

XML

Osornio Gutierrez Juan Damian
Varela Aguilar Luis Pavel

Definicion XML

Especificación para diseñar lenguajes de marcado, que permite definir etiquetas personalizadas para descripción y organización de datos.

Es un metalenguaje que permite definir lenguajes de marca que se utiliza para almacenar datos en forma legible.

Objetivo XML

Su objetivo es facilitar la representación, almacenamiento y transmisión de información por parte de aplicaciones informáticas, computadoras y medios de comunicación digital en general. Da soporte a bases de datos, siendo útil cuando varias aplicaciones deben comunicarse entre sí o integrar información.

XML se propone como un estándar para el intercambio de información estructurada entre diferentes plataformas, bases de datos, editores de texto, hoja de cálculo, etc.

¿Qué NO es XML?

- No es una versión mejorada de HTML.
- HTML es una aplicación de SGML por lo tanto de XML.
- No es un lenguaje para hacer paginas WEB.

Ventajas de ocupar XML

- Fácilmente procesable
- Separa radicalmente el contenido y el formato de presentación
- Diseñado para cualquier lenguaje y alfabeto. (encoding)

Características XML

XML es un subconjunto de SGML (Standard Generalised Mark-up Language), el cual es lenguaje de marcado generalizado estándar para definir lenguajes de marcado generalizados para documentos. Incorpora las tres características más importantes de este:

- Extensibilidad
- Estructura
- Validación

Características

- Basado en texto.
- Orientado a los contenidos no presentación.
- Las etiquetas se definen para crear los documentos, no tienen un significado preestablecido.
- No es sustituto de HTML.
- No existe un visor genérico de XML.

Estructura de un documento XML

Un documento XML está formado por datos de caracteres y marcado, el marcado lo forman las etiquetas:

Prologo { `<?xml version="1.0" encoding="ISO-8859-1" standalone="no"?>`
 `<! DOCTYPE persona SYSTEM "persona.dtd"`
Cuerpo { `<persona>`
 `<nombre>Luis</nombre>`
 `<apellidos>Pérez</apellidos>`
 `</persona>`

Estructura de las etiquetas :

Nombre del atributo Contenido del elemento

`<autor pais = "España"> Jose Ramón </autor>`

Nombre de elemento Valor de atributo Etiqueta de fin

The diagram illustrates the components of the XML tag `<autor pais = "España"> Jose Ramón </autor>`. Arrows point from labels to specific parts of the tag: 'Nombre del atributo' points to 'pais', 'Contenido del elemento' points to 'Jose Ramón', 'Nombre de elemento' points to '<autor', 'Valor de atributo' points to '="España"', and 'Etiqueta de fin' points to '</autor>'.

Componentes de un documento XML

En un documento XML existen los siguientes componentes:

Elementos: Pieza lógica del marcado, se representa con una cadena de texto(dato) encerrada entre etiquetas. Pueden existir elementos vacíos (
). Los elementos pueden contener atributos.

Instrucciones: Órdenes especiales para ser utilizadas por la aplicación que procesa
<?xml-stylesheet type="text/css" href="estilo.css">

Las instrucciones XML. Comienzan por <? Y terminan por ?>.

Comentarios: Información que no forma parte del documento. Comienzan por <!-- y terminan por -->.

Declaraciones de tipo: Especifican información acerca del documento:

```
<!DOCTYPE persona SYSTEM "persona.dtd">
```

Secciones CDATA: Se trata de un conjunto de caracteres que no deben ser interpretados por el procesador:

`<![CDATA[Aquí se puede meter cualquier carácter, como <, &, >, ... Sin que sean interpretados como marcación]]>`

DTD(document type definition)

Una definición de tipo de documento o DTD (document type definition) es una descripción de estructura y sintaxis de un documento XML.

Una DTD es un documento SGML que incluye las reglas sintácticas para un tipo de documento específico. Incluye los elementos que se permiten y sus atributos, así como reglas que afectan a la anidación de los primeros y a los valores de los segundos. Contrastando un documento con su DTD se puede comprobar si éste es válido o no.

Declaraciones de elemento

Declaraciones de elemento Para declarar los elementos, que pueden ser utilizados por los documentos que se ajusten a esa DTD, se utiliza la expresión:

<!ELEMENT NombreElemento Contenido>

Un elemento representa el contenido del elemento y puede ser:

Una lista de elementos secundarios: Cuando un elemento está formado por otros elementos, estos se declaran entre paréntesis separados por comas o por | si se trata de elementos alternativos:

<!ELEMENT planeta (nombre+, tipo, (orbita | satélite+), composición*)>

Cada subelemento puede llevar, además, los siguientes símbolos:

- **Sin símbolo:** El elemento secundario debe aparecer una sola vez.
- **Interrogación:** Puede aparecer una o ninguna vez.
- **Asterisco (*):** Puede aparecer cualquier número de veces.
- **Signo más (+):** Debe aparecer por lo menos una vez.

Datos de carácter. Cuando un elemento solo va a contener datos de carácter analizados sintácticamente, se utiliza la declaración (*#PCDATA*):

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

Es posible declarar elementos mixtos, es decir, que contengan datos de carácter y elementos secundarios:

```
<!ELEMENT capitulo (#PCDATA|subcapitulo)*>
```

Elemento vacío. Los elementos sin contenido se declaran:

`<!ELEMENT saltolinea EMPTY>`

Declaraciones de atributos

Los atributos se utilizan para especificar información adicional del elemento.

```
<libro>  
  <capitulo numero="1">  
    <subcapitulo>  
      :
```

Se declaran utilizando la siguiente sintaxis:

```
<!ATTLIST Elemento NomAtr Tipo Valpred>
```

NomAtr. Es el nombre que se le da al atributo.

Tipo. Es el tipo de atributo. Puede tomar uno de los siguientes valores:

- CDATA: Datos de caracteres no analizados sintácticamente.
- Lista de valores: Entre paréntesis y separados por |.

Ejemplo: (Este | Oeste | Sur | Norte)

ENTITY: Entidad binaria externa.

NOTATION: Notación declarada en la DTD.

ID: Identificador único.

Valorpred. Representa el valor predeterminado del atributo. Puede ser una de las siguientes posibilidades:

#REQUIRED: El atributo es obligatorio.

#IMPLIED: El atributo es opcional.

#FIXED valor: El atributo tiene ese valor fijo

```
<!ELEMENT addressbook (contact)+>
<!ELEMENT contact (name, address+, city, state, zip, phone, email, web, company)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT address (#PCDATA)>
<!ELEMENT city (#PCDATA)>
<!ELEMENT state (#PCDATA)>
<!ELEMENT zip (#PCDATA)>
<!ELEMENT phone (voice, fax?)>
<!ELEMENT voice (#PCDATA)>
<!ELEMENT fax (#PCDATA)>
<!ELEMENT email (#PCDATA)>
<!ELEMENT web (#PCDATA)>
<!ELEMENT company (#PCDATA)>
<!ELEMENT peliculas (pelicula)+>
<!ELEMENT pelicula(titulo, writer+, productor+, director+, actor*, comentarios?)>
<!ATTLIST pelicula
tipo (drama | comedy | adventure | sci-fi | mistery | horror | romance | documentary)
"drama"
clasificacion (G | PG | PG-13 | R | X) "PG"
review (1 | 2 | 3 | 4 | 5) "3"
año CDATA #IMPLIED>
<!ELEMENT titulo(#PCDATA)>
<!ELEMENT writer (#PCDATA)>
<!ELEMENT productor(#PCDATA)>
<!ELEMENT director (#PCDATA)>
<!ELEMENT actor (#PCDATA)>
<!ELEMENT comentarios(#PCDATA)>
```

Entidades

Representa una unidad de almacenamiento de datos que se utilizan en un documento.

Las entidades se encuentran definidas fuera del documento.

Para utilizar los datos de una entidad dentro de un documento se utiliza la referencia a entidad. Se trata de un nombre

asociado a los datos contenidos en la entidad.

Los datos contenidos en una entidad pueden ser:

- Una cadena de texto.
- Una sección de la DTD.
- Una referencia externa a un archivo que contiene texto XML.
- Una referencia externa a un archivo que contiene datos binarios.

Declaración de una entidad

Las entidades que van a ser utilizadas por los documentos XML deben ser declaradas en la DDT: **<!ENTITY NombreEntidad ContenidoEntidad>**

ContenidoEntidad puede ser:

Texto de sustitución de la entidad. Representa el contenido por el que se sustituirá la referencia a entidad dentro del documento:

<!ENTITY firma "Madrid a 23 de Julio de 2001">

Entidad Externa. Ubicación del archivo que contiene la entidad. Debe ir precedido por las palabras

SYSTEM, si el archivo se encuentra en el sistema local de archivos o en una red.

PUBLIC, si es un archivo de dominio público que se encuentra accesible públicamente. Se usa con SYSTEM

```
<!ENTITY imagen SYSTEM "anexo.xml" >
```

Referencia a entidades

Para hacer referencia a una entidad que ha sido declarada en una DTD, dentro de un documento XML, se utiliza la expresión: **&NombreEntidad;**

Existen entidades predefinidas que no hay que declarar:

& Ampersand(&) ' Apóstrofo(')

" Comillas("") > Signo mayor(>)

< Signo menor (<)

Entidades Parámetro

Son entidades que sólo pueden utilizarse dentro de una DTD:

```
<!ENTITY % patrón "nombre, dirección">  
<!ELEMENT usuario (%patrón;)>  
<ELEMENT tienda (%patrón;)>
```

Vocabularios XML

Conjunto de definiciones y reglas que indican cómo deben ser utilizados los elementos y cómo deben ser utilizados.

Permiten crear aplicaciones que “entiendan” un determinado tipo de documentos.

Se crean mediante las DTD's.

Existen gran cantidad de vocabularios ya diseñados, de interés para desarrolladores web.

Vocabularios XML más importantes

SVG (Gráficos de Vectores escalables): Permite describir imágenes basadas en vectores utilizando sintaxis XML.

SMIL (Lenguaje Sincronizado de Integración Multimedia): Define una serie de elementos que permiten crear presentaciones multimedia que combinan audio, vídeo, animación, imágenes y texto.