



CODIFICA DELL'INFORMAZIONE

CONCETTI DI BASE SU COME LE INFORMAZIONI VENGONO CODIFICATE IN UNA SEQUENZA DI BIT

FONTI

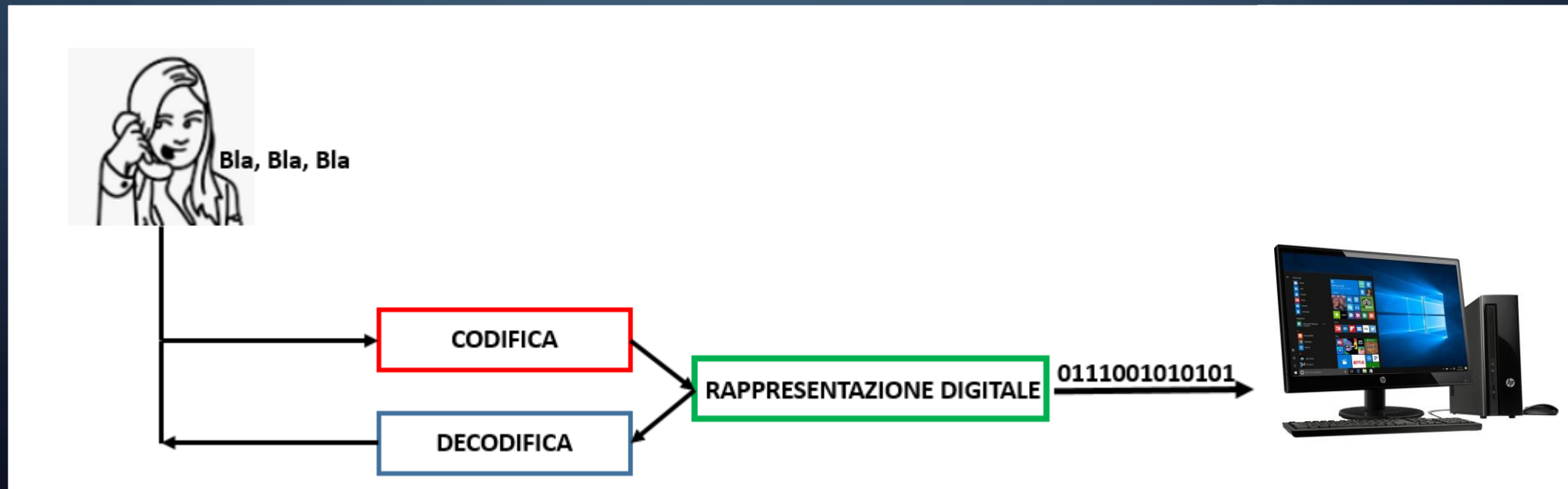
- www.hoepliscuola.it, «tabella dei caratteri ASCII»
- <http://www.unicode.org>
- <http://bilbo.di.unipi.it/~gervasi/EdE06/Lezione%203-a.pdf>

INDICE

- Codifica delle informazioni
 - Codifica ASCII e UNICODE dei caratteri
 - Codifica raster e vettoriale delle immagini

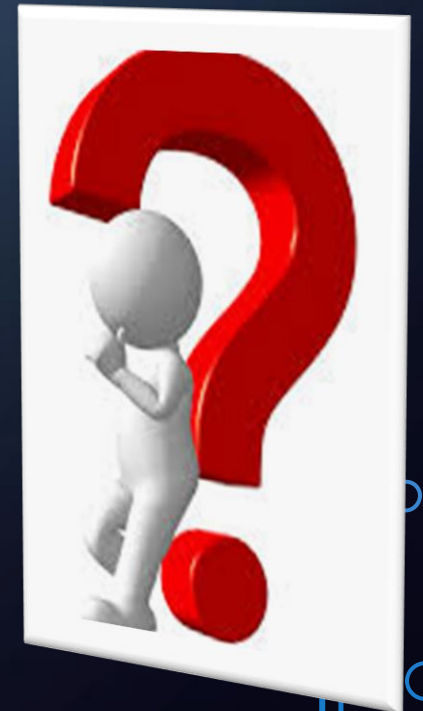
LA CODIFICA DELL'INFORMAZIONE

- Si parla di **codifica dell'informazione** quando si vuole trasformare un qualsiasi fenomeno fisico in una sequenza di bit in modo da averne un modello equivalente da poter immagazzinare in un dispositivo informatico per essere elaborato o trasmesso a distanza
 - L'operazione si può chiamare anche digitalizzazione
- L'operazione inversa prende il nome di **decodifica**



LA CODIFICA DELL'INFORMAZIONE

- A seconda che si voglia digitalizzare un sequenza di caratteri, un'immagine, un filmato, un segnale elettrico è necessario stabilire delle regole uguali per tutti
- Le regole prendono il nome di STANDARD
 - Se non ci fossero gli STANDARD l'informazione perderebbe di significato perché nella sequenza di 0 e 1 non è memorizzata alcuna informazione sull'origine dei dati



CODICE ASCII

ASCII è l'acronimo di **American Standard Computer Information Exchange** ed è la tabella più usata universalmente per la codifica dei dati alfanumerici

- proposto dall'ANSI nel 1963 (definitivo nel 1968) è divenuto lo standard ISO 646
- il codice ASCII standard, codificato su 7 bit, può rappresentare al massimo $2^7=128$ simboli
 - essendo stata concepita in America, la tabella ASCII standard non comprende le lettere accentate
- la versione ASCII ESTESA usa per la codifica un byte (8 bit) e permette la rappresentazione di $2^8=256$ simboli
 - sono quindi rappresentabili i 128 simboli standard più le varianti relative alle diverse lingue nonché i caratteri speciali e grafici
 - alcuni esempi sono: vocali accentate (accento grave à, acuto á, circumflesso â, dieresi ä, tilde ã), lettere modificate (lettere con barrette, cediglie, segni), lettere speciali usate solo in una lingua, segni di punteggiatura particolari (il punto interrogativo ed il punto esclamativo capovolti usati nello spagnolo), simboli di valuta

CODICE ASCII ESTESO

Esistono diverse estensioni della tabella ASCII

- lo standard ISO 8859 prevede 15 diverse estensioni, comprese quelle per gli alfabeti diversi dal latino
 - esistono anche ulteriori estensioni non riconosciute dall'ISO e create per esempio dalla Microsoft per i sistemi Windows o dalla Apple per i Macintosh
- la tabella ASCII estesa tipicamente utilizzata in Italia è quella dell'Europa occidentale, creata per le lingue germaniche e neolatine (escluso il rumeno)
 - altre estensioni usate in Europa sono la Centro Europea per i paesi dell'Europa orientale (lingue slave, ungherese, rumeno), la Turca, la Cirillica e la Greca
 - [ISO-8859-1 \(Latin-1\)](#), utilizzato nella Zona Europea Occidentale
 - [ISO-8859-2 \(Latin-2\)](#), utilizzato zona Europea Orientale (Serbia, Albania, Ungheria, Romania)

CODICE ASCII ISO-8859-1 (LATIN-1)

Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char
0	00	Null	32	20	Space	64	40	Ø	96	60	`	128	80	Ç	160	A0	á	192	C0	Ł	224	E0	α
1	01	Start of heading	33	21	!	65	41	Å	97	61	a	129	81	ù	161	A1	í	193	C1	ł	225	E1	β
2	02	Start of text	34	22	"	66	42	B	98	62	b	130	82	é	162	A2	ó	194	C2	ṛ	226	E2	Γ
3	03	End of text	35	23	#	67	43	C	99	63	c	131	83	â	163	A3	ú	195	C3	ṛ	227	E3	π
4	04	End of transmit	36	24	\$	68	44	D	100	64	d	132	84	ä	164	A4	ñ	196	C4	—	228	E4	Σ
5	05	Enquiry	37	25	%	69	45	E	101	65	e	133	85	à	165	A5	Ñ	197	C5	†	229	E5	σ
6	06	Acknowledge	38	26	&	70	46	F	102	66	f	134	86	å	166	A6	ª	198	C6	‡	230	E6	μ
7	07	Audible bell	39	27	'	71	47	G	103	67	g	135	87	ç	167	A7	º	199	C7	‡	231	E7	τ
8	08	Backspace	40	28	(72	48	H	104	68	h	136	88	ê	168	A8	¿	200	C8	Ł	232	E8	Φ
9	09	Horizontal tab	41	29)	73	49	I	105	69	i	137	89	ë	169	A9	ƒ	201	C9	ŕ	233	E9	Θ
10	0A	Line feed	42	2A	*	74	4A	J	106	6A	j	138	8A	è	170	AA	ƒ	202	CA	Ł	234	EA	Ω
11	0B	Vertical tab	43	2B	+	75	4B	K	107	6B	k	139	8B	ÿ	171	AB	½	203	CB	ŕ	235	EB	δ
12	0C	Form feed	44	2C	,	76	4C	L	108	6C	l	140	8C	î	172	AC	¼	204	CC	‡	236	EC	∞
13	0D	Carriage return	45	2D	-	77	4D	M	109	6D	m	141	8D	ì	173	AD	ı	205	CD	=	237	ED	∞
14	0E	Shift out	46	2E	.	78	4E	N	110	6E	n	142	8E	Ë	174	AE	«	206	CE	‡	238	EE	ε
15	0F	Shift in	47	2F	/	79	4F	O	111	6F	o	143	8F	Ā	175	AF	»	207	CF	±	239	EF	∩
16	10	Data link escape	48	30	0	80	50	P	112	70	p	144	90	É	176	B0	░	208	D0	Ł	240	FO	≡
17	11	Device control 1	49	31	1	81	51	Q	113	71	q	145	91	æ	177	B1	▒	209	D1	ŕ	241	F1	±
18	12	Device control 2	50	32	2	82	52	R	114	72	r	146	92	Æ	178	B2	▓	210	D2	ŕ	242	F2	≥
19	13	Device control 3	51	33	3	83	53	S	115	73	s	147	93	ó	179	B3		211	D3	Ł	243	F3	≤
20	14	Device control 4	52	34	4	84	54	T	116	74	t	148	94	ö	180	B4	┆	212	D4	Ł	244	F4	[
21	15	Neg. acknowledge	53	35	5	85	55	U	117	75	u	149	95	ò	181	B5	┆	213	D5	ŕ	245	F5]
22	16	Synchronous idle	54	36	6	86	56	V	118	76	v	150	96	û	182	B6	┆	214	D6	ŕ	246	F6	÷
23	17	End trans. block	55	37	7	87	57	W	119	77	w	151	97	ù	183	B7	ŕ	215	D7	‡	247	F7	≈
24	18	Cancel	56	38	8	88	58	X	120	78	x	152	98	ÿ	184	B8	ŕ	216	D8	‡	248	F8	°
25	19	End of medium	57	39	9	89	59	Y	121	79	y	153	99	ÿ	185	B9	‡	217	D9	┆	249	F9	·
26	1A	Substitution	58	3A	:	90	5A	Z	122	7A	z	154	9A	Û	186	BA	‡	218	DA	ŕ	250	FA	·
27	1B	Escape	59	3B	;	91	5B	[123	7B	{	155	9B	◊	187	BB	ŕ	219	DB	▀	251	FB	√
28	1C	File separator	60	3C	<	92	5C	\	124	7C		156	9C	£	188	BC	┆	220	DC	▀	252	FC	∂
29	1D	Group separator	61	3D	=	93	5D]	125	7D	}	157	9D	¥	189	BD	┆	221	DD	▀	253	FD	∂
30	1E	Record separator	62	3E	>	94	5E	^	126	7E	~	158	9E	€	190	BE	┆	222	DE	▀	254	FE	▀
31	1F	Unit separator	63	3F	?	95	5F	□	127	7F	□	159	9F	ƒ	191	BF	┆	223	DF	▀	255	FF	□

CODICE UNICODE

L'**UNICODE** è uno standard internazionale nato per superare le limitazioni del codice ASCII

- è definito dalla Unicode Consortium e dalla International Organization for Standardization (ISO 10646)
- rappresenta i caratteri usando 2 byte (16 bit)
 - il numero di combinazioni possibili diventa $256 \times 256 = 65.536$, perciò Unicode supporta 65.536 diversi segni, al posto dei 256 del set ASCII esteso
 - si riescono così a rappresentare non solo tutte le varianti dell'alfabeto latino, ma anche tutti gli altri alfabeti (greco, cirillico, arabo, ebraico...) oltre all'insieme degli ideogrammi cinesi e giapponesi (che sono in tutto circa 30.000, anche se poi ne vengono effettivamente utilizzati solo poche centinaia)
- lo standard definitivo è ancora in corso di definizione
- lo svantaggio dell'Unicode, rispetto all'ASCII, è che le dimensioni dei file di testo risultano comunque raddoppiate (vengono usati 2 byte per carattere, invece di 1 solo)
- <http://www.unicode.org/charts/> è l'indirizzo al quale si possono trovare le tabelle UNICODE

ESEMPIO DI CONFRONTO TRA ASCII E UNICODE

ASCII/8859-1 Text

A	0100 0001
S	0101 0011
C	0100 0011
I	0100 1001
I	0100 1001
/	0010 1111
8	0011 1000
8	0011 1000
5	0011 0101
9	0011 1001
-	0010 1101
l	0011 0001
	0010 0000
t	0111 0100
e	0110 0101
x	0111 1000
t	0111 0100

Unicode Text

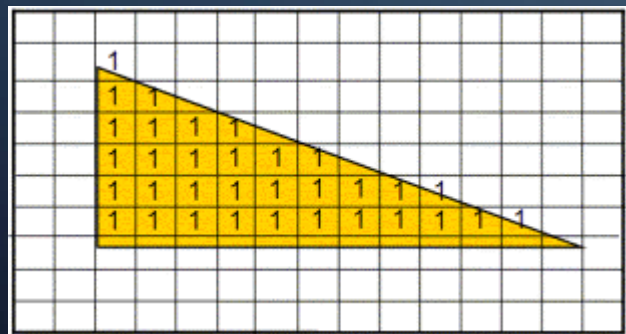
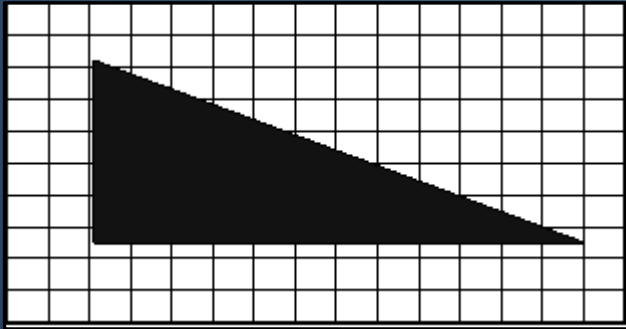
A	0000 0000 0100 0001
S	0000 0000 0101 0011
C	0000 0000 0100 0011
I	0000 0000 0100 1001
I	0000 0000 0100 1001
	0000 0000 0010 0000
天	0101 1001 0010 1001
地	0101 0111 0011 0000
	0000 0000 0010 0000
س	0000 0110 0011 0011
ل	0000 0110 0100 0100
ا	0000 0110 0010 0111
م	0000 0110 0100 0101
	0000 0000 0010 0000
α	0000 0011 1011 0001
£	0010 0010 0111 0000
γ	0000 0011 1011 0011

t	0111 0100
---	-----------

λ	
---	--

0000 0011 1011 0011

CODIFICA BITMAP DELLE IMMAGINI



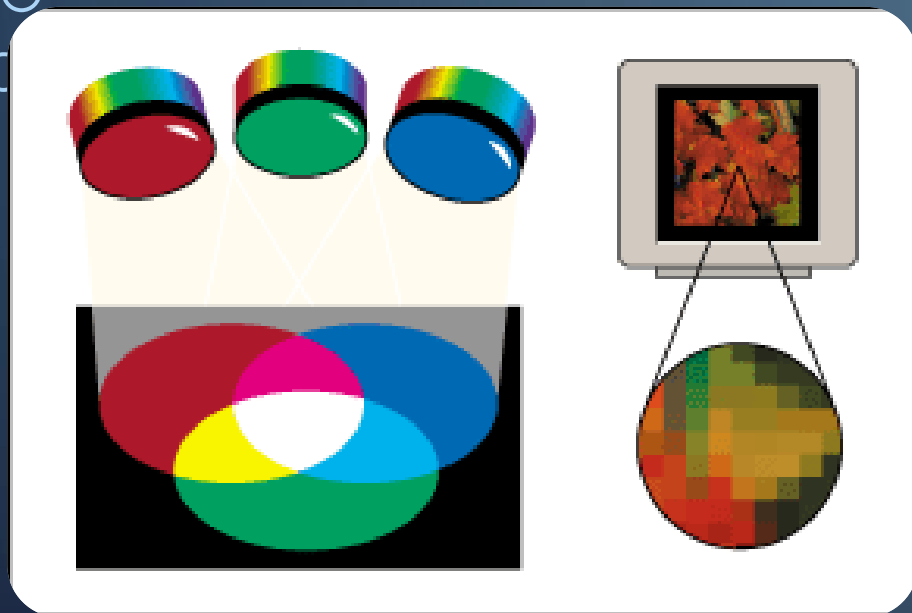
- La codifica bitmap (o raster) prevede di dividere l'immagine in una fitta maglia e di assegnare un colore ad ogni quadretto che viene chiamato pixel
- Viene memorizzato il colore del singolo pixel
 - per esempio nel caso di immagini in bianco e nero potrebbe essere memorizzato un «1» nel caso di pixel nero e «0» nel caso di pixel bianco
- Il numero di punti, calcolato come

Numero Colonne x Numero Righe

rappresenta la risoluzione di un'immagine

- per esempio 640x480
- la risoluzione può venire espressa anche in ppi o dpi (rispettivamente point per inch e dot per inch); per es. 300 dpi

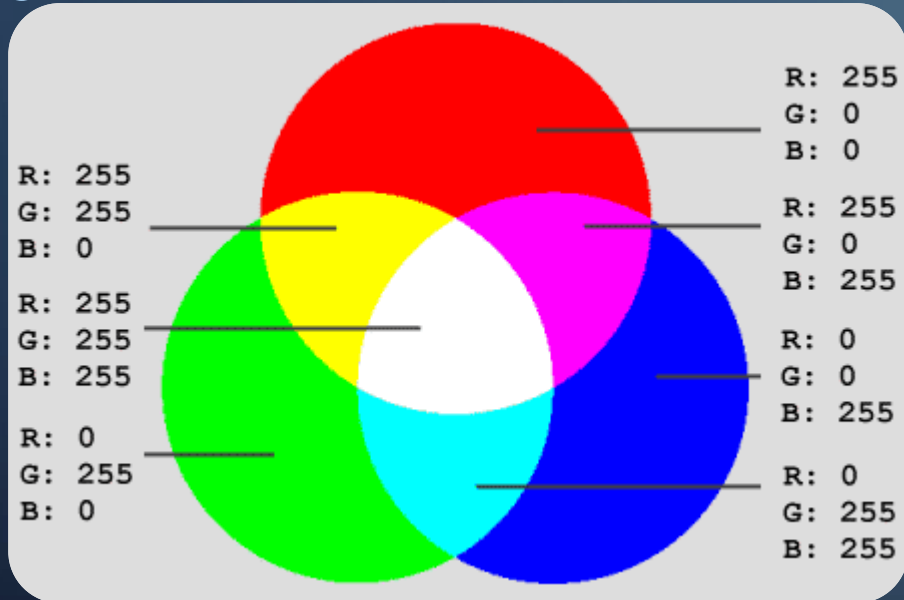
CODIFICA BITMAP DELLE IMMAGINI



- Nel caso di immagini a colori, il colore di ogni pixel può essere pensato come la somma delle intensità dei tre colori fondamentali
 - Nel caso di «colori luce», ovvero i colori emessi da una sorgente luminosa come un display, i tre colori fondamentali sono rosso (red: R), verde (green G) e blu (blu B)
 - Il bianco è dato dalla somma dei tre colori fondamentali, il nero dalla loro assenza
- Il numero di bit utilizzati per rappresentare il colore di un singolo pixel si chiama PROFONDITÀ DI COLORE.
- Disponendo di un byte per ogni componente di colore (24 bit di profondità di colore), è possibile rappresentare $256 * 256 * 256 = 16.777.216$ colori

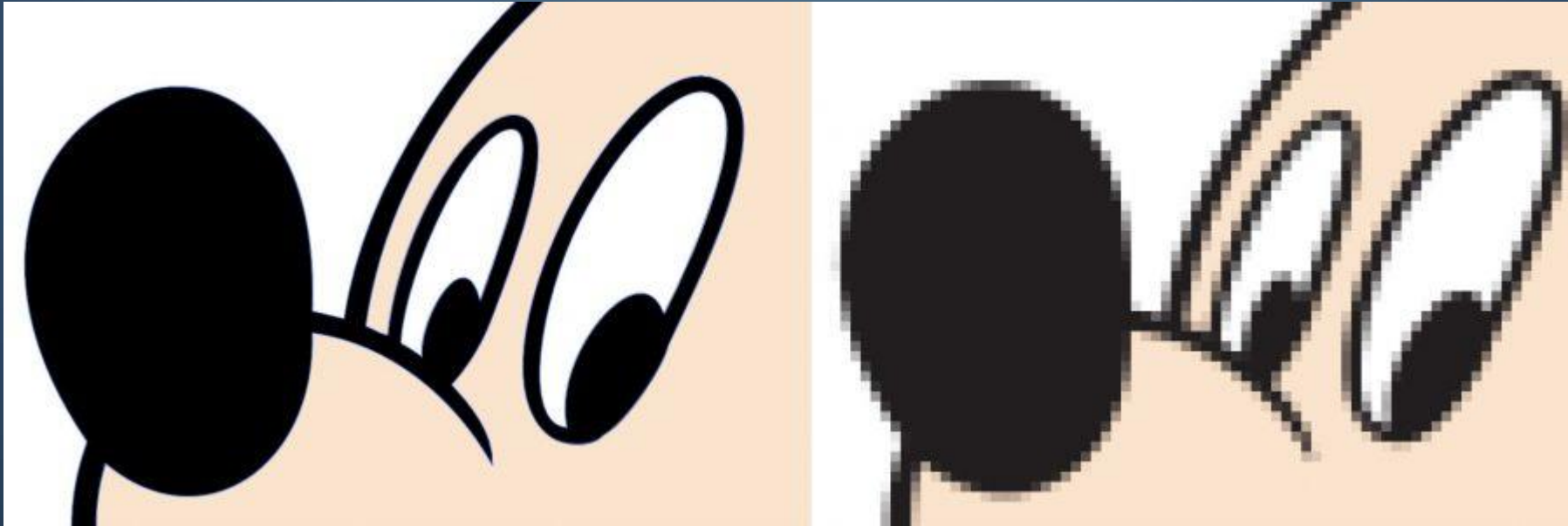
CODIFICA DEI COLORI RGB

- La codifica RGB prevede di assegnare 1 byte a ciascun colore primario per un totale di 3 byte



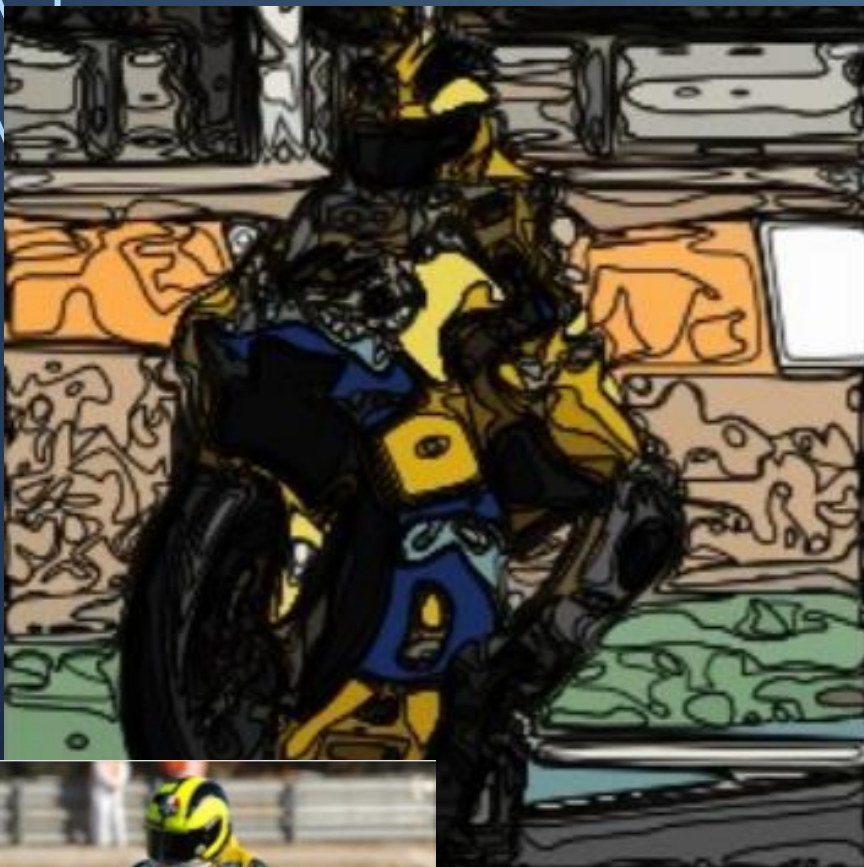
- 1 byte equivale a 8 bit e quindi sono rappresentabili 256 intensità di ogni colore, da 0 (assenza di colore) a 255 (massima intensità di colore)
- per esempio il colore rosso è rappresentato dalla sequenza 255 0 0, il blu dalla sequenza 0 0 255, il magenta dalla sequenza 255 0 255
- i colori totali rappresentabili sono 16.777.216

CODIFICA VETTORIALE DELLE IMMAGINI



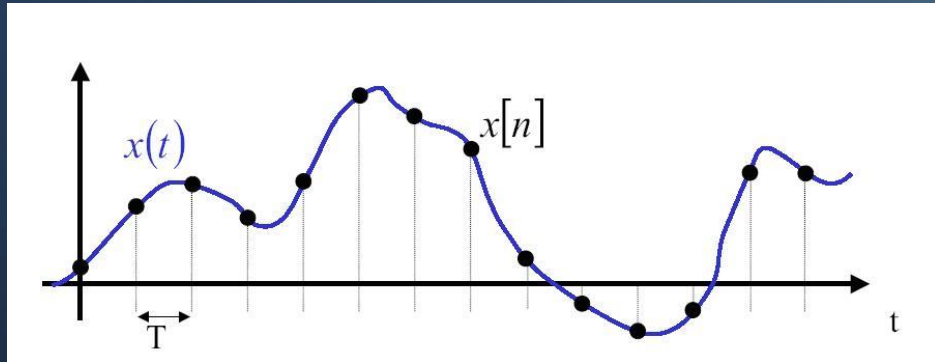
- La risoluzione delle immagini ha un ruolo fondamentale nelle operazioni di ridimensionamento. Se si ingrandisce eccessivamente un'immagine a bassa risoluzione appare evidente il fenomeno della quadrettatura
- Per ovviare a questo problema si è introdotta la codifica vettoriale

CODIFICA VETTORIALE DELLE IMMAGINI



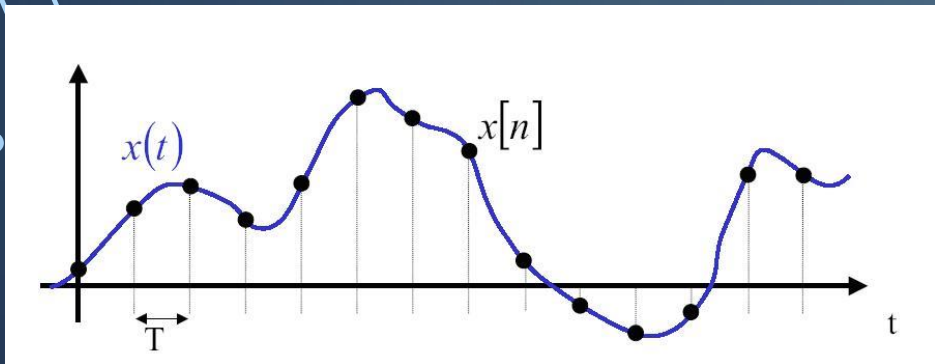
- Le immagini vettoriali si possono pensare come sovrapposizione di curve chiuse semplici (per esempio rettangoli, ellissi, linee spezzate chiuse) chiamate primitive grafiche di cui si conosce la rappresentazione matematica
- Le primitive vengono memorizzate come una sequenza di numeri binari che codificano la loro forma, la dimensione e il colore
- Questa tipologia di immagini può venire ingrandita a piacimento senza perdere di risoluzione poiché lo zoom viene effettuato scalando le primitive di cui si conosce la rappresentazione matematica

CAMPIONAMENTO

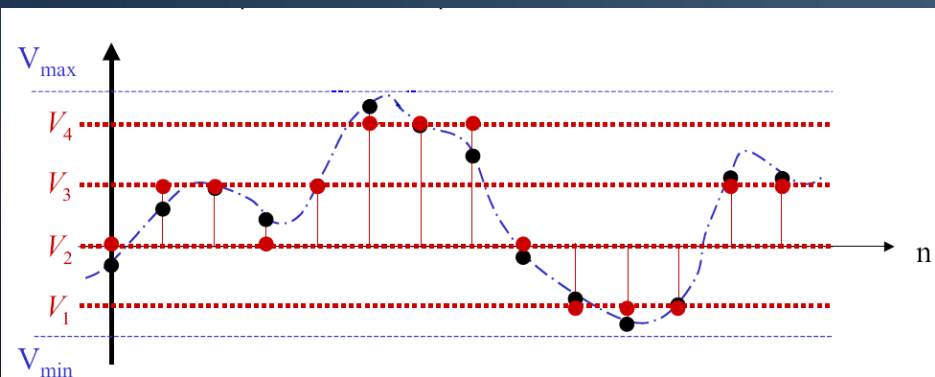


- Il campionamento è una tecnica che consiste nel convertire un segnale continuo nel tempo in uno discreto valutandone l'ampiezza a intervalli di tempo regolari
- Il teorema del campionamento (o di Nyquist-Shannon) garantisce che se la frequenza di campionamento è almeno il doppio della banda del segnale con un semplice filtro passa basso è possibile ricostruire senza distorsioni il segnale di partenza
 - A livello pratico si campiona con frequenza pari a circa 5 volte la banda del segnale
 - Per banda si intende l'insieme di frequenze che compongono la trasformata di Fourier di un segnale

QUANTIZZAZIONE E CODIFICA



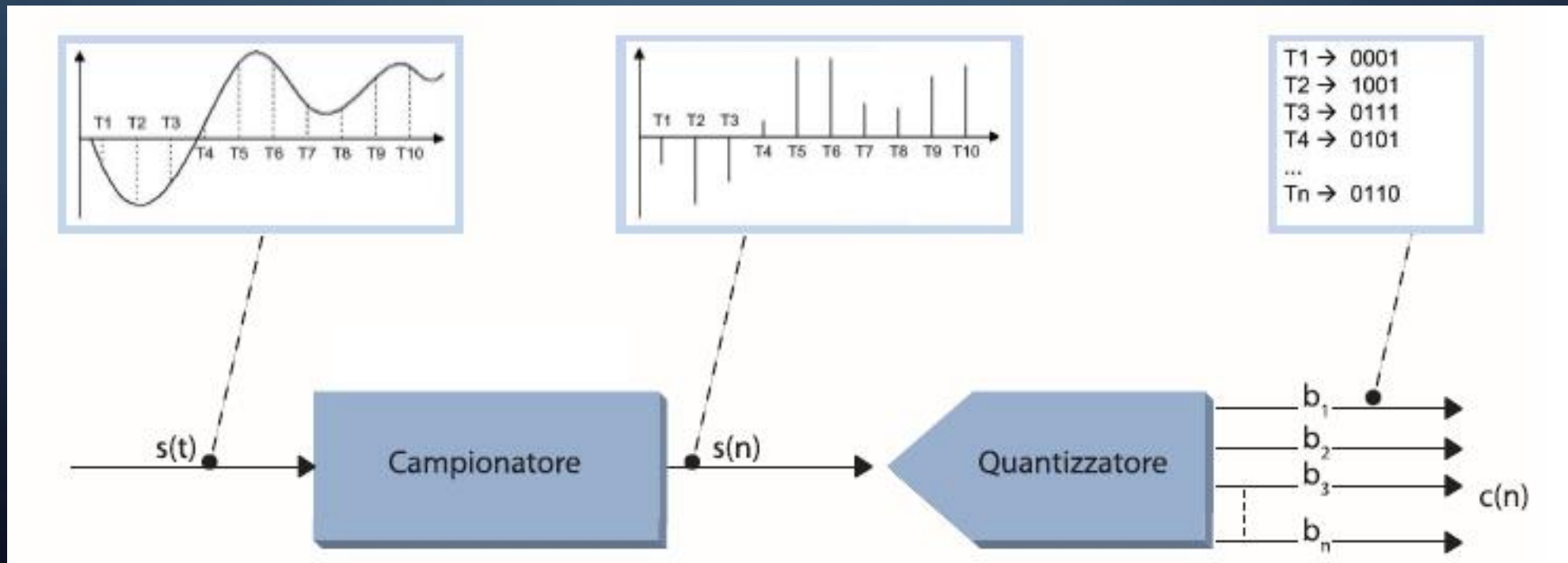
- A causa dell'impossibilità di una rappresentazione reale a precisione infinita, nelle comunicazioni di tipo "numerico" o "digitale", il valore della grandezza in questione deve essere convertito in formato discreto. L'operazione prende il nome di quantizzazione
- Poiché l'insieme dei valori V_i è finito è possibile associare ad ognuno un valore numerico binario che può essere memorizzato (codifica)



Valore quantizzato	Valore binario
V_1	00
V_2	01
V_3	10
V_4	11

QUANTIZZAZIONE

- Mentre l'operazione di campionamento è lineare e invertibile, l'operazione di quantizzazione introduce sempre un errore di «approssimazione» che prende il nome di errore di quantizzazione
- Per ridurre l'errore di quantizzazione è necessario aumentare il numero dei livelli
 - Un campionatore reale ha tipicamente un'uscita a 10-12 bit corrispondenti rispettivamente a 1024-4096 livelli di quantizzazione



CONCLUSIONI

- Per poter essere trasmesse le informazioni devono essere convertite in una sequenza di bit
- L'insieme delle regole per trasformare le informazioni in sequenze di bit si chiama codifica
- I caratteri si codificano in ASCII e più recentemente in UNICODE
- Nelle immagini raster ad ogni pixel è assegnato un colore
- La codifica RGB assegnando 1 byte a ciascun colore primario permette di rappresentare 16.777.216 colori
- I valori di tensione e corrente vengono codificati tramite il processo che prende il nome di conversione analogico-digitale (A/D)