
 - DocBook: Permite generar documentos separando la estructura lógica del documento de su formato.
- Tecnologías de internet:
 - HTML, XHTML: Para creación de páginas web.
 - RSS: Para la sindicación de contenidos en internet, permite la difusión de contenidos.
- Otros lenguajes especializados:
 - MathML: Para expresar el formalismo matemático e tal manera que pueda ser entendido por distintos sistemas y aplicaciones.
 - VoiceXML: Para intercambio de información entre el usuario y una aplicación con capacidad de reconocimiento de voz.
 - MusicXML: Permite intercambio de partituras entre distintos editores de las mismas.

Por último, el ámbito de aplicación de los lenguajes de marcas es muy diverso, ya que permite el intercambio de datos entre distintas aplicaciones, independientemente de la plataforma usada y la tecnología en la que estén creadas. Desde un fichero XML, pueden generarse vistas como HTML o pdf. Java EE usa ficheros XML Para poder especificar datos de configuración. Visual Studio, a la hora de crear servicios web, genera varios documentos con estructura XML. Puede usarse para bases de datos, frameworks de desarrollo, sistemas de publicación de contenidos, definición de interfaces gráficas, etc.

2. Características comunes

Vamos a ver las características comunes a la mayoría de lenguajes de marca que tienen cierta relevancia:

- Texto plano:
 - Se trata de un tipo de ficheros que contienen únicamente caracteres de texto, no permiten almacenar información con formato (color, tipo de letra...). El espacio en blanco, el retorno de carro y las tabulaciones se consideran caracteres)
 - Al ser solo dependiente del alfabeto que se utilice se utilizara una codificación específica como puede ser ASCII, UTF-8, Unicode o, por ejemplo, ISO 8859-1 para la codificación de alfabeto latino(ñ, acentos, etc.)
- Interoperabilidad o independencia:
 - Se pueden ver y editar con un simple editor de texto como el bloc de notas de Windows (aunque hay otros que facilitan mucho el trabajo)Por lo que es independiente de sistemas operativos y programas de software por tanto del dispositivo final que se utilice. Se puede abrir y editar desde cualquier máquina, teniendo en cuenta eso si la codificación empleada.
- Flexibles y fáciles de crear:
 - Como hemos dicho con un simple editar de pueden crear y editar en la extensión que se necesite.
 - Es posible también que en un solo documento se combinen distintos lenguajes de marcas, lo que le da una mayor funcionalidad. Y no solo combinando lenguajes de marcas entre si como puede ser XML con HTML, sino que es posible con otro tipo de lenguajes como HTML con PHP y Javascript.

3. Concepto de metalenguaje

Un metalenguaje es un lenguaje que se utiliza para decir algo acerca de otro. Por ejemplo XML no es un lenguaje en sí, sino que es más bien un metalenguaje que permite definir lenguajes de acuerdo a las necesidades.

4. Evolución de los lenguajes de marcas

El origen de los lenguajes de marcas radica en el lenguaje de marcado generalizado estándar SGML, un estándar que permite definir lenguajes de marcado, considerándose un metalenguaje.

SGML es descendiente de GML(Generalized markup language) de IBM. Se trataba de un formato de documentos cuyo objetivo era describir los textos de tal modo que el resultado fuese independiente de la plataforma y la aplicación utilizada . En 1986 GML se convirtió en SGML y dio lugar al estándar ISO 8879. Éste era un lenguaje muy complejo y requería de unas herramientas de software caras. Por ello su uso ha quedado relegado a grandes aplicaciones industriales.

En 1989/90 Tim Berners-Lee creó el World Wide Web y se encontró con la necesidad de organizar, enlazar y compatibilizar gran cantidad de información procedente de diversos sistemas. Para resolverlo creó un lenguaje de descripción de documentos llamado HTML, que, en realidad, era una combinación de dos estándares ya existentes:

- **ASCII:** Es el formato que cualquier procesador de textos sencillo puede reconocer y almacenar. Por tanto es un formato que permite la transferencia de datos entre diferentes ordenadores.
- **SGML:** Lenguaje que permite dar estructura al texto, resaltando los títulos o aplicando diversos formatos al texto.

HTML es una versión simplificada de SGML, ya que sólo se utilizaban las instrucciones absolutamente imprescindibles. Era tan fácil de comprender que rápidamente tuvo gran aceptación logrando lo que no pudo SGML. HTML se convirtió en un estándar general para la creación de páginas web. Además, tanto las herramientas de software como los navegadores que permiten visualizar páginas HTML son cada vez mejores.

A pesar de todas estas ventajas, HTML no es un lenguaje perfecto, sus principales desventajas son:

- No soporta tareas de impresión y diseño.
- El lenguaje no es flexible, ya que las etiquetas son limitadas.
- No permite mostrar contenido dinámico.
- La estructura y el diseño están mezclados en el documento.

Para resolver estos problemas se estableció el estándar internacional XML, un lenguaje de marcas puramente estructural que no incluye ninguna información relativa al diseño. Se convierte con rapidez en estándar para el intercambio de datos en la web. A diferencia de HTML las etiquetas indican el significado de los datos en lugar del formato con el que se van a visualizar éstos.