

UT01. INTRODUCCIÓN A LOS LENGUAJES DE MARCAS

Curso: 2DAW-CIFP Carlos III-Cartagena

Lenguajes de Marcas y sistemas de gestión de información

Índice

1. Conceptos y ventajas
2. SGML. El origen
3. Características de los lenguajes de marcas
4. Clasificación de los lenguajes de marcas
5. Organizaciones y estándares
6. Introducción a los principales lenguajes de marcas
 1. HTML
 2. XML



1. Conceptos y ventajas

Conceptos y ventajas

- Aplicación de **formato** → **marcas**
- Intérprete → navegador
- Programar
- **Ejemplo.** Guardar el siguiente código como **saludo.txt** y volver a abrir. Guardar ahora como **saludo.html** y volver a abrir

```
<h1>Hola, este texto aparece más grande</h1>  
<h3>Este texto aparece más pequeño, adiós</h3>
```

Hola, este texto aparece más grande

Este texto aparece más pequeño, adiós

Conceptos y ventajas

- Escribir código html, ¿es programar?
- Usuarios finales
- Diferencias en los navegadores. Instalar Chrome, Firefox, Opera
- Diferencias en los diferentes dispositivos
- CSS, Lenguaje de estilos.
 - Independencia de la web

Conceptos y ventajas

1.2. <http://www.w3c.es> y usa las vistas de la parte superior de la página e indica las diferencias que observes entre ellas.



The screenshot shows the W3C España website. At the top, there is a navigation bar with the W3C logo and the text "España". To the right of the logo, there are three view options: "Vistas: escritorio", "móvil", and "impresión". The "móvil" option is highlighted with a red circle. To the right of these options, there is a dropdown menu labeled "El W3C por Región" and a search bar with the Google logo.

Below the navigation bar, there are four main sections: "ESTÁNDARES", "PARTICIPAR", "UNIRSE", and "SOBRE EL W3C".

On the left side, there is a sidebar with the following sections:

- DIVULGACIÓN**
 - W3DEVOCAMPUS
 - Documentos y Guías
 - Accesibilidad
 - Traducciones [inglés]
- ESTÁNDARES**
 - Diseño Web y Aplicaciones [inglés]
 - Arquitectura Web [inglés]
 - Web Semántica [inglés]
 - Tecnología XML [inglés]
 - Web de los Servicios [inglés]
 - Web de los Dispositivos [inglés]
 - Navegadores y Herramientas

The main content area features a news item titled "Nueva Suite de Validación del W3C para mejorar la calidad de la Web" dated 18 Septiembre 2013. The article includes the W3C Validator Suite logo and text describing the new premium validation service. Below the article, there is a section titled "Día del W3C en Uruguay. ¡Inscríbete ya!" dated 11 Septiembre 2013.

On the right side, there is a section titled "VALIDADORES Y SOFTWARE [inglés]" featuring the W3C Validator Suite logo. Below this, there is a section titled "TESTIMONIOS MIEMBROS W3C" featuring the Zaragoza Ayuntamiento logo and text about their commitment to accessibility.



2. SGML. El origen

Años 60

IBM → lenguaje de marcas GML
(almacenamiento de información
de diferentes plataformas)

Marcas descriptivas de
estructura del documento: texto,
listas, tablas



1986

GML pasa a ISO y se convierte
en SGML (GML estándar), libre
y de código abierto

Ejemplo 1: Los módulos del ciclo ASIR

Vocabulario: ASIR, módulo, título, contenido, unidad.

Reglas: **ASIR** contiene varios **módulos**, un **módulo** tiene un elemento simple **título** y un elemento **contenido**, **contenido** tiene varias **unidades** y toda **unidad** debe estar en un **contenido**, las **unidades** son **texto simple**, detrás de cada **unidad** solo puede ir otra **unidad** o **fin de contenido**, detrás de un **módulo** solo puede ir otro **módulo** o **fin de ASIR**.

Ejemplo 1: Los módulos del Ciclo ASIR

Vocabulario: ASIR, módulo, título, contenido, unidad.

Reglas: ASIR contiene varios módulos, un módulo tiene un elemento simple título y un elemento contenido, contenido tiene varias unidades y toda unidad debe estar en un contenido, las unidades son texto simple, detrás de cada unidad solo puede ir otra unidad o fin de contenido, detrás de un módulo solo puede ir otro módulo o fin de ASIR.

```
<asir>
  <módulo>
    <título>Lenguajes de Marcas</título>
    <contenido>
      <unidad>Introducción</unidad>
      <unidad>HTML</unidad>
      <unidad>CSS</unidad>
      <unidad>XML</unidad>
      <unidad>XLST</unidad>
      <unidad>Sindicación</unidad>
      <unidad>ERP</unidad>
    </contenido>
  </módulo>
  ...
</asir>
```

Ejemplo 2: Mini HTML

Vocabulario: html, head, title, body, p.

Reglas: html contiene un elemento head y un elemento body, head contiene un elemento simple title, body contiene varios elementos p y todo elemento p debe estar en body, p es texto simple, detrás de p solo puede ir otro p o fin de body, detrás de body solo puede ir fin de html.

```
<html>
  <head><title>Mi página</title></head>
  <body>
    <p>Hola mundo</p>
    <p>Esta es mi página</p>
  </body>
</html>
```


Características

- Texto plano
 - compuestos únicamente por caracteres de texto.
 - codificar con distintos códigos dependiendo del idioma o alfabeto que se necesite, por ejemplo: ASCII, 180-8859-15, UTF-8.
 - pueden ser interpretados directamente por un simple editor de texto.
 - independientes del sistema operativo o programa con el que fueron creados
- Mezcla de texto con **<tags>**
- Independencia del dispositivo final
 - Interpretado de diferentes formas dependiendo del dispositivo final.
- Especialización, gráficos vectoriales, sindicación de contenidos, notación científica, interfaces de usuario
- Flexibilidad, combinación con otros lenguajes, PHP, JavaScript

Características

- Han destacado por una serie de características que los han convertido en los tipos de **lenguajes más usados** en la informática actual para **almacenar y representar los datos**.
 - Que se basan en el texto plano.
 - Que permiten utilizar metadatos.
 - Que son fáciles de interpretar y procesar.
 - Que son fáciles de crear y suficientemente flexibles para representar datos muy diversas.
- Las aplicaciones de Internet y muchos de los programas de ordenador que se utilizan habitualmente utilizan de alguna manera u otra algún lenguaje de marcas.

Características

- **Basados en texto plano**
 - Codificados en diferentes códigos de caracteres: ASCII , ISO-8859-1, UTF-8, etc.
 - Una de las **ventajas**: se pueden interpretar directamente, y sólo es posible si usamos el formato de texto (los binarios requieren un programa para interpretarlos).
 - Otra **ventaja**: son independientes de la plataforma, del sistema operativo o del programa.
 - Son fáciles de crear y modificar ya que sólo requieren un simple editor de textos.

Características

- **Uso de metadatos**
 - Las marcas se intercalan entre el contenido del documento → etiquetas descriptivas → indica el contenido de los datos que contienen (semántica).
 - Estas marcas son la forma en que se añaden los metadatos a los documentos de texto.

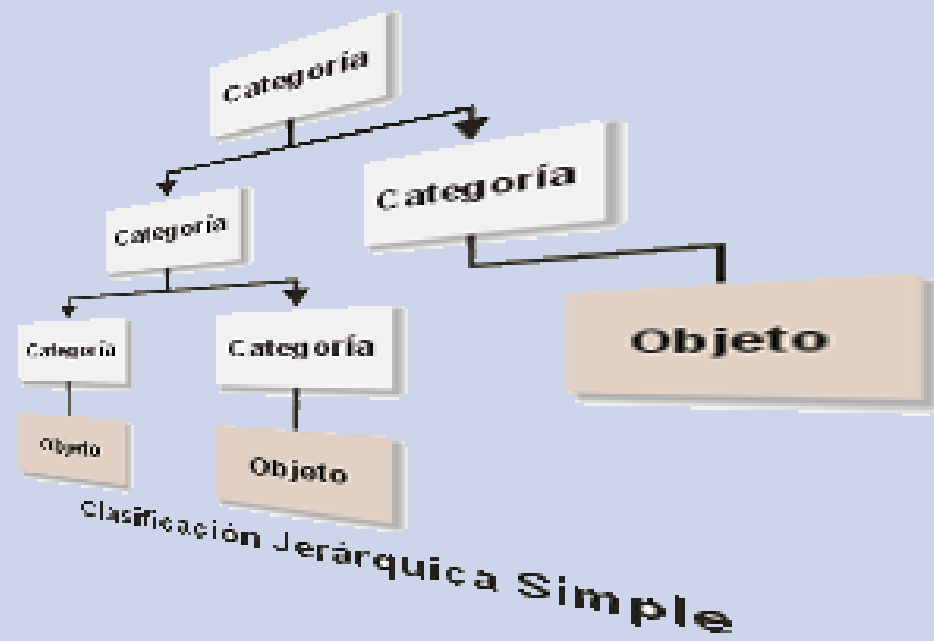
Características

- **Facilidad de proceso**

- Los lenguajes de marcas permiten que el procesamiento de los datos que contengan pueda ser automatizado de alguna manera, ya que el archivo contiene la **estructura** de los datos que contiene.
- El hecho de incluir la estructura permitirá que un programa pueda interpretar cada uno de los datos de un fichero de marcas para representarlo o tratarlo convenientemente, ya que muestran la estructura de los datos que contienen.
- Posteriormente un programa podrá interpretar gracias a las marcas que es lo que significa cada uno de los datos del documento.

Características

- **Facilidad de creación y representación de datos diversos**
 - A pesar de que fueron pensados para contener datos de texto, los lenguajes de marcas han demostrado que son capaces de contener datos de muchos tipos diferentes.
 - Actualmente se están utilizando archivos de marcas para representar imágenes vectoriales, fórmulas matemáticas, crear páginas web, ejecutar funciones remotas mediante servicios web, representar música o sonidos, etc.
 - Y sin importar qué tipo de datos se representen siempre habrá la posibilidad de crear estos archivos desde un editor de texto básico.



4. Clasificación de los lenguajes de marcas

Atendiendo al **objetivo básico**

- Atendiendo al **objetivo básico** del lenguaje de marcas:
 - **Lenguajes procedimentales y de presentación** , orientados a especificar **cómo** se debe representar la información.
 - **Lenguajes descriptivos o semánticos**: orientados a describir **la estructura** de los datos que contiene.
- Esta es la clasificación más aceptada
- Nos podemos encontrar lenguajes que tengan aspectos de los dos grupos y permitan tanto definir la manera de presentar la información como definir su estructura.

Atendiendo al objetivo básico

- **Procedimentales y de presentación**
 - En estos lenguajes lo que se hace es indicar **de qué manera** se hará la presentación de los datos.
 - El ejemplo más popular de estos lenguajes es el HTML pero hay muchos más: TeX, Wikitext...
 - **Ejemplo:** lenguaje de marcas ligero *Wiki markup* , que utiliza Mediawiki (programa con el que se ha desarrollado la Wikipedia):

==Página Principal==
===Hardware===
Parte "'física"' del ordenador
=== Software ===
Parte "'lógica"' del "'ordenador"'
====VideoJuegos====

==, ===
para
establecer
los niveles

"' negrita
"'negrita
con comillas
simples



[Página principal](#)
[Cambios recientes](#)
[Página aleatoria](#)
[Ayuda](#)

Herramientas
[Lo que enlaza aquí](#)
[Cambios relacionados](#)
[Subir archivo](#)
[Páginas especiales](#)
[Versión para imprimir](#)
[Enlace permanente](#)
[Información de la página](#)

[Marc](#) [Discusión](#) [Preferencias](#) [Lista d](#)

[Página principal](#) [Discusión](#) [Leer](#) [Editar](#) [★](#) [Más](#)

Página principal

Contenido [\[ocultar\]](#)

- 1 [Págian Principal](#)
 - 1.1 [Hardware](#)
 - 1.2 [Software](#)
 - 1.2.1 [VideoJuegos](#)

Págian Principal [\[editar\]](#)

Hardware [\[editar\]](#)

Parte **física** del ordenador

Software [\[editar\]](#)

Parte **lógica** del **'ordenador'**

VideoJuegos [\[editar\]](#)

Atendiendo al objetivo básico

- **Descriptivos o semánticos**
 - Se describe qué estructura lógica tiene el documento, ignorando de qué manera será representada en los programas.
 - Sólo se ponen las marcas con el objetivo de definir las partes que dan estructura al documento.
 - El ejemplo más importante es el XML pero hay algún otro que está teniendo mucho apoyo, como por ejemplo JSON.
 - Ejemplo:

```
<alumnos>
  <alumno>
    <nombre>María</nombre>
    <apellidos>García Ruíz</apellidos>
  </alumno>
  <alumno>
    <nombre>Carlos</nombre>
    <apellidos>Pérez Pérez</apellidos>
  </alumno>
</alumnos>
```

Etiquetas más
información

Estructuras de
datos

Atendiendo al tipo de marcas

De presentación

- Especifica formato sin estructura.
- Procesadores de texto y aplicaciones de edición profesional como nroff, troff, RTF, TeX, Docbook

Descriptivo, estructural o semántico.

- Estructura del documento, en árbol que almacena información (bases de datos semiestructuradas)
- ASN.1, YAML, EBML, RDF, XFML, OWL, XTM (derivados de XML)

Híbrido

- Marcas de los dos tipos anteriores
- HTML, XHTML, WML

Atendiendo a su funcionalidad.

Para crear documentación electrónica

- RTF, TeX, troff, nroff
- ASN.1, EBML, YAML
- Wikitexto, DocBook, LinuxDoc

Tecnologías de internet

- HTML, XHTML, WML (páginas web)
- gladeXML, Xforms, XAML (formularios/interfaces de usuario)
- RSS; Atom (sindicación de contenidos)
- WSDL, SOAP, UDDI (servicios web)
- SMPP (mensajería instantánea)

De propósito específico

- MathML, CML
- SSML, SRGS, VoiceXML
- MusicXML
- SVG, VML, X3D
- SMIL (archivos multimedia)
- XLL (enlaces)
- XSLT (transferencia de documentos)
- XTM (mapas conceptuales)
- RDF, XFML, OWL, XMP (catalogación y clasificación de documentos)
- GML (información geográfica)
- OFX



5. Organizaciones y estándares

Organización y estándares

- Especificación de normas.
- Organismos internacionales, nacionales y privadas para la definición de estas normas
 - ISO
 - Open Source
 - W3C
 - Desarrolla estándares que aseguran el crecimiento de la web a largo plazo
 - Recibe ingresos de las cuotas de sus miembros, becas de investigación, subvenciones y donaciones privadas ➔ sin fines lucrativos
 - Miembros: Microsoft, Apple, Google, etc
 - Ha creado SW de código abierto, como el validador W3C

Actividades

1.4. Visita la página de W3C y consulta el estado de la publicación de HTML5

Consulta los objetivos de W3C

Consulta el listado de miembros españoles



6. Introducción a los principales lenguajes de marcas

Terminología

- **HTML** es el lenguaje en el que se construyen las páginas web, mezcla texto y contenido multimedia, hipervínculos.
- Creado por **Tim Berners Lee** a principio de los 90
- **HTML** es la base y la esencia sobre la que está construida la *World Wide Web* en la actualidad.
- Los navegadores web son capaces de interpretarlas.
- **Servidor**. Ordenador que contiene algún tipo de recurso o servicio que ofrece a otros ordenadores (**clientes**).
- Servicio *World Wide Web* → servidores web.
- Las páginas web internamente se estructuran como archivos de texto interpretados con sentido por un navegador web, como *Firefox, Chrome, Opera, Safari, ...*
- **HTML (*HyperText Markup Language*)**. Lenguaje basado en etiquetas que se escriben mediante signos de mayor y menor. Por ejemplo, `<p>` para indicar el comienzo de párrafo y `</p>` para indicar su final.

Terminología II.

- **CSS (*Cascade Style Sheet*)**: otro conjunto de normas que complementa de HTML. Hojas de estilo.
- Para añadir interactividad a las páginas y convertir nuestras páginas web en verdaderas aplicaciones, se usan otros lenguajes como *Javascript* o *PHP*.
- W3C (<http://www.w3c.org>). Organismo que define el conjunto de normas para HTML, CSS.

Direcciones URL

- **IP** (*Internet Protocol*): Identificador único. Identificador del servidor web en Internet.
 - Cuatro bytes separados por puntos.
 - P.E. <http://209.85.148.104/> = <http://www.google.es>
- **DNS** (*Domain Name System*): servidores que toman una dirección como la anterior y la convierten en el formato numérico de su IP.
- **URL** (*Uniform Resource Locator*): es una dirección que nos permite acceder de forma inequívoca a un determinado servidor.

¿Qué necesitamos para trabajar con HTML?

- Crear archivos de texto sencillos → editor de textos planos.
- Navegador que interprete los archivos
- Existen editores de texto más completos para facilitar la edición

Herramientas

- Varios tipos de herramientas.
- **Editor de texto.**
 - Editor de texto plano.
- **Navegador.**
 - Conviene probar en varios navegadores.
 - <http://www.mozilla.com/es-ES/firefox>
 - <http://www.google.com/chrome?hl=es>
- **Editores de páginas web**
 - Editor visual de archivos HTML
 - BlueGriffon: gratuita, multiplataforma y código abierto.
 - <http://bluegriffon.org/pages/Download>

- **Editores de imágenes**

- Web visual ➔ necesidad de imágenes en la web.
- Gimp es la opción libre.
- <http://www.gimp.org.es/descargar-gimp.html>

- **Aplicación FTP**

- Para subir la aplicación a un servidor
- Filezilla.
- <http://filezilla-project.org/download.php>

- **Versiones portables**

- Archivos que no necesitan ser portables y se pueden ejecutar desde un dispositivo USB

XML

- Simplificación y adaptación de SGML
- Forma de definir lenguajes → metalenguaje.
- **Características**
 - **Extensible.** Se pueden definir nuevas etiquetas
 - **Versátil.** Separa contenido, estructura y presentación
 - **Estructurado.** Se pueden modelar datos a cualquier nivel de complejidad
 - **Validable.** Cada documento se puede validar frente a un DTD/Schema
 - **Abierto.** Independiente de empresas, sistemas operativos, lenguajes de programación o entornos de desarrollo
 - **Sencillo.** Fácil de aprender y de usar.
- Internet, intercambio de información entre diferentes plataformas
- **Uso:** BD ligeras, editores de texto , hojas de cálculo, transacciones comerciales

Ejercicio

- Analizamos una web: accede a una web
- Visualiza el código (Ver código fuente de la página)
- Accede al Inspector
- Chequea la web con el Validator