

## Qué es una DTD

Una DTD es un documento que define la estructura de un documento XML: los elementos, atributos, entidades, notaciones, etc, que pueden aparecer, el orden y el número de veces que pueden aparecer, cuáles pueden ser hijos de cuáles, etc. El procesador XML utiliza la DTD para verificar si un documento es válido, es decir, si el documento cumple las reglas del DTD.

## Referencia a una DTD en un documento XML

La DTD que debe utilizar el procesador XML para validar el documento XML se indica mediante la etiqueta DOCTYPE. La DTD puede estar incluida en el propio documento, ser un documento externo o combinarse ambas.

- La DTD puede incluirse en el propio documento, con la siguiente sintaxis:
  - `<!DOCTYPE nombre [`
  - `... declaraciones ...`
  - `>]`
- La DTD puede estar en un documento externo y, si sólo va a ser utilizada por una única aplicación, la sintaxis es la siguiente:

```
<!DOCTYPE nombre SYSTEM "uri">
```

Se puede combinar una DTD externa con una DTD interna, con la siguiente sintaxis:

```
<!DOCTYPE nombre SYSTEM "uri" [  
... declaraciones ...  
>]
```

- La DTD puede estar en un documento externo y, si va a ser utilizada por varias aplicaciones, la sintaxis es la siguiente:

```
<!DOCTYPE nombre PUBLIC "fpi" "uri">
```

Se puede combinar una DTD externa con una DTD interna, con la siguiente sintaxis:

```
<!DOCTYPE nombre PUBLIC "fpi" "uri" [  
... declaraciones ...  
>]
```

En todos estos casos:

- "nombre" es el nombre del tipo de documento XML, que debe coincidir con el nombre del elemento raíz del documento XML.
- "uri" es el camino (absoluto o relativo) hasta la DTD.
- "fpi" es un identificador público formal (Formal Public Identifier).

## Declaración de elementos

Las declaraciones de los elementos siguen la siguiente sintaxis:

```
<!ELEMENT nombreElemento (contenido)>
```

en la que "nombreElemento" es el nombre del elemento, y "(contenido)" una expresión que describe el contenido del elemento.

Para definir el contenido del elemento se pueden utilizar los términos EMPTY, (#PCDATA) o ANY o escribir expresiones más complejas:

- **EMPTY**: significa que el elemento es vacío, es decir, que no puede tener contenido. Los elementos vacíos pueden escribirse con etiquetas de apertura y cierre sin nada entre ellos, ni siquiera espacios, o con una etiqueta vacía. EMPTY debe escribirse sin paréntesis.

```
<!DOCTYPE ejemplo [  
  <!ELEMENT ejemplo EMPTY>  
>
```

```
<ejemplo></ejemplo>  
  
<ejemplo />  
  
<ejemplo>Esto es un ejemplo</ejemplo>  
  
<ejemplo><a></a></ejemplo>
```

- **(#PCDATA)**: significa que el elemento puede contener texto. #PCDATA debe escribirse entre paréntesis.

```
<!DOCTYPE ejemplo [  
  <!ELEMENT ejemplo (#PCDATA)>  
>  
  
<ejemplo />  
  
<ejemplo>Esto es un ejemplo</ejemplo>  
  
<ejemplo><a></a></ejemplo>
```

- **ANY:** significa que el elemento puede contener cualquier cosa (texto y otros elementos). ANY debe escribirse sin paréntesis.

```
<!DOCTYPE ejemplo [
  <!ELEMENT ejemplo ANY>
  <!ELEMENT a ANY>
]>

<ejemplo />

<ejemplo>Esto es un ejemplo</ejemplo>

<ejemplo>Esto es <a>un ejemplo</a></ejemplo>
```

Para indicar que un elemento puede o debe contener otros elementos se deben indicar los elementos, utilizando los conectores y modificadores siguientes:

- **, (coma):** significa que el elemento contiene los elementos en el orden indicado.

```
<!DOCTYPE ejemplo [
  <!ELEMENT ejemplo (a, b)>
  <!ELEMENT a EMPTY>
  <!ELEMENT b EMPTY>
]>

<ejemplo><a /><b /></ejemplo>

<ejemplo><a /><b /><c /></ejemplo>

<ejemplo><a /></ejemplo>

<ejemplo><b /><a /></ejemplo>
```

- **| (o lógico):** significa que el elemento contiene uno de los dos elementos.

```
<!DOCTYPE ejemplo [
  <!ELEMENT ejemplo (a | b)>
  <!ELEMENT a EMPTY>
  <!ELEMENT b EMPTY>
]>

<ejemplo><a /></ejemplo>

<ejemplo><b /></ejemplo>

<ejemplo><a /><b /></ejemplo>

<ejemplo></ejemplo>
```

- **?:** significa que el elemento puede aparecer o no, pero sólo una vez.

```
<!DOCTYPE ejemplo [  
  <!ELEMENT ejemplo (a, b?)>  
  <!ELEMENT a EMPTY>  
  <!ELEMENT b EMPTY>  
  
<ejemplo><a /></ejemplo>  
  
<ejemplo><a /><b /></ejemplo>  
  
<ejemplo><b /></ejemplo>  
  
<ejemplo><b /><b /></ejemplo>
```

- **\***: significa que el elemento puede no aparecer o aparecer una o más veces.

```
<!DOCTYPE ejemplo [  
  <!ELEMENT ejemplo (a*, b)>  
  <!ELEMENT a EMPTY>  
  <!ELEMENT b EMPTY>  
  
<ejemplo><b /></ejemplo>  
  
<ejemplo><a /><b /></ejemplo>  
  
<ejemplo><a /><a /><b /></ejemplo>  
  
<ejemplo><b /><a /></ejemplo>
```

- **+**: significa que el elemento tiene que aparecer una o más veces (no puede no aparecer).

```
<!DOCTYPE ejemplo [  
  <!ELEMENT ejemplo (a+, b)>  
  <!ELEMENT a EMPTY>  
  <!ELEMENT b EMPTY>  
  
<ejemplo><a /><b /></ejemplo>  
  
<ejemplo><a /><a /><b /></ejemplo>  
  
<ejemplo><b /></ejemplo>
```

- **()**: permite agrupar expresiones.

```
<!DOCTYPE ejemplo [  
  <!ELEMENT ejemplo (a, (a|b))>  
  <!ELEMENT a EMPTY>  
  <!ELEMENT b EMPTY>  
>
```

```
<ejemplo><a /><a /></ejemplo>  
<ejemplo><a /><b /></ejemplo>  
<ejemplo><a /></ejemplo>
```

```
<!DOCTYPE ejemplo [  
  <!ELEMENT ejemplo ((a, b)|(b, a))>  
  <!ELEMENT a EMPTY>  
  <!ELEMENT b EMPTY>  
>
```

```
<ejemplo><a /><b /></ejemplo>  
<ejemplo><b /><a /></ejemplo>
```

```
<ejemplo><a /><a /></ejemplo>
```