

Informatica e sistemi di elaborazione delle informazioni

La codifica delle informazioni

Informatica e sistemi di elaborazione delle informazioni | slide 1

La codifica delle informazioni

```
graph LR; A[Codifica delle informazioni] --> B[Codifica del testo]; A --> C[Codifica delle immagini]; A --> D[Codifica del video]; A --> E[Codifica dell'audio];
```

Informatica e sistemi di elaborazione delle informazioni | slide 2

La codifica del testo

Informatica e sistemi di elaborazione delle informazioni | slide 3

Domanda

Informatica e sistemi di elaborazione delle informazioni | slide 4

Codifica dei caratteri

- Alfabeto anglosassone
 - Lettere maiuscole e minuscole
 - Cifre numeriche (0, 1, 2, ..., 9)
 - Simboli di punteggiatura (, . ; : ! " ? ...)
 - Segni matematici (+, -, {, [, >, ...)
 - Caratteri di controllo
- In totale abbiamo circa 120 caratteri, che possono essere codificati utilizzando 7 bit ($2^7 = 128$)

Informatica e sistemi di elaborazione delle informazioni | slide 5

Codifica ASCII standard

- Il metodo di codifica più diffuso tra i produttori di hardware e di software prende il nome di **codifica ASCII (American Standard Code for Information Interchange)**.
- Nel caso si utilizzino 7 bit, si parla di **codifica ASCII standard**

Informatica e sistemi di elaborazione delle informazioni | slide 6

Tabella dei codici ASCII standard

Byte	Code	Char	Byte	Code	Char	Byte	Code	Char	Byte	Code	Char
00000000	0	Null	00110000	32	SpC	01000000	64	@	01100000	96	a
00000001	1	Start of heading	00110001	33	!	01000001	65	A	01100001	97	b
00000010	2	Start of text	00110010	34	"	01000010	66	B	01100010	98	c
00000011	3	End of text	00110011	35	#	01000011	67	C	01100011	99	d
00000100	4	End of transmit	00110100	36	\$	01000100	68	D	01100100	100	e
00000101	5	Enquiry	00110101	37	%	01000101	69	E	01100101	101	f
00000110	6	Acknowledge	00110110	38	&	01000110	70	F	01100110	102	g
00000111	7	Audible bell	00110111	39	'	01000111	71	G	01100111	103	h
00001000	8	Backspace	00110100	40	(01001000	72	H	01101000	104	i
00001001	9	Horizontal tab	00110101	41)	01001001	73	I	01101001	105	j
00001010	10	Line feed	00110110	42	*	01001010	74	J	01101010	106	k
00001011	11	Vertical tab	00110111	43	+	01001011	75	K	01101011	107	l
00001100	12	Form Feed	00111000	44	,	01001100	76	L	01101100	108	m
00001101	13	Carriage return	00111001	45	-	01001101	77	M	01101101	109	n
00001110	14	Shift out	00111010	46	.	01001110	78	N	01101110	110	o
00001111	15	Shift in	00111011	47	/	01001111	79	O	01101111	111	p
00010000	16	Data link escape	00111100	48	0	01010000	80	P	01110000	112	q
00010001	17	Device control 1	00111001	49	1	01010001	81	Q	01110001	113	r
00010010	18	Device control 2	00111010	50	2	01010010	82	R	01110010	114	s
00010011	19	Device control 3	00111011	51	3	01010011	83	S	01110011	115	t
00010100	20	Device control 4	00111100	52	4	01010100	84	T	01110100	116	u
00010101	21	Neg. acknowledge	00111010	53	5	01010101	85	U	01110101	117	v
00010110	22	Synchronous idle	00110110	54	6	01010110	86	V	01110110	118	w
00010111	23	End trans. block	00110111	55	7	01010111	87	W	01110111	119	x
00011000	24	Cancel	00111000	56	8	01011000	88	X	01111000	120	y
00011001	25	End of medium	00111001	57	9	01011001	89	Y	01111001	121	z
00011010	26	Substitution	00111010	58	:	01011010	90	Z	01111010	122	{
00011011	27	Escape	00111011	59	;	01011011	91	[01111011	123	
00011100	28	File separator	00111100	60	<	01011100	92	\	01111100	124	}
00011101	29	Group separator	00111101	61	=	01011101	93]	01111101	125	~
00011110	30	Record Separator	00111110	62	>	01011110	94	^	01111110	126	
00011111	31	Unit separator	00111111	63	?	01011111	95	_	01111111	127	Del

Codifica ASCII estesa

- Per rappresentare, oltre ai caratteri precedentemente visti, anche i caratteri nazionali (à, è, ì, ò, ù, ç, ñ, ö, ...), occorrono circa 220 caratteri, che possono essere codificati utilizzando 8 bit ($2^8 = 256$)
- In questo caso si parla di **codifica ASCII estesa**

Tabella dei codici ASCII estesa

Byte	Code	Char									
10000000	128	À	10100000	160	à	11000000	192	+	11100000	224	ó
10000001	129	Á	10100001	161	á	11000001	193	+	11100001	225	õ
10000010	130	Â	10100010	162	â	11000010	194	+	11100010	226	ö
10000011	131	Ã	10100011	163	ã	11000011	195	+	11100011	227	ø
10000100	132	Ä	10100100	164	ä	11000100	196	+	11100100	228	õ
10000101	133	Å	10100101	165	å	11000101	197	+	11100101	229	ö
10000110	134	Æ	10100110	166	æ	11000110	198	+	11100110	230	ß
10000111	135	Ç	10100111	167	ç	11000111	199	+	11100111	231	þ
10001000	136	È	10101000	168	è	11001000	200	+	11101000	232	ÿ
10001001	137	É	10101001	169	é	11001001	201	+	11101001	233	ÿ
10001010	138	Ê	10101010	170	ê	11001010	202	+	11101010	234	ÿ
10001011	139	Ë	10101011	171	ë	11001011	203	+	11101011	235	ÿ
10001100	140	Ì	10101100	172	ì	11001100	204	+	11101100	236	ÿ
10001101	141	Í	10101101	173	í	11001101	205	+	11101101	237	ÿ
10001110	142	Î	10101110	174	î	11001110	206	+	11101110	238	ÿ
10001111	143	Ï	10101111	175	ï	11001111	207	+	11101111	239	ÿ
10010000	144	Ð	10110000	176	ð	11001000	208	+	11110000	240	ÿ
10010001	145	Ñ	10110001	177	ñ	11010001	209	+	11110001	241	ÿ
10010010	146	Ò	10110010	178	ò	11010010	210	+	11110010	242	ÿ
10010011	147	Ó	10110011	179	ó	11010011	211	+	11110011	243	ÿ
10010100	148	Ô	10110100	180	ô	11010100	212	+	11110100	244	ÿ
10010101	149	Õ	10110101	181	õ	11010101	213	+	11110101	245	ÿ
10010110	150	Ö	10110110	182	ö	11010110	214	+	11110110	246	ÿ
10010111	151	Ï	10110111	183	ï	11010111	215	+	11110111	247	ÿ
10011000	152	Ù	10111000	184	ù	11011000	216	+	11111000	248	ÿ
10011001	153	Ú	10111001	185	ú	11011001	217	+	11111001	249	ÿ
10011010	154	Û	10111010	186	û	11011010	218	+	11111010	250	ÿ
10011011	155	Ü	10111011	187	ü	11011011	219	+	11111011	251	ÿ
10011100	156	Ý	10111100	188	ý	11011100	220	+	11111100	252	ÿ
10011101	157	Þ	10111101	189	þ	11011101	221	+	11111101	253	ÿ
10011110	158	ß	10111110	190	ÿ	11011110	222	+	11111110	254	ÿ
10011111	159	ä	10111111	191	ä	11011111	223	+	11111111	255	ÿ

Codifica delle parole

- Le parole sono sequenze di caratteri
- Esempio: Come è codificata la parola "cane"?
c a n e
01100011 01100001 01101110 01100101
- 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 (byte)
 - si determina il carattere corrispondente ad ogni byte

Codifica dei caratteri

- Abbiamo considerato i codici:
 - Codice ASCII standard: 7 bit per carattere
 - Codice ASCII esteso: 8 bit per carattere
- Esiste poi un altro codice molto utilizzato:
 - Codice UNICODE, 16 bit per carattere (ASCII + caratteri etnici)
 - Microsoft Windows usa un codice proprietario a 16 bit per carattere, simile ad UNICODE

Codifiche a confronto

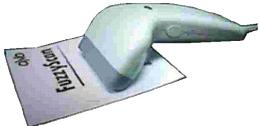
- Il carattere "A" è rappresentato:
 - nella codifica ASCII standard come:
1000001
 - nella codifica ASCII estesa come:
01000001
 - nella codifica UNICODE come:
000000001000001

Domande

- Quanti bit si devono utilizzare per rappresentare 300 informazioni distinte?
- Quanti byte occorrono per codificare la parola "economia" se si utilizza il codice ASCII esteso?
- Quanti byte occorrono per codificare la parola "economia" se si utilizza il codice UNICODE?
- La parola "economia" ha la stessa codifica sia se si utilizza il codice ASCII che il codice UNICODE?
- Dati 12 bit per la codifica, quante informazioni distinte si possono rappresentare?

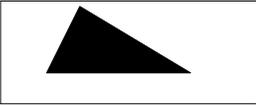
Informatica e sistemi di elaborazione delle informazioni I slide 13

La codifica delle immagini

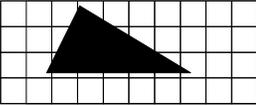



Informatica e sistemi di elaborazione delle informazioni I slide 14

Codifica delle immagini



Suddividiamo l'immagine mediante una griglia formata da righe orizzontali e verticali a distanza costante



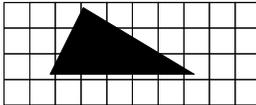
Informatica e sistemi di elaborazione delle informazioni I slide 15

Codifica delle immagini

- Ogni quadratino derivante da tale suddivisione prende il nome di **pixel** (picture element) e può essere codificato in binario secondo la seguente convenzione:
 - Il simbolo "0" viene utilizzato per la codifica di un pixel corrispondente ad un quadratino in cui il bianco è predominante
 - Il simbolo "1" viene utilizzato per la codifica di un pixel corrispondente ad un quadratino in cui il nero è predominante

Informatica e sistemi di elaborazione delle informazioni I slide 16

Codifica delle immagini



0	0	0	1	0	0	0	0	0	0
0	0	1	1	1	0	0	0	0	0
0	0	1	1	1	1	1	0	0	0
0	0	0	0	0	0	0	0	0	0

Informatica e sistemi di elaborazione delle informazioni I slide 17

Codifica delle immagini

0	0	0	1	0	0	0	0	0	0
0	0	1	1	1	0	0	0	0	0
0	0	1	1	1	1	1	0	0	0
0	0	0	0	0	0	0	0	0	0

Poiché una sequenza di bit è lineare, è necessario definire convenzioni per ordinare la griglia dei pixel in una sequenza. Assumiamo che i pixel siano ordinati dal basso verso l'alto e da sinistra verso destra

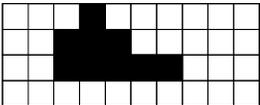
000000000 0011111000 0011100000 0001000000

Informatica e sistemi di elaborazione delle informazioni I slide 18

Codifica delle immagini

Non sempre il contorno della figura coincide con le linee della griglia. Quella che si ottiene nella codifica è un'approssimazione della figura originaria

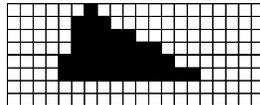
Se riconvertiamo la sequenza di stringhe
0000000000 0011111000 0011100000 0001000000
in immagine otteniamo



Informatica e sistemi di elaborazione delle informazioni I slide 19

Codifica delle immagini

La rappresentazione sarà più fedele all'aumentare del numero di pixel, ossia al diminuire delle dimensioni dei quadratini della griglia in cui è suddivisa l'immagine



Informatica e sistemi di elaborazione delle informazioni I slide 20

Codifica delle immagini

- Assegnando un bit ad ogni pixel è possibile codificare solo **immagini in bianco e nero**
- Per codificare le immagini con diversi **livelli di grigio** oppure **a colori** si usa la stessa tecnica: per ogni pixel viene assegnata una sequenza di bit

Informatica e sistemi di elaborazione delle informazioni I slide 21

Codifica delle immagini (grigio e colore)

- Per memorizzare un pixel non è più sufficiente un solo bit
 - Per esempio, se utilizziamo quattro bit possiamo rappresentare $2^4=16$ livelli di grigio o 16 colori diversi
 - Mentre con otto bit ne possiamo distinguere $2^8 = 256$, ecc.

Informatica e sistemi di elaborazione delle informazioni I slide 22

L'uso del colore

- Il colore può essere generato componendo 3 colori (sintesi additiva): red, green, blue (RGB)
- Ad ogni colore si associa una possibile sfumatura
- Usando **2 bit** per ogni colore si possono ottenere 4 sfumature per il rosso, 4 per il blue e 4 per il verde che, combinate insieme, danno origine a 64 colori diversi
- Ogni pixel per essere memorizzato richiede **6 bit**

Informatica e sistemi di elaborazione delle informazioni I slide 23

L'uso del colore

- Usando **8 bit** per ogni colore si possono ottenere 256 sfumature per il rosso, 256 per il blu e 256 per il verde che, combinate insieme, danno origine a circa 16,8 milioni di colori diversi (precisamente 16777216 colori)
- Ogni pixel per essere memorizzato richiede **3 byte**

Informatica e sistemi di elaborazione delle informazioni I slide 24

Risoluzione

- Il numero di pixel presenti sullo schermo (colonne x righe) prende il nome di **risoluzione**
- Risoluzione tipiche sono: 640 x 480 1024 x 768 1280 x 1024
- Esempio: Per distinguere 256 colori sono necessari 8 bit per la codifica di ciascun pixel
 - La codifica di un'immagine formata da 640 x 480 pixel richiederà $640 \times 480 \times 8 = 2.457.600 \text{ bit} = 307.200 \text{ B} = 300 \text{ KB}$



Informatica e sistemi di elaborazione delle informazioni I slide 25

Grafica bitmap

- Le immagini codificate pixel per pixel sono dette immagini in grafica **bitmap**
 - Le immagini bitmap occupano parecchio spazio
- Esistono delle tecniche di compressione che permettono di ridurre le dimensioni
 - Ad esempio, se più punti vicini di un'immagine assumono lo stesso colore, si può memorizzare la codifica del colore una sola volta e poi ricordare per quante volte deve essere ripetuta
- I formati come GIF, JPEG e PNG sono formati compressi
- Argomento correlato: formati come Postscript e PDF per i documenti

Informatica e sistemi di elaborazione delle informazioni I slide 26

Domande

- Quanti byte occupa un'immagine di 100 x 100 pixel in bianco e nero?
- Quanti byte occupa un'immagine di 100 x 100 pixel a 256 colori?
- Se un'immagine a $16777216 (= 2^{24})$ di colori occupa 2400 byte, da quanti pixel sarà composta?

Informatica e sistemi di elaborazione delle informazioni I slide 27

La codifica del video (immagini in movimento)

Informatica e sistemi di elaborazione delle informazioni I slide 28

Codifica di immagini in movimento

- Un filmato è una sequenza di immagini statiche (dette fotogrammi o frame)
- Per codificare un filmato si "digitalizzano" i suoi fotogrammi
- Esempio:
 - 30 immagini con una risoluzione di 640 x 480 al secondo (30 fps)
 - 1 secondo: $30 \times 300 \text{ KB} = 9000 \text{ KB/s}$
 - 1 minuto: $60 \times 9000 \text{ KB} = 540000 \text{ KB} = 527,34 \text{ MB}$
- Esempi di formati per il video: AVI, MOV
- Compressione: MPEG (Moving Picture Expert Group), differenza tra fotogrammi

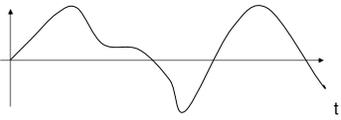
Informatica e sistemi di elaborazione delle informazioni I slide 29

La codifica dei suoni

Informatica e sistemi di elaborazione delle informazioni I slide 30

Codifica dei suoni

- Fisicamente un suono è rappresentabile da un'onda che descrive la variazione della pressione dell'aria nel tempo (onda sonora)

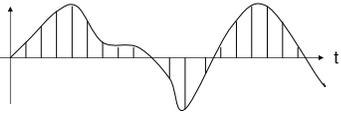


- Sull'asse delle ascisse viene rappresentato il tempo e sull'asse delle ordinate viene rappresentata la variazione di pressione corrispondente al suono stesso

Informatica e sistemi di elaborazione delle informazioni | slide 31

Codifica dei suoni

- Si effettuano dei **campionamenti** sull'onda (cioè si misura il valore dell'onda a intervalli di tempo costanti) e si codificano in forma digitale le informazioni estratte da tali campionamenti

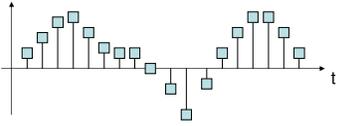


- Quanto più frequentemente il valore di intensità dell'onda viene campionato, tanto più precisa sarà la sua rappresentazione
- Il numero di campioni raccolti per ogni secondo definisce la **frequenza di campionamento** che si misura in Hertz (Hz)

Informatica e sistemi di elaborazione delle informazioni | slide 32

Codifica dei suoni

- La sequenza dei valori numerici ottenuti dai campioni può essere facilmente codificata con sequenze di bit



Una approssimazione!

- La rappresentazione è tanto più precisa quanto maggiore è il numero di bit utilizzati per codificare l'informazione estratta in fase di campionamento

Informatica e sistemi di elaborazione delle informazioni | slide 33

Codifica dei suoni (esempio)

- Se volessimo codificare la musica di qualità CD dovremmo:
 - Usare due registrazioni corrispondenti a due microfoni distinti per ottenere l'effetto stereo
 - Campionare il segnale musicale a 44100 Hz, producendo quindi 44100 campioni al secondo
 - Codificare ogni campione con 16 bit
 - Per cui, il numero di bit che sarebbero necessari per codificare ogni secondo è pari a: $2 \times 44100 \text{ campioni} \times 16 \text{ bit/campione} = 1414200 \text{ bit} = 176775 \text{ B} = 172,6 \text{ KB}$
 - Per un minuto occorrono: $60 \times 172,6 \text{ KB} \sim 10 \text{ MB/minuto}$
 - Questo è il motivo per cui su un CD da 650 MB è possibile registrare circa 74 minuti di musica e su un CD da 700 MB è possibile registrare 80 minuti.

Informatica e sistemi di elaborazione delle informazioni | slide 34

Codifica dei suoni

- **Codifiche standard**
 - WAV (MS-Windows), AIFF (Audio Interchange File Format, Apple)
 - MIDI
 - MP3
- **MIDI**
 - Codifica le note e gli strumenti che devono eseguirle
 - Efficiente, ma solo musica, non voce
- **MP3**
 - MPEG-3: variante MPEG per suoni
 - Grande diffusione, molto efficiente

Informatica e sistemi di elaborazione delle informazioni | slide 35

Domanda

- Qual è la dimensione del file che si ottiene campionando a 100 Hz (100 campioni al secondo) un suono (non stereo) della durata di 10 secondi, ipotizzando di utilizzare campioni da 4 Byte?

Informatica e sistemi di elaborazione delle informazioni | slide 36