

Tema V. XML

II. XML Bien Formado

Desarrollo de Aplicaciones para Internet
Curso 20|21

Índice

- 1.Introducción
- 2.Componentes XML
- 3.Reglas de los Elementos
- 4.Referencias

Índice

1.Introducción

2.Componentes XML

3.Reglas de los Elementos

4.Referencias

Introducción

- Los **documentos XML** deben seguir estrictamente una serie de **reglas**:
 - Se dice que un documento XML está **bien formado** cuando cumple todas las reglas de la recomendación del W3C.
 - Los documentos bien formados pueden ser procesados por cualquier *parser* XML.
- **Procesadores/parsers XML**
 - “A software module called an XML processor is used to read XML documents and provide access to their content and structure. It is assumed that an XML processor is doing its work on behalf of another module, called the application.” - **XML recommendation** (<http://www.w3.org/TR/REC-xml/#sec-intro>).

Índice

1.Introducción

2.Componentes XML

3.Reglas de los Elementos

4.Referencias

Componentes XML I

- **Etiquetas**

- Texto delimitado por `< y >`.
- Dos **formatos**:
 - **Inicio y fin**: `<nombre>Antonio</nombre>`
 - **Autocerrada**: `
`

- **Elementos**

- Conjunto formado por las **etiquetas de inicio y fin** más el contenido que delimitan.

- **Contenido del elemento**

- **Texto** delimitado por las etiquetas de inicio y de fin de un elemento.

Componentes XML II

- **Parsed Character Data (PCDATA)**

- **Contenido textual** de un elemento.
- También llamado “contenido textual” o “nodo de texto”.

- **Atributos**

- Parejas de **valor/nombre** asociadas a un elemento.
- Se incluyen en la etiqueta de inicio.
- Se admite comillas simples o dobles para delimitar el valor.
- No se admite atributos duplicados (es decir, un atributo no puede aparecer dos o más veces en una etiqueta).
- Ejemplo:

```
<persona id="12345678R">...</persona>
```

Componentes XML III

- **Comentarios**

- Información que **no forma parte del contenido ni estructura** de un documento XML.
- Tienen la misma función que los comentarios de programación.
- Se delimitan por `<!-- y -->`.
- No pueden escribirse dentro de una etiqueta:

```
<nombre <!-- Comentario inválido -->> ...  
</nombre>
```
- No pueden contener `--`:
 - `<!-- Comentario -- inválido -->`

Componentes XML IV

- **Declaraciones**

- Identifican un documento XML como tal y permiten proporcionar cierta información a los *parsers*.

- Ejemplo

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

- **Atributos** (debe respetarse el orden)

- **Versión** (obligatorio): indica la versión de la recomendación utilizada.
- **Encoding** (opcional): indica la codificación de caracteres utilizada.
- **Standalone** (opcional): si es *yes* entonces el documento no depende de ninguna especificación. Si es *no*, el documento depende de un DTD externo.

Componentes XML V

- **Caracteres de escape**

- Algunos caracteres tales como < o & no pueden utilizarse directamente en el contenido de un elemento.

- Para poder utilizarlas debemos usar entidades (*entity references*):

http://en.wikipedia.org/wiki/List_of_XML_and_HTML_character_entity_references

- **Secciones CDATA**

- Permiten definir bloques de texto que no deben ser procesados por el *parser*.
- Evitan tener que hacer uso de las entidades.
- Se delimitan con <![CDATA[y]]>.

Componentes XML VI

- **Instrucciones de procesado**

- Añaden información que no forma parte del contenido del documento, pero que se le pasará a la aplicación que lo procese.
- Se delimitan con `<? y ?>`, donde la apertura debe ir seguida del nombre de la instrucción.
- Su uso es poco frecuente.
- El texto que sigue a la etiqueta de apertura es el **objetivo** de la instrucción de procesado.
- El texto que sigue al objetivo es el **valor** de la instrucción.
- Ejemplo:

```
<?nameprocessor PRINT nombre?>
```

Ejemplo XML

```
<?xml version="1.0" encoding="UTF-8"?>
<persona dni="12345678-X">
  <nombre>Pedro</nombre>
  <apellido1>Pérez</apellido1>
  <apellido2>García</apellido2>

  <?procesadornombres PRINT nombre?>

  <edad>25</edad><!-- El 01/03/2012 -->

  <empresa>Pérez & Pérez, S.L.</empresa>
  <descripción><![CDATA[
    Pedro es un cliente de nuestra empresa con el que hemos
    trabajado desde antes de que fundase Pérez & Pérez, S.L.
  ]]></descripción>
</persona>
```

Índice

1.Introducción

2.Componentes XML

3.Reglas de los Elementos

4.Referencias

Reglas de los Elementos I

- **Cada etiqueta de inicio debe tener una etiqueta de fin**
 - Facilita el procesamiento y la detección de errores.
 - Una etiqueta “autocerrada” se considera de inicio y fin al mismo tiempo.
 - Ejemplo de **error**:

`<p>Este es un párrafo estilo HTML`

`
`

`Aquí finaliza el párrafo</p>`

Reglas de los Elementos II

- **Las etiquetas no pueden solaparse**
 - XML es estrictamente jerárquico.
 - Los elementos pueden contener etiquetas, pero deben anidarse correctamente.
 - Ejemplo de **error**:

<p>

Este ****es ****otro**** párrafo****
estilo HTML incorrecto en XML

</p>

Reglas de los Elementos III

- **Un documento solo puede tener una raíz**

- Debe existir un único **elemento raíz**.

- Ejemplo de documento **mal formado**:

- `<nombre>Pepe</nombre>`

- `<nombre>Juan</nombre>`

- Ejemplo de documento **bien formado**:

- `<nombres>`

- `<nombre>Pepe</nombre>`

- `<nombre>Juan</nombre>`

- `</nombres>`

Reglas de los Elementos IV

- **Los nombres de las etiquetas deben seguir las convenciones**
 - Pueden comenzar con letras o guión (–).
 - No pueden comenzar con números u otros signos de puntuación.
 - No pueden contener espacios.
 - Es recomendable que no contengan dos puntos (:).
 - No pueden comenzar con `xml` (en cualquier combinación de mayúsculas y minúsculas).
 - No puede haber un espacio entre `<` y el nombre de la etiqueta.

Reglas de los Elementos V

- **Sensible a mayúsculas y minúsculas**
 - El nombre de las etiquetas y atributos es sensible a mayúsculas y minúsculas.
 - Es importante tener esto en cuenta porque puede dar lugar a etiquetas no cerradas.
 - En el caso de nombres compuestos por varias palabras se recomienda el uso de *camel-case* o de un carácter que sustituya el espacio. Por ejemplo:

`<nombreCompleto>`

`<NombreCompleto>`

`<nombre_completo>`

`<nombre-completo>`

Reglas de los Elementos VI

- **Espacios en blanco en los PCDATA**
 - Espacios, tabuladores y saltos de línea se consideran espacios en blanco.
 - Al contrario que en HTML, los espacios en blanco se mantienen.
 - En el caso de los saltos de línea, solo se mantiene el carácter de salto de línea (el retorno de carro no).
 - Puede ocurrir que al visionar el XML en algún navegador se eliminen los espacios en blanco “sobrantes”.

Resumen de Reglas en los Elementos

- Cada etiqueta de inicio debe tener una etiqueta de fin o ser una etiqueta de autocierre.
- Las etiquetas no pueden solaparse.
- Los documentos solo pueden tener un nodo raíz.
- Los elementos deben respetar unas convenciones de nombrado.
- XML es sensible a mayúsculas y minúsculas.
- XML mantiene los espacios en blanco en los PCDATA.

Índice

1.Introducción

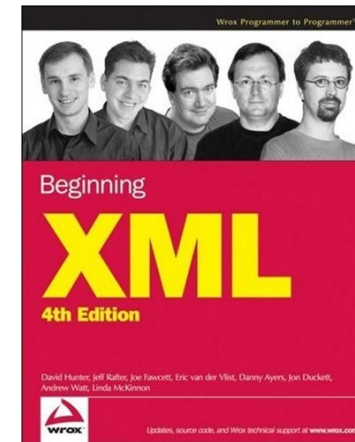
2.Componentes XML

3.Reglas de los Elementos

4.Referencias

Referencias

- **Beginning XML, 4th Edition**
– D. Hunter *et al.*



- **World Wide Web Consortium (W3C)**
[<http://www.w3.org/> - última visita 25/10/2020]