

Qué es XML?

XML es un lenguaje de marcas que se ha estandarizado y se ha convertido en uno de los formatos más populares para el almacenamiento e intercambio de datos estructurados entre distintas plataformas.

Nace como consecuencia de los problemas de HTML como formato de intercambio de información.

Por ello al crear XML se plantearon estos objetivos:

- [1] Debía de ser similar a HTML (de hecho, se basa en el lenguaje SGML base para el formato HTML), dada su enorme aceptación.
- [2] Debía de ser extensible, es decir, que sea posible añadir nuevas etiquetas sin problemas. Esa es la base del lenguaje XML.
- [3] Debía de tener unas reglas concisas y fáciles, además de estrictas.
- [4] Debía de ser fácil de implantar en todo tipo de sistemas. XML nace con una vocación multiplataforma, como base de intercambio de información entre sistemas de toda índole.
- [5] Debía ser fácil de leer para todas las personas.
- [6] Debía ser fácil crear herramientas capaces de procesar XML.

Diferencia entre HTML y XML

Definición

HTML es un lenguaje de marcado estándar para crear páginas web y aplicaciones web. XML es un lenguaje de marcado que define un conjunto de reglas para codificar documentos en un formato que sea legible por humanos y por máquinas. Esta es la diferencia básica entre HTML y XML.

Sensibilidad de caso

Mientras que HTML no distingue entre mayúsculas y minúsculas, XML distingue entre mayúsculas y minúsculas. Esta es una diferencia importante entre HTML y XML.

Etiquetas

Las etiquetas crean otra diferencia entre HTML y XML. Es decir; HTML tiene etiquetas predefinidas. Por otro lado, en XML, el programador define su propio conjunto de etiquetas.

Etiquetas de cierre

Además, en HTML, algunas etiquetas no tienen una etiqueta de cierre. Pero, en XML, es obligatorio cerrar cada etiqueta que se ha utilizado.

Enfoque principal

Otra diferencia entre HTML y XML es el enfoque principal de cada uno. HTML se enfoca en mostrar datos mientras que XML se enfoca en transportar información.

Uso

Además, HTML ayuda a desarrollar la estructura de las páginas web, mientras que XML ayuda a intercambiar datos entre diferentes plataformas.

Funcionamiento de XML

Cuestiones básicas

Los documentos XML, en definitiva, son documentos de texto cuyos metadatos se indican mediante etiquetas (marcas) que permiten clasificar el texto a través de su significado.

Las etiquetas en XML se deciden a voluntad, no hay una lista concreta de etiquetas que se pueden utilizar. La idea es que con XML estamos definiendo un tipo de documento cuya forma, significado y estructura dependerán solo de nuestras necesidades.

Ejemplo de XML:

```
<persona>
  <nombre>Jorge</nombre>
  <apellido>Sánchez</apellido>
</persona>
```

En este código, se define un elemento llamado **persona**, que contiene a dos elementos más: nombre y apellido, cuyos valores son Jorge y Sánchez respectivamente.

Algunos detalles sobre el funcionamiento son:

- Un elemento se define con un nombre entre los símbolos **<** y **>**. Eso forma una etiqueta de apertura.
- El elemento termina con la etiqueta de cierre, la cual se indica con el símbolo **/** antes del nombre de la etiqueta y todo ello nuevamente encerrado entre los símbolos **<** y **>**.
- El contenido entre la apertura y el cierre queda afectado por esas etiquetas. De hecho se tratará del contenido del elemento.

- XML distingue entre mayúsculas y minúsculas, siendo buena práctica escribir las etiquetas en minúsculas.
- Se pueden espaciar y tabular las etiquetas a voluntad. Es buena práctica que un elemento interno a otro aparezca en el código con una sangría mayor. Por eso en el código anterior, nombre, que está dentro de persona, aparece con una tabulación en su lado izquierdo.
- Los comentarios en el código se inician con los símbolos `<!--` y terminan con `-->` El texto dentro de esos símbolos no se tienen en cuenta como parte del documento XML y se usa para hacer anotaciones.
- Según la **W3C** (organismo de estandarización de XML), el texto en un documento XML debe de estar codificado en Unicode (normalmente UTF-8).

Estructura de los documentos

Los documentos XML se dividen en:

- **Prólogo.** Se trata de la primera zona del documento y sirve para indicar qué tipo de documento es. Los apartados del prólogo suelen ser:
 - Declaración del documento, que permite indicar el tipo de documento XML, es decir su versión y la forma en la que está codificado.
 - Instrucciones para el procesado del documento
 - Comentarios
 - Ruta hacia el documento **DTD**, **XSD** O **Relax NG** que sirve para validar el documento XML actual.
 - Indicación de otros documentos que afectan al actual. Como por ejemplo los documentos XSLT que sirven para dar formato al documento XML.
- **Elemento raíz.** Tras el prólogo tenemos el contenido en sí del documento. Todo ese contenido debe de estar incluido dentro de un mismo elemento llamado elemento raíz. Se trata de un elemento obligatorio que se abre tras el prólogo y se debe cerrar justo al final del documento. Es decir, no puede haber otro elemento después del elemento raíz.
- Dentro de un elemento (sea raíz o no), puede haber:
 - Más elementos
 - Atributos
 - Texto normal
 - Entidades
 - Comentarios

```
<?xml version="1.0" encoding="UTF-8"?>
<deportistas>
  <deportista>
    <deporte nombre="Atletismo" />
    <nombre>Jesse Owens</nombre>
  </deportista>
  <deportista>
    <deporte nombre="Natación" />
    <nombre>Mark Spitz</nombre>
  </deportista>
</deportistas>
```

Normas de sintaxis básicas

En un documento XML, todos los nombres de los elementos tienen que cumplir las siguientes normas:

- Pueden contener letras minúsculas, letras mayúsculas, números, puntos “.”, guiones medios “-” y guiones bajos “_”.
- Asimismo, pueden contener el carácter dos puntos “:”. No obstante, su uso se reserva para cuando se definan espacios de nombres.
- El primer carácter tiene que ser una letra o un guion bajo “_”.
- Las letras no inglesas (á, ñ, Ñ ...) están permitidas. Sin embargo, es recomendable no utilizarlas para reducir posibles incompatibilidades con programas que puedan no reconocerlas.

Carácteres especiales

Entidad	Carácter
<	<
>	>
&	&
"	“
'	'

Atributos

Los elementos de un documento XML pueden tener atributos definidos en la etiqueta de inicio. Un atributo sirve para proporcionar información extra sobre el elemento que lo contiene.

EJEMPLO Dados los siguientes datos de un producto:

Código: G45
Nombre: Gorro de lana
Color: negro
Precio: 12.56

Su representación en un documento XML podría ser, por ejemplo:

```
<producto codigo="G45">  
  <nombre color="negro" precio="12.56">Gorro de lana</nombre>  
</producto>
```

En este ejemplo se han escrito tres atributos: código, color y precio. Obsérvese que, sus valores ("G45", "negro" y "12.56") se han escrito entre comillas dobles ("). No obstante, también pueden ir entre comillas simples (').

Si, por ejemplo, el atributo código se quisiera representar como un elemento, se podría escribir:

```
<producto>  
  <codigo>G45</codigo>  
  <nombre color="negro" precio="12.56">Gorro de lana</nombre>  
</producto>
```

¿Elemento o atributo?

Al diseñar un documento XML puede surgir la duda de cuándo usar elementos y cuándo atributos. Aunque no se puede dar un criterio válido para todos los casos si podemos dar algunas pautas que ayuden a tomar una decisión:

- Si la información tiene una estructura interna debe ser un elemento. Si contiene una gran cantidad de información, parece más adecuado, un elemento.
- Podemos hacer el símil que un elemento es un sustantivo y un atributo un adjetivo.
- Los mecanismos de procesamiento y presentación de documentos, permiten tener mejor control sobre los elementos. Por tanto aquella información que tenga un procesamiento o presentación complejos debe ser un elemento.

- Como consejo final, en caso de duda, se debe utilizar un elemento.