

8. I codici ASCII e Unicode

A ogni carattere (alfabetico, numerico, speciale) della tastiera si assegna un codice convenzionale.

La prima codifica universalmente accettata fu la cosiddetta **ASCII**

(ASCII = American Standard Code for Information Interchange).

Ogni carattere vi veniva rappresentato per mezzo di 1 Byte (=sequenza di 8 bit).

Osserviamo che una sequenza di 8 bit consente esattamente $2^8=256$ combinazioni diverse; quindi, una codifica a 8 bit può coinvolgere al massimo 256 differenti caratteri, non di più.

La **TABELLA ASCII STANDARD**

contiene i caratteri, alfanumerici e speciali, di uso più comune; in essa, i caratteri codificati sono in numero di 128.

Eccone qui a destra una parte.

Se vuoi vedere la tabella completa, clicca QUI →.

Si può ottenere ogni carattere ASCII tenendo premuto il tasto Alt e digitando il codice *decimale* corrispondente col *tastierino numerico*.

Per esempio, si ha la graffa aperta tenendo premuto il tasto Alt e contemporaneamente digitando sul tastierino numerico la sequenza 1-2-3.

Char	Byte	Dec	Hex	Char	Byte	Dec	Hex
0	0011 0000	48	30	[0101 1011	91	5B
1	0011 0001	49	31	\	0101 1100	92	5C
2	0011 0010	50	32]	0101 1101	93	5D
3	0011 0011	51	33	^	0101 1110	94	5E
4	0011 0100	52	34
5	0011 0101	53	35
6	0011 0110	54	36	a	0110 0001	97	61
7	0011 0111	55	37	b	0110 0010	98	62
8	0011 1000	56	38	c	0110 0011	99	63
9	0011 1001	57	39	d	0110 0100	100	64
...
@	0100 0000	64	40	{	0111 1011	123	7B
A	0100 0001	65	41		0111 1100	124	7C
B	0100 0010	66	42	}	0111 1101	125	7D
C	0100 0011	67	43	~	0111 1110	126	7E
D	0100 0100	68	44	Del	0111 1111	127	7F
...

Char = carattere, Dec = codice decimale, Hex = codice esadecimale

Si passò poi alla **TABELLA ASCII ESTESA**, di cui nacquero (sigh!) diverse versioni, a seconda dell'area geografica (e non solo).

Ecco qui a destra una piccola parte della tabella ASCII ESTESA in uso in EUROPA OCCIDENTALE.

Se vuoi vedere la tabella completa, clicca QUI →.

Noterai la presenza di caratteri tipici delle lingue europee, ad esempio le vocali con la dieresi usate in Tedesco, la "c" con "cediglia" del Francese, il punto interrogativo rovesciato dello Spagnolo ...

Char	Byte	Dec	Hex
ü	1000 0001	129	81
é	1000 0010	130	82
...
ç	1000 0111	135	87
...
¿	1010 1000	168	A8
...
½	1010 1011	171	AB

A partire dal 1991 fu sviluppato un nuovo standard, lo standard **UNICODE**.

L'idea iniziale era di rappresentare ogni carattere non più con 1, ma con 2 byte.

Con 2 byte, ossia con una sequenza di 16 bit, le combinazioni possibili sono $2^{16}=65536$.

Si sarebbero così potuti comprendere tutti gli ideogrammi cinesi e giapponesi, tutti i caratteri dell'alfabeto cirillico, di quello arabo, di quello ebraico, ecc. ... in un codice unico a livello mondiale. Anche il problema della coesistenza di più tabelle ASCII estese differenti sarebbe stato superato. L'aumento enorme della capienza delle memorie di massa, che nel frattempo si era verificato, avrebbe reso poco rilevante la questione della maggiore occupazione di memoria.

Il progetto **UNICODE** subì negli anni svariati aggiornamenti, ed è tuttora in evoluzione.

Il suo campo d'azione include non solo le varie lingue vive (anche quelle meno diffuse), ma pure: le lingue morte, simboli di varia natura (es. matematici e chimici), l'alfabeto Braille, ideogrammi ecc.

In definitiva, **Unicode prevede oggi una codifica che va a utilizzare fino a 21 bit.**

Di tutte queste questioni si occupa l'associazione internazionale **Unicode Consortium**.