
Tema 3. Lenguajes de marcado

Ofimática Avanzada

Profesor: Víctor Fresno Fernández

Lenguajes de marcado

- Un “*Lenguaje de marcado*” o “*lenguaje de marcas*” se puede definir como una forma de codificar un documento donde, junto con el texto, se incorporan etiquetas, marcas o anotaciones con información adicional relativa a la estructura del texto, su presentación,...

Lenguajes de marcado

- **Anotación** (metadatos): información añadida al documento que no forman parte del texto en sí mismo
- **Lenguajes de marcado** (de anotaciones): conjunto de reglas que describen cómo deben realizarse anotaciones, bajo qué condiciones se permiten y su significado.

Lenguajes de marcado

- Los lenguajes de marcado permiten hacer explícita la estructura de un documento, su contenido semántico o cualquier otra información lingüística o extralingüística que se quiera hacer patente

Ejemplos:

- `<fecha>El día <date>22/11/2006 </date>tuvo lugar ...</fecha>`
- `<subr>Móstoles</subr>a<subr>2 de noviembre</subr>`

Lenguajes de marcado

- Los lenguajes de marcado se pueden clasificar en:
 - **Procedimental:**
 - Describen operaciones tipográficas
 - **Estructural:**
 - Describen la estructura lógica de un documento, pero no su tipografía
 - **Híbrido:**
 - Combinación de ambos
- Las hojas de estilo o lenguajes de transformación permiten la “traducción” de anotaciones de tipo estructural a anotaciones de carácter tipográfico.

Lenguajes de marcado

- Otra posible clasificación sería:
 - **De presentación:**
 - Indica el formato del texto (información para el maquetado).
 - **De procedimientos:**
 - Orientado también a la presentación pero, en este caso, se indican los procedimientos que deberá realizar el SW de representación.
 - **Descriptivo o semántico:**
 - Describen las diferentes partes en las que se estructura el documento pero sin especificar cómo deben representarse.

Lenguajes de marcado

- Algunos lenguajes de marcado específicos:
 - **Documentación electrónica**
 - RTF
 - TeX
 - Wikitexto
 - DocBook
 - **Tecnologías de internet**
 - HTML, XHTML
 - RDF (recurso-propiedad(relación)-valor)
 - RSS
 - **Otros lenguajes especializados**
 - MathML
 - VoiceXML
 - SVG
 - MusicXML

Lenguajes de marcado

- Haciendo un poco de historia..
 - La idea de introducir un marcado en un documento electrónico viene heredada de la corrección manual de manuscritos
 - En la década de los 60' se empieza a desarrollar la idea de separar presentación y estructura.
 - Desde IBM se implusa la creación del lenguaje GML, que resultó se la semilla de una versión posterior estandarizada: SGML.
 - La potencia de SGML implica una dificultad en su aprendizaje y uso.
 - El HTML se crea a partir del SGML.
 - XML surge como respuesta al desorden que supuso el rápido crecimiento del HTML.

Lenguajes de marcado

- **SGML** (Standard Generalized Markup Language, 1986): es un metalenguaje que permite definir lenguajes de marcado.
 - Especifica la sintáxis para la inclusión de marcas en los textos, así como la sintáxis del documento que especifica qué etiquetas están permitidas y dónde: el Document Type Definition.
 - La definición de la estructura y el contenido de un tipo de documento se realiza por medio de su DTD (Document Type Definition)

Lenguajes de marcado

- **Ventajas de SGML:**
 - Reutilización de los datos
 - Integridad y mayor control sobre los datos
 - Portable
 - Flexible
 - Perdurabilidad de la información
- **Inconvenientes de SGML:**
 - Alta complejidad

Ejemplo de SGML

```
<EMail>
  <sender>
    <person>
      <firstname> Karen </firstname>
      <lastname> Lemone </lastname>
    </person>
  </sender>
  <receiver>
    <person>
      <distributionList> cs525@cs.com </distributionList>
    </person>
  </receiver>
  <contents>¿no es sencillo?</contents>
</EMail>
```

Lenguajes de marcado

- **HTML:** lenguaje de marcado definido en SGML
 - **Origen:** 1989 en el Laboratorio Europeo de Física de Partículas (CERN)
 - **Objetivo inicial:** presentar información estática.

Jugó un papel fundamental en el crecimiento de Internet.
- Presenta limitaciones relacionadas con:
 - Tratamiento de información dinámica.
 - No es un metalenguaje, por lo que dispone de un número fijo de etiquetas.
 - Su vocabulario es muy limitado.

Lenguajes de marcado

- **Ventajas de HTML:**
 - Es muy simple y sencillo de aprender y usar.
 - No requiere herramientas especiales.
 - Está muy difundido.

Lenguajes de marcado

- **Inconvenientes de HTML:**
 - Carecer de chequeo sintáctico.
 - Carecer de estructura lógica.
 - Estar orientado fundamentalmente a la representación de los datos y no a su estructura.
 - Carecer de una semántica estándar.
 - No ser adecuado para el intercambio de datos.
 - No ser extensible.
 - No permitir la reutilización de la información.

Ejemplo de HTML

```
<html>
<head>
  <meta http-equiv="content-type" content="text/html" charset="ISO-8859-1"/>
  <meta name="generator" content="Adobe GoLive 5"/>
  <title>Archivo L&eacute;ame de Adobe Acrobat para Windows</title>
</head>
<body bgcolor="#ffffff">
  <p></p>
  <p>
    <b><font size="+1">21 de septiembre de 2003</font></b>
  </p>
  <p><b><font size="+2">Archivo L&eacute;ame de Adobe Acrobat para
  Windows</font></b></p>
  <p>Bienvenido al archivo L&eacute;ame de Adobe&reg; Acrobat&reg; 5.0.5.
  Si lo desea, puede acceder al <a href="http://www.adobe.com/supportservice/">
  soporte t&eacute;cnico</a><br/> </p>
  <p> Este archivo está dividido en los siguientes apartados:<br/>
  .... </p>
</body>
</html>
```

Lenguajes de marcado

- **XML** (Extensible Markup Language): forma restringida de SGML optimizada para su utilización en Internet.
 - **Origen:** 1996 World Wide Web Consortium (W3C)
- **Objetivos iniciales:**
 - Lenguaje estructurado, extensible y que se pueda validar.
 - Permitir la transmisión de información realmente estructurada.

Lenguajes de marcado

- **Características de XML:**

- Es un subconjunto de SGML (toma el 80% de sus ventajas y le resta el 20% de complejidad).
- Es simple de usar y se basa en etiquetas de texto.
- Es una tecnología madura puesto que se basa en SGML.
- Soporta Unicode.
- Se orienta a los datos, su semántica y no a la representación.
- Se está convirtiendo en el lenguaje de Bases de Datos de la Web.
- Permite un fácil intercambio de información entre aplicaciones.

Lenguajes de marcado

- **Características de XML:**
 - Al tratarse de un metalenguaje tiene un vocabulario extensible:
 - Permite definir lenguajes de marcado por medio de DTD's (Document Type Definition) o de XML-Schemas
 - Sirve para representar datos estructurados en un fichero de texto.
 - Usa etiquetas para delimitar los datos pero deja su interpretación a la aplicación que lee el código XML.

Lenguajes de marcado

- **Ventajas:**
 - Tecnologías asociadas:
 - XML (Estructura de los datos)
 - XSL= XSLT+XSL-FO's + XPath (hojas de estilo)
 - XLL = XLink + XPointer+ Xpath (hiperenlaces)
 - XQL (consultas a bases de datos)
 - DOM (Document Object Model)
 - SAX (Simple Api for XML)
 - ...
 - La Web Semántica se está construyendo sobre XML.

Ejemplo de XML

```
<?xml version="1.0" standalone="no"?>
<!DOCTYPE movies SYSTEM "Movies.dtd" [<!NOTATION JPEG SYSTEM "Iexplore.exe">
<!ENTITY raposter SYSTEM "RAPoster.jpg" NDATA JPEG> <!ENTITY mrposter SYSTEM "MRPoster.jpg" NDATA
JPEG>]>
<movies>
  <movie rating="PG-13" review="5" type="comedy" year="1987">
    <title>Raising Arizona</title>
    <writer>Ethan Coen</writer>
    <writer>Joel Coen</writer>
    <producer>Ethan Coen</producer>
    <director>Joel Coen</director>
    <actor>Nicolas Cage</actor><actor>Holly Hunter</actor><actor>John Goodman</actor>
    <poster image="raposter"/>
    <comments>A classic one-of-a-kind screwball love story.</comments>
  </movie>
  <movie rating="R" review="5" type="comedy" year="1988">
    <title>Midnight Run</title>
    <writer>George Gallo</writer>
    <producer>Martin Brest</producer>
    <director>Martin Brest</director>
    <actor>Robert De Niro</actor><actor>Charles Grodin</actor>
    <poster image="mrposter"/>
    <comments>The quintessential road comedy.</comments>
  </movie>
</movies>
```

Lenguajes de marcado

- **Diferencia entre XML (estructura) y HTML(presentación):**

```
<cancion>
  <titulo>Hot cop</titulo>
  <compositor> Jacques Morali</compositor>
  <compositor> Henry Beolo </compositor>
  <compositor> Victor Wills </compositor>
  <productor> Jacques Morali </productor>
  <compañia> PolyGram </compañia>
  <duracion> 6:20 </duracion>
  <año> 1978 </año>
  <artista> Village People </artista>
</cancion>
```

```
<html>
<body>
  <h3>Hot Cop</h3>
  <p>by Jacques Morali,
    Henry Beolo, and Victor Wills</p>
  <ul>
    <li>productor: Jacques Morali </li>
    <li>compañia: PolyGramRecords</li>
    <li>duracion: 6:20</li>
    <li>año: 1978 </li>
    <li>artista: Village People</li>
  </ul>
</body>
</html>
```