

Tema 4 - Creación, validación y uso de XML

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

Introducción, evolución y estado actual

Introducción

— — —

- XML (eXtensible Markup Language, “lenguaje de marcas extensible”, en castellano).
- XML es un metalenguaje (permite construir otros lenguajes y dialectos) de propósito general.
- XML no define etiquetas: define reglas.
- Tiene como orígenes los lenguajes GML y SGML.
- Características:
 - Sencillo.
 - Robusto.
 - Formato abierto (no tiene coste).

Las versiones de XML son gestionadas por el W3C

— — —

Año	Versión	Observaciones
1998	XML 1.0	Primera edición
2000	XML 1.0	Segunda edición
2004	XML 1.0	Tercera edición
2004	XML 1.1	Primera edición
2006	XML 1.0	Cuarta edición
2006	XML 1.1	Segunda edición
2008	XML 1.0	Quinta edición

Estructura y sintaxis de XML

Estructura y sintaxis

- Un documento XML tiene la siguiente estructura:
 - Declaración.
 - Versión.
 - Codificación.
 - Indicador de dependencias.

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

- Contenido.
 - Elemento raíz.
 - Resto de elementos.

- Un documento XML es una estructura en forma de árbol.
- Contiene una raíz.
- De la raíz descienden elementos que a su vez pueden ser contenedores de otros elementos.



Las relaciones entre los elementos son:

1. Existe un único elemento raíz (**root**).
2. Cada uno de los elementos descendientes directos de un elemento se llama hijo (**children**).
3. El elemento ascendiente directo de un elemento se llama padre (**parent**).
4. Los elementos que tienen un padre común se denominan hermanos (**sibling**).

- Los elementos tienen una etiqueta de apertura, una etiqueta de cierre y, opcionalmente, uno o más atributos.
- Los atributos están compuestos de pares parámetro-valor.

<nombre-elemento atributo-1="valor1" atributo-2="valor2"
>

</nombre-elemento>

- Los elementos pueden estar vacíos, contener texto o a otros elementos.

Reglas para asignar nombre a los elementos o etiquetas

1. Diferencian entre mayúsculas y minúsculas.
2. Deben comenzar con una letra o un guión bajo.
3. Los nombres de elementos tienen que ser idénticos en las etiquetas de apertura y de cierre.
4. Los nombres pueden estar formados por caracteres alfanuméricos, guiones, guiones bajos y puntos.
5. Los nombres no pueden contener espacios (interpreta como nombre del elemento la primera palabra y el resto como parámetros)

```
<?xml version="1.0" encoding="UTF-8"?>
<animales>
  <animal id="1">
    <nombre>Tiburón blanco</nombre>
    <nombrecientifico>Carcharodon
    carcharias</nombrecientifico>
  </animal>
  <animal id="2">
    <nombre>Atún rojo</nombre>
    <nombrecientifico>Thunnus thynnus</nombrecientifico>
    extinción</estadodeconservacion>
  </animal>
</animales>
```

Comentarios

- Contienen información que no es procesada ni interpretada.
- Comienzan por `<!--`
- Terminan por `-->`

`<!-- Este texto es un comentario -->`

Espacios de nombres

Permiten resolver las ambigüedades en caso de que varios elementos tengan el mismo nombre.

Declaración: **xmlns:nombre-espacio-nombres=URL**

Uso: **<nombre-espacio-nombres:nombre-elemento>**

Ejemplo:

xmlns:autorestecnicos='<http://unaurlcualquiera>**':**
<autorestecnicos:autor>

Entidades

- Son un mecanismo para representar información dentro de un documento mediante referencias en lugar de incluirla directamente.
- Tipos:
 - Generales
 - Internas. Contienen el valor en el propio documento.
 - Externas. Contienen el valor en un documento externo.
 - De parámetros

Entidades internas predefinidas

Entidad	Significado	Carácter
&lt;	Símbolo de “menor que”	<
&gt;	Símbolo de “mayor que”	>
&amp;	Símbolo del <u>ampersand</u> (en español et)	&
&apos;	Comilla simple	'
&quot;	Comilla doble	“

Sección CDATA

- Contiene un conjunto de caracteres que no debe ser tratado por el analizador.
- Sintaxis:
 - Comienzo: **<![CDATA[**
 - Fin: **]]>**


```
<?xml version="1.0" encoding="UTF-8"?>
<ficha>
  <nombre>Lenguaje C</nombre>
  <info-html>
    <![CDATA[
      <h1>LENGUAJES DE PROGRAMACIÓN</h1>
      <hr>
      <a href="url-enlace">C</a>
    ]]>
  </info-html>
</ficha>
```

Validaciones de XML

Mecanismos de validación

Dos perspectivas:

- XML bien formado. Es conforme a las reglas generales de XML.
- XML válido. Está bien formado y conforme a las reglas definidas por el creador o diseñador del documento.

XML bien formado. Reglas:

- Debe haber uno y solo un elemento raíz.
- Todos los elementos deben estar cerrados.
- Los elementos deben estar anidados correctamente: no se pueden intercalar aperturas y cierres de elementos distintos.
- Todos los valores de los atributos están entrecomillados.
- Los nombres de elementos y atributos deben cumplir con sus respectivas reglas.

XML válido. Alternativas:

- Alternativa 1: Debe cumplir las reglas especificadas en un DTD.
- Alternativa 2: Debe cumplir las reglas especificadas en un XML-Schema.

DTD

¿Qué es un DTD?

Document Type Definition (Definición de Tipo de Documento).

Contiene una serie de reglas que sirven para validar un documento XML.

Alternativas de uso:

- DTD interno. Incluido en el propio documento XML.
- DTD externo. Almacenado en un fichero externo y referenciado desde el documento XML.

DTD interno. Ejemplo:

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE aviso [
  <!ELEMENT aviso (de,para,mensaje)>
  <!ELEMENT de (#PCDATA)>
  <!ELEMENT para (#PCDATA)>
  <!ELEMENT mensaje (#PCDATA)>
]>
<aviso>
  <de>David</de>
  <para>Rosalía</para>
  <mensaje>Mañana nos vemos en el estudio a las 10.</mensaje>
</aviso>
```


DTD externo. Ejemplo:

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE aviso SYSTEM "dtdExterno.dtd">
<aviso>
    <de>David</de>
    <para>Rosalía</para>
    <mensaje>Mañana nos vemos en el estudio a las
10.</mensaje>
</aviso>
```

DTD. Entidades:

En un DTD, una entidad es el equivalente a una constante en programación.

Declaración:

```
<!ENTITY nombre-entidad "texto de reemplazo">
```

Uso:

```
&nombre-entidad;
```

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE fichas [
    <!ENTITY poblacion "Tejeda de Tiétar, Cáceres">
]>
<fichas>
    <ficha>
        <empresa nombre="Tractores
Toledo"></empresa>
        <direccion>&poblacion;</direccion>
    </ficha>
</fichas>
```

Declaración entidad externa privada:

```
<!ENTITY provincia SYSTEM "provincia.txt">
```

Uso: `<poblacion>&poblacion;</poblacion>`

DTD. Anotaciones:

Se utilizan para identificar el formato de entidades que no son XML y que, por lo tanto, no se van a procesar o como valor válido para un atributo.

DTD. Elementos;

Un elemento en un DTD indica en qué condiciones debe encontrarse dicho elemento en el documento XML que se está validando.

<!ELEMENT nombre-elemento contenido>

Tipos de contenido

— — —

Tipo de contenido	Descripción
EMPTY	Indica que el elemento referenciado debe estar vacío.
ANY	Indica que el elemento referenciado puede contener cualquier contenido.
(#PCDATA)	Indica que el elemento referenciado puede datos de tipo carácter (<u>Parsed Character Data</u>).
(<u>nombreElemento</u>)	Indica que el elemento referenciado puede contener al elemento indicado.
(nombreElemento1, nombreElemento2, ...)	Indica que el elemento referenciado puede contener una sucesión de elementos indicados como una lista separada por comas.

Cardinalidades

Notación	Descripción	Ejemplo
(<u>nombreElemento</u>)	Una única ocurrencia del elemento.	<!ELEMENT aviso (de)>
(<u>nombreElemento?</u>)	Cero o una única ocurrencia del elemento.	<!ELEMENT aviso (<u>de?</u>)>
(<u>nombreElemento+</u>)	Una o más ocurrencias del elemento.	<!ELEMENT aviso (mensaje+)>
(<u>nombreElemento*</u>)	Cero o más ocurrencias del elemento.	<!ELEMENT aviso (mensaje*)>
(nombreElemento1, nombreElemento2,...)	Debe contener todos los elementos de la lista.	<!ELEMENT aviso (de, para, mensaje)>
(nombreElemento1 nombreElemento2)	Debe contener uno u otro elemento	<!ELEMENTO aviso (de para, mensaje)>

DTD. Atributos:

Los DTD permiten restringir las reglas y condiciones de los atributos usados en los documentos XML que validan.

Sintaxis:

```
<!ATTLIST elemento nombre-atributo tipo-atributo valor-atributo>
```

Tipos de atributos

Tipo	Descripción
CDATA	Cadena de caracteres.
(valor1 valor2 ...)	Una lista de posibles valores.
ID	Un identificador único
IDREF	Una referencia a un identificador único de otro elemento
IDREFS	Una lista de referencias a identificadores de otros elementos
NMTOKEN	Un nombre XML válido.
NMTOKENS	Una lista de nombres XML válidos separadas por espacios.
ENTITY	Una referencia a una entidad
ENTITIES	Una referencia a un conjunto de entidades
NOTATION	Un nombre de una notación
<u>xml:lang</u>	Indica el idioma del contenido.
<u>xml:space</u>	Indica qué hacer con espacios, tabulaciones y retornos de carro múltiples.

Valores de atributos

— — —

Valor	Descripción
valor	El valor por defecto del atributo
#REQUIRED	Indica que el atributo es obligatorio
#IMPLIED	Indica que el atributo es opcional
#FIXED valor	El valor del atributo

XML-Schema

¿Qué es XML-Schema?

Es el lenguaje utilizado para describir la estructura, relaciones y restricciones de los documentos XML.

- Mejora la precisión proporcionada por los DTD.
- Se definen en ficheros XML como extensión .xsd.
- Se referencias desde el documento XML a validar:

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

```
<element-raiz
```

```
xmlns:xs="http://www.w3.org/2001/XMLSchema-instance"
```

```
xs:noNamespaceSchemaLocation="fichero-esquema.xsd">
```

Ejemplo de XML-Schema

```
<?xml version="1.0" encoding="UTF-8"?>
<mensaje
xmlns:xs="http://www.w3.org/2001/XMLSchema-instance"
  xs:noNamespaceSchemaLocation="mensaje.xsd">
  <de>Anna</de>
  <para>Rocío</para>
  <mensaje>Ya estoy en la biblioteca.</mensaje>
</mensaje>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="mensaje">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="de" type="xs:string"/>
        <xs:element name="para" type="xs:string"/>
        <xs:element name="mensaje" type="xs:string"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Estructura del XML-Schema

- Declaración del documento XML.
 - `<?xml version="1.0" encoding="UTF-8"?>`
- Declaración del elemento raíz y del espacio de nombres.
 - `<xs:schema`
`xmlns:xs="http://www.w3.org/2001/XMLSchema">`
- Declaración de los elementos.
- Declaración de los atributos.

Tipos de elementos

- **Locales:** Son hijos de elementos que no son raíz y solo se pueden utilizar una vez.
- **Globales.:** Son hijos del elemento raíz y pueden ser reutilizados.
- **Simples:** Pueden contener textos (cadenas de caracteres, fechas, valores lógicos, etc.).
 - `<xs:element name=""></xs:element>`
- **Complejos:** Pueden contener a otros elementos, estar vacíos o contener otros elementos y textos.
 - `<xs:element name="" type="">
 <xs:complexType/>
 </xs:complexType>
</xs:element>`

Sub-elementos

- Son contenidos por elementos.
- Tipos:
 - xs:sequence. Secuencia de elementos obligatorios.
 - xs:choise. Secuencias de elementos de los que debe aparecer uno.
 - xs:all. Secuencia de elementos opcionales.

```
<xs:element name="direccion">
  <xs:complexType>
    <xs:choice>
      <xs:element name="particular" type="xs:string"/>
      <xs:element name="profesional" type="xs:string"/>
    </xs:choice>
  </xs:complexType>
</xs:element>
```

Atributos

- Se definen a nivel de elemento.
- Componentes:
 - Tipo (type): por defecto, anySimpleType.
 - Uso (use): required, optional o prohibited.
 - Valor por defecto (default).
 - Valor fijo si existe (fixed).

<xs:attribute name="" type="" />

Tipos de datos

— — —

Tipo (utilizando <u>xs</u> como prefijo)	Descripción
<u>xs:string</u>	Cadena de caracteres.
<u>xs:integer</u>	Números enteros.
<u>xs:decimal</u>	Números decimales.
<u>xs:boolean</u>	Tipo de dato lógico.
<u>xs:date</u>	Fecha.
<u>xs:time</u>	Instante de tiempo en <u>en</u> formato " <u>hh:mm:ss</u> ".
<u>xs:duration</u>	Un periodo de tiempo.

Restricciones de los atributos

— — —

Faceta	Descripción
<u>xs:length</u>	Determina una longitud fija.
<u>xs:minLength</u>	Determina una longitud mínima.
<u>xs:maxLength</u>	Determina una longitud máxima.
<u>xs:totalDigits</u>	Determina el máximo número de dígitos que puede tener un número.
<u>xs:fractionDigits</u>	Determina el máximo número de decimales que puede tener un número.
<u>xs:minExclusive</u>	Determina que el valor debe ser mayor que el valor indicado.
<u>xs:maxExclusive</u>	Determina que el valor debe ser menor que el valor indicado.

Faceta	Descripción
<u>xs:minInclusive</u>	Determina que el valor debe ser mayor o igual que el valor indicado.
<u>xs:maxInclusive</u>	Determina que el valor debe ser menor o igual que el valor indicado.
<u>xs:enumeration</u>	Determina una lista de valores posibles.
<u>xs:whiteSpace</u>	Determina como tratar los espacios en blanco, las tabulaciones y los saltos de línea.
<u>xs:pattern</u>	Determina un patrón de caracteres permitidos.

Comentarios en los XML-Schema

Se indican con el elemento `<annotation>`

```
<xs:annotation>
```

```
    <xs:appinfo>Validador de mensajes</xs:appinfo>
```

```
    <xs:documentation xml:lang="es">
```

```
        Este esquema valida mensajes privados seguros.
```

```
    </xs:documentation>
```

```
</xs:annotation>
```

Otras herramientas

Otras alternativas de validación

— — —

- RELAX NG
- Schematron

XML Aplicado

La versatilidad de XML es tal que se admiten usos en múltiples entornos:

- XML en aplicaciones, como estructura de datos.
- XML en la Web, como herramienta de diseño de páginas web.
- XML Servicios Web, como formato de intercambio.

Herramientas de visualización y edición de XML

- Editores de texto.
- Navegadores web.
- XML Notepad.
- Liquid Studio.
- XMLPad.
- Eclipse.
- Editores web.

Ejemplos y Ejercicios

Ejemplos

— — —

- Estructura de un documento XML
- Estructuración de la información
- Estructura de árbol
- Ejemplo de creación de un DTD
- Ejemplo de XML-Schema

Ejercicios

— — —

- [Ejercicios XML Básico](#)
- [Ejercicios DTD Básico](#)
- [Ejercicios DTD Avanzado](#)
- [Ejercicios XML-Schema Básico](#)
- [Ejercicios XML-Schema Avanzado](#)

Documentación Extra

Documentación Extra

— — —

- [Google Site Todo XML](#)
- [Estándares](#)
- [Más sobre XML](#)
- [Tutorial de W•Schools de XML](#)