


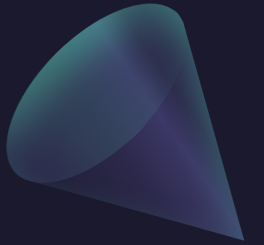
Tema 1.

Introducción a los
lenguajes de marcas



1. Definición y clasificación de lenguajes de marcas.

Los lenguajes de marcas (o lenguajes de marcado) sirven para 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 forma de presentarlo. El lenguaje de marcas es el que especifica cuáles serán las etiquetas posibles, donde deben colocarse y el significado que tendrá cada una de ellas. Permiten hacer explícita la estructura de un documento, su contenido semántico.



1. Definición y clasificación de lenguajes de marcas.

Aunque en la práctica, en un mismo documento pueden combinarse varios tipos diferentes de lenguajes de marcas, éstos se pueden clasificar en tres grupos:

Orientados a presentación: Son los utilizados generalmente por los procesadores de texto y codifican cómo ha de presentarse el documento, es decir, define el formato del texto. Por ejemplo: indicando que una determinada palabra se debe presentar en negrita, o que se debe dejar un espacio entre caracteres determinado. Generalmente, las marcas se ocultan al usuario, lo que permite obtener un efecto WYSIWYG .

De procedimientos: Orientados también a la presentación pero, en este caso, dentro de un marco procedural que permite definir macros (secuencias de acciones), es decir, el programa que representa el documento debe interpretar el código en el mismo orden en que aparece. Entre los ejemplos más comunes, encontramos TeX, LaTeX y Postscript.

Descriptivos o semánticos: Este tipo no define qué se debe hacer con un trozo o sección del documento, sino que por el contrario, las marcas sirven para indicar qué es esa información, es decir, describen qué es lo que se está representando. Describen las diferentes partes pero sin especificar cómo deben representarse.

1. Definición y clasificación de lenguajes de marcas.

Algunos ejemplos de lenguajes de marcado agrupados por su ámbito de utilización son:

Documentación electrónica

RTF (Rich Text Format): Formato de Texto Enriquecido, fue desarrollado por Microsoft en 1987. Permite el intercambio de documentos de texto entre distintos procesadores de texto.

TeX: Su objetivo es la creación de ecuaciones matemáticas complejas.

Wikitexto: Permite la creación de páginas wiki en servidores preparados para soportar este lenguaje.

DocBook: Permite generar documentos separando la estructura lógica del documento de su formato. De este modo, dichos documentos, pueden publicarse en diferentes formatos sin necesidad de realizar modificaciones en el documento original.

Tecnologías de internet

HTML, XHTML: (Hypertext Markup Language, eXtensible Hypertext Markup Language): Su objetivo es la creación de páginas web.

RSS: Permite la difusión de contenidos web

2. Evolución de los lenguajes de marcas

A finales de los años 60, para poder introducir anotaciones dentro de documentos electrónicos, surgen unos lenguajes informáticos, distintos de los lenguajes de programación, orientados a la gestión de información. Con el desarrollo de los editores y procesadores de texto surgen los primeros lenguajes informáticos especializados en tareas de descripción y estructuración de información: los lenguajes de marcas. Paralelamente, también, surgen otros lenguajes informáticos orientados a la representación, almacenamiento y consulta eficiente de grandes cantidades de datos: lenguajes y sistemas de bases de datos.

Los lenguajes de marcas surgieron, inicialmente, como lenguajes formados por el conjunto de códigos de formato que los procesadores de texto introducen en los documentos para dirigir el proceso de presentación (impresión) mediante una impresora. Como en el caso de los lenguajes de programación, inicialmente estos códigos de formato estaban ligados a las características de una máquina, programa o procesador de textos concreto y, en ellos, inicialmente no había nada que permitiese al programador (formateador de documentos en este caso) abstraerse de las características del procesador de textos y expresar de forma independiente a éste la estructura y la lógica interna del documento.

2. Evolución de los lenguajes de marcas

```
<times 14><color verde><centrado> Este texto es un ejemplo para mostrar la utilización primitiva de las marcas</centrado></color></times 14>  
<color granate><times 10><cursiva>Para realizar este ejemplo se utilizan etiquetas de nuestra invención. </cursiva>  
Las partes importantes del texto pueden resaltarse usando la  
<negrita>negrita</negrita>, o el <subrayar>subrayado</subrayar></times 10></color>
```

Al imprimirlo se obtendría:

Este texto es un ejemplo para mostrar la utilización primitiva de las marcas

*Para realizar este ejemplo se utilizan etiquetas de nuestra invención. Las partes importantes del texto pueden resaltarse usando la **negrita**, o el subrayado*

3. La popularización: HTML

En 1989/90 Tim Berners-Lee creó el World Wide Web y conociendo SGML, se encontró con la necesidad de organizar, enlazar y compatibilizar gran cantidad de información procedente de diversos sistemas. Para resolverlo, a partir de la sintaxis SGML, creó un lenguaje de descripción de documentos llamado HTML, siendo 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, siendo un rotundo éxito en la World Wide Web. 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.

4. XML

Como respuesta a problemas surgidos en torno al HTML, el W3C establece, en 1998, el estándar internacional XML, un lenguaje de marcas puramente estructural que no incluye ninguna información relativa al diseño, que permite crear etiquetas adaptadas a las necesidades (de ahí lo de "extensible"). Está convirtiéndose 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 los datos.

XML es un metalenguaje caracterizado por:

Permitir definir etiquetas propias.

Permitir asignar atributos a las etiquetas.

Utilizar un esquema para definir de forma exacta las etiquetas y los atributos.

La estructura y el diseño son independientes.