

Ciclos Formativos de Grado Superior

Lenguajes de Marcas y Sistemas de Gestión de la Información

Capítulo 1: Reconocimiento de las características de
lenguajes de marcas

Índice

1. Definición y clasificación de lenguajes de marcas
 2. Tipos de lenguajes de marcas
 3. Evolución de los lenguajes de marcas
 4. Etiquetas, elementos y atributos
 5. Organizaciones desarrolladoras
 6. Utilización de lenguajes de marcas en entornos web
 7. Gramáticas
-

Definición y clasificación de lenguajes de marcas

Los **lenguajes de marcas** (también llamados **lenguajes de marcado**) son aquellos que combinan la información, generalmente textual, que contiene un documento con marcas o anotaciones relativas a la estructura del texto o a la forma de representarlo. El lenguaje de marcas es el que especifica cuáles serán las etiquetas posibles, dónde deben colocarse y el significado que tendrá cada una de ellas.

Ejemplo:

<noticia>

<lugar>Madrid</lugar>

<fecha>27/08/2010</fecha>

<desc>Se ha inaugurado una estación de tren</desc>

</noticia>

Tipos de lenguajes de marcas

Lenguajes orientados a presentación. Este tipo de lenguajes son los usados tradicionalmente por los procesadores de texto como puede ser Microsoft Word® y codifican cómo ha de presentarse el documento, por ejemplo, indicando que una determinada palabra debe presentarse en fuente itálica o que se debe dejar un espacio de 10 puntos al terminar el párrafo.

Lenguajes procedurales. En este tipo de lenguajes las etiquetas son también orientadas a presentación pero se integran dentro de un marco procedural que permite definir macros (secuencias de acciones) y subrutinas. Entre los ejemplos más comunes de lenguajes procedurales podemos encontrar TeX, LaTeX y Postscript.

Lenguajes descriptivos. Este tipo de lenguajes no definen 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, esto es, describen que es lo que se está representando. La mayoría de los lenguajes de marcas que se usan hoy en día se encuentran dentro de este grupo como por ejemplo, el SGML y sus derivados (HTML, XML, etc.).

Evolución de los lenguajes de marcas

En la década de los 60 apareció el lenguaje **SGML** (*Standard Generalized Markup Language*), que es un descendiente directo del lenguaje **GML** propuesto por IBM. Este lenguaje surgió para permitir compartir información por parte de sistemas informáticos. Este estándar tuvo una gran aceptación pero no consiguió asentarse del todo debido principalmente a su complejidad lo que provocaba que el software que usará SGML terminaba siendo excesivamente extenso y complejo.

A finales de los 80 dentro del CERN se creó un lenguaje de marcado pensado para compartir información usando las redes de computadores y, de forma más general, a través de Internet. Este lenguaje se basaba en algunos principios de SGML y lo denominaron HTML (*Hyper-text Markup Language*).

Finalmente, en el 1998, W3C hizo público un nuevo estándar que denominaron XML (*eXtended Markup Language*), más sencillo que SGML y más potente que HTML.

Etiquetas, elementos y atributos

Una **etiqueta** (*tag*) es un texto que va entre el símbolo menor que (<) y el símbolo mayor que (>). Existen etiquetas de inicio (como <nombre>) y etiquetas de fin (como </nombre>).

Los **elementos** representan estructuras mediante las que se organizará el contenido del documento o acciones que se desencadenan cuando el programa navegador interpreta el documento. Constan de la etiqueta de inicio, la etiqueta de fin y de todo aquello que se encuentra entre ambas. Algunos elementos no tienen contenido. Se les denomina elementos vacíos y no deben llevar etiqueta de fin.

Un **atributo** es un par nombre-valor que se encuentra dentro de la etiqueta de inicio de un elemento e indican las propiedades que pueden llevar asociadas los elementos.

Etiquetas, elementos y atributos

EJEMPLO

```
<direccion>  
  <nombre>  
    <titulo>Mrs.</titulo>  
    <nombre> Mary </nombre>  
    <apellidos> McGoon </apellidos>  
  </nombre>  
  <calle> 1401 Main Street </calle>  
  <ciudad estado="NC"> Anytown</ciudad>  
  <codigo-postal> 34829 </codigo-postal>  
</direccion>
```

En el ejemplo anterior, el elemento *<nombre>* contiene tres elementos hijos: *<titulo>*, *<nombre>* y *<apellidos>* y *estado* es un atributo del elemento *<ciudad>*.

Organizaciones desarrolladoras

Organización Internacional para la Estandarización (ISO, *International Organization for Standardization*). Se formó después de la Segunda Guerra Mundial (23 de febrero de 1947) y es el organismo encargado de promover el desarrollo de normas internacionales de fabricación, comercio y comunicación para todas las ramas industriales a excepción de la eléctrica y la electrónica. Su función principal es la de buscar la estandarización de normas de productos y seguridad para las empresas u organizaciones a nivel internacional.

World Wide Web Consortium (W3C). El W3C se creó en 1994 por Tim Berners-Lee en el MIT, actual sede central del consorcio. Posteriormente se unió, en abril de 1995, el INRIA en Francia, reemplazado por el ERCIM en el 2003 como el huésped europeo del consorcio y la Universidad de Keiō (ShonanFujisawaCampus) en Japón en septiembre de 1996 como huésped asiático. Su función principal es tutelar el crecimiento y organización de la web.

Utilización de lenguajes de marcas en entornos web

Una **página web** es un documento electrónico adaptado para la *World Wide Web* que, normalmente, forma parte de un sitio web. Está compuesta, principalmente, por información (solo texto o módulos multimedia) así como por hiperenlaces; además, puede contener o asociar datos de estilo para especificar cómo debe visualizarse, y también aplicaciones embebidas para hacerla interactiva.

Las páginas web están escritas en un lenguaje de marcas que proporciona la capacidad de manejar e insertar hiperenlaces, generalmente, HTML.

El contenido de la página puede ser predeterminado (**página web estática**) o generado en el momento de su visualización o al solicitarla a un servidor web (**página web dinámica**).

Respecto a la estructura de las páginas web, algunos organismos, en especial el W3C, suelen establecer directivas con la intención de normalizar el diseño, para así facilitar y simplificar la visualización e interpretación del contenido.

Gramáticas

- La **DTD (Definición de Tipo de Documento)** que establece las reglas de formación del lenguaje formal, es decir, qué combinaciones de símbolos elementales son sintácticamente correctas. La especificación del W3C para HTML 4.0 contempla:
 - **DTD estricta** : incluye todos los elementos y atributos que no han sido declarados “desaprobados” (*deprecated*).
 - **DTD transicional o flexible -loose-**: incluye todo lo que la anterior más los elementos y atributos desaprobados (*deprecated*).
 - **DTD para documentos con marcos** : engloba todo lo incluido en la transicional más lo relativo a la creación de documentos con marcos (*frames*).
 - **XML Schema** es la evolución de la DTD descrita por el W3C, también denominado XSD (*XML SchemaDefinition*). Es un lenguaje de esquema más complejo y más potente, basado en la gramática para proporcionar una potencia expresiva mayor que la DTD. Utiliza sintaxis XML que le permite especificar de forma más detallada un extenso sistema de tipos de datos.
 - **RELAX NG** es un lenguaje de esquema basado en la gramática, muy intuitivo y más fácil de entender que el XML Schema. Tiene un alto poder expresivo, ya que, por ejemplo, permite validar elementos intercalados que pueden aparecer en cualquier orden.
-