



UNIVERSITA' DEGLI STUDI DI FOGGIA

DIPARTIMENTI

DI AREA MEDICA

CdLS in Odontoiatria e Protesi Dentarie

Corso di Informatica

Prof. Crescenzo Gallo

crescenzo.gallo@unifg.it

La Codifica

Informazioni

Numeri

- *Naturali*
- *Relativi*
- *Reali*

Testi



Informazioni
tradizionali

Immagini

- *Bitmap*
- *Vettoriali*

Audio

Video



Informazioni
multimediali

Bit, Byte e Word

- L'unità atomica è il **bit** (**B**inary **D**igi**T**)
- L'insieme di 8 bit è detto **byte**
- **Word**
 - ✓ Tipicamente 16, 32 o 64 bit
 - ✓ Insieme di bit la cui dimensione è una importante caratteristica del calcolatore considerato. Essa influenza:
 - La larghezza degli indirizzi
 - La dimensione dei registri del processore
 - Larghezza dei bus (word o multipli di essa)

Il problema della codifica

- Un calcolatore può trattare **diversi tipi di dati**: numeri (interi, reali), testo, immagini, suoni, etc. che vanno comunque memorizzati in registri di memoria.
- È quindi necessario adottare una **codifica** del tipo di dato considerato: occorre, cioè, mettere in corrispondenza biunivoca i valori del tipo con gli stati che può assumere il registro.

Registro da un byte $\Rightarrow 2^8 = 256$ stati possibili. Che cosa è possibile codificare ?

Numeri naturali [0,255]

0	↔	00000000
1	↔	00000001
...		
255	↔	11111111

Numeri interi [-128,127]

-128	↔	00000000
-127	↔	00000001
0	↔	10000000
+127	↔	11111111

Numeri reali [0,1[

0.0000	↔	00000000
0.0039	↔	00000001
0.0078	↔	00000010
...		
0.9961	↔	11111111

Caratteri

A	↔	01000001
a	↔	01100001
0	↔	00110000
1	↔	00110001

La codifica implica una rappresentazione dei dati limitata e discreta

Codifica binaria

- Esiste una particolare aggregazione di bit che è costituita da **8 bit** ($2^8 = 256$ informazioni) e prende il nome di **byte (B)**
- Di solito si usano i multipli del byte

Kilo	KB	2^{10} (~ un migliaio, 1024 byte)
Mega	MB	2^{20} (~ un milione, 1KB x 1024)
Giga	GB	2^{30} (~ un miliardo, 1MB x 1024)
Tera	TB	2^{40} (~ mille miliardi, 1GB x 1024)
Peta	PB	2^{50} (~ miliardo miliardi, 1TB x 1024)

Codifica binaria

- ▶ Oltre ai dati, è necessario memorizzare anche le **istruzioni**, cioè le singole azioni elementari che l'unità centrale può eseguire.
- ▶ Nello specificare un'istruzione, bisogna precisare l'**operazione** da compiere e i **dati** coinvolti nell'operazione.

Esempio: somma 3 e 4
operazione dati

- ▶ Come rappresentare le operazioni? L'insieme delle diverse operazioni che l'unità centrale è in grado di eseguire è **finito** e quindi è possibile codificarlo con un certo numero di bit (**codice operativo**):

0000 somma
0001 sottrai
0010 moltiplica
0011 dividi
...

Codifica binaria

Una istruzione sarà quindi rappresentabile da una sequenza di bit divisa in due parti:

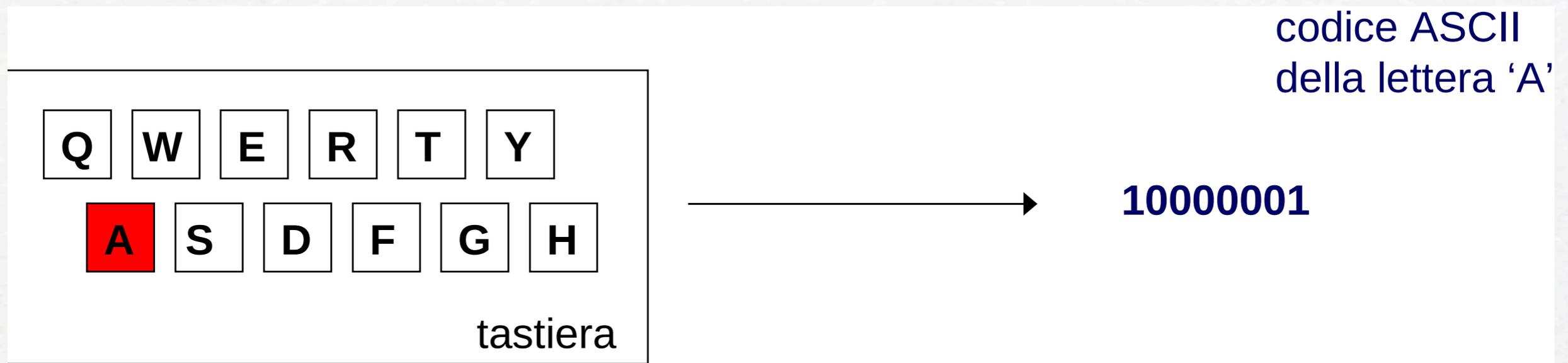
- un **codice operativo**
- un campo **operandi** (1, 2 o più operandi)



Codifica dei testi

- ➔ Si utilizza una tabella (arbitraria)
- ➔ Standard oggi (quasi) universalmente riconosciuto il codice ASCII (**A**merican **S**tandard **C**ode for **I**nformation **I**nterchange)
 - ✓ Nella versione base ogni carattere (simbolo) è codificato con 7 bit
 - 128 simboli diversi
 - ✓ Versione estesa di 8 bit
 - 256 simboli diversi

La codifica ASCII



La Codifica ASCII serve a codificare i caratteri alfanumerici.

Il Formato RTF (Rich Text Format) memorizza alcune caratteristiche aggiuntive dei caratteri.

Tabella codici ASCII base

Dec	Hx	Oct	Char	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr
0	0	000	NUL (null)	32	20	040	 	Space	64	40	100	@	@	96	60	140	`	`
1	1	001	SOH (start of heading)	33	21	041	!	!	65	41	101	A	A	97	61	141	a	a
2	2	002	STX (start of text)	34	22	042	"	"	66	42	102	B	B	98	62	142	b	b
3	3	003	ETX (end of text)	35	23	043	#	#	67	43	103	C	C	99	63	143	c	c
4	4	004	EOT (end of transmission)	36	24	044	$	\$	68	44	104	D	D	100	64	144	d	d
5	5	005	ENQ (enquiry)	37	25	045	%	%	69	45	105	E	E	101	65	145	e	e
6	6	006	ACK (acknowledge)	38	26	046	&	&	70	46	106	F	F	102	66	146	f	f
7	7	007	BEL (bell)	39	27	047	'	'	71	47	107	G	G	103	67	147	g	g
8	8	010	BS (backspace)	40	28	050	((72	48	110	H	H	104	68	150	h	h
9	9	011	TAB (horizontal tab)	41	29	051))	73	49	111	I	I	105	69	151	i	i
10	A	012	LF (NL line feed, new line)	42	2A	052	*	*	74	4A	112	J	J	106	6A	152	j	j
11	B	013	VT (vertical tab)	43	2B	053	+	+	75	4B	113	K	K	107	6B	153	k	k
12	C	014	FF (NP form feed, new page)	44	2C	054	,	,	76	4C	114	L	L	108	6C	154	l	l
13	D	015	CR (carriage return)	45	2D	055	-	-	77	4D	115	M	M	109	6D	155	m	m
14	E	016	SO (shift out)	46	2E	056	.	.	78	4E	116	N	N	110	6E	156	n	n
15	F	017	SI (shift in)	47	2F	057	/	/	79	4F	117	O	O	111	6F	157	o	o
16	10	020	DLE (data link escape)	48	30	060	0	0	80	50	120	P	P	112	70	160	p	p
17	11	021	DC1 (device control 1)	49	31	061	1	1	81	51	121	Q	Q	113	71	161	q	q
18	12	022	DC2 (device control 2)	50	32	062	2	2	82	52	122	R	R	114	72	162	r	r
19	13	023	DC3 (device control 3)	51	33	063	3	3	83	53	123	S	S	115	73	163	s	s
20	14	024	DC4 (device control 4)	52	34	064	4	4	84	54	124	T	T	116	74	164	t	t
21	15	025	NAK (negative acknowledge)	53	35	065	5	5	85	55	125	U	U	117	75	165	u	u
22	16	026	SYN (synchronous idle)	54	36	066	6	6	86	56	126	V	V	118	76	166	v	v
23	17	027	ETB (end of trans. block)	55	37	067	7	7	87	57	127	W	W	119	77	167	w	w
24	18	030	CAN (cancel)	56	38	070	8	8	88	58	130	X	X	120	78	170	x	x
25	19	031	EM (end of medium)	57	39	071	9	9	89	59	131	Y	Y	121	79	171	y	y
26	1A	032	SUB (substitute)	58	3A	072	:	:	90	5A	132	Z	Z	122	7A	172	z	z
27	1B	033	ESC (escape)	59	3B	073	;	;	91	5B	133	[[123	7B	173	{	{
28	1C	034	FS (file separator)	60	3C	074	<	<	92	5C	134	\	\	124	7C	174	|	
29	1D	035	GS (group separator)	61	3D	075	=	=	93	5D	135]]	125	7D	175	}	}
30	1E	036	RS (record separator)	62	3E	076	>	>	94	5E	136	^	^	126	7E	176	~	~
31	1F	037	US (unit separator)	63	3F	077	?	?	95	5F	137	_	_	127	7F	177		DEL

www.lookuptables.com



Codice ASCII esteso

0		32		64	@	96	`	128	Ç	160	á	192	Ł	224	Ó
1	☺	33	!	65	A	97	a	129	ü	161	í	193	ł	225	ó
2	☹	34	"	66	B	98	b	130	é	162	ó	194	Ł	226	Ô
3	♥	35	#	67	C	99	c	131	â	163	ú	195	ł	227	Ò
4	♦	36	\$	68	D	100	d	132	ä	164	ñ	196	—	228	ô
5	♣	37	%	69	E	101	e	133	à	165	Ñ	197	+	229	Õ
6	♠	38	&	70	F	102	f	134	á	166	ª	198	â	230	µ
7	·	39	'	71	G	103	g	135	ç	167	º	199	Ã	231	þ
8	▣	40	(72	H	104	h	136	ê	168	¿	200	Ł	232	ƒ
9	○	41)	73	I	105	i	137	ë	169	©	201	ł	233	Ú
10	◼	42	*	74	J	106	j	138	è	170	¬	202	ł	234	Û
11	♂	43	+	75	K	107	k	139	ï	171	½	203	ł	235	Ü
12	♀	44	,	76	L	108	l	140	î	172	¼	204	ł	236	ý
13	♪	45	-	77	M	109	m	141	ï	173	ı	205	=	237	Ý
14	♫	46	.	78	N	110	n	142	Ä	174	«	206	≠	238	—
15	☼	47	/	79	O	111	o	143	Å	175	»	207	◻	239	´
16	▶	48	0	80	P	112	p	144	É	176	◻	208	◻	240	-
17	◀	49	1	81	Q	113	q	145	æ	177	◻	209	Đ	241	±
18	↑	50	2	82	R	114	r	146	Æ	178	◻	210	Ê	242	_
19	!!	51	3	83	S	115	s	147	ø	179		211	Ë	243	¾
20	¶	52	4	84	T	116	t	148	ö	180	┆	212	È	244	¶
21	§	53	5	85	U	117	u	149	ò	181	Á	213	ı	245	§
22	—	54	6	86	V	118	v	150	û	182	Â	214	í	246	÷
23	↑	55	7	87	W	119	w	151	ù	183	À	215	î	247	,
24	↑	56	8	88	X	120	x	152	ÿ	184	©	216	ï	248	°
25	↓	57	9	89	Y	121	y	153	Ö	185	¶	217	┆	249	¨
26	→	58	:	90	Z	122	z	154	Ü	186		218	┆	250	·
27	←	59	;	91	[123	{	155	ø	187	¶	219	◼	251	¹
28	L	60	<	92	\	124		156	£	188	┆	220	◼	252	³
29	↔	61	=	93]	125	}	157	Ø	189	φ	221	ı	253	²
30	▲	62	>	94	^	126	~	158	×	190	¥	222	ı	254	■
31	▼	63	?	95	_	127	△	159	f	191	┆	223	◼	255	



Codice EBCDIC

In informatica, la sigla **EBCDIC** (dall'inglese Extended Binary Coded Decimal Interchange Code) indica un sistema di codifica dell'informazione a 8 bit usato in numerosi sistemi operativi di produzione IBM, sia per elaboratori di classe mainframe che per minicomputer.

Viene inoltre utilizzato da varie piattaforme di altri produttori.

Deriva dalla codifica a 6 bit binary-coded decimal, utilizzata nelle schede perforate e nella maggior parte delle periferiche IBM della fine degli anni 1950 e dell'inizio degli anni 1960.

	-0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-A	-B	-C	-D	-E	-F
0-	NUL 00	SOH 01	STX 02	ETX 03	SEL 04	HT 09	RNL 0A	DEL 0B	GE 0C	SPS 0D	RPT 0E	VT 0F	FF 10	CR 11	SO 12	SI 13
1-	DLE 14	DC1 15	DC2 16	DC3 17	RES 18	NL 19	BS 1A	POC 1B	CAN 1C	EM 1D	UBS 1E	CU1 1F	IFS 20	IGS 21	IRS 22	IUS 23
2-	DS 24	SOS 25	FS 26	WUS 27	BYP 28	LF 29	ETB 2A	ESC 2B	SA 2C	SFE 2D	SM 2E	SW 2F	CSP 30	MFA 31	ENQ 32	ACK 33
3-			SYN 34	IR 35	PP 36	TRN 37	NBS 38	EOT 39	SBS 3A	IT 3B	RFF 3C	CU3 3D	DC4 3E	NAK 3F		SUB 40
4-	SP 41	RSP 42	à 43	ã 44	ä 45	á 46	ä 47	ç 48	ñ 49	[4A	. 4B	< 4C	(4D	+ 4E	! 4F	
5-	& 50	é 51	ê 52	ë 53	è 54	ì 55	î 56	ï 57	í 58	β 59] 5A	\$ 5B	* 5C) 5D	; 5E	^ 5F
6-	~ 60	/ 61	Á 62	Ã 63	À 64	Ä 65	Ã 66	Ç 67	Ñ 68	! 69	, 6A	% 6B	- 6C	> 6D	? 6E	
7-	ø 6F	É 70	Ê 71	Ë 72	È 73	Ì 74	Î 75	Ï 76	Ì 77	ˆ 78	: 79	# 7A	@ 7B	' 7C	= 7D	" 7E
8-	Ø 7F	a 80	b 81	c 82	d 83	e 84	f 85	g 86	h 87	i 88	« 89	» 8A	ò 8B	ý 8C	þ 8D	± 8E
9-	° 8F	j 90	k 91	l 92	m 93	n 94	o 95	p 96	q 97	r 98	ª 99	º 9A	æ 9B	, 9C	Æ 9D	¤ 9E
A-	µ 9F	~ A0	s A1	t A2	u A3	v A4	w A5	x A6	y A7	z A8	 A9	¿ AA	Ð AB	Ý AC	Þ AD	@ AE
B-	€ AF	£ B0	¥ B1	· B2	© B3	§ B4	¶ B5	¼ B6	½ B7	¾ B8	¬ B9	 BA	ˆ BB	ˆ BC	ˆ BD	x BE
C-	{ BF	A C0	B C1	C C2	D C3	E C4	F C5	G C6	H C7	I C8	SHY C9	ó CA	ô CB	ò CC	ó CD	õ CE
D-	} CF	J D0	K D1	L D2	M D3	N D4	O D5	P D6	Q D7	R D8	¹ D9	ú DA	û DB	ù DC	ú DD	ÿ DE
E-	\ DF	÷ E0	S E1	T E2	U E3	V E4	W E5	X E6	Y E7	Z E8	² E9	Ó EA	Ö EB	Ò EC	Ó ED	Õ EE
F-	0 EF	1 F0	2 F1	3 F2	4 F3	5 F4	6 F5	7 F6	8 F7	9 F8	³ F9	Ù FA	Ú FB	Û FC	Ü FD	EO FE



UNICODE

Un altro codice molto diffuso oggi è lo UNICODE, che impiega 16 bit per carattere (Extended ASCII + caratteri etnici):

$2^{16} = 65.536$ simboli diversi

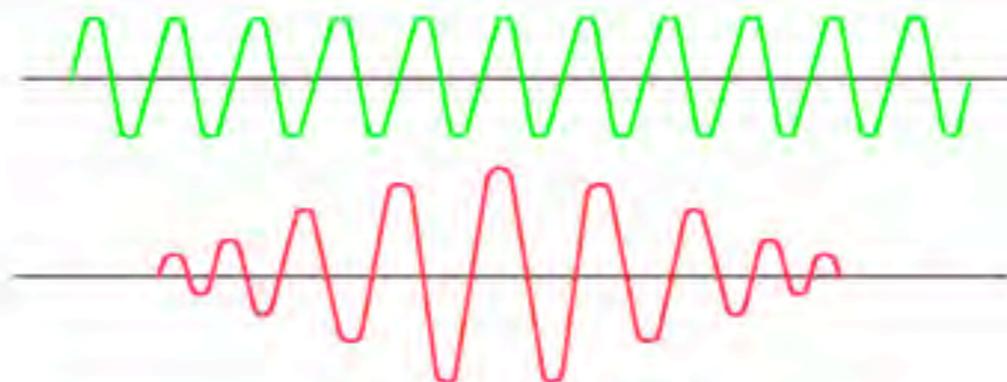
Derivazioni: UTF-8,16,32 per il web

Da Wikipedia

Unicode era stato originariamente pensato come una codifica a 16 bit (quattro cifre esadecimali) che dava la possibilità di codificare 65.536 caratteri. Tanto si riteneva essere sufficiente per rappresentare i caratteri impiegati in tutte le lingue scritte del mondo.

Ora invece lo standard Unicode, che tendenzialmente è perfettamente allineato con la norma [ISO/IEC 10646](#), prevede una codifica fino 21 bit e supporta un repertorio di codici numerici che possono rappresentare circa un milione di caratteri. Ciò appare sufficiente a coprire anche i fabbisogni di codifica di scritti del patrimonio storico dell'umanità, nelle diverse lingue e negli svariati sistemi di segni utilizzati.

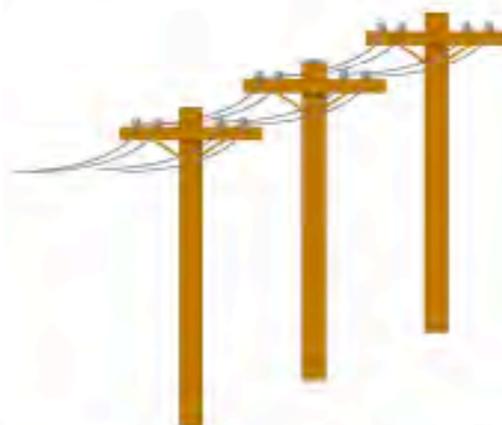
Trasmissione delle informazioni



SEGNALI



EMITTENTE



**mezzo trasmissivo
(canale)**



RICEVENTE

Errori di trasmissione



DISTURBI: occasionali alterazioni dei messaggi → **ERRORI** di trasmissione

TASSI DI ERRORE: su linee telefoniche
~ uno ogni cento milioni di bit trasmessi

CONTROLLO DEGLI ERRORI DI TRASMISSIONE

- controllo di parità verticale (**VRC**)
- controllo di ridondanza longitudinale (**LRC**)
- controllo polinomiale



Errori di trasmissione

blocco

	0	0	1	1	0	1	1	0	0	0
	1	0	1	0	1	0	1	1	0	1
	0	1	0	1	1	1	0	0	1	1
	0	0	0	1	1	1	0	0	0	1
caratteri →	1	0	1	0	0	1	1	0	0	0
	0	1	0	1	0	1	0	1	1	1
	1	1	1	0	1	1	0	0	1	0
	1	0	1	1	0	1	1	1	0	0
	0	1	1	1	0	1	0	1	1	0

VRC ↑

← carattere di controllo LRC

Errori di trasmissione

⇒ ritrasmissione

- *controllo semplice*
- *costo di trasmissione maggiore*

⇒ **correzione** (ricostruzione del messaggio a partire dai bit ricevuti e dagli errori riscontrati):

- *più complesso e costoso*
- *applicabile anche per trasmissioni monodirezionali*
- *utile se l'indice di affidabilità è basso (ricorrere sempre alla ritrasmissione del messaggio può voler dire di fatto rallentare notevolmente la velocità del canale)*

Errori di trasmissione

