
Lenguaje de marcado XML

Ofimática Avanzada

Profesor: Víctor Fresno Fernández

Lenguaje de marcado XML

XML

“El Lenguaje Extensible de Marcas, abreviado XML, describe una clase de objetos de datos llamados XML y parcialmente describe el comportamiento de programas de computadores que pueden procesarlos ”

Lenguaje de marcado XML

La especificación 1.0 de XML proporciona:

- .. la propia especificación XML
- .. los estándares adoptados
 - Unicode e ISO/IEC 10646 para caracteres
 - Internet RFC 1766 para identificación de lenguajes
 - ISO 639 para códigos de nombres de lenguajes
 - ISO 3166 para códigos de nombres de países

Lenguaje de marcado XML

Tecnologías asociadas, estándares que complementan la definición de XML

- XSL
- XML Linking Language (Xlink, Xpath, Xpointer)
- XML Namespaces
- XML-Schemas

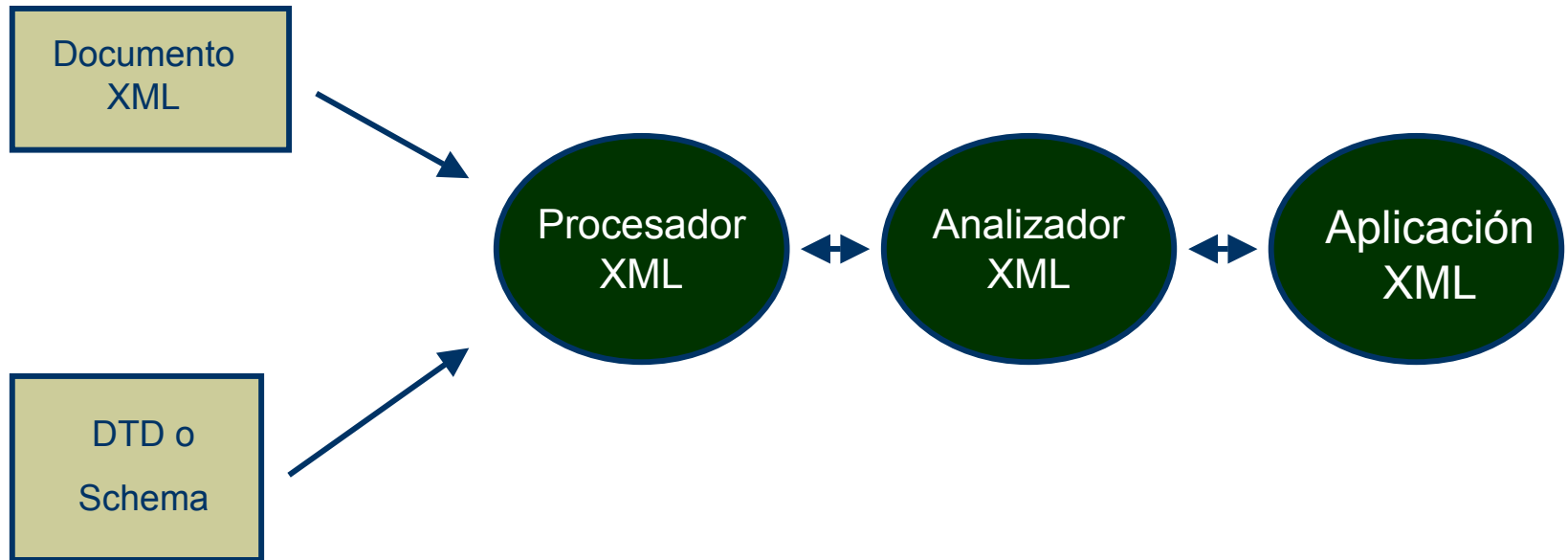
Lenguaje de marcado XML

Aplicaciones XML:

- Aplicaciones de búsqueda de información
- Procesos inteligentes
- Cambios de medio.. XML es un soporte neutro.. transformación a otros formatos
- Personalización
- Información reutilizable
- Automatización.. comunicaciones entre máquinas

Lenguaje de marcado XML

Esquema de una aplicación XML



Lenguaje de marcado XML

Procesador XML: módulo de software que proporciona acceso al contenido y estructura de un documento XML

Analizador XML: es un componente que requiere el procesador para determinar:

- La estructura de un documento XML
- Si es válido o conforme a una DTD o XML-Schema

Aplicación XML: emplea un procesador XML para acceder al contenido y estructura de un documento XML

Lenguaje de marcado XML

- **Estructura física..** el bloque de construcción básica en el XML es la *entidad* y contiene *datos analizados* (las marcas) y *no analizados sintácticamente* (texto asociado a esas marcas). Los datos de caracteres representan el contenido de la información real y el marcado la estructura del mismo.
- **Estructura lógica..** la estructura lógica de un documento XML la forman: el *elemento* -bloque de construcción básico-, y las *relaciones que hay entre los elementos*.

Lenguaje de marcado XML

El marcado se utiliza para proporcionar una descripción de..

- .. la **estructura de almacenamiento** de un documento (entidades)
- .. la **estructura lógica** (elementos)

Los distintos componentes de los documentos XML que soporta la especificación XML 1.0 son:

- **Etiquetas** de elemento
- **Instrucciones de procesamiento**
- **Declaración de tipo de documentos**
- Referencias de **entidades**
- **Comentarios**
- **Secciones marcadas**

Lenguaje de marcado XML

Componentes de un documento XML

◆ Elementos

- Un elemento es un conjunto de datos del documento, delimitado por **etiquetas de comienzo y fin** de elemento
- Cada elemento representa un componente lógico del documento
- Un elemento puede contener otros elementos y normalmente contendrá texto, figuras, etc (*datos carácter*).

Lenguaje de marcado XML

Componentes de un documento XML

- ◆ **Etiquetas** (nombres de elementos):
 - Delimitan los elementos de un documento XML. Su sintaxis es:

`< Identificador-elemento >`

- Etiqueta de comienzo de elemento: `<id-elemento>`
- Etiqueta de fin de elemento: `</id-elemento>`

Lenguaje de marcado XML

Componentes de un documento XML

- ◆ **Etiquetas** (nombres de elementos)

- Las etiquetas pueden incluir uno o más **atributos**
- Un atributo hace explícita una propiedad del elemento
- Los atributos se expresan en la etiqueta de comienzo y su sintaxis es:

```
nombre_atributo="valor del atributo"
```

- Ejemplo: <s id="1">

Lenguaje de marcado XML

Componentes de un documento XML

- ◆ **Etiquetas** (nombres de elementos)

- En XML puede haber **elementos vacíos**

- En XML un elemento vacío consta de:

```
<identificador_elemento atrib="va".../>
```

- Ejemplo: `<imagen fichero="imagen.gif"/>`

Lenguaje de marcado XML

Componentes de un documento XML

◆ Comentarios

- No forman parte del texto del documento
- No serán procesados por el analizador y su sintaxis es:

```
<!-- comentario -->
```

- Ejemplo: `<!-- esto es un comentario -->`

Lenguaje de marcado XML

Componentes de un documento XML

♦ Comentarios

- La cadena "--" no puede aparecer dentro de un comentario
- Los comentarios pueden ir en cualquier parte de un documento XML excepto:
 - Dentro de las declaraciones
 - Dentro de las etiquetas
 - Dentro de otros comentarios

Lenguajes de marcado

Componentes de un documento XML

◆ Sección CDATA

- Se utilizan para expresar bloques de texto que contienen caracteres que de otra manera serían reconocidos como etiquetas (se suele utilizar para introducir código que podría ser analizado)
- El procesador de XML mostrará su contenido, pero las anotaciones que contenga no serán analizadas

- Sintaxis:

```
<![CDATA[ texto que no se quiere analizar]]>
```

Lenguajes de marcado

Componentes de un documento XML

◆ Sección CDATA

- Ejemplo:

```
<![CDATA[ <saludo> Hola </saludo> ]]>
```

- `<saludo>` y `</saludo>` serían reconocidos como caracteres del texto y no como etiquetas
- No se pueden anidar las secciones CDATA

Lenguajes de marcado

Componentes de un documento XML

◆ Entidades

- Una entidad permite asignar un nombre a un subconjunto de datos (texto) y utilizar ese nombre para referirnos a dicho subconjunto.
- Las entidades pueden aparecer en dos contextos:
 - el documento
 - la DTD

Lenguajes de marcado

Componentes de un documento XML

◆ Entidades

- Las entidades se pueden clasificar en dos categorías:
 - generales
 - parámetro
- La marca de comienzo de una referencia a una entidad es “&” y la del final “;” (referencias generales)

Lenguajes de marcado

Componentes de un documento XML

◆ Entidades

- XML especifica 5 entidades predefinidas:

<code>&amp;</code>	para el carácter “&”
<code>&lt;</code>	para el carácter “<”
<code>&gt;</code>	para el carácter “>”
<code>&apos;</code>	para el carácter “ ’ ”
<code>&quot;</code>	para el carácter “ “ ”

Lenguajes de marcado

Componentes de un documento XML

◆ Instrucciones de procesamiento

- Se utilizan para proporcionar información a la aplicación y su sintaxis es:

```
<? Instrucción procesamiento ?>
```

- Los analizadores XML no hacen nada con las instrucciones de procesamiento, las pasan a la aplicación
- Están prohibidas las que comiencen por XML (salvo la del prólogo)

Lenguajes de marcado

Componentes de un documento XML

◆ Prólogo

- Los documentos XML pueden empezar con un prólogo en el que se define:
 - Una declaración XML
 - Una declaración de tipo de documento

Lenguajes de marcado

Componentes de un documento XML

◆ Prólogo

- La **declaración XML** es una instrucción de procesamiento:

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

Donde se indica:

- la versión de XML que estamos utilizando (la versión 1.0)
- Información sobre el tipo de codificación de caracteres que estamos utilizando
- En el ejemplo es el código ASCII de 7 bits (subconjunto del código Unicode denominado UTF-8, los parsers lo manejan por defecto)

Lenguajes de marcado

Componentes de un documento XML

◆ Prólogo

- Otra posibilidad de **declaración XML**:

```
<?xml version="1.0" encoding="UTF-16"?>
```

Hay más codificaciones de caracteres además de UTF-8 y UTF-16 aunque estas dos son las únicas soportadas por todos los procesadores XML

Lenguajes de marcado

Componentes de un documento XML

- ◆ **Prólogo**

- **Codificación de caracteres**

- Referencias a caracteres (valor numérico)

- Referencia decimal:

- &#xNum; ũ (signo copyright)

- Referencia hexadecimal:

- © © (signo copyright)

Lenguajes de marcado

Componentes de un documento XML

◆ Prólogo

- En la **declaración del tipo de documento**, se asocia la DTD respecto a la cual el documento es conforme (o válido)
- La DTD puede estar en un fichero distinto del documento:

```
<!DOCTYPE documento SYSTEM "ejemplo.dtd"> o
```

```
<!DOCTYPE coche SYSTEM
```

```
"http://www.sitio.com/dtd/coche.dtd">
```

Lenguajes de marcado

Componentes de un documento XML

◆ Prólogo

- El prólogo quedaría:

```
<?xml version="1.0"?>
```

```
<!DOCTYPE coche SYSTEM
```

```
"http://www.sitio.com/dtd/coche.dtd">
```

Lenguajes de marcado

Componentes de un documento XML

◆ Prólogo

- La DTD también puede incorporarse en el propio documento XML:

```
<!DOCTYPE documento [  
  <!ELEMENT documento (p | imagen | ejemplo)*>  
  <!ELEMENT p (#PCDATA|destacar)*>  
  <!ELEMENT destacar (#PCDATA)>  
  <!ATTLIST destacar  
    importancia CDATA #REQUIRED>  
  <!ELEMENT imagen EMPTY>  
  <!ATTLIST imagen  
    fichero CDATA #REQUIRED>  
  <!ELEMENT ejemplo (#PCDATA)> ]>
```

Lenguajes de marcado

Componentes de un documento XML

◆ Prólogo

- Ambas partes del prólogo son opcionales, aunque en el caso de incluir ambas, la declaración XML tiene que ir antes que la DTD

Lenguajes de marcado

Modelado de datos

- ◆ **Documento XML:** medio estructurado para almacenar información
- ◆ **Esquema:**
 - modelo utilizado para describir la estructura de los documentos
 - permiten modelar clases de documentos (vocabularios)

Lenguajes de marcado

Modelado de datos

- ◆ Cualquier documento que cumpla la especificación XML es **bien formado** por defecto
- ◆ Un documento XML debe adherirse a un esquema para ser **válido**
- ◆ Sin un esquema un documento puede estar bien formado pero no será válido

Lenguajes de marcado

Modelado de datos

- ◆ Un esquema establece restricciones en la estructura del contenido del documento:
 - El modelo de contenido determina el orden y anidación de los elementos
 - Establece los tipos de datos del documento (las DTD no)
- ◆ Formas de modelar datos en XML: **DTD** y **XML-Schema**

Lenguajes de marcado

Modelado de datos

- ◆ Diferencias principales entre las **DTD** y los **XML-Schema**:
 - DTD tienen una sintaxis específica mientras que Schema utiliza sintaxis XML
 - Un Schema se puede manipular como cualquier otro documento XML
 - Hay muchas más herramientas para trabajar con DTD que con Schema

Lenguajes de marcado

Modelado de datos

- ◆ Diferencias principales entre las **DTD** y los **XML-Schema**:
 - Un Schema soporta tipos de datos (int, float, boolean, date, ...) las DTD tratan todos los datos como cadenas
 - Un Schema soporta la integración de los espacios de nombres permitiendo asociar nodos de un documento con declaraciones de tipo de un esquema
 - La DTD solo permite una asociación entre un documento y su DTD

