

Validación y definición de documentos

DTD: Document Type Definition

- Crear una definición del tipo de documento (DTD) es como crear nuestro propio lenguaje de marcado, para una aplicación específica.
- La DTD define los tipos de elementos, atributos y entidades permitidas, y puede expresar algunas limitaciones para combinarlos.
- La DTD puede residir en un fichero externo, y quizá compartido por varios (puede que miles) de documentos. O bien, puede estar contenida en el propio documento XML, como parte de su declaración de tipo de documento.

Los documentos XML que se ajustan a su DTD, se denominan "válidos". El concepto de "validez" no tiene nada que ver con el de estar "bien-formado". Un documento "bien-formado" simplemente respeta la estructura y sintaxis definidas por la especificación de XML. Un documento "bien-formado" puede además ser "válido" si cumple las reglas de una DTD determinada.

También existen documentos XML sin una DTD asociada, en ese caso no son "válidos", pero tampoco "inválidos"... simplemente "bien-formados"... o no.

DTD (ejemplo)

```
<!DOCTYPE etiqueta[
<!ELEMENT etiqueta (nombre, calle, ciudad, pais, codigo)>
<!ELEMENT nombre (#PCDATA)>
<!ELEMENT calle (#PCDATA)>
<!ELEMENT ciudad (#PCDATA)>
<!ELEMENT pais (#PCDATA)>
<!ELEMENT codigo (#PCDATA)>
]>
```

```
<etiqueta>
<nombre>Fulano Mengánez</nombre>
<calle>c/ Mayor, 27</calle>
<ciudad>Valderredible</ciudad>
<pais>España</pais>
<codigo>39343</codigo>
</etiqueta>
```

En el ejemplo anterior, todas las declaraciones DTD que definen "etiqueta" residen dentro del documento. Sin embargo, la DTD se puede definir parcial o completamente en otro lugar. Por ejemplo:

**AVILÉS**

```
<?xml version="1.0"?>
<!DOCTYPE coche SYSTEM "http://sitio.com/coche.dtd">
<coche>
<modelo>...</modelo>
...
</coche>
```

Declaraciones de tipo Elemento

Los elementos son la base de las marcas XML, y deben ajustarse a un tipo de documento declarado en una DTD para que el documento XML sea considerado válido.

Las declaraciones de tipo de elemento deben empezar con "<!ELEMENT" seguidas por el identificador genérico del elemento que se declara. A continuación tienen una especificación de contenido.

```
<!ELEMENT receta (titulo, ingredientes, procedimiento)>
```

Siguiendo la definición de elemento anterior, este ejemplo de documento XML sería válido:

```
<receta>
<titulo>...</titulo>
<ingredientes>...</ingredientes>
<procedimiento>...</procedimiento>
</receta>
```

Pero no este:

```
<receta>
<parrafo>Esto es un párrafo</parrafo>
<titulo>...</titulo>
<ingredientes>...</ingredientes>
<procedimiento>...</procedimiento>
</receta>
```

La especificación de contenido puede ser de cuatro tipos:

- **EMPTY**

Puede no tener contenido. Suele usarse para los atributos.

```
<!ELEMENT salto-de-pagina EMPTY>
```

- **ANY**

Puede tener cualquier contenido. No se suele utilizar, ya que es conveniente estructurar adecuadamente nuestros documentos XML.

```
<!ELEMENT batiburrillo ANY>
```

- **(#PCDATA)**

Contiene datos (caracteres).

```
<!ELEMENT enfasis (#PCDATA)>
```



AVILÉS

▪ Mixed

Puede tener caracteres de tipo datos o una mezcla de caracteres y sub-elementos especificados.

<!ELEMENT parrafo (#PCDATA|enfasis)*>

Para declarar que un tipo de elemento tenga contenido de elementos se especifica un modelo de contenido en lugar de una especificación de contenido.

Sólo puede contener sub-elementos especificados en la especificación de contenido.

<!ELEMENT mensaje (remite, destinatario, texto)>

Modelos de contenido

<!ELEMENT aviso (parrafo)>

Esto indica que <aviso> sólo puede contener un solo <parrafo>

<!ELEMENT aviso (titulo, parrafo)>

La coma, en este caso, denota una secuencia. Es decir, el elemento <aviso> debe contener un <titulo> seguido de un <parrafo>.

<!ELEMENT aviso (parrafo | grafico)>

La barra vertical "|" indica una opción. Es decir, <aviso> puede contener o bien un <parrafo> o bien un <grafico>. El número de opciones no está limitado a dos, y se pueden agrupar usando paréntesis.

<!ELEMENT aviso (titulo, (parrafo | grafico))>

En este último caso, el <aviso> debe contener un <titulo> seguido de un <parrafo> o de un <grafico>.

Además, cada partícula de contenido puede llevar un indicador de frecuencia, que siguen directamente a un identificador general, una secuencia o una opción, y no pueden ir precedidos por espacios en blanco.

Indicadores de frecuencia	
?	Opcional (0 o 1 vez)
*	Opcional y repetible (0 o más veces)
+	Necesario y repetible (1 o más veces)



AVILÉS

Ejemplo

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

En este caso, <aviso> puede tener <titulo>, o no (pero sólo uno), y puede tener cero o más conjuntos <parrafo><grafico>, <parrafo><parrafo><grafico>, etc.

