

DTD (Document Type Definition)

Declaración de Tipo de Documento

DTD (Document Type Definition)

- Conjunto de reglas estándar que deben cumplir un documento XML de un determinado tipo
- Pueden ser vistos como “plantillas” o como gramáticas
- Representados en archivos **.dtd**
- Determinan:
 - Qué elementos están permitidos en un tipo de documento
 - Relaciones entre los elementos
- Ejemplo: Un artículo de revista puede ser definido así:
 - Título
 - Autores
 - Resumen
 - Palabras clave
 - Nombre de la revista
 - Número de la revista
 - Página inicio-Página final
 - Fecha

Vincular DTDs a documentos XML

- El documento XML que se ajusta a su DTD, se denomina “**valido**” ≠ “**bien-formado**”
- Hay 3 formas de vincular un DTD con un documento XML:
 - Incluyendo el DTD al comienzo del XML con una etiqueta **!DOCTYPE**
 - *Ejemplo:*

```
<!DOCTYPE poema [  
  <!ELEMENT poema (titulo, verso+)>  
  . . .  
>  
<poema>. . . </poema>
```
 - Colocarlos en un archivo aparte utilizando la misma etiqueta **!DOCTYPE** pero con distinta sintaxis, y así se puede utilizar para validar múltiples documentos XML.
 - *Ejemplo:*

```
<!DOCTYPE factura SYSTEM "factura.dtd">
```
 - Algunas DTDs pueden tener identificadores públicos
 - *Ejemplo:*

```
<!DOCTYPE html  
PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"  
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
```

DTD: Tipos de declaraciones

- ELEMENT
 - Elementos del documento XML
- ATTLIST
 - Lista de atributos de un elemento
- ENTITY
 - Entidad (similares a las macros)
- NOTATION
 - Definen otros tipos de contenidos
 - Facilitan la inclusión de formatos binarios (imágenes, vídeos, sonidos, ...)

DTD: Declaración de Elementos

- Formato:

<!ELEMENT NombreElemento Regla>

- La regla describe el contenido que es almacenado en el elemento:
 - **PCDATA**: el contenido es solo texto, NO tiene anidación dentro de él.
Ej. **<!ELEMENT enfasis (#PCDATA)>**
 - **ANY**: Puede tener cualquier contenido. Es mejor no usarla y estructurar adecuadamente los documentos
Ej. **<!ELEMENT batiburrillo ANY>**
 - **EMPTY**: puede ser vacío
Ej. **<!ELEMENT linea-horizontal EMPTY>**
 - **MIXED**: Puede tener caracteres o una mezcla de caracteres y sub-elementos especificados
Ej. **<!ELEMENT parrafo (#PCDATA|enfasis)*>**
 - **Element**: Especifica uno o más sub-elementos contenidos en el elemento que se está definiendo
Ej. **<!ELEMENT html (head, body)>**

Ejemplo DTD

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

Documento XML válido

```
<receta>  
  <titulo>Arroz cocido</titulo>  
  <ingredientes>Arroz</ingredientes>  
  <procedimiento>Cocer el arroz</procedimiento>  
</receta>
```

Documento XML no válido

```
<receta>  
  <párrafo>La siguiente receta me la pasó Álvaro</párrafo>  
  <titulo>Arroz cocido</titulo>  
  <ingredientes>Arroz</ingredientes>  
  <procedimiento>Cocer el arroz</procedimiento>  
</receta>
```



Error

DTD: Cardinalidad de Elementos

(?) = 0, 1 elemento

(*) = 0 ó más elementos

(+) = 1 ó más elementos

(|) = alternativa

(,) = secuencia

EMPTY = vacío

ANY = cualquier estructura de subelementos

#PCDATA = cadena de caracteres analizados

```
<!ELEMENT pizza (ingrediente*, inventor?)>  
<!ELEMENT servicio (domicilio | restaurante) >  
<!ELEMENT ingrediente EMPTY>  
<!ELEMENT inventor (#PCDATA)>
```

Ejemplo: libros.xml y libro.dtd

```
<?xml version="1.0">
<libros>
  <libro isbn="9788420633114">
    <titulo>El Aleph</titulo>
    <autor>Jorge Luis Borges</autor>
    <any>1946</any>
    <precio>10</precio>
  </libro>
</libros>
```

- El DTD correspondiente es:

```
<!ELEMENT libros (libro+)>
<!ELEMENT libro (titulo, autor+, any?, precio)>
<!ELEMENT titulo (#PCDATA)>
<!ELEMENT autor (#PCDATA)>
<!ELEMENT any (#PCDATA)>
<!ELEMENT precio (#PCDATA)>
<!ATTLIST libro isbn ID #REQUIRED>
```


DTD: Modelos de contenido

- **Indicador de frecuencia**

- Siguen directamente a un identificador general, una secuencia o una opción
- No pueden ir precedidos por espacios en blanco
- Ejemplo:

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

- **Agrupamientos** utilizando los paréntesis curvos:

Ejemplo:

`<!ELEMENT p (font+, (img, br?)*, a*, ul*, ol*)>`

el elemento p tiene 1 o más ocurrencias del elemento font, el elemento img seguido por 0 o 1 elemento br que halla pueden aparecer 0 o más veces, y así sucesivamente.

- Cuando hay elementos que pueden ser conjuntos **no ordenados**, se representan:

Ejemplo: `<!ELEMENT p (font | (img, br?) | a | ul | ol) +>`

DTD: Declaración de lista de atributos

- Todos los atributos son declarados como:

```
<!ATTLIST elemento nombre-del-atributo TYPE Palabra-clave>
```

- Donde palabra clave puede ser:
 - **#REQUIRED**: Es obligatorio especificar el atributo. No tiene valor por defecto.
 - **#IMPLIED**: Se puede omitir el atributo, sin que se adopte automáticamente un valor por defecto
 - **#FIXED**: le da un valor por defecto al atributo

DTD: Atributos

- Atributos**

#REQUIRED Obligatorio

#IMPLIED Opcional

#FIXED Constante

Tipos de datos

- CDATA** = Cadena de caracteres

- NMTOKEN** = Palabra (sin espacios)

- NMTOKENS** = Lista de palabras

- Enumeración separada por |

- ID** = Nombre único (sin duplicados)

- IDREF** = Su valor debe apuntar a un ID

```
<!ATTLIST pizza nombre CDATA #REQUIRED>
```

```
<!ATTLIST ingrediente nombre CDATA #REQUIRED  
calorías CDATA #IMPLIED>
```

```
<!ATTLIST precio moneda (euros|dólares) #REQUIRED  
valor CDATA #REQUIRED>
```

```
<!ATTLIST persona código ID #REQUIRED>
```

```
<!ATTLIST dueño código IDREF #REQUIRED>
```

```
<!ATTLIST impuesto tipo CDATA #FIXED "IVA">
```

```
<pizza nombre="4 estaciones" >  
  <ingrediente nombre="Jamón" />  
  <precio moneda="euros" valor="7" />  
</pizza>
```

```
<persona código="23" nombre="Juan" />  
<persona código="35" nombre="Pepe" />  
<persona código="37" nombre="Luis" />
```

```
<dueño código="35" />
```

```
<impuesto tipo="IVA" />
```

DTD: Ejemplo

```
<!ELEMENT mensaje (de, a, texto)>  
<!ATTLIST mensaje prioridad (normal | urgente) normal>  
<!ELEMENT texto (#PCDATA)>  
<!ATTLIST texto idioma CDATA #REQUIRED>
```

```
<mensaje prioridad="urgente">  
  <de>Menganito</de>  
  <a>Juanita</a>  
  <texto idioma="español">  
    Lo que le tenga que contar Menganito a Juanita  
  </texto>  
</mensaje>
```

DTD: Tipos de atributos

- **CDATA** (character data): texto

```
<!ATTLIST mensaje fecha CDATA #REQUIRED>
```

```
< mensaje fecha="21 de Mayo de 2003">
```

- **Enumerados**: Sólo pueden contener un valor de entre un número reducido de opciones ("|")

```
<!ATTLIST mensaje prioridad (normal | urgente) normal>
```

- **ID y IDREF**

- El tipo **ID** permite que un atributo determinado tenga un nombre único que podrá ser referenciado por un atributo de otro elemento que sea de tipo **IDREF**
- Permite implementar un sistema de hipervínculos en un documento XML

```
<!ELEMENT enlace EMPTY>
```

```
<!ATTLIST enlace destino IDREF #REQUIRED>
```

```
<!ELEMENT capitulo (parrafo)*>
```

```
<!ATTLIST capitulo referencia ID #IMPLIED>
```

DTD: Tipos de atributos (II)

- **NMTOKEN Y NMTOKENS:**

- Los atributos NMTOKEN (NaMe TOKEN) sólo aceptan los caracteres válidos para nombrar cosas (letras, números, puntos, guiones, subrayados y los dos puntos)

```
<!ATTLIST mensaje fecha CDATA #REQUIRED>  
<mensaje fecha="15 de Diciembre de 1999">
```

```
<!ATTLIST mensaje fecha NMTOKEN #REQUIRED>  
<mensaje fecha="15-12-1999">
```

- **ENTITY** (permiten definir strings reusables) y **ENTITIES**

Proceso de validación documento XML

