

# CODIFICA DATI

**Ing. Daniele Corti**



copyright

all rights reserved

Copyright © Ing. Daniele Corti 2013

[www.ingdanielecorti.it](http://www.ingdanielecorti.it)

Tutti i diritti sono riservati a norma di legge e a norma delle convenzioni internazionali.

Ver.1.0

## **PREREQUISITI**

- ✓ Linguaggio binario.

## **OBIETTIVI**

- ✓ Conoscere la codifica dei dati.

## **ARGOMENTI**

- ✓ Codifica dei dati: il codice.
- ✓ Tipologia di informazioni.
- ✓ Tipologia di codifica.
- ✓ La digitalizzazione dei dati.
- ✓ Perché codificare.
- ✓ I diversi modi di interpretazione dei dati.
- ✓ Il sistema di codifica.

## CAP 3 - CODIFICA DATI

### CODIFICA DEI DATI: IL CODICE

Le **informazioni** gestite dai sistemi di elaborazioni (es il computer) devono essere **codificate** (convertite) in modo tale che siano facilmente:

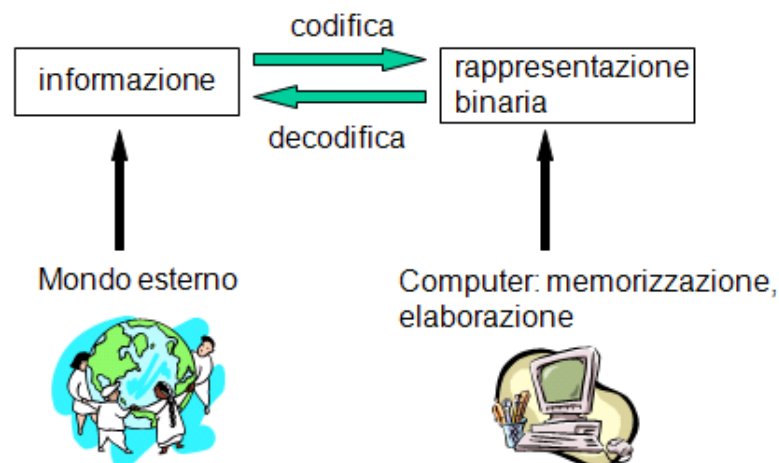
- archiviabili (memorizzabili) nella memoria principale o secondaria,
- modificabili (elaborabili) dal processore,
- recuperabili e scambiabili utilizzando i dispositivi di input/output.

**La codifica è una tecnica con la quale un dato viene rappresentato mediante un insieme finito di simboli.**

### TIPOLOGIA DI INFORMAZIONI

Gli esseri umani utilizzano e interpretano le informazioni con innumerevoli strumenti:

- Suoni.
- Colori.
- Disegni, immagini, video
- Simboli.
- Lettere.



Le informazioni che intendiamo trattare non sono costituite solo da numeri. Ad esempio, voi state leggendo delle lettere e vedete sullo schermo delle immagini.

Per rappresentare lettere, immagini, suoni partendo dai numeri occorre utilizzare dei sistemi di codifica. Questi sistemi stabiliscono una corrispondenza fra un determinato numero e una particolare informazione. L'informazione rappresentata con numeri viene detta digitalizzata.

Testo, numeri, immagini, video → codice binario
---

Le possibili informazioni rappresentabili al computer sono:

- Numeri.
- Caratteri (testi alfanumerici).
- Dati multimediali (immagini, filmati, suoni).

## TIPOLOGIA DI CODIFICA

In base al tipo di dato da convertire parleremo di

- Codifica dei numeri
  - Numeri naturali (interi senza segno – positivi)
  - Relativi (interi con segno)
  - Numeri razionali.
- Codifica alfanumerica - Codifica ASCII.
- Codifica di dati multimediali
  - Immagini.
  - Video.
  - Suoni.

## LA DIGITALIZZAZIONE DEI DATI

Qualsiasi sia il dato da trattare il calcolatore lo codifica in binario. Parliamo allora di codice. L'informazione rappresentata con codici binari viene detta **digitalizzata**.

I calcolatori, infatti, utilizzano l'aritmetica binaria per la rappresentazione dei dati (numeri, lettere, immagini, video, ecc.) basandosi su solo due simboli di base (il bit può assumere solo due stati logici: 0 o 1).

Spesso alla codifica viene associata la “**compressione dei dati**”; la compressione dati è una tecnica che consente di ridurre le dimensioni dei dati al fine di rendere più veloce la trasmissione dei dati stessi.

### **Esempio**

La spia dell’olio in un macchinario quando è spenta indica che tutto è a posto, mentre quando si accende indica un’anomalia.

Ovviamente l’utilizzo di sole due diverse informazioni non è molto utile e quindi si utilizzano più bit.

Analogamente al nostro metodo che ci permette di codificare le informazioni componendo parole e frasi accostando più lettere, usando più bit si aumenta il numero di informazioni diverse rappresentabili.

### **Esempio**

Indicando con il primo bit la spia dell’olio e con il secondo quella dei freni:

00 => olio ok!, freni ok!

01 => olio ok!, freni ko!

10 => olio ko!, freni ok!

11 => olio ko!, freni ko!

In generale, quindi, avendo a disposizione n bit si possono rappresentare  $2^n$  sequenze diverse e quindi codificare altrettante informazioni.

Si è concordato di utilizzare sequenze di 8 bit per formare l’unità effettiva di misura della memoria e si è attribuito a questa quantità il nome di Byte e quindi, come già osservato:

8 bit = 1 Byte

Utilizzando un Byte è quindi possibile rappresentare 256 diverse informazioni.

## Esempi

- **Codice Morse:** ad ogni lettera viene assegnata una sequenza di punti e linee.
- **Codifica digitale:** utilizzata per la conversione di un segnale analogico in digitale.

In molti contesti della vita di tutti i giorni abbiamo a che fare con i codici, per esempio, nella codifica dei prodotti a peso variabile, quelli che vengono confezionati, prezzati e posti sugli scaffali a disposizione dei consumatori finali per l'acquisto:



Il codice visualizzato nella precedente figura è quello utilizzato in Italia, basato su 13 cifre e include anche il prezzo di vendita del prodotto.

## PERCHE' CODIFICARE

La codifica, o conversione, è necessaria per far sì che le informazioni siano facilmente

- Archiviabili,
- Modificabili,
- Elaborabili e,
- Recuperabili.

Abbiamo in precedenza definito il bit come l'unità di memoria di base in campo informatico. Usando un bit si possono definire (codificare) solo due differenti informazioni.

Il computer è molto più semplice dell'essere umano e grazie alla sua struttura fisica (l'hardware) è quindi destinato ad usare una logica di tipo binario del genere:

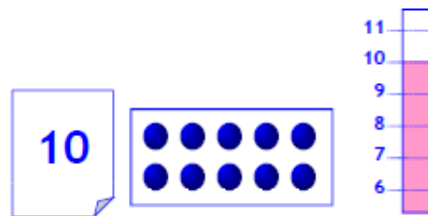
- Sì/no.
- 0/1.
- Accesso/Spento.

- On/Off.

## I DIVERSI MODI DI INTERPRETAZIONE DEI DATI

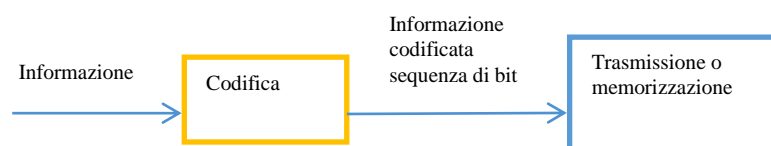
Per realizzare un algoritmo (e conseguentemente un programma) i dati e le informazioni devono essere rappresentati in un formato (modalità con cui i dati vengono rappresentati elettronicamente) tale che l'esecutore sia in grado di salvarli e manipolarli.

Un dato può essere rappresentato in modi differenti. Per esempio il valore numerico 10 può essere codificato in questi modi:



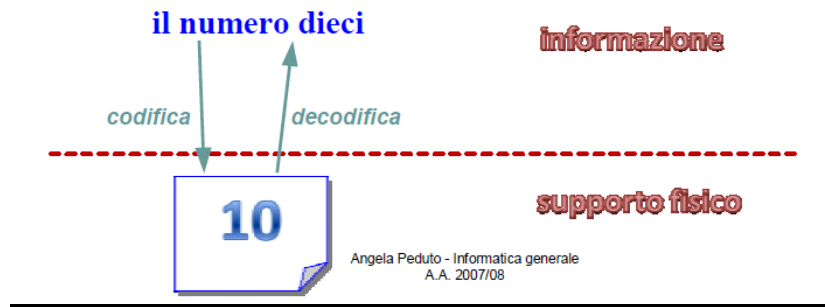
**Tramite il codice possiamo specificare come un dato debba essere interpretato affinché fornisca informazione.**

**La conversione (codifica) di un dato in bit (codifica binaria) permette la gestione dei dati con il calcolatore e li rende utilizzabili a velocità sempre maggiori.**



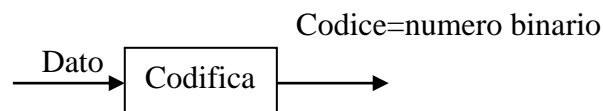
**La codifica è l'operazione con cui l'informazione viene scritta (o trasmessa) su un supporto fisico.**

**La decodifica è l'operazione con cui l'informazione viene letta (o ricevuta) da un supporto fisico.**



## IL SISTEMA DI CODIFICA (O CODIFICA, O CODICE)

È necessario stabilire delle regole di corrispondenza tra dato e codice, dette **codifiche**.

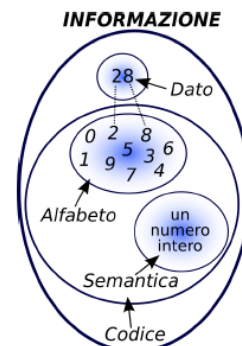


Un calcolatore necessita di dati codificati in un formato adatto alla memorizzazione e manipolazione da parte di un esecutore automatico. Per formalizzare dati come numeri e caratteri alfabetici, si utilizzano *successioni di simboli*, scelti da un insieme finito detto *alfabeto*.

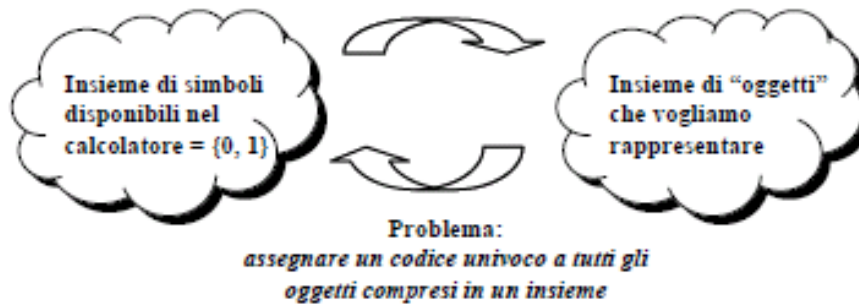
Il sistema di codifica utilizza un insieme di simboli di base (alfabeto).

- I simboli dell'alfabeto possono essere combinati attraverso differenti configurazioni (o codici, o stati), distinguibili l'una dall'altra.
- Associa ogni configurazione ad una particolare entità di informazione (la configurazione diventa un modo per rappresentarla).

- Nell'esempio, l'alfabeto relativo all'età di un paziente (alfabeto decimale) è l'insieme dei simboli  $\{0, 1, 2, \dots, 8, 9\}$
- Il dato "28" è rappresentato nell'alfabeto decimale dalla parola composta dalla sequenza 2 e 8







### Sistema di codifica: Numeri Interi (Decimali)

- **Alfabeto**
  - Cifre "0", "1", "2", ..., "9"
  - separatore decimale (".")
  - separatore delle migliaia (",")
  - Segni positivo ("+") e negativo ("-")
- **Regole di composizione (sintassi)**
  - Definiscono le combinazioni ben formate
  - 12.318,43
  - ✓ 123.18.43 ?
- **Codice (semantica)**
  - Associano ad ogni configurazione un'entità di informazione
  - $12.318,43 = 1 \times 10^4 + 2 \times 10^3 + 3 \times 10^2 + 1 \times 10^1 + 1 \times 10^0 + 4 \times 10^{-1} + 3 \times 10^{-2}$
- **Lo stesso alfabeto può essere usato per codici diversi**
  - $123,456 = 1 \times 10^2 + 2 \times 10^1 + 3 \times 10^0 + 4 \times 10^{-1} + 5 \times 10^{-2} + 6 \times 10^{-3}$  [IT]
  - $123,456 = 1 \times 10^5 + 2 \times 10^4 + 3 \times 10^3 + 4 \times 10^2 + 5 \times 10^1 + 6 \times 10^0$  [UK]

### RIVEDERE

#### Codifica binaria – parte 1

L'alfabeto binario è l'unico alfabeto utilizzabile nel calcolatore, costituito dai simboli {0,1}, per:

- Semplicità: i circuiti elettronici non sono complessi e sono costituiti da elementi a due stati.
- Efficienza: esecuzione veloce di operazioni.

Tuttavia, è un sistema di difficile e non immediata comprensione per l'uomo.

I computer sanno eseguire solo operazioni nel linguaggio binario.

- **Alfabeto binario: due simboli**
- **Utilizzata nei sistemi informatici**
  - Si utilizza una grandezza fisica (luminosità, tensione elettrica, la corrente elettrica), per rappresentare informazione
- **BIT (BInary digiT)**
  - unità elementare di informazione rappresentabile con dispositivi elettronici
  - ✓ Due soli stati
  - con 1 bit si possono rappresentare 2 stati
  - 0/1, on/off, si/no

Per funzionare i computer hanno bisogno di:

- Dati sui quali agire.
- Un programma che li istruisca su cosa fare In quale modo operare sui dati).

Abbiamo quindi bisogno di tradurre sia i programmi che i dati in un linguaggio “comprensibile” al computer.

Parliamo, infatti, di: **rappresentazione e codifica dell’informazione.**

- **Quanti oggetti/entità si possono codificare con k bit?**
  - 1 bit  $\Rightarrow$  2 stati {0,1}  $\Rightarrow$  2 oggetti (e.g., Vero, Falso)
  - 2 bit  $\Rightarrow$  4 stati {00,01,10,11}  $\Rightarrow$  4 oggetti
  - 3 bit  $\Rightarrow$  8 stati {000,001,010,...,111}  $\Rightarrow$  8 oggetti
  - ...
  - K bit  $\Rightarrow 2^k$  stati  $\Rightarrow 2^k$  oggetti
- **Quanti bit servono per codificare N oggetti?**
  - $N \leq 2^k \Rightarrow k \geq \log_2 N \Rightarrow k = \lceil \log_2 N \rceil$  (intero superiore)
- **Attenzione: Ipotesi implicita che i codici abbiano tutti la stessa lunghezza**

### Esempio di Codifica Binaria

- **Problema: assegnare un codice binario univoco a tutti i giorni della settimana**
- **Giorni della settimana:  $N=7$ ,  $\Rightarrow k \geq \lceil \log_2 N \rceil \Rightarrow k=3$** 
  - Con 3 bit, 8 configurazioni
- **Possibile soluzione**
  - Lunedì            000
  - Martedì           001
  - Mercoledì        010
  - Giovedì            011
  - Venerdì            100
  - Sabato             101
  - Domenica         110

### Esempio: i mesi dell’anno

Abbiamo  $N=12$  differenti configurazioni:  $N \leq 2^K \rightarrow \lg_2 N \leq K \rightarrow K \geq 4$

Configurazioni	Codice
Gennaio	0000
Febbraio	0001

.....	.....
Dicembre	1100

### Tipo di dati di base del calcolatore

- Di base, il calcolatore tratta – utilizzando il solo alfabeto binario – tre tipologie di informazione
  - Informazione booleana: vero/falso
  - Informazione numerica: valore numerico intero/reale
  - Informazione testuale: carattere
- Per aggregazione di questi tipi di dato *elementari*, si esprimono gli altri tipi di dato (tipi *derivati* o *strutturati*)

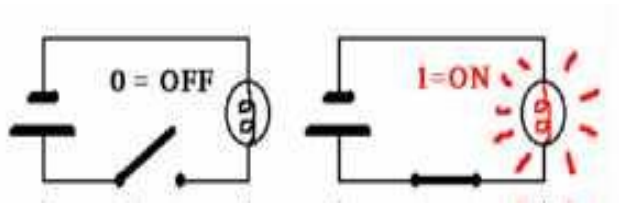
### Codifica binaria – parte 2

La codifica binaria è la rappresentazione dell'informazione (caratteri alfanumerici, immagini, suoni etc.) effettuata utilizzando un alfabeto limitato a soli due caratteri (0, 1), a causa della capacità dei circuiti di un elaboratore di utilizzare o memorizzare solo cifre binarie (corrispondenti ai due possibili stati di un circuito elettrico: aperto o chiuso).

Il sistema di numerazione che utilizziamo comunemente è quello decimale: dieci simboli (0,1,...,9) che raggruppati in opportune sequenze rappresentano i numeri naturali. Il sistema che utilizziamo è inoltre posizionale, perché il significato di ogni cifra dipende dalla sua posizione nella sequenza che costituisce il numero da rappresentare (6435 è diverso da 3465). Possiamo scegliere un insieme diverso di simboli per costituire un sistema di numerazione posizionale. Genericamente possiamo avvalerci di B simboli (sistema in base B). La codifica binaria è basta su un sistema di numerazione binario, in base 2: i simboli che compongono l'alfabeto sono solo 2 (0 e 1). I dati in un sistema di codifica binaria vengono infatti rappresentati attraverso '0' e '1', un sistema cioè che utilizza due sole cifre e ha un importante vantaggio: i dati binari sono facilmente rappresentabili (e manipolabili) all'interno di un computer. Per distinguere la base utilizzata, viene utilizzato un pedice:

- I numeri in base 10 saranno scritti come  $853_{10}$  e  $675_{10}$ ;
- I numeri in base 2 saranno scritti come  $100_2$  e  $110_2$ .

Ogni elemento che assume un valore binario viene indicato con il termine bit (da binarydigit, cifra binaria). Il bit rappresenta l'unità elementare di informazione, una scelta tra "si" e "no", due risposte che possono essere associate ai valori "0" e "1". Il funzionamento dei circuiti elettrici di tutti i calcolatori moderni è basato su due stati elementari: la presenza oppure l'assenza di un segnale elettrico.



Il passaggio della corrente elettrica con l'interruttore chiuso e la lampadina accesa rappresenta il simbolo 1, mentre il non passaggio di corrente elettrica con l'interruttore aperto e la lampadina spenta rappresenta il simbolo 0. Ovviamente, la rappresentazione di informazioni più complesse richiede l'uso di un insieme di bit. In breve la risposta "si" o "no" a una domanda porta dunque 1 bit di informazione. La risposta a due domande di tale genere (che pone una soluzione con 4 alternative: "si-si", "no-no", "si-no", "no-si") porta 2 bit di informazione. A una successione di 8 bit, con cui possono quindi essere rappresentati  $2^8=256$  dati diversi, si dà il nome di Byte, e di questo si considerano i multipli KiloByte, MegaByte, GigaByte e TeraByte. Questi prefissi sono associati generalmente a potenze crescenti di 10: rispettivamente  $10^3$ ,  $10^6$ ,  $10^9$ ,  $10^{12}$ ; nel caso binario invece questi multipli sono definiti in termini di potenze di 2 (essendo la base 2). Un KByte corrisponde quindi a  $2^{10}$  Byte, un MByte a  $2^{10}$ KByte e così via.

1 bit = unità elementare di informazione, ovvero la risposta "si" o "no" a una domanda;

1 Byte = 8 bit

1 KByte =  $2^{10}$  Byte = 1024 Byte;

1 MByte =  $2^{10}$  KByte =  $2^{20}$  Byte = 1 048 576 Byte;

1 GByte =  $2^{10}$ MByte =  $2^{20}$ KByte =  $2^{30}$  Byte;

1 TByte =  $2^{10}$ GByte =  $2^{20}$ MByte =  $2^{30}$ KByte =  $2^{40}$  Byte.

## ESERCIZI

Dati 6 bit per la codifica, quante informazioni distinte si possono rappresentare?

$2^6=64$  informazioni distinti

Dato un byte per la codifica, quante informazioni distinte si possono rappresentare?

Un byte = 8 bit,  $2^8=256$  informazioni distinti

Quanti bit si devono utilizzare per rappresentare 20 informazioni distinte?

Almeno 5 bit, perché  $2^5=32 \Rightarrow 20$  (4 bit non sono sufficiente, perché  $2^4=16 < 20$ )

Quanti byte occupa la frase "cervello" scritta in ASCII?

8 (in ASCII, un carattere corrisponde a un byte)

Quanti byte occupa la frase "dipartimento di psicologia" scritta in ASCII?

26

Quanti byte occupa un suono della durata di 5 secondi campionato a 30 Hz (30 campioni per secondo), in cui ogni campione occupa 6 byte?

$5 \times 30 \times 6 = 900$  Byte

Hai ricevuto un messaggio di posta elettronica da un amico. Il messaggio contiene:

un testo di 300 caratteri scritto in ASCII,

un'immagine di 120x150 pixel con 1024 colori.

Quanti byte occupa il messaggio?

Testo: 300 Byte. Immagine: ogni pixel richiede 10 bit (perché  $2^{10}=1024$ ); l'immagine ha  $120 \times 150 = 18000$  pixel; l'immagine occupa  $10 \times 18000 = 180000$  bit = 22500 Byte. Testo + immagine:  $300 + 22500 = 22800$  Byte

Un secondo di suono campionato a 512 Hz occupa 1 KB. Quanti valori distinti possono avere i campioni?

1 KB = 1024 byte; numero di campioni =  $1 \times 512 = 512$ ; ogni campione contiene  $1024/512 = 2$  byte; 2 byte = 16 bit;  $2^{16}$  valori distinti