

Aplicaciones Web y Lenguajes de Marcas

Tema 1a: Introducción a los Lenguajes de Marcas

1º Parcial - LMSGI

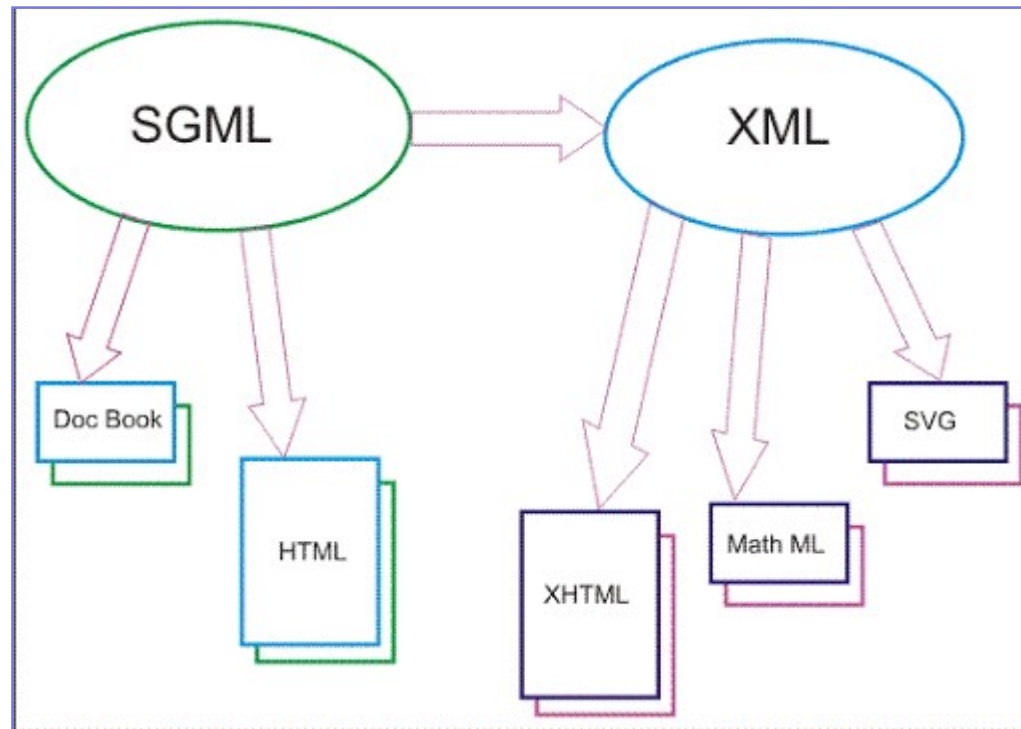
1º de Ciclo Superior de DAM

les de Teis (Vigo) – Curso 2020/21

Tema 1a: Introducción a los Lenguajes de Marcas

Objetivos: *conocer cómo surgen los lenguajes de marcas y su evolución, e identificar las características generales.*

- Identificar las características generales de los lenguajes de marcas
- Clasificar los lenguajes de marcas e identificar los más importantes
- Estudiar las ventajas y ámbitos de aplicación de los lenguajes de marcas



Tema 1a: Introducción a los Lenguajes de Marcas

■ Origen de los Lenguajes de Marcas:

- ✓ Tradicionalmente, en la época de la imprenta, los **manuscritos** de autor incluían instrucciones que indicaban el tipo de letra, el estilo y el tamaño con que debía ser representado el texto, etc.
- ✓ A estas indicaciones se les llamaba *marcas*, y existía un buen número de ellas conocidas y manejadas informalmente por los tipógrafos.
- ✓ Con la introducción de las computadoras, y sobre todo de la web, se trasladó este concepto al mundo de la informática.
- ✓ Cualquier información se puede representar en un ordenador: imágenes, texto, música, vídeos, mapas, ...

➤ **Recuerda:** ¿sabes cómo se almacenan los datos en un ordenador?

Tema 1a: Introducción a los Lenguajes de Marcas

■ Formas de representar la información en un ordenador:

✓ Los ordenadores digitales son máquinas binarias => trabaja con todos los datos utilizando 0 y 1.

✓ Toda información almacenada lo hace de forma binaria.

■ Ej (utilizando un sistema 8 bits):

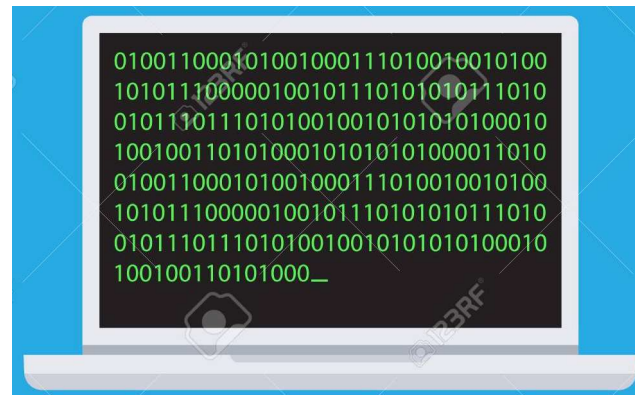
✓ **Nº 19:** 00010011

✓ **Letra A:** 01000001

✓ **Color Azul**, modelo RGB (0,0,255): 0..0 0..0 11111111

✓ **Imagen:** cada píxel de la imagen a binario

✓ **Vídeo:** codificar cada muestra de sonido e imagen.



Tema 1a: Introducción a los Lenguajes de Marcas

■ ¿Como surge la necesidad de crear lenguajes de marcas?

*Un ordenador sólo es capaz de representar información de forma binaria => toda información a guardar debe ser previamente **codificada** en binario*

✓ **Codificación:** proceso de transformación de datos a otro código (binario).

✓ La codificación crea problemas si no existe un estándar.

✓ Existan varios estándares de codificación.

■ Datos binarios:

✓ **Todo es binario menos el texto:** cualquier fichero que no sea texto se considera binario.

✓ La forma de **codificación** es **variable**, dependiendo del programa que la produce y el sistema elegido.

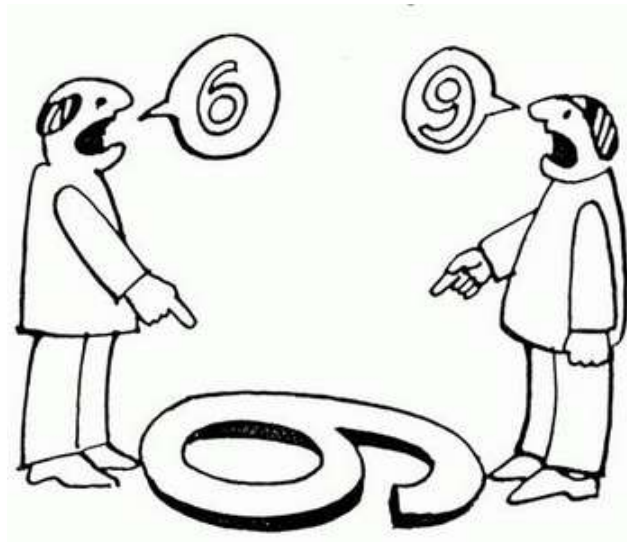
✓ Para poder acceder a información existente es necesario un proceso de **decodificación**: pasar de datos binarios a algo entendible por las personas.

■ Ej: presentar en pantalla una fotografía.

¿Como surge la necesidad de crear lenguajes de marcas?

■ Texto:

- ✓ Forma más natural para representar la información desde el punto de vista de las personas.
 - Ej: libro en papel, imagen borrosa que representa una palabra,
- ✓ Para introducir los textos en un ordenador es necesario **codificar** los caracteres en secuencias binarias.
- **Debate:** ¿cómo crees que un ordenador interpreta el “número” inferior? ¿qué ocurre si dos ordenadores utilizan distinto sistema de codificación?



¿Como surge la necesidad de crear lenguajes de marcas?

- **Problemas de codificar texto:** cada sistema codifica de una manera diferente.
 - Ej: ñ es el 241 (11101101) y el 164 (10100100)
- **Sistemas de Codificación:**
 - **ASCII** (*American Standard Code for Information Interchange*):
 - ✓ Código estándar unificado que permite la compatibilidad entre ordenadores de diferentes fabricantes.
 - ✓ Hecho en 1967 por *ANSI*.
 - ✓ Usa 7 bits para codificar las letras=> representa 128 caracteres.
 - ✓ Utiliza el alfabeto inglés.
 - **Problema:** cuando se usa en países de habla no inglesa.
 - **ASCII extendido:** se expande a 8 bits => representa 256 caracteres.
 - ✓ Tiene caracteres obsoletos (los de control para impresoras y otros periféricos).
 - **Problema:** hay muchos ASCII extendidos.

Tabla ASCII

Caracteres ASCII de control			Caracteres ASCII imprimibles			ASCII extendido		
00	NULL	(carácter nulo)	32	espacio	64	@	96	`
01	SOH	(inicio encabezado)	33	!	65	A	97	a
02	STX	(inicio texto)	34	"	66	B	98	b
03	ETX	(fin de texto)	35	#	67	C	99	c
04	EOT	(fin transmisión)	36	\$	68	D	100	d
05	ENQ	(consulta)	37	%	69	E	101	e
06	ACK	(reconocimiento)	38	&	70	F	102	f
07	BEL	(timbre)	39	'	71	G	103	g
08	BS	(retroceso)	40	(72	H	104	h
09	HT	(tab horizontal)	41)	73	I	105	i
10	LF	(nueva línea)	42	*	74	J	106	j
11	VT	(tab vertical)	43	+	75	K	107	k
12	FF	(nueva página)	44	,	76	L	108	l
13	CR	(retorno de carro)	45	-	77	M	109	m
14	SO	(desplaza afuera)	46	.	78	N	110	n
15	SI	(desplaza adentro)	47	/	79	O	111	o
16	DLE	(esc.vínculo datos)	48	0	80	P	112	p
17	DC1	(control disp. 1)	49	1	81	Q	113	q
18	DC2	(control disp. 2)	50	2	82	R	114	r
19	DC3	(control disp. 3)	51	3	83	S	115	s
20	DC4	(control disp. 4)	52	4	84	T	116	t
21	NAK	(conf. negativa)	53	5	85	U	117	u
22	SYN	(inactividad sinc)	54	6	86	V	118	v
23	ETB	(fin bloque trans)	55	7	87	W	119	w
24	CAN	(cancelar)	56	8	88	X	120	x
25	EM	(fin del medio)	57	9	89	Y	121	y
26	SUB	(sustitución)	58	:	90	Z	122	z
27	ESC	(escape)	59	;	91	[123	{
28	FS	(sep. archivos)	60	<	92	\	124	
29	GS	(sep. grupos)	61	=	93]	125	}
30	RS	(sep. registros)	62	>	94	^	126	~
31	US	(sep. unidades)	63	?	95	_		
127	DEL	(suprimir)						
128	Ç		160	á	192	Ł	224	Ó
129	ü		161	í	193	ł	225	ô
130	é		162	ó	194	Ł	226	Ô
131	â		163	ú	195	ł	227	Ò
132	ä		164	ñ	196	—	228	ö
133	à		165	Ñ	197	†	229	Õ
134	á		166	º	198	ä	230	μ
135	ç		167	°	199	Ä	231	þ
136	ê		168	¿	200	Ł	232	ð
137	è		169	@	201	Œ	233	ù
138	ë		170	¬	202	ℒ	234	û
139	ï		171	½	203	Ť	235	ü
140	î		172	¼	204	Ŧ	236	ý
141	ï		173	ı	205	=	237	ÿ
142	Ä		174	«	206	≠	238	—
143	Å		175	»	207	□	239	'
144	É		176	⋮	208	δ	240	≡
145	æ		177	⋮	209	Ð	241	±
146	Æ		178	⋮	210	Ê	242	—
147	ô		179		211	Ë	243	¼
148	ö		180	†	212	È	244	¶
149	ò		181	À	213	Ì	245	§
150	û		182	Â	214	Í	246	÷
151	ù		183	Ã	215	Î	247	°
152	ÿ		184	©	216	Ï	248	°
153	Ö		185	≡	217	Ɔ	249	—
154	Ü		186	≡	218	Ɔ	250	·
155	ø		187	Ɔ	219	■	251	·
156	£		188	Ɔ	220	■	252	·
157	Ø		189	¢	221	·	253	·
158	×		190	¥	222	·	254	■
159	f		191	¬	223	■	255	nbsp

Sistemas de Codificación

- **ISO 8859-1 o Alfabeto Latino n.º 1 o ISO Latín 1:**

- ✓ Norma de la ISO que define la codificación del alfabeto latino.
- ✓ Los 8 bits de *ASCII* no llegan para representar todas las lenguas del planeta.
- ✓ Cada región usaba una parte extendida de *ASCII* distinta.
- ✓ Se crean nuevos estándares de 256 caracteres.
- ✓ Los 128 primeros son iguales a *ASCII*.
- ✓ Incluye los diacríticos (como letras acentuadas, ñ, ç), y letras especiales (como ß, Ø), necesarios para la escritura de las lenguas originarias de Europa occidental.

- **Problema:**

- ✓ En cada documento es necesario especificar qué codificación se usa.

Sistemas de Codificación

▪ Unicode:

- ✓ Trata de resolver los problemas anteriores y almacenar todos los caracteres de todas las lenguas del planeta.
- ✓ **Estándar de codificación de caracteres universal** utilizado para la representación de texto para procesamiento del equipo.
- ✓ Proporciona una manera consistente de codificación de texto multilingüe y facilita el intercambio de archivos de texto internacionales.
- ✓ El tamaño del carácter es variable, usando hasta un máximo de 32 bits.
- ✓ Los 128 primeros son iguales a ASCII
- ✓ Los 256 primeros son iguales a ISO-8859-1

▪ Formas de codificar los caracteres:

- UTF-8:** tamaño variable para cada carácter (8, 16, 24 o 32 bits) según carácter que se quiera representar.
- UTF-16:** tamaño variable para cada carácter (16 o 32).
- UTF-32:** cada carácter ocupa 32 bits.

Archivos binarios versus archivos de texto

■ Ventajas de archivos binarios:

- ✓ Ocupan **menos espacio** que los archivos de texto, ya que optimizan mejor su codificación a binario.
 - **Ej:** el número 213 ocupa un byte y no tres si se codificase como texto
- ✓ Son más **rápidos** de manipular por parte del ordenador, ya que se parecen más al lenguaje nativo de los ordenadores.
- ✓ Permiten el **acceso directo** a los datos, el cual es más rápido que el acceso secuencial de los archivos de texto.
- ✓ Permiten **cifrar** el contenido y que no sea visible de accesos indebidos.

■ Ventajas de archivos de texto:

- ✓ Fácilmente interpretables: ideales para almacenar datos para **exportar/importar** información a cualquier dispositivo electrónico.
- ✓ Compatibles con cualquier dispositivo.
- ✓ Directamente modificables, sin necesidad de tener software específico.
- ✓ Manipulación más sencilla que la de los archivos binarios.
- ✓ Transportables y entendibles por todo tipo de redes.

El problema de compartir datos

- ¿Qué ocurre cuando creamos un cierto tipo de datos en una aplicación y necesito/quiero compartirlos con otra aplicación/ordenador diferente?
 - ✓ Para **exportar/importar** datos **binarios** => utilizar **convertidores de formato**.
 - Ej: una foto almacenada como tipo .jpg, .tiff, .gif, ...
- **Creación de estándares para compartir datos:**
 - ✓ Si se universalizan dichos estándares, compartir información entre diferentes aplicaciones y dispositivos es más sencillo.
 - ✓ Cada formato tiene sus peculiaridades.
 - Ej: para música utilizar el mp3, para documentos el pdf, para imágenes el jpg
- **Problema:**
 - ✓ Creación de nuevos estándares de formato mejores que sustituyan a los anteriores.
 - ✓ Formatos propietarios de empresas.
 - **Averigua:** ¿sabes qué ocurrió con **Flash Player de Adobe**? Lee este [artículo](https://www.xataka.com/aplicaciones/siempre-flash-player-adobe-confirma-que-dejara-funcionar-a-finales-2020) de la web Xataka para saber cuándo desaparecerá:
<https://www.xataka.com/aplicaciones/siempre-flash-player-adobe-confirma-que-dejara-funcionar-a-finales-2020>

Evolución de los Lenguajes de Marcas

- Formato de archivo que cualquier dispositivo es capaz de entender: el **texto**
- En los **archivos de texto plano** se intenta **almacenar otro tipo de datos**.
 - **Metadatos**: datos que sirven para describir otros datos
 - **Ej**: información de formato de texto, escrita como palabras marcadas de forma especial
- Hay **procesadores** que comienzan a almacenar los datos de formato en texto:
 - ✓ Utilizan **caracteres marcados** de forma **especial**.
 - ✓ El programa interpreta esos datos como operaciones de formato.
- **¿De dónde surgió la idea? ¿ y el nombre?**
 - ✓ El nombre proviene del inglés *marking up* que consistía en hacer marcas de colores en manuscritos.
 - ✓ Supone una evolución de las marcas en el margen que se empleaban para anotar manuscritos.
 - ✓ En los sistemas digitales, estas marcas a mano fueron sustituidas por *etiquetas*.
 - ✓ Las distintas formas de hacer las *marcas* dieron lugar a los llamados *lenguajes de marcas*.

Definición de los Lenguajes de Marcas

- **Definición de los lenguajes de marcas:**

sistema que permite hacer anotaciones en un documento de forma que puedan distinguirse sintácticamente del texto del documento

- Es un modo de codificar un documento donde, junto con el **texto**, se incorporan etiquetas, marcas o anotaciones con **información adicional** relativa a la **estructura** del texto o su **formato** de presentación.
- Cada lenguaje de marcas debe especificar las siguientes **características**:
 - ✓ **Etiquetas** posibles
 - ✓ **Dónde** colocarlas
 - ✓ **Significado** de cada una de ellas
 - ✓ Hacer **explícita** la **estructura** de los documentos
 - ✓ Definir su contenido semántico o cualquier otra información lingüística o extralingüística que se quiera expresar.

Definición de los Lenguajes de Marcas

- Todo lenguaje de marcas está definido en un documento denominado **DTD** (*Document Type Definition*).
- En el **DTD** se establecen las marcas, los elementos utilizados por dicho lenguaje y sus correspondientes etiquetas y atributos, su sintaxis y normas de uso.
 - **Ej:** Aspecto de un documento realizado con un lenguaje de marcas

```
<carta>
    <fecha>22/11/2006</fecha>
    <presentacion>Estimado cliente:</presentacion>
    <contenido>bla bla bla bla ...</contenido>
    <firma>Don José Gutiérrez González</firma>
</carta>
```

Evolución de los Lenguajes de Marcas

- **GML** (*Generalized Markup Language*):

- ✓ Desarrollado entre 1969 y 1970 por **Charles Goldfarb** de IBM.
- ✓ En el propio documento de texto, se almacenaba el formato del mismo.
- ✓ Los elementos marcados con símbolos ":" y "." delimiten marcas de formato.
- ✓ Base para el **SGML**.

- Ej:

```
:h0.El reino de los animales
:h1.Mamíferos
:p.Los mamíferos (:hp1.Mammalia:ehp1.) son una clase de
vertebrados :hp2.amniotas homeotermos:ehp2. que poseen
glándulas mamarias productoras de leche con las que
alimentan a las crías
:h1.Aves
:p. Las aves son animales vertebrados, de sangre
caliente, que caminan, saltan o se mantienen solo sobre
las extremidades posteriores
```

Evolución de los Lenguajes de Marcas

- **TeX ():**
 - ✓ Una de las primeras aplicaciones prácticas de esta idea fue el estándar TeX, creado y mantenido por *Donald Knuth* en los años 70 y 80.
 - ✓ Se creó para hacer textos científicos, utilizando una tipografía y capacidades que fuera igual en todos los ordenadores, asegurando gran calidad en los resultados.
 - ✓ Se necesita un programa que traduzca el fichero TeX a un formato de impresión.
 - ✓ TeX requiere amplios conocimientos para ser utilizado, por lo que sólo se utiliza en entornos científicos a través de **LaTeX**, el cual simplifica su uso.
 - ✓ La aparición de procesadores de texto tipo WYSIWYG relegó estos sistemas al olvido.
 - WYSIWYG -*What You See Is What You Get*:- se aplica a procesadores de texto y otros editores de texto con formato, como los editores de HTML, que permiten escribir un documento viendo directamente el resultado final.
- **Amplía:** lee el siguiente [artículo](http://www.cervantex.es/queestex) completo sobre TeX para entender mejor su utilidad (<http://www.cervantex.es/queestex>).

Ejemplos de LaTeX y su resultado impreso:

```
\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 f\'ormula
\begin{align}
E &= mc^2
\end{align}
\end{document}
```

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

$$E = mc^2$$

- Ej facilitado por un alumno de Daw:

```
\begin{equation}
y = \int_{x=0}^{x=2\pi + 10} f(x) \cdot dx
\end{equation}
```

$$y = \int_{x=0}^{x=2\pi+10} f(x) \cdot dx$$

Evolución de los Lenguajes de Marcas

- **RTF** (*Rich Text Format*):

- ✓ Creado por *Microsoft* en 1987 por varios miembros del equipo de desarrollo de *Microsoft Word*.
- ✓ Crea documentos de texto que incluye anotaciones de formato.
- ✓ Es un formato casi estándar para procesadores de texto.
- ✓ El formato de los documentos lo almacena como *metadatos*.
- ✓ Se utiliza como formato de intercambio entre diferentes procesadores.
- ✓ Su éxito procede de que las indicaciones de formato son potentes y son más legibles por las personas que el formato nativo de los procesadores de textos.
- ✓ Su inconveniente es que es un lenguaje de marcado complejo.

Evolución de los Lenguajes de Marcas

- **RTF** (*Rich Text Format*):
 - Ej: contenido de documento rtf y cómo se visualizaría.

```
{\rtf\ansicpg1252\deff0\deflang3082
{\fonttbl
{\f0\fcharset0\froman Times New Roman}
{\f1\fcharset0\fswiss Arial Black}
}
{\pard \f1\fs48
El reino de los animales
\par}
{\pard \f1\fs40
Mamíferos
\par}

{\pard \f0\fs25
Los mamíferos ({\b Mammalia}) son una clase de
vertebrados {\i amniotas homeotermos} que poseen
glándulas mamarias productoras de leche con las que
alimentan a las crías
\par}
{\pard \f1\fs40
```

El reino de los animales

Mamíferos

Los mamíferos (**Mammalia**) son una clase de vertebrados *amniotas homeotermos*. que poseen glándulas mamarias productoras de leche con las que alimentan a las crías

Evolución de los Lenguajes de Marcas

- **SGML** (*Standard Generalized Markup Language*):

- ✓ La idea de crear un lenguaje **estándar** de marcado de texto fue expuesta por *William W. Tunnicliffe* en 1967. La mayor novedad consistía en **separar presentación y estructura del texto**.
- ✓ Realmente, *Charles Goldfarb* es el padre de la idea. Dirigió el comité que elaboró el estándar del lenguaje de marcado.
- ✓ Es una mejora notable del lenguaje de **GML** que estandarizaba el lenguaje de marcado y que fue definida finalmente por ISO como **estándar mundial** en documentos de texto con etiquetas de marcado.
- ✓ Es un **metalenguaje**: *lenguaje para la definición de lenguajes de marcas*.
- ✓ Todos los lenguajes de marcas actuales descienden de él directamente, **XML** y **HTML**.
- ✓ **No tiene** etiquetas concretas => los elementos SGML se crean a voluntad.
- ✓ Su uso se materializó en los 80.

Evolución de los Lenguajes de Marcas

- **SGML** (*Standard Generalized Markup Language*):
 - ✓ Tras crear los elementos necesarios, hay que **determinar las reglas** e indicar cómo se va a presentar el formato final, es decir, hay que **definirlos**.
 - **Ej: ¿cómo se van a mostrar los elementos <titulo1>**

```
<articulo>
    <titulo1>El reino de los animales</titulo1>
    <titulo2>Mamíferos</titulo2>
    <normal>Los mamíferos (<negrita>Mammalia
</negrita>) son una clase de vertebrados
< cursiva>amniotas homeotermos</ cursiva>. que poseen
glándulas mamarias productoras de leche con las que
alimentan a las crías</normal>
    <titulo2>Aves</titulo2>
    <normal>Las aves son animales vertebrados, de
sangre caliente, que caminan, saltan o se mantienen solo
sobre las extremidades posteriores</normal>
</articulo>
```

Evolución de los Lenguajes de Marcas

- **PostScript:**

- ✓ Lenguaje de **descripción** de páginas.
- ✓ Desarrollado en 1976 por *John Warnock*, continuando desde *Xerox*, hasta que en 1985 funda *Adobe Systems*.
- ✓ Es el lenguaje más utilizado por los sistemas de impresión de alta gama.
- ✓ Es un **lenguaje de programación** que **marca acciones a realizar**, indicando cómo se va a mostrar la información en el dispositivo final.

- **Ej:**

```
%!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
```



Hola mundo!

Evolución de los Lenguajes de Marcas

- **HTML** (*HiperText Markup Language*):
 - ✓ En 1991, la situación cambia drásticamente cuando **Tim Berners-Lee**, el inventor de la **Web**, utiliza la sintaxis **SGML** para crear un nuevo lenguaje de etiquetas: **HTML**.
 - ✓ La extraordinaria **sencillez** del HTML, junto con el empleo de **URLs** y la **distribución libre** de los primeros **navegadores**, están en el origen del éxito de la World Wide Web.
 - ✓ *Documentos transportables por Internet en los que es posible el **hipertexto***: palabras marcadas que abren un documento relacionado.
 - ✓ Se usa en la Web, siendo el tipo de documento más empleado en el mundo => **páginas web**.
 - ✓ Inicialmente los documentos se veían con ayuda de *intérpretes o visores de texto* -**Lynx** de Unix- que simplemente coloreaban el texto y remarcaban el hipertexto.
 - ✓ Después el software mejoró y aparecieron **navegadores** con capacidad más gráfica para mostrar formatos más avanzados y visuales.
 - ✓ El motivo de su éxito es también su principal **handicap**.
- **Razona**: ¿cuáles crees que son las desventajas de HTML?

Evolución de los Lenguajes de Marcas

- HTML (*HiperText Markup Language*):
- Ej página Html desde *Lynx*:

```
Portada - Wikipedia, la enciclopedia libre (pl of 11)
#search copyright

Los resultados de la votación para el Consejo de Administración de
la Fundación Wikimedia ya están disponibles.

Portada
De Wikipedia, la enciclopedia libre

Saltar a navegación, búsqueda

¡Bienvenido a Wikipedia, la enciclopedia libre!
Edición en español de Wikipedia, iniciada en el 2001.
Quarto - Libro de visitas - Acceso WAP - Contacto

Explora Wikipedia

Exploración Exploración Clasificación Sistemas de clasificación
Búsqueda · índice alfabético · índice de categorías · Portales
temáticos · Todos los artículos · Acceso WAP Decimal Universal* ·
Unesco* (4 dígitos* · 6 dígitos*) · Listas (Acrónimos · Biografías ·
Países)

Buscar título exacto Buscar en el texto
En la columna de la izquierda de cada artículo hay una caja de
búsqueda equivalente a esta.
<NORMAL LINK> Use right-arrow or <return> to activate.
Arrow keys: Up and Down to move. Right to follow a link; Left to go back.
H>elp O>ptions P>rint G>o M>ain screen Q>uit /=search [delete]=history list
```

Evolución de los Lenguajes de Marcas

- **HTML** (*HiperText Markup Language*):
 - **Desventajas de HTML:**
 - ✓ Su extrema **simplicidad** en declaración e interpretación, impide restringir la generación de **documentos mal formados**.
 - ✓ Además, se comenzaron a **mezclar presentación y contenido**, por lo que hubo de crearse **CSS** como solución.
 - ✓ No soporta tareas de impresión y diseño.
 - ✓ El lenguaje no es flexible, ya que las **etiquetas** son **limitadas**.
 - ✓ No permite mostrar contenido dinámico.

Evolución de los Lenguajes de Marcas

- ✓ Para crear un documento HTML llega con tener un **editor de texto**.
- Programas capaces de **visualizar** un documento HTML: **navegadores**.
 - ✓ **Mosaic**: primer navegador, a principios de los 90.
 - ✓ *Guerra de navegadores*, a finales de los 90, en los que cada uno ampliaba HTML añadiendo **etiquetas propias**.
 - ✓ Se crea la **W3C** -*World Wide Web Consortium*- para estandarizar.
- **Evolución de HTML**:
 - ✓ Aparición de **nuevas etiquetas** -tablas, imágenes, capas, ...-
 - ✓ Permite añadir “**dinamismo**” a las páginas gracias a lenguajes como **JavaScript**.
 - ✓ Se puede generar de forma automática desde el *lado del servidor*.
 - ✓ Se crean **hojas de estilo** que separan la presentación del contenido: **CSS**.
 - ✓ Se permite la incrustación de objetos: **Flash**.
 - ✓ Aparecen elementos semánticos.
 - ✓ Permite el dibujo libre y generación de gráficos mediante el elemento *canvas*.
 - ✓ Se puede añadir vídeo y audio de forma sencilla.

Evolución de los Lenguajes de Marcas

- **HTML** (*HiperText Markup Language*):
- **Ej de HTML5:**

```
<!DOCTYPE html>
<html lang="es">
<head>
  <meta charset="UTF-8">
  <title>Document</title>
</head>
<body>
  <h1>El reino de los animales</h1>
  <h2>Mamíferos</h2>
  <p>Los mamíferos (<strong>Mammalia</strong>) son una clase de vertebrados <em>amniotas homeotermo</em>
  <h2>Aves</h2>
  <p>Las aves son animales vertebrados, de sangre caliente, que caminan, saltan o se mantienen solo
</body>
</html>
```

- **Ej de HTML 4:**

```
!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">
```

Evolución de los Lenguajes de Marcas

- **XML** (*eXtensible Markup Language*):

- ✓ Desarrollado por el **World Wide Web Consortium**, es un **subconjunto** de **SGML**, ideado para **mejorarlo**, simplificándolo y adaptándolo a los documentos de Internet.
- ✓ Es más **entendible**, logrando equilibrio entre simplicidad y flexibilidad, además de ser uno de los lenguajes de etiquetas más importantes.
- ✓ Define lenguajes con sintaxis **más estricta**, indicando lo que es permitido y lo que no.
- ✓ Se trata de un **metalenguaje** que permite crear **etiquetas** adaptadas a las necesidades, **definiendo cómo pueden ser y qué se puede hacer con ellas**.
- ✓ Todo **documento** debe cumplir dos **condiciones**: ser **válido** y estar **bien formado**.
- ✓ Permite asignar **atributos** a las etiquetas.
- ✓ La **estructura** y el **diseño** son **independientes**.
- ✓ **Soluciona problemas** surgidos de la internacionalización y de la imposibilidad de validar (problemas de HTML).
- ✓ Es uno de los formatos de documentos más populares para **exportación e importación** de datos.

Evolución de los Lenguajes de Marcas

- **XML** (*eXtensible Markup Language*):

- Ej:

```
1 <persona>
2     <nombres>Elsa</nombres>
3     <apellidos>Zambrano</apellidos>
4     <fecha-de-nacimiento>
5         <día>18</día>
6         <mes>6</mes>
7         <año>1996</año>
8     </fecha-de-nacimiento>
9     <ciudad>Pamplona</ciudad>
10</persona>
```

Evolución de los Lenguajes de Marcas

- **XHTML**(*eXtensible HyperText Markup Language*):

- ✓ Es una redefinición del HTML en clave XML => HTML **compatible** con XML: *un documento XHTML es un documento XML válido.*
- ✓ Se creó como sustituto del HTML.
- ✓ Al cumplir las restricciones de XML, mantiene sus etiquetas y características, añadiendo algunas restricciones y elementos propios de XML.
- ✓ Los lenguajes basados en XML tienen múltiples aplicaciones => como el **intercambio de datos** entre servidores.

- Ej:

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"
  "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="es">
  <head>
    <title>Título del documento</title>
  </head>
  <body>
    <h1>Cabecera de documento</h1> <!-- Título de sección -->
    <p id="inicial">Un párrafo en el cuerpo<br />del documento.</p>
    <hr />
  </body>
</html>
```

Evolución de los Lenguajes de Marcas

- **JSON** (*JavaScript Object Notation*):
 - ✓ Es la forma en la que Javascript **serializa** los objetos.
 - ✓ Es **versátil**: permite definir datos complejos como **arrays** y código de funciones.
 - ✓ Los navegadores le daban soporte nativo.
 - ✓ Es uno de los competidores del **XML**. Muy popular para **almacenar** datos.
 - ✓ NO es exactamente un lenguaje de marcas.
 - El texto se divide en **datos** y **metadatos**:
 - El símbolo de los dos puntos separa el metadato del dato.
 - Los símbolos de llave y corchete permiten agrupar de diversas formas los datos.

Evolución de los Lenguajes de Marcas

- **JSON** (*JavaScript Object Notation*):

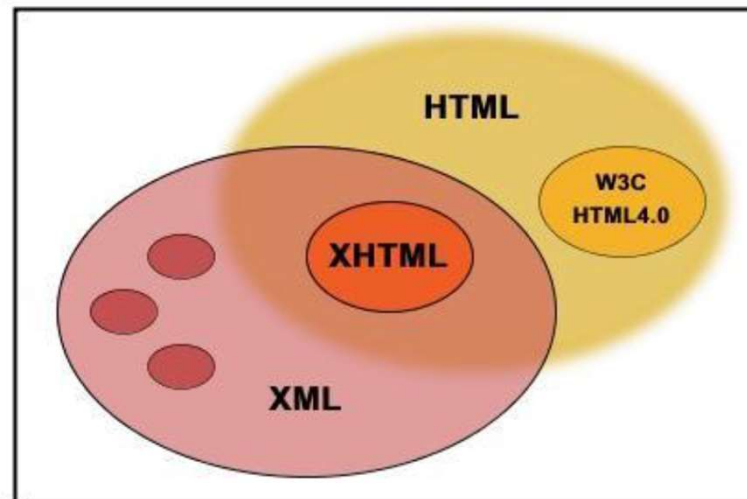
- Ej:

```
{ "empinfo" :  
  {  
    "employees" : [  
      {  
        "name" : "Scott Philip",  
        "salary" : £44k,  
        "age" : 27,  
      },  
      {  
        "name" : "Tim Henn",  
        "salary" : £40k,  
        "age" : 27,  
      },  
      {  
        "name" : "Long Yong",  
        "salary" : £40k,  
        "age" : 28,  
      }  
    ]  
  }  
}
```

Evolución de los Lenguajes de Marcas



- Relación entre HTML y XML: XHTML



Comparativa entre XML y HTML:

XML	HTML
✓ Es un perfil de SGML.	✓ Es una aplicación de SGML.
✓ Especifica cómo deben definirse conjuntos de etiquetas aplicables a un tipo de documento.	Aplica un conjunto limitado de etiquetas sobre un único tipo de documento.
✓ Modelo de  hiperenlaces complejo.	✓ Modelo de hiperenlaces simple.
✓ El navegador es una plataforma para el desarrollo de aplicaciones.	✓ El navegador es un visor de páginas.
✓ Fin de la guerra de los navegadores y etiquetas propietarias.	El problema de la 'no compatibilidad' y las diferencias entre navegadores ha alcanzado un punto en el que la solución es difícil.

- **Debate:** ¿qué lenguaje describe el contenido de lo que etiqueta? ¿qué lenguaje usa etiquetas para presentar información?

Un Ejemplo de HTML:

```
1 <b>Elizabeth Castro</b>
2 <b><i>XML Guía de Aprendizaje</i></b>
3 <b>precio: 30 €</b>
```

El mismo Ejemplo en XML:

```
1 <libro>
2   <autor>Elizabeth Castro</autor>
3   <titulo>XML Guía de Aprendizaje</titulo>
4   <precio moneda="euros">30</precio>
5 </libro>
```

Tipos de Lenguajes de Marcas

- **Tipos de lenguajes de marcas:**
 - a) **Orientados a la presentación:** sólo se muestra la presentación, pero es difícil extraer información.
 - b) **Orientados a la descripción** *-descriptivos o semánticos-*: no se especifica cómo hay que procesar el texto sino que se utilizan etiquetas que describen el texto - indican el significado del contenido-.
 - c) **Orientados a los procedimientos:** se incluyen instrucciones de cómo hay que procesar el texto.
- **Debate:** ¿crees que en un mismo documento se pueden combinar diferentes tipos de lenguajes de marcas?. A partir de los lenguajes de marcas que hemos visto, facilita un ejemplo de cada tipo. ¿Dónde encajarías a HTML en sus primeras versiones? ¿Y al HTML 5?. Utiliza un muro de padlet para añadir las conclusiones.

Tipos de Lenguajes de Marcas

- a) **De presentación:** Define el formato del texto. Al texto común se añaden palabras encerradas en símbolos especiales que contienen indicaciones de formato que permiten a los *traductores* generar un documento final en el que el texto aparece con el formato indicado.
 - **Ej:** primeras versiones de **HTML**, **RTF**, en el que se indica cómo debe presentarse el texto y no lo que significa, archivos generados por los procesadores de texto tradicionales, en los que al texto del documento se le acompaña de indicaciones de formato –negrita–
- b) **Descriptivo o semántico:** Describen las diferentes partes en las que se estructura el documento sin especificar cómo deben representarse.
 - **Ej:** **XML**, **JSON** en los que indica una semántica de contenido que lo hace ideal para almacenar datos, como el nombre de persona, su Nif,
- c) **De procedimientos:** Orientado también a la presentación pero, el programa que representa el documento debe *interpretar* el código en el mismo orden en que aparece. Incluye instrucciones a realizar con el texto.
 - **Ej:** **LaTeX**, **PostScript**, en los que se puede indicar una fórmula matemática.

Características de los Lenguajes de Marcas

- **Características que representan ventajas en los lenguajes de marcas:**
 - ✓ **Texto plano:** Un documento escrito con lenguajes de marcas puede ser editado con un simple **editor de textos**, ya que las *marcas* se intercalan entre el contenido.
 - También se pueden utilizar programas más complejos que faciliten el trabajo.
 - Al ser sólo texto, los **documentos** son **independientes** de la **plataforma**, sistema operativo o programa con el que fueron creados.
 - ✓ **Compactos:** las instrucciones de marcado se intercalan con el propio contenido, en un **único archivo** o flujo de datos.
 - ✓ **Facilidad de procesamiento:** las etiquetas de estos lenguajes son fácilmente interpretables y tratables informáticamente.
 - ✓ **Flexibilidad:** aunque originalmente estos lenguajes se idearon para documentos de texto, se empezaron a utilizar con éxito en otras áreas, como gráficos vectoriales, servicios web e interfaces de usuario, aprovechando su **simplicidad y potencia**.
- **Razona:** ¿crees que las características indicadas representan una ventaja o un inconveniente?

Ámbitos de aplicación de los Lenguajes de Marcas

- El lenguaje de marcas por excelencia ha sido **XML**: descendiente directo de SGML, es multipropósito y permite definir otros lenguajes de marcas, estandarizados por **W3C**.
- **Clasificación de los lenguajes de marcas según su ámbito de aplicación:**
 - a) **Contenidos web**: permiten definir las estructuras de los documentos web. Ej: **HTML** y **XHTML**.
 - b) **Procesadores de textos**: emplean lenguajes de marcas para definir el formato y apariencia de los documentos. Ej: formatos de archivo **ODF** o **RTF**.
 - c) **Sindicación de contenidos**: permite difundir información actualizada frecuentemente a usuarios suscritos a una fuente de contenidos. Ej: **RSS** utiliza lenguajes de marcas para este propósito.
 - d) **Servicios web**: se utiliza para definir la información que intercambian distintos servicios web, facilitando la transferencia de información. Ej: el lenguaje **SOAP**
 - e) **Gráficos vectoriales**: Ej: **SVG** (*Scalable Vector Graphics*), permite describir completamente complejas imágenes de 2D empleando únicamente texto.

Ámbitos de aplicación de los Lenguajes de Marcas

- **Clasificación de los lenguajes de marcas según su ámbito de aplicación (cont.):**
 - f) **Listas de reproducción:** los reproductores multimedia emplean lenguajes de marcas para definir sus listas de reproducción. Ej: **ASX, PLA, WPL**.
 - g) **Archivos de configuración:** muchos programas y servicios utilizan lenguajes de marcas para sus archivos de configuración. Ej: el servidor de aplicaciones Tomcat.
 - h) **Lenguaje de marcas de propósito general:** los lenguajes anteriormente citados están basadas en **XML**. Este es un lenguaje de marcas de propósito general que permite definir lenguajes concretos de representación de documentos.

Ámbitos de aplicación de los Lenguajes de Marcas

- Ej de lenguajes de marcas especializados:

Área	Lenguaje	Descripción / Ejemplo
Matemática:	MathML y OpenMath	Su objetivo es expresar notación matemática
Geomática:	Geography ML	Su objetivo es el modelaje, transporte y almacenamiento de información geográfica
Aeronáutica:	Spacecraft ML	
Multimedia:	Synchronized Multimedia Integration Language	El lenguaje SMIL permite integrar audio, video, imágenes, texto o cualquier otro contenido multimedia
Voz:	VoiceXML	
Mensajería instantánea:	XMPP	Con el protocolo XMPP queda establecida una plataforma para el intercambio de datos XML que puede ser usada en aplicaciones de mensajería instantánea
Gráficos 3D:	VRML/X3D, STEP	formato de archivo normalizado que tiene como objetivo la representación de escenas u objetos interactivos tridimensionales