

Lenguajes de Marcas y Sistemas de Gestión de la Información

## UNIDAD 2

XML

# Lenguajes de Marcas y Sistemas de Gestión de la Información

## Capítulo 2: XML

- Conceptos básicos
- Prologo
- Atributos
- Ejercicio
- Entidades y referencia a entidades
- Instrucciones de procesamiento
- Comentarios
- Documentos bien formados

# XML : Conceptos básicos

- **Caracteres y codificación:**

Los documentos XML pueden estar codificados en distintos juegos de caracteres (ISO-8859-1, UTF-8, etc)

- **Procesador XML (XML parser):**

La aplicación recurre a un procesador XML para leer un documento XML. Este procesador es el que lee, analiza y le pasa la información en un formato estructurado a la aplicación.

- **Etiquetas (tags) o Marcas (mark-up):**

Las etiquetas son marcas que sirven para identificar un contenido concreto del resto del contenido del documento.

Existen tres tipos de etiquetas:

- **Etiquetas de apertura (start-tag).**

`<apartado>`

- **Etiquetas de cierre (end-tag),** que empiezan por `"/"`.

`</apartado>`

- **Etiquetas vacías (empty-tag),** que terminan por `"/"`.

`<salto-de-linea />`

# XML : Conceptos básicos

- Los documentos XML pueden estar formados por dos partes:
  - **Prólogo** (opcional)
    - Declaración XML
    - Declaración del tipo de documento
  - **Ejemplar** (Contenido del documento)

# XML : Conceptos básicos - Prologo

El prólogo debe preceder al ejemplar del documento. En la 1ª línea del doc, y primer carácter.

Facilita el procesado de la información del ejemplar. Está dividido en dos partes:

- La declaración XML:

- Atrib. **version** `<?xml version="1.0" ?>`
- Atrib. **encoding** `<?xml version="1.0" encoding="iso-8859-1" ?>`
- Atrib. **standalone** `<?xml version="1.0" encoding="iso-8859-1" standalone="yes" ?>`

La declaración XML es opcional. Pero, si se escribe, el atributo **version** es obligatorio.

Los atributos **encoding** y **standalone** son opcionales y sus valores por defecto son "UTF-8" y "no".

Cuando se escriba el atributo **encoding**, siempre después de **version**.

El atributo standalone, siempre que se escriba, deberá ir en último lugar, después de **encoding**

- La declaración del tipo de documento: `<!DOCTYPE nombre_tipo...>`

`<!DOCTYPE persona>`

**standalone:** Especifica si la validez del documento depende de otro documento externo (DTD)

# XML : Conceptos básicos – Ejemplar y elementos

Contiene los datos reales del documento XML, **el contenido**. Está formado por un elemento raíz que a su vez contendrá otros elementos anidados.

- **Elemento**: componente lógico de un documento XML.
  - Comienza por una etiqueta de apertura y termina por la etiqueta de cierre correspondiente.
  - O bien consiste en una única etiqueta vacía.
  - El contenido de un elemento es todo lo que se encuentra entre las etiquetas de apertura y cierre
  - También existen elementos vacíos, los cuales no tienen contenido.

- Elemento con texto como contenido:  
`<autor>TicArte</autor>`
- Elemento con otros elementos como contenido:  
`<web>  
    <autor>TicArte</autor>  
</web>`
- Elemento vacío o sin contenido:  
`<ingreso />` o también puede escribirse así: `<ingreso></ingreso>`

# XML : Conceptos básicos – Ejemplar y elementos

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<!DOCTYPE persona>
<persona>
  <nombre>Carlos</nombre>
  <apellido>Carnero</apellido>
  <nacimiento>
    <anyo>1956</anyo>
    <mes>Abril</mes>
    <dia>13</dia>
  </nacimiento>
  <ciudad>Alicante</ciudad>
</persona>
```

**Busquemos el elemento raíz y resto de elementos**

# XML : Atributos

Un atributo es un componente de las etiquetas que consiste en una pareja nombre (name) / valor (value).

Se puede encontrar en las etiquetas de apertura o en las etiquetas vacías, pero no en las de cierre.

En una misma etiqueta no pueden existir dos atributos con el mismo nombre.

La sintaxis es siempre **nombreAtributo="valorAtributo"**

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<!DOCTYPE persona>
<persona nombre="Alfonso" apellidos="García Fernández">
    <nacimiento dia='13' mes='mar' anyo='2019'></nacimiento>
    <ciudad cp="03001">Alicante</ciudad>
</persona>
```



## XML : Ejercicio – Encontrar errores

```
<?XML version="1.0" encoding="UTF-8" standalone="yes" ?>
<!DOCTYPE biblioteca>
<biblioteca>
  <ejemplar tipo_ejem='libro' titulo='XML práctico' editorial='Ediciones Eni'>
    <tipo>
      <libro isbn='978-2-7460-4958-1' edicion=paginas='347'></Libro>
    </tipo>
    <autor nombre='Sebastien Lecomte'></autor>
    <autor nombre='Thierry Boulanger'></autor>
    <autor nombre=Angel Belinchon Calleja funcion='traductor'></Autor>
    <prestado lector='Pepito Grillo'>
      <fecha_pres dia='13' mes='mar' anyo='2019'></fecha_pres>
      <fecha_devol/>
    </prestado>
  </ejemplar>
</biblioteca>
```

## XML : Ejercicio – Encontrar errores (errores)

```
<?XML version="1.0" encoding="UTF-8" standalone="yes" ?>
<!DOCTYPE biblioteca>
<biblioteca>
  <ejemplar tipo_ejem='libro' titulo='XML práctico' editorial='Ediciones Eni'>
    <tipo>
      <libro isbn='978-2-7460-4958-1' edicion=paginas='347'></Libro>
    </tipo>
    <autor nombre='Sebastien Lecomte'></autor>
    <autor nombre='Thierry Boulanger'></autor>
    <autor nombre=Angel Belinchon Calleja funcion='traductor'></Autor>
    <prestado lector='Pepito Grillo'>
      <fecha_pres dia='13' mes='mar' anyo='2019'></fecha_pres>
      <fecha_devol/>
    </prestado>
  </ejemplar>
</biblioteca>
```

## XML : Ejercicio – Encontrar errores (corrección)

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<!DOCTYPE biblioteca>
<biblioteca>
  <ejemplar tipo_ejem='libro' titulo='XML práctico' editorial='Ediciones Eni'>
    <tipo>
      <libro isbn='978-2-7460-4958-1' edicion="1ª" paginas='347'></Libro>
    </tipo>
    <autor nombre='Sebastien Lecomte'></autor>
    <autor nombre='Thierry Boulanger'></autor>
    <autor nombre="Angel Belinchon Calleja" funcion='traductor'></autor>
    <prestado lector='Pepito Grillo'>
      <fecha_pres dia='13' mes='mar' anyo='2019'></fecha_pres>
      <fecha_devol><fecha_devol/>
    </prestado>
  </ejemplar>
</biblioteca>
```

# XML : Entidades y referencia a entidades

Una entidad consiste en un nombre y su valor (son similares a las constantes en los lenguajes de programación). Las entidades se definen mediante la etiqueta ENTITY.

```
<!ENTITY web "IGFormación">
```

Una referencia a una entidad empieza con el caracter "&", sigue con el nombre de la entidad y termina con ";".

Al abrir el documento XML el procesador sustituye la referencia a la entidad por su valor.

```
<autor>&web;</autor>
```

El procesador XML la convertiría en:

```
<autor>IGFormación</autor>
```

# XML : Instrucciones de procesamiento

Una instrucción de procesamiento es una etiqueta que empieza por "<?" y acaba por "?>" y que contiene instrucciones dirigidas a las aplicaciones que leen el documento. Lo normal es encontrarlas al comienzo del documento.

```
<?xml version="1.0" encoding="utf-8"?>
```

O por ejemplo para asociar una hoja de estilos a un documento XML:

```
<?xml-stylesheet type="text/xsl" href="estilo.xsl" ?>
```

# XML : Comentarios

Los documentos XML pueden contener comentarios que no son interpretados por el motor XML.

- Se incluyen entre las cadenas `<!--` y `-->`
- No puede haber comentarios antes del prólogo
- No puede haber comentarios dentro de etiquetas
- No pueden contener la secuencia `--`
- No se reconocen las entidades

```
<!-- esto es un comentario-->
```

```
<nombre> Alfredo </nombre> <!-- pon un nombre-->
```

```
<nombre> Alfredo <!-- comentario mal puesto--> </nombre>
```

```
<nombre <!-- error garrafal --> > Alfredo </nombre>
```

# XML : documentos bien formados

Todos los documentos XML deben verificar las reglas sintácticas que define la recomendación del W3C para su estándar. Las normas básicas son:

- El documento ha de tener definido un prólogo con la declaración xml completa.
- Existe un único elemento raíz para cada documento, y el resto de elementos estarán anidados a este.
- Cumplir las reglas sintácticas del lenguaje XML para definir elementos, atributos, entidades, etc.

```
<?xml version="1.0"?>
```

```
<mensaje>
```

```
  <destinatario>Tomas</destinatario>
```

```
  <remitente>Juan</ remitente>
```

```
  <asunto>
```

```
    <contenido> No olvides ir a recogerme!</contenido>
```

```
</mensaje>
```

# XML : documentos bien formados

Todos los documentos XML deben verificar las reglas sintácticas que define la recomendación del W3C para su estándar. Las normas básicas son:

- El documento ha de tener definido un prólogo con la declaración xml completa.
- Existe un único elemento raíz para cada documento, y el resto de elementos estarán anidados a este.
- Cumplir las reglas sintácticas del lenguaje XML para definir elementos, atributos, entidades, etc.

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
```

```
<!DOCTYPE mensaje>
```

```
<mensaje>
```

```
  <destinatario>Tomas</destinatario>
```

```
  <remitente>Juan</ remitente>
```

```
  <asunto> No te olvides de mí</asunto>
```

```
  <contenido> No olvides ir a recogerme!</contenido>
```

```
</mensaje>
```