



UD1.1 Introducción a los lenguajes de marcas

LMSXI

IES de Teis

Vigo

Curso 2021-2022

¿Qué son los lenguajes de marcas?

- Lenguajes escritos en **texto plano** que, además de la propia información del texto incorporan **marcas o etiquetas** que aportan información sobre la propia información del texto (**metadatos**)
- Los **metadatos** pueden ser de cualquier tipo:
 - Relativos a la estructura del texto:
`<p> Esto es un párrafo </p>`
 - Relativos al formato de presentación, por ejemplo para mostrar en negrita (*bold*):
Los `metadatos` pueden...
- Normalmente las marcas o etiquetas usarán los caracteres “<” y “>”

¿Qué son los lenguajes de marcas?

- Proviene de la industria editorial, donde se añadían notaciones para indicar instrucciones de impresión sobre el propio texto.
 - Podéis ver un ejemplo de símbolos de corrección de escritos aquí del Manual de estilo de Chicago

Archivos de texto plano vs archivos binarios

	Archivos texto plano	Archivos binarios
Estructura	Caracteres alfanuméricos y caracteres de control codificados en binario	Información de distinta naturaleza (documentos con formato, imágenes, audio, código en lenguajes de programación, etc.) codificada en binario
Herramientas de edición/visualización	Simple procesador de texto (Bloc de notas)	SW específico
Carece de efectos tipográfico	Sí	No
Legible por el ser humano al abrirlo con un bloc de notas	Sí	No
Ejemplos de extensiones	.txt, .xml, .html, .tex	.docx, .pdf, .jpeg, .exe, .bin

- Si intentamos abrir un documento PDF con el Bloc de notas el contenido será ininteligible:



Codificación de caracteres

- Los ordenadores trabajan con información binaria, es decir, con 2 tipos de información que comúnmente representamos como 0 y 1.
- **Codificación de caracteres:** método que permite convertir un carácter de un lenguaje natural en un símbolo de otro sistema de representación. Por ejemplo:
 - Líneas y puntos en código Morse
 - 0's y 1's en codificación binaria

Codificación de caracteres: Evolución

- Surgieron diferentes tipos de codificación de caracteres:
 - Norma ASCII: 7 bits, inicialmente solo para el alfabeto inglés
 - Sistemas ASCII extendido: 8 bits. Por ejemplo, la norma Latin-1 (ISO-8859-1) que define la codificación del alfabeto latino, incluyendo letras acentuadas, ñ, ç, ..., y letras especiales (como ß, Ø) para las lenguas de Europa occidental y países escandinavos.
 - Otros idiomas asiáticos desarrollaron sus propias codificaciones: chino GB2312, Chino tradicional BIG5 o Japón JIS
 - UNICODE: En un intento de unificación para estandarizar la representación de cualquier lengua del planeta

Codificación de caracteres: UNICODE

- UNICODE asigna a cada posible carácter un número (code-point) escrito como U+00639 (donde U viene de 'Unicode') y el número se representa en formato hexadecimal. Para mantener la compatibilidad con ASCII, los 7 primeros bits en UNICODE coinciden con ASCII.
- Tiene varias implementaciones:
 - UTF-8
 - UTF-16
 - UTF-32
- Se pueden consultar los números asignados a cada carácter [aquí](#)
- Nosotros trabajaremos principalmente con UTF-8

Implementaciones <u>UNICODE</u>	Tamaño de codificación (bytes)
<u>UTF-8</u>	Variable: 1,2,3 ó 4 bytes
<u>UTF-16</u>	2 ó 4 bytes
<u>UTF-32</u>	4 bytes



Ventajas/inconvenientes del uso de texto plano

- Interpretable por el ser humano
- Fácil de manipular por el ser humano con un simple editor de texto plano
- Facilita el intercambio de información entre distintos elementos electrónicos
 - Por el contrario:
 - Ocupa más que un archivo binario
 - Ofrece menos seguridad



¿Para qué se utilizan los lenguajes de marcas?

- Definir la estructura de los datos de un documento
- Permitir el intercambio de ficheros entre diferentes aplicaciones y plataformas
- Dar formato a los documentos de texto

Ejemplos de lenguajes de marcas

- GML
 - HTML
 - PostScript
 - SGML
 - LaTeX
 - Markdown
 - XML
 - RTF
 - JSON
- Breve descripción de cada uno con ejemplos disponible [aquí](#) (a partir del apartado 1.4.2 en adelante)



Tipos de lenguajes de marcas

- Orientados a la **presentación**
- Orientados a la **descripción**
- Orientados a **procedimientos**
 - Breve descripción de cada tipo con ejemplos disponible en [el mismo recurso](#) (apartado 1.5)



Heramientas de edición

- Un simple editor de texto bastaría:
 - Bloc de notas o Notepad++
- IDEs (Integrated Development Environment):
 - Visual Studio Code, Eclipse, NetBeans, ...
- Editores más específicos:
 - XML Copy Editor, Brackets, Oxygen XML Editor