



# XML

---

## Guía básica para el alumno

## Documentos XML



- Contenidos:

[Introducción.](#)

[Sintaxis de un documento XML.](#)

[Arquitectura de manipulación de XML-data.](#)

[Verificación de un documento XML.](#)

1. Validación de documentos XML:
  - a. [Validación de documentos XML: DTD.](#)
  - b. [Validación de documentos XML: XML Schema.](#)
2. Lenguajes de acceso a la información para documentos XML.
3. Transformación de documentos XML.

XML es el acrónimo de eXtensible Markup Language (Lenguaje Extensible de Marcado).

### Sintaxis de un documento XML.

Todo documento XML tendrá la siguiente **estructura**:

**1.- El prólogo o Cabecera:** contendrá información determinante sobre el resto de datos del documento, instrucciones de procesamiento (informa al intérprete encargado de procesar el documento de todos aquellos datos que necesita para realizar su trabajo), y estará compuesto por:

- La declaración del XML, `<? ... ¿>`,  
`<?xml version="1.0" encoding="utf-8" standalone="yes"?>`
  - version: indica la versión de XML (la más extendida sigue siendo la versión 1.0, aunque ya existe la versión 1.1)
  - encoding: indica la codificación de caracteres.
  - standalone: indica si el documento está listo para procesarse independientemente o requiere de archivos externos (por defecto: no).  
standalone="yes": el documento XML es independiente.
- De manera opcional, también podemos encontrar la declaración del tipo de documento, que puede ser una referencia a una DTD externa o parte de una DTD, que será usada por el parser XML para comprobar la validez del documento

durante el proceso de validación, seguida de cero o más comentarios o instrucciones de procesado.

```
<!-- Esto es un comentario en XML -->
```

```
<!DOCTYPE nombre SYSTEM "arch.dtd" >
```

- Visualización: podemos establecer la forma de representar el documento mediante una hoja de estilos (CSS) o una hoja de transformaciones (XSLT).

```
<?xml-stylesheet type="text/css" href="ejemplo.css" ?>
```

```
<?xml-stylesheet type="text/xsl" href="transforma.css" ?>
```

## 2.- El cuerpo del documento (datos que se quieren procesar):

```
<Raíz>
```

```
...
```

```
</Raíz>
```

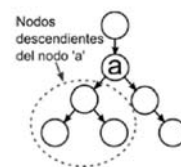
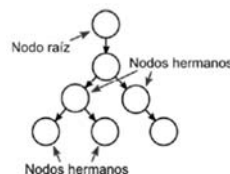
- Elementos o Etiquetas anidadas:

```
<ETIQ1>...
```

```
<ETIQ2>... </ETIQ2>
```

```
...
```

```
</ETIQ1>
```



- Atributos:

```
<ETIQ atributo1="valor1" atributo2="valor2"...>
```

- Contenido: los elementos pueden contener texto, otro elemento, o estar vacíos.

- Otro elemento: anidamiento de elementos formando la estructura jerárquica.
- Texto: representan la información de un elemento y pueden contener cualquier carácter, aunque para algunos hay que utilizar su entidad:

| Carácter | Entidad | Carácter | Entidad |
|----------|---------|----------|---------|
| &        | &amp;   | '        | &apos;  |
| <        | &lt;    | "        | &quot;  |
| >        | &gt;    |          |         |

- Secciones CDATA (Character DATA): texto que no se desea que sea analizado

```
<![CDATA[...]]>
```

El contenido que no se analizará se introducirá en el espacio entre corchetes

- Elementos sin contenido

- Entidades:

```
<!ENTITY nombre "valor">
```

Una entidad se define mediante una etiqueta que comienza por <!ENTITY y termina por > y contiene el nombre y el valor de la entidad.

Para hacer referencia a una entidad se escribe sin espacios intermedios el carácter "&", el nombre de la entidad y el carácter ";"

```
&nombre;
```

Al abrir el documento XML el procesador sustituye la referencia a la entidad por su valor.

## 3.- Epílogo de un documento XML. (opcional)

Está formado por cero o más comentarios o instrucciones de procesado.

## ▪ Espacios de Nombres XML o XML Namespaces.

Un espacio de nombres XML es una recomendación W3C para proporcionar elementos y atributos con nombre único en una instancia XML.

- Declaración: Podemos usar XML namespaces, añadiendo el atributo **xmlns**, a cualquier elemento de nuestro archivo, y especificando cómo su valor una URI.

`<elemento xmlns:prefijo="URI_espacio_nombres">`

prefijo: sirve como referencia. Puede ser cualquier texto, pero se recomienda que sea descriptivo.

URI: debe ser único, aunque no es más que un identificador que no se comprueba mediante conexión alguna.

- Utilización: se antepone el prefijo definido en la declaración y dos puntos (:) al nombre del elemento o atributo.

`<prefijo:elemento prefijo:atributo="valor">`

El espacio de nombres puede ser referenciado anteponiendo su prefijo en el elemento en que se ha declarado y en sus descendientes.

También se puede declarar un espacio de nombres nuevo para un elemento descendiente de otro en el que ya se había declarado un espacio de nombres.

Se puede desasignar un elemento de un espacio de nombres por defecto asignando el valor vacío al atributo xmlns:

`<elemento xmlns="">`

**Todo documento válido está bien formado, pero no todo documento bien formado es válido.**