

Corso Integrato di Statistica Informatica e Analisi dei dati Informatica - 4a

Dr Carlo Meneghini

Dip. di Fisica "E. Amaldi"
via della Vasca Navale 84
meneghini@fis.uniroma3.it

st. - 83 -

tel.: 06 55177217

Home page

<http://www.fis.uniroma3.it/~meneghini>

Codificare le informazioni

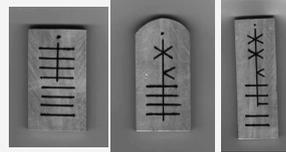
C. Meneghini - C.I.-Informatica

1

La quantità

E una proprietà comune ad ogni insieme di oggetti,
è facilmente rappresentabile e catalogabile,
l'elaborazione richiede 4 operazioni fondamentali + - × /

Tavolette per registrare la quantità di latte (Austria)



44

190

277

da: <http://www.tecnoteca.it/contenuti/museo/>

Tavoletta Sumera con valori numerici



C. Meneghini - C.I.-Informatica

2

Rappresentazione dei numeri

I numeri rappresentano quantità

Il sistema decimale

è un sistema di numerazione che utilizza 10 simboli (cifre):

0, 1, 2, ..., 9

è un sistema **posizionale**

il significato dei simboli (cifre) dipende dalla loro posizione:

$$\begin{array}{r}
 \begin{array}{ccc}
 7 & 9 & 2 \\
 \swarrow & \uparrow & \searrow \\
 \text{centinaia} & \text{decine} & \text{unità}
 \end{array} \\
 700 + 90 + 2 \\
 7 \cdot 100 + 9 \cdot 10 + 2 \cdot 1 \\
 7 \cdot 10^2 + 9 \cdot 10^1 + 2 \cdot 10^0
 \end{array}$$

3

Sistemi di numerazione posizionale

Ingredienti:

- insieme di simboli (cifre) C_i
- base B = numero di simboli

C_i = cifre: valori interi progressivi

Cifra più significativa \swarrow $C_n \dots C_2 C_1 C_0$ \nwarrow Cifra meno significativa

$$C_n \cdot B^n + \dots + C_2 \cdot B^2 + C_1 \cdot B^1 + C_0 \cdot B^0 = \text{quantità}$$

Attenzione: C_n è il valore numerico della cifra n-esima

4

Il sistema Binario

E un sistema di numerazione posizionale

che utilizza 2 cifre: 0, 1

Base = 2

1101101^{bin}

$$\begin{aligned}
 &1 \cdot 2^6 + 1 \cdot 2^5 + 0 \cdot 2^4 + 1 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 = \\
 &= 1 \cdot 64 + 1 \cdot 32 + 0 \cdot 16 + 1 \cdot 8 + 1 \cdot 4 + 0 \cdot 2 + 1 \cdot 1 =
 \end{aligned}$$

= 109^{dec}

C. Meneghini - C.I.-Informatica

5

Sistemi di numerazione per PC

B = 2: sistema binario (bin)

B = 10: sistema decimale (utenti umani)

B = 8: ottale (oct)

B = 16: esadecimale (HEX)

Dec	Binario	ottale	HEX
0	0 0000	0	0
1	0 0001	1	1
2	0 0010	2	2
3	0 0011	3	3
4	0 0100	4	4
5	0 0101	5	5
6	0 0110	6	6
7	0 0111	7	7
8	0 1000	10	8
9	0 1001	11	9
10	0 1010	12	A
15	0 1111	17	F
16	0 0001 0000	20	10
255	0 1111 1111	377	FF
256	1 0000 0000	400	100

C. Meneghini - C.I.-Informatica

6

Proprietà dei sistemi di numerazione posizionale

- in base **B** un **1** seguito da **n-0** rappresenta **B^N**

es.: N=3

binario (B=2): $1000 = 2^3 = 8$

Hex (B=16): $1000 = 16^3 = 4096$

- con **N** cifre posso rappresentare **B^N** valori diversi, compreso lo 0 e quindi contare fino a **B^N-1**

es.: N=3

binario: $111^{bin} = 2^3 - 1 = 7$

decimale: $999^{dec} = 10^3 - 1 = 999$

Hex: $FFF^{hex} = 16^3 - 1 = 4095$

Perchè il sistema binario?

informazione:

elemento che consente di avere conoscenza di fatti e situazioni

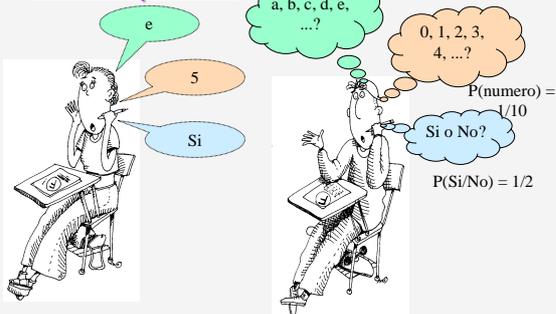
Scopo dell'informazione:

rimuovere uno stato di incertezza

Come quantificare l'informazione?

La quantità di informazione (**oggettivo**) deve essere indipendente dal contenuto (**soggettivo**) del messaggio.

Esempio 1

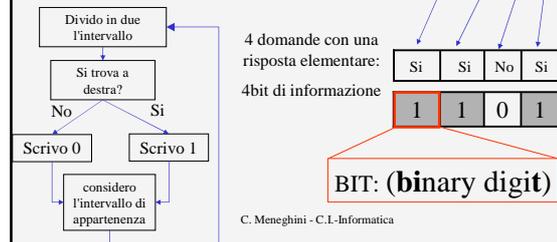


Più la scelta tra le possibili risposte è ampia più l'informazione è importante. La quantità di informazione è inversamente proporzionale alla probabilità di "indovinare" il contenuto del messaggio.

Esempio 2

Indovinare un numero scelto a caso tra 0 e 15

	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
1																
2																
3																
4																



Unità di misura dell'informazione

Il **bit**, introdotto da J. W. Tukey (1946), rappresenta la minima quantità di informazione; è quindi l'unità di misura naturale per quantificare l'informazione. Deriva da **binary digit** e rappresenta quanta informazione posso ottenere da una risposta ad una **domanda elementare**.

Multipli

byte = 8bit

1 KB = 2^{10} bytes = 1024 bytes ~ 10^3 bytes

1 MB = 2^{20} bytes = 1 048 576 bytes ~ 10^6 bytes

1 GB = 2^{30} bytes = 1 073 741 824 bytes ~ 10^9 bytes

1 TB = 2^{40} bytes = 1 099 511 627 776 bytes ~ 10^{12} bytes

11

Quantificare l'informazione

La quantità di informazione **Q** contenuta in un dato **s** è inversamente proporzionale alla probabilità **p_s** di ottenere quel dato:

$$Q(s) = \log_2 1/p_s = -\log_2 p_s$$

Quindi possiamo dire che la scelta tra **N** possibili "eventi" equiprobabili ci fornisce **$\log_2 N$ bit** di informazione che possono essere rappresentati con altrettanti simboli binari. Dall'esempio 1:

$$Q(E) = \log_2 26 = 4.7$$

$$Q(5) = \log_2 10 = 3.32$$

$$Q(Si) = \log_2 2 = 1$$

Quindi la quantità di informazione contenuta in un carattere alfabetico è maggiore di quella contenuta in una cifra decimale, che è maggiore di quella di una risposta "binaria"

C. Meneghini - C.I.-Informatica

12

Per un messaggio costituito di N simboli la quantità di informazione è espressa dalla formula di Shannon (1948) (entropia di Shannon):

$$Q(x_1, x_2, \dots, x_N) = K \left(p_{x_1} \log_2 \frac{1}{p_{x_1}} + p_{x_2} \log_2 \frac{1}{p_{x_2}} + \dots + p_{x_N} \log_2 \frac{1}{p_{x_N}} \right)$$

K= costante di normalizzazione che dipende dall'alfabeto usato

L'entropia di Shannon è massima se i simboli sono equiprobabili ed è nulla se uno dei simboli è certo

Un codice efficiente utilizza simboli con probabilità simili

C. Meneghini - C.I.-Informatica

13

Per approfondimenti: I. Bialynicki-Birula & I. Bialynicka-Birula *Modelling the reality* Oxford univ. press 2004

Notazione Binaria

In notazione **binaria** il numero 13 si esprime mediante la stringa:

1101

Usando il codice binario:

L'entropia di Shannon associata a questa stringa è:

$$Q_{1101}^{\text{bin}} = 4 \log_2 2 = 4$$

La stringa 1101 contiene 4 bit di informazione binaria

Perchè usare la notazione binaria?

C. Meneghini - C.I.-Informatica

14

Codice e rappresentazione

Decimale	Binario	ottale	esadecimale
00	0000	00	0
01	0001	01	1
02	0010	02	2
03	0011	03	3
04	0100	04	4
05	0101	05	5
06	0110	06	6
07	0111	07	7
08	1000	10	8
09	1001	11	9
10	1010	12	A
11	1011	13	B
12	1100	14	C
13	1101	15	D
14	1110	16	E
15	1111	17	F
16	10000	20	10

C. Meneghini - C.I.-

13_{dec} 1101_{bin} 15_{oct} D_{hex}

Rappresentano lo stesso numero

Attenzione: non confondere il dato con la sua rappresentazione!

I numeri sono entità matematiche che rappresentano quantità, non vanno confusi con la loro rappresentazione (che dipende dal codice)

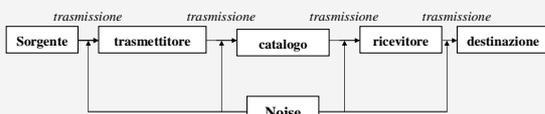
Il codice Binario è semplice da implementare

1	0
Si	No
Acceso	Spento
Aperto	Chiuso
Vero	Falso
Su	Giú
Sopra	Sotto
...	...

C. Meneghini - C.I.-Informatica

16

Informazione e comunicazione



Schematizzazione del processo di comunicazione (C.E. Shannon 40')

Un **messaggio** contiene una certa quantità di **informazione** che viene **trasmessa** da una sorgente ad un destinatario.

C. Meneghini - C.I.-Informatica

17

Integrità dell'informazione

00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15

10 Nel caso della rappresentazione e decimale l'incertezza può essere maggiore

1 ?

1 0 1

9

13

Se "perdo" una delle cifre nella rappresentazione binaria ho sempre due sole possibilità (perdo 1 bit di informazione)

