



# Introducción a los Lenguajes de MarCas

FUNDAMENTOS DE DESARROLLO WEB

Curso académico

2023 – 2024

# Introducción LM: Índice

---

1. Definición
2. Evolución
3. Clasificación y uso
  - Lenguajes orientados a la presentación
  - Lenguajes procedurales
  - Lenguajes descriptivos
4. Conceptos
  - Etiquetas
  - Elementos
  - Atributos
5. Gramáticas
6. Estructura

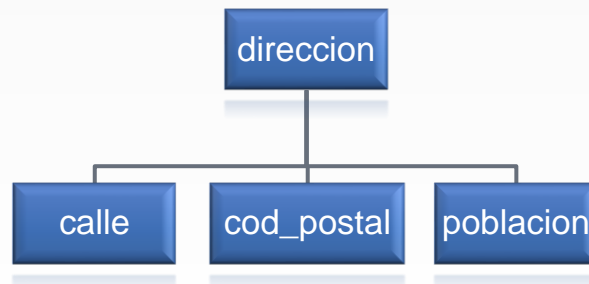


## INTRODUCCIÓN A LOS LM

# Lenguajes de Marcas

# Introducción LM: Definición

- Los lenguajes de marcas son aquellos que combinan la información que contiene un documento con anotaciones relativas a:
  - La estructura del texto.
  - La forma de presentarlo.
  - La forma de procesarlo.
- A las anotaciones se les llama **marcas** o **etiquetas** (*tags*), y siguen una estructura jerárquica en forma de árbol.
- Ejemplo



```
<direccion>
  <calle> Calle Princesa </calle>
  <codigo_postal> 28290 </codigo_postal>
  <poblacion> Madrid </poblacion>
</direccion>
```

# Introducción LM: Definición

---

```
<direccion>  
  <calle> Calle Princesa </calle>  
  <codigo_postal> 28290 </codigo_postal>  
  <poblacion> Madrid </poblacion>  
</direccion>
```

Etiquetas o tags

# Introducción LM: Evolución

---

## ■ Origen

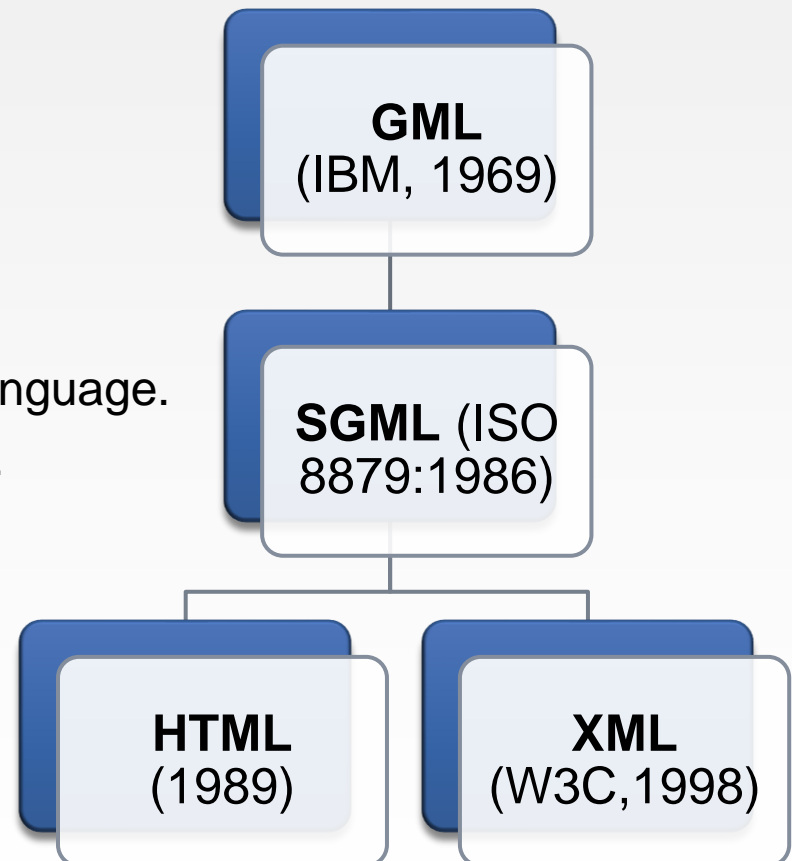
- Falta de estandarización en los formatos de información usados por distintos programas.
- Surge en IBM por la necesidad para almacenar grandes cantidades de información.
  - Se crea un sistema de edición, almacenamiento y búsqueda de documentos para establecer un formato estándar para todos los documentos de la empresa, esto es, un formato flexible que se ajuste a cualquier tipo de documento.

## ■ Objetivo

- Describir documentos con independencia de la plataforma o aplicación utilizada.

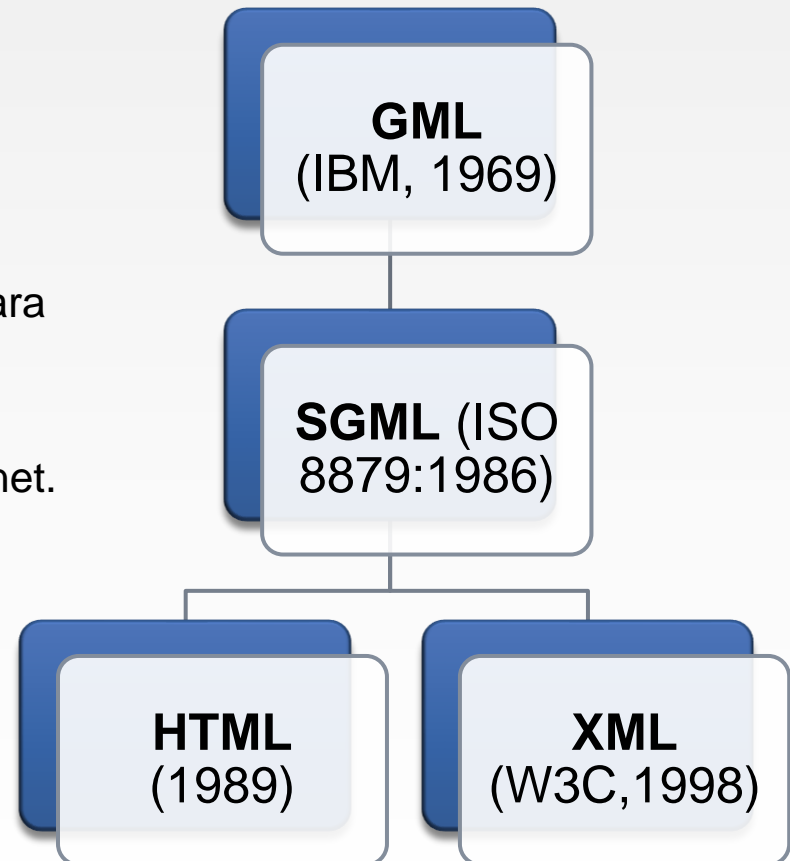
## Introducción LM: Evolución

- Comienzan a usarse a finales de los 60 con el objetivo de separar presentación y estructura.
- GML – Generalized Markup Language.
- SGML – Standard Generalized Markup Language.
  - Extenso y completo, pero muy complicado.



# Introducción LM: Evolución

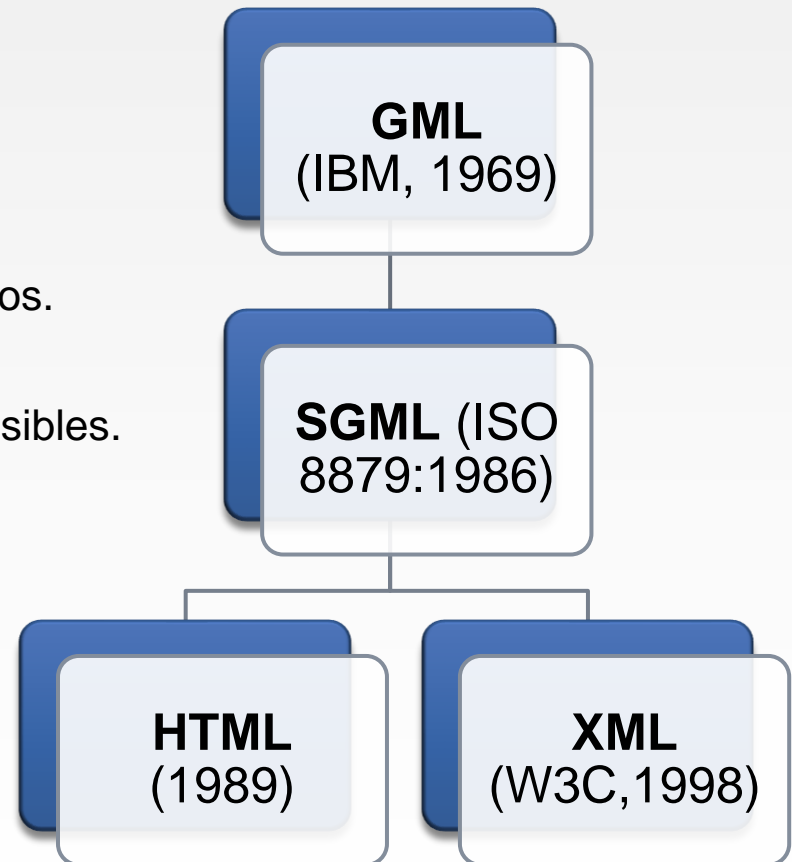
- A finales de los 80 se crea HTML.
  - HyperText Markup Language.
  - Tim Berners-Lee (1.989).
  - ISO 15445:2000
    - HTML se convierte en un estándar general para la creación de páginas webs.
    - Compartir información a través de internet.
    - Papel fundamental en el crecimiento de internet.
    - Sintaxis sencilla.
- No es flexible.
  - Las etiquetas son limitadas.





## Introducción LM: Evolución

- A mediados de los 90 se elabora XML.
  - eXtensible Markup Language.
  - Más sencillo que SGML.
  - Más potente que HTML (semántica).
    - Las etiquetas indican el significado de los datos.
  - Flexible, extensible.
    - No limita el número de marcas o etiquetas posibles.
  - Muchos usos: lenguajes derivados.
- W3C – World Wide Web.
  - Consorcio internacional creado por Tim Berners-Lee, que crea recomendaciones para la www.



# Introducción LM: Evolución

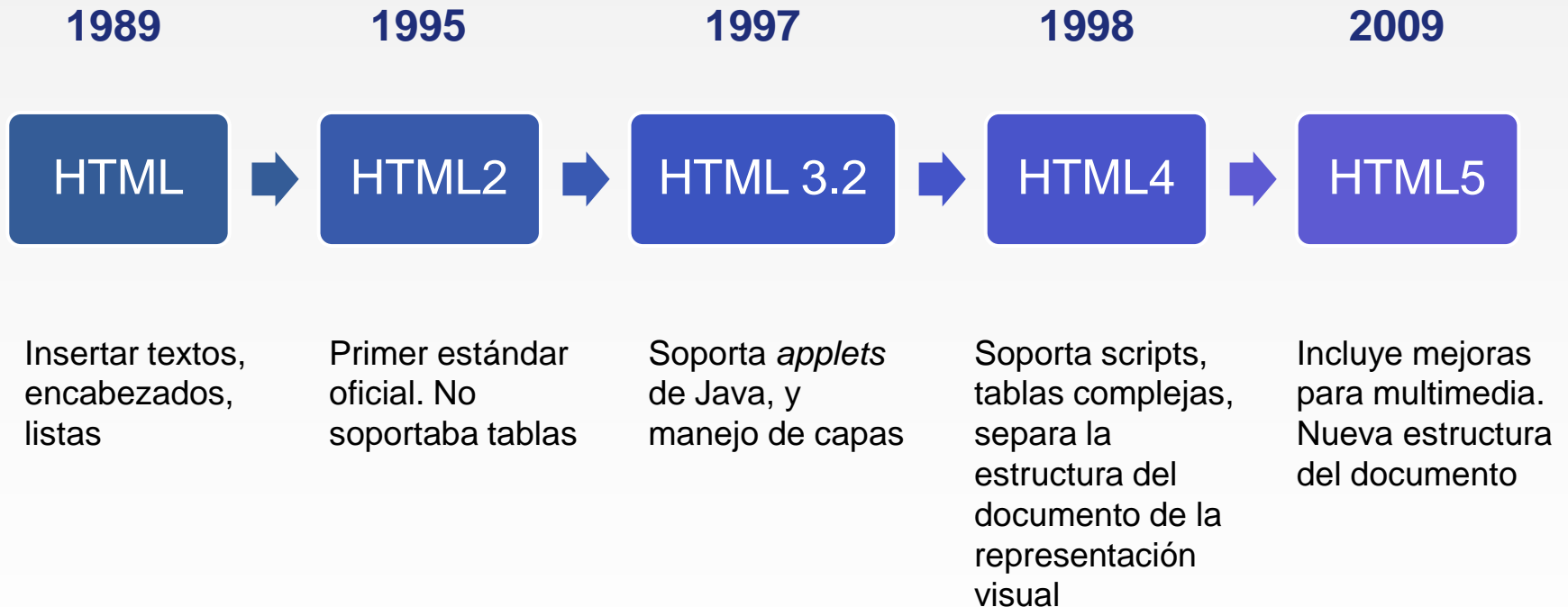
---

- HTML
- HTTP
- URL



# Introducción LM: Evolución HTML

---



# Introducción LM: HTML vs XML

---

## ■ ¿HTML o XML?

```
<p>Miguel de Cervantes</p>
<p>Don Quijote de la Mancha</p>
<p>Precio: 30€ </p>
```

```
<libro>
  <autor>Miguel de Cervantes</autor>
  <titulo>Don Quijote de la Mancha</titulo>
  <precio moneda="euros">30</precio>
</libro>
```

# Introducción LM: HTML vs XML

---

## ■ HTML

```
<p>Miguel de Cervantes</p>  
<p>Don Quijote de la Mancha</p>  
<p>Precio: 30€ </p>
```

- Las etiquetas presentan la información.

## ■ XML

```
<libro>  
  <autor>Miguel de Cervantes</autor>  
  <titulo>Don Quijote de la Mancha</titulo>  
  <precio moneda="euros">30</precio>  
</libro>
```

- Las etiquetas definen el contenido y significado de la información.

## INTRODUCCIÓN A LOS LM

# Clasificación

- \* De presentación
- \* De procedimiento
- \* Descriptivo o semántico

# Introducción LM: Clasificación

---

- Lenguajes orientados a la presentación.
  - El marcado de presentación define el formato del texto (negrita, interlineado,...), esto es, cómo se presenta el documento.
  - Formato del texto Marcas o etiquetas se pueden ocultar al usuario.
    - WYSIWYG (What You See Is What You Get).
  - Insuficiente para el procesamiento automático de información.
  - No son flexibles ni reusables.
  - Ej.: procesadores de texto (MSWord), DreamWeaver.

# Introducción LM: Clasificación

---

- Lenguajes procedurales o marcado de procedimiento.
  - También orientado a presentación.
  - Funciona mediante secuencia de acciones.
  - Etiquetas visibles para el usuario que edita el texto.
  - Permiten procesamiento según el tipo de etiqueta.
  - Pueden ser muy complejos.
  - Ej.: LaTeX, Postscript, HTML.



# Introducción LM: Clasificación

---

- Lenguajes descriptivos.
  - También llamados de marcado semántico.
    - Significado de las expresiones lingüísticas.
  - Las marcas indican qué es la información, detallan lo que se está representando.
  - No especifican cómo representarla.
  - Son flexibles y reusables.
  - Ej.: XML y derivados.
- En general, los lenguajes tienen características de más de una clase.

# Introducción LM: Uso

---

- **Objetivo general**
  - Describir documentos con independencia de la plataforma o aplicación utilizada.
  - Procesamiento automático de la información representada.
- **Cualquier tipo de documento**
  - Textos
  - Presentaciones
  - Gráficos
  - Música
  - Matemáticas
  - Entornos Web
  - Almacenamiento de información

## INTRODUCCIÓN A LOS LM

# Conceptos

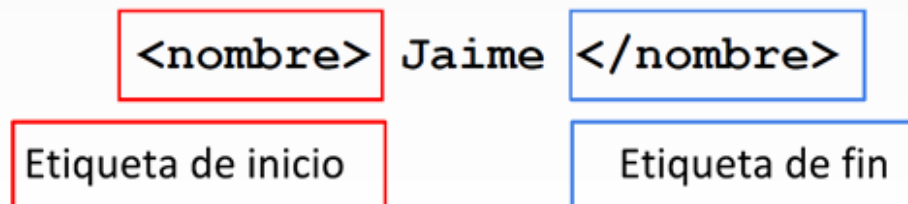
- \* Etiquetas
- \* Elementos
- \* Atributos

# Introducción LM: Conceptos

---

## ■ Etiquetas

- Se utilizan para organizar un archivo de texto dentro de una estructura.
- Etiqueta de inicio
  - Compuesta por el símbolo menor '<', un texto y el símbolo mayor '>'.
- Etiqueta de fin
  - Contiene el símbolo '</', un texto y '>'.
  - El texto debe ser igual que el de la etiqueta de inicio.
- Ej. <nombre> Jaime </nombre>



# Introducción LM: Conceptos

---

- Ejemplos incorrectos

`<nombre> Pablo </Nombre>`

`<nombre> Pablo </finNombre>`

`</apellido> Pérez <apellido>`

`<!nota> 10 </!nota>`



# Introducción LM: Conceptos

---

## ■ Elementos

- Constan de etiqueta de inicio, un contenido y etiqueta de fin.
- Puede haber elementos vacíos.
  - No tienen contenido ni etiqueta de fin.
- Los elementos pueden ser:
  - Simples: el contenido está formado por texto o números.
  - Complejos: formados por más elementos 'hijos'.

Elemento

`<nombre> Jaime </nombre>`

# Introducción LM: Conceptos

---

## ■ Elementos

Elemento vacío

```
<libro></libro>
```

```
<libro/>
```

```
<br/>
```

# Introducción LM: Conceptos

---

## ■ Atributos

- Dupla nombre-valor.
- Indican propiedades de los elementos.
- Se encuentran dentro de la etiqueta de inicio.
- El valor suele ir entre comillas.
- Valor obligatorio en muchos lenguajes.
- Las comillas pueden ser simples (‘ ’) o dobles (“ ”).



# Introducción LM: Conceptos

---

## ■ Ejemplo

Atributo

```
<evento fecha="20/09/2015" >  
    España gana la final del  
    Eurobasket de baloncesto a  
    Lituania  
</evento>
```

# Introducción LM: Conceptos

---

- Ejemplo
  - Etiquetas, Elementos y Atributos.

```
<noticia>
  <lugar> Lille </lugar>
  <evento fecha ="20/09/2015">
    España gana la final del Eurobasket de baloncesto
    a Lituania
  </evento>
</noticia>
```

# Introducción LM: Conceptos

---

## ■ Ejemplo

`<noticia>`

`<lugar>` Lille `</lugar>`

`<evento fecha ="20/09/2015">`

España gana la final del Eurobasket de baloncesto  
a Lituania

`</evento>`

`</noticia>`

Etiquetas

# Introducción LM: Conceptos

## ■ Ejemplo

```
<noticia>
```

```
<lugar> Lille </lugar>
```

```
<evento fecha ="20/09/2015">
```

```
    España gana la final del Eurobasket de baloncesto  
    a Lituania
```

```
</evento>
```

```
</noticia>
```

ELEMENTOS (etiqueta inicio+contenido+etiqueta fin)

# Introducción LM: Conceptos

---

## ■ Ejemplo

```
<noticia>
  <lugar> Lille </lugar>
  <evento fecha ="20/09/2015" >
    España gana la final del Eurobasket de baloncesto
    a Lituania
  </evento>
</noticia>
```

Atributos

# Introducción LM: Conceptos

---

## ■ Elementos vs. Atributos

- Los elementos se utilizan para encapsular porciones de datos.
- Los atributos se utilizan para proporcionar información adicional (propiedades).
- El uso de uno u otro depende de las necesidades.
  - No hay una regla fija sólo 'buenas prácticas'.
  - Depende del lenguaje de marcas utilizado.

# Introducción LM: Conceptos

---

## ■ Etiquetas: Recomendaciones XML

- Utilizar nombres descriptivos.
- Utilizar nombres cortos.
- Evitar acentos y caracteres extraños.
- Evitar caracteres de puntuación.

```
<nombreAlumno>  
    mejor que  
<el_Nombre_del_alumno>
```

## INTRODUCCIÓN A LOS LM

# Gramáticas



# Introducción LM: Gramáticas

---

- Conjunto de reglas de un lenguaje
  - Las gramáticas definen el marcado permitido, el marcado requerido, y cómo debe ser utilizado en la instancia del documento.
    - Especifican cuáles serán las etiquetas posibles, las obligatorias.
    - Dónde deben colocarse las etiquetas y el significado que tiene cada una de ellas.
- Documentos bien formados
  - Aquellos que son sintácticamente correctos según las siguientes reglas:
    - Etiquetas de inicio y fin han de ser iguales.
    - Elementos correctamente cerrados y anidados.
    - Atributos en la etiqueta de inicio.
    - Cuidado con las mayúsculas y los caracteres especiales.
    - En muchos casos (XML y HTML) deben tener un único elemento raíz.

# Introducción LM: Gramáticas

---

- Documentos válidos

- Aquellos que, además de estar bien formados, cumplen los requisitos de una definición de estructura (gramática).
- Existen distintas formas de definir la gramática de un lenguaje:
  - DTD (Definición de Tipo de Documento)
    - Formalismo para especificar un lenguaje.
    - Descripción de la estructura y sintaxis, proporcionando un formato común.
    - Es un fichero aparte vinculado.
  - Esquemas
    - Similar a DTD.
    - Mayor especificación del tipo de datos.

# Introducción LM: Gramáticas

---

## ■ Aclaración

- La tabulación y el nombre de las etiquetas (XML) no influye en que el documento esté bien formado ni sea válido.
- Simplemente facilitan la lectura del documento por parte de un humano.

```
<a>  
<b>Paco</b>  
<c>Defensa</c>  
</a>
```

```
<jugador>  
  <nombre>Paco</nombre>  
  <posicion>Defensa</posicion>  
</jugador>
```

## INTRODUCCIÓN A LOS LM

# Estructura

# Introducción LM: Estructura

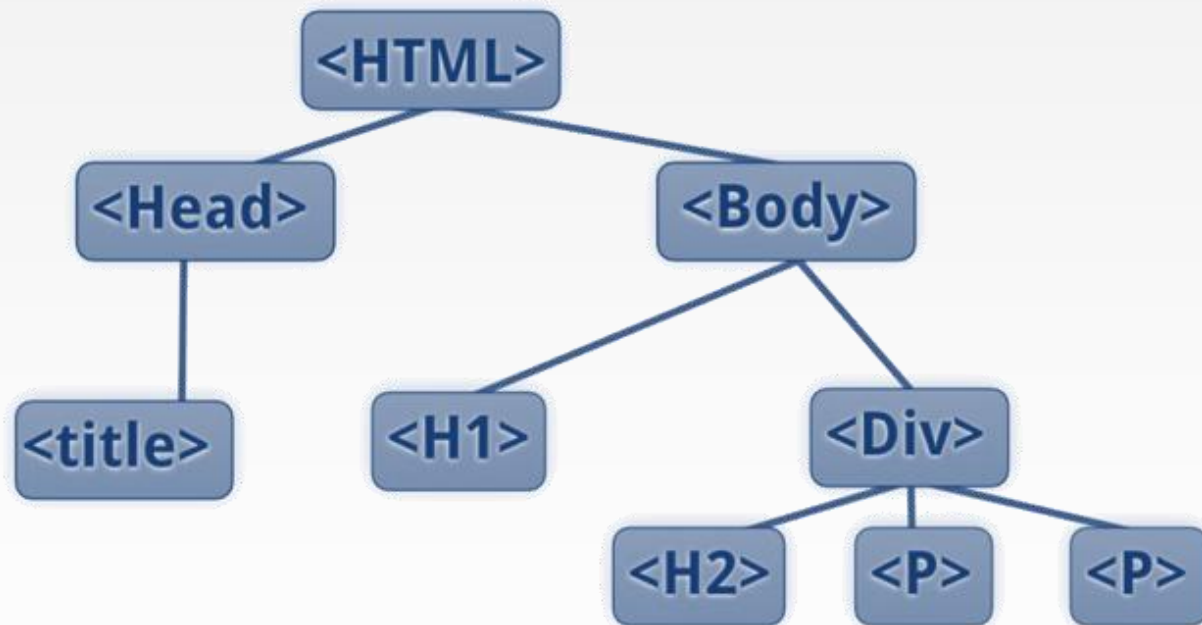
---

- Muchos documentos de marcas, como XML o HTML, tienen estructura en forma de ‘árbol’.
- En primer lugar aparece una o varias instrucciones de procesamiento.
  - Dependen del lenguaje concreto.
- A continuación el elemento raíz que es el ‘ancestro’ del resto de elementos.
- Los elementos pueden tener subelementos.
  - Elementos hijos.

# Introducción LM: Estructura

---

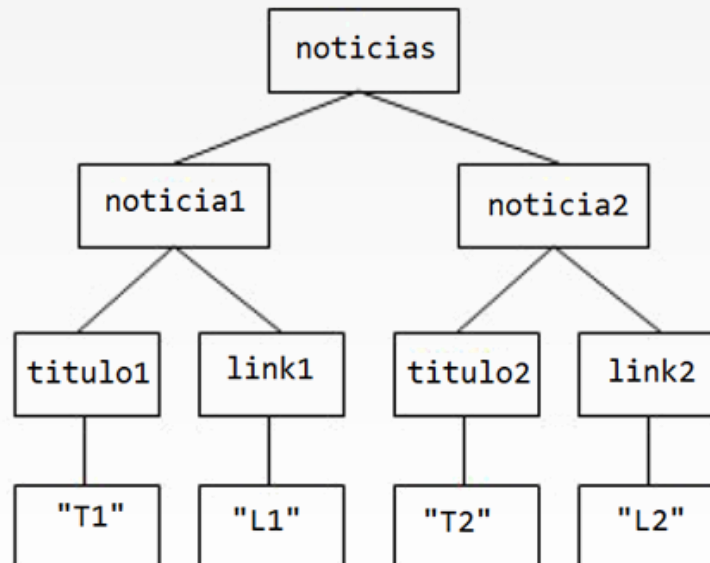
- Ejemplo



# Introducción LM: Estructura

## ■ Ejercicio

- Escribir el lenguaje de marcas correspondiente al siguiente árbol.



# Introducción LM: Estructura

---

## ■ Ejercicio

### ■ Solución

```
<noticias>
  <noticial>
    <titulo1> T1 </titulo1>
    <link1> L1 </link1>
  </noticial>

  <noticia2>
    <titulo2> T2 </titulo2>
    <link2> L2 </link2>
  </noticia2>
</noticias>
```



# Grado en Ingeniería del Software

Asignatura: Fundamentos de Desarrollo Web

Curso: 2023 – 2024

ramona.ruiz@u-tad.com