

## SIS2-C3: XML

### Índice

SIS2-C3: XML.....	1
¿Qué es XML? .....	1
XML no hace nada.....	1
XML es sólo texto plano .....	2
Con XML inventas tus propias etiquetas .....	2
XML es una Recomendación del W3C .....	2
XML está en todas partes .....	2
XML tiene estructura de árbol .....	2
Ejemplo de árbol XML: .....	3
Sintaxis del documento XML.....	4
Prólogo.....	4
Elemento.....	5
Atributos.....	5
Entidades predefinidas.....	5
Visualización de documentos XML.....	5

### ¿Qué es XML?

- XML son las siglas de **Extensible Markup Language**
- XML es un **lenguaje** de **marcas** muy similar a HTML
- XML fue diseñado **para** el transporte de **datos**, no para mostrar datos
- Las etiquetas de XML no están predefinidas. Cada cual debe **definir sus propias etiquetas**
- XML está diseñado para ser **auto-descriptivo**
- XML es una **Recomendación del W3C**

### XML no hace nada

Tal vez esto es un poco difícil de entender, pero XML no hace nada. XML fue creado para estructurar, almacenar, y transportar información.

El ejemplo siguiente es una nota a Tove de Jani, almacenados en formato XML:

```
<nota>
  <de>Tove</de>
  <para>Jani</para>
  <titulo>Recordatorio</titulo>
  <cuerpo> No se olvide de mí este fin de semana!</cuerpo>
</nota>
```

En la nota anterior el documento XML es bastante auto-descriptivo. Se muestra el remitente y el receptor de la información, así como un título y un cuerpo del

mensaje.

Pero aún así, este documento XML no hace nada. Es sólo información pura envuelta en las etiquetas. Alguien debe escribir un software para enviar, recibir o mostrar el mensaje.

### ***XML es sólo texto plano***

XML no tiene nada especial. Es simplemente texto. Un software que permita tratar texto plano también puede tratar XML.

Sin embargo, existen aplicaciones específicas que permiten tratar las etiquetas XML. El significado funcional de las etiquetas depende de la naturaleza de la solicitud.

### ***Con XML inventas tus propias etiquetas***

Las etiquetas en el ejemplo anterior (como <para> y <de>) no están definidos en cualquier estándar XML. Estas etiquetas han sido "inventadas" por el autor del documento XML.

Esto se debe a que el lenguaje XML no tiene etiquetas predefinidas.

Otros códigos que utilizan etiquetas, como HTML (y su estructura) están predefinidos. Los documentos HTML sólo pueden utilizar las etiquetas definidas en el estándar HTML (como <p>, <h1>, etc.)

XML permite al autor a definir sus propias etiquetas y su propia estructura del documento.

### ***XML es una Recomendación del W3C***

El "Extensible Markup Language" (XML) se convirtió en una recomendación del W3C el 10 de Febrero de 1998.

### ***XML está en todas partes***

Ha sido asombroso ver lo rápidamente que el estándar XML se ha desarrollado y con qué rapidez un gran número de proveedores de software han adoptado la norma.

XML es ahora tan importante para la Web, como HTML es la fundación de la Web.

XML está en todas partes. Es la herramienta más común para las transmisiones de datos entre todo tipo de aplicaciones, y se convierte más y más popular en la zona de almacenamiento y descripción de la información.

### ***XML tiene estructura de árbol***

Documentos XML debe contener un **elemento raíz (root)**. Este elemento es "el padre" de todos los demás elementos.

Los elementos XML en un documento un documento árbol. El árbol empieza en la raíz y se expande por las ramas hasta el nivel más bajo.

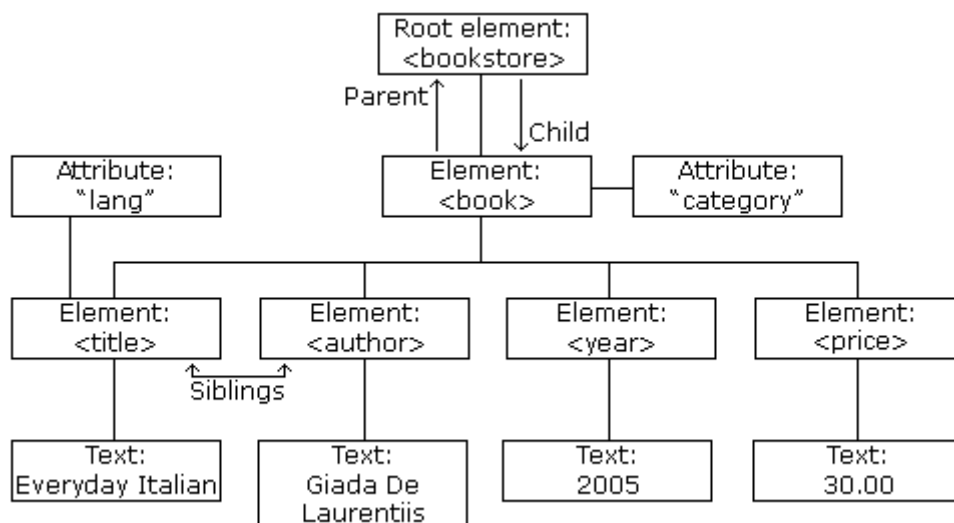
Todos los elementos pueden tener subelementos (elementos niño o *child elements*):

```
<root>
  <child>
    <child>
      <subchild>.....</subchild>
      <subchild> .....</ subchild>
    </child>
  </child>
</root>
```

Los términos padre, hijo, hermano y se utilizan para describir las relaciones entre los elementos. Los elementos padre (*parent*) tienen hijos (*children*). Los niños en el mismo nivel se llaman hermanos (hermanos o hermanas, *siblings*).

Todos los elementos pueden tener contenido de texto y atributos (al igual que en HTML).

### Ejemplo de árbol XML:



La imagen de arriba representa el documento XML mostrado a continuación:

```
<bookstore>
<book category="COOKING">
  <title lang="en">Everyday Italian</title>
  <author>Giada De Laurentiis</author>
  <year>2005</year>
  <price>30.00</price>
</book>
<book category="CHILDREN">
  <title lang="en">Harry Potter</title>
  <author>J K. Rowling</author>
  <year>2005</year>
  <price>29.99</price>
</book>
<book category="WEB">
  <title lang="en">Learning XML</title>
  <author>Erik T. Ray</author>
  <year>2003</year>
  <price>39.95</price>
</book>
</bookstore>
```

El elemento raíz en el ejemplo es <bookstore> (librería). Todos los elementos <book> (libro) en el documento están contenidas dentro de <bookstore>.

El elemento <book> tiene 4 hijos: <title> (título), <author> (autor), <year> (año) y <price> (precio).

## Sintaxis del documento XML

Un documento XML se construye con diferentes componentes:

### Prólogo

El prólogo define los datos principales que permitan a las aplicaciones “leer” el documento XML sin cometer errores.

La cabecera mínima debería definir al menos los siguientes datos:

- **versión:** indica que especificación del w3c se ha seguido al escribir el documento. Existen actualmente dos especificaciones:
  - *XML 1.0:* Inicialmente definida en 1998, aunque ligeramente modificada en cuatro ocasiones sin cambiar su número de versión. Es la especificación recomendada para uso general.
  - *XML 1.1:* Fue inicialmente definida en 2004, incluyendo algunas características de utilidad en algunos entornos concretos. Solo es recomendado su uso en casos estrictamente necesarios.
- **codificación:** indica la codificación de caracteres en la que se ha escrito el

contenido. La codificación recomendable para cualquier documento es *UTF-8*, o *unicode*, pues permite la coexistencia en un mismo documento de todos los caracteres existentes. Pero existen todavía muchas aplicaciones que no reconocen esa codificación. La alternativa sería usar la codificación de Europa Occidental: *ISO-8859-1*. O mejor su versión más actual, *ISO-8859-15*, que incluye el símbolo €.

El siguiente ejemplo muestra una cabecera que se podría usar sin problemas para un documento XML cuyo contenido está en lengua castellana o catalana:

```
<?xml version="1.0" encoding="ISO-8859-15">
```

La cabecera se sitúa al inicio del documento, en la primera línea. Adicionalmente podría contener más líneas indicando la definición tipo de documento (DTD – *document type definition*) o otras instrucciones para su procesamiento en aplicaciones.

## Elemento

Entendemos por elemento, etiqueta o *tag* a cada uno de los nodos que forman el árbol del documento.

Un documento XML bien formado debe cumplir algunos requisitos. Ha de tener un elemento raíz (*root*) que contenga al resto de elementos. Todos los elementos abiertos deben cerrarse, antes de que su elemento padre (*parent node*) sea cerrado.

Finalmente los nombres de los elementos han de ser escritos sin espacios, usando únicamente caracteres del alfabeto inglés y sin contener caracteres numéricos o propios de otras lenguas, como serían por ejemplo la ñ, la ç, o los caracteres acentuados.

## Atributos

Los elementos pueden tener atributos, que son una manera de incorporar características o propiedades a los elementos de un documento. Los valores de los atributos deben escribirse entre comillas:

```
<elemento nombre_atributo="valor_atributo"></elemento>
```

## Entidades predefinidas

Entidades para representar caracteres especiales para que, de esta forma, no sean interpretados como marcado en el procesador XML.

## Visualización de documentos XML

Los documentos XML pueden ser visualizados en algunos navegadores web, entre otras aplicaciones.

Adicionalmente, pueden ir vinculados a una hoja de estilos en cascada, CSS, que definirá como deben ser mostrados por el navegador.