

T4 - XML. CREACIÓN, VALIDACIÓN Y UTILIZACIÓN

Lenguaje de Marcas y Sistemas de Gestión de la
Información

CONTENIDOS

4.1. INTRODUCCIÓN, EVOLUCIÓN Y ESTADO ACTUAL

4.2. ESTRUCTURA Y SINTAXIS DE XML

4.3. VALIDACIÓN DE XML

4.4. XML APLICADO

Introducción

XML es, junto con HTML, uno de los lenguajes de marcas más característicos. En realidad, XML es un metalenguaje y la base para la construcción de otros lenguajes más específicos.

XML define las reglas de construcción de los documentos, pero no se especifican las etiquetas tal y como se hace en HTML. En XML, en lugar de haber elementos o atributos predefinidos, hay mecanismos para definirlos.

Al tratarse de un estándar, alrededor de XML hay un buen número de tecnologías y herramientas: mecanismos de validación, editores, programas de visualización o librerías de programación para integrarlo en desarrollos de software propios.

4.1. Introducción, evolución y estado actual

El objetivo de XML es, en cambio, menos concreto. No está orientado a la web, pero puede utilizarse para presentar datos en un navegador como HTML. Tampoco está orientado a la representación de interfaces gráficas de usuario, pero se utiliza en un buen número de tecnologías para ello. También se puede utilizar para intercambiar información entre sistemas, guardar la configuración de una aplicación o difundir información mediante RSS.

El lenguaje de marcado XML (eXtensible Markup Language, «lenguaje de marcas extensible», en castellano) es, en realidad, un metalenguaje de propósito general desarrollado por el W3C (World Wide Consortium) y se utiliza en multitud de contextos.

4.1. Introducción, evolución y estado actual

XML proviene del lenguaje GML (Generalized Markup Language) y de SGML (Standard Generalized Markup Language). Fue lanzado en 1998 y se encuentra en la versión 1.1. Es un formato abierto, lo que implica que se puede utilizar de manera libre y sin coste. La extensión de los ficheros XML es .xml y el tipo MIME es application/text y text/xml.

El W3C anima a crear los documentos en XML 1.0 salvo que se necesiten las nuevas funciones añadidas en XML 1.1. En la gran mayoría de los casos estas funciones no son necesarias, por lo que habitualmente los documentos XML se encuentran en versión 1.0.

4.1. Introducción, evolución y estado actual

A continuación, se muestra un ejemplo de un documento XML 1.0:

```
<?xml version="1.0" encoding="UTF-8"?>
<animal>
  <nombre>Tiburón blanco</nombre>
  <nombrecientifico>Carcharodon carcharias</nombrecientifico>
  <familia>Lamnidae</familia>
  <estadodeconservacion>Vulnerable</estadodeconservacion>
</animal>
```

4.1. Introducción, evolución y estado actual

XML tiene fundamentos robustos y una sintaxis muy simple, lo que lo convierte en una herramienta muy versátil y que goza de una gran aceptación. XML y sus dialectos se utilizan en múltiples contextos; creación de ficheros de configuración, diseño de interfaces gráficas, publicación de contenidos, intercambio de información entre aplicaciones o diseño de páginas web son solo algunas de las áreas de aplicación en las que se emplea.

A pesar de la sencillez de XML, existe un importante conjunto de tecnologías alrededor que convierten el lenguaje en una herramienta poderosa: validadores, transformadores, editores... La estandarización a la que se somete permite que el ecosistema de aplicaciones relacionadas con el lenguaje sea uno de sus activos más importantes.

4.2. Estructura y sintaxis de XML

Los documentos XML se estructuran en dos partes principales: prólogo y cuerpo. El prólogo contiene información relativa al conjunto del documento y el cuerpo recoge los elementos con la información propiamente dicha. También permite incluir comentarios que no serán procesados.

Un DTD (Document Type Definition) contiene reglas que determinan que un documento XML es válido en un contexto determinado. Más adelante se explicará en detalle qué son y cómo se construyen las validaciones.

4.2. Estructura y sintaxis de XML

4.2.1. Declaración

La declaración del documento XML se realiza en el prólogo, siendo este opcional. Si existe, debe ser la primera línea del documento. En él se indica que el documento es XML, la versión y la codificación utilizada mediante los siguientes atributos:

- **version**. Versión de XML.
- **encoding**. Codificación utilizada en el documento.
- **standalone**. Indica si el documento es independiente o si existe un DTD externo para realizar la validación. Admite los valores yes (es independiente) y no (está asociado a un DTD de validación). Este atributo es opcional.

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

4.2. Estructura y sintaxis de XML

4.2.2. Estructura, sintaxis, elementos y atributos

El cuerpo del documento XML está compuesto por elementos. Los elementos son el componente fundamental de XML. Se parecen mucho a los elementos de HTML, pues tienen una etiqueta de apertura y otra de cierre, que en el caso de XML es obligatoria.

Las etiquetas de apertura tienen un nombre y, opcionalmente, uno o más atributos separados por espacios. La etiqueta de cierre debe tener el mismo nombre que la de apertura, pero este estará precedido por el símbolo / .

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

4.2. Estructura y sintaxis de XML

4.2.2. Estructura, sintaxis, elementos y atributos

Las reglas para la **asignación de nombres de elementos** son las siguientes:

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

4.2. Estructura y sintaxis de XML

4.2.2. Estructura, sintaxis, elementos y atributos

Ejemplos **elementos erróneos:**

```
<color>Azul</Color>
```

```
<-color>Azul</-color>
```

```
<.color>Azul</ .color>
```

```
<el color>Azul</el color>
```

Ejemplos **elementos correctos:**

```
<color1>Azul</color1>
```

```
<_color>Azul</_color>
```

```
<COLOR>Azul</ COLOR>
```

```
<el_color>Azul</el_color>
```

4.2. Estructura y sintaxis de XML

4.2.2. Estructura, sintaxis, elementos y atributos

Las reglas de **creación de atributos** se muestran a continuación:

- Deben tener **asignado un valor**.
- Los valores siempre van **entrecomillados**, admitiendo **comillas simples y dobles** (deben coincidir el tipo de comilla de cierre con el de apertura).
- **Diferencian** entre **mayúsculas y minúsculas**.
- Deben comenzar con una **letra** o un **guion bajo**.
- Pueden estar formados por caracteres **alfanuméricos, guiones, guiones bajos y puntos**.

Ejemplo de **atributo válido**:

```
<nombre familia="Lamnidae">Tiburón blanco</nombre>
```

4.2. Estructura y sintaxis de XML

4.2.2. Estructura, sintaxis, elementos y atributos

En XML se puede caracterizar un elemento mediante atributos o elementos hijos. No existe una regla para determinar en qué caso se deben utilizar uno u otro, pero sí casuísticas en la que los atributos no son suficientes para crear la estructura de datos que se desea.

Los elementos pueden contener texto u otros elementos. En el segundo caso, la relación entre los elementos es jerárquica: un elemento puede contener a otros elementos y estos, a su vez, ser contenedores de otros más. Esta estructura se denomina «de árbol» y se caracteriza por tener un único elemento en la parte alta de la jerarquía (denominado raíz) que se ramifica hasta llegar a los nodos finales u hojas.

4.2. Estructura y sintaxis de XML

4.2.2. Estructura, sintaxis, elementos y atributos

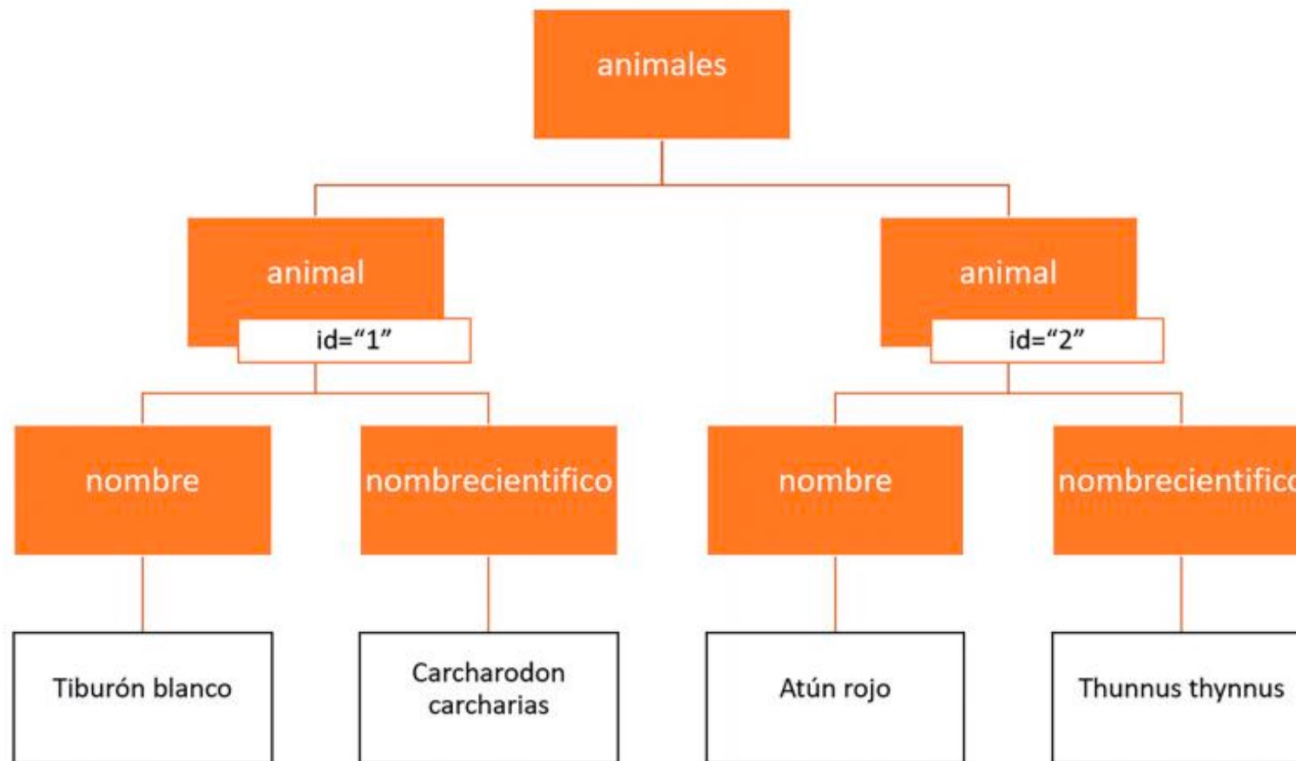
Las relaciones entre los elementos de un documento XML se denominan de la siguiente manera:

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

4.2. Estructura y sintaxis de XML

4.2.2. Estructura, sintaxis, elementos y atributos

Ejemplo **estructura de árbol**:



4.2. Estructura y sintaxis de XML

4.2.2. Estructura, sintaxis, elementos y atributos

Ejemplo estructura de xml:

```
<?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>
  </animal>
</animales>
```

4.2. Estructura y sintaxis de XML

4.2.2. Estructura, sintaxis, elementos y atributos

En el ejemplo anterior que se muestra a continuación las relaciones entre los elementos son las siguientes:

- `<animales>` es el elemento raíz.
- Los elementos `<animal>` son hijos (**children**) del elemento `<animales>`.
- El elemento `<animales>` es padre (**parent**) de los elementos `<animal>`.
- Los elementos `<animal>` son hermanos (**sibling**) entre sí.
- Los elementos `<nombre>` y `<nombrecientifico>` son hijos del elemento `<animal>`.