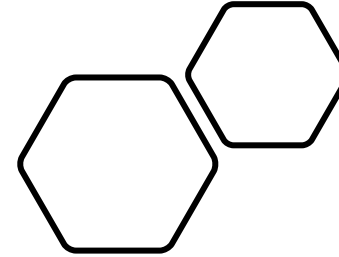


UD 1: Introducción a los lenguajes de marcas



Lenguajes de
marcas y sistemas
de gestión de
información

Introducción a los lenguajes de marcas

- Los lenguajes de marcas existen desde los comienzos de la computación, y tienen el objetivo de **intercambiar información entre sistemas**, usando un lenguaje sencillo.
- Su uso es imprescindible para el funcionamiento de la *World Wide Web*.
- Existen muchos lenguajes de marcas diferentes, pero nos centraremos en los más empleados en la actualidad.

¿Qué es un lenguaje de marcas?

- Un lenguaje de marcas es aquel que permite almacenar, además de la información “en bruto”, **marcas o etiquetas que nos proporcionan información sobre esa información**: su estructura, significado, cómo debe representarse, etc.
- Un lenguaje de marcas **no es un lenguaje de programación**.
- Uno de los lenguajes de marcas más empleado es **HTML**. Con HTML indicamos la **estructura de los datos** de una página web (es decir, le da formato a la página web).
- Existen otros lenguajes de marcas más orientados al almacenamiento e intercambio de información como XML, JSON, etc.

Ejemplo de uso de etiquetas en HTML

Código HTML

```
<h1>Esto es un encabezado  
de primer nivel</h1>  
<p>Esto es un párrafo</p>
```

```
<p>Esto es un encabezado  
de primer nivel</p>  
<h1>Esto es un  
párrafo</h1>
```

Resultado

**Esto es un encabezado de
primer nivel**

Esto es un párrafo

Esto es un encabezado de primer nivel

Esto es un párrafo

Ejemplo de uso de etiquetas en HTML

- En el ejemplo anterior, la etiqueta <h1> indica que el contenido es una cabecera de primer nivel, esto quiere decir que es, por ejemplo, el título de un artículo, mientras que los capítulos y/o secciones del artículo usarían etiquetas de cabecera de menor nivel (<h2>). La etiqueta <p> sirve para indicar un párrafo.
- Al intercambiar las etiquetas se puede observar que cambia la presentación del texto, ya que **el navegador solo entiende las etiquetas:** no interpreta el contenido de las mismas.
- Por tanto, HTML nos sirve para definir la **estructura** de un documento, así como también para obtener **información semántica** del contenido.

Ejemplo de lenguaje de marcas propio

- Un lenguaje de marcas es, en su forma más simple, un conjunto de **etiquetas** y las **reglas** para usar e interpretar esas etiquetas.
- Podemos definir nuestro propio lenguaje de marcas si definimos las etiquetas y las reglas. Un ejemplo sencillo:
 - En nuestro lenguaje de marcas usaremos la etiqueta :titulo: para indicar el título de un libro. Se representará con tipografía Arial, color rojo y tamaño de 18 puntos.
 - Usaremos la etiqueta :autor: para indicar el autor de un libro. Se representará con tipografía Arial, color azul y tamaño de 14 puntos.
 - Se usarán :/titulo: y :/autor: como etiquetas de cierre.

Ejemplo de lenguaje de marcas propio

- Si desarrollamos un programa que sea capaz de entender y representar nuestro lenguaje de marcas obtendremos el siguiente resultado:

Codificación

:titulo: El origen de las especies :/titulo:

:autor: Charles Darwin :/autor:

:titulo: Piensa en Java :/titulo:

:autor: Bruce Eckel :/autor:

Resultado

El origen de las especies

Charles Darwin

Piensa en Java

Bruce Eckel

Estandarización de los lenguajes de marcas

- Como hemos visto, crear un lenguaje de marcas es bastante sencillo, pero el problema es que tenemos que desarrollar el software capaz de interpretar y representar el lenguaje.
- Si cada uno crea su propio lenguaje de marcas para representar la información no sería posible el intercambio de información como se conoce hoy en día en Internet.
- Para solucionar este problema, los lenguajes de marcas más usados en la red están **estandarizados**.
- La estandarización permite que todos los programas que implementen el estándar puedan intercambiar información entre sí.
- HTML está estandarizado por la *World Wide Web Consortium (W3C)*.

Lenguajes y metalenguajes

- Hasta ahora hemos visto los principios generales de funcionamiento de un lenguaje de marcas. Ahora es necesario establecer una distinción entre lenguajes y metalenguajes:
- Un lenguaje de marcas establece las etiquetas necesarias para su interpretación por un software. Un ejemplo es HTML, que tiene sus etiquetas definidas en el estándar, con su sintaxis y significado.
- Un metalenguaje, sin embargo, no define las etiquetas, sino el conjunto de reglas que debe seguir un lenguaje para definir sus etiquetas. Es, en esencia, un conjunto de instrucciones para construir un lenguaje de marcas. Podemos considerar a XML un metalenguaje.

Estructura de un lenguaje de marcas

- Tomaremos como ejemplo XML, que veremos con detalle en la [UD 4: Creación y validación XML](#).
 - Todo elemento en XML tiene que ir precedido de una **etiqueta de apertura**, consistente en un **nombre** rodeado de los símbolos **<** y **>** y seguido por una **etiqueta de cierre**, consistente en el mismo nombre de la etiqueta de apertura y rodeado por los símbolos **</** y **>**.
 - El contenido de un elemento puede ser texto u otro(s) elemento(s).
 - Existe un único elemento raíz, en el que están contenidos todos los elementos del documento.
 - Un elemento puede tener entre 0 y n atributos, que siguen el formato nombre="valor".

Estructura de un lenguaje de marcas

- Podemos poner un ejemplo de documento XML siguiendo estas normas:

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<libros>
  <libro ebook="si">
    <titulo> El origen de las especies </titulo>
    <autor> Charles Darwin </autor>
  </libro>
  <libro ebook="si">
    <titulo> Guerra y paz </titulo>
    <autor> León Tolstoi </autor>
  </libro>
  <libro ebook="no">
    <titulo> El conde de Montecristo </titulo>
    <autor> Alejandro Dumas </autor>
  </libro>
</libros>
```

Estructura de un lenguaje de marcas

- A diferencia de HTML, las etiquetas que hemos usado en el anterior ejemplo no están definidas en ningún estándar. Es el creador del documento XML el que decide el nombre de las etiquetas, los atributos y la estructura. Esto es lo que convierte a XML en un metalenguaje.
- Como las reglas para la creación de un documento XML han sido respetadas, el anterior es un **documento XML válido**.
- Aunque HTML tiene una estructura bien definida y es similar a XML, tiene sus etiquetas perfectamente definidas en el estándar y no permite la creación de etiquetas nuevas por parte del usuario.

Actividad LM1.1

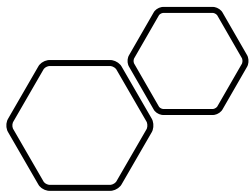
Crea un documento discos.xml, usando las etiquetas que consideres necesarias, para almacenar la siguiente información:

Discografía de Daft Punk

Título	Ventas	Fecha salida
Homework	2300000	20-01-1997
Discovery	3000000	13-03-2001
Human After All	1200000	14-03-2005
Random Access Memories	5000000	17-05-2013

Características fundamentales de los lenguajes de marcas

- **Independencia** de plataforma hardware y software.
- **Codificado en texto plano.**
- **Flexibilidad** para adaptarse a varios contextos.
- **Compactabilidad:** tanto las marcas como el contenido están contenidos en un mismo documento.
- **Facilidad de procesamiento:** su formato permite que el procesador automatizado sea sencillo.



Lenguajes de marcas más comunes

HTML

- Es el lenguaje de marcas más conocido y empleado en la actualidad.
- Actualmente en su versión HTML5, aparecida por primera vez en 2014 y actualizada desde entonces.
- Se usa para definir la estructura de las páginas web.
- Es necesario un navegador web para procesarlas.
- La extensión de los documentos HTML es .html o .htm
- La versión HTML5 es muy popular y ha desplazado el uso de XHTML.
- Estandarizado por W3C.

XML

- Se usa principalmente para organizar la información para el procesamiento automatizado.
- Si HTML se usa para estructurar la información para el entendimiento humano, XML la estructura para el entendimiento de los sistemas de procesamiento.
- No solo se usa en ámbito web, se puede usar también, por ejemplo, para almacenar información de la configuración de una aplicación.
- La extensión de un documento XML es .xml
- Al igual que HTML está estandarizado por W3C.

JSON

- Acrónimo de *JavaScript Object Notation*.
- Es el formato en el que JavaScript representa los objetos.
- Aunque tiene vinculación con JavaScript, se ha extendido a otros ámbitos como alternativa a XML.
- Es más ligero y tiene una sintaxis más sencilla que XML.
- Pero no tiene tanta potencia como XML.
- Dependiendo del contexto será más apropiado usar JSON o XML.
- La extensión de un documento JSON es .json.

SVG

- *Scalable Vector Graphics.*
- Se usa para representar gráficos vectoriales bidimensionales.
- Su extensión es .svg o .svgz.
- Podemos editar documentos SVG mediante un editor de texto o con programas gráficos como *Inkscape*.

Interfaces de usuario

- **FXML**

- Dialecto de XML.
- Se usa para representar las interfaces de usuario en la tecnología JavaFX.

- **XAML**

- *eXtensible Application Markup Language*.
- Es el lenguaje en el que se definen las interfaces de usuario en las aplicaciones de Windows desarrolladas con la tecnología de Microsoft.

- **Interfaces de usuario de Android**

- Generalmente las interfaces de usuario en Android se diseñan mediante un dialecto de XML.