
UT1. Lenguajes de Marcas

1. Introducción a los lenguajes de marcas
 2. Clasificación de los lenguajes de marcas
 3. Orígenes. SGML(Standard Generalized Markup Language)
 4. Organizaciones desarrolladoras:
 - ISO (International Standard Organization)
 - W3C (Word Wide Web Consortium)
 5. Utilización de lenguajes de marcas en entornos Web
-

¿Por qué surgen los lenguajes de marcas?

- ❑ Uno de los primeros problemas que se plantearon en informática era como traspasar o enviar información entre equipos.
 - ❑ Para solventar esta situación se decidió codificar la información, en formato Solo texto, para su envío.
 - ❑ Una vez solucionado el paso de información, se planteó la posibilidad de enviar junto a la información el formato correspondiente a la misma.
 - ❑ Una de las soluciones propuestas fue la de crear una serie de marcas que insertadas en la información a enviar establecieran dicho formato.
 - ❑ A estas marcas se les llama *Etiquetas* y al conjunto de las mismas que determinan el formato se le conoce como un lenguaje de marcado.
-

Definición de lenguaje de marcas

Un lenguaje de marcas es una forma de codificar un documento que, junto con el texto, incorpora etiquetas o marcas que contienen información adicional acerca de la estructura del texto o su presentación.

Ejemplos de lenguajes de marcas:

- ❑ Documentación electrónica: RTF, TeX, DocBook,...
 - ❑ Tecnologías de internet: HTML, XHTML, XML, RDF, RSS
 - ❑ Otros lenguajes especializados: MathML, voiceXML, SGV, MusicXML
-

Clasificación de los lenguajes de marcas

- **Orientados a la presentación.** En ellos al texto común se añaden palabras encerradas en símbolos especiales que contienen indicaciones de formato que permiten a los traductores de este tipo de documentos generar un documento final en el que el texto aparece con el formato indicado.
Es el caso de HTML y de los procesadores de texto tradicionales como puede ser Microsoft Word en los que al texto del documento se le acompaña de indicaciones de formato (como negrita, cursiva,...)
- **Orientados a la descripción.** En ellos las marcas especiales permiten dar significado al texto pero no indican cómo se debe presentar en pantalla el mismo. Sería el caso de XML o de SGML en el que la presentación nunca se indica en el documento, simplemente se indica una semántica de contenido que lo hace ideal para almacenar datos.
- **Orientados a procedimientos.** Se trata de documentos en los que hay texto marcado especialmente que en realidad se interpreta como órdenes a seguir y así el archivo en realidad contiene instrucciones a realizar con el texto. Es el caso de LaTeX o PostScript.

Orígenes. SGML

En los años 60 IBM creó un lenguaje de marcas denominado GML (Generalized markup Language). Más tarde GML pasó a manos de ISO y se convirtió en SGML (ISO 8879). Esta norma es la que se aplica desde entonces a todos los lenguajes de marcas.

SGML es un metalenguaje que permite definir lenguajes de marcado:

- ❑ Especifica la sintaxis para la inclusión de marcas en los textos, así como la sintaxis del documento que especifica qué etiquetas están permitidas y dónde: el Document TypeDefinition (DTD).
- ❑ La definición de la estructura y el contenido de un tipo de documento se realiza por medio de su DTD.

Orígenes. SGML

Ventajas de SGML:

- Reutilización de los datos.
- Integridad y mayor control sobre los datos.
- Portable.
- Flexible.
- Perdurabilidad

Inconvenientes de SGML:

- Alta complejidad.
-

Organizaciones desarrolladoras: W3C.

W3C, World Wide Web Consortium, es un consorcio internacional que produce recomendaciones para la Web. Fue creada el 1 de octubre de 1994 por Tim Berners-Lee, está formado por:

- ❑ Miembros del W3C. A abril de 2010 contaba con 330 miembros
- ❑ Equipo W3C (W3C Team) 65 investigadores y expertos de todo el mundo
- ❑ Oficinas W3C (W3C Offices). Centros regionales establecidos en Alemania y Austria (oficina conjunta), Australia, Benelux (oficina conjunta), China, Corea del Sur, España, Finlandia, Grecia, Hong Kong, Hungría, India, Israel, Italia, Marruecos, Suecia y Reino Unido e Irlanda (oficina conjunta)

Organizaciones desarrolladoras. ISO.

ISO, Organización Internacional para la Estandarización nacida tras la Segunda Guerra Mundial, es una red de los institutos de normas nacionales de 163 países, sobre la base de un miembro por país, con una Secretaría Central en Ginebra (Suiza) que coordina el sistema.

Las normas desarrolladas por ISO son voluntarias y se conocen como normas ISO y su finalidad es contribuir con normas comunes al desarrollo y a la transferencia de tecnologías. Algunas normas ISO relacionadas con la informática:

- ❑ ISO 8859-1, que permite codificar las lenguas originales de Europa occidental, como el español.
- ❑ ISO 9899 - Lenguaje de programación C.
- ❑ ISO 8879- Lenguaje SGML.

Utilización de lenguajes de marcas en entornos Web

HTML:

- ❑ 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.

Ventajas de HTML:

- ❑ Es muy simple y sencillo de aprender y usar.
 - ❑ No requiere herramientas especiales.
 - ❑ Está muy difundido.
-

Utilización de lenguajes de marcas en entornos Web

Inconvenientes de HTML:

- ❑ La mayoría de etiquetas HTML no son semánticas; es decir, no sirven para decir el tipo de contenido que tenemos sino para indicar el formato.
- ❑ HTML es un lenguaje rígido, no es extensible. Es decir no podemos añadir etiquetas ya que ningún navegador las reconocerá. Cada vez que se decide añadir hay que cambiar el estándar y los navegadores se deben de adaptar a los cambios.
- ❑ Requiere de arreglos “extraños” para añadir potencia y funcionalidad, por lo que los diseñadores tienden a incrustar dentro del código HTML código de lenguajes como PHP o Javascript que dificultan su legibilidad y comprensión.

Ejemplo de HTML

```
<!DOCTYPE html>
```

```
<html>
```

```
  <head>
```

```
    <meta http-equiv="content-type" content="text/html; charset=utf-8">
```

```
    <title>HTML</title>
```

```
  </head>
```

```
  <body>
```

```
    <h1>CURSO DE HTML</h1>
```

```
    <p>Tema 1</p>
```

```
  </body>
```

```
</html>
```

Utilización de lenguajes de marcas en entornos Web

- ❑ **XML** es un lenguaje de marcas que se ha estandarizado y se ha convertido en uno de los formatos más populares para intercambiar información.
Al crear XML se plantearon estos objetivos:

- ❑ Debía de ser similar a HTML (de hecho se basa en el lenguaje SGML base para el formato HTML)
- ❑ Debía de ser extensible, es decir que sea posible añadir nuevas etiquetas sin problemas. Esa es la base del lenguaje XML.
- ❑ Debía de tener unas reglas concisas y fáciles, además de estrictas.
- ❑ 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.
- ❑ Debía ser fácil de leer por los humanos y fácil crear procesadores XML software (llamados parsers)

Utilización de lenguajes de marcas en entornos Web

Algunos de los lenguajes estándares de marcado basados en XML:

- ❑ **XHTML**. Adaptación de HTML a las normas de XML.
- ❑ **RSS**. (*Really Simple Syndication, aunque hay otras interpretaciones de los acrónimos*) Para producir contenidos sindicables, se utiliza fundamentalmente para producir noticias. Es una de las aplicaciones XML más utilizada
- ❑ **SMIL**. *Synchronized Multimedia Integration Language, Lenguaje sincronizado de integración multimedia. Utilizado para producir presentaciones de TV en la web, fundamentalmente.*
- ❑ **MathML**. Pensado para representar documentos con expresiones matemáticas.
- ❑ **VoiceXML**. Se utiliza para representar diálogos vocales.

Utilización de lenguajes de marcas en entornos Web

Tecnologías relacionadas con XML:

- ❑ **DTD.** Permite aplicar validación avanzada a XML
- ❑ **XML Schema.** La función que cumple esta tecnología es la misma que la anterior, la diferencia está en que este lenguaje es un lenguaje XML.
- ❑ **Namespacing,** espacios de nombres. Permite conseguir nombres de elementos que carecen de ambigüedad. Es decir nombres únicos.
- ❑ **XPath.** Lenguaje de consulta que permite seleccionar o acceder a partes de un documento XML.
- ❑ **XSLT.** Sirve para lo mismo que CSS, sólo que posee más posibilidades.
- ❑ **XQuery.** Permite consultar datos de los documentos XML.
- ❑ **DOM.** Document Object Model, permite acceder a la estructura jerárquica del documento normalmente para utilizarla dentro de un lenguaje de programación
- ❑ **SAX.** Simple API for XML, permite el uso de herramientas para acceder a la estructura jerárquica del documento XML, se usa mucho en Java.

Ejemplo de XML

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<librosDepart>
  <libro>
    <titulo>XHTML y CSS </titulo>
    <autor>Elizabeth Castro</autor>
    <editorial>Anaya Multimedia</editorial>
    <fecha-compra>2010-1-9</fecha-compra>
    <comentario>nivel básico</comentario>
  </libro>
  <libro>
    <titulo>UNIX</titulo>
    <autor>Kaare Chistian</autor>
    <descripcion>Introducción al s. operativo Unix</descripcion>
    <fecha-compra>2000-22-2</fecha-compra>
    <comentario>nivel intermedio</comentario>
  </libro>
</librosDepart>
```

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

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

```
<librosDepart>
```

```
  <libro>
```

```
    <titulo>XHTML y CSS </titulo>
```

```
    <autor>Elizabeth Castro</autor>
```

```
    <editorial>Anaya M.</editorial>
```

```
    <fecha-compra>2010-1-9</fecha-compra>
```

```
    <comentario>nivel básico</comentario>
```

```
  </libro>
```

```
  <libro>
```

```
    <titulo>UNIX</titulo>
```

```
    <autor>Kaare Chistian</autor>
```

```
    <descripcion>S.O. Unix</descripcion>
```

```
    <fecha-compra>2011-2-2</fecha-compra>
```

```
    <comentario>nivel alto</comentario>
```

```
  </libro>
```

```
</librosDepart>
```

```
<!DOCTYPE html>
```

```
<html>
```

```
  <head>
```

```
    <meta http-equiv="content-type"  
content="text/html; charset=utf-8">
```

```
    <title>HTML</title>
```

```
  </head>
```

```
  <body>
```

```
    <h1>CURSO DE HTML</h1>
```

```
    <p>Tema 1</p>
```

```
  </body>
```

```
</html>
```


Tecnologías Web

