

Unidad 1

Introducción a los Lenguajes de Marcas

LENGUAJES DE MARCAS Y SISTEMAS DE GESTIÓN DE INFORMACIÓN

ÍNDICE DE CONTENIDOS

1. Introducción
2. Definición de lenguaje de marcas
3. Tipos de lenguajes de marcas
4. Etiquetas, elementos y atributos
5. Evolución de los lenguajes de marcas
6. Características de los lenguajes de marcas
7. Gramáticas
8. Ejemplos de uso
9. Bibliografía

ÍNDICE DE CONTENIDOS

1. Introducción
2. Definición de lenguaje de marcas
3. Tipos de lenguajes de marcas
4. Etiquetas, elementos y atributos
5. Evolución de los lenguajes de marcas
6. Características de los lenguajes de marcas
7. Gramáticas
8. Ejemplos de uso
9. Bibliografía

1.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.

1.Introducción

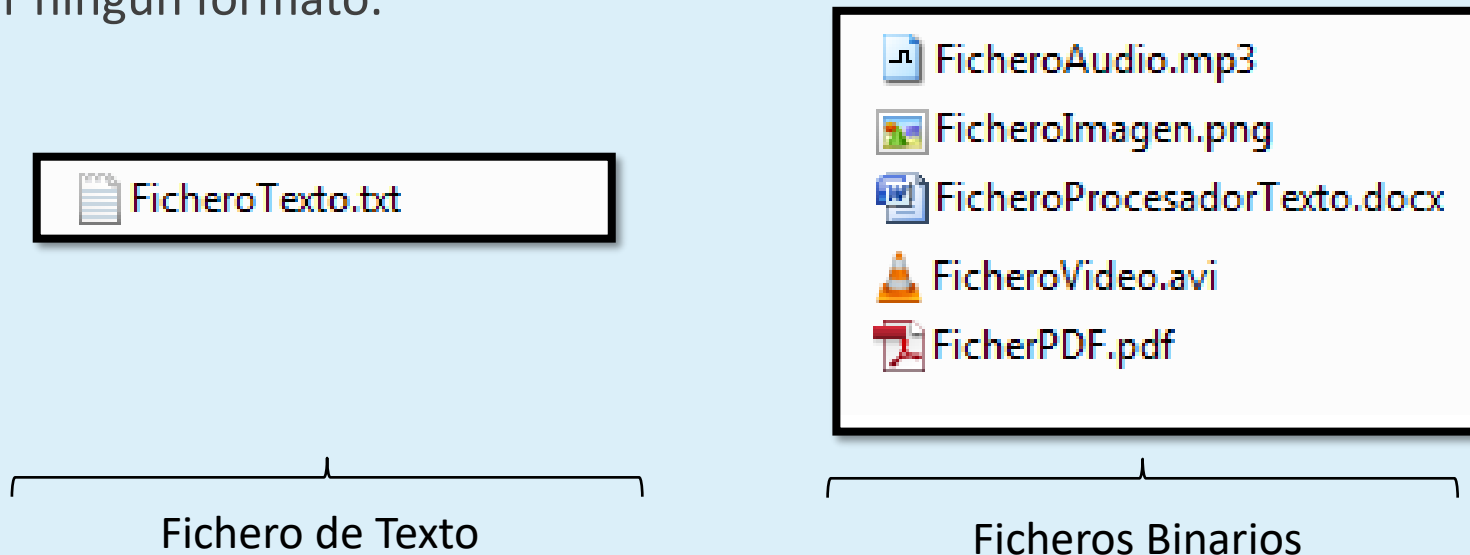
- Por ejemplo:
 - ¿Qué ocurre si creas una montaje fotográfico usando la herramienta de Adobe Photoshop, la guardas en el formato 'ps' (formato propio de dicha aplicación) y se la envías a otra persona para que la vea?
 - ¿Puede verla? ¿Por qué?
 - ¿Qué ocurre ahora si guardas ese montaje en formato 'jpg' y se la envías a la misma persona?
 - ¿Puede verla? ¿Por qué?

1.Introducción

- 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%.

1.Introducción

- 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.



1. Introducción

- 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.

ÍNDICE DE CONTENIDOS

1. Introducción
- 2. Definición de lenguaje de marcas**
3. Tipos de lenguajes de marcas
4. Etiquetas, elementos y atributos
5. Evolución de los lenguajes de marcas
6. Características de los lenguajes de marcas
7. Gramáticas
8. Ejemplos de uso
9. Bibliografía

2. 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.
 - 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.**

2. Definición de lenguaje de marcas

- ¿Qué es un lenguaje de marcas?
 - Lo más interesante de esto es que tú puedes **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). Algunos de estos lenguajes ya existentes son:
 - XHTML (usado para presentar información en la web).
 - MathML (usado para intercambiar información en ámbito matemático).
 - Open eBook (usado en el formato ePub por libros electrónicos).
 - SVG (usado para representar gráficos vectoriales).
 - Open Document (usado para almacenar y presentar información en un paquete ofimático).
 - RSS (usando para difundir información actualizada a usuarios suscritos a una web).

2. Definición de lenguaje de marcas

- ¿Qué es un lenguaje de marcas?
 - Ejemplo de fichero de texto escrito usando MathML:

```
1 <math>
2 <mrow>
3   <mi>x</mi>
4   <mo>=</mo>
5   <mfrac>
6     <mrow>
7       <mrow>
8         <mo>-</mo>
9         <mi>b</mi>
10      </mrow>
11      <mo>&PlusMinus;</mo>
12    <msqrt>
13      <mrow>
14        <msup>
15          <mi>b</mi>
16          <mn>2</mn>
17        </msup>
18      <mo>-</mo>
19    </mrow>
20      <mn>4</mn>
21      <mo>&InvisibleTimes;</mo>
22      <mi>a</mi>
23      <mo>&InvisibleTimes;</mo>
24      <mi>c</mi>
25    </mrow>
26  </mrow>
27 </msqrt>
28 </mrow>
29 <mrow>
30   <mn>2</mn>
31   <mo>&InvisibleTimes;</mo>
32   <mi>a</mi>
33 </mrow>
34 </mfrac>
35 </mrow>
36 </math>
```

2. Definición de lenguaje de marcas

- ¿Qué es un lenguaje de marcas?

- Si abrimos ese fichero de texto usando una aplicación capaz de entender MathML se vería la siguiente fórmula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

- Al ser un fichero de texto plano que se rigen por unas normas definidas a través de un estándar internacional, dicho fichero podrá ser interpretado perfectamente por cualquier aplicación matemática.

2. Definición de lenguaje de marcas

- ¿Qué es un lenguaje de marcas?
 - Ejemplo de fichero texto usando un lenguaje de marcas que hemos creado nosotros mismos para almacenar noticias y resolver así un problema determinado que nos ha surgido en la empresa o en el proyecto fin de ciclo o al trabajar por cuenta propia:

```
1 <noticia>
2   <lugar>Madrid</lugar>
3   <fecha>11/08/2013</fecha>
4   <desc>Se ha inaugurado una estación de trenes.</desc>
5 </noticia>
```

- ¿Por qué nos puede interesar crear un lenguaje de marcas así?

2. Definición de lenguaje de marcas

- ¿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.
 - En segundo lugar, suponer que tenemos un texto con tres noticias:

```
1 - El 11 de Agosto se ha inaugurado una estación de
2   trenes en Madrid.
3 - En Granada las obras del tranvía continúan con el plan
4   previsto a día 29/09/2013.
5 - El tranvía de Jaén continúa sin estar operativo y
6   estamos a 7 de Octubre.
```

2. Definición de lenguaje de marcas

- ¿Qué ventajas tiene un lenguaje de marcas?
 - Ahora tenemos las tres noticias usando un lenguaje de marcas:

```
1 <noticia>
2   <lugar>Madrid</lugar>
3   <fecha>11/08/2013</fecha>
4   <desc>Se ha inaugurado una estación de trenes.</desc>
5 </noticia>
6 <noticia>
7   <lugar>Granada</lugar>
8   <fecha>29/09/2013</fecha>
9   <desc>Las obras del tranvía continúan con el plan previsto.</desc>
10 </noticia>
11 <noticia>
12   <lugar>Jaén</lugar>
13   <fecha>07/10/2013</fecha>
14   <desc>El tranvía continúa sin estar operativo.</desc>
15 </noticia>
```


2. Definición de lenguaje de marcas

- **Muy Importante:** 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.

ÍNDICE DE CONTENIDOS

1. Introducción
2. Definición de lenguaje de marcas
- 3. Tipos de lenguajes de marcas**
4. Etiquetas, elementos y atributos
5. Evolución de los lenguajes de marcas
6. Características de los lenguajes de marcas
7. Gramáticas
8. Ejemplos de uso
9. Bibliografía

3. Tipos de lenguajes de marcas

- Los lenguajes de marcas pueden agruparse en tres grandes grupos:
 - a) Lenguajes orientados a presentación.
 - b) Lenguajes procedurales.
 - c) Lenguajes descriptivos.

3. 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.

- b) Lenguajes procedurales.

- c) Lenguajes descriptivos.

3. Tipos de lenguajes de marcas

- Los lenguajes de marcas pueden agruparse en tres grandes grupos:
 - a) Lenguajes orientados a presentación.

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**.

3. Tipos de lenguajes de marcas

- Los lenguajes de marcas pueden agruparse en tres grandes grupos:

- a) Lenguajes orientados a presentación.

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.¶

3. Tipos de lenguajes de marcas

- Los lenguajes de marcas pueden agruparse en tres grandes grupos:

- a) Lenguajes orientados a presentación.

- 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.

- c) Lenguajes descriptivos.

3. Tipos de lenguajes de marcas

- Los lenguajes de marcas pueden agruparse en tres grandes grupos:

b) Lenguajes procedurales.

```
\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

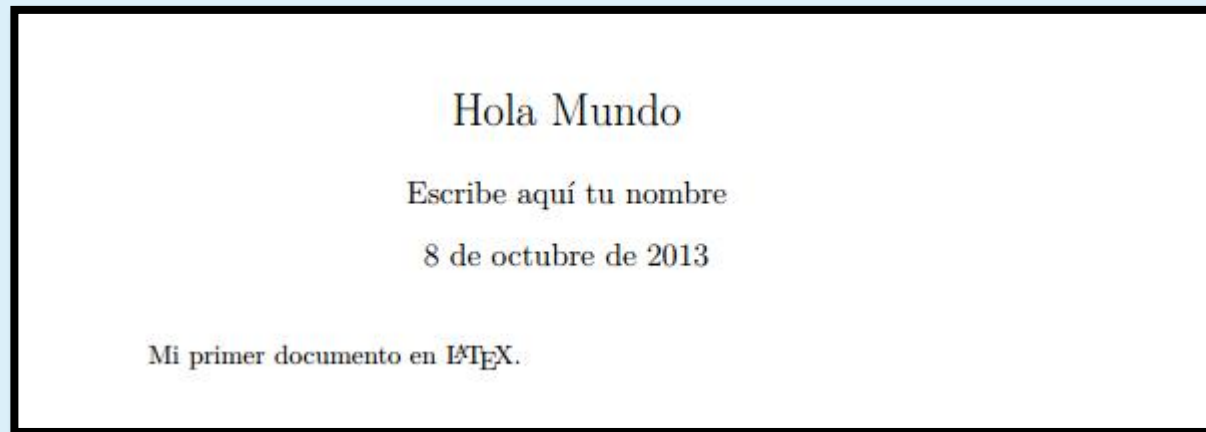
\maketitle

Mi primer documento en \LaTeX{}.

\end{document}
```


3. Tipos de lenguajes de marcas

- Los lenguajes de marcas pueden agruparse en tres grandes grupos:
 - b) Lenguajes procedurales.**



3. Tipos de lenguajes de marcas

- Los lenguajes de marcas pueden agruparse en tres grandes grupos:

- a) Lenguajes orientados a presentación.

- b) Lenguajes procedurales.

- 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.

3. Tipos de lenguajes de marcas

- Los lenguajes de marcas pueden agruparse en tres grandes grupos:

c) Lenguajes descriptivos.

```
<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>
```

ÍNDICE DE CONTENIDOS

1. Introducción
2. Definición de lenguaje de marcas
3. Tipos de lenguajes de marcas
- 4. Etiquetas, elementos y atributos**
5. Evolución de los lenguajes de marcas
6. Características de los lenguajes de marcas
7. Gramáticas
8. Ejemplos de uso
9. Bibliografía

4. 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.

4. Etiquetas, elementos y atributos

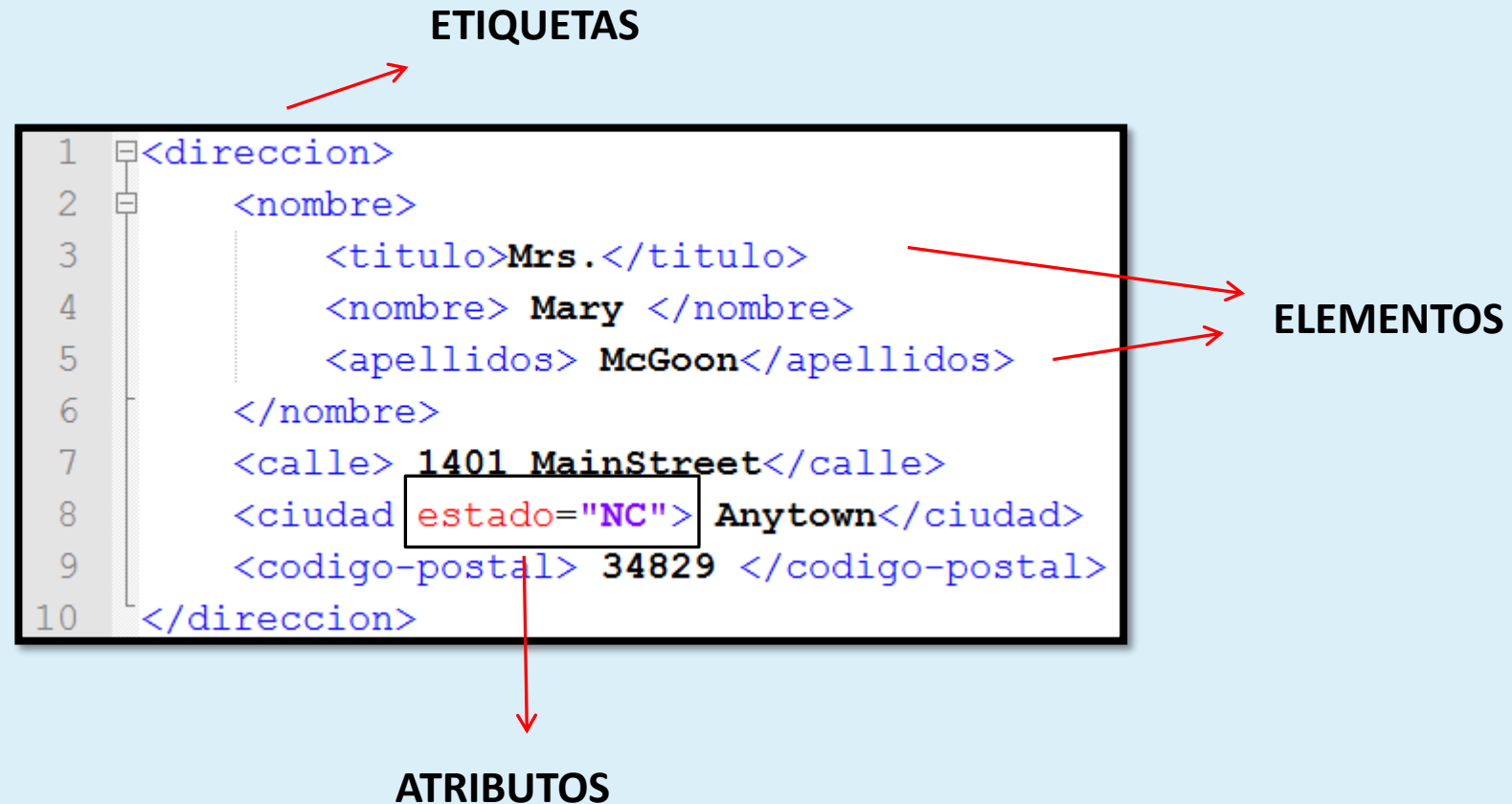
ETIQUETAS

```
1 <direccion>
2   <nombre>
3     <titulo>Mrs.</titulo>
4     <nombre> Mary </nombre>
5     <apellidos> McGoon</apellidos>
6   </nombre>
7   <calle> 1401 MainStreet</calle>
8   <ciudad estado="NC"> Anytown</ciudad>
9   <codigo-postal> 34829 </codigo-postal>
10 </direccion>
```

ELEMENTOS

ATRIBUTOS

4. Etiquetas, elementos y atributos



ÍNDICE DE CONTENIDOS

1. Introducción
2. Definición de lenguaje de marcas
3. Tipos de lenguajes de marcas
4. Etiquetas, elementos y atributos
- 5. Evolución de los lenguajes de marcas**
6. Características de los lenguajes de marcas
7. Gramáticas
8. Ejemplos de uso
9. Bibliografía

5. 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 usaba 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.

5. Evolución de los lenguajes de marcas

■ 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.

5. Evolución de los lenguajes de marcas

■ AÑOS 90:

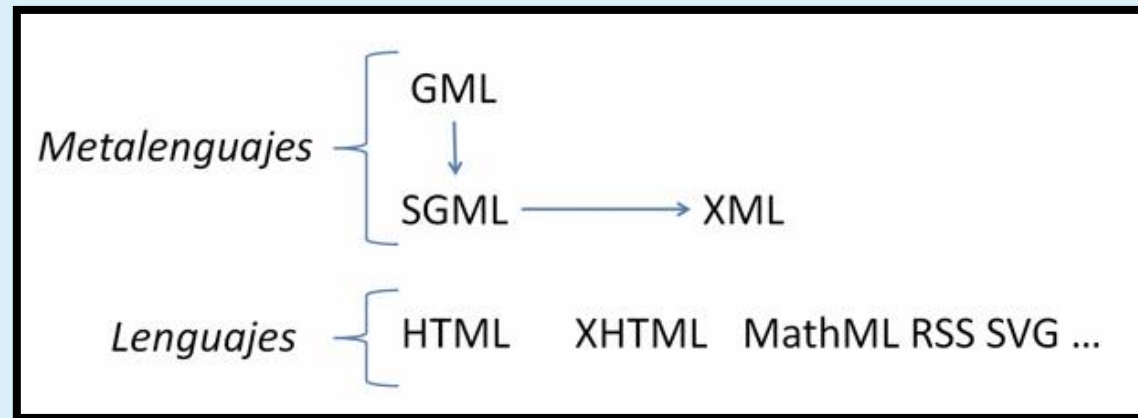
- Finalmente, en el 1998, W3C hizo público un nuevo estándar que denominaron XML (eXtended Markup Language), más sencillo que SGML y más potente que HTML.
- Pero... ¿qué es XML en realidad?

<https://www.youtube.com/watch?v=bCd2xaQrTAo>

Vídeo de Sergio Luján Mora, profesor de la Universidad de Alicante.

5. Evolución de los lenguajes de marcas

- AÑOS 90:
 - Pero... ¿qué es XML en realidad?



Captura del vídeo de Sergio Luján Mora: "XML – Introducción".

ÍNDICE DE CONTENIDOS

1. Introducción
2. Definición de lenguaje de marcas
3. Tipos de lenguajes de marcas
4. Etiquetas, elementos y atributos
5. Evolución de los lenguajes de marcas
- 6. Características de los lenguajes de marcas**
7. Gramáticas
8. Ejemplos de uso
9. Bibliografía

6. 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.

6. Características de los lenguajes de marcas

- **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.

6. Características de los lenguajes de marcas

■ 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 una página web) y de XHTML+MathML+SVG (permite integrar fórmulas matemáticas y gráficos vectoriales en una página web).

ÍNDICE DE CONTENIDOS

1. Introducción
2. Definición de lenguaje de marcas
3. Tipos de lenguajes de marcas
4. Etiquetas, elementos y atributos
5. Evolución de los lenguajes de marcas
6. Características de los lenguajes de marcas
- 7. Gramáticas**
8. Ejemplos de uso
9. Bibliografía

7. 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.

7. Gramáticas

- **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.

7. Gramáticas

- **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.

ÍNDICE DE CONTENIDOS

1. Introducción
2. Definición de lenguaje de marcas
3. Tipos de lenguajes de marcas
4. Etiquetas, elementos y atributos
5. Evolución de los lenguajes de marcas
6. Características de los lenguajes de marcas
7. Gramáticas
- 8. Ejemplos de uso**
9. Bibliografía

8. Ejemplos de uso

XML: Ejemplos de uso

Parte 1: <https://www.youtube.com/watch?v=KehhDEfxnhI>

Parte 2: https://www.youtube.com/watch?v=l8jHhbhdOl8&list=PLKLLGxMgKZ_QrK6eY106uNWOYmzmA61_B

Vídeo de Sergio Luján Mora, profesor de la Universidad de Alicante.

ÍNDICE DE CONTENIDOS

1. Introducción
2. Definición de lenguaje de marcas
3. Tipos de lenguajes de marcas
4. Etiquetas, elementos y atributos
5. Evolución de los lenguajes de marcas
6. Características de los lenguajes de marcas
7. Gramáticas
8. Ejemplos de uso
- 9. Bibliografía**

9. Bibliografía

- **Lenguajes de Marcas y Sistemas de Gestión de Información**

Sánchez, F. J. y otros. Editorial: RA-MA.

- **Varios tutoriales - w3schools.com**

<http://www.w3schools.com/>

- **Varios artículos - Wikipedia, la enciclopedia libre**

<http://es.wikipedia.org>

- **World Wide Web Consortium - w3c.org**

<http://www.w3.org>