

INTRODUCCIÓN

Los problemas relacionados con el intercambio de información entre aplicaciones y máquinas informáticas es tan viejo como la propia informática. El problema parte del hecho de haber realizado un determinado trabajo con un software en un ordenador concreto y después querer pasar dicho trabajo a otro software en ese u otro ordenador. En la informática actual esto sigue siendo un problema porque además del problema mencionado se tiene una necesidad de disponibilidad global del trabajo y una necesidad de ver dicho trabajo en dispositivos de todo tipo como mini ordenadores, tablets, smartphones, etc.

Como vemos poco a poco han aparecido formatos binarios de archivos que han sido estándares de facto como por ejemplo el formato documental PDF, el formato de imagen JPEG, el formato de música MP3 o el formato MPEG de vídeo. Estos formatos se han extendido tanto y son tan populares que cualquier dispositivo actual proporciona al usuario una aplicación capaz de interpretarlos. Gracias a esto podemos compartir archivos entre distintos dispositivos.

Pero esto solo es una solución parcial al problema porque sigue habiendo empresas que utilizan su formato binario propio por la idea de que sus formatos de archivo están directamente relacionados con la calidad de su software o simplemente porque necesitan un formato más potente. Por tanto vuelven a aparecer problemas cuando se quieren compartir datos usando esos formatos propios. En estos casos, la solución para exportar e importar datos es utilizar conversores capaces de convertir los datos de un formato a otro (por ejemplo de Microsoft Word a Open Office Writer, de AVI a MOV, etc.). El problema reside en que esas conversiones no son fieles al original al 100%.

A pesar de todo esto, existe un formato de archivo que cualquier dispositivo y sistema operativo sí es capaz de entender: el texto. Los archivos llamados de texto, sólo son capaces de almacenar texto plano; es decir sólo texto sin indicar ningún formato.

Debido a la facilidad de ser leído con cualquier dispositivo, se pensó que el propio texto podría servir para almacenar otros datos. Evidentemente no es posible usar texto para almacenar por ejemplo imágenes o audio, pero sí otras cosas tal y como veremos a lo largo del tema.

De ahí surge la idea de los lenguajes de marcas. Esta idea consiste en que dentro del archivo de texto habrá contenido que no se interpretará como texto informativo, sino que hay texto en el archivo que se marca de manera especial haciendo que signifique otra cosa.

DEFINICIÓN DE LENGUAJE DE MARCAS

¿Qué es un lenguaje 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. Es el que especifica cuáles serán las etiquetas posibles, dónde deben colocarse y el significado que tendrá cada una de ellas.

Lo más interesante de esto es que tú puedes crear tu propio lenguaje de marcas definiendo las etiquetas que necesitas y las reglas que se deben cumplir para usar esas etiquetas adaptadas al problema determinado que debes resolver o bien, puedes usar un lenguaje de marcas ya existente respetando las normas que definen para su uso. Existen numerosos lenguajes de marcas y suelen ser definidos y mantenidos por distintas organizaciones, como por ejemplo el W3C, Sun Microsystems o DCMI (Dublin Core Metadata Initiative).

LENGUAJES DE MARCAS-TEMA 1

¿Qué ventajas tiene un lenguaje de marcas?

En primer lugar para intercambiar información entre aplicaciones tanto web como de escritorio porque estamos almacenando la información en ficheros de texto plano.

Los lenguajes de marcado suelen confundirse con lenguajes de programación. Sin embargo, no son lo mismo, ya que el lenguaje de marcado no tiene funciones aritméticas o variables, como sí poseen los lenguajes de programación.

EVOLUCIÓN DE LOS LENGUAJES DE MARCAS

AÑOS 60:

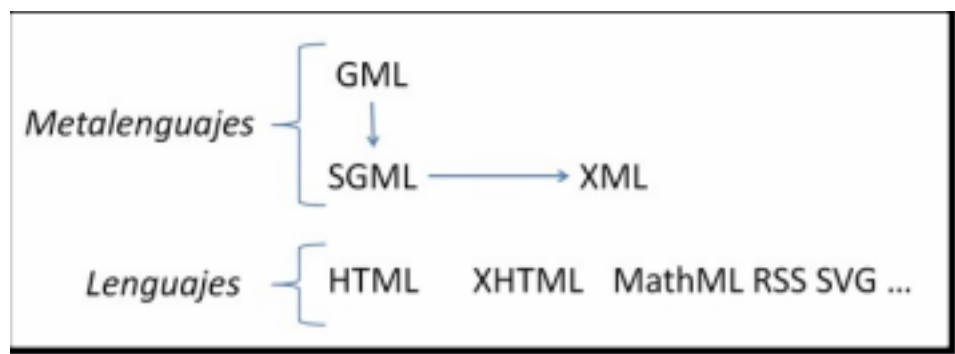
- El primer lenguaje que usaba marcas fue el lenguaje GML (General Markup Language) creado por IBM. No se trataba de un lenguaje de marca propiamente dicho, pero fue la semilla. Se usa para que las empresas pudiesen guardar grandes cantidades de información sobre cualquier tema. El GML nació a finales de los años 60 y tuvo una gran aceptación general, pero era realmente complejo de utilizar.
- Poco después apareció el lenguaje SGML (Standard Generalized Markup Language), que es un descendiente directo del lenguaje GML definido por la ISO. El objetivo de este lenguaje era también compartir información por parte de sistemas informáticos, pero tampoco consiguió asentarse del todo debido a que provocaba que el software que usaba SGML terminaba siendo excesivamente extenso y complejo.

AÑOS 80:

- A finales de los 80 dentro del CERN se creó un lenguaje de marcado más sencillo que SGML y sobre todo pensado para compartir información usando redes de computadores. Este lenguaje se basaba en algunos principios de SGML y lo denominaron HTML (Hyper-text Markup Language).
- Con este lenguaje se quiso hacer más fluido el intercambio de información entre empresas y universidades usando las redes de computadores que estaban empezando a surgir.

AÑOS 90:

- 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.
- Pero... ¿qué es XML en realidad?



LENGUAJES DE MARCAS-TEMA 1

CARACTERÍSTICAS DE LOS LENGUAJES DE MARCAS

Texto plano:

- ✓ Una de las principales ventajas de los ficheros creados usando lenguajes de marcas es que pueden ser interpretados directamente, dado que son archivos de texto plano.
- ✓ Esto es una ventaja evidente respecto al los sistemas de archivos binarios, que requieren siempre de un programa intermediario para trabajar con ellos. Un documento escrito con lenguajes de marcado puede ser editado por un usuario con un sencillo editor de textos.
- ✓ Al tratarse solamente de texto, los documentos son independientes de la plataforma, sistema operativo o programa con el que fueron creados. Esta fue una de las premisas de los creadores de GML en lo años 70, para no añadir restricciones innecesarias al intercambio de información.

Facilidad de procesamiento:

- ✓ Los lenguajes de marcas son muy fáciles de procesar y esto ha permitido que las organizaciones de estándares hayan desarrollado lenguajes especializados para los tipos de documentos de comunidades o industrias concretas.
- ✓ Uno de los primeros fue el CALS, utilizado por las fuerzas armadas de EE.UU. para sus manuales técnicos. Otras industrias con necesidad de gran cantidad de documentación, como las de aeronáutica, telecomunicaciones, automoción o hardware, ha elaborado lenguajes adaptados a sus necesidades.

Flexibilidad

- ✓ Aunque originalmente los lenguajes de marcas se idearon para documentos de texto, se han empezado a utilizar en áreas como gráficos vectoriales, servicios web, sindicación web o interfaces de usuario.
- ✓ Estas nuevas aplicaciones aprovechan la sencillez y potencia del lenguaje XML. Esto ha permitido que se pueda combinar varios lenguajes de marcas diferentes en un único archivo, como en el caso de XHTML+SMIL (permite integrar audio, video, imágenes, texto o cualquier otro contenido multimedia en un página web) y de XHTML+MathML+SVG (permite integrar fórmula matemáticas y gráficos vectoriales en una página web).

Compacidad

- ✓ Las instrucciones de marcado se entremezclan con el propio contenido <p>Hola Mundo</p>
- ✓ El texto entre las marcas es el propio contenido del documento.

LENGUAJES DE MARCAS-TEMA 1

TIPOS DE LENGUAJES DE MARCAS

Los lenguajes de marcas pueden agruparse en tres grandes grupos:

a) 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. Generalmente en este tipo de lenguajes las marcas son ocultas al usuario.

Un lenguaje de marcado o lenguaje de marcas es una forma de codificar un documento que, junto con el texto, incorpora etiquetas o marcas que contienen información adicional acerca de la estructura del texto o su presentación.¶¶

¶¶

El lenguaje de marcas más extendido es el HTML ("HyperText Markup Language", Lenguaje de marcado de hipertexto), fundamento del World Wide Web (entramado de comunicación de alcance mundial).¶¶

¶¶

Los lenguajes de marcado suelen confundirse con lenguajes de programación. Sin embargo, no son lo mismo, ya que el lenguaje de marcado no tiene funciones aritméticas o variables, como sí poseen los lenguajes de programación. Históricamente, el marcado se usaba y se usa en la industria editorial y de la comunicación, así como entre autores, editores e impresores.¶¶

b) 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.

```
\documentclass{article}
% pre\ambulo

\usepackage{lmodern}
\usepackage[T1]{fontenc}
\usepackage[spanish,activeacute]{babel}
\usepackage{mathtools}

\title{Hola Mundo}
\author{Escribe aquí 'i tu nombre}

\begin{document}
% cuerpo del documento

\maketitle

Mi primer documento en \LaTeX{}.

\end{document}
```

c) 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.

```
<CATALOG>
  <CD>
    <TITLE>Empire Burlesque</TITLE>
    <ARTIST>Bob Dylan</ARTIST>
    <COUNTRY>USA</COUNTRY>
    <COMPANY>Columbia</COMPANY>
    <PRICE>10.90</PRICE>
    <YEAR>1985</YEAR>
  </CD>
  <CD>
    <TITLE>Maggie May</TITLE>
    <ARTIST>Rod Stewart</ARTIST>
    <COUNTRY>UK</COUNTRY>
    <COMPANY>Pickwick</COMPANY>
    <PRICE>8.50</PRICE>
    <YEAR>1990</YEAR>
  </CD>
</CATALOG>
```

ETIQUETAS, ELEMENTOS Y ATRIBUTOS

Etiquetas (tag):

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>).

Elementos:

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.

Atributos:

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.

GRAMÁTICAS

¿Qué es una gramática?

Según la R.A.E., la gramática es la ciencia que estudia los elementos de una lengua y sus combinaciones. En lenguajes de marcas tiene el mismo significado. Todo documento de un lenguaje de marcas tiene en común una gramática que define el marcado permitido, requerido y cómo debe usarse el marcado en ese determinado documento.

DTD (Definición de Tipo de Documento)

Establece las reglas de formación del lenguaje formal, es decir, qué combinaciones de símbolos elementales son sintácticamente correctas.

XML Schema

Es la evolución de la DTD descrita por el W3C, también denominado XSD (XML Schema Definition). 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.

LENGUAJES DE MARCAS-TEMA 1

PRINCIPALES LENGUAJES DE MARCAS

HTML: Destinado a la creación de páginas web. Permite mezclar texto y contenido multimedia en un mismo documento. EL W3C se encarga de su mantenimiento y estandarización.

XML: Permite definir lenguajes específicos para diferentes necesidades ->metalenguaje. Se define por reglas y no por etiquetas. Algunos de los lenguajes basados en XML son:

- ✓ XHTML: usado para presentar información en la web.
- ✓ MathML: usado para intercambiar información en ámbito matemático.
- ✓ SVG: usado para representar gráficos vectoriales.
- ✓ RSS y Atom: usado para difundir información actualizada a usuarios suscritos a una web.
- ✓ Open Document: usado para almacenar y presentar información en un paquete ofimático.
- ✓ Open eBook: usado en el formato ePub por libros electrónicos.

JSON: Alternativa a XML para almacenamiento e intercambio de datos entre sistemas, independientemente de la plataforma y del lenguaje de programación.