

l'informazione non numerica

caratteri, simboli ...

Cristiana Bolchini

3Ma09!2£ gi0c4r3
13294021
acquario h

informazione da rappresentare

informazione
dati e istruzioni

numerica

non numerica

24 6 11
naturali

-24 $+1$ $+315$
interi
relativi

$+2.409$
 -14.25
razionali

casa
n0-0ne!
testi


suoni


immagini

codifica dei caratteri

testi

- obiettivo: definire una codifica per ogni possibile carattere rappresentabile
 - inizialmente:
 - alfabeto base a b ... z A B ... Z
 - caratteri numerici 0 1 ... 9
 - simboli interpunzione . , : ...
 - caratteri "speciali" a-capo nuova-riga spazio
- } ~120 elementi
- lunghezza codifica: $\lceil \log_{|S|} n \rceil = \lceil \log_2 120 \rceil = 7$

codifica dei caratteri

testi

- **codice ASCII** (American Standard Code for Information Interchange) **base**: 7 bit
 - **codice ASCII esteso**: 8 bit
 - codice ASCII base per le prime 128 configurazioni
 - caratteri nazionali à è é
 - simboli
 - cornici
- } caratteri
semigrafici

codice ASCII base

0	Null character	00000000
1	Start of Header	00000001
2	Start of Text	00000010
3	End of Text	00000011
4	End of Transmission	00000100
5	Enquiry	00000101
6	Acknowledge	00000110
7	Bell	00000111
8	Backspace	00001000
9	Horizontal Tab	00001001
10	Line Feed	00001010
11	Vertical Feed	00001011
12	Form Feed	00001100
13	Carriage Return	00001101
14	Shift Out	00001110
15	Shift In	00001111
16	Data Link Escape	00010000
17	Device Control 1	00010001
18	Device Control 2	00010010
19	Device Control 3	00010011
20	Device Control 4	00010100
21	Negative Ack	00010101
22	Synchronous Idle	00010110
23	End of trans. block	00010111
24	Cancel	00011000
25	End of medium	00011001
26	Substitute	00011010
27	Escape	00011011
28	File Separator	00011100
29	Group Separator	00011101
30	Record Separator	00011110
31	Delete	00011111

caratteri di controllo

32	(spazio)	00100000
33	!	00100001
34	"	00100010
35	#	00100011
36	\$	00100100
37	%	00100101
38	&	00100110
39	'	00100111
40	(00101000
41)	00101001
42	*	00101010
43	+	00101011
44	,	00101100
45	-	00101101
46	.	00101110
47	/	00101111
48	0	00110000
49	1	00110001
50	2	00110010
51	3	00110011
52	4	00110100
53	5	00110101
54	6	00110110
55	7	00110111
56	8	00111000
57	9	00111001
58	:	00111010
59	;	00111011
60	<	00111100
61	=	00111101
62	>	00111110
63	?	00111111

simboli di interpunzione e simboli

64	@	01000000
65	A	01000001
66	B	01000010
67	C	01000011
68	D	01000100
69	E	01000101
70	F	01000110
71	G	01000111
72	H	01001000
73	I	01001001
74	J	01001010
75	K	01001011
76	L	01001100
77	M	01001101
78	N	01001110
79	O	01001111
80	P	01010000
81	Q	01010001
82	R	01010010
83	S	01010011
84	T	01010100
85	U	01010101
86	V	01010110
87	W	01010111
88	X	01011000
89	Y	01011001
90	Z	01011010
91	[01011011
92	\	01011100
93]	01011101
94	^	01011110
95	_	01011111

96	`	01100000
97	a	01100001
98	b	01100010
99	c	01100011
100	d	01100100
101	e	01100101
102	f	01100110
103	g	01100111
104	h	01101000
105	i	01101001
106	j	01101010
107	k	01101011
108	l	01101100
109	m	01101101
110	n	01101110
111	o	01101111
112	p	01110000
113	q	01110001
114	r	01110010
115	s	01110011
116	t	01110100
117	u	01110101
118	v	01110110
119	w	01110111
120	x	01111000
121	y	01111001
122	z	01111010
123	{	01111011
124		01111100
125	}	01111101
126	~	01111110
127		01111111

codice ASCII esteso

ISO-8859-1 - Latin-1
(IBM equivalente: 819)

128	€	10000000
129		
130	,	10000010
131	f	10000011
132	„	10000100
133	...	10000101
134	†	10000110
135	‡	10000111
136	^	10001000
137	‰	10001001
138	Š	10001010
139	Š	10001011
140	ƒ	10001100
141		10001101
142	Ž	10001110
143		10001111
144		10010000
145	‘	10010001
146	’	10010010
147	“	10010011
148	”	10010100
149	•	10010101
150	–	10010110
151	—	10010111
152	ˆ	10011000
153	™	10011001
154	š	10011010
155	›	10011011
156	œ	10011100
157		10011101
158	ž	10011110
159	ÿ	10011111

160	ı	10100000
161	¢	10100001
162	£	10100010
163	¤	10100011
164	¥	10100100
165	¦	10100101
166	§	10100110
167	¨	10100111
168	©	10101000
169	ª	10101001
170	«	10101010
171	¬	10101011
172	SHY	10101100
173	®	10101101
174	¯	10101110
175	°	10101111
176	±	10110000
177	²	10110001
178	³	10110010
179	´	10110011
180	µ	10110100
181	¶	10110101
182	·	10110110
183	,	10110111
184	¹	10111000
185	º	10111001
186	»	10111010
187	¼	10111011
188	½	10111100
189	¾	10111101
190	¿	10111110
191	À	10111111

192	Á	11000000
193	Â	11000001
194	Ã	11000010
195	Ä	11000011
196	Å	11000100
197	Æ	11000101
198	Ç	11000110
199	È	11000111
200	É	11001000
201	Ê	11001001
202	Ë	11001010
203	Ì	11001011
204	Í	11001100
205	Î	11001101
206	Ï	11001110
207	Ð	11001111
208	Ñ	11010000
209	Ò	11010001
210	Ó	11010010
211	Ô	11010011
212	Õ	11010100
213	Ö	11010101
214	×	11010110
215	Ø	11010111
216	Ù	11011000
217	Ú	11011001
218	Û	11011010
219	Ü	11011011
220	Ý	11011100
221	Þ	11011101
222	ß	11011110
223	à	11011111

224	á	11100000
225	â	11100001
226	ã	11100010
227	ä	11100011
228	å	11100100
229	æ	11100101
230	ç	11100110
231	è	11100111
232	é	11101000
233	ê	11101001
234	ë	11101010
235	ì	11101011
236	í	11101100
237	î	11101101
238	ï	11101110
239	ð	11101111
240	ñ	11110000
241	ò	11110001
242	ó	11110010
243	ô	11110011
244	õ	11110100
245	ö	11110101
246	÷	11110110
247	ø	11110111
248	ù	11111000
249	ú	11111001
250	û	11111010
251	ü	11111011
252	ý	11111100
253	þ	11111101
254	ÿ	11111110
255		11111111

codice ASCII esteso

ISO/IEC 8859-7
Latin/Greek

128		10000000	160		10100000
129		10000001	161	´	10100001
130		10000010	162	´	10100010
131		10000011	163	£	10100011
132		10000100	164	€	10100100
133		10000101	165	□	10100101
134		10000110	166	¡	10100110
135		10000111	167	§	10100111
136		10001000	168	™	10101000
137		10001001	169	©	10101001
138		10001010	170	□	10101010
139		10001011	171	«	10101011
140		10001100	172	¬	10101100
141		10001101	173	SHY	10101101
142		10001110	174		10101110
143		10001111	175	—	10101111
144		10010000	176	°	10110000
145		10010001	177	±	10110001
146		10010010	178	²	10110010
147		10010011	179	³	10110011
148		10010100	180	´	10110100
149		10010101	181	^	10110101
150		10010110	182	À	10110110
151		10010111	183	·	10110111
152		10011000	184	Ê	10111000
153		10011001	185	Ë	10111001
154		10011010	186	Ì	10111010
155		10011011	187	»	10111011
156		10011100	188	Ï	10111100
157		10011101	189	½	10111101
158		10011110	190	Υ	10111110
159		10011111	191	Ω	10111111

192	î	11000000	193	A	11000001
194	B	11000010	195	Γ	11000011
196	Δ	11000100	197	E	11000101
198	Z	11000110	199	H	11000111
200	Θ	11001000	201	I	11001001
202	K	11001010	203	Λ	11001011
204	M	11001100	205	N	11001101
206	Ξ	11001110	207	O	11001111
208	Π	11010000	209	P	11010001
210		11010010	211	Σ	11010011
212	T	11010100	213	Υ	11010101
214	Φ	11010110	215	X	11010111
216	Ψ	11011000	217	Ω	11011001
218	Ϊ	11011010	219	Ÿ	11011011
220	ά	11011100	221	έ	11011101
222	ή	11011110	223	ί	11011111

224	ú	11100000	225	α	11100001
226	β	11100010	227	γ	11100011
228	δ	11100100	229	ε	11100101
230	ζ	11100110	231	η	11100111
232	θ	11101000	233	ι	11101001
234	κ	11101010	235	λ	11101011
236	μ	11101100	237	ν	11101101
238	ξ	11101110	239	ο	11101111
240	π	11110000	241	ρ	11110001
242	ς	11110010	243	σ	11110011
244	τ	11110100	245	υ	11110101
246	φ	11110110	247	χ	11110111
248	ψ	11111000	249	ω	11111001
250	ϊ	11111010	251	ϋ	11111011
252	ό	11111100	253	ύ	11111101
254	ώ	11111110	255		11111111

128		10000000	129		10000001
130		10000010	131		10000011
132		10000100	133		10000101
134		10000110	135		10000111
136		10001000	137		10001001
138		10001010	139		10001011
140		10001100	141		10001101
142		10001110	143		10001111
144		10010000	145		10010001
146		10010010	147		10010011
148		10010100	149		10010101
150		10010110	151		10010111
152		10011000	153		10011001
154		10011010	155		10011011
156		10011100	157		10011101
158		10011110	159		10011111

160		10100000	161	À	10100001
162	˘	10100010	163	Ł	10100011
164	▯	10100100	165	Ł	10100101
166	Ś	10100110	167	Ş	10100111
168	™	10101000	169	Š	10101001
170	Ş	10101010	171	Ť	10101011
172	Ž	10101100	173	SHY	10101101
174	Ž	10101110	175	Ž	10101111
176	°	10110000	177	ą	10110001
178	˙	10110010	179	ł	10110011
180	´	10110100	181	ŕ	10110101
182	ś	10110110	183	˘	10110111
184	˙	10111000	185	š	10111001
186	ş	10111010	187	š	10111011
188	ž	10111100	189	™	10111101
190	ž	10111110	191	ž	10111111

ISO/IEC 8859-4
Latin-4 or North European

192	Ř	11000000	193	Á	11000001
194	Â	11000010	195	Ă	11000011
196	Ä	11000100	197	Ĺ	11000101
198	Č	11000110	199	Ç	11000111
200	Č	11001000	201	É	11001001
202	Ę	11001010	203	Ě	11001011
204	Ě	11001100	205	Í	11001101
206	Î	11001110	207	Ď	11001111
208	Ð	11010000	209	Ň	11010001
210	Ñ	11010010	211	Ó	11010011
212	Ô	11010100	213	Õ	11010101
214	Ö	11010110	215	×	11010111
216	Ř	11011000	217	Ů	11011001
218	Ú	11011010	219	Û	11011011
220	Ü	11011100	221	Ý	11011101
222	Ť	11011110	223	ß	11011111

224	ř	11100000	225	á	11100001
226	â	11100010	227	ă	11100011
228	ä	11100100	229	ĺ	11100101
230	č	11100110	231	ç	11100111
232	č	11101000	233	é	11101001
234	ę	11101010	235	ě	11101011
236	ě	11101100	237	í	11101101
238	î	11101110	239	ď	11101111
240	đ	11110000	241	ň	11110001
242	ñ	11110010	243	ó	11110011
244	ô	11110100	245	õ	11110101
246	ö	11110110	247	×	11110111
248	ř	11111000	249	ů	11111001
250	ú	11111010	251	ű	11111011
252	ü	11111100	253	ý	11111101
254	ť	11111110	255	·	11111111

codice ASCII

un carattere

- ad ogni singolo carattere corrisponde un byte

A ► 01000001

- quindi possiamo pensare ad un carattere ANCHE come ad un valore numerico ► naturale

A ► 01000001 ► 65

codice ASCII

- caratteri adiacenti nell'alfabeto, sono adiacenti nel codice ASCII
 - caratteri maiuscoli e minuscoli non sono la stessa cosa

...

3	▶	00110011	▶	51
4	▶	00110100	▶	52

...

A	▶	01000001	▶	65
B	▶	01000010	▶	66

...

g	▶	01100111	▶	103
h	▶	01101000	▶	104

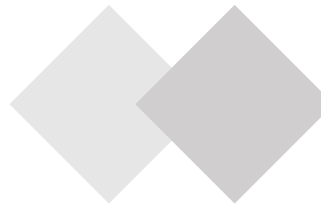
...

codice ASCII

- non è necessario imparare il codice ASCII a memoria
 - ricordare:
 - caratteri numerici prima dei caratteri alfabetici
 - maiuscole prima delle minuscole
 - considerazioni:
 - la distanza tra due caratteri numerici è pari alla differenza tra il valore dei corrispondenti codici ASCII
 - $4 - 3 = 1$ ▶ $52 - 51 = 1$
 - stesso ordinamento alfabetico
 - F segue C ▶ codice di F (70) > codice di C (67)
- ▶ in entrambi i casi NON SERVE conoscere il codice del carattere

approfondimento

la tastiera del calcolatore

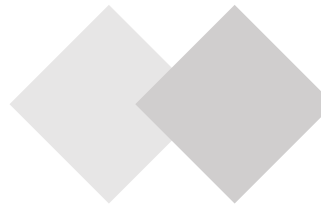


- quando premiamo un tasto sulla tastiera inviamo al calcolatore una sequenza di bit che corrispondono al codice ASCII del tasto premuto

A ► 01000001

- quando si installa un calcolatore e chiede la lingua chiede anche quali tastiere installare
 - italiana: il codice ASCII esteso è quello LATIN-1
- quando si cambia tastiera, si cambia la mappa tasto-sequenza binaria che viene inviata

codice UNICODE



- Standard internazionale che permette di rappresentare tutti i carattere attraverso un codice a 16 bit che raggruppa tutti gli alfabeti esistenti
- Compatibile con il codice ASCII
- <https://symbl.cc/it/unicode/blocks/>
- Rispetto al codice ASCII, le dimensioni dei file risultano raddoppiate