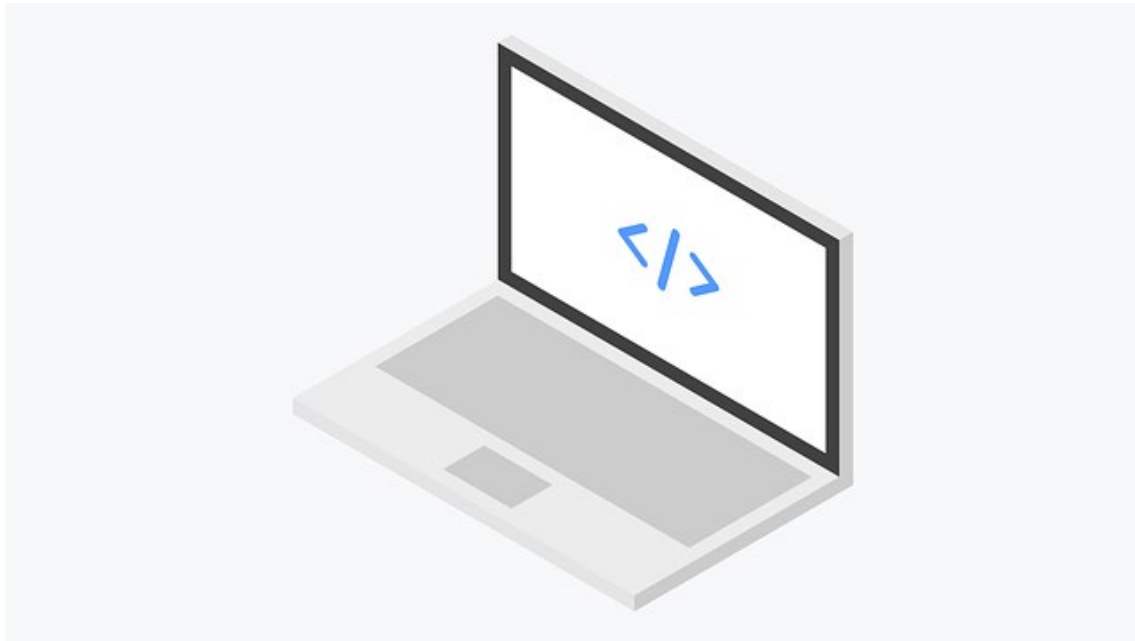


UT1. Introducción a los Lenguajes de Marcas.



Índice

UT1. Introducción a los Lenguajes de Marcas.....	1
1.1 Definición y clasificación de lenguajes de marcas	2
1.2 Tipos de lenguajes de marcas.....	2
1.3 Evolución de los lenguajes de marcas	3
1.5 Organizaciones desarrolladoras.....	4
1.6 Utilización de lenguajes de marcas en entornos web	4
1.7 Gramáticas	5
1.7.1 DTD	5
1.7.2 Esquema XML	5
1.7.3 RELAX NG	6

1.1 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 o cualquier otra información lingüística o extralingüística que se quiera hacer patente.

A continuación, se muestra un ejemplo en el que mediante una serie de marcas o etiquetas se ha representado una información relativa a una noticia:

```
<noticia>  
  <lugar>Madrid</ lugar>  
  <fecha>27/08/2010</ fecha>  
  <desc>Se ha inaugurado una estación de tren</desc>  
</noticia>
```

Los lenguajes de marcas **han de diferenciarse de los lenguajes de programación**. El lenguaje de marcas no tiene funciones aritméticas o variables, como sí poseen los lenguajes de programación.

1.2 Tipos 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.**

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**. 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.).

1.3 Evolución de los lenguajes de marcas

Los lenguajes de marcas **comenzaron a usarse** a finales de la década de los 60 **para poder introducir anotaciones dentro de documentos electrónicos**, de la misma forma que se hacía cuando la documentación estaba en papel. De esta posibilidad de incorporar marcas es de donde reciben su nombre. Es en esas fechas cuando se estandariza 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 **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)**.

La aparición de este lenguaje supuso de alguna manera una revolución en la forma de compartir información, gracias principalmente a la **sencillez de sus sintaxis y del software necesario para interpretarlo**. En poco tiempo el lenguaje HTML se extendió y empezó a crecer de forma en ocasiones descontrolada y casi siempre influenciado por razones meramente comerciales.

En el 1998, **W3C (World Wide Web Consortium)** hizo público un nuevo estándar que denominaron **XML (eXtended Markup Language)**, más sencillo que SGML y más potente que HTML.

Resumiendo, el **XML sirve para describir información** y **HTML sirve para darle formato y presentarla a través de un navegador**.

1.4 Etiquetas, elementos y atributos

Existen tres términos comúnmente usados para describir las partes de un documento de lenguajes de marcas: **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** 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.

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>.

Actividad 1.2: ¿Como seria el marcado de una biblioteca para un libro?

1.5 Organizaciones desarrolladoras

Dentro de las organizaciones que se han encargado de desarrollar los lenguajes de marcas se encuentran:

- **Organización Internacional para la Estandarización (ISO, International Organization for Standardization).** Su función principal es la de buscar la estandarización de normas de productos y seguridad para las empresas u organizaciones a nivel internacional. Las normas desarrolladas por ISO son voluntarias, ya que es un organismo no gubernamental y no depende de ningún otro organismo internacional, por tanto, no tiene autoridad para imponer sus normas a ningún país. Esta organización después del éxito que tuvo **GML** y, publicó en 1986 el Standard Generalized Markup Language (**SGML**) con rango de Estándar Internacional con el código **ISO 8879**.
- **World Wide Web Consortium (W3C).** Su función principal es tutelar el crecimiento y organización de la web. Su primer trabajo fue normalizar el lenguaje **HTML**, el lenguaje de marcas con el que se escriben las páginas web y posteriormente **XML**.

1.6 Utilización de lenguajes de marcas en entornos web

Una **página web** es un documento electrónico adaptado para la **World Wide Web (Internet)** 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**).

La W3C, establece directivas con la intención de normalizar el diseño, para así facilitar y simplificar la visualización e interpretación del contenido.

1.7 Gramáticas

Todo documento de un lenguaje de marcas tiene en común una gramática que **define el marcado permitido en esa clase, el marcado requerido y cómo debe ser utilizado** dicho marcado en la instancia del documento.

Al igual que un idioma tiene una gramática, un lenguaje de marcas posee la suya.

1.7.1 DTD

La gramática **DTD** (Definición de Tipo de Documento), **establece las reglas de formación del lenguaje formal**, es decir, qué **combinaciones de símbolos elementales** son sintácticamente correctas.

Para definir un párrafo se emplea los símbolos elementales <, p y > para formar <p> y no otro como por ejemplo >p<.

En la DTD se **identifica la estructura del documento**, es decir, aquellos elementos que son necesarios en la elaboración de un documento o un grupo de documentos estructurados de manera similar. **Contiene las reglas de dichos elementos: el nombre, su significado, dónde pueden ser utilizados y qué pueden contener.**

Leer: <https://disenowebakus.net/doctype-html5-definicion-de-tipo-de-documento.php>

1.7.2 Esquema XML

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, cosa que le permite especificar de forma más detallada un extenso sistema de tipos de datos. A diferencia de las DTD, soporta la extensión del documento sin problemas.

A la hora de la validación del documento, la utilización de **XSD supone un gran consumo en recursos y tiempo debido a su gran especificación y complejidad en la sintaxis** (los esquemas son más difíciles de leer y de escribir).

El modelo de datos de XML Schema incluye:

- **El vocabulario (nombres de elemento y atributo).**
- **El contenido modelo (relaciones y estructura).**
- **Los tipos de datos.**

1.7.3 RELAX NG

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.