

## LENGUAJES DE MARCAS

Desarrollo de Aplicaciones Web 1er Curso

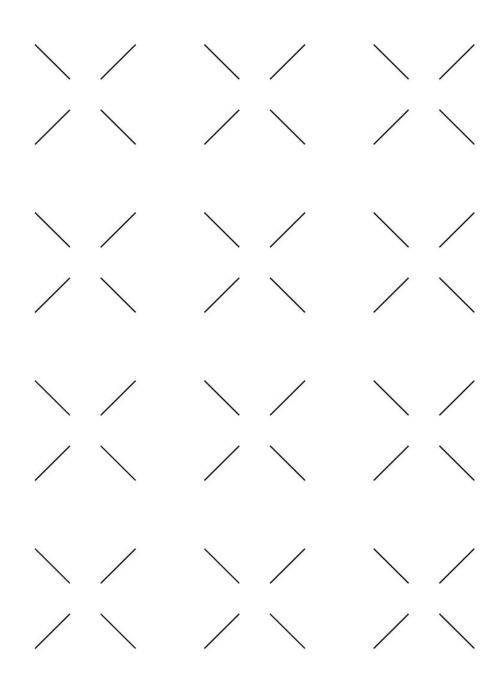
Profesores: Gissela García/ Óscar Villar



# Índice de contenido

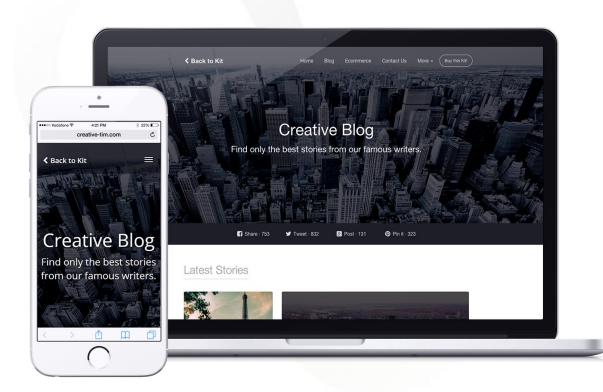
## Índice de contenido

- 1. ¿Cómo funciona internet?
- 2. Ordenador e información
- 3. Exportar / Importar datos
- 4. Conjuntos de caracteres
- 5. Lenguajes de Marcas



1\_ ¿Cómo funciona internet?

## Evolución de las tecnologías



VS

#### La Web en los 90





Tiramillas

🔛 Tienda

Juegos



<u>FÚTBOL</u> <u>LIGA 1999/2000</u>

Pichichi

Quinielas 1ª División

2ª División

EUROCOPA

Fórmula 1

Liga ACB

BALONCESTO

BALONMANO

Liga Asobal

CICLISMO

Mundial 99 Copa Mundo

**FUTBOL SALA** 

Liga 99-00

Grand Slam C. Davis

VOLEIBOL

Superliga

Mundial

TENIS

MOTOCICLISMO

AZA

Comp. Europeas

Comp. Europeas

Calend. Caza Mayor

Rallyes

Jornada en juego

2a B: <u>I / II</u> / <u>III</u> / <u>IV</u>

· La pág. de tu equipo

L. DE CAMPEONES

COPA DE LA UEFA

AUTOMOVILISMO

Noticias de...

▶ Polideportivo

Quinielas Add Active Channel

Foros

lig a99-00

Cibergimnasio

Liga Fantástica

Tu deporte

Tu equipo

· · Webs Especiales · · 

✓

ا به باطنیمار باشار در بازگره از در <u>واسانها ۱۳۳۳ آشاد ام بیم از میدراز بران از بیما</u>

#### ¿Buscas información sobre alguna competición en Juego?. Prueba aquí

#### CRIVILLÉ FUE SEXTO EN LOS ENTRENAMIENTOS DE SURÁFRICA

#### Cita con la historia

Álex Crivillé partirá desde el sexto puesto de la parrilla en el G.P. de Suráfrica, antepenúltima prueba del Mundial, en la que el piloto de Seva puede convertirse en el primer español que logra un título de 500 c.c. Sus dos rivales lo hicieron mejor que él en los entrenamientos: Okada logró la 'pole' y Roberts, el cuarto mejor registro.



#### IAPÚNTATE! mail MARCA

¿Te gustaría tener una cuenta de correo GRATIS nombre@marca.es? Haz clic aquí

#### BARCA - R. MADRID

Toda la información sobre el partido del próximo miércoles

#### CAZA Calendario de la Temporada completa de Caza Mayor

### LOS ALLBLACKS

#### Esta es la danza ritual del equipo de rugby Nueva Zelanda

### **YO** 😢 EURO · 2004

#### PÍDESELO A LA UEFA DESDE MARCA



La selección cerrará el domingo ante Israel una fase de clasificación de la Eurocopa que comenzó en drama y termina con euforia, tras asegurarse la presencia en la fase final y recuperar la estima de la



España cumplió con los pronósticos y derrotó a Irlanda en la Dunhill Cup. De esta manera, García, Olazábal y Jiménez concluyen como íderes de su grupo. Ahora se verán las caras en semifinales con



El defensa del Real Madrid Fernando Hierro aseguró que le "duele" no poder disputar el próximo miércoles el encuentro ante el Barcelona, pero que no le da más vueltas a una situación que no queda más que acatar.

#### Noticias de fútbol

- · Venturin no se siente aludido por las palabras de Jesús Gil
- · Pedro Cortés confía en retener al 'Piojo' López
- El sevillista Prieto afirma que el árbitro Esquinas Torres es un mentiroso
- Todas las noticias

#### Noticias de polideportivo

- Torneo de Palermo. Berasategui jugará la final
- Mundial de judo. Isabel Fernández, plata
- Mundial de ciclismo. Vandenbroucke, el hombre a batir
- Todas las noticias

#### FORO

El próximo martes se decide la sede de la próxima Eurocopa 2004. Mándanos tu mensaje de apoyo a la candidatura española y se lo haremos llegar a la UEFA.

#### ENCUESTA

¿Deberían deiar jugar a Serena Williams contra hombres?



O No

Votar

Así va la encuesta



@ESCRIBENOS







20:45

**UEFA NATIONS LEAGUE** 

Regultar

No despierta de la pesadilla

C. NAVARRO 3 58



Ver todo >

REAL MADRID Solari, de vuelta al Real Madrid JOSÉ FÉLIX DÍAZ = 404



Colección DISNEY ENGLISH VAUGHAN. Este sábado 4ª entrega "VAIANA" por 6.95 € iInfórmate!×

### ¿Cómo funciona internet?

Las principales tecnologías que permiten hacer funcionar un sitio web son:

- ➤ Lenguaje de marcas de hipertexto o HTML (+ CSS + JS..)
- > Sistema de nombres de dominio o **DNS**
- > Protocolo de transferencia de hipertexto o HTTP / modo seguro HTTPS
- > Servidores web
- ➤ Navegadores web (Firefox, Chrome, Safari..)



### La web en la actualidad

Lenguajes de marcas



HTML: define el contenido de la web. Es decir, los elementos que la componen.

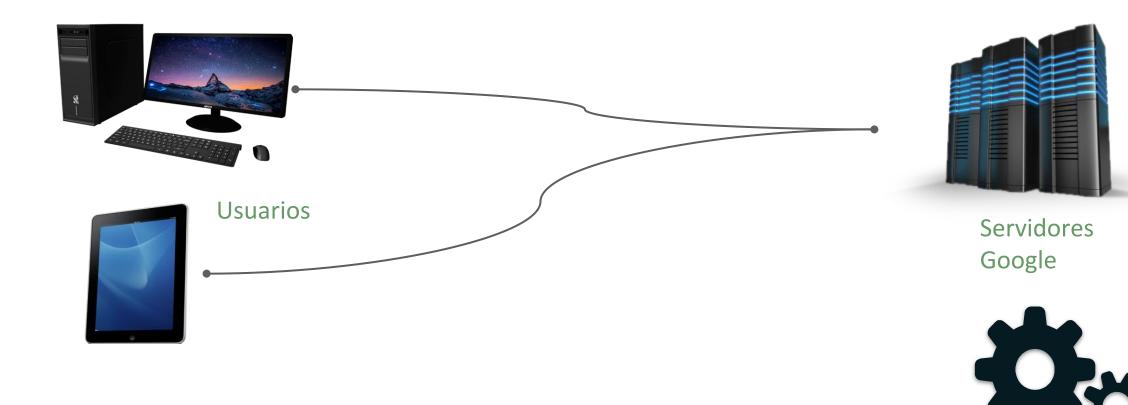
**CSS**: define los estilos de la web. Aspectos visuales, así como disposición de elementos (<u>layout</u>).

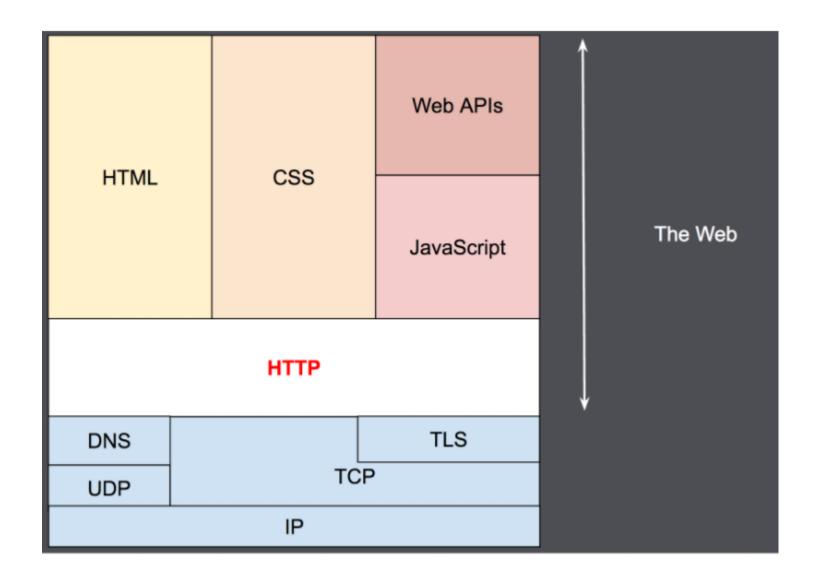
Lenguajes de programación

<u>Lenguajes front-end</u>: del lado cliente (navegador) muestran el contenido web solicitado a un servidor. (JS)

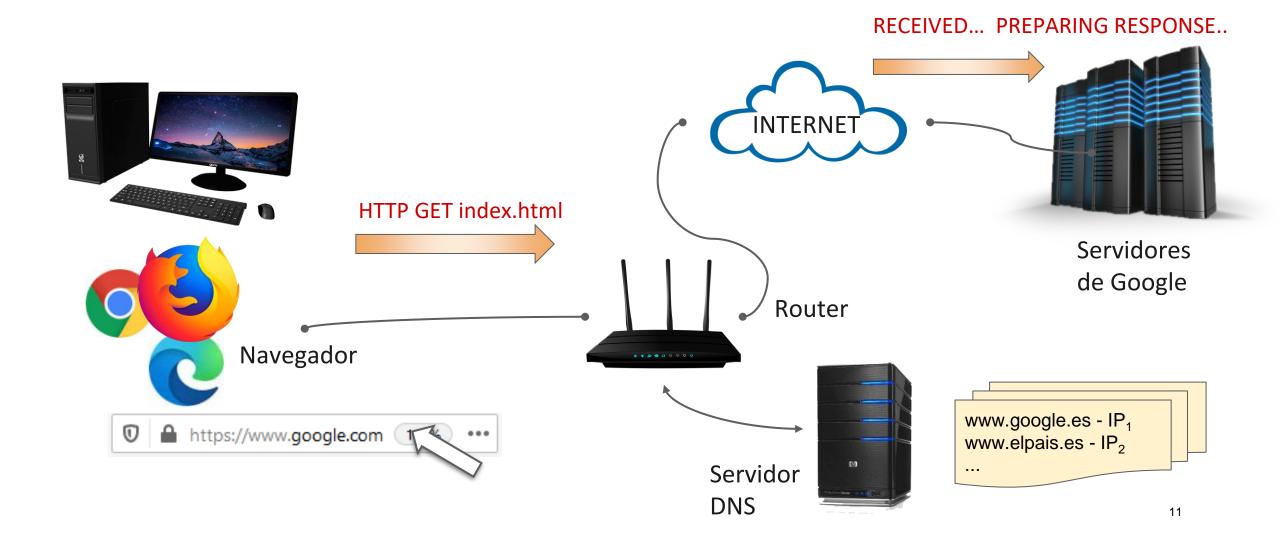
<u>Lenguajes back-end:</u> gestionan del lado servidor las peticiones que hacen los usuarios con sus aplicaciones cliente.

## ¿Cómo funciona internet?

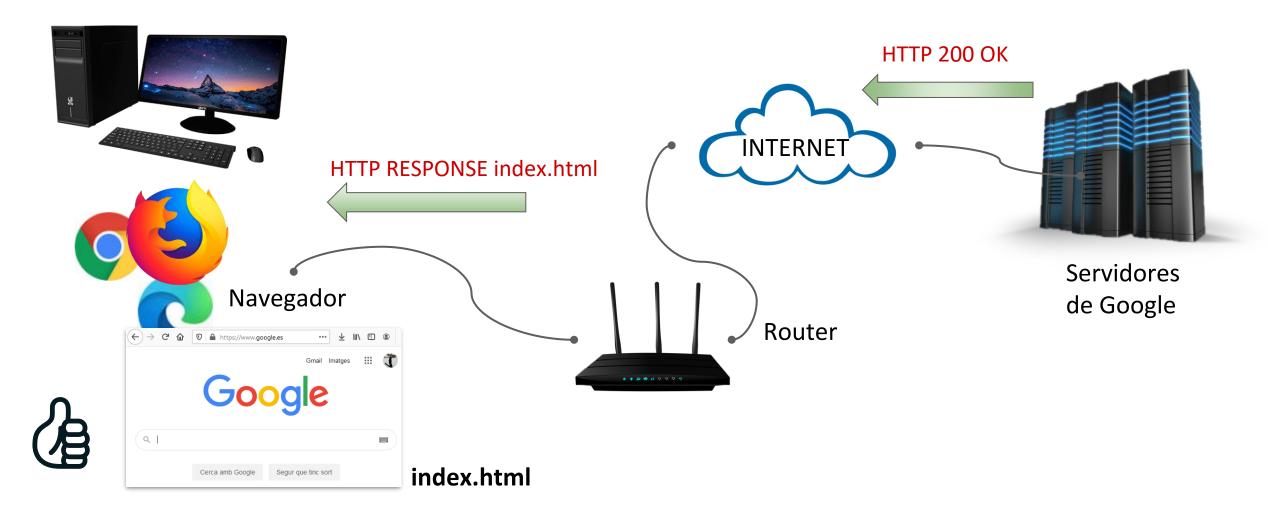




### Petición al servidor



### Respuesta del servidor (200 OK)



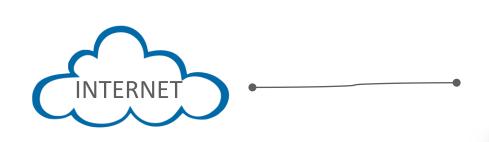
#### Petición

```
GET / HTTP/1.1
Host: youtube.com
Accept: text/html
Accept-Language: es-ES
```

#### Respuesta

```
HTTP/1.1 200 OK
Server: Node.JS
Content-type: text/html
<html>
  <body>
    <h1> Hola!!! </h1>
  </body>
</html>
```

### Respuesta del servidor (404 NOT FOUND)



HTTP GET URL: www.google.com/pruebas



RECEIVED... PREPARING RESPONSE..

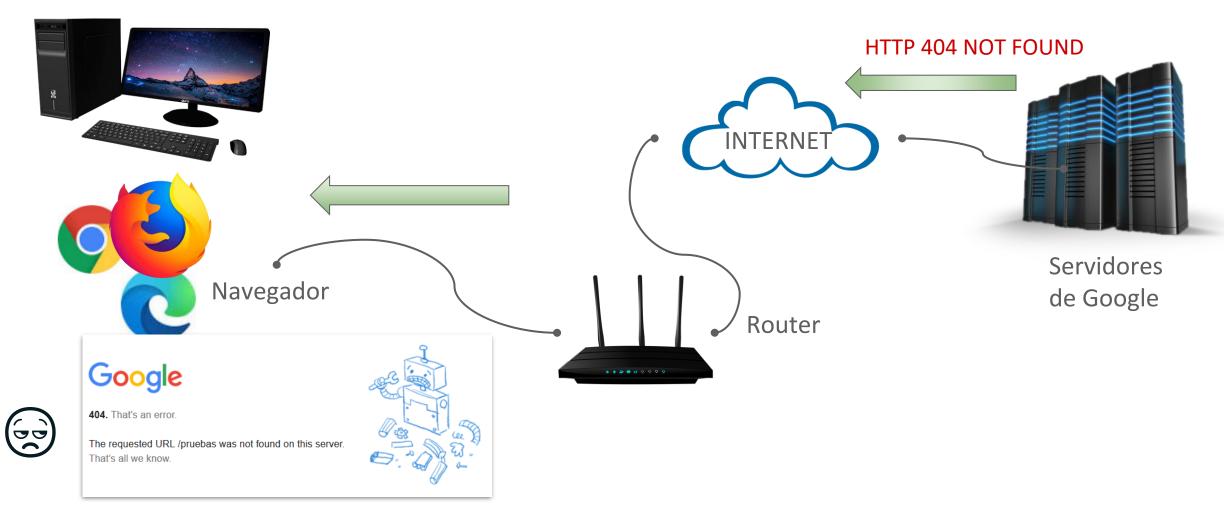


Servidores de Google I DON'T KNOW

NOTHING ABOUT

THE RESOURCE /pruebas..

## Respuesta del servidor (200 OK)



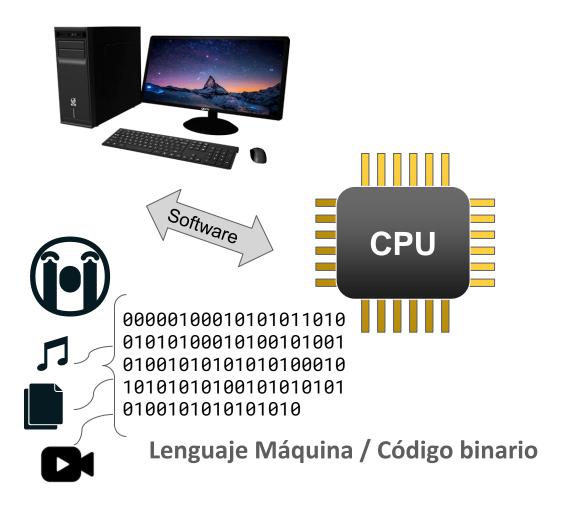
Lenguaje de Marcas

## Ejemplo de peticiones

https://www.google.com/	0 0
https://www.google.com/gmail	0 0
https://www.google.com/translate	0 0
https://www.google.com/pruebas ————	

# 2. Ordenador e información

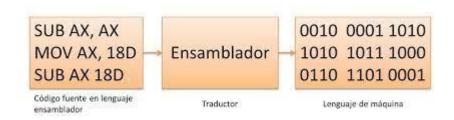
### Representación de la información



### Código binario -> Entendido por máquina

Lenguaje propio del procesador

Excesivamente complejo para el humano.



### **Texto -> Entendido por los humanos**



Forma más habitual de transmitir la información

Sistema decimal / cadenas de caracteres

## **Archivos binarios / texto**

### **Archivos binarios**

### 1100 1010 0101

- Menos espacio
- > Ejecución más rápida
- Acceso aleatorio a datos
- Cierta protección a modo de cifrado



### **Archivos texto**

- Facilidad para exportar/importar datos entre dispositivos.
- Múltiples posibilidades de modificación
- Manipulación más sencilla
- Posibilidad de establecer un proceso de telecomunicación basado en texto

### Problema codificación del texto

00000100010101011010 Bienvenidos 01010100010100101001 queridos alumnos 01001010101010100010 de Lenguajes de Marcas! 10101010100101010101 0100101010101010 01000001 01000010 ?????

**Textos** incompatibles entre sí

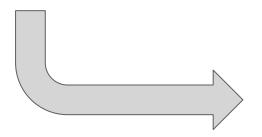
> Estándares de codificación (1967)

### Importancia de estándares

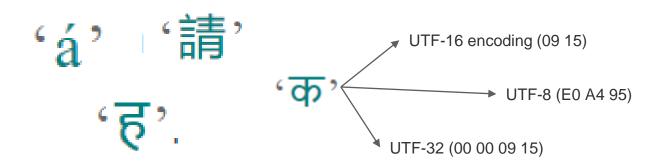


### Problema codificación del texto

Author: Guðrún Guðmundsdóttir. Title: Introduction to character encoding (文字符号化入門). Copyright © 2004-2007 W3C® (MIT, ERCIM, Keio).



Author: Guðrðn Guðmundsdðttir. Title: Introduction to character encoding (æ–‡å—符å-化入門). Copyright © 2004-2007 W3C® (MIT, ERCIM, Keio).



## **Codigo ASCII**

### **American Standard Code for Information Interchange**

- Creado por ANSI (American National Standards Institute) en 1967.
- ➤ Codificación del **alfabeto inglés** con **7 bits** (128 símbolos).
- ➤ Incluye: minúsculas / mayúsculas, caracteres de puntuación, símbolos especiales / control.

**PROBLEMA:** ¿Y la codificación del resto de alfabetos del mundo?

Ampliaciones sucesivas tabla ASCII

Carácteres no imprimibles				Carácteres imprimibles								
Nombre	Dec	Hex	Car.	Dec	Hex	Car.	Dec	Hex	Car.	Dec	Hex	Car.
Nulo (null)	0	00	NUL	32	20	Espacio	64	40	<b>@</b>	96	60	•
Inicio de cabecera	1	01	SOH	33	21	!	65	41	A	97	61	a
Inicio de texto	2	02	STX	34	22		66	42	В	98	62	b
Fin de texto	3	03	ETX	35	23	#	67	43	C	99	63	c
Fin de transmision	4	04	EOT	36	24	\$	68	44	D	100	64	d
Enquiry	5	05	ENQ	37	25	%	69	45	E	101	65	e
Acknowledge	6	06	ACK	38	26	&	70	46	F	102	66	f
Beep (sonido)	7	07	BEL	39	27		71	47	G	103	67	g
Backspace	8	08	BS	40	28	(	72	48	H	104	68	h
Tabulador horizontal	9	09	HT	41	29	)	73	49	I	105	69	i
Salto de línea	10	0A	LF	42	2A	*	74	4A	J	106	6A	j
Tabulador vertical	11	0B	VT	43	2B	+	75	4B	K	107	6B	k
Salto de página	12	0C	FF	44	2C	,	76	4C	L	108	6C	1
Retorno de carro	13	0D	CR	45	2D	_	77	4D	M	109	6D	m
Shift fuera	14	0E	SO	46	2E		78	4E	N	110	6E	n
Shift dentro	15	0F	SI	47	2F	/	79	4F	0	111	6F	0
Escape línea de datos	16	10	DLE	48	30	0	80	50	P	112	70	р
Control dispositivo 1	17	11	DC1	49	31	1	81	51	Q	113	71	q
Control dispositivo 2	18	12	DC2	50	32	2	82	52	R	114	74	r
Control dispositivo 3	19	13	DC3	51	33	3	83	53	S	115	73	S
Control dispositivo 4	20	14	DC4	52	34	4	84	54	T	116	74	t
Acknowledge negativo	21	15	NAK	53	35	5	85	55	U	117	75	u
Síncronismo	22	16	SYN	54	36	6	86	56	V	118	76	v
Fin bloque transmitido	23	17	ETB	55	37	7	87	57	W	119	77	w
Cancelar	24	18	CAN	56	38	8	88	58	X	120	78	X
Fin medio	25	19	EM	57	39	9	89	59	Y	121	79	y
Sustituto	26	1A	SUB	58	3A	:	90	5A	Z	122	7A	Z
Escape	27	1B	ESC	59	3B	;	91	5B	]	123	7B	{
Separador archivos	28	1C	FS	60	3C	<	92	5C	1	124	7C	ì
Separador grupos	30	1E	RS	62	3E	>	94	5E	٨	126	7D	}
Separador unidades	31	1F	US	63	3F	?	95	5F	_	127	7F	DEL

## **Ampliación ASCII**

### **American Standard Code for Information Interchange**

- ➤ ASCII extendido de 8 bits (+128 símbolos)
- > ASCII extendido por zonas:
  - O Alfabetos: Europa (Occidental, Central..), Árabe, Griego, Hebreo, Turco..
  - Regularizado por la <u>ISO</u> (International Organization for Standardization)

El problema persiste en la actualidad. Debe indicarse en los documentos de texto el sistema de codificación utilizado.

sucesivas ampliaciones..

+ 128 símbolos

ASCII

**ASCII** extendido

128 símbolos originales

### Unicode

Estándar internacional de codificación de caracteres que incluye los símbolos de todas las lenguas del planeta.

Cada carácter codificado con al menos un byte (8 bits)

Hereda los 128 primeros caracteres originales de ASCII y los 256 primeros de ISO8859\_1 (Latin 1)

La Unicode Technical Committee (UTC) define 3 tipos de codificación:

- ➤ UTF-8: La más utilizada ( y más compleja para el ordenador ). Usa 1-4 bytes.
- ➤ UTF-16: Más sencilla que la anterior. Usa 2 o 4 bytes.
- > UTF-32: Todo carácter utiliza 4 bytes. No utilizada.

- 1.ASCII Original
- 2. Lenguas latinas, árabes..
- 3. Chino / Japonés
- 4. Matemáticas..

## **Ejemplos**

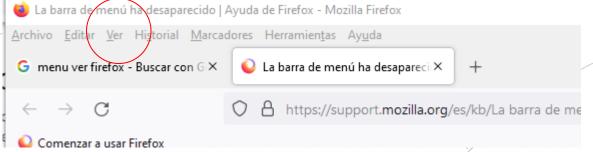
Sistemas de codificación de caracteres configurables en el navegador Mozilla Firefox.

#### Accesible desde:

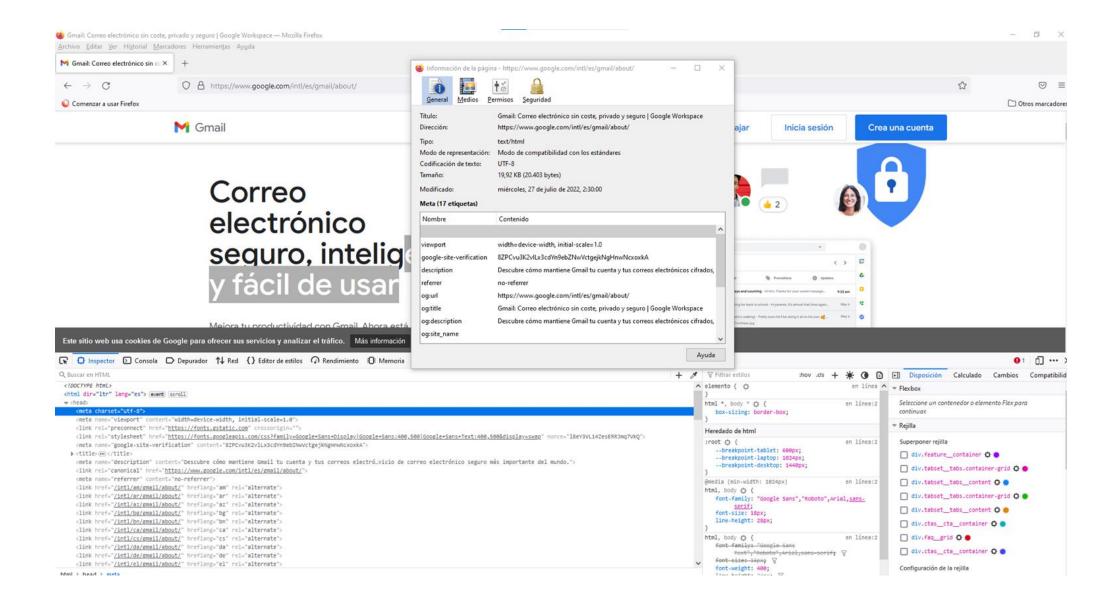
➤ Menú ver -> Repara la codificación del texto

### Otros navegadores.









# 3. Exportar / importar datos

## La compartición de datos entre aplicaciones

### Principales handicaps de las comunicaciones

- ➤ Gran diversidad de dispositivos y sistemas implantados sobre tecnologías de la información en constante cambio.
- > Alta disponibilidad en servicios destinados a sistemas intermedios o usuarios finales.
- ➤ Conflicto entre productos privativos y los estándares propuestos por las principales organizaciones de estandarización. (pdf, MPEG, mp3, JPG, docx, odt, mov ...)



## La compartición de datos entre aplicaciones

### Texto como el formato más versátil de intercambio

- Cualquier dispositivo es capaz de entenderlo.
- ➤ Posibilidad de aplicar un formato para tomar un sentido determinado.
- > Sin posibilidad de almacenar imágenes / vídeos.
- ➤ Parte se interpreta como "otra cosa"
- Queda expuesto desde el punto de la seguridad informática.



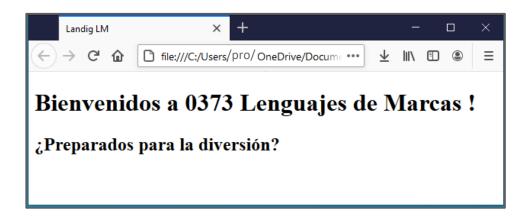
Remitente: Juan.Diego@empresa.com

Teléfono: (+34) 6XX7XX3XX Cuerpo: Hola Miguel!...

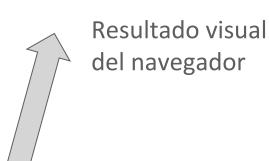
4. Ejemplos de uso de los conjuntos de caracteres

## Páginas HTML

```
<!DOCTYPE html>
<html>
<head>
    <title>Landig LM</title>
   <meta charset="UTF-8">
</head>
<body>
    <h1>Bienvenidos a 0373 Lenguajes de Marcas !</h1>
    <h2>¿Preparados para la diversión?<h2>
</body>
</html>
```

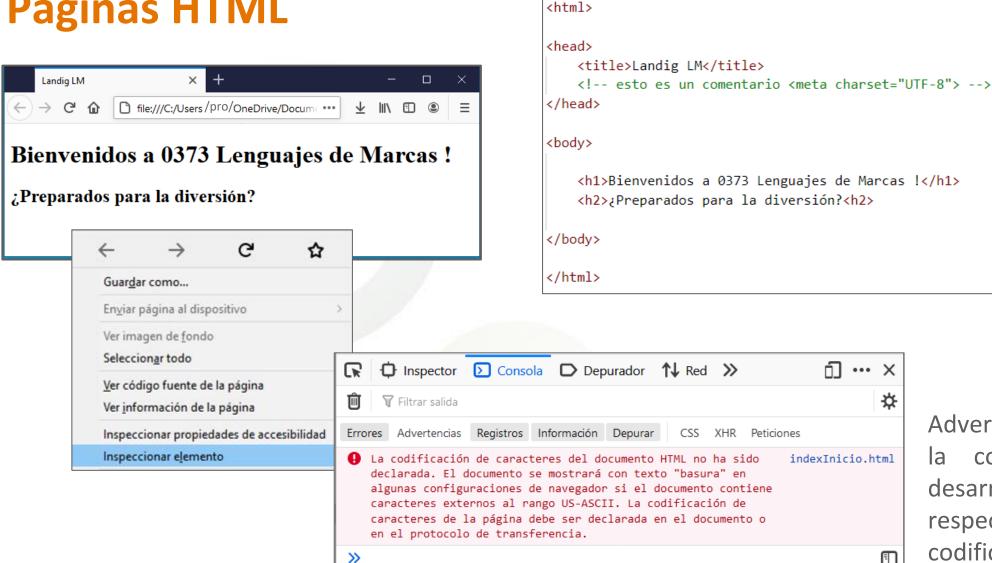


Tipo de codificación empleada en este ejemplo de HTML



Interpretación de la información

## Páginas HTML



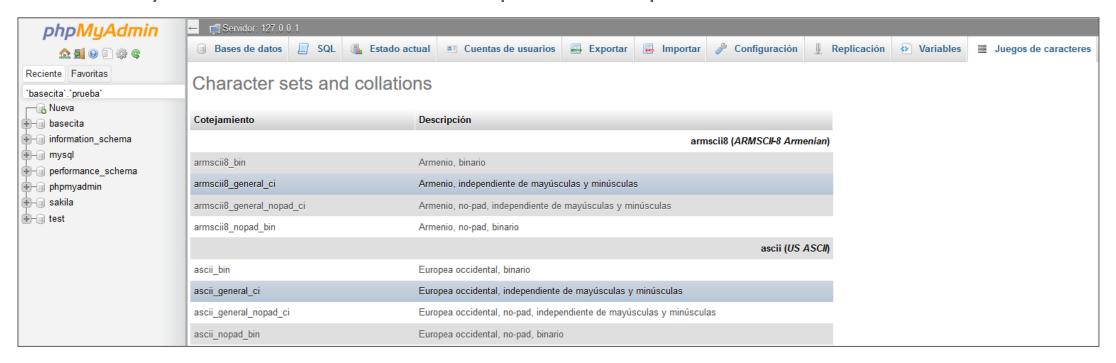
<!DOCTYPE html>

Código fuente sin definición del tipo de codificación

Advertencia la consola desarrollador respecto a la codificación

### Bases de datos

**Cotejamiento (collation)**: cómo la información está ordenada dentro de su conjunto de caracteres y cuál será el criterio de comparación a aplicar sobre sus datos.



Lectura: https://database.guide/what-is-collation-in-databases/

## **Sistemas Operativos**



Compartición de proyectos Java con el IDE Eclipse donde el código fuente (con acentos) podría verse de formas distintas en los equipos

## Solución al conflicto presentado...



Aseguraos que configuráis la codificación de caracteres de los proyectos en vuestro IDE a UTF-8

"Error del euro" que cometeréis en más de una ocasión si no estáis atentos!



<sup>&</sup>lt;sup>1</sup> Error del euro: error que aparece recurrentemente y por el cual los profes de informática cobran 1€ por cada vez que el alumno los comete (el alumno paga, claro está).

# 5. Lenguajes de marcas

#### **Actividad colaborativa**

Seleccionaremos 4 lenguajes de marcas, en este caso van a ser los siguientes:

- Latex
- HTML
- XML
- JSON

Y generaremos 4 grupos de trabajo. Cada uno de los grupos de trabajo será el encargado de codificar en su lenguaje de marcado la información contenida en la siguiente tabla:

Iremos añadiendo en la pizarra colaborativa la información y características de nuestro lenguaje de marcado.

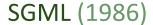
Y para finalizar, un miembro del grupo explicará en 2 minutos al resto de compañeros/as que características tiene nuestro lenguaje de marcado.

Disco	Canción	Cantante	Año
Hey Jude	Can't buy me love	The Beatles	1970
Saturday Night Fever	More than a woman	Bee Gees	1977

## Historia de los lenguajes de marcas

**GML** (1970)





RTF (1987)

HTML (1992)

XML(1998)

JSON(2001)



















Charles Goldfarb:

padre de GML/SGML

Donald Knuth:

Creador de TeX/LateX



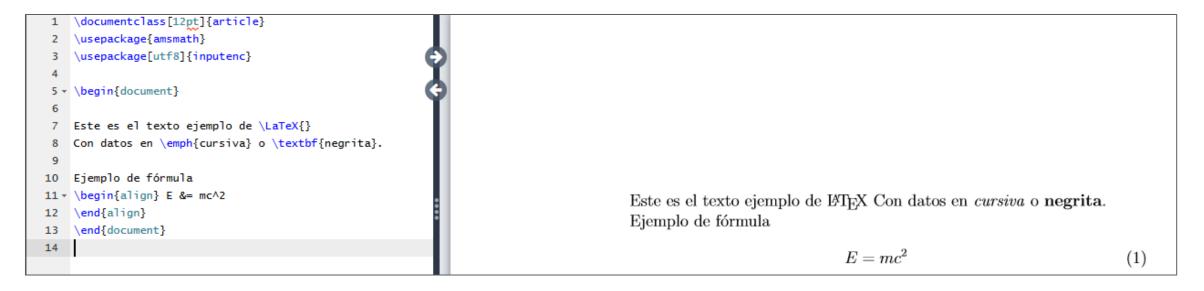
Tim Bernes:

Creador de HTML



# Tex/Latex

Lenguaje de marcas creado en la década de los 70 por Donald Knuth para la producción de documentos científicos con total homogeneidad entre equipos y de gran calidad en sus resultados. Herramienta frecuentemente utilizada en la comunidad científica en la actualidad.



Aspecto de fórmula matemática generada con <u>Overleaf</u> (utiliza LaTeX)

#### **RTF**

Rich Text Format o RTF: lenguaje ideado por Microsoft en 1987 para producir documentos de texto que incluyan anotaciones de formato.



AllTheThings.rtf Rich Text Document 322 KB



file.rtf Rich Text Document 834 bytes



note\_error.rtf Rich Text Document 283 bytes



cd\_catalog.rtf Rich Text Document 1.10 KB



note.rtf Rich Text Document 266 bytes



plant\_catalog.rtf Rich Text Document 1.25 KB





#### **SGML**

- > Versión de GML que estandariza el lenguaje de marcado. Creado por Charles Goldfarb (IBM).
- ➤ Lenguaje predecesor a XML y base de HTML (Notación de etiquetas encerradas entre '<' '>' y concepto de etiqueta de cierre).





#### **HTML**

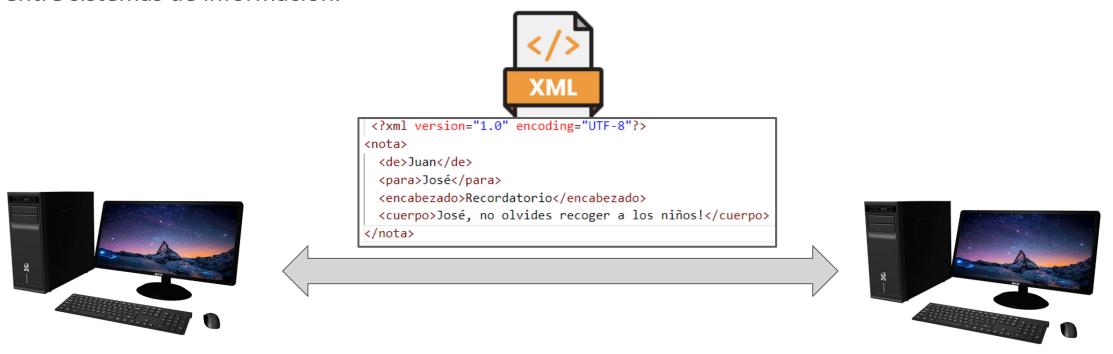
Hypertext Markup Language (lenguaje de marcado de hipertexto). Creado por Tim Bernes con el propósito de generar documentos transportables a través de internet donde fuera posible el hipertexto. Lenguaje de marcas que actúa como piedra angular en las páginas web.



Código fuente en html

#### **XML**

XML o eXtensible Markup Language (Lenguaje de Marcado Extensible). Subconjunto de SGML. Uno de los principales lenguajes de marcado utilizado en operaciones de importación/exportación de datos entre sistemas de información.



#### **JSON**

Javascript Object Notation o JSON: notación procedente del lenguaje JavaScript (ECMAScript 1999). Texto con formato que compite con XML para el intercambio de datos entre sistemas de la información por ser más ligero. No se considera un lenguaje de marcas como tal.



### Tipos de lenguajes de marcas

**Orientados a la presentación**: texto común + palabras encerradas en símbolos especiales (indicaciones de formato) HTML, Texto con indicaciones de negrita/cursiva. Las etiquetas no son visibles.

**Orientados a la descripción**: símbolos especiales que dan significado al texto (pero no aportan características de formato). XML / JSON

Orientados a procedimientos: texto marcado que es interpretado como órdenes a procesar (LateX / PostScript).

## Bibliografía / Webgrafía

W3Schools: <a href="https://www.w3schools.com//">https://www.w3schools.com//</a>

Overleaf: <a href="https://www.overleaf.com/">https://www.overleaf.com/</a>

Tabla ASCII: <a href="https://ca.wikipedia.org/wiki/ASCII">https://ca.wikipedia.org/wiki/ASCII</a>



