

INTRODUCCIÓN A LOS LENGUAJES DE MARCAS

LMSGI

ÍNDICE

- Codificación de la información.
- Lenguajes de marcas.
- HTML.
- XML.

CODIFICACIÓN DE LA INFORMACIÓN

(1/2)

- Los ordenadores se comunican en lenguaje binario (0 y 1).
 - Las personas CODIFICAMOS la información (la convertimos a lenguaje máquina) para que los ordenadores puedan procesarla.
- Por lo general, todos los datos del ordenador son BINARIOS.
 - Pero a nivel formal, curiosamente se considera BINARIO a la información que NO sea de TEXTO.
- Los datos se organizan en archivos (o ficheros).
 - Estructuras de información que suelen tener <nombre>.<extensión>

CODIFICACIÓN DE LA INFORMACIÓN (2/2)

- **Planos.**

- Ficheros de texto
- ASCII, UNICODE
 - Estándares de codificación internacionales para que no haya problemas entre diferentes países.
- Tipos de ficheros planos:
 - Configuración: .ini .inf .conf
 - Código fuente: .sql .c .java
 - Página Web: .html .php .asp .xml
 - Enriquecidos: .rtf .ps. tex

- **Binarios.**

- No son de texto.
- Necesitan un formato para interpretarse.
 - Imagen: jpg .gif .bmp
 - Vídeo: .mpg .mov .avi
 - Comprimidos: .zip .gz .rar .tar
 - Ejectables: .exe .com .cgi
 - Procesadores de texto: .doc .odt

LENGUAJES DE MARCAS (1/6)

¿Por qué los lenguajes de marcas?

- Antiguamente los ficheros planos solo transmitían información sin formato ni características especiales, solo era un texto “básico”.
 - Ej: .txt
- La alternativa, los binarios, que sí permitían formatos, requerían de software específico para su decodificación.
 - Ej: .exe
- La solución fue darle formato mediante **etiquetas** a los **ficheros planos**.
 - Esto dio origen a los **lenguajes de marcas**.
 - Siguen siendo ficheros planos, pero poseen etiquetas que el ordenador puede entender para dar un formato (tamaños, colores, tipo de letra...)
 - Ej: .html

LENGUAJES DE MARCAS (2/6)

Algo de historia:

- **Procesadores de texto.** Primer software en añadir marcas para dar info extra al texto plano.
- **Marking up.** Anotaciones que se hacían a los documentos. Las anotaciones incluidas en los documentos de texto han dado lugar a los lenguajes de marcas.
- **Charles Goldfarb.** Padre de los lenguajes de marcas. Ayudó a realizar el GML de IBM. Ideó posteriormente el SGML.
- **TeX y LaTeX.** En los 70 Donald Knuth crea TeX para documentos científicos. Unos 300 comandos. Leslie Lamport define LaTeX como evolución de TeX.
- **RTF** acrónimo de Rich Text Format. Microsoft en 1987. WordPad lo utiliza.
- **SGML.** Versión de GML. Estándar de documentos de texto con etiquetas de marcado. Padre del XML y base del HTML
- **PostScript.** John Warnock lo inicia en 1976. En 1985 funda Adobe donde lo desarrolla.
- **HTML.** Tim Bernes Lee a partir del SGML. Su éxito radica en el de Internet. Aparecen los Navegadores para mostrar formatos más avanzados y visuales.
- **XML.** En 1998 el W3C (World Wide Web Consortium) publica el XML (eXtendedMarkup Language). Más sencillo que SGML, más potente que HTML. Creado para el intercambio de información.

LENGUAJES DE MARCAS (3/6)

Tipos de Lenguajes de marcas:

- **Orientados a la presentación.** Al texto se añaden marcas que indican cómo presentarlo. **Procesadores de Texto y HTML.**
- **Orientados a la descripción.** Las marcas dan significado al texto pero no indican como presentarlo. Ej.- Indican si el texto es un nombre , una matrícula de coche. Ideal para almacenar datos. **SGML, XML**
- **Orientados a procedimientos.** Las marcas se interpretan como órdenes. El archivo contiene instrucciones a realizar con el texto. **LaTeX o PostScript.**

LENGUAJES DE MARCAS (4/6)

Organizaciones Desarrolladoras de Lenguajes de Marcas:

- **Organización Internacional para la Estandarización (ISO).**
 - Se crea en 1947. 163 países. Sede en Ginebra
 - Normas voluntarias y no obligatorias.
 - En 1986 publica SGML (Standard Generalized Markup Language)
- **World Wide Web Consortium (W3C).**
 - Se crea en 1994 por Tim Berbers-Lee
 - Función: Tutelar el crecimiento y organización de la web
 - Normalizó el HTML
 - En 1998 desarrolla **XML**. Sintaxis para compatibilizar los Leng.Marcas

LENGUAJES DE MARCAS (5/6)

Ejemplo de código LaTeX:

- `\documentclass[12pt]{article}`
- `\usepackage{amsmath}`
- `\title{\Ejemplo}`
- `\begin{document}`
- Este es el texto ejemplo de `\LaTeX{}`
- Con datos en `\emph{cursiva}` o `\textbf{negrita}`.
- Ejemplo de fórmula
- `\begin{align}`
- $E = mc^2$
- `\end{align}`
- `\end{document}`

Este es el texto ejemplo de \LaTeX
Con datos en *cursiva* o **negrita**. Ejemplo de fórmula

$$E = mc^2$$

LENGUAJES DE MARCAS (6/6)

Ejemplo de PostScript:

- %!PS
- /Courier % Elige el tipo de letra
- 20 selectfont % Establece el tamaño de la letra y la toma como el tipo de letra en uso
- 72 500 moveto % Coloca el cursor en las coordenadas 72, 500 (contando los píxeles desde % la esquina izquierda de la página)
- (Hola mundo!) show % Escribe el texto entre paréntesis,
- showpage % Imprime el resultado

HTML (1/2)

- Es un lenguaje de marcas de hipertexto para la elaboración de páginas web.
- **Utiliza Etiquetas, Elementos y Atributos:**
 - **Etiqueta**
 - Texto entre < >. Existen etiquetas de inicio y de fin.
 - Ej. <head> , </head>; <h1> , </h1> <p>, </p>
 - **Elementos.**
 - Constan de la etiqueta de inicio de la de fin y de todo aquello que se encuentra entre ellas.
 - Ej.- Este texto aparecerá con letra Arial en la pantalla.
 - Algunos elementos no tienen contenido. Se les denomina elementos vacíos y no llevan etiqueta de fin: Ej. <hr>
 - **Atributo.**
 - Par nombre-valor dentro de la etiqueta de inicio.
 - Indican las propiedades que pueden llevar asociadas un elemento
 - Ej.- family="Arial" href="http://www.google.es"

HTML (2/2)

```
<html>
<head>
<meta http-equiv="Content-Type" content="text/html;
charset=utf-8">
</meta>
</head>
<body>
<b>Este texto aparecerá en negrita en la pantalla.</b>
<i>Este texto aparecerá en cursiva en la pantalla.</i>
<font family="Arial">Este texto aparecerá con letra Arial en
la pantalla. </font>
<p>Este texto forma parte del párrafo 1. </p>
<p>Este texto forma parte del párrafo 2. Por eso está en
otra línea.</p>
<h1>Encabezado de nivel 1.</h1>
<h2>Encabezado de nivel 2.</h2>
<a href="http://www.google.es">Texto del enlace</a> <!--
esto es un comentario -->
</body>
</html>
```

Este texto aparecerá en negrita en la pantalla.

Este texto aparecerá en cursiva en la pantalla.

Este texto aparecerá con letra Arial en la pantalla.

Este texto forma parte del párrafo 1.

Este texto forma parte del párrafo 2. Por eso está en otra línea.

Encabezado de nivel 1.

Encabezado de nivel 2.

[Texto del enlace](http://www.google.es)

XML (1/4)

- Es un metalenguaje que permite definir lenguajes de marcas.
- Para crearlo se puede utilizar cualquier editor de texto (notepad++, vs code, atom...)
- **Características destacables:**
 - Proviene de SGML y es compatible con él.
 - Representa datos de forma estructurada (jerárquica).
 - Simple y fácil de leer.
 - Gramática de OBLIGADO CUMPLIMIENTO.
- **Usos:**
 - Contenido web. Podría reemplazar a HTML (XML tiene una sintaxis rígida y HTML no).
 - Intercambio de información entre aplicaciones.
 - Computación distribuida. Intercambio de información a través de redes.
 - Información empresarial. Facilita la estructuración de datos empresariales.

XML (2/4)

Lenguajes basados en XML:

- **XHTML**. Adaptación de HTML a las normas de XML.
- **RDF**, Resource Description Format, Sirve para desarrollar documentos que describan recursos. Se basa en los modelos conceptuales como el modelo entidad/relación. Actualmente se utiliza para describir recursos web.
- **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.
- **SVG**. Scalable Vector Graphics, gráficos de vectores escalables. Permite definir imágenes vectoriales pensadas para ser publicadas en una página web.
- **MathML**. Pensado para representar documentos con expresiones matemáticas.
- **VoiceXML**. Se utiliza para representar diálogos vocales.
- **OSDF**. Open Software Description Format, basado en especificaciones de las empresas Marimba y Microsoft, sirve para describir la tecnología y los componentes con los que se ha desarrollado un determinado software, a fin de facilitar el proceso de instalación.

XML (3/4)

Otras tecnologías relacionadas que utilizan XML:

- **DTD.** Nos permite componer nuestras propias estructuras y nombres de elementos para nuestros documentos.
- **XML Schema.** Misma función que DTD pero es un lenguaje en XML.
- **Namespacing,** espacios de nombres. Nos permite crear documentos más consistentes mediante el uso de múltiples vocabularios en un mismo documento.
- **XPath.** Lenguaje de consulta que permite seleccionar o acceder a partes de un documento XML.
- **CSS.** (Cascading Style Sheets) Permiten dar formato al documento.
- **XSL.** (extensible Style Sheet Language) Como CSS con más posibilidades.
- **XSLT.** Para transformar nuestros documentos de un tipo a otro.
- **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.

XML (4/4)

XML

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE note SYSTEM "Note.dtd">
<note>
<to>Tove</to>
<from>Jani</from>
<heading>Reminder</heading>
<body>Don't forget me this weekend!</body>
</note>
```

NOTE.DTD

```
<!DOCTYPE note
[
<!ELEMENT note (to,from,heading,body)>
<!ELEMENT to (#PCDATA)>
<!ELEMENT from (#PCDATA)>
<!ELEMENT heading (#PCDATA)>
<!ELEMENT body (#PCDATA)>
]>
```

!Doctype → define elemento raíz

!Element → define los elementos (estructura árbol)

¿PREGUNTAS?

INTRODUCCIÓN A LOS LENGUAJES DE MARCAS

LMSGI