
LABORATORIO TECNOLOGÍAS AUDIOVISUALES EN LA WEB

José María Cañas Plaza

jmplaza@gsyc.es



Grado Sistemas Audiovisuales y Multimedia
curso 2015-2016

Introducción a XML

Contenido

- Introducción
- Documento XML, sintaxis
- Tipos de Documentos XML: DTD, XML-Schema
- XML DOM
- Herramientas: XPath, XSLT

Introducción

- **eXtensible Markup Language (XML)** es un lenguaje de marcado que define unas reglas para expresar documentos en un formato que es fácil de entender por las personas y por las máquinas.
- Estandarizado por el consorcio W3C's XML, especificación 1.0 y otras muchas.
- Está diseñado para facilitar la simplicidad, generalidad y usabilidad en Internet
- Está basado en texto plano, con soporte de Unicode

- Aunque se diseñó para documentos se usa mucho para representación de estructuras arbitrarias de datos (en AJAX, en servicios Web)
- No tiene un juego de etiquetas (elementos) fijo
- Es extensible, hay varios mecanismos para definir lenguajes específicos basados en XML, con sus propias etiquetas
- Se han desarrollado muchas APIs que ayudan a procesar datos XML
- XML no hace nada, representa información
- Separa contenido, datos, de presentación

Lenguaje de marcado

- Orígenes en mundo de la publicación: **marcas** indican cómo componer la zona marcada para imprimir
- Los documentos consisten en una serie de datos marcados mediante etiquetas

Sólo texto

ALBA Abril de 1915 Granada Mi corazón oprimido
siente junto a la alborada el dolor de sus
amores y el sueño de las distancias.

Texto marcado

```
]ALBA[ ← Título, negrita, centrado, 14pt  
]Abril de 1915[← SubTítulo, negrita, centrado  
]Granada[← SubTítulo cursiva, centrado  
]Mi corazón oprimido [← Verso, 10pt  
]siente junto a la alborada [← Verso  
]el dolor de sus amores [← Verso  
]y el sueño de las distancias. [← Verso
```

Resultado

ALBA
Abril de 1915
Granada

Mi corazón oprimido
siente junto a la alborada
el dolor de sus amores
y el sueño de las distancias.

Documento XML, sintaxis

- *Lenguaje de marcado para describir documentos (y datos)*
- Las etiquetas describen la estructura del documento
- Un **elemento** es etiqueta inicial, otra final y contenido entre ambas
- La etiqueta inicial puede incluir **atributos**

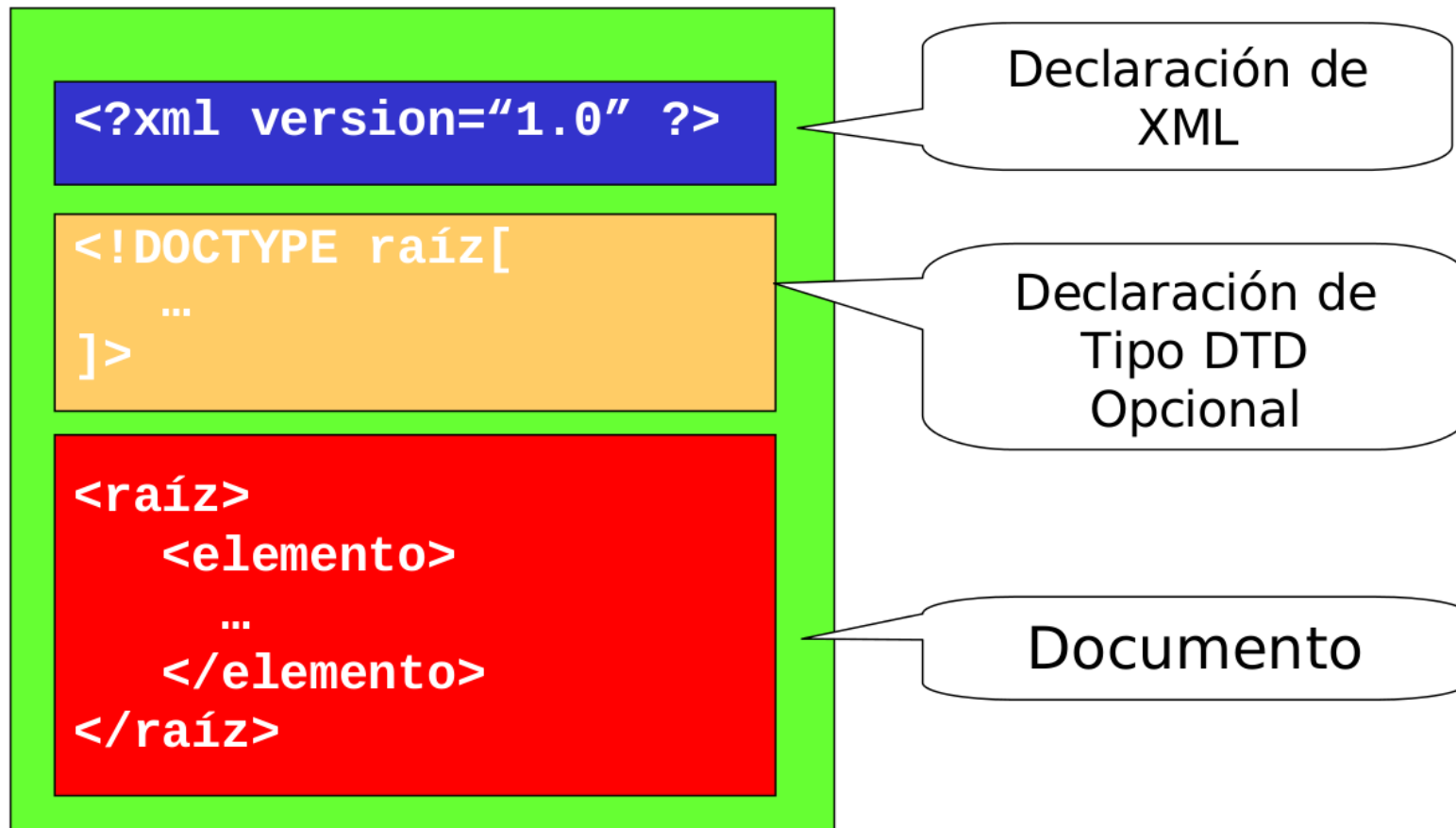
`<etiqueta atributo= valor >.....</etiqueta>`

```
<?xml version="1.0" encoding="UTF-8"?>
<bookstore>
  <book category="cooking">
    <title lang="en">Everyday Italian</title>
    <author>Giada De Laurentiis</author>
    <year>2005</year>
    <price>30.00</price>
  </book>

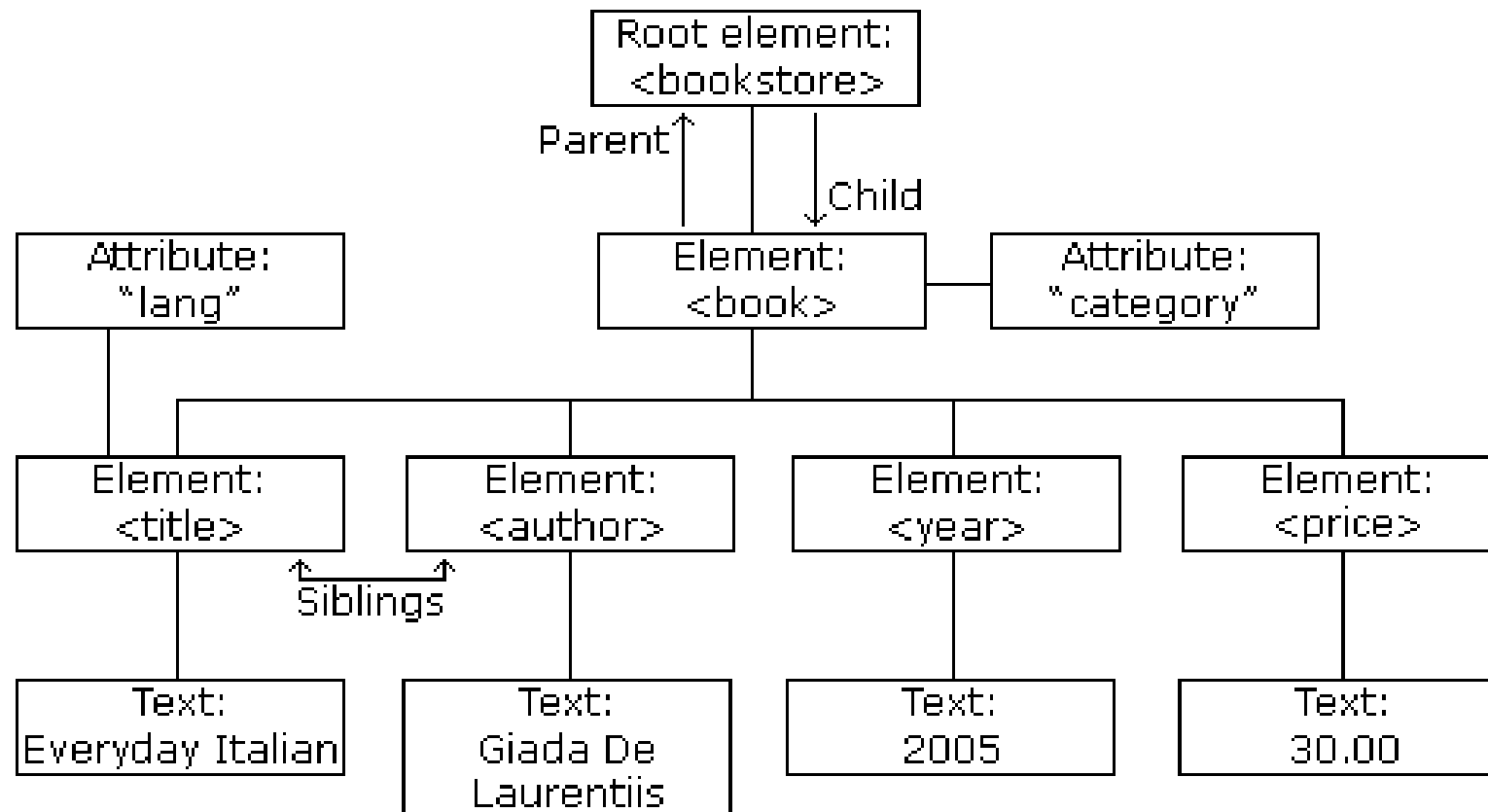
  <book category="children">
    <title lang="en">Harry Potter</title>
    <author>J K. Rowling</author>
    <year>2005</year>
    <price>29.99</price>
  </book>

  <book category="web" cover="paperback">
    <title lang="en">Learning XML</title>
    <author>Erik T. Ray</author>
    <year>2003</year>
    <price>39.95</price>
  </book>
</bookstore>
```


Estructura de un documento XML



- Un documento XML está formado por un **árbol de elementos**
- Empieza por un nodo raíz (el primero) y se tienen nodos hijo



Sintaxis XML

- El documento XML tiene un elemento raíz
- Las etiquetas se cierran
- Distingue mayúsculas de minúsculas
- Las etiquetas se pueden anidar, pero sin cruzarse
- Los atributos van entrecomillados
- Un documento que cumple la sintaxis se dice que está *bien formado*
- No se permiten errores, un error en XML puede detener la aplicación

Espacios de nombres

- El desarrollador define los nombres de los elementos. Puede crear conflictos cuando se mezclan documentos XML de diferentes aplicaciones
- Se solucionan usando prefijos en los nombres

```
<root>
```

```
<h:table xmlns:h="http://www.w3.org/TR/html4/">
```

```
  <h:tr>
```

```
    <h:td>Apples</h:td>
```

```
    <h:td>Bananas</h:td>
```

```
  </h:tr>
```

```
</h:table>
```

```
<f:table xmlns:f="http://www.w3schools.com/furniture">
```

```
  <f:name>African Coffee Table</f:name>
```

```
  <f:width>80</f:width>
```

```
  <f:length>120</f:length>
```

```
</f:table>
```

```
</root>
```

Tipos de documentos XML

- Un documento XML se dice *válido* si es conforme a cierta definición de tipo de documento
- Hay dos maneras de definir un tipo de documento:
 - DTD (Document Type Definition), la original
 - XML Schema, una alternativa (XML a su vez) a DTD
- Una definición de tipo de documento establece las reglas y los elementos y atributos posibles.

DTD

- La declaración DOCTYPE es una referencia a un fichero externo DTD
- DTD define la estructura con una lista de elementos permitidos

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE note SYSTEM "Note.dtd">
<note>
<to>Tove</to>
<from>Jani</from>
<heading>Reminder</heading>
<body>Don't forget me this weekend!</body>
</note>
```

```
<!DOCTYPE note
[
<!ELEMENT note (to,from,heading,body)>
<!ELEMENT to (#PCDATA)>
<!ELEMENT from (#PCDATA)>
<!ELEMENT heading (#PCDATA)>
<!ELEMENT body (#PCDATA)>
]>
```

XML-Schema

- Más potentes: permite definir espacios de nombres
- Soporta tipos de datos, ajusta mejor restricciones y conversiones
- Un Schema es a su vez un XML

```
<xs:element name="note">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="to" type="xs:string"/>
      <xs:element name="from" type="xs:string"/>
      <xs:element name="heading" type="xs:string"/>
      <xs:element name="body" type="xs:string"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
```

XML DOM

- Define una manera estandarizada para acceder y manipular documentos XML
- Todos los elementos del documento XML se pueden acceder a través del DOM, localizarlos, cambiarlos, añadir nuevos o borrar existentes.
- XML DOM define los objetos, propiedades y métodos de todos los elementos XML
- Se puede ver como un modelo estandar de objetos para XML
- Se puede ver como un API para programar usando documentos XML
- Es independiente de lenguaje y de plataforma
- Está estandarizado por el consorcio W3C

Herramientas asociadas

XPath

- XPath (XML Path language) es un lenguaje para encontrar información dentro de un documento XML
- XPath es una sintaxis para definir partes de un documento XML
- XPath usa expresiones para navegar en el documento
- XPath contiene una biblioteca con funciones standard
- Está estandarizado por el W3C

XSLT

- XSLT (eXtensible Stylesheet Language Transformations) es el lenguaje de presentación para documentos XML
- XSLT más sofisticado que CSS, más potente.
- XSLT use XPath para encontrar información en el documento XML

Referencias

- <https://en.wikipedia.org/wiki/XML>
- <http://www.w3schools.com/xml/default.asp>