

XML: estrutura e sintaxe

Sumario

- 1 XML: estrutura e sintaxe
 - ♦ 1.1 **RECORDA** XML é unha linguaxe de marcas independente do sistema operativo deseñada para almacenar e transportar datos entre aplicacións. Por iso simplifica a compartición de datos entre programas e a realización de cambios nas configuracións dos sistemas. Permite tamén crear outras linguaxes a partir del, como o XHTML, polo que se di que é unha metalinguaxe.
 - ♦ 1.2 Estrutura
 - ♦ 1.3 **RECORDA** Os documentos XML forman unha **estrutura en árbore** que comeza na raíz e que se bifurca cara ás follas.
 - ◊ 1.3.1 Atributos
 - ♦ 1.4 Sintaxe

XML: estrutura e sintaxe

A linguaxe de marcas extensible (*eXtensible Markup Language* ou XML) deseñouse para transportar e almacenar datos. Aínda que ten parecidos co HTML as súas aplicacións son ben distintas. De feito, o HTML creouse simplemente para a visualización de datos a través da web e non para o seu tratamento automatizado. O XML céntrase no **que son os datos** (recorda que é unha linguaxe descritiva), mentres que o HTML encárgase de como se teñen que visualizar eses datos. Por tanto, XML non é un substituto do HTML, senón un complemento.



RECORDA

XML é unha linguaxe de marcas independente do sistema operativo deseñada para almacenar e transportar datos entre aplicacións. Por iso simplifica a compartición de datos entre programas e a realización de cambios nas configuracións dos sistemas. Permite tamén crear outras linguaxes a partir del, como o XHTML, polo que se di que é unha metalinguaxe.

En XML **non existen etiquetas predefinidas**, senón que é o usuario quen as crea. Por exemplo, unha nota de Xoán a Elvira pode almacenarse como un ficheiro XML:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<nota>
  <destinatario>Xoán</destinatario>
  <remitente>Elvira</remitente>
  <cabeceira>Recorda</cabeceira>
  <corpo>Chámame a fin de semana!</corpo>
</nota>
```

As etiquetas do exemplo anterior, como <destinatario> ou <remitente> non están definidas en ningún estándar XML. Son marcas inventadas polo autor do documento XML. Porén, os documentos HTML están construídos mediante etiquetas predefinidas no estándar, como <p>, <h1>, etc. O usuario define as etiquetas XML, así como a estrutura do documento.

Os ficheiros XML, en si mesmos, non fan nada, máis que describir os datos. Son ficheiros de texto plano editables con calquera editor. O exemplo anterior é bastante descritivo. Podemos ler que hai un destinatario e un remitente da nota, así como unha cabeceira e o corpo da mensaxe, pero non son máis que datos estruturados mediante etiquetas. Con todo, alguén podería escribir un programa para o envío de notas e usar este ficheiro XML para recibilas ou amosalas na pantalla do usuario.

Dende que o XML se convertera nunha recomendación do W3C, en febreiro de 1998, a súa popularidade foi medrando ata ser, actualmente, un dos piares da web. Asemade, o seu uso fóra das tecnoloxías web está moi estendido.

Estrutura

No exemplo anterior, a primeira liña é a declaración XML, na que se define a versión e a **codificación a empregar** (ISO-8859-1 = Latin-1/Conxunto de caracteres de Europa Occidental). A versión por defecto é a 1.0, aínda que dende o ano 2006 xa está dispoñible a 1.1.

A segunda liña describe o **elemento raíz** do documento (algo así como dicir "este documento é unha nota"):

```
<nota>
```

As seguintes 4 liñas describen 4 **elementos fillos** do elemento raíz (destinatario, remitente, cabeceira, corpo).

A última liña define onde remata o elemento raíz:

```
</nota>
```



RECORDA

Os documentos XML forman unha **estrutura en árbore** que comeza na raíz e que se bifurca cara ás follas.

O elemento raíz é o pai do resto dos elementos. Á súa vez, un elemento fillo pode ter outros subelementos (elementos fillos):

```
<raiz>
  <fillo>
    <subfillo>.....</subfillo>
  </fillo>
</raiz>
```

Os termos pai, fillo e irmán úsanse para describir as relacións entre elementos. Os elementos pais teñen fillos e os fillos do mesmo nivel son irmáns.

Atributos

Os elementos poden ter **atributos** (como en HTML). Un atributo proporciona información adicional sobre un elemento:

```
<materia numHoras="12">Redes de computadores</materia>
```

Un atributo tamén se pode representar como un elemento fillo:

```
<materia>
  <numHoras>12</numHoras>
  <nome>Redes de computadores</nome>
</materia>
```

Non existen regras sobre cando usar atributos ou elementos pero, en xeral, é preferible usar elementos no seu lugar xa que:

- Os atributos non poden conter múltiples valores (os elementos si)
- Os atributos non teñen estrutura de árbore (os elementos si)
- Os atributos son difíciles de manter (para cambios futuros no ficheiro XML)

Con todo, existe unha excepción na que si se aconsella o uso de atributos. Esta é cando se usan para representar metadatos. Por exemplo, en ocasións hai que asignar un identificador único (ID) a un elemento:

```
<facturacion>
  <factura id="501">
    <data>01/01/2010</data>
  </factura>
  <factura id="502">
    <data>01/01/2010</data>
  </factura>
</facturacion>
```

Os atributos id úsanse para identificar facturas pero non son parte da mesma. É dicir, os datos represéntanse mediante elementos e os datos sobre os propios datos adoitan representarse como atributos.

Sintaxe

Existen unha serie de regras sintácticas que se deben recordar á hora de escribir ficheiros en XML:

- **Todos os elementos XML deben ter unha etiqueta de peche.**

- **O XML distingue entre maiúsculas e minúsculas.** A etiqueta <Dispositivo> é diferente a <dispositivo>. Por tanto, as etiquetas de apertura e peche deben ser idénticas.

```
<Mensaxe>Incorrecto</mensaxe>
<mensaxe>Correcto</mensaxe>
```

- **Os elementos XML teñen que estar correctamente aniñados.** O seguinte exemplo é erróneo:

```
<can>
<raza>Boxer</can>
</raza>
```

- **Os documentos XML deben ter sempre un único elemento raíz.**
- **Os atributos XML deben ir entre comiñas, simples ou dobres.** O primeiro exemplo é incorrecto e o segundo correcto:

```
<nota data=12/11/2007>
  <destinatario>Tove</destinatario>
  <remitente>Jani</remitente>
</nota>

<nota data="12/11/2007">
  <destinatario>Tove</destinatario>
  <remitente>Jani</remitente>
</nota>
```

- **Caracteres especiais.** Existen caracteres que teñen un significado especial para o XML. Por exemplo, o carácter < pode confundirse cun principio de etiqueta. Para evitalo úsanse **entidades de referencia** (máis adiante verase como crealas). As entidades de referencia predefinidas son:

Carácter	Referencia
&	&
<	<
>	>
"	"
'	'

- **Comentarios.** A sintaxe para escribir un comentario dentro dun ficheiro XML é igual que en HTML