

Extensible Markup Language

Documentos XML



Componentes de un documento XML

2

- XML permite expresar información estructurada de la manera más abstracta y reutilizable posible.
 - La información estructurada se compone de partes bien definidas, y esas partes se componen a su vez de otras partes.
 - Estas partes se llaman *elementos*, y se las señala mediante *etiquetas*.
- Los documentos XML se componen de *etiquetas de marcado* y de *contenido*. Todo texto que no sea marcación constituye los datos de carácter del documento.



Componentes de un documento XML

3

- Los documentos XML pueden contener:
 - ✓ Elementos (*elements*)
 - ✓ Atributos (*attributes*)
 - ✓ Comentarios
 - ✓ Entidades (*entities*)
 - ✓ Instrucciones de procesamiento (*PI*)
 - ✓ Datos de tipo carácter (*CData*)
 - ✓ Texto XML y espacios en blanco
 - ✓ *DTD's*



Componentes de un documento XML

4

- Los documentos XML se dividen en dos grupos, documentos *bien formados* y *documentos válidos*.
 - ▣ Bien formados: Son todos los que cumplen las especificaciones del lenguaje respecto a las reglas sintácticas sin estar sujetos a unos elementos fijados en un DTD. De hecho los documentos XML deben tener una estructura jerárquica muy estricta y los documentos bien formados deben cumplirla.



Componentes de un documento XML

5

- Los documentos XML se dividen en dos grupos, documentos *bien formados* y *documentos válidos*.
 - ▣ Válidos: Además de estar bien formados, siguen una estructura y una semántica determinada por un DTD o un Esquema: sus elementos y sobre todo la estructura jerárquica que define el DTD/Esquema, además de los atributos, deben ajustarse a lo que el DTD/Esquema dicte.



Componentes de un documento XML

6

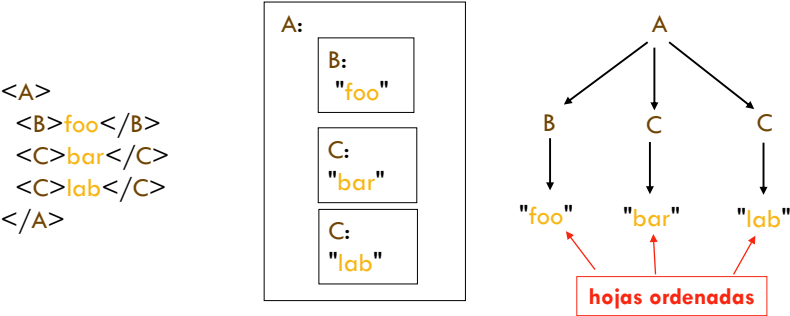
- Modelos de Instancia
 - ▣ En el estándar XML 1.0 no hay un modelo de datos explícito.
 - Sólomente marca la sintaxis de los documentos *bien formados* y de los *válidos* (a través de DTD's o Esquemas)
 - ▣ Modelo implícito:
 - *Contenedores anidados* ("cajas dentro de cajas")
 - *Árboles etiquetados y ordenados* (un modelo de datos semiestructurado)
 - Relacional, orientado a objetos, otros: fáciles de codificar



Componentes de un documento XML

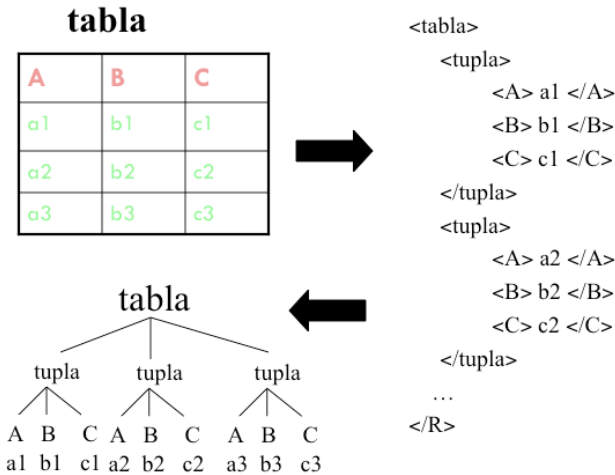
7

□ Modelo implícito:



Componentes de un documento XML

8



Componentes de un documento XML

10

● Elementos

- Son las etiquetas principales de la arquitectura de XML. La mayoría están orientados a contener datos u otros elementos.
- Identifican y determinan la naturaleza del contenido que encierran.
- Identifican componentes estructurales y/o semánticos.
- Su formato es el siguiente:

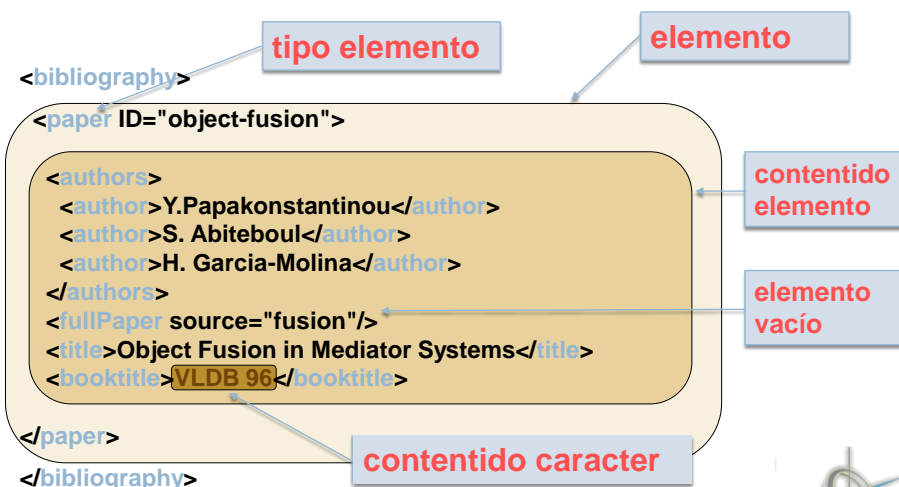
<etiqueta> contenido **</etiqueta>**

<etiqueta/> (para etiquetas vacías)



Componentes de un documento XML

11



Componentes de un documento XML

12

- Atributos
 - Son piezas de información, típicamente pequeñas, que están asociadas con los elementos XML
 - Aumentan la semántica de las etiquetas.
 - Especificar las propiedades / características de los elementos
 - Siempre aparecen en la etiqueta de inicio de un elemento.
 - Los valores son cadenas atómicas y han de ir encerrados por comillas dobles.



Componentes de un documento XML

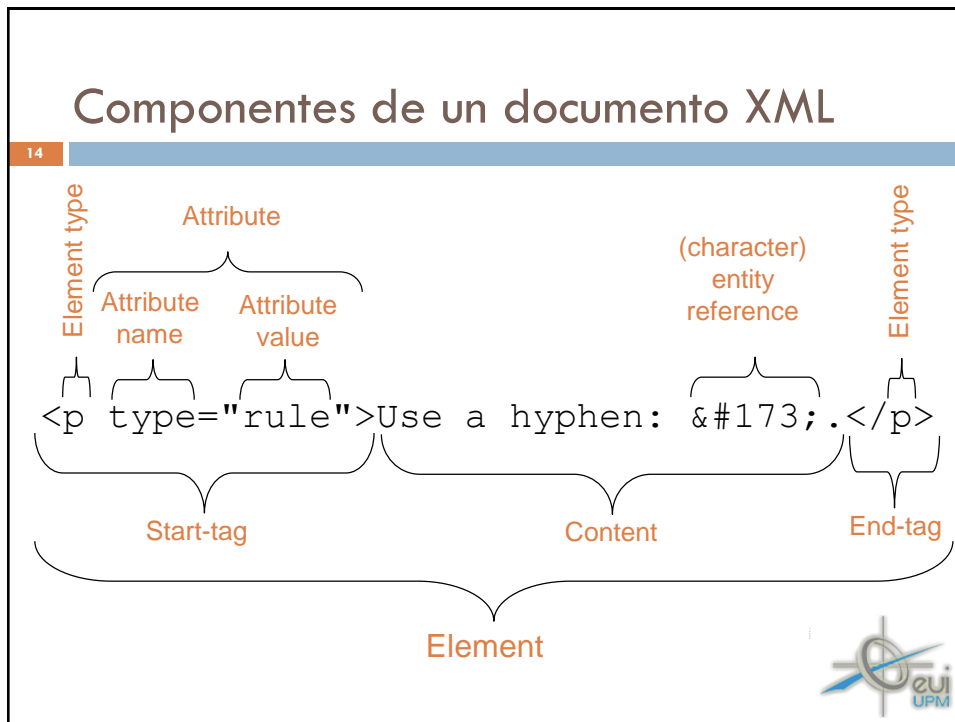
13

- Atributos

```
<etiqueta atributo_1="valor_1"
           atributo_n="valor_n"/>
```

```
<etiqueta atributo_1="valor_1"
           atributo_n="valor_n">
...</etiqueta>
```






Componentes de un documento XML

16

- Elementos son los mejores para:
 - Indicadores de jerarquía (padre/hijo)
 - Contenedores
 - Pasajes de texto
 - Secuenciación
 - Contenido sin restricciones
- Los atributos son los mejores para:
 - Modificación de la información
 - Metadatos
 - Enumeración
 - Valores restringidos o tipos



Creación de Documentos XML

17

- ¿Cuándo utilizar **atributos**?
 - Los atributos representan una forma de agregar metadatos a los elementos.
 - Los atributos no son extensibles.
 - Los atributos no poseen estructura.
 - Los atributos pueden provocar ambigüedad.
 - En general es necesario utilizar XSLT para mostrarlos en navegadores convencionales.
 - Por convenio, el texto debe aparecer en los elementos y los atributos deben contener información sobre el marcado o sobre el texto.



Componentes de un documento XML

18

- Entidades
 - Son contenedores de información
 - Interna (desde un único carácter, a una cadena) o externa (documentos, imágenes, sonido, videos...)
 - Pueden contener tanto datos analizados (*parsed*) como sin analizar (*unparsed*).
 - Están identificadas por un nombre.
 - Entidades **predefinidas**: utilizan para escapar los caracteres del marcado.

< “<” & “&” > “>” ' “” "e; “”
 - Entidades **generales**, utilizadas como mecanismos léxicos para la inclusión de información.



Componentes de un documento XML

20

• Comentarios

- Mantienen la misma forma que los comentarios definidos en HTML, es decir:

`<!-- esto es un comentario -->`

- No forman parte del contenido textual de los documentos XML.
- No se permite la anidación de comentarios.
- No pueden dividir etiquetas de un mismo elemento.
- No se pueden incluirse dentro de las etiquetas.



Componentes de un documento XML

21

• Secciones CDATA

- A veces un documento contiene gran cantidad de caracteres considerados especiales por el XML, a través de las secciones CDATA se pueden aislar dichos contenidos.
- Estas secciones indican al parser que ignore todos los caracteres de marcas contenidos en ellas.
- También permiten inclusiones condicionales.

`<![CDATA[contenido]]>`



Componentes de un documento XML

22

- Instrucciones de Procesamiento (PI's)
 - Son mecanismos usados para almacenar información específica de aplicaciones en un documento SGML.
 - Proveen de un mecanismo para poder “hablar” con otras aplicaciones.
 - No forman parte del contenido textual del documento XML, pero los parsers están obligados a pasarlas a las aplicaciones finales.

<?nombre datoPI?>



Componentes de un documento XML

23

- Documento de declaración de tipos – DTD
 - Proporcionan un mecanismo para definir reglas que nos permiten describir restricciones en la estructura lógica de los documentos XML.
 - Un documento XML es válido si tiene un asociada una declaración de tipo de documento o un esquema y es conforme con las restricciones que en ellos se expresan.



Componentes de un documento XML

24

- Documento de declaración de tipos
 - También nos permiten especificar las entidades a utilizar dentro del documento.
 - El DTD asociado a un documento debe aparecer referenciado en la segunda línea del documento XML de la siguientes formas:

`<!DOCTYPE nombre* [...]>`

`<!DOCTYPE nombre* SYSTEM|PUBLIC
"URI_fichero.dtd">`

`<!DOCTYPE nombre* SYSTEM|PUBLIC
"URI_fichero.dtd" [...]>`



Creación de Documentos XML

25

- Reglas de buena formación de un documento:
 - La declaración XML debe ser la primera línea del documento (opcional pero recomendable).

`<?xml version="1.0"?>`
 - Debe existir un elemento raíz que contiene al resto de elementos del documento.
 - Los elementos deben tener etiquetas de apertura y cierre si poseen contenido*.
 - Los elementos no pueden solaparse, pero sí anidarse.



Creación de Documentos XML

26

- Reglas de buena formación de un documento:
 - La declaración `<?xml ... ?>` es la forma en la que se identifica un documento XML. Su sintaxis completa es:

```
<?xml version="versión"
      [encoding = "juego_de_caracteres"]
      [standalone="yes|no"] ?>
```

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="yes" ?>
```



Creación de Documentos XML

27

- Reglas de buena formación de un documento, **nombres de elementos** :
 - Debe comenzar por una letra o un guión de subrayado.
 - Después puede continuar con cualquier combinación de letras, números, puntos y guiones de subrayado.
 - Dentro de un nombre de elemento no pueden aparecer espacios en blanco o cualquier otro tipo de carácter diferente de los indicados anteriormente.
 - Los nombres de elemento de XML distinguen entre mayúsculas y minúsculas.

```
<_elemento.01> ... </_elemento.01>
```



Creación de Documentos XML

28

- Reglas de buena formación de un documento, **nombres de atributos**:
 - Debe comenzar por una letra o un guión de subrayado y continuar con cualquier combinación de letras, números, puntos y guiones de subrayado.
 - No pueden contener espacios en blanco o cualquier otro tipo de carácter diferente de los anteriormente indicados.
 - Los nombres de atributo de XML distinguen entre mayúsculas y minúsculas.
 - Los valores deben aparecer entre comillas dobles.
 - No pueden existir dos atributos con el mismo nombre dentro de una misma etiqueta.

`<elemento _atributo.01="valor">... </elemento>`



Creación de Documentos XML

29

- Utilización de **espacios de nombres**
 - Un documento XML puede contener nombres de elementos o atributos procedentes de más de un vocabulario XML. Si a cada uno de estos vocabularios se le da un espacio de nombres, se resuelve la ambigüedad existente entre elementos o atributos que se llamen igual.
 - Los nombres de elementos dentro de un espacio de nombres deben ser únicos.



Creación de Documentos XML

31

● Utilización de espacios de nombres

```
<?xml version="1.0"?>
<cli:cliente xmlns:cli='http://miempresa.org/cliente'
             xmlns:ped='http://miempresa.org/pedido'>
  <cli:numero_ID>1232654</cli:numero_ID>
  <cli:nombre>Fulanito de Tal</cli:nombre>
  <cli:telefono>999999999</cli:telefono>
  <ped:pedido>
    <ped:numero_ID>6523213</ped:numero_ID>
    <ped:articulo>Caja de herramientas</ped:articulo>
    <ped:precio>187,90</ped:precio>
  </ped:pedido>
</cli:cliente>
```

