

Codifica ASCII e Codifica UNICODE

ASCII

ASCII (*American Standard Code for Information Interchange*) è una codifica base per i caratteri di un testo a **7-bit** che consente di rappresentare **128** caratteri diversi tra loro, di cui i primi 32 sono caratteri speciali di formattazione del testo.

E' stato poi esteso a **8-bit** (1 byte) con il nome di *Extended ASCII* per poter rappresentare **256** caratteri.

Ogni byte, rappresentato dalla combinazione di due caratteri esadecimali (semi-byte o nibble), facilita la decodifica per mezzo di spiazzamento all'interno della tabella ASCII.

ASCII

Bit Mask				6	0	0	0	0	1	1	1	1
				5	0	0	1	1	0	0	1	1
				4	0	1	0	1	0	1	0	1
				Col→	0	1	2	3	4	5	6	7
3	2	1	0	Row↓								
0	0	0	0	0	NUL	DLE	SP	0	@	P	`	p
0	0	0	1	1	SOH	DC1	!	1	A	Q	a	q
0	0	1	0	2	STX	DC2	"	2	B	R	b	r
0	0	1	1	3	ETX	DC3	#	3	C	S	c	s
0	1	0	0	4	EOT	DC4	\$	4	D	T	d	t
0	1	0	1	5	ENQ	NAK	%	5	E	U	e	u
0	1	1	0	6	ACK	SYN	&	6	F	V	f	v
0	1	1	1	7	BEL	ETB	'	7	G	W	g	w
1	0	0	0	8	BS	CAN	(8	H	X	h	x
1	0	0	1	9	HT	EM)	9	I	Y	i	y
1	0	1	0	A	LF	SUB	*	:	J	Z	j	z
1	0	1	1	B	VT	ESC	+	;	K	[k	{
1	1	0	0	C	FF	FS	,	<	L	\	l	
1	1	0	1	D	CR	GS	-	=	M]	m	}
1	1	1	0	E	SO	RS	.	>	N	^	n	~
1	1	1	1	F	SI	US	/	?	O	_	o	DEL

UNICODE

UNICODE è stato originariamente pensato come una codifica a **16-bit** (2 byte) dando la possibilità di codificare **65,535** caratteri di cui i primi sono esattamente gli stessi della codifica ASCII, in maniera tale da mantenere la compatibilità con il sistema preesistente.

Nelle versioni successive, la codifica è evoluta per ammettere un maggior numero di bit (fino a 32) in maniera variabile, e suddividendo le varie formattazioni in quelli che sono denominati *piani*.

Il **piano (zero) di base multilinguistico** (BMP) comprende i caratteri originali UNICODE a **16-bit**.