

Introducción a XML



IES Gonzalo Nazareno
CONSEJERÍA DE EDUCACIÓN

Alberto Molina Coballes

Lenguajes de Marcas y Sistemas de
Gestión de la Información



18 de septiembre de 2012

¿Qué es XML?

- *Wikipedia*: "Es un conjunto de reglas para codificar un documento en un formato legible por una máquina"
- No es un lenguaje de marcas, sino que hay cientos de lenguajes de marcas que cumplen con las especificaciones de XML y se manejan de la misma forma
- Ejemplos: XHTML, RSS, Atom, Docbook, OpenDocument, OOXML, SVG, MathML, SOAP, ...¹
- Prácticamente todos los nuevos lenguajes de la web están basados en XML

¹http://en.wikipedia.org/wiki/List_of_XML_markup_languages

Características de XML

- Sus siglas provienen de *eXtensible Markup Language*
- Puede almacenar y organizar cualquier tipo de información
- Es un estándar internacional abierto (W3C)
- Utiliza Unicode por defecto → Cualquier idioma
- Permite revisar sintaxis y validar documentos
- Es fácil de leer por personas y por aplicaciones o programas
- Es posible exportarlo a numerosos formatos

Estructura (sintaxis) de un documento XML (I)

- En la primera línea se incluye la declaración XML:
`<?xml version="1.0" encoding="UTF-8" ?>`
- Normalmente incluye un prólogo con las "instrucciones"
- Incluye elementos que empiezan por una etiqueta o tag de apertura, a continuación se pone el contenido y termina con la etiqueta o tag de cierre:

```
<alumno>Pepa Ramírez Heredia</alumno>
```

- Los elementos pueden tener atributos (entrecomillados):
`<alumno dni="45.123.123-J">Pepa Ramírez Heredia</alumno>`
- Tiene estructura jerárquica:

```
<alumnos>
```

```
  <alumno dni="45.123.123-J">Pepa Ramírez Heredia</alumno>
```

```
  <alumno dni="41.321.321-H">José González Pons</alumno>
```

```
</alumnos>
```

Estructura (sintaxis) de un documento XML (II)

- Los elementos vacíos tienen que cerrarse siempre:
`<alumno dni="45.123.123-J">Pepa Ramírez Heredia</alumno>
<becado/>`
- Las etiquetas distinguen mayúsculas de minúsculas
- Los elementos deben estar correctamente anidados
- Los documentos deben tener un elemento raíz
- Ejemplo:
- Las etiquetas XML:
 - Pueden contener letras, números y caracteres especiales
 - No pueden empezar por ningún carácter de puntuación
 - No pueden empezar por *xml* o *XML*
 - No pueden contener espacios

http://www.editgrid.com/user/michkhk/Card_Status.xml

¿Atributos o elementos?

- Los atributos son más naturales para expresar metainformación
- La utilización de atributos hace más sencilla la estructura del documento
- Pero ...
 - Los atributos no pueden incluir valores múltiples
 - No pueden incluir estructura compleja
 - No son fáciles de extender

Documentos Bien Formados y Válidos

- Los documentos bien formados (*well formed*) son aquellos que cumplen con las reglas de sintaxis de XML.
- Lo mínimo que se exige a un documento XML es que esté bien formado.
- Si además cumple otra serie de relaciones y restricciones referidas a un dialecto concreto XML (XHTML, SVG, OpenDocument, Docbook, etc.), se dice que el documento es válido.
- Las relaciones y restricciones que especifican un determinado dialecto XML se especifican en un documento externo.

Lenguajes de Definición de Documentos o esquemas

- Se utilizan para especificar los nombres de los elementos, los atributos que pueden tener, la estructura del documento, tipos de datos que pueden contener, etc.)
- No hay un solo tipo :-), los más usados son:
 - DTD** *Document Type Definition*. Heredado de SGML y con algunas limitaciones
 - XML Schema** Evolución de DTD específico para XML y descrito por el W3C
 - RELAX NG** Más intuitivo que XML Schema, desarrollado por OASIS y muy utilizado actualmente

Espacios de nombres (XML namespaces)

- Al ser XML un lenguaje extensible en el que cada uno puede definir su dialecto y poner los nombres de elementos y etiquetas que desee, puede haber coincidencias de nombres con diferente significado.
- Un espacio de nombres es una especificación de un dialecto en el que los nombres de los elementos y atributos son únicos y se especifica en el documento XML mediante una URI utilizando el atributo reservado **xmlns**, por ejemplo:

```
<html xmlns:xhtml="http://www.w3.org/1999/xhtml">
```