# Entidades de carácter y entidades numéricas

En esta lección se explica cómo incluir cualquier carácter o secuencia de caracteres Unicode en una página web, mediante entidades de carácter o entidades numéricas.

Esta lección es la segunda lección de un grupo de tres lecciones que se recomienda leer en el siguiente orden:

- <u>El juego de caracteres UTF-8</u>, que explica qué es un juego de caracteres informático y, concretamente, qué es el juego de caracteres UTF-8.
- Entidades de carácter y entidades numéricas, que explica cómo incluir caracteres Unicode en una página web
- Secuencias, que muestra cómo se definen las secuencias de caracteres Unicode que corresponden a dibujos (banderas, emojis, etc.)

Como complemento a las lecciones anteriores, las páginas siguientes muestran diferentes símbolos Unicode agrupados por temas:

- <u>Símbolos gráficos</u>, que muestra los caracteres Unicode que corresponden a símbolos gráficos (incluye tanto los símbolos en blanco y negro como los emojis que tiene versión en modo texto)
- Emojis, que muestra únicamente los caracteres Unicode que corresponden a los emojis (en color)
- Géneros, que muestra las secuencias Unicode que corresponden a personas con un género (masculino o femenino) determinado
- Colores de piel, que muestra las secuencias Unicode que corresponden a personas con un color de piel determinado
- Parejas, que muestra las secuencias Unicode que corresponden a parejas y familias
- Banderas, que muestra las secuencias Unicode que corresponden a banderas
- Otras, que muestra las secuencias Unicode no incluidas en las páginas anteriores
- <u>Twemoji</u>, que muestra los dibujos de emojis creados por el proyecto <u>Twemoji</u> con licencia libre <u>CC-BY 4.0</u> (es un fichero grande que puede tardar un poco en cargarse)

## Qué son las entidades de carácter y las entidades numéricas

En las primeras versiones de HTML se especificaba que el juego de caracteres utilizado por las páginas web debía ser el juego de caracteres <u>ISO 8859-1</u>, más conocido como Latin-1.

Para poder incluir en un página web caracteres no incluidos en ese juego de caracteres se debían utilizar las entidades de carácter. Una entidad de carácter es una referencia a un carácter y se escribe con la sintaxis &nombre; es decir, empieza por el símbolo & (en inglés, ampersand, en español, et), a continuación se escribe el nombre del carácter y termina por un punto y coma.

Por ejemplo, la letra griega alfa (α) se podía incluir en una página web con la entidad de carácter α.

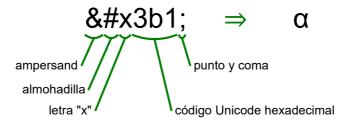


En las primeras versiones de HTML se fueron añadiendo entidades de carácter hasta llegar en HTML 4 a 252 entidades de carácter, que se muestran al final de esta lección, en el apartado <u>Lista de entidades de carácter</u>.

La norma RFC 2070, publicada en enero de 1997, permitió el uso de Unicode como juego de caracteres en HTML, por lo que se amplió el concepto de entidad de carácter a entidad numérica, para permitr la referencia a cualquier carácter Unicode, con la sintaxis &número; es decir empieza por el símbolo &, a continuación se escribe el código Unicode del carácter (en notación decimal o en notación hexadecimal precedida por una letra "x") y termina por un punto y coma.

Por ejemplo, la letra griega alfa (α), cuyo código de carácter Unicode es 945 (3B1 en hexadecimal) se podía incluir en una página web con la entidad de carácter α o con las entidades numéricas α o α.





Otras recomendaciones (por ejemplo, MathML), han definido más entidades de carácter que los navegadores también reconocen. La recomendación XML Entity Definitions for Characters (3ª edición, publicada en marzo de 2023) ha ido recopilando las entidades de carácter más empleadas en páginas web.

En HTML 5 la lista de entidades de carácter se amplió a más 2200. En la página Entidades de carácter HTML se muestra la lista completa.

Las entidades de carácter son más fáciles de utilizar porque es más fácil recordar un nombre que un código, pero al no existir entidades de carácter para todos los caracteres Unicode, en la mayoría de casos se tiene que utilizar la entidad numérica.

Las entidades de carácter más antiguas (las incluidas en HTML 3.2 y anteriores) se pueden escribir sin el punto y coma final, lo que puede dar lugar a resultados inesperados, como muestra el siguiente ejemplo. En él, los navegadores no muestran la palabra Microsoft porque **&micro** es la entidad de carácter del carácter  $\mu$ . La página, además, no es válida por el mismo motivo (el validador requiere que la entidad de carácter termine por punto y coma).



Durante años la informática personal ha estado dominada por intel&microsoft.

Durante años la informática personal ha estado dominada por intelµsoft.



Oprante años la informática personal ha estado dominada por intel& microsoft.

Durante años la informática personal ha estado dominada por intel&microsoft.

Las fuentes no suelen incluir todos los caracteres Unicode existentes, por lo que los navegadores suelen recurrir a distintas fuentes dependiendo de los caracteres a mostrar. Estas fuentes pueden ser alguna de las disponibles en el sistema operativo o estar incluidas en el navegador. Por eso hay que tener cuidado al incluir caracteres Unicode, sobre todo los incluidos en las versiones más recientes de Unicode, porque los visitantes pueden no ver esos caracteres.

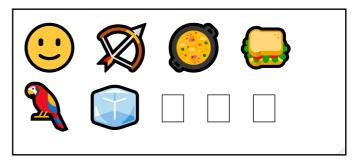
En el caso de Windows, para mostrar emojis, Microsoft distribuye gratuitamente la fuente <u>Segoe UI Emoji</u>, que contiene muchos caracteres Unicode. Pero Microsoft, por desgracia, ya no actualiza la fuente Segoe UI Emoji en Windows 10 una vez se ha publicado Windows 11. Windows 10 usa <u>la versión 1.29</u> de la fuente y Windows 11 usa actualmente (octubre de 2022) <u>la versión 1.33</u>, que contiene más caracteres y con un estilo diferente.

En el caso de Android (y cada vez más las distribuciones Linux), se incluyen las <u>fuentes Noto</u>, una familia de fuentes libres creada por Google. Para mostrar emojis, están disponibles las fuentes <u>Noto Emoji</u>.

Desde Firefox 61 (junio de 2018), Firefox utiliza la fuente <u>Twemoji-colr</u>, una versión en formato COLR de la fuente <u>Twemoji</u>, creada por Twitter, para mostrar los emojis no disponibles en la fuente del sistema. Anteriormente, desde Firefox 50 (noviembre de 2016) hasta Firefox 60 (mayo de 2018) Firefox utilizó la versión libre de la fuente EmojiOne. Aunque Twemoji contiene actualmente (octubre de 2022) los emojis hasta Unicode 14, Twemoji-colr todavía no se ha actualizado desde Unicode 13.1.

El ejemplo siguiente contiene unos caracteres Unicode (concretamente emojis) para que usted mismo compruebe cómo se ven en su navegador y sistema operativo:

🙂 🏹
🥘 🥪 🦜 🧊 🪢
🛝 🪼



Las capturas siguientes muestran cómo se ve actualmente (octubre de 2022) el ejemplo anterior dependiendo de la fuente empleada (lo que a su vez depende del navegador y el sistema operativo). La fecha indica el año en que cada uno de los emojis de la imagen se añadió a Unicode:

• Fuente Noto (Android y otros):

2014

2015

2016

2017

2018

2019

2020

2021

2022



















• Fuente Segoe UI Emoji (Windows 11):



















• Fuente Segoe UI Emoji (Windows 10):



















• Fuente Apple Color Emoji (iOS, macOS):



















• Fuente Twemoji (Firefox en Windows y otros):



















### Nota:

<u>Bugs de Firefox relacionados con Twemoji</u>

# Los caracteres especiales de HTML

En HTML hay cinco caracteres que tienen un significado especial. Si se quieren mostrar esos caracteres en una página web, escribirlos directamente puede hacer que la página no sea válida o puede hacer que el navegador no muestre el resultado esperado. En lugar de escribirlos directamente, se deben utilizar las entidades de carácter o las entidades numéricas correspondientes, que se muestran en la tabla siguiente. En el caso de los dos primeros (ampersand y menor que) siempre hay que escribirlos como entidad, en el caso del tercero (mayor que) es aconsejable, en el caso de los dos últimos (comillas dobles o simples) hay que escribirlas como entidad si queremos escribirlas en el valor de un atributo (lo que no es muy habitual).

Carácter	Nombre	Entidad de carácter	Entidad numérica	Recomendación	Descripción
&	amp	&	&	HTML 2.0	et (en inglés, ampersand)
<	It	<	<	HTML 2.0	menor
>	gt	>	>	HTML 2.0	mayor
"	quot	"	"	HTML 2.0	comillas
•	apos	'	'	XHTML 1.0	apóstrofo

Estos son los problemas que provoca utilizar los caracteres en vez de las entidades de carácter:

• El carácter *ampersand* (&) es el carácter inicial de las entidades de carácter o de las entidades numéricas. En HTML se puede escribir el carácter & directamente y la página será válida, salvo si el texto que sigue al & se puede interpretar como una entidad de carácter. En ese caso, además, el navegador puede mostrar un resultado no esperado. Por ello se aconseja utilizar siempre la entidad de carácter & amp; cuando se quiera mostrar el carácter &, para no tener que pensar en cada caso si el navegador va a mostrar el resultado esperado.

En el ejemplo siguiente el navegador muestra el resultado esperado puesto que el texto que sigue al carácter & no corresponde a ninguna entidad de carácter o numérica. La página es válida.

Nota: El ejemplo se ha marcado con un aspa amarilla porque, aunque es correcto, se desaconseja no utilizar la entidad de carácter.



Calvin & Hobbes es una de las grandes tiras cómicas del siglo XX.

Calvin & Hobbes es una de las grandes tiras cómicas del siglo XX.



26/10/23, 20:20

Calvin & mp; Hobbes es una de las grandes tiras cómicas del siglo XX.

Calvin & Hobbes es una de las grandes tiras cómicas del siglo XX.

En el ejemplo siguiente el navegador no muestra el resultado esperado porque las entidades de carácter más antiguas (las incluidas en HTML 3.2 y anteriores) se pueden escribir sin el punto y coma final. Los navegadores no muestran la palabra Microsoft porque &micro es la entidad de carácter del carácter µ. La página, además, no es válida por el mismo motivo (el validador requiere que la entidad de carácter termine por punto y coma).



Ourante años la informática personal ha estado
dominada por intel&microsoft.

Durante años la informática personal ha estado dominada por intelµsoft.



Ourante años la informática personal ha estado dominada por intel& microsoft.

Durante años la informática personal ha estado dominada por intel&microsoft.

• El carácter "menor que" (<) en HTML indica que se abre una etiqueta, por lo que no se debe escribir en una página web si no es por ese motivo. Aunque los navegadores sean capaces de mostrar el resultado esperado, la página no es válida. Para que la página sea correcta, se debe escribir la entidad de carácter &It;.

En el ejemplo siguiente, los navegadores consiguen mostrar el resultado esperado, pero la página no es válida:



0 < 1</p>

0 < 1



0 < 1

0 < 1

En el ejemplo siguiente, los navegadores ni siquiera consiguen sacar el resultado esperado:



0<x<5</p>

0



0<x&lt;5

0<x<5

• El carácter "mayor que" (>) en HTML indica que se cierra una etiqueta y puede dar problemas en navegadores antiguos, por lo que se aconseja escribir la entidad de carácter >, aunque su uso no es incorrecto (las páginas son válidas).



1 > 0

1 >



1 > 0

1 > 0

• Los caracteres comillas dobles (") y simples (') en HTML indica el principio y final de los valores de los atributos, por tanto no se pueden escribir en un atributo y en su lugar se deben escribir las entidades de carácter **&quot**; o **&apos**;.



Sitúe el cursor sobre este
párrafo para ver un mensaje oculto.

Sitúe el cursor sobre este párrafo para ver un mensaje oculto.



Sitúe el cursor sobre
este párrafo para ver un mensaje oculto.

Sitúe el cursor sobre este párrafo para ver un mensaje oculto.

#### Lista de entidades de carácter HTML

En HTML 5 la lista de entidades de carácter se amplió a más 2200. En la página Entidades de carácter HTML se muestra la lista completa.

# Lista de entidades de carácter HTML 4

26/10/23, 20:20 Entidades de carácter y entidades numéricas. HTML. Páginas web HTML y hojas de estilo CSS. Bartolomé Sintes Marco. www...

La tabla siguiente muestra las 252 entidades de carácter definidas en la recomendación HTML 4. De cada carácter se muestra:

- el carácter
- el nombre del carácter
- la entidad de carácter con la notación &nombre;
- la entidad numérica con la notación &código;, en la que el código se expresa como número hexadecimal
- la recomendación HTML en la que se introdujo la entidad de carácter
- la descripción del carácter

Referencia: Lista de entidades de carácter de HTML en Wikipedia

Ο

HTML 4.0

letra griega omicron mayúscula

Omicron Ο

Carácter	Nombre	Entidad de carácter	Entidad numérica	Recomendación	Descripción
	frasl	⁄	⁄	HTML 4.0	barra de fracción
	euro	€	€	HTML 4.0	euro
3	image	ℑ	ℑ	HTML 4.0	parte imaginaria
$\mathcal{O}$	weierp	℘	℘	HTML 4.0	función elíptica de Weierstrass
$\overline{\mathfrak{R}}$	real	ℜ	ℜ	HTML 4.0	parte real
TM	trade	™	™	HTML 4.0	marca registrada
×	alefsym	ℵ	ℵ	HTML 4.0	primer cardinal transfinito (numerable)
<b>←</b>	larr	←	←	HTML 4.0	flecha hacia la izquierda
$\uparrow$	uarr	↑	↑	HTML 4.0	flecha hacia arriba
$\rightarrow$	rarr	→	→	HTML 4.0	flecha hacia la derecha
$\overline{}$	darr	↓	↓	HTML 4.0	flecha hacia abajo
$\leftrightarrow$	harr	↔	↔	HTML 4.0	flecha derecha izquierda (relación en notación z)
<u></u>	crarr	↵	↵	HTML 4.0	retorno de carro
	IArr	&IArr	⇐	HTML 4.0	flecha doble hacia la izquierda
<u> </u>	uArr	⇑	⇑	HTML 4.0	flecha doble hacia arriba
$\Rightarrow$	rArr	⇒	⇒	HTML 4.0	flecha doble hacia la derecha
	dArr	⇓	⇓	HTML 4.0	flecha doble hacia abajo
$\Longrightarrow$	hArr	⇔	⇔	HTML 4.0	flecha doble derecha izquierda
<u>∀</u>	forall	∀	∀	HTML 4.0	para todo
д	part	∂	∂	HTML 4.0	diferencial parcial
	exist	∃	∃	HTML 4.0	existe
<u>Ø</u>	empty	∅	∅	HTML 4.0	conjunto vacío
∇	nabla	∇	∇	HTML 4.0	operador nabla
€	isin	∈	∈	HTML 4.0	pertenece a
∉	notin	∉	∉	HTML 4.0	no pertenece a
<u> </u>	ni	∋	∋	HTML 4.0	contiene a
<del>-</del>	prod	∏	∏	HTML 4.0	productorio
	sum	∑	∑	HTML 4.0	sumatorio
<u> </u>	minus	−	−	HTML 4.0	menos
*	lowast	∗	∗	HTML 4.0	operador asterisco raíz cuadrada
	radic	√	√ ∝	HTML 4.0	proporcional a
<u>∝</u>	prop infin	∝ ∞	∞	HTML 4.0	infinito
	ang	∠	∠	HTML 4.0	ángulo
	and	∧	∧:	HTML 4.0	operador lógico y
$\frac{\lambda}{V}$	or	∨	∨	HTML 4.0	operador lógico o
Ť	cap	∩	∩	HTML 4.0	intersección
U	cup	∪	∪	HTML 4.0	unión
Ť	int	∫	∫	HTML 4.0	integral
	there4	∴	∴	HTML 4.0	por lo tanto
~	sim	∼	∼	HTML 4.0	operador tilde
$\cong$	cong	≅	≅	HTML 4.0	aproximadamente igual a
~	asymp	≈	≈	HTML 4.0	casi igual a
<del></del>	ne	≠	≠	HTML 4.0	distinto de
	equiv	≡	≡	HTML 4.0	idéntico a
≤	le	≤	≤	HTML 4.0	inferior o igual
≥	ge	≥	≥	HTML 4.0	superior o igual
$\subset$	sub	⊂	⊂	HTML 4.0	subconjunto de
	sup	⊃	⊃	HTML 4.0	superconjunto de
⊄	nsub	⊄	⊄	HTML 4.0	no subconjunto de
	sube	⊆	⊆	HTML 4.0	subconjunto de o igual a
⊇	supe	⊇	⊇	HTML 4.0	superconjunto de o igual a
	oplus	⊕	⊕	HTML 4.0	signo más rodeado por un círculo
$-\otimes$	otimes	⊗	⊗	HTML 4.0	signo por rodeado por un círculo
	perp	⊥	⊥	HTML 4.0	perpendicular
	sdot	⋅	⋅	HTML 4.0	operador punto
	lceil	⌈	⌈	HTML 4.0	corchete superior izquierdo
	rceil	⌉	⌉	HTML 4.0	corchete superior derecho
	Ifloor	⌊	⌊	HTML 4.0	corchete inferior izquierdo
	rfloor	⌋	⌋	HTML 4.0	corchete inferior derecho
	lang	⟨	〈	HTML 4.0	corchete angular izquierdo
	rang	⟩	〉	HTML 4.0	corchete angular derecho

26/10/23, 20:20 Entidades de carácter y entidades numéricas. HTML. Páginas web HTML y hojas de estilo CSS. Bartolomé Sintes Marco. www...

Caracter	Nombre	Entidad de carácter	Entidad numérica	Recomendación	Descripción
$\Diamond$	loz	◊	◊	HTML 4.0	rombo
<b>♦</b>	spades	♠	♠	HTML 4.0	picas
*	clubs	♣	♣	HTML 4.0	tréboles
*	hearts	♥	♥	HTML 4.0	corazones
<b>*</b>	diams	♦	♦	HTML 4.0	rombos

Última modificación de esta página: 8 de marzo de 2023



Esta página forma parte del curso <u>Páginas web HTML y hojas de estilo CSS</u> por <u>Bartolomé Sintes Marco</u> que se distribuye bajo una <u>Licencia Creative Commons Reconocimiento-CompartirIgual 4.0 Internacional (CC BY-SA 4.0)</u>.