

UD1.1 Introducción a los lenguajes de marcas

Introducción

Un **lenguaje de marcas** (o lenguaje de marcado) 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.

Todo lenguaje de marcas está definido en un documento denominado **DTD** (*Document Type Definition*). En él se establecen:

- Marcas: los elementos utilizados por dicho lenguaje
- Etiquetas y atributos
- Sintaxis
- Normas de uso

EJEMPLO

Supongamos que tenemos la siguiente información:

```
22/11/2006
Estimado cliente:
bla bla bla bla ...
Don José Gutiérrez González
```

El aspecto de un documento realizado en un lenguaje de marcas es el siguiente:

```
<carta>
  <fecha>22/11/2006</fecha>
  <presentacion>Estimado cliente:</presentacion>
  <contenido>bla bla bla bla ...</contenido>
  <firma>Don José Gutiérrez González</firma>
</carta>
```

Tipos de lenguajes de marcas

En la práctica, en un mismo documento pueden combinarse varios tipos diferentes de lenguajes de marcas.

Los lenguajes de marcas se pueden clasificar como sigue:

- **De presentación:** define el formato del texto.
- **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.
- **Descriptivo o semántico:** describen las diferentes partes en las que se estructura el documento pero sin especificar cómo deben representarse.

Ámbitos de aplicación

A continuación, se muestran algunos ejemplos de lenguajes de marcado agrupados por su ámbito de utilización.

Documentación electrónica

- **Markdown:** creado en 2004, es un lenguaje de marcado ligero ampliamente en la actualidad, especialmente en el ámbito de desarrollo de software.
- **RTF** (*Rich Text Format*) o *Formato de Texto Enriquecido*: fue desarrollado por Microsoft en 1987. Permite el intercambio de documentos de texto entre distintos procesadores de texto.
- **TeX** (*Text EXecutive processor*): 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** (*Hypertext Markup Language*): uno de los más populares con diferencia. Su objetivo es la creación de páginas web.
- **XHTML** (*eXtensible Hypertext Markup Language*): es idéntico a HTML, pero expresado de otra forma (representado como XML válido).

- **RSS** (*Really Simple Syndication*): permite la difusión de contenidos web. Utilizado por agregadores de noticias para extraer la información de diferentes medios de comunicación, blogs, portales de información, etc.

Multimedia

- **SVG** (*Scalable Vector Graphics*): permite definir gráficos vectoriales de dos dimensiones y soporta animación.
- **VoiceXML** (*Voice Extended Markup Language*): tiene como objetivo el intercambio de información entre un usuario y una aplicación con capacidad de reconocimiento de habla.
- **MusicXML** (*Music Extended Markup Language*): permite el intercambio de partituras entre distintos editores de partituras.

Otras aplicaciones

- **MathML** (*Mathematical Markup Language*): su objetivo es expresar el formalismo matemático de tal modo que pueda ser entendido por distintos sistemas y aplicaciones.