

Documentos XML

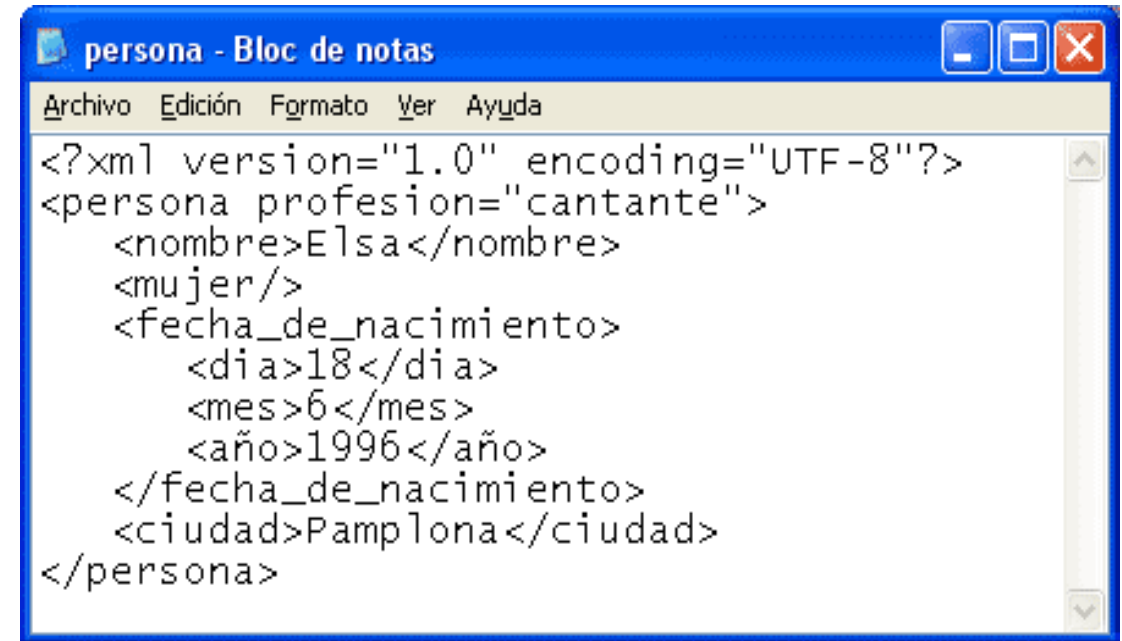
LENGUAJES DE MARCAS Y SISTEMAS DE GESTIÓN DE INFORMACIÓN

¿Qué es XML?

- **XML** (*eXtensible Markup Language*) no es lenguaje. Es un meta-lenguaje que nos permite definir lenguajes de marcas.

[Recuerda: anteriormente vimos HTML, que es un lenguaje de marcas; y Javascript, que es un lenguaje de programación]

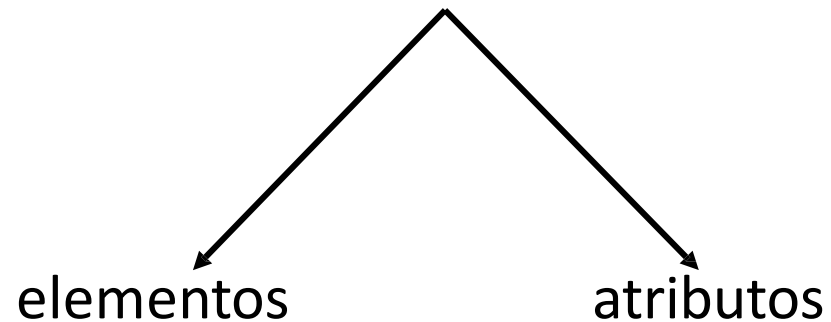
- No se usa sólo en el desarrollo web, sino que se utiliza como un estándar para intercambio de información estructurada entre diferentes plataformas. Por ejemplo, podría servirnos para pasar datos de una aplicación Linux con BD Oracle a una Windows con BD MySQL. También se puede usar en hojas de cálculo, y en muchas más plataformas.



```
<?xml version="1.0" encoding="UTF-8"?>
<persona profesion="cantante">
  <nombre>Elsa</nombre>
  <mujer/>
  <fecha_de_nacimiento>
    <dia>18</dia>
    <mes>6</mes>
    <año>1996</año>
  </fecha_de_nacimiento>
  <ciudad>Pamplona</ciudad>
</persona>
```

¿Qué es un documento XML?

- Un **documento XML** es un fichero de texto en el que la información viene etiquetada (etiquetas + datos).
- Debe cumplir unas normas para que funcione bien → Documentos bien formados
- Dos conceptos claves en documentos XML



```
<?xml version="1.0" encoding="UTF-8"?>
<persona profesion="cantante">
  <nombre>Elsa</nombre>
  <mujer/>
  <fecha_de_nacimiento>
    <dia>18</dia>
    <mes>6</mes>
    <año>1996</año>
  </fecha_de_nacimiento>
  <ciudad>Pamplona</ciudad>
</persona>
```

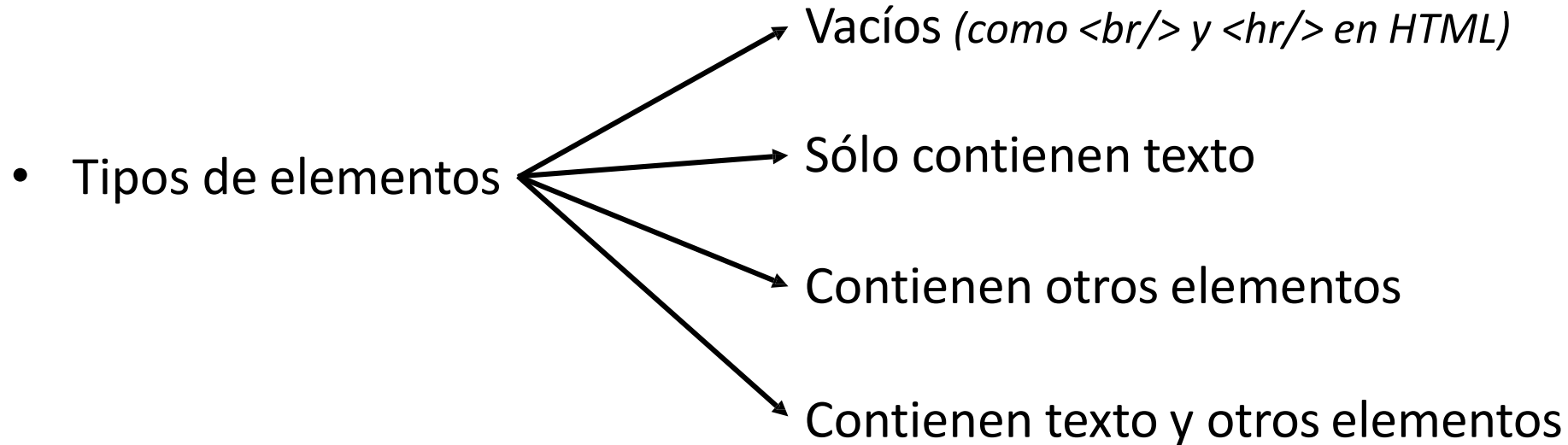
A screenshot of a Windows Notepad window titled "persona - Bloc de notas". The window has a menu bar with "Archivo", "Edición", "Formato", "Ver", and "Ayuda". The text area contains the following XML code:

Elementos

- Un **elemento** es el componente semántico básico de un documento. En el ejemplo anterior, por ejemplo: *nombre* es un elemento. También *persona*, *mes*, *ciudad*...
 - Cada elemento está formado por:
 - Etiqueta de inicio → **<NombreElemento>**
 - **Contenido**
 - Etiqueta final → **</NombreElemento>**
- <NombreElemento>Contenido</NombreElemento>**
<peso>5,73</peso>
- Un documento XML tendrá una jerarquía de elementos a partir de un elemento raíz único (en el ejemplo anterior, todos parten del elemento *persona*).

Elementos

- El nombre de un elemento, para ser válido, debe empezar por letra o _ y luego puede tener letras, dígitos, ., _ o -. No puede empezar por XML o xml.



Atributos

- Un **atributo** es una información complementaria asociada a un elemento. En el ejemplo anterior, por ejemplo, *profesion* es un atributo.
- Cada elemento puede (no es obligatorio) tener una lista de atributos asociada.
- El orden de los atributos no importa. No pueden estar repetidos.

```
<NombreElemento atributo1="valor1" atributo2="valor2">Contenido</NombreElemento>  
<peso unidad="gramos">5,73</peso>
```

Documentos XML → Distintas soluciones

```
<persona>Miguel Navarro Fernández</persona>
```

```
<persona>
```

```
  <nombre>Miguel</nombre>
```

```
  <apellidos>Navarro Fernández</apellidos>
```

```
</persona>
```

```
<persona>
```

```
  <nombre>Miguel</nombre>
```

```
  <primer_apellido>Navarro</primer_apellido>
```

```
  <segundo_apellido>Fernández</segundo_apellido>
```

```
</persona>
```



¿Cuál de estas formas es correcta?

TODAS PUEDEN SERLO

**NO HAY UNA
SOLUCIÓN ÚNICA**

Documentos XML → Distintas soluciones

<persona>Miguel Navarro Fernández</persona>

<persona>
 <nombre>Miguel</nombre>
 <apellidos>Navarro Fernández</apellidos>
</persona>

<persona>
 <nombre>Miguel</nombre>
 <primer_apellido>Navarro</primer_apellido>
 <segundo_apellido>Fernández</segundo_apellido>
</persona>

-contenido semántico
-complejidad

¿Necesidades futuras?

+contenido semántico
+complejidad

¿Atributos o elementos?

```
<diapositiva>  
  <titulo>Introducción a XML</titulo>  
  <texto>XML es un meta-lenguaje...</texto>  
</diapositiva>
```

```
<diapositiva titulo="Introducción a XML">  
  <texto>XML es un meta-lenguaje...</texto>  
</diapositiva>
```

¿Cuándo usar atributos?



DUDAS

¿Atributos o elementos? → Pautas

Cada caso es un mundo. Por lo tanto, no hay reglas universales y válidas para todos los casos. Pero algunas pautas pueden ayudarnos a decidir si algo es un **elemento** o un **atributo**:

- Si tiene estructura interna o mucha información, debe ser un **elemento**.
- También se puede hacer la comparación:

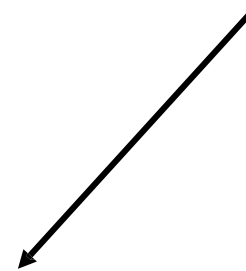
sustantivo	→ elemento	} (no es 100% fiable)
adjetivo	→ atributo	
- Si se tienen dudas, mejor utilizar un **elemento**.

```
<diapositiva tiempo="medio">  
  <titulo>Introducción a XML</titulo>  
  <texto>XML es un meta-lenguaje...</texto>  
</diapositiva>
```

En el caso planteado antes, quizás sería más conveniente que **titulo** fuera un elemento.
Un ejemplo de atributo sería **tiempo**, que puede tomar los valores "poco", "medio" y "mucho".

Primera línea del documento XML

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
```



Indica si el documento es autocontenido. O sea, si funciona por sí sólo, sin depender de otro documento externo.

Cuando el documento utilice etiquetas que están definidas en otro archivo (ya sea DTD o esquema... esto lo veremos) habrá que poner **standalone="no"**

Comentarios en XML

`<!-- Esto es un comentario para organizarme el código -->`

(Misma sintaxis que en HTML)

Caracteres especiales

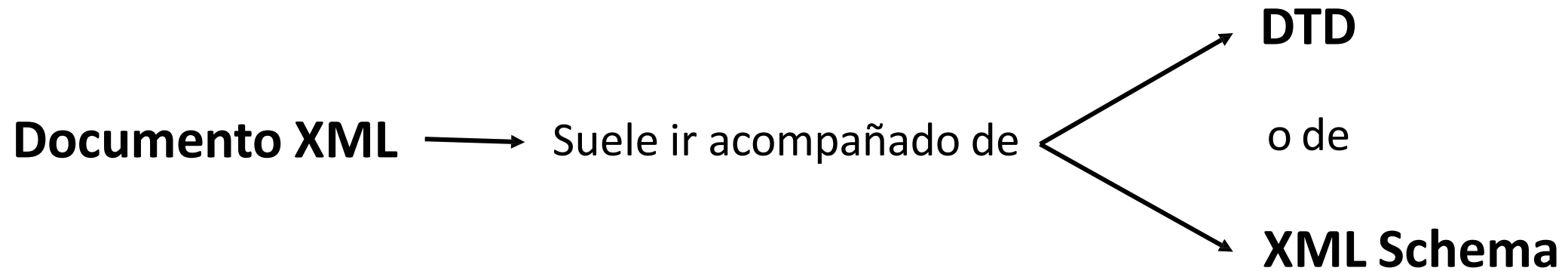
< > & ' "

Si queremos que aparezca alguno de estos caracteres, tenemos varias opciones:

- Incluirlos entre `<![CDATA[y]]>` (lo que vaya dentro no se interpreta como etiquetas)
- Utilizar entidades predefinidas

Entidad	Carácter	Entidad	Carácter
<	<	>	>
&	&	'	'
"	"	à	à

Más adelante, veremos...



Próximamente veremos como, en ese documento que acompaña al XML (DTD o Schema), se definirán: qué elementos que pueden usarse (y por tanto, qué etiquetas habrá), qué atributos pueden tener, qué puede tener dentro cada elemento, etc.

De momento... ¿Documento bien formado?

Por lo tanto, sin la presencia de un DTD (o Schema), no podemos comprobar la validez de un documento (cualquier etiqueta es válida). Pero sí podemos comprobar si está bien formado.

Reglas para ser un DOCUMENTO BIEN FORMADO

- ✓ Los elementos deben estar anidados adecuadamente.
- ✓ Los valores de los atributos deben ponerse entre comillas simples o dobles.
- ✓ Todo elemento debe tener una etiqueta de final o utilizar la etiqueta de elemento vacío.
- ✓ El documento debe tener un único elemento raíz.
- ✓ Todo texto debe estar incluido en un elemento.