Elaboración de documentos XML ben formados

Sumario

- 1 Elaboración de documentos XML ben formados
- 2 Documentos XML válidos
  - 2.1 **RECORDA** Un XML coa sintaxe correcta é un documento "ben formado". Un XML validado contra un DTD é un XML "válido".
  - 2.2 Declaración do DTD
  - 2.3 Caracteres de frecuencia
  - 2.4 Or exclusivo
  - 2.5 Por que usar DTD
  - 2.6 Esquemas XML

Elaboración de documentos XML ben formados

Un documento XML **sintacticamente correcto** dise que está "ben formado". As regras de sintaxe xa se describiron:

- Os documentos XML deben ter un elemento raíz.
- As etiquetas hai que pechalas.
- Distínguese entre maiúsculas e minúsculas.
- As etiquetas deben estar correctamente aniñadas.
- Os atributos deben ir entrecomiñados.

Documentos XML válidos

Un documento XML válido é un documento "ben formado" que, ademais, se axusta ás regras dun DTD (*Document Type Definition*).



RECORDA

Un XML coa sintaxe correcta é un documento "ben formado".

Un XML validado contra un DTD é un XML "válido".

Declaración do DTD

O obxectivo dun DTD é definir a estrutura dun documento XML, é dicir, establecer restricións sobre os elementos do XML. O DTD pode especificarse **mediante un ficheiro externo** que contén as restricións sobre os datos. Por exemplo:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE nota SYSTEM "nota.dtd">
<nota>
<destinatario>Tove</destinatario>
<remitente>Jani</remitente>
<cabeceira>Recordatorio</cabeceira>
<corpo>Chámame!</corpo>
</nota>
```

A declaración DOCTYPE é unha referencia a un ficheiro DTD externo, chamado nota.dtd, xunto co nome do elemento raíz do documento. A ubicación do ficheiro especifícase mediante un URI e, por tanto, pode ser unha dirección web:

```
<!DOCTYPE notas SYSTEM "http://www.servidor.org/dtd/notas.dtd">
```

Independentemente da súa ubicación física o ficheiro conterá información como a seguinte:

```
<!ELEMENT nota (destinatario,remitente,cabeceira,corpo)>
<!ELEMENT destinatario (#PCDATA)>
<!ELEMENT remitente (#PCDATA)>
```

```
<!ELEMENT cabeceira (#PCDATA)>
<!ELEMENT corpo (#PCDATA)>
```

A información do DTD tamén **pode incluírse directamente no ficheiro XML**:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE nota
[
<!ELEMENT nota (destinatario,remitente,cabeceira,corpo)>
<!ELEMENT destinatario (#PCDATA)>
<!ELEMENT remitente (#PCDATA)>
<!ELEMENT cabeceira (#PCDATA)>
<!ELEMENT corpo (#PCDATA)>
]>
<nota>
<destinatario>Tove</destinatario>
<remitente>Jani</remitente>
<cabeceira>Recordatorio</cabeceira>
<corpo>Chámame!</corpo>
</nota>
```

En ambos os dous casos, o DTD anterior ten o seguinte significado:

- **!DOCTYPE nota** indica que o elemento raíz deste documento é nota.
- **!ELEMENT nota** indica que o elemento nota contén catro elementos: destinatario, remitente, cabeceira e corpo.
- **!ELEMENT destinatario** indica que o elemento destinatario é de tipo #PCDATA, é dicir, texto.
- **!ELEMENT remitente** indica que o elemento remitente é de tipo #PCDATA.
- **!ELEMENT cabeceira** indica que o elemento cabeceira é de tipo #PCDATA.
- **!ELEMENT corpo** indica que o elemento remitente é de tipo #PCDATA.

Caracteres de frecuencia

Na especificación de elementos de contido pódese incorporar un símbolo de frecuencia, de acordo coa seguinte táboa de significados:

Carácter	Descrición
Ningún	O elemento aparece só unha vez
+	O elemento aparece 1 ou máis veces
*	O elemento aparece 0 ou máis veces
?	O elemento aparece 0 ou 1 vez

Así, ao escribir, por exemplo:

```
<!ELEMENT aviso ( titulo?, (paragrafo+, grafico)*)>
```

Especifícase que aviso pode ter titulo ou non (pero só un), e ter cero ou máis conxuntos da forma (paragrafo, grafico), (paragrafo,paragrafo,grafico), etc.

Or exclusivo

Úsase o carácter (|) para separar as distintas opcións, co significado de "ou exclusivo". Por exemplo:

```
<!ELEMENT sobremesa ( xeado | pastel )>
```

Indica que sobremesa pode conter ben un elemento xeado ben un elemento pastel. O numero de opcións non está limitado a dúas, e pódense agrupar usando parénteses, da forma:

```
<!ELEMENT sobremesa ( sorbete, (xeado | pastel))>
```

Por que usar DTD

Resumindo, usamos DTD porque así:

- Cada ficheiro XML incorpora unha descrición do seu propio formato.
- Diferentes institucións poden poñerse de acordo e usar un DTD estándar para o intercambio de datos.
- Un programa pode usar un DTD estándar para verificar que os datos que recibe doutro programa externos son válidos.
- Nós mesmos podemos verificar a validez dos nosos propios datos.

Esquemas XML

O W3C apoia unha alternativa ao DTD chamada esquemas XML. É outra forma de validar documentos coa vantaxe de que os esquemas están baseados tamén en XML e permiten especificar **tipos de datos** para os elementos. Os DTD son anteriores aos esquemas XML e, previsiblemente, irán sendo substituídos polos esquemas pero, de momento, conviven as dúas tecnoloxías.

Un ficheiro do esquema XML para o documento que almacena notas podería chamarse notas.xsd e ter o seguinte aspecto:

```
<xs:element name="nota">

<xs:complexType>
  <xs:sequence>
    <xs:element name="destinatario" type="xs:string"/>
    <xs:element name="remitente" type="xs:string"/>
    <xs:element name="cabeceira" type="xs:string"/>
    <xs:element name="corpo" type="xs:string"/>
  </xs:sequence>
</xs:complexType>

</xs:element>
```

Os programas de axuda para a edición de documentos XML, como o XML Copy Editor, inclúen ferramentas para validalos, ben sexa a través de DTD ou de esquemas XML.

Os esquemas XML e os DTD estudaranse pormenorizadamente máis adiante neste curso.

--Arribi 12:31 29 sep 2010 (BST)