

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

1.	DOCUMENTO XML. ESTRUCTURA Y SINTAXIS	2
1.1.	DECLARACIÓN DE TIPO DE DOCUMENTO	2
1.2.	DEFINICIÓN DE LA SINTAXIS DE DOCUMENTOS XML.....	3
2.	DEFINICIONES DE TIPO DE DOCUMENTO, DTD.	5
2.1.	DECLARACIÓN DE LA DTD	6
2.2.	DECLARACIÓN DE TIPOS DE ELEMENTOS TERMINALES	7
2.3.	DECLARACIONES DE TIPOS DE ELEMENTOS NO TERMINALES	10
2.4.	DECLARACIONES DE LISTAS DE ATRIBUTOS PARA LOS TIPOS DE ELEMENTOS	11
2.5.	DECLARACIÓN DE ENTIDADES	14
2.6.	DECLARACIÓN DE NOTACIÓN	16
2.7.	SECCIONES CONDICIONALES	17
3.	XML SCHEMA.....	18
3.1.	TIPOS DE ELEMENTOS EN XML SCHEMA	20
3.2.	ATRIBUTO EN XML SCHEMA	23
3.3.	TIPOS DE DATOS	25
3.4.	FACETAS DE LOS TIPOS DE DATOS	28
3.5.	DEFINICIÓN DE TIPOS DE DATOS SIMPLES XML SCHEMA	30
3.6.	DEFINICIÓN DE TIPOS DE DATOS COMPLEJOS XML SCHEMA	32
3.7.	ASOCIACIÓN DE DOCUMENTOS XML	35
3.8.	DOCUMENTACIÓN DEL ESQUEMA	37
4.	HERRAMIENTAS DE CREACIÓN Y VALIDACIÓN	38
	ANEXO 1. DOCUMENTACIÓN DE APOYO.....	39

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

1. DOCUMENTO XML ESTRUCTURA Y SINTAXIS

Hasta ahora hemos trabajado con documentos básicos de XML. Esto significa que dichos documentos están incompletos ya que solo hemos declarado el tipo de documento que va a ser, es decir que ejemplar vamos a definir, pero no hemos definido qué cualidades tiene ese tipo.

En la primera unidad vimos que un documento XML básico estaba formado por un prólogo y un ejemplar.

Recordamos que cada una de esas partes tiene el siguiente cometido:

- **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:** Donde se indica la versión de XML que se utiliza, el código de los datos a procesar y la autonomía del documento. Este último dato hasta ahora siempre ha sido "yes" ya que los documentos generados eran independientes.
 - **Declaración del tipo de documento:** Hasta el momento solo hemos dicho que es el nombre del ejemplar precedido de la cadena "<!DOCTYPE" y separado de ésta por, al menos un espacio y seguido de ">".

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

- **Ejemplar:** Contiene los datos del documento que se quiere procesar. **Es el elemento raíz del documento y ha de ser único.** Está compuesto de elementos estructurados según una estructura de árbol en la que el elemento raíz es el ejemplar y las hojas los elementos terminales, es decir, aquellos que no contienen más elementos. Los elementos pueden estar a su vez formados por atributos.

1.1. DECLARACIÓN DE TIPO DE DOCUMENTO

Ya habíamos visto que permite **al autor definir restricciones y características en el documento**, aunque no habíamos profundizado en las partes que la forman:

- **La declaración del tipo de documento propiamente dicha.** Comienza con el texto que indica el nombre del tipo, precedido por la cadena "<!DOCTYPE" separado del nombre del tipo por, al menos, un espacio. El nombre del tipo ha de ser idéntico al del ejemplar del documento XML en el que se está trabajando.
- **La definición del tipo de documento.** Permite asociar al documento una definición de tipo DTD, la cual se encarga de definir las cualidades del tipo. Es decir, define los tipos de los elementos, atributos y notaciones que se pueden utilizar en el documento así como las restricciones del documento, valores por defecto, etc. Para formalizar todo esto, XML está provisto de ciertas estructuras llamadas **declaraciones de marcado**, las cuales pueden ser internas o externas. Normalmente un documento XML se compone de una mezcla de declaraciones de marcado internas y externas. En este último caso debe expresarse en el documento dónde encontrar las declaraciones, así como indicar en la declaración de XML que el documento no es autónomo. Las diferencias entre estos tipos de declaraciones de marcado dan lugar a dos subconjuntos, el interno y el externo. Conviene saber que

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

primero se procesa el subconjunto interno y después el externo, lo que permite sobrescribir declaraciones externas compartidas entre varios documentos y ajustar el DTD a un documento específico.

- **Subconjunto interno:** Contiene las **declaraciones que pertenecen exclusivamente a un documento** y no es posible compartirlas. Se localizan dentro de unos corchetes que siguen a la declaración de tipo del documento.
- **Subconjunto externo:** Están localizadas en un documento con extensión **dtd** que puede situarse en el mismo directorio que el documento XML. Habitualmente son declaraciones que pueden ser compartidas entre múltiples documentos XML que pertenecen al mismo tipo. En este caso la declaración de documento autónomo ha de ser negativa, ya que es necesario el fichero del subconjunto externo para la correcta interpretación del documento. Con ello el procesamiento del documento será más lento, ya que antes de procesar el documento el procesador ha de obtener todas las entidades. Ahora los corchetes pierden sentido, para localizar las declaraciones del tipo de documento externo mediante una declaración explícita de subconjunto externo se utiliza:

```
<!DOCTYPE nombre_ejemplar SYSTEM "URI">
```

En este caso, se especifica un URI donde pueden localizarse las declaraciones.

```
<!DOCTYPE nombre_ejemplar PUBLIC "id_publico" "URI">
```

En este caso también se especifica un identificador, que puede ser utilizado por el procesador XML para intentar generar un URI alternativo, posiblemente basado en alguna tabla. **Como se puede observar también es necesario incluir algún URI.**

1.2. DEFINICIÓN DE LA SINTAXIS DE DOCUMENTOS XML

Recordamos que en estos documentos las etiquetas de marcado describen la estructura del documento.

Un elemento es un grupo formado por una etiqueta de apertura, otra de cierre y el contenido que hay entre ambas.

En los documentos de lenguajes de marcas, la distribución de los elementos está jerarquizada según una estructura de árbol, lo que implica que es posible anidarlos pero no entrelazarlos.

Hemos visto que en los elementos el orden es importante, ¿lo es también para los atributos? En este caso el orden no es significativo. Lo que hay que tener presente es que no puede haber dos atributos con el mismo nombre en el mismo elemento.

Sabemos que los atributos no pueden tener nodos que dependan de ellos, por tanto solo pueden corresponder con hojas de la estructura de árbol que jerarquiza los datos. ¿Significa esto que todas las hojas van a ser atributos? Pues no, es cierto que los atributos son hojas, pero las hojas pueden ser atributos o elementos.

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

En ese caso, ¿qué criterios podemos utilizar para decidir si un dato del documento que se pretende estructurar ha de representarse mediante un elemento o un atributo? Aunque no siempre se respetan, podemos usar los siguientes criterios:

- **El dato será un elemento si cumple alguna de las siguientes condiciones:**
 - Contiene subestructuras.
 - Es de un tamaño considerable.
 - Su valor cambia frecuentemente.
 - Su valor va a ser mostrado a un usuario o aplicación.
- **Los casos en los que el dato será un atributo son:**
 - El dato es de pequeño tamaño y su valor raramente cambia, aunque hay situaciones en las que este caso puede ser un elemento.
 - El dato solo puede tener unos cuantos valores fijos.
 - El dato guía el procesamiento XML pero no se va a mostrar.

Los espacios de nombres, o namespaces, ¿qué nos permiten?

- Diferenciar entre los elementos y atributos de distintos vocabularios con diferentes significados que comparten nombre.
- Agrupar todos los elementos y atributos relacionados de una aplicación XML para que el software pueda reconocerlos con facilidad.

¿Cómo se declaran? **xmlns:"URI_namespace"**

```
<nombre
xmlns="https://educacionadistancia.juntadeandalucia.es/EspacioNombres">Ejemplo</
nombre>
```

¿Y si se usa un prefijo que nos informe sobre cuál es el vocabulario al que está asociada esa definición?

xmlns:prefijo="URI_namespace"

```
<EN:nombre
xmlns:EN="https://educacionadistancia.juntadeandalucia.es/EspacioNombres">Ejempl
o</EN:nombre>
```

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

2. DEFINICIONES DE TIPO DE DOCUMENTO, DTD.

Una **definición de tipo de documento o DTD** es una descripción de estructura y sintaxis de un documento XML o SGML. Su función básica es la descripción de la estructura de datos, para usar un diseño común y mantener la consistencia entre todos los documentos que utilicen la misma DTD. De esta forma, dichos documentos pueden ser validados, conocen la estructura de los elementos y la descripción de los datos que trae consigo cada documento.

Dos o más documentos XML que tengan la misma DTD, se construyen de forma similar, tienen el mismo tipo de etiquetas y en el lugar, orden y cantidad especificados en la DTD.

En temas anteriores se explicó cuándo un documento XML estaba bien formado o era correcto. Pero, igual que para construir una frase en castellano de forma correcta no solo hace falta escribirla sin faltas de ortografía, un documento XML bien formado no es necesariamente válido. Para que un documento XML sea válido debe primero estar bien formado y después seguir las especificaciones dictadas por la DTD.

Las DTD están formadas por una **relación precisa de qué elementos pueden aparecer en un documento y dónde, así como el contenido y los atributos del mismo. Garantizan que los datos del documento XML cumplen las restricciones que se les haya impuesto en el DTD**, ya que estas últimas permiten:

- Especificar la estructura del documento.
- Reflejar una restricción de integridad referencial mínima utilizando (ID e IDREF).
- Utilizar unos pequeños mecanismos de abstracción comparables a las macros, que son las entidades.
- Incluir documentos externos.

¿Cuáles son los inconvenientes de los DTD? Los principales son:

- Su sintaxis no es XML.
- No soportan espacios de nombres.
- No definen tipos para los datos. Solo hay un tipo de elementos terminales, que son los datos textuales.
- No permite las secuencias no ordenadas.
- No es posible formar claves a partir de varios atributos o elementos.
- Una vez que se define un DTD no es posible añadir nuevos vocabularios.

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML



Ejercicio Resuelto

Al siguiente documento XML tenemos que agregarle un DTD que lo valide.

Con "<!DOCTYPE película [" comienza la DTD y finaliza con "]>". El nombre que aparece a continuación debe ser la raíz del documento, es de decir, película, a su vez se indica que la etiqueta título está dentro de la etiqueta película.

```
<?xml version="1.0"?>
<película>
  <título>Titanic</título>
</película>
```

Mostrar retroalimentación

```
<?xml version="1.0"?>
<!DOCTYPE película [
  <!ELEMENT película (título)>
  <!ELEMENT título (#PCDATA)>
]>
<película>
  <título>Titanic</título>
</película>
```

2.1.DECLARACIÓN DE LA DTD

La DTD está compuesta por lo que aparece entre la etiqueta "<!DOCTYPE nombre [" y la etiqueta "]>". El nombre corresponde con la etiqueta raíz que debe tener el documento XML.

Existen dos formas de definir la DTD que describirá la estructura de un documento XML. Se puede incluir dentro del mismo documento, o incluir la información sobre su ubicación (enlace a un documento DTD).

- **DTD incrustada:** Es posible incluir la DTD en el mismo documento como se ha visto anteriormente. En el siguiente ejemplo está resaltada en negrita:

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE película [
  <!ELEMENT película (título)>
  <!ELEMENT título (#PCDATA)>
]>
<película>
  <título>Titanic</título>
</película>
```

Se puede proporcionar una ayuda extra al analizador de XML, si a través de las instrucciones de proceso presentes en el prólogo, se indica que el documento es independiente y que todo lo que necesita está contenido en el mismo. Para ello basta con añadir el atributo standalone="yes", como puede verse a continuación:

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

El valor por defecto del atributo standalone es no, es decir, en caso de omitir el atributo es como si se pusiera standalone="no".

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

- **DTD externa:** En el ejemplo anterior, la DTD estaba incrustada en el documento XML. Es posible separar ambos elementos, guardándolos en archivos diferentes. Así, se puede situar en un archivo la DTD, en este caso se llama cine.dtd

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?> <!-- Enlace  
a la DTD -->  
<!DOCTYPE pelicula SYSTEM "cine.dtd">  
<pelicula>  
  <titulo>Titanic</titulo>  
</pelicula>
```

No hace falta poner standalone="no" ya que por defecto toma ese valor.



Debes conocer

Las ventajas de incluir una DTD externa son las siguientes:

- ✓ Si la DTD que se va a incluir es compartida por muchos documentos XML, es preferible que se encuentre en un archivo independiente ya que si hay que hacer algún cambio en la DTD, solo tendrá que hacerse en un archivo y no en cada XML en el caso de tener una DTD incrustada en el mismo documento XML.
- ✓ La DTD puede ubicarse en un servidor web, de forma que cualquier persona con acceso a Internet puede validar el documento XML que está creando, lo que garantiza que todos los documentos XML usen la última versión de la DTD. Para declarar que nuestra DTD está en un servidor de Internet basta con especificarlo en el DOCTYPE de la siguiente forma:

```
<!DOCTYPE cine SYSTEM "http://cine.com/filmoteca.dtd">
```

Aunque es más correcto si se pone de forma pública. Se hace así:

```
<!DOCTYPE cine PUBLIC "filmoteca" "http://cine.com/filmoteca.dtd">
```

Siendo filmoteca el nombre de la DTD.

2.2.DECLARACIÓN DE TIPOS DE ELEMENTOS TERMINALES

Los **tipos de elementos terminales** son aquellos elementos que se corresponden con hojas de la estructura de árbol formada por los datos del documento XML asociado al DTD. La declaración de tipos de elementos está formada por la cadena "**<!ELEMENT**" separada por, al menos un espacio del nombre del elemento XML que se declara, y seguido de la declaración del contenido que puede tener dicho elemento.

En el caso de elementos terminales, es decir, aquellos que no contienen más elementos, esta declaración de contenido es dada por uno de los siguientes valores:

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

- **EMPTY:** Indica que el elemento no es contenedor, es vacío, es decir, que no puede tener contenido. Por ejemplo, la siguiente definición muestra un elemento "ejemplo" que no contiene nada: `<!ELEMENT ejemplo EMPTY>`.

- XML asociado **correcto:**

```
<ejemplo></ejemplo> ó <ejemplo />
```

- XML asociado **incorrecto:**

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

- **(#PCDATA):** Indica que los datos son analizados en busca de etiquetas, resultando que el elemento no puede contener elementos, es decir solo puede contener datos de tipo carácter exceptuando los siguientes: `<`, `[`, `&`, `]`, `>`. Si es de este tipo, el elemento "ejemplo" tendrá una definición como:

```
<!ELEMENT ejemplo (#PCDATA)>.
```

- XML asociado **correcto:**

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

- XML asociado **incorrecto:**

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

- **ANY:** Permite que el contenido del elemento sea cualquier cosa (texto y otros elementos). **No es recomendable usar este tipo de elemento.** Ejemplos de XML correcto con elemento ANY:

```
<?xml version="1.0" encoding="UTF-8"?>  
<!DOCTYPE mascota[  
<!ELEMENT mascota ANY>  
<!ELEMENT nombre (#PCDATA)>  
<!ELEMENT tipo(#PCDATA)>  
  <!ELEMENT raza(#PCDATA)>  
>]
```


TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

Para este vocabulario, estos documentos xml son válidos:

```
<mascota>
  <nombre>Coco</nombre> es mi mascota. Es una <tipo>chinchilla</tipo>
  <raza>blanca</raza>.
</mascota>
```

```
<mascota>
  Coco es mi mascota. Es una <tipo>chinchilla</tipo> <raza>blanca</raza>.
</mascota>
```

```
<mascota>
  <nombre>Coco</nombre>.
</mascota>
```

```
<mascota/>
```



Ejercicio resuelto

Creación de una DTD correspondiente a la siguiente estructura de datos de un documento XML:

```
<alumno>Olga Velarde Cobo</alumno>
```

Mostrar retroalimentación

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

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

2.3.DECLARACIONES DE TIPOS DE ELEMENTOS NO TERMINALES

Una vez que sabemos el modo de definir las hojas de un árbol de datos, **veamos cómo definir sus ramas**, es decir los elementos que están formados por otros elementos. Para definirlos utilizamos referencias a los grupos que los componen tal y como muestra el ejemplo:

`<!ELEMENT A (B, C)>`

En este caso se ha definido un elemento A que está formado por un elemento B seguido de un elemento C.

¿Y qué sucede cuando un elemento puede aparecer en el documento varias veces, hay que indicarlo de algún modo? Pues sí, también hay que indicar cuando un elemento puede no aparecer. Para ello usamos los siguientes operadores, que nos permiten definir la cardinalidad de un elemento:

- **Operador opción, ?**. Indica que el elemento no es obligatorio. En el siguiente ejemplo el subelemento trabajo es opcional:

`<!ELEMENT telefono (trabajo?, casa)>`

- **Operador uno-o-más, +**. Define un componente presente al menos una vez. En el ejemplo definimos un elemento formado por el nombre de una provincia y otro grupo, que puede aparecer una o varias veces:

`<!ELEMENT provincia (nombre, (cp, ciudad)+)>`

- **Operador cero-o-mas, ***. Define un componente presente cero, una o varias veces. En el ejemplo el grupo (cp, ciudad) puede no aparecer o hacerlo varias veces:

`<!ELEMENT provincia (nombre, (cp, ciudad)*)>`

- **Operador de elección, |**. Cuando se utiliza sustituyendo las comas en la declaración de grupos indica que para formar el documento XML hay que elegir entre los elementos separados por este operador.

- En el ejemplo siguiente, el documento XML tendrá elementos provincia que estarán formados por el elemento nombre y el cp (código postal), o por el elemento nombre y la ciudad:

`<!ELEMENT provincia (nombre, (cp | ciudad))>`

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML



Ejercicio resuelto

Creación de un DTD correspondiente a la siguiente estructura de datos de un documento XML:

```
<alumno>
  <nombre>Olga</nombre>
  <dirección>El Percebe 13</dirección>
</alumno>
```

Mostrar retroalimentación

```
<!ELEMENT alumno (nombre, direccion)>
<!ELEMENT nombre (#PCDATA)>
<!ELEMENT dirección (#PCDATA)>
```

2.4.DECLARACIONES DE LISTAS DE ATRIBUTOS PARA LOS TIPOS DE ELEMENTOS

Ya sabemos cómo declarar elementos, ahora veamos el modo de **declarar los atributos asociados a un elemento**.

Para ello utilizamos la cadena **<!ATTLIST** seguida del nombre del elemento asociado al atributo que se declara, luego el nombre de éste último seguido del tipo de atributo y del modificador.

```
<! ATTLIST nombre_elemento_del_atributo nombre_atributo tipo_del_atributo "Valor_por_defecto">
```

Este elemento puede usarse para declarar una lista de atributos asociada a un elemento, o repetirse el número de veces necesario para asociar a dicho elemento esa lista de atributos, pero individualmente. Si un elemento tiene más de un atributo se puede expresar:

```
<! ATTLIST nombre_elemento_del_atributo1 nombre_atributo1 tipo_del_atributo1 "Valor_por_defecto"
          nombre_atributo2 tipo_del_atributo2 "Valor_por_defecto"
          nombre_atributo3 tipo_del_atributo3 Valor_por_defecto"
          .....>
```

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

Al igual que los elementos no todos los atributos son del mismo tipo, los más destacados son:

- **Enumeración**, es decir, el atributo solo puede tomar uno de los valores determinados dentro de un paréntesis y separados por el operador |.

- DTD:

```
<!ATTLIST fecha dia_semana (lunes|martes|miércoles|jueves|viernes|sábado|domingo) #REQUIRED>
```

- **CDATA**, se utiliza cuando el atributo es una cadena de texto.

- DTD:

```
<!ATTLIST ejemplo color CDATA #REQUIRED>
```

- XML asociado:

```
<ejemplo color="verde" />
```

- **ID**, permite declarar el valor del atributo (no el nombre) debe ser único y no se puede repetir en otros elementos o atributos. Hay que tener en cuenta que los números no son nombres válidos en XML, por tanto no son un identificador legal de XML. Para resolverlo suele incluirse un prefijo en los valores y separarlo con un guion o una letra.

- DTD:

```
<!ATTLIST libro codigo ID #REQUIRED>
```

- XML asociado:

```
<libro codigo="Q1">El Quijote</libro>
```

- **IDREF**, permite hacer referencias a identificadores. En este caso el valor del atributo ha de corresponder con el de un identificador de un elemento existente en el documento.
- **NMTOKEN**, permite determinar que el valor de un atributo ha de ser una sola palabra compuesta por los caracteres permitidos por XML, es decir letras, números y los caracteres ":", "_", "-" o ".".
- ¿También hemos de declarar si el valor de un atributo es obligatorio o no? Si, para ello se usan los siguientes modificadores:

- **#IMPLIED**, determina que el atributo sobre el que se aplica es opcional.

- DTD:

```
<!ATTLIST ejemplo color CDATA #IMPLIED>
```

- XML asociado:

```
<ejemplo color="verde" /> o <ejemplo color="" />
```

- **#REQUIRED**, determina que el atributo tiene carácter obligatorio.

- DTD:

```
<!ATTLIST ejemplo color CDATA #REQUIRED>
```

- XML asociado NO válido:

```
<ejemplo color="" />
```

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

- XML asociado válido: `<ejemplo color="verde" />`

- **#FIXED**, permite definir un valor fijo para un atributo independientemente de que ese atributo se defina explícitamente en una instancia del elemento en el documento XML.

- DTD: `<!ATTLIST ejemplo color CDATA #FIXED "verde">`

- XML asociado NO válido: `<ejemplo color="rojo" />`

- XML asociado válido: `<ejemplo color="verde" />`

- **Literal**, asigna por defecto a un atributo el valor dado por una cadena entre comillas, pero puede tomar otro valor.

- DTD: `<!ATTLIST ejemplo color CDATA (rojo|verde|amarillo) "verde">`

- XML asociado: `<ejemplo color="verde" />`



Ejercicio resuelto

Creación de un DTD correspondiente a la siguiente estructura de datos de un documento XML:

```
<alumno edad=15>
  <nombre>Olga</nombre>
  <apellidos>Velarde Cobo</apellidos>
  <dirección>El Percebe 13</dirección>
</alumno>
```

Mostrar retroalimentación

```
<!ELEMENT alumno (nombre, apellidos, direccion)>
<!ATTLIST alumno edad CDATA #REQUIRED>
<!ELEMENT nombre (#PCDATA)>
<!ELEMENT apellidos (#PCDATA)>
<!ELEMENT dirección (#PCDATA)>
```

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

2.5.DECLARACIÓN DE ENTIDADES

¿Qué sucede si queremos **declarar valores constantes dentro de los documentos**? ¿Podemos?

Las entidades nos permiten definir constantes en un documento XML. Cuando se usan dentro del documento XML se limitan por "&" y ";", por ejemplo **& entidad**;

¿Cómo trabaja el intérprete con ellos? Al procesar el documento XML, el intérprete sustituye la entidad por el valor que se le ha asociado en el DTD.

No admiten recursividad, es decir, una entidad no puede hacer referencia a ella misma.

Las entidades básicamente pueden ser de dos tipos, **generales** y de **parámetro**.

- **Generales:**
 - **Internas:** Son las entidades declaradas en el DTD.
 - Existen cinco entidades predefinidas en el lenguaje, son:
 - **<;**: Se corresponde con el signo menor que, <.
 - **>;**: Hace referencia al signo mayor que, >.
 - **";**: Son las comillas rectas dobles, ".
 - **';**: Es el apóstrofe o comilla simple, '.
 - **&;**: Es el et o ampersand, &.
 - Podemos definir nuestras propias entidades usando la estructura:<!ENTITY nombre_entidad "valor_entidad">, donde:
 - nombre_entidad es el nombre que recibe la entidad
 - "valor_entidad" es el valor que toma dicha entidad
 - se hace referencia a la entidad usando **&nombre_entidad**;

Veamos un ejemplo. Dado el DTD:

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <!DOCTYPE libros [
3      <!ELEMENT libros (libro)+>
4      <!ELEMENT libro (#PCDATA)>
5
6      <!ENTITY autor "Cervantes">
7      <!ENTITY editorial "Alfaguara">
8  ]>
9
10 <libros>
11   <libro>Don Quijote de la Mancha fue escrito por &autor;</libro>
12   <libro>SIDI fue escrito por Arturo Pérez-Reverte y publicado por &editorial;</libro>
13   <libro>Tiempos recios fue escrito por Mario Vargas Llosa y publicado por &editorial;</libro>
14 </libros>
```

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

Al abrir el fichero en un navegador, se vería:

Este fichero XML no parece tener ninguna información de estilo asociada. Se muestra debajo el árbol del documento.

```
<libros>
  <libro>Don Quijote de la Mancha fue escrito por Cervantes</libro>
  <libro>
    SIDI fue escrito por Arturo Pérez-Reverte y publicado por Alfaguara
  </libro>
  <libro>
    Tiempos recios fue escrito por Mario Vargas Llosa y publicado por Alfaguara
  </libro>
</libros>
```

- **Externas:** Permiten establecer una relación entre el documento XML y otro documento a través de la URL de éste último.

- La declaración de una entidad externa cuando sólo va a ser utilizada por una única

aplicación es: `<!ENTITY nombre_entidad SYSTEM "http://localhost/docsxml/fichero_entidad.xml">`

En este caso el contenido de los ficheros es analizado, por lo que deben seguir la sintaxis XML.

Cuando es necesario incluir ficheros con formatos binarios, es decir ficheros que no se analicen, se utiliza la palabra reservada NDATA en la definición de la entidad y habrá que asociar a dicha entidad una declaración de notación, tal y como muestra el ejemplo del apartado siguiente.

- La declaración de una entidad externa cuando va a ser utilizada por varias aplicaciones es:

```
<!ENTITY nombre_entidad PUBLIC "identificador público formal" "camino hasta la DTD (uri)">
```

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

- De parámetro:

- **Internas:** Permiten dar nombres a partes de un DTD y hacer referencia a ellas a lo largo del mismo. Son especialmente útiles cuando varios elementos del DTD comparten listas de atributos o especificaciones de contenidos. Se denotan por % entidad;

```
<!ENTITY % direccion "calle, numero?, ciudad, cp"> y se puede usar como sigue:  
<!ELEMENT almacen (%direccion, web)>  
<!ELEMENT oficina (%direccion, movil)>  
<!ELEMENT central (%direccion, telefono)>  
<!ELEMENT tienda (%direccion, fax)>
```

- **Externas:** Permite incluir en un DTD elementos externos, lo que se aplica en dividir la definición DTD en varios documentos.

```
<!ENTITY %persona SYSTEM "persona.dtd">
```

2.6.DECLARACIÓN DE NOTACIÓN

Cuando se incluyen ficheros binarios en un fichero, ¿cómo le decimos qué aplicación ha de hacerse cargo de ellos? La respuesta es utilizando notaciones. La sintaxis para declarar **notaciones** es:

```
<!NOTATION nombre SYSTEM aplicacion>
```

Por ejemplo, una notación llamada **gif** donde se indica que se hace referencia a un editor de formatos gif para visualizar imágenes será:

```
<!NOTATION gif SYSTEM "gifEditor.exe"
```

Para asociar una entidad externa no analizada, a esta notación basta declarar dicha entidad del siguiente modo:

```
<!ENTITY dibujo SYSTEM "imagen.gif" NDATA gif>
```


TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

2.7. SECCIONES CONDICIONALES

Permiten incluir o ignorar partes de la declaración de un DTD. Para ello se usan dos tokens:

- **INCLUDE**, permite que se vea esa parte de la declaración del DTD. Su sintaxis es:

```
<![INCLUDE [Declaraciones visibles] ] >
```

Por ejemplo:

```
<![INCLUDE [ <!ELEMENT nombre (#PCDATA)> ] ] >
```

- **IGNORE**, permite ocultar esa sección de declaraciones dentro del DTD. La forma de uso es:

```
<![IGNORE [Declaraciones ocultas] ] >
```

Por ejemplo:

```
<![IGNORE [ <!ELEMENT clave (#PCDATA)> ] ] >
```

El uso de las secciones condicionales suele estar ligado a entidades paramétricas.



Ejercicio resuelto

Vamos a crear un dtd para guardar la información sobre un alumno y queremos tener la posibilidad de obtener el registro de dos maneras diferentes, una corta que incluya únicamente el nombre y la edad del alumno y la otra que incluya además los apellidos, la dirección donde vive y su DNI.

Creamos un archivo llamado "*alumno.dtd*" y escribimos:

```
<![ %datos_basicos; [  
  <!ELEMENT alumno(nombre, edad)>  
]]>  
  
<![ %datos_ampliados; [  
  <!ELEMENT alumno(nombre, apellidos, edad, direccion, DNI)>  
]]>  
  
<!ELEMENT nombre (#PCDATA)>  
<!ELEMENT apellidos (#PCDATA)>  
<!ELEMENT edad (#PCDATA)><!ELEMENT direccion (#PCDATA)>  
<!ELEMENT DNI (#PCDATA)>  
  
El siguiente documento XML sería válido:  
<?xml version="1.0" encoding="UTF-8" standalone="no"?>  
<!DOCTYPE alumno SYSTEM "alumno.dtd" [  
  <!ENTITY % datos_basicos "INCLUDE">  
  <!ENTITY % datos_ampliados "IGNORE">  
>
```

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

```
<alumno>
  <nombre>Silvia</nombre>
  <edad>13</edad>
</alumno>

También sería válido el documento:
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!DOCTYPE persona SYSTEM "alumno.dtd" [
  <!ENTITY % datos_basicos "IGNORE">
  <!ENTITY % datos_ampliados "INCLUDE">
]>

<alumno>
  <nombre>Ana</nombre>
  <apellidos>Gil Lechuga</apellidos>
  <edad>14</edad>
  <direccion>C/ Granada sn. Almería</direccion> <DNI>22222222-S</DNI>
</alumno>
```

Obsérvese que, en este ejemplo, en los dos documentos XML asociados a la DTD externa escrita en el archivo "**alumno.dtd**", se ha indicado –por medio de **IGNORE** e **INCLUDE** si el elemento "alumno" tiene que contener 2 ó 5 hijos, es decir, ("nombre" y "edad") o ("nombre", "apellidos", "edad", "direccion" y "DNI").

3. XML SCHEMA

Los **DTD** permiten diseñar un **vocabulario para ficheros XML**, pero, ¿qué sucede cuando los valores de los elementos y atributos de esos ficheros han de corresponder a datos de un tipo determinado, o cumplir determinadas restricciones que no pueden reflejarse en los DTD? Para ello se definen **XML Schemas**, que se componen de elementos y atributos, al igual que los DTD.

¿También **se definen en ficheros planos**? Si, ya que son documentos XML, pero en este caso la extensión de los archivos es **xsd**, motivo por el cual también se les denomina documentos **XSD**.

Los elementos XML que se utilizan para generar un esquema han de pertenecer al espacio de nombres XML Schema, que es: <http://www.w3.org/2001/XMLSchema>. En esta especificación se usa el prefijo **<xsd:schema>**, pero para abreviar se usa **<xs:schema>**, aunque en la práctica cualquier prefijo puede ser usado, siempre que se use el mismo prefijo en todo el documento.

Las estructuras que se definen en XML Schema definen a su vez numerosos atributos para uso directo en cualquier documento XML y están en un espacio de nombres diferente, <http://www.w3.org/2001/XMLSchema-instance>. En esta especificación se usa el prefijo **<xsi:schema>**, pero para abreviar se usa **<xs:schema>**, aunque en la práctica cualquier prefijo puede ser usado, siempre que se use el mismo prefijo en todo el documento.

De esta manera, los prefijos **<xsd:schema>**, **<xsi:schema>** y **<xs:schema>** se pueden utilizar indistintamente al definir el esquema siempre que en el mismo se utilice únicamente uno de ellos.

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

El ejemplar de estos ficheros es **<xs:schema>**, contiene declaraciones para todos los elementos y atributos que puedan aparecer en un documento XML asociado válido. Los elementos hijos inmediatos de este ejemplar son **<xs:element>** que nos permiten crear globalmente un elemento. Esto significa que el elemento creado puede ser el ejemplar del documento XML asociado.

El elemento **<xs:schema>** puede tener algunos atributos. La declaración de un esquema suele tener el siguiente aspecto:

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

<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
targetNamespace="https://www.w3schools.com"
xmlns="https://www.w3schools.com"
elementFormDefault="qualified">
...
...
</xs:schema>
```

El siguiente fragmento:

```
xmlns:xs="http://www.w3.org/2001/XMLSchema"
```

indica que los elementos y tipos de datos usados en el esquema vienen del espacio de nombres "http://www.w3.org/2001/XMLSchema". También especifica que los elementos y los tipos de datos que provengan de dicho espacio de nombres deben tener el prefijo **xs:**. **Este fragmento es el único obligatorio para que la definición sea correcta.**

El siguiente fragmento:

```
targetNamespace="https://www.w3schools.com"
```

indica que los elementos definidos en el esquema pertenecen al espacio de nombres de "https://www.w3schools.com". Por defecto toma este valor.

El siguiente fragmento:

```
xmlns="https://www.w3schools.com"
```

indica que el espacio de nombres por defecto es "https://www.w3schools.com".

El siguiente fragmento:

```
elementFormDefault="qualified"
```

indica que cualquier elemento usado en una instancia xml que ha sido declarada con este esquema debe ser identificado con el espacio de nombres. Por defecto toma este valor.

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML



Ejercicio resuelto

Creación de un esquema correspondiente al siguiente documento XML:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<alumno>Olga Velarde Cobo</alumno>
```

Mostrar retroalimentación

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="alumno" type="xs:string"/>
</xs:schema>
```

Para saber más

En este primer enlace encontrarás los fundamentos del estándar XML Schema. [Fundamentos XML Schema](#)

3.1.TIPOS DE ELEMENTOS EN XML SCHEMA

Un elemento se utiliza para describir datos. Recordemos que su etiqueta más habitual suele ser `xsd:element`, `xsi:element` o `xs:element`.

Los elementos se utilizan para especificar las etiquetas válidas en un documento XML. Todos los elementos que se vayan a utilizar en el ejemplar XML tienen que estar declarados en el esquema. Las declaraciones de elementos en XML Schema tienen una estructura diferente dependiendo de si son **simples** o **complejos**:

- **Tipo simple:** No pueden contener otros elementos o atributos. Su estructura es la siguiente:

```
<xs:element name="nombreElemento"
  ref="elementoReferenciado"
  type="tipoDato"
  minOccurs="valor"
  maxOccurs="valor"
  fixed="valor"
  default="valor"/>
```

Donde:

- **name:** es el nombre del elemento.
- **ref:** el elemento al que hace referencia está declarado en otro lugar del esquema. No puede aparecer junto con "name" y ni si el elemento padre es `<xs:schema>`
- **type:** el tipo de dato del elemento. No puede aparecer si se usa "ref".

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

- **minOccurs y maxOccurs** (Opcionales): estos dos atributos indican el mínimo (minOccurs) y máximo (maxOccurs) número de ocurrencias del elemento. El valor por defecto en ambos casos es 1. Para indicar que el elemento puede aparecer un número ilimitado de veces, el atributo maxOccurs toma el valor "unbounded". Para indicar que un elemento puede no aparecer, el atributo minOccurs toma el valor 0. Ninguno de los atributos se puede usar si el elemento padre es <xs:schema>.
- **fixed** (Opcional): especifica un valor fijo para el elemento.
- **default** (Opcional): especifica un valor por defecto para el elemento.
- Se pueden crear nuevos elementos "simpleType" a partir de uno ya existente, añadiendo condiciones a alguno de los tipos predefinidos en el XML Schema. Para ello se usa el elemento <xs:restriction>

```
<xsd:simpleType name="precio">  
  <xsd:restriction base="xsd:decimal">  
    <xsd:fractionDigits value="2"/>  
  </xsd:restriction>  
</xsd:simpleType>
```

- **Tipo complejo:** Pueden estar compuestos por otros elementos y/o atributos. Su contenido está definido entre las etiquetas de inicio y de cierre del elemento .
 - **Esquema, xs:schema**, contiene la definición del esquema, es el elemento tipo complejo básico.
 - **xs:complexType**, entre sus etiquetas de inicio y cierre se definen los elementos de tipo complejo. Pueden estar formados por subelementos predefinidos en XML Schema como:
 - **Secuencias, xs:sequence**, permite construir elementos complejos mediante la enumeración de los elementos que los forman en un orden concreto. Si se altera dicho orden en el documento xml, dicho documento no será válido.
 - **Alternativa, xs:choice**, representa alternativas, hay que tener en cuenta que es una o-exclusiva. Especifica una lista concreta de elementos de los que sólo puede aparecer uno.
 - **Secuencias no ordenadas, xs:all**, representa a todos los elementos que componen el elemento de tipo compuesto en cualquier orden. Pueden aparecer en el documento xml en cualquier orden y dicho documento es válido.
 - **Contenido mixto**, definido dando valor true al atributo mixed del elemento <xs:complexType mixed="true">, permite mezclar texto con elementos hijo. Los elementos hijo se definen con las opciones anteriores; xs:sequence, xs:choice o xs:all.
 - **Elemento vacío**, el elemento no puede contener texto ni otros subelementos; únicamente atributos.

- XSD:

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

```
<xsd:element name="asignatura">
  <xsd:complexType>
    <xsd:attribute name='codigo' type='xs:integer' />
  </xsd:complexType>
</xsd:element>
```

- XML válido:

```
<asignatura codigo='MAT-1920'>(vacío)</asignatura>
```

O lo que es lo mismo y más correcto:

```
<asignatura codigo='MAT-1920' />
```

- **Referencias a otros elementos:** Tal y como ocurre en otros lenguajes, en un documento xsd podemos definir elementos de forma global y luego hacer referencias a ellos desde otros elementos. Esto es muy útil si a lo largo de un documento se repiten determinados elementos.

```
<?xml version="1.0" encoding="UTF-8"?>
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:element name="Direccion">
    <xsd:complexType>
      <xsd:sequence>
        <xsd:element ref="via"/>
        <xsd:element ref="numero"/>
        <xsd:element ref="poblacion"/>
        <xsd:element ref="provincia"/>
        <xsd:element ref="cp"/>
      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>
  <xsd:element name="via" type="xsd:string"/>
  <xsd:element name="numero" type="xsd:integer"/>
  <xsd:element name="poblacion" type="xsd:string"/>
  <xsd:element name="provincia" type="xsd:string"/>
  <xsd:element name="cp" type="xsd:string"/>
</xsd:schema>
```

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML



Ejercicio resuelto

Ejemplo de esquema correspondiente a un documento XML para estructurar la información personal sobre los alumnos de un centro educativo.

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">

  <!-- elemento raíz -->

  <xs:element name="alumnos" type="datosAlum"/>

  <!-- Definición del tipo datosAlum -->

  <xs:complexType name="datosAlum">
    <xs:sequence>
      <xs:element name="alumno" type="datos" minOccurs="1" maxOccurs="unbounded"/>
    </xs:sequence>
  </xs:complexType>

  <!-- Definición del tipo datos -->

  <xs:complexType name="datos">
    <xs:sequence>
      <xs:element name="nombre" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="apellidos" type="xs:string" minOccurs="1" maxOccurs="1"/>
      <xs:element name="direccion" type="datosDireccion" minOccurs="1" maxOccurs="1"/>
      <xs:element name="contactar" type="datosContactar" minOccurs="1" maxOccurs="1"/>
    </xs:sequence>
  </xs:complexType>

  <!-- Atributos del elemento alumno -->
```

```
<xs:attribute name="id" type="xs:string"/>
</xs:complexType>

<xs:complexType name="datosDireccion">
  <xs:sequence>
    <xs:element name="domicilio" type="xs:string" minOccurs="0" maxOccurs="1"/>
    <xs:element name="codigo_postal" minOccurs="0" maxOccurs="1">
      <xs:complexType>
        <xs:attribute name="cp" type="xs:string"/>
      </xs:complexType>
    </xs:element>
    <xs:element name="localidad" type="xs:string" minOccurs="0" maxOccurs="1"/>
    <xs:element name="provincia" type="xs:string" minOccurs="0" maxOccurs="1"/>
  </xs:sequence>
</xs:complexType>

<xs:complexType name="datosContactar">
  <xs:sequence>
    <xs:element name="telf_casa" type="xs:string" minOccurs="0" maxOccurs="1"/>
    <xs:element name="telf_movil" type="xs:string" minOccurs="0" maxOccurs="1"/>
    <xs:element name="telf_trabajo" type="xs:string" minOccurs="0" maxOccurs="1"/>
    <xs:element name="email" minOccurs="0" maxOccurs="unbounded">
      <xs:complexType>
        <xs:attribute name="href" type="xs:string"/>
      </xs:complexType>
    </xs:element>
  </xs:sequence>
</xs:complexType>
</xs:schema>
```

3.2. ATRIBUTO EN XML SCHEMA

El elemento "atributo" permite definir los atributos de los elementos en el documento xsd para usarlos adecuadamente en el documento XML.

```
<xsd:attribute name="nombreAtributo"
  ref="atributoReferenciado"
  type="tipoAtributo"
  use="valor"
  fixed="valor"
  default="valor"/>
```

Donde:

- **name:** nombre del atributo.
- **ref:** el atributo referenciado se encuentra definido en otra parte del esquema. No puede aparecer al mismo tiempo que **name**.

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

- **type**: tipo de dato del atributo. No puede aparecer al mismo tiempo que **ref**.
- **use** (opcional): indica si la aparición del atributo es opcional (optional), obligatoria (required) o prohibida (prohibited). Por defecto toma el valor "optional".
- **default** (opcional): valor que tomará el atributo al ser procesado si en el documento XML no hay ningún valor. Sólo se puede usar con tipos de datos cadena de caracteres. No puede aparecer al mismo tiempo que **fixed**.
- **fixed** (opcional): valor fijo que toma el atributo. No puede aparecer al mismo tiempo que **default**.

Los atributos sólo pueden aparecer en los elementos de tipo compuesto y su declaración debe aparecer siempre al final de la definición del elemento del que es atributo, justo antes del cierre de `</xs:complexType>`.



Ejercicio Resuelto

Crear el código xsd para el siguiente xml

```
<libro codigo="001">
  <titulo>Don Quijote</titulo>
  <autor>Miguel de Cervantes</autor>
</libro>
```

Mostrar retroalimentación

```
<xs:element name="libro">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="nombre" type="xs:string"/>
      <xs:element name="autor" type="xs:string"/>
    </xs:sequence>
    <xs:attribute name="codigo" type="xs:string"/>
  </xs:complexType>
</xs:element>
```


TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

3.3. TIPOS DE DATOS

Son los distintos valores que puede tomar el **atributo type** cuando se declara un elemento o un atributo y representan el tipo de dato que tendrá el elemento o atributo asociado a ese **type** en el documento XML.

Algunos de estos valores predefinidos son:

- **string**, se corresponde con una cadena de caracteres UNICODE. Puede contener caracteres, saltos de línea y tabulaciones.

- XSD:

```
<xs:element name="poblacion" type="xs:string"/>
```

- XML válido:

```
<poblacion>La Puebla de VÍcar</poblacion>
```

O lo que es lo mismo:

```
<poblacion>    La Puebla de VÍcar    </poblacion>
```

- **boolean**, representa valores lógicos, es decir que solo pueden tomar dos valores, true o false.

- XSD:

```
<xs:attribute name="cancelado" type="xs:boolean"/>
```

- XML válido:

```
<vuelo cancelado="true">LK345</vuelo>
```

- **integer**, número entero positivo o negativo.

- XSD:

```
<xs:element name="precio" type="xs:integer"/>
```

- XML válido:

```
<precio>94</precio>
```

- **positiveInteger**, número entero positivo.
- **negativeInteger**, número entero negativo.
- **decimal**, número decimal, por ejemplo, 8,97.

- XSD:

```
<xs:element name="precio" type="xs:decimal"/>
```

- XML válido:

```
<precio>8,97</precio>
```

o también

```
<precio>8</precio>
```

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

- **dateTime**, representa una fecha y hora absolutas. Tiene el siguiente formato: "YYYY-MM-DDThh:mm:ss". Sólo es válido si se especifican todos los componentes.

- XSD:

```
<xs:element name="fecha" type="xs:dateTime"/>
```

- XML válido:

```
<fecha>2020-05-20T08:20:00</fecha>
```

o bien

```
<fecha>2020-05-20T08:20:00.5</fecha>
```

- **duration**, representa una duración de tiempo expresado en años, meses, días, horas, minutos segundos. El formato utilizado es: PnYnMnDTnHnMnS. Para indicar una duración negativa se pone un signo – precediendo a la P.

- XSD:

```
<xs:element name="periodo" type="xs:duration"/>
```

- XML válido:

Por ejemplo para representar una duración de 2 años, 4 meses, 3 días, 5 horas, 6 minutos y 7 segundos habría que poner:

```
<periodo>P2Y4M3DT5H6M7S</periodo>
```

Se pueden omitir los valores nulos, luego una duración de 2 años será

```
<periodo>P2Y</periodo>
```

Un ejemplo de duración negativa sería:

```
<periodo>-P2D</periodo> indica un periodo de menos de dos días.
```

- **time**, hora en el formato hh:mm:ss.
- **date**, fecha en formato CCYY-MM-DD.
- **gYearMonth**, representa un mes de un año determinado mediante el formato CCYY-MM.

- XSD:

```
<xs:element name="fecha" type="xs:gYearMonth"/>
```

- XML válido:

```
<fecha>2020-05-20T08:20:00</fecha> Mayo de 2020
```

- XML NO válidos:

```
<fecha>20-05</fecha> El año no puede divid
```

o bien

```
<fecha>2020</fecha> El mes es requerido
```

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

- **gYear**, indica un año gregoriano, el formato usado es CCYY.
- **gMonthDay**, representa un día de un mes mediante el formato --MM-DD.
 - XSD:

```
<xs:element name="fecha" type="xs:gMonthDay"/>
```
 - XML válido:

```
<fecha>--05-19</fecha> 19 de May
```
 - XML NO válidos:

```
<fecha>05-19</fecha> Los guiones delante del mes son obligatorios  
o bien  
<fecha>--05-32</fecha> Debe ser un día válido del año; mayo tiene 31 días.
```
- **gDay**, indica el ordinal del día del mes mediante el formato -DD, es decir el 4º día del mes será -04.
- **gMonth**, representa el mes mediante el formato -MM. Por ejemplo, febrero es -02.
- **anyURI**, representa una URI.
 - XSD:

```
<xs:element name="web" type="xs:anyURI"/>
```
 - XML válido:

```
<web>www.iesaguadulce.es</web> URI absoluta, incluso una URI  
o bien  
<web>www.iesaguadulce.es#texto</web> URI con un identificador #
```
 - XML NO válidos:

```
<web>www.iesaguadulce.es#texto#texto1</web> Demasiados identificadores #
```
- **language**, representa los identificadores de lenguaje, sus valores están definidos en RFC 1766.
- **ID, IDREF, ENTITY, NOTATION, NMTOKEN**. Representan lo mismo que en los DTD's (ver apartado 2.4).

Para saber más

En este enlace encontrarás los tipos de datos admitidos por el estándar. [XML Schema Tipos de datos](#)

En el siguiente enlace podrás ver ejemplos válidos y no válidos de los distintos tipos de datos. [Tipos de datos](#)


TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

3.4. FACETAS DE LOS TIPOS DE DATOS

¿Cuáles son las restricciones que podemos aplicar sobre los valores de los datos de un elemento o atributo? Están definidas por las **facetas**, que solo pueden aplicarse sobre tipos simples utilizando el elemento **xs:restriction**, que tiene como atributo "**base**" en el que se indica el tipo de dato a partir del cual se define la restricción.

Las facetas se expresan como un elemento dentro de una restricción y se pueden combinar para lograr restringir más el valor del elemento. Son, entre otros:

- **length, minlength, maxlenght**: Longitud del tipo de datos.
- **enumeration**: Restringe a un determinado conjunto de valores.




Ejercicio resuelto

Creación de una cadena de texto con una longitud máxima de 9 caracteres y dos valores posibles

Mostrar retroalimentación

```
<xs:simpleType name="estado">
  <xs:restriction base="xs:string">
    <xs:maxLength value="9"/>
    <xs:enumeration value="conectado"/>
    <xs:enumeration value="ocupado"/>
  </xs:restriction>
</xs:simpleType>
```

- **whitespace**: Define el tratamiento de espacios (preserve/replace, collapse).



Ejercicio resuelto

Creación de un elemento en el que se respetan los espacios tal y como se han introducido.

Mostrar retroalimentación

```
<xs:simpleType name="nombre">
  <xs:restriction base="xs:string">
    <xs:whitespace value="preserve"/>
  </xs:restriction>
</xs:simpleType>
```

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

- **(max/min)(In/Ex)clusive:** Límites superiores/inferiores del tipo de datos. Cuando son Inclusive el valor que se determine es parte del conjunto de valores válidos para el dato, mientras que cuando se utiliza Exclusive, el valor dado no pertenece al conjunto de valores válidos.
- **totalDigits, fractionDigits:** número de dígitos totales y decimales de un número decimal.



Ejercicio resuelto

Creación de un elemento calificaciones de dos dígitos cuyo valor es un número entero comprendido entre 1 y 10, ambos inclusive.

Mostrar retroalimentación

```
<xs:simpleType name="calificaciones">
  <xs:restriction base="xs:integer">
    <xs:totalDigits value="2"/>
    <xs:minExclusive value="0"/>
    <xs:maxInclusive value="10"/>
  </xs:restriction>
</xs:simpleType>
```

- **pattern:** Permite construir máscaras que han de cumplir los datos de un elemento. La siguiente tabla muestra algunos de los caracteres que tienen un significado especial para la generación de las máscaras.

Elementos para hacer patrones.

Patrón	Significado
[A-Z a-z]	Letra.
[A-Z]	Letra mayúscula.
[a-z]	Letra minúscula.
[0-9]	Dígitos decimales.
\D	Cualquier carácter excepto un dígito decimal.
(A)	Cadena que coincide con A.
A B	Cadena que es igual a la cadena A o a la B.

Elementos para hacer patrones.

Patrón	Significado
AB	Cadena que es la concatenación de las cadenas A y B.
A?	Cero o una vez la cadena A.
A+	Una o más veces la cadena A.
A*	Cero o más veces la cadena A.
[abcd]	Alguno de los caracteres que están entre corchetes.
[^abcd]	Cualquier carácter que no esté entre corchetes.

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML



Ejercicio resuelto

El ejemplo siguiente muestra la utilización de pattern para crear la máscara de un DNI.

Mostrar retroalimentación

```
<xs:simpleType name="dni">
  <xs:restriction base="xs:string">
    <xs:pattern value="[0-9]{8}[A-Z]"/>
  </xs:restriction>
</xs:simpleType>
```

Recomendación

Visita el siguiente enlace para conocer un poco más sobre las expresiones regulares: [Expresiones regulares](#)

3.5.DEFINICIÓN DE TIPOS DE DATOS SIMPLES XML SCHEMA

En los DTD se diferencia entre los elementos terminales y los no terminales ¿en este caso también? Si, este lenguaje **permite trabajar tanto con datos simples como con estructuras de datos complejos**, es decir, compuestos por el anidamiento de otros datos simples o compuestos. En este apartado vamos a indicar las tres diferentes maneras que existen de extender un tipo de datos simple:

- **Restricción.**

Se hace una restricción sobre un tipo de datos **XSD** ya definido y se establece el rango de valores que puede tomar. Las restricciones son conocidas como **facet**as.



Ejercicio resuelto

Creación de un elemento simple de nombre edad que representa la edad de un alumno de la ESO, por tanto su rango está entre los 12 y los 18 años

Mostrar retroalimentación

```
<xs:simpleType name="edad">
  <xs:restriction base="xsd:positiveInteger">
    <xs:maxExclusive value="19"/>
    <xs:minInclusive value="12"/>
  </xs:restriction>
</xs:simpleType>
```

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

- **Union.**

Consiste en combinar dos o más tipos de datos diferentes en uno único.



Ejercicio Resuelto

Crear un tipo de dato para las medidas de longitud que una el sistema internacional (cm, m y km) con el sistema inglés (pulgada, pie, yarda)

Mostrar retroalimentación

```
<xs:simpleType name="LongitudInternacional">
  <xs:restriction base="xs:string">
    <xs:enumeration value="cm" />
    <xs:enumeration value="m" />
    <xs:enumeration value="km" />
  </xs:restriction>
</xs:simpleType>

<xs:simpleType name="LongitudSajona">
  <xs:restriction base="xs:string">
    <xs:enumeration value="pulgada" />
    <xs:enumeration value="pie" />
    <xs:enumeration value="yarda" />
  </xs:restriction>
</xs:simpleType>
```

Unimos los dos tipos en uno único:

```
<xs:simpleType name="Longitud">
  <xs:union memberTypes="LongitudInternacional LongitudSajona">
  </xs:union>
</xs:simpleType>
```

- **Lista.**

Permite asignar a un elemento un valor dentro de un número de valores válidos separados por espacios en blanco. Puede ser creada de manera similar a una unión, con la diferencia de que sólo puede contener un tipo de elementos.



Ejercicio Resuelto

Crear una lista con las unidades de medida de volumen sajonas.


Mostrar retroalimentación

```
<xs:simpleType name="LongitudSajona">
  <xs:restriction base="xs:string">
    <xs:enumeration value="pulgada" />
    <xs:enumeration value="pie" />
    <xs:enumeration value="yarda" />
  </xs:restriction>
</xs:simpleType>

<xs:simpleType name="Longitud">
  <xs:list itemType="LongitudSajona">
  </xs:list>
</xs:simpleType>
```

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

- También podemos usar el atributo de simpleType, derivedBy, para crear una lista de elementos.



Ejercicio resuelto

Creación de una lista con los días de la semana en letras.

Mostrar retroalimentación

```
<xs:simpleType name="dia_semana" base="xs:string" derivedBy="list"/>
  <dia_semana>Lunes Martes Miercoles Jueves Viernes Sabado Domingo</dia_semana>
</xs:simpleType>
```

Otra opción:

```
<xs:simpleType name="dia_semana">
  <xs:restriction base="xs:string">
    <xs:pattern value="Lunes|Martes|Miercoles|Jueves|Viernes|Sabado|Domingo"/> </xs:restriction>
  </xs:simpleType>
```

3.6.DEFINICIÓN DE TIPOS DE DATOS COMPLEJOS XML SCHEMA

Tipo complejo: Pueden estar compuestos por otros elementos y/o atributos. Su contenido está definido entre las etiquetas de inicio y de cierre del elemento.

- Esquema, xs:schema**, contiene la definición del esquema, es el elemento tipo complejo básico. Contiene el atributo xmlns (XML NameSpace), que indica el espacio de nombres para el esquema.

- XSD:

```
<xs:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
</xs:schema>
```

- xs:complexType**, entre sus etiquetas de inicio y cierre se definen los elementos de tipo complejo. Pueden estar formados por subelementos predefinidos en XML Schema como:
 - **Secuencias, xs:sequence**, permite construir elementos complejos mediante la enumeración de los elementos que los forman en un orden concreto. Si se altera dicho orden en el documento xml, dicho documento no será válido.

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

- XSD:

```
<xs:element name="Direccion">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="calle" type="xs:string" />
      <xs:element name="poblacion" type="xs:string" />
      <xs:element name="provincia" type="xs:string" />
      <xs:element name="codigo_postal" type="xs:int" />
    </xs:sequence>
  </xs:complexType>
</xs:element>
```

o lo que es lo mismo:

```
<xs:element name="Direccion">
  <xs:complexType>
    <xs:complexContent>
      <xs:sequence>
        <xs:element name="calle" type="xs:string" />
        <xs:element name="poblacion" type="xs:string" />
        <xs:element name="provincia" type="xs:string" />
        <xs:element name="codigo_postal" type="xs:int" />
      </xs:sequence>
    </xs:complexContent>
  </xs:complexType>
</xs:element>
```

- XML válido:

```
<Direccion>
  <calle>Lago de Enol, nº 32</calle>
  <poblacion>Aguadulce</poblacion>
  <provincia>Almeria</provincia>
  <codigo_postal>04720</codigo_postal>
</Direccion>
```

- XML **NO** válido:

```
<Direccion>
  <poblacion>Aguadulce</poblacion>
  <provincia>Almeria</provincia>
  <calle>Lago de Enol, nº 32</calle>
  <codigo_postal>04720</codigo_postal>
</Direccion>
```

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

- **Alternativa, xs:choice**, representa alternativas, hay que tener en cuenta que es una o-exclusiva. Especifica una lista concreta de elementos de los que sólo puede aparecer uno.

- XSD:

```
<xs:element name="FormaPago">
  <xs:complexType>
    <xs:choice>
      <xs:element name="paypal" type="xs:string" />
      <xs:element name="transferencia" type="xs:string" />
      <xs:element name="bizum" type="xs:string" />
      <xs:element name="contado" type="xs:int" />
    </xs:choice>
  </xs:complexType>
</xs:element>
```

- XML válido:

```
<FormaPago>transferencia</FormaPago>
```

- **Secuencias no ordenadas, xs:all**, representa a todos los elementos que componen el elemento de tipo compuesto en cualquier orden. Pueden aparecer en el documento xml en cualquier orden y dicho documento es válido.
- **Contenido mixto**, definido dando valor true al atributo mixed del elemento **<xs:complexType mixed="true">**, permite mezclar texto con elementos hijo. Los elementos hijo se definen con las opciones anteriores; xs:sequence, xs:choice o xs:all.
- **Elemento vacío**, el elemento no puede contener texto ni otros subelementos; únicamente atributos.

- XSD:

```
<xs:element>
  <xs:complexType>
    <xs:attribute name='codigo' type='xs:integer'/>
  </xs:complexType>
</xs:element>
```

- XML válido:

```
<asignatura codigo='MAT-1920'>(vacío)</asignatu
O lo que es lo mismo y más correcto:
<asignatura codigo='MAT-1920'/>
```

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

- **Elemento simple con atributo**, es un elemento simple porque no contiene a otros elementos, sólo datos, pero es de tipo complejo (complexType) porque tiene un atributo.

- XSD:

```
<xs:element name="cuota">
  <xs:complexType>
    <xs:simpleContent>
      <xs:attribute name='moneda' type='xs:string'/>
    </xs:simpleContent>
  </xs:complexType>
</xs:element>
```

- XML válido:

```
<cuota moneda='Euro'>1200</cuota>
```

3.7.ASOCIACIÓN DE DOCUMENTOS XML

Una vez que tenemos creado el **fichero XSD** ¿cómo lo **asociamos a un fichero XML**?

El modo de asociar un esquema a un documento XML es un espacio de nombres al ejemplar del documento, donde se indica la ruta de localización de los ficheros esquema mediante su [URI](#), precedida del prefijo " **xs:**".

Para que un documento XML siga las reglas definidas en un esquema, no disponemos de etiqueta !DOCTYPE; en su lugar utilizamos atributos especiales en el elemento raíz del documento XML.

Primero, al igual que en el documento XMLSchema, necesitamos definir los dos espacios de nombres, el correspondiente al documento XML (que se suele usar sin abreviatura, es decir como espacio por defecto) y el espacio de nombres de XMLSchema (que suele utilizar el prefijo xs, aunque se puede utilizar otro).

Además es necesario indicar dónde está el archivo XMLSchema que contiene las reglas de validación que se aplican al documento. Esto se hace gracias al atributo llamado schemaLocation (perteneciente al espacio de nombres del esquema, por lo que se usa normalmente como xs:schemaLocation).

```
<?xml version="1.0" encoding="UTF-8"?>
<documento xmlns="http://www.iesaguadulce.es/doc" xmlns:xs="http://w3.org/2001/XMLSchema-instance"
xs:schemaLocation="esquema.xsd">
....
</documento>
```

Se indica el espacio por defecto de nombres en el documento (coincide con el declarado en el propio archivo del esquema), se indica el espacio de nombres correspondiente al esquema (siempre es la misma dirección de Internet) y se asocia a este espacio el prefijo xs (se puede elegir otro prefijo, pero no es nada conveniente).

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

En el caso de que en el Schema únicamente definamos el espacio de nombres `xmlns:xs="http://www.w3.org/2001/XMLSchema"`, en el documento xml habrá que indicar:

```
<?xml version="1.0" encoding="UTF-8"?>
<documento xmlns:xs="http://w3.org/2001/XMLSchema-instance"
xs:noNamespaceSchemaLocation="esquema.xsd">
....
</documento>
```



Ejercicio resuelto

Un documento XML asociado al esquema que se ha realizado anteriormente para estructurar la información personal sobre los alumnos de un centro educativo (apartado 3.1) puede ser:

1

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<alumnos xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:SchemaLocation="alumnos.xsd">
  <alumno>
    <nombre>Jose Ramón</nombre>
    <apellidos>García González</apellidos>
    <direccion>
      <domicilio>El Pez, 12</domicilio>
      <codigo_postal>85620</código_postal>
      <localidad>Suances</localidad>
      <provincia>Cantabria</provincia>
    </direccion>
    <contactar>
      <telf_casa>985623165</telf_casa>
      <telf_movil>611233544</telf_movil>
      <telf_trabajo>965847536</telf_trabajo>
      <email>pepito@educadistancia.com</email>
    </contactar>
  </alumno>
</alumnos>
```

2

```
<alumno>
  <nombre>Carlos</nombre>
  <apellidos>López Pérez</apellidos>
  <direccion>
    <domicilio>El Cangrejo, 25</domicilio>
    <codigo_postal>86290</código_postal>
    <localidad>Santillana</localidad>
    <provincia>Cantabria</provincia>
  </direccion>
  <contactar>
    <telf_casa>931132565</telf_casa>
    <telf_movil>623863544</telf_movil>
    <telf_trabajo>984657536</telf_trabajo>
    <email>carlos@educadistancia.com</email>
  </contactar>
</alumno>
</alumnos>
```

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

3.8. DOCUMENTACIÓN DEL ESQUEMA

Una vez que hemos visto como crear un esquema vamos a ver el modo de incorporar cierta documentación (quién es el autor, limitaciones de derechos de autor, utilidad del esquema, etc.) al mismo.

Podemos pensar que un método para añadir esta información es utilizar comentarios. El problema es que los analizadores no garantizan que los comentarios no se modifiquen al procesar los documentos y por tanto, que los datos añadidos no se pierdan en algún proceso de transformación del documento.

En lugar de usar los comentarios, XML Schema tiene definido un elemento **xs:annotation** que permite **guardar información adicional**. Este elemento a su vez puede contener una combinación de otros dos que son:

- **xs:documentation**, además de contener elementos de esquema puede contener elementos XML bien estructurados.
También permite determinar el idioma del documento mediante el atributo `xml:lang`.
- **xs:appinfo**, se diferencia muy poco del elemento anterior, aunque lo que se pretendió inicialmente era que `xs:documentation` fuese legible para los usuarios y que `xs:appinfo` guardase información para los programas de software.

También es usado para generar una ayuda contextual para cada elemento declarado en el esquema.



Ejercicio resuelto

Ejemplo de documentación de un esquema.

Mostrar retroalimentación

```
<xs:schema xmlns:xsi="http://www.w3.org/2001/XMLSchema">

  <xs:annotation>
    <xs:documentation xml:lang="es-es">
      Materiales para formación e-Learning
    <modulo>Lenguajes de marcas y sistemas de gestión de información.</modulo>
    <fecha_creación>2011</fecha_creación>
    <autor>Nuky La Bruji</autor>
  </xs:documentation>
</xs:annotation>

  <xs:element name="lmsgi" type="xs:string">
    <xs:annotation>
      <xs:appinfo>
        <texto_de_ayuda>Se debe de introducir el nombre completo del tema</texto_de_ayuda>
      </xs:appinfo>
    </xs:annotation>
  </xs:element>
</xs:schema>
```

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

4. HERRAMIENTAS DE CREACIÓN Y VALIDACIÓN

Igual que hasta ahora, para crear y validar los documentos XML y los esquemas, basta con un editor de texto plano y un navegador. ¿Pero, no hay ninguna herramienta que nos facilite el trabajo? Pues sí, existen aplicaciones que permiten al usuario visualizar, validar y editar documentos en el lenguaje XML. Algunos de estos productos de escritorio son:

- Notepad ++ (Gratuito)
- Editix XML Editor (Gratuito)
- XML Copy Editor (Gratuito)
- NetBeans (Gratuito)

También tenemos herramientas de validación online, a continuación algunas web que nos validan un xml, dtd y schema:

- <http://www.xmlvalidation.com>
- <http://www.utilities-online.info/xsdvalidation/>

TEMA 4: DEFINICIÓN DE ESQUEMAS Y VOCABULARIOS EN XML

ANEXO 1. DOCUMENTACIÓN DE APOYO

- DTD: http://www.mclibre.org/consultar/xml/lecciones/xml_dtd.html
- Resumen DTD: [Descargar PDF](#)
- Ejercicios DTD: <http://www.mclibre.org/consultar/xml/ejercicios/dtd.html>
- DTD y XML Schema: [Descargar PDF](#)
- [Video] XML Schema (Estructura): http://www.youtube.com/watch?list=PLKLLGxMgKZ_QCcAMbq8L9dEXa6hMdb1q5&v=JKhfLpkVh3o&feature=player_detailpage
- [Video] Mi primer DTD: https://www.youtube.com/watch?feature=player_embedded&v=EfnWCeQNTQI
- [Video] Elementos DTD: https://www.youtube.com/watch?feature=player_embedded&v=ryoW-B_6cGs
- [Video] Creación DTD sencillo Parte I: https://www.youtube.com/watch?feature=player_embedded&v=fPU1ex7bSgg
- [Video] Creación DTD sencillo Parte II: https://www.youtube.com/watch?feature=player_embedded&v=4NB89iXyxMU
- [Descarga] Referencia rápida Tipos de Datos: [SchemaDataTypes](#) - [Datatype Hierarchy](#)