

INTRODUCCIÓN A XML

JOSÉ DOMINGO MUÑOZ

IES GONZALO NAZARENO

ENERO 2020



¿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, ...



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

- 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>
```



ESTRUCTURA (SINTAXIS) DE UN DOCUMENTO XML

- 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>
```

- Los elementos vacíos tienen que cerrarse siempre:

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



ESTRUCTURA (SINTAXIS) DE UN DOCUMENTO XML

- Las etiquetas distinguen mayúsculas de minúsculas
- Los elementos deben estar correctamente anidados
- Los documentos deben tener un elemento raíz
- 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



¿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



- 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.



- 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.)
- Hay varios tipos:
 - ▶ DTD: Document Type Definition.
 - ▶ XML Schema
 - ▶ RELAX NG



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">
```

