

1.- Documentos XML.

Hasta ahora hemos trabajado con documentos básicos XML. Sólo declarábamos el tipo de documento que iba a ser (el ejemplar) pero no definíamos que cualidades tenía ese tipo de documento.

Ya vimos que un documento XML básico estaba formado por un **prólogo** y por un **ejemplar**.

Prólogo

Informa al intérprete encargado de procesar el documento de todos aquellos datos que necesita para realizar su trabajo. Consta de dos partes:

- **Definición de XML.** Indicamos la versión de XML, la codificación y la autonomía. Hasta ahora todos los generados han sido independientes.

<?xml version="1.0" encoding="utf-8" standalone="yes" ?>

- **Declaración del tipo de documento.** Hasta ahora sólo indicábamos
<!DOCTYPE nombre_del_ejemplar>

<!DOCTYPE biblioteca>

Ejemplar

Contiene los datos del documento que se quiere procesar. **Es el elemento raíz y debe ser único.**

Está compuesto de elementos estructurados según una estructura de árbol. Los elementos pueden contener atributos.

```
<biblioteca>
  <libro>
    <titulo edición="2" pagina="540">La búsqueda</titulo>
    <autor>Sebastián Torres</autor>
    <editorial>Ediciones Plum</editorial>
    <isbn>978-2-7460-4344-1</isbn>
  </libro>
</biblioteca>
```

Documentos bien formados.

Son sintácticamente correctos. Cumplen las reglas de sintaxis del lenguaje.

Documentos válidos.

Además de bien formados cumplen los requisitos de la definición de estructura que se haya indicado en la definición del documento.

Es decir que el ejemplar debe ceñirse a un formato y estructura definido previamente. **Necesitamos entonces especificar este formato de comunicación en XML.**

Lenguaje de marcas XML

Si cumple la sintaxis decimos **está bien formado**.

Para definir la estructura de datos podemos usar DTD o XML Schema

Si el documento XML cumple lo especificado en el DTD o XML Schema decimos que **es válido**.

2.- Declaración de tipos de documentos XML.

El DTD (*Document Type Definition*) establece que elementos son aceptados y en qué posición deben estar dentro de un documento XML.

La declaración de una DTD puede establecerse de dos formas.

1. **Interna.** La DTD se ubica dentro del propio documento XML.
2. **Externa.** La DTD se ubica en un archivo externo al documento XML.

Para validar más de un documento XML con la misma DTD debemos usar la *externa* para no tener que repetir la DTD dentro de cada documento XML. En el caso de que la DTD solo se utilice para validar un único documento XML, la DTD es habitual escribirla *internamente*.

Interna. Para referenciarlo como DTD interno, el atributo *standalone* en la declaración XML se debe marcar con un **Sí**. Esto significa que la declaración funciona al margen de la fuente externa.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
```

```
<!DOCTYPE autor[
```

```
  <!ELEMENT autor (nombre, empresa, telefono)>
```

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

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

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

```
<!--DOCTYPE root-element [element-declarations]-->
```



```
<nombre>Alejandro Romero</nombre>
```

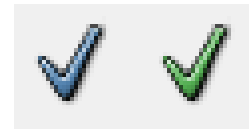
```
<empresa>Software Alcuza</empresa>
```

```
<telefono>954119023</telefono>
```

01interna.xml

Comprobación en XML Copy Editor

Bien formado



Válido

Externa

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
```

```
<!DOCTYPE autor SYSTEM "02autor.dtd">
```

```
<autor>
```

```
  <nombre>Alejandro Romero</nombre>
```

```
  <empresa>Software Alcuza</empresa>
```

```
  <telefono>954119023</telefono>
```

```
</autor>
```

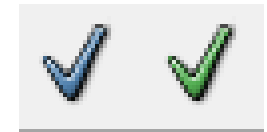
<!DOCTYPE root-element SYSTEM "file-name">

```
<!ELEMENT autor (nombre, empresa, telefono)>
```

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

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

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



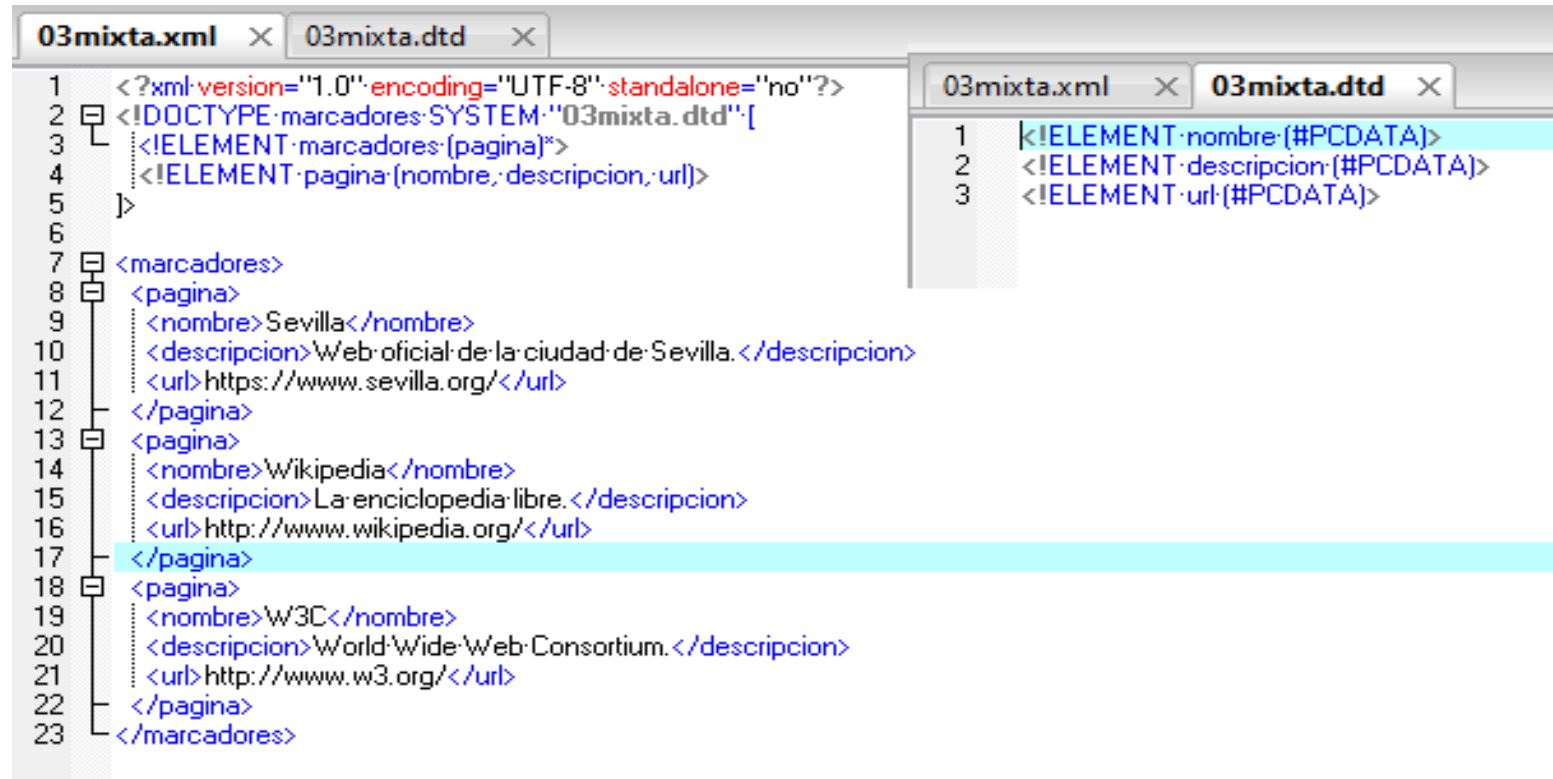
02externa.xml

02autor.dtd

.dtd no es un fichero XML

Combinando Interna / Externa

Existe la posibilidad de combinar ambas propuestas. Se concatenarían las líneas de código. Si una regla aparece en ambos lugares, las internas tienen prioridad sobre las externa.



```
03mixta.xml x 03mixta.dtd x
1 <?xml:version="1.0" encoding="UTF-8" standalone="no"?>
2 <!DOCTYPE marcadores SYSTEM "03mixta.dtd"[
3   <!ELEMENT marcadores (pagina)*>
4   <!ELEMENT pagina (nombre, descripcion, url)>
5 ]>
6
7 <marcadores>
8   <pagina>
9     <nombre>Sevilla</nombre>
10    <descripcion>Web oficial de la ciudad de Sevilla.</descripcion>
11    <url>https://www.sevilla.org/</url>
12  </pagina>
13  <pagina>
14    <nombre>Wikipedia</nombre>
15    <descripcion>La enciclopedia libre.</descripcion>
16    <url>http://www.wikipedia.org/</url>
17  </pagina>
18  <pagina>
19    <nombre>W3C</nombre>
20    <descripcion>World Wide Web Consortium.</descripcion>
21    <url>http://www.w3.org/</url>
22  </pagina>
23 </marcadores>

03mixta.xml x 03mixta.dtd x
1 <!ELEMENT nombre (#PCDATA)>
2 <!ELEMENT descripcion (#PCDATA)>
3 <!ELEMENT url (#PCDATA)>
```

03mixta.xml

03mixta.dtd

Ejemplo. Documento XML que guarda *sms* (cero o más de uno)

04EjemploSMS.xml

04BSsms.dtd

<!ELEMENT BDsms (sms)>*

Ejemplo. Creación de un DTD y documento XML que recoja información (nombre y dirección) de diferentes alumnos de una clase.

05Ejemplo.xml

05alumno.dtd

Los bloques de un documento DTD:

- **Entidades.** Son caracteres especiales que permiten incluir datos que podrían confundirse con entidades propias del lenguaje de marcado. También puede usarse para predefinir valores (constantes). **!ENTITY**
- **Elemento.** Es el bloque principal con el que se desarrollan los documentos. **!ELEMENT**
- **Atributos.** Es la manera de añadir más información a un elemento. **!ATTLIST**

Entidades (*!ENTITY*)

Espacio en blanco	
>	>
<	<
“	"
‘	'
&	&

Para caracteres
especiales

06entidadesUTF.xml

06entidadesISO.xml

Entidades generales internas analizables

<!ENTITY pi “3,141592”>	π
<!ENTITY alf “Alien Life Form”>	&alf;

Para declarar constantes

Una **sección CDATA** es una etiqueta que comienza por <![CDATA[y termina por]]> y cuyo contenido **el procesador XML no interpreta como marcas sino como texto**. Es una alternativa en los textos con muchas entidades como < y & que dificultan la lectura del documento

Entidades (***!ENTITY***)

Uso de una entidad dentro de otra. En la DTD se han declarado dos entidades generales internas analizables (color y frase). La primera aparece como parte del valor de la segunda.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<!DOCTYPE frase [
  <!ELEMENT frase (#PCDATA)>
  <!ENTITY color "azul">
  <!ENTITY frase "El cielo es &color;.">
]>
<frase>&frase;
</frase>
```

06entidadesanidadas.xml

Entidades (**!ENTITY**)

Entidades paramétricas internas analizables

Nos permite dar nombres a partes de un DTD y hacer referencia a ellas a lo largo del mismo como una constante. Útiles cuando varios elementos del DTD comparten listas de atributos o especificaciones de contenidos. Se denotan por **%entidad**;

```
<!ENTITY % direccion "calle, ciudad, cp">
```

```
<!ELEMENT beca (alumno, ies)>
```

```
<!ELEMENT alumno (dni, %direccion;)>
```

```
<!ELEMENT ies (nombre, %direccion;)>
```

← Defino "dirección"

Uso "dirección"

06Entidadesparametricas.xml

06Entidadesparametricas.dtd