

Codificación de caracteres

Código Morse

1844

Código inventado por Samuel F. B. Morse.



Primero en utilizarse para las comunicaciones de larga distancia a través del telégrafo.

Más adelante

Código Baudot o Código Murray utilizado en por las teleimpresoras

Podían codificar y decodificar caracteres utilizando 5 bits y sólo se disponía de 32 caracteres.

Década de 1960

Se adoptó el código **ASCII** (American Standard Code for Information Interchange)





A · -	J · - - -	S · · ·	2 · · - - -
B - · · ·	K - · -	T -	3 · · · - -
C - · - ·	L · - · ·	U · · -	4 · · · · -
D - · ·	M - -	V · · · -	5 · · · · ·
E ·	N - ·	W · - -	6 - · · · ·
F · · - ·	O - - -	X - · · -	7 - - · · ·
G - - ·	P · - - ·	Y - · - -	8 - - - · ·
H · · · ·	Q - - · -	Z - - · ·	9 - - - - ·
I · ·	R · - ·	1 · - - - -	0 - - - - -

Representación de datos alfanuméricos

- *Problema*
- La memoria de un ordenador no guarda caracteres
- La información la guarda en formato **binario**
- Cada carácter tiene un **código digital**

Código ASCII

Ascii fue creado en 1963, y es una correspondencia entre cadenas de bits y una serie de símbolos (alfanuméricos y otros).

Esto permite comunicación entre dispositivos digitales, así como procesar y almacenar la información.

Caracteres:

```
!"#$%&'()*+,-./
0123456789:;<=>?
@ABCDEFGHIJKLMNO
PQRSTUVWXYZ[\]^_
`abcdefghijklmnopqrstuvwxyz{|}~
```

Ejemplo:

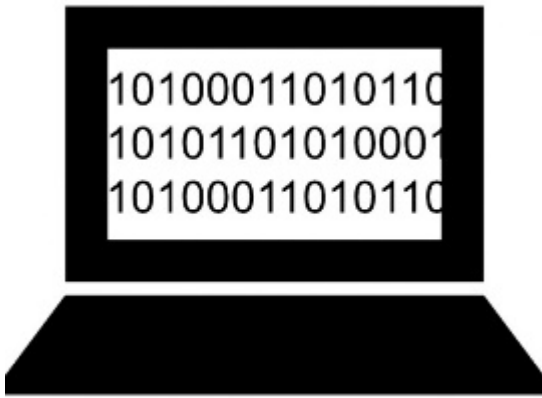
Ascii Text Value	Binary Value
Esta información se traduce en el ordenador al sistema binario, en forma de 1 y 0.	01000101 01110011 01110100 01100001 00100000 01101001 01101110 01100110 01101111 01110010 01101101 01100001 01100011 01101001 11110011 01101110 00100000 01110011 01100101 00100000 01110100 01110010 01100001 01100100 01110101 01100011 01100101 00100000 01100101 01101110 00100000 01100101 01101100 00100000 01101111 01110010 01100100 01100101 01101110 01100001 01100100 01101111 01110010 00100000 01100001 01101100 00100000 01110011 01101001 01110011 01110100 01100101 01101101 01100001 00100000 01100010 01101001 01101110 01100001 01110010 01101001 01101111 00101100 00100000 01100101 01101110 00100000 01100110 01101111 01110010 01101101 01100001 00100000 01100100 01100101 00100000 00110001 00100000 01111001 00100000 00110000 00101110

- Representaba caracteres utilizando __ 7 bits __
- 128 caracteres posibles, enumerados del 0 al 127
 - _Códigos 0 al 31 _
 - Se denominan *caracteres de control*
 - Tienen efecto sobre cómo se procesa el texto.
 - _Códigos 65 al 90 _
 - Representan las letras *mayúsculas*
 - _Códigos 97 al 122 _
 - Representan las letras *minúsculas*

<i>H</i>	<i>O</i>	<i>L</i>	<i>A</i>
01001000	01101111	01101100	01100001

	·0	·1	·2	·3	·4	·5	·6	·7	·8	·9	·A	·B	·C	·D	·E	·F
0·	U+0000	U+0001	U+0002	U+0003	U+0004	U+0005	U+0006	U+0007	U+0008	U+0009	U+000A	U+000B	U+000C	U+000D	U+000E	U+000F
1·	U+0010	U+0011	U+0012	U+0013	U+0014	U+0015	U+0016	U+0017	U+0018	U+0019	U+001A	U+001B	U+001C	U+001D	U+001E	U+001F
2·	U+0020	!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3·	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4·	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5·	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6·	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7·	p	q	r	s	t	u	v	w	x	y	z	{		}	~	

Caracteres de control				Símbolos gráficos											
Nombre	Dec	Binario	Hex	Símbolo	Dec	Binario	Hex	Símbolo	Dec	Binario	Hex	Símbolo	Dec	Binario	Hex
NUL	0	0000000	00	space	32	0100000	20	@	64	1000000	40	`	96	1100000	60
SOH	1	0000001	01	!	33	0100001	21	A	65	1000001	41	a	97	1100001	61
STX	2	0000010	02	"	34	0100010	22	B	66	1000010	42	b	98	1100010	62
ETX	3	0000011	03	#	35	0100011	23	C	67	1000011	43	c	99	1100011	63
EOT	4	0000100	04	\$	36	0100100	24	D	68	1000100	44	d	100	1100100	64
ENQ	5	0000101	05	%	37	0100101	25	E	69	1000101	45	e	101	1100101	65
ACK	6	0000110	06	&	38	0100110	26	F	70	1000110	46	f	102	1100110	66
BEL	7	0000111	07	'	39	0100111	27	G	71	1000111	47	g	103	1100111	67
BS	8	0001000	08	(40	0101000	28	H	72	1001000	48	h	104	1101000	68
HT	9	0001001	09)	41	0101001	29	I	73	1001001	49	i	105	1101001	69
LF	10	0001010	0A	*	42	0101010	2A	J	74	1001010	4A	j	106	1101010	6A
VT	11	0001011	0B	+	43	0101011	2B	K	75	1001011	4B	k	107	1101011	6B
FF	12	0001100	0C	,	44	0101100	2C	L	76	1001100	4C	l	108	1101100	6C
CR	13	0001101	0D	-	45	0101101	2D	M	77	1001101	4D	m	109	1101101	6D
SO	14	0001110	0E	.	46	0101110	2E	N	78	1001110	4E	n	110	1101110	6E
SI	15	0001111	0F	/	47	0101111	2F	O	79	1001111	4F	o	111	1101111	6F
DLE	16	0010000	10	0	48	0110000	30	P	80	1010000	50	p	112	1110000	70
DC1	17	0010001	11	1	49	0110001	31	Q	81	1010001	51	q	113	1110001	71
DC2	18	0010010	12	2	50	0110010	32	R	82	1010010	52	r	114	1110010	72
DC3	19	0010011	13	3	51	0110011	33	S	83	1010011	53	s	115	1110011	73
DC4	20	0010100	14	4	52	0110100	34	T	84	1010100	54	t	116	1110100	74
NAK	21	0010101	15	5	53	0110101	35	U	85	1010101	55	u	117	1110101	75
SYN	22	0010110	16	6	54	0110110	36	V	86	1010110	56	v	118	1110110	76
ETB	23	0010111	17	7	55	0110111	37	W	87	1010111	57	w	119	1110111	77
CAN	24	0011000	18	8	56	0111000	38	X	88	1011000	58	x	120	1111000	78
EM	25	0011001	19	9	57	0111001	39	Y	89	1011001	59	y	121	1111001	79
SUB	26	0011010	1A	:	58	0111010	3A	Z	90	1011010	5A	z	122	1111010	7A
ESC	27	0011011	1B	;	59	0111011	3B	[91	1011011	5B	{	123	1111011	7B
FS	28	0011100	1C	<	60	0111100	3C	\	92	1011100	5C		124	1111100	7C
GS	29	0011101	1D	=	61	0111101	3D]	93	1011101	5D	}	125	1111101	7D
RS	30	0011110	1E	>	62	0111110	3E	^	94	1011110	5E	~	126	1111110	7E
US	31	0011111	1F	?	63	0111111	3F	_	95	1011111	5F	Del	127	1111111	7F



Código ASCII Extendido

Los problemas con el ASCII original eran que fue desarrollado para utilizarse con inglés y no poseía caracteres acentuados o caracteres específicos de otros idiomas, por lo que para codificar estos caracteres se necesitaba un sistema de códigos distinto.

La solución fue la extensión del código ASCII para 8 bits (byte) y la codificación de hasta 256 caracteres (ASCII extendido). Esto permitió añadir caracteres acentuados, símbolos y otras letras de idiomas distintos al inglés.

Sin embargo, el ASCII extendido sigue siendo limitado y no soporta caracteres complejos como los usados en el **chino o el japonés**, para los cuales se usan otros sistemas de codificación.



- Este código asigna los valores del __0 al 255 __ para las mayúsculas, las minúsculas, los dígitos, las marcas de puntuación y otros símbolos.
- *Compatibilidad con ASCII*
- Para no romper la compatibilidad con ASCII, se hace que el primer bit signifique
 - 0: Los 7 bits inferiores siguen la tabla ASCII
 - 1: Los 7 bits inferiores siguen otra tabla
- *Caracteres ASCII extendido*
 - Caracteres alfabéticos no ingleses
 - Símbolos de moneda no ingleses
 - Letras griegas
 - Símbolos matemáticos
 - Caracteres para gráficos
 - Caracteres para gráficos de barras
 - Caracteres sombreados

DEC	BIN	Symbol
248	11111000	ø
249	11111001	ù
250	11111010	ú
251	11111011	û
252	11111100	ü
253	11111101	ý
254	11111110	þ
255	11111111	ÿ

É	Ê	Ë	Ì	Í	Î	Ï
Ù	Ú	Û	Ü	Ý	Þ	ß
é	ê	ë	ì	í	î	ï
ù	ú	û	ü	ý	þ	ÿ
Ĉ	Ċ	ċ	Č	č	Ď	ď
Ė	Ė	ė	Ĝ	ĝ	Ğ	ğ
Ī	Ĭ	ī	Ĳ	ĩ	İ	ı
Ł	ł	Ł	ł	Ł	ł	Ł
Ń	Ń	ń	Ō	ō	Ŏ	ő
Ŕ	Ŗ	ŗ	Ŝ	ŝ	Ş	ş
Ũ	Ū	ū	Ů	ů	Ű	ű
Ž	ž	Ž	ž	Ž	ž	ƒ
Ð	Đ	đ	đ	đ	đ	đ
Ɓ	Ƀ	Ʉ	Ʌ	Ɇ	ɇ	Ɉ
Σ	λ	τ	T	f	τ	U
ε	ε	ε	ε	ε	ε	p

UTF

En 1991 fue establecida la norma de codificación universal de caracteres (UTF) que permite codificar todos los caracteres de cualquier lenguaje usando 16 bits por carácter (hasta 65.536 caracteres).

La organización de los códigos se organiza en “**planos**” cada uno con 216 caracteres, limitando el número de planos a 17, con un máximo de 1.114.112 caracteres (17 x 216).

Unicode® 8.0.0

Released: 2015 June 17 ([Announcement](#))

Código Unicode	Carácter	Nombre
U+0041	A	Letra A Latina Mayúscula
U+00D1	Ñ	Letra Ñ Latina Mayúscula
U+03B4	δ	Letra Delta Griega Minúscula
U+20AC	€	Símbolo del Euro
U+0259	ə	Letra "Schwa" del IPA
U+13A3	Ꭰ	Letra O Cherokee
U+263A	☺	Smiley
U+6F22	漢	Carácter chino "Han"
U+3042	あ	Letra A del alfabeto hiragana (japonés)
U+FEB6	ش	Letra Sheen en forma final (árabe)
U+2815	⠕	Letra O en Código Braille
U-00010330	ⱦ	Letra A en Gótico
U-0001D4DB	ℒ	Letra L caligráfica matemática (transformada de Laplace)
U-0001D571	ℱ	Letra F gótica matemática (transformada de Fourier)
U-0001D160	♪	Nota Musical Corchea