



La rappresentazione delle Informazioni

- Nella vita di tutti i giorni siamo abituati ad avere a che fare con vari tipi di informazioni, di natura e forma diversa, così come siamo abituati a diverse rappresentazioni della stessa informazione

Esempio: un testo può essere in forma stampata oppure la stessa informazione può essere registrata su un nastro

- Rappresentazioni *equivalenti* della stessa informazione e meccanismi di trasformazione tra differenti rappresentazioni
- La scelta della rappresentazione è in genere vincolata al tipo di utilizzo ed al tipo di operazioni che devono essere fatte sulle informazioni stesse



La rappresentazione delle informazioni

- I computer memorizzano ed elaborano le informazioni sotto forma di bit (Binary Digit)
- Un bit è l'unità di informazione base e può rappresentare due informazioni:
 - vero o falso
 - acceso o spento
 -
- Rappresentazione **binaria** (o **digitale**). Il linguaggio di base mediante il quale ogni informazione deve essere codificata è costituito da due soli simboli (**0** e **1**)



La rappresentazione delle informazioni

- Rappresentazione **binaria** (o **digitale**). Il linguaggio di base mediante il quale ogni informazione deve essere codificata è costituito da due soli **simboli** (**0** e **1**)
- NOTA BENE: i *simboli* 0 e 1 NON devono essere confusi con i *numeri* 0 e 1 (la tecnica per rappresentare i numeri in formato digitale verrà studiata più avanti), con le *cifre del sistema decimale* 0 e 1 (ricordo che il concetto di cifra e quello di numero sono diversi), con i *caratteri* 0 e 1 (i caratteri sono usati per la rappresentazione di testi, come vedremo più avanti)



La rappresentazione delle informazioni

Le ragioni di questa scelta sono prevalentemente di tipo tecnologico e i due simboli corrispondono a:

- due stati di polarizzazione di una sostanza magnetizzabile;
- due stati di carica elettrica di una sostanza;
- al passaggio/non passaggio di corrente attraverso un cavo conduttore;
- al passaggio/non passaggio di luce attraverso un cavo ottico



Codifica binaria

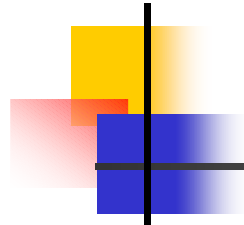
- Per poter rappresentare un numero maggiore di informazioni è necessario utilizzare sequenze di bit.

Per esempio, per rappresentare quattro informazioni diverse possiamo utilizzare due bit che ci permettono di ottenere quattro configurazioni distinte

00 01 10 11

- Il processo secondo cui si fa corrispondere ad un'informazione una configurazione di bit prende il nome di **codifica dell'informazione**

Esempio: un esame può avere quattro possibili esiti: ottimo, discreto, sufficiente, insufficiente



Codifica binaria

- Codifico

ottimo con *00*

discreto con *01*

sufficiente con *10*

insufficiente con *11*

- Con N bit si possono codificare 2^N informazioni differenti

N	Informazioni
2	4
3	8
4	16
5	32
6	64
7	128
8	256



Codifica binaria

- Se invece il mio problema è quello di dover rappresentare M informazioni differenti devo selezionare il numero di N bit in modo tale che

$$2^N \geq M$$

Esempio: per rappresentare 40 informazioni differenti devo utilizzare 6 bit perché $2^6 = 64$, 5 bit non sono sufficienti perché $2^5 = 32$

- Esiste una particolare aggregazione di bit che è costituita da 8 bit ($2^8 = 256$ informazioni) e prende il nome di **byte**



La rappresentazione delle Informazioni

- Noi vedremo le tecniche per rappresentare in formato digitale:
 - i caratteri
 - i numeri naturali
 - le immagini fisse
 - il suono
 - le immagini in movimento
- Parleremo di come si procede per ottenere una rappresentazione digitale (cioè, usando bit) mentre vedremo nella parte dedicata all'architettura di un calcolatore quali sono i dispositivi che materialmente realizzano le tecniche descritte.



La codifica dei caratteri

- L'obiettivo è quello di comunicare con il calcolatore usando il nostro linguaggio. Dobbiamo rappresentare le lettere dell'alfabeto
- L'insieme di simboli comunemente usati nell'alfabeto anglosassone, incluse le cifre numeriche, lettere maiuscole e minuscole, simboli di punteggiatura, parentesi e operatori aritmetici, può essere codificato usando 7 bit ($2^7 = 128$)
- Il metodo di codifica più diffuso tra i produttori di hardware e di software prende il nome di codice **ASCII** (**A**merican **S**tandard **C**ode for **I**nformation **I**nterchange)



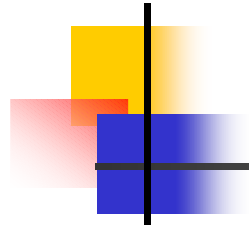
Il codice ASCII

ASCII	Simb.	ASCII	Simb.	ASCII	Simb.
0000000	NUL	0001110	SO	0011100	FS
0000001	SOH	0001111	SI	0011101	GS
0000010	STX	0010000	DLE	0011110	RS
0000011	ETX	0010001	DC1	0011111	US
0000100	EOT	0010010	DC2	0100000	SP
0000101	ENQ	0010011	DC3	0100001	!
0000110	ACK	0010011	DC4	0100010	"
0000111	BEL	0010101	NAK	0100011	#
0001000	BS	0010110	SYN	0100100	\$
0001001	HT	0010111	ETB	0100101	%
0001010	NL	0011000	CAN	0100110	&
0001011	VT	0011001	EM	0100111	'
0001100	NP	0011010	SUB	0101000	(
0001101	CR	0011011	ESC	0101001)



Il codice ASCII

ASCII	Simb.	ASCII	Simb.	ASCII	Simb.
0101010	*	0111001	9	1000111	G
0101011	+	0111010	:	1001000	H
0101100	,	0111011	;	1001001	I
0101101	-	0111100	<	1001010	J
0101110	.	0111101	=	1001011	K
0101111	/	0111110	>	1001100	L
0110000	0	0111111	?	1001101	M
0110001	1	1000000	@	1001110	N
0110010	2	1000001	A	1001111	O
0110011	3	1000010	B	1010000	P
0110100	4	1000011	C	1010001	Q
0110101	5	1000100	D	1010010	R
0110110	6	1000101	E	1010011	S
0111000	8	1000110	F	1010100	T



Il codice ASCII

ASCII	Simb.	ASCII	Simb.	ASCII	Simb.
1010101	U	1100011	c	1110001	q
1010110	V	1100100	d	1110010	r
1010111	W	1100101	e	1110011	s
1011000	X	1100110	f	1110100	t
1011001	Y	1100111	g	1110101	u
1011010	Z	1101000	h	1110110	v
1011011	[1101001	i	1110111	w
1011100	\	1101010	j	1111000	x
1011101]	1101011	k	1111001	y
1011110	^	1101100	l	1111010	z
1011111	_	1101101	m	1111011	{
1100000	`	1101110	n	1111100	
1100001	a	1101111	o	1111101	}
1100010	b	1110000	p	1111110	~
1111111	DEL				



Il codice ASCII

Sebbene 7 bit siano sufficienti per codificare l'insieme di caratteri di uso comune, in genere il codice ASCII standard utilizza 8 bit, il primo dei quali è sempre 0

Codifica della parola **cane**

01100011	01100001	01101110	01100101
c	a	n	e

Tra i simboli speciali del codice ASCII vi è anche il simbolo spazio bianco "NUL"(codice 00000000), il simbolo di fine riga "CR" (00001101)

In questo modo è possibile rappresentare mediante una sequenza di codici ASCII un testo strutturato in righe e pagine



Il codice ASCII

Consideriamo il problema inverso: data una sequenza di bit, il testo che essa codifica può essere ottenuto nel modo seguente:

- si divide la sequenza in gruppi di otto bit (un byte);
- si determina il carattere corrispondente ad ogni byte

Esempio:

01101001 01101100 00000000 01110000 01101111 00101110

01101001 01101100 00000000 01110000 01101111 00101110

i

l

p

o

.



La codifica dei caratteri

- 52 lettere alfabetiche maiuscole e minuscole
- 10 cifre (0, 1, 2, ..., 9)
- Segni di punteggiatura (, . ; : ! " ? ' ^ \ ...)
- Segni matematici (+, -, \times , \pm , {, [, >, ...)
- Caratteri nazionali (à, è, ì, ò, ù, ç, ñ, ö, ...)
- Altri segni grafici (©, ←, ↑, ⊕, @, €, ...)
- In totale 220 caratteri circa



La codifica dei caratteri

- Codifiche standard:
 - **ASCII**, 8 bit per carattere, rappresenta 256 caratteri
 - **UNICODE**, 16 bit per carattere, (ASCII + caratteri etnici)
- Codifiche proprietarie:
 - **MSWindows**, 16 bit per carattere, simile ad UNICODE