Lezioni 1, 2 e 3
"Nozioni di Base, Rappresentazione ed Elaborazione delle Informazioni"

Laboratorio di Informatica per l'Educazione

A. A. 2014/2015

17/04/15 Dott. Aniello Castiglione

Lezioni 1 e 2

Nozioni di Base e Rappresentazione delle Informazioni (Introduzione)

Il termine “Informatica”

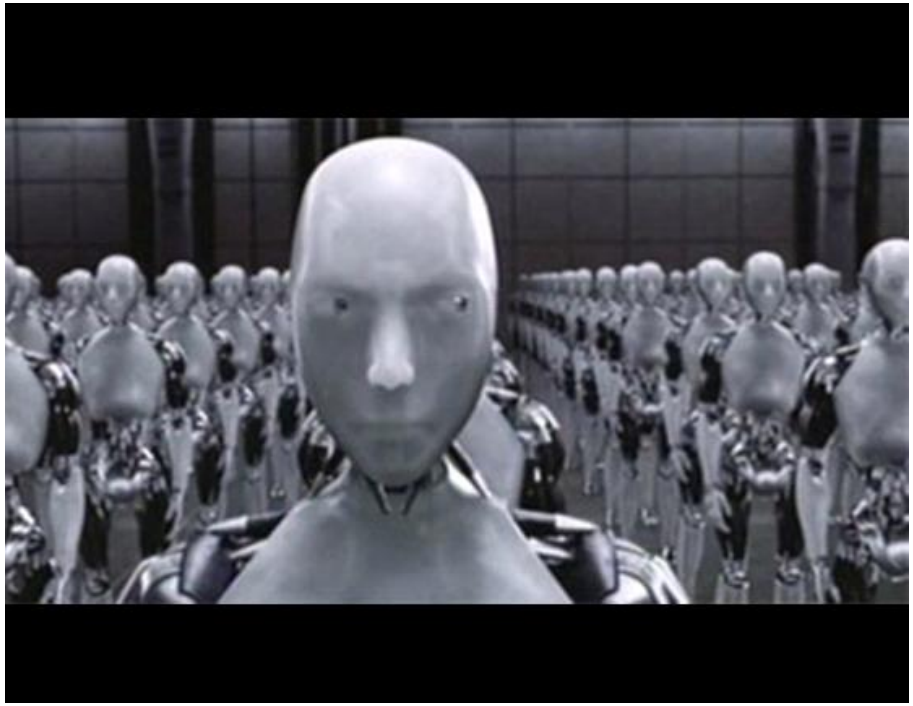
- Deriva dal francese *informatique* coniato negli anni ’60 da **Philippe Dreyfus**
- Ottenuto come contrazione dei termini “information” e “automatique”
- Usato per indicare la disciplina tecnico-scientifica che si occupa della progettazione e costruzione di macchine in grado di **trattare**, o **elaborare**, in **modo automatico** l’informazione

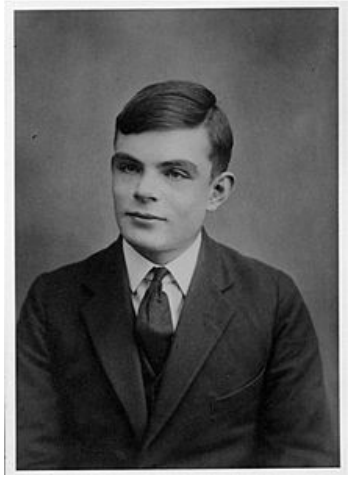
Trattamento dell'Informazione

- In generale, il trattamento dell'informazione è
  - l'elaborazione dell'informazione in qualunque modo rilevabile da un osservatore
    - stampa di un file di testo da parte di un sistema informatico (cambia il supporto)
    - eseguire operazioni aritmetiche (si trasformano i numeri)
    - archiviare foto digitali
    - ricerca su Google
    - scrivere un post sulla bacheca di Facebook
    - …

La realizzazione parziale di un sogno

- Disporre di **elaboratori automatici**, ovvero macchine in grado di compiere autonomamente attività di elaborazione dell'informazione
- Garantire dei benefici rispetto all'elaborazione condotta dall'uomo:
  - Riduzione dei tempi
  - Maggiore affidabilità
  - Incombenze "noiose"
  - Esecuzione automatica di operazioni che richiedono competenze poco comuni





Alan Mathison Turing (Londra, 23 giugno **1912** – Wilmslow, 7 giugno **1954**) è stato un matematico, logico e crittografo britannico, considerato uno dei padri dell'informatica e uno dei più grandi matematici del XX secolo.

Turing è solitamente considerato il padre della scienza informatica e dell'intelligenza artificiale per la sua formalizzazione dei concetti di **algoritmo** e calcolo mediante la **macchina di Turing**

I Calcolatori (Elaboratori) Elettronici (1/2)

- Svolgono un lavoro che per l'uomo è naturale, ma:
  - lo fanno rapidamente
  - sono infaticabili
  - riescono a gestire enormi quantità di dati
  - in alcuni casi permettono di superare vincoli temporali e spaziali
  - …
- A differenza degli elettrodomestici sono **programmabili** e possono essere adattati a risolvere qualsiasi tipo di problema che possa tradursi in un numero finito di operazioni

I Calcolatori (Elaboratori) Elettronici (2/2)

- L'informazione che ha attirato l'attenzione dell'uomo verso la realizzazione di elaboratori è quella espressa in **forma quantitativa** e, in particolare, numerica
- Per tale motivo si è diffuso il termine "**calcolatore**" (in inglese **computer**) che viene usato spesso in alternativa a "**elaboratore**"
- In inglese, **computer science** è la traduzione del nostro termine "informatica"

Information & Communication Technology (1/2)

- In alternativa al termine “informatica” viene utilizzato spesso il termine “tecnologia dell’informazione” (dall’inglese **Information Technology** o **IT**)
- Negli ultimi decenni le capacità dei singoli calcolatori sono notevolmente aumentate grazie alla loro interconnessione in **rete**
  - I calcolatori sono diventati degli strumenti di comunicazione

Information & Communication Technology (2/2)

- La convergenza tra la tecnologia dell'informazione e le **telecomunicazioni** ha dato vita ad una disciplina che prende il nome di **Information & Communication Technology (ICT)**

Il termine “Informazione” (1/2)

- Nei nostri discorsi:
  - Mezzi di informazione
  - Società dell’informazione
  - Tecnologie dell’informazione
  - Potere dell’informazione
  - Monopolio dell’informazione

- Nell’ambito dell’informatica il significato del termine “**informazione**” messo spesso in relazione con quello di “**dato**” e di “**conoscenza**”

Il termine “Informazione” (2/2)

La temperatura corporea di 39,5 di Tizio viene considerata un segnale d’allarme

La temperatura corporea di Tizio è 39,5 misurata in gradi centigradi

39,5

Dati, Informazioni e Conoscenza

- Dati
  - Insieme di **simboli** tracciati su supporto fisico e rappresentano una proprietà di un oggetto nel mondo reale ma non contengono alcun riferimento alla proprietà a cui si riferiscono
- Informazioni
  - Dati messi in relazione con la proprietà a cui essi si riferiscono
- Conoscenza
  - Regole che permettono di trarre vantaggio dalle informazioni

Ancora sui Dati

- Sono la materia prima del trattamento dell'informazione
- Possono essere classificati in:
  - dati **semplici**, come i numeri, i caratteri, etc.
  - dati **complessi**, come i film, i suoni, le immagini, etc. La gestione di questo tipo di dati è resa possibile dall'incredibile potenza raggiunta dagli elaboratori nell'ultimo decennio

Dati e Informazioni

- Il 50% (dato)
- E' stato registrato il 50% di assenze degli studenti di Scienze della Formazione nel giorno 12/10/2011 (informazione)

- 29 settembre 1936 (dato)
- Il 29 settembre 1936 è la data di nascita di Silvio Berlusconi (informazione)

Dati e Informazioni

Dato

Dati e Informazioni

Informazione

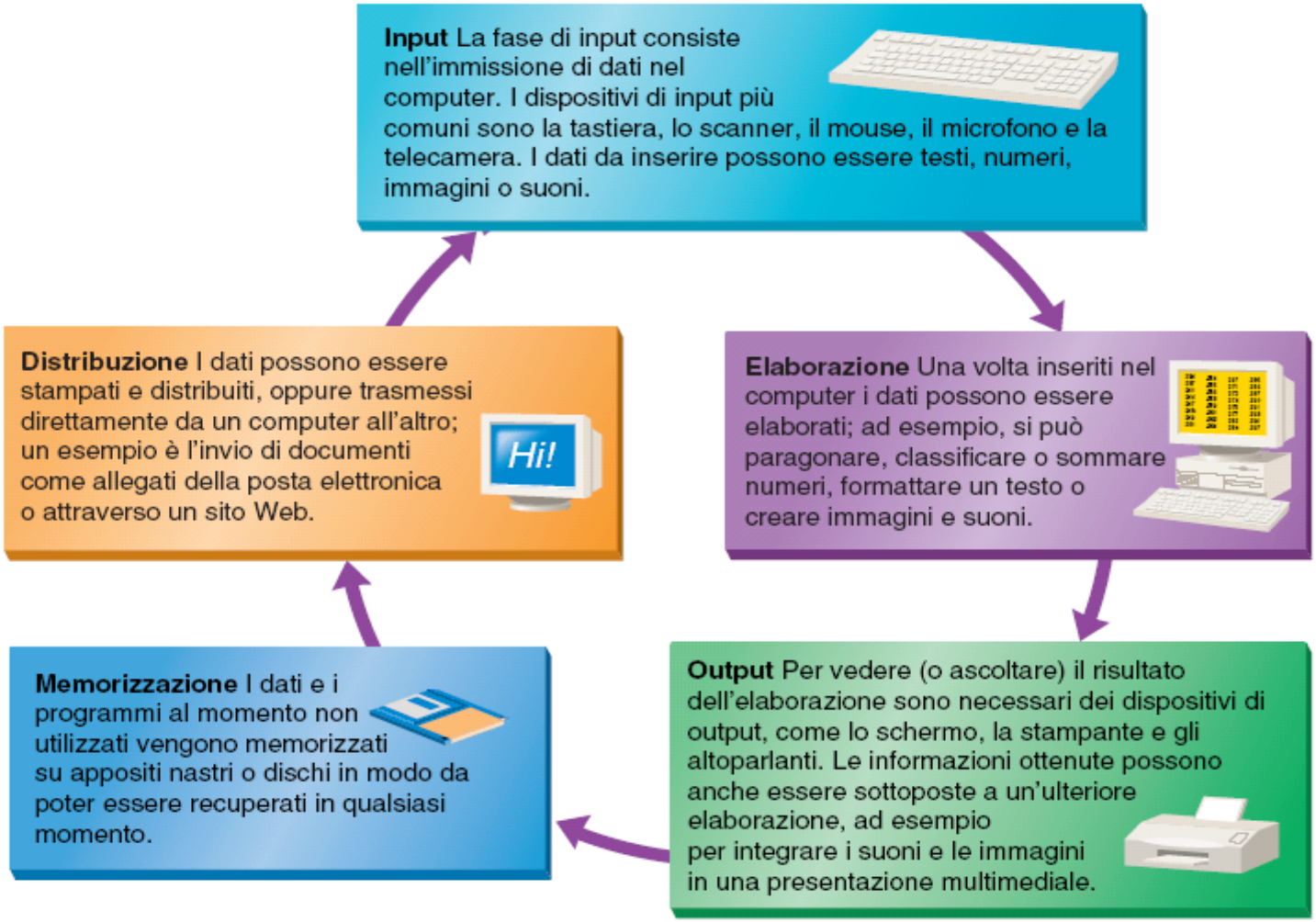
La Montagna Sacra
(Ayers Rock o Uluru in Australia)

Dati e Informazioni

Informazione

l'ultimo viaggio di Mario

Il Ciclo di Elaborazione dei Dati



Il ciclo di elaborazione dell'informazione

Ancora su Dati e Informazioni

- Per codificare le informazioni abbiamo bisogno necessariamente dei DATI e di una “logica” che permetta loro di diventare INFORMAZIONI
  - …
  - …

Perché codificare dati e informazioni?

Risoluzione dei Problemi basata sull'Informazione

- E' un approccio a cui ricorriamo senza nemmeno accorgercene
- Consente di compiere ragionamenti e di prendere decisioni sugli oggetti del mondo reale acquisendo solo un numero limitato di informazioni
- Tale approccio è alla base di tutte le discipline tecnico-scientifiche
  - Algebra, geometria
  - Fisica
  - …

Suddivisione delle caramelle

- Problema: *come distribuire, in egual numero, X caramelle a Y persone?*

- Si rappresentano le caramelle e le persone (oggetti del mondo reale) in forma simbolica
  - In questo caso i simboli scelti sono le cifre arabe e vengono utilizzati per rappresentare il numero **sessanta** ed il numero **quattro**: 60 e 4
- Sottoponiamo questi due dati ad una **elaborazione** o **trattamento**
  - Operazione di divisione: 60 / 4 = **15**
- Ripensiamo al simbolo ottenuto come alla rappresentazione dell'informazione cercata
  - la quantità di caramelle spettanti a ciascuno è **quindici**

In sintesi

- Con i COMPUTER, le macchine diventano strumenti per risolvere (o aiutare a risolvere) problemi basati sulle informazioni
  1. Abbiamo bisogno di codificare e memorizzare opportunamente dati e informazioni
  2. Abbiamo bisogno di impartire le giuste istruzioni (programmi o software) per risolvere correttamente i problemi dati
  3. Abbiamo bisogno di dotare queste macchine di una logica di funzionamento in grado di interpretare le nostre istruzioni ed eseguirle sui dati forniti

Problemi legati al linguaggio

- Linguaggio:
  - Alfabeto
    - è una collezione di simboli grafici, aventi di solito un ordine ben preciso, che servono a rappresentare le parole di una lingua
      - i simboli spesso (ma non sempre) sono lettere ovvero rappresentazione scritta di suoni linguistici
  - Vocabolario (o lessico)
    - è il complesso delle parole ammissibili di una lingua
  - Grammatica
    - è un insieme di regole utili alla corretta costruzione di frasi, sintagmi e parole (il termine si riferisce allo studio delle suddette regole)
  - Semantica
    - è quella parte della linguistica che studia il significato delle parole (semantica lessicale), degli insiemi delle parole, delle frasi (semantica frasale) e dei testi

LINGUAGGI

- I linguaggi (verbali, non verbali, orali, scritti, etc.) sono strumenti che contribuiscono a:
  - Rappresentare le informazioni
    - i concetti, i pensieri, le emozioni, etc. vengono formalizzati attraverso i linguaggi per poter essere memorizzati, trasferiti, elaborati
  - Memorizzare le informazioni
    - la scrittura
  - Trasferire le informazioni
    - la comunicazione
  - Elaborare le informazioni
    - le deduzioni nella logica

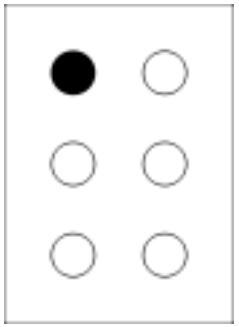
Problemi relativi alla Comunicazione

- Accordo sui simboli
  - a b c d e f g …
- Accordo sul lessico
  - casa, gatto, automobile, vado, …
- Accordo sulla grammatica
  - *<soggetto verbo complemento>*
- Accordo sulla semantica
  - La nonna chiude la porta (ok)
  - La porta chiude la nonna (no)

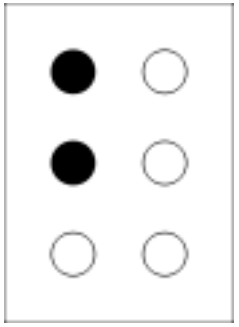
- Accordo sulla codifica (regole per trasformare simboli, parole e frasi di un linguaggio in una nuova rappresentazione con possibilità di effettuare in maniera corretta anche l'operazione inversa)
  - "a" in codice Morse (Samuel Morse, pittore e storico inglese) è " . — "
  - "b" in codice Morse è " — . . . "

Codice Braille

- Lettera “a”



- Lettera “b”



Codifica dei Numeri

- Linguaggio di partenza:
  - i numeri
- Codifica 1
  - numerazione decimale
    - 5, 45, 670
- Codifica 2
  - numerazione binaria
    - 101, …
- Codifica 3

I Linguaggi Naturali: Ambiguità

- Per comunicare tra loro gli uomini hanno sviluppato i **linguaggi naturali**: italiano, inglese, francese, …
- Una caratteristica negativa di tali linguaggi è la loro inerente ambiguità:
  - Una qualsiasi frase formulata è potenzialmente **polisemica**
  - Il significato che viene dato alla frase da chi riceve il messaggio può essere diverso da quello datogli dal mittente

Un esempio: "Sposta il Cavallo"

I Linguaggi Naturali: Disambiguazione

- Comprendere il significato di una frase in linguaggio naturale è più semplice se questa viene pronunciata in un dato **contesto**:
  - Concorso ippico
  - Circolo Scacchistico
  - Sartoria
  - Palestra

I Linguaggi Naturali nella Comunicazione con i Calcolatori

- Per comunicare con un calcolatore, l'ambiguità dei linguaggi naturali rappresenta una grosso problema
- Bisognerebbe fornire al calcolatore anche il contesto e la conoscenza (le regole) per disambiguare le frasi
  - La rappresentazione (comprensibile al calcolatore) di contesto e regole è molto complicata
  - In alcuni casi, la disambiguazione è difficile anche per gli uomini

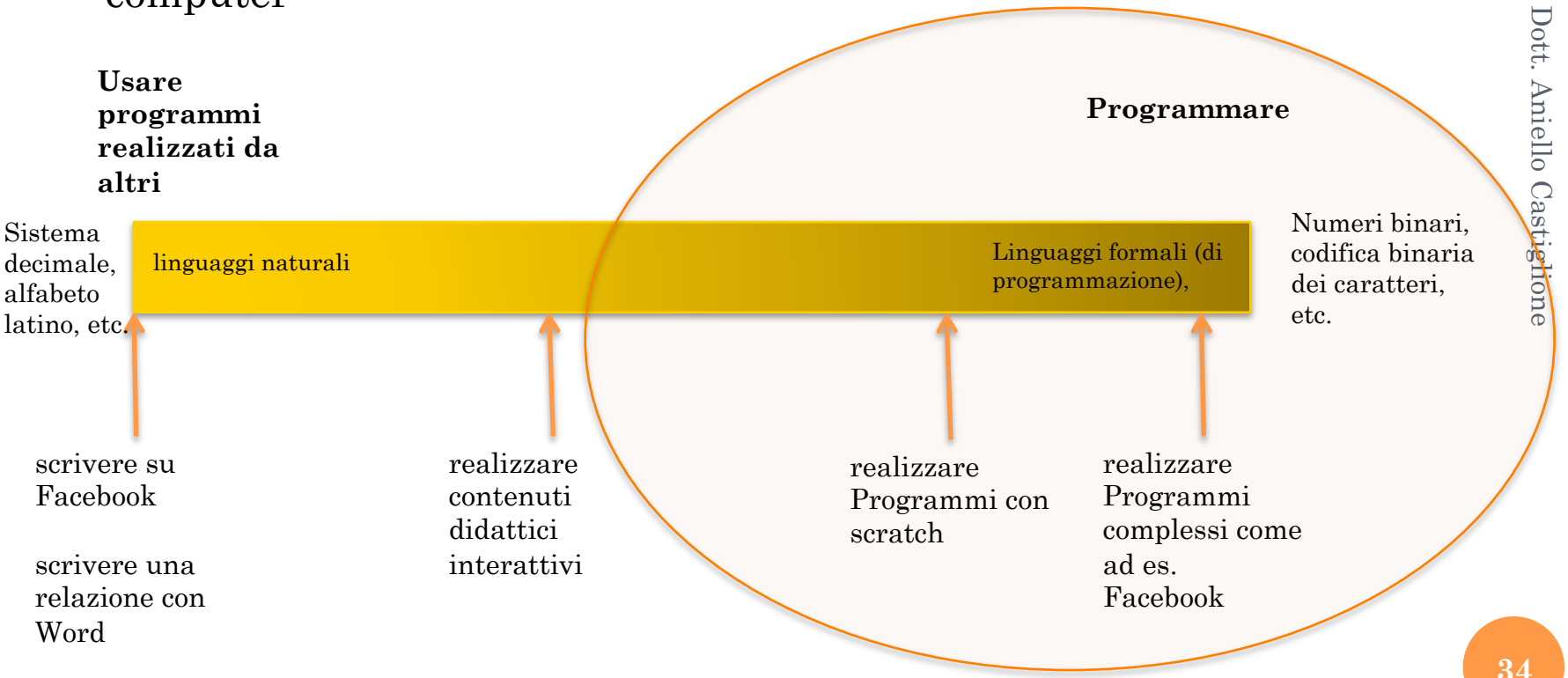
Tizio vide Caio nel parco con il cannocchiale

I Linguaggi Formali

- Vengono sviluppati e sono impiegati in tutti gli ambiti, come l'Informatica, la Matematica e la Logica, in cui è importante **evitare l'ambiguità**
- La definizione di un Linguaggio Formale prevede:
  - L'individuazione di un **alfabeto**, ovvero, di un elenco (finito) di simboli.
  - La definizione di una **grammatica formale**, ovvero un insieme di regole sintattiche che specificano come i simboli dell'alfabeto possono essere combinati tra loro per costituire le frasi ben formate all'interno del linguaggio stesso.
- Inoltre, le **semantiche formali** consentono di attribuire un significato non ambiguo alle frasi in un linguaggio formale

Usare e Programmare il Computer

- I Programmi (o software) risolvono problemi specifici con approccio basato sulle informazioni e vengono eseguiti dai computer



I due livelli delle tecnologie informatiche (1/2)

"L'hardware è la parte del computer che puoi prendere a calci; il software quella contro cui puoi solo imprecare."

"L'hardware è un attore ed il software una sceneggiatura: l'hardware esegue operazioni diverse a seconda del tipo di software usato."

I due livelli delle tecnologie informatiche (2/2)

- Il termine **hardware** indica la struttura fisica dei dispositivi (o macchine informatiche), costituita di norma da componenti elettronici (computer, monitor, stampante, cavi telefonici, antenne, etc.) che svolgono specifiche funzioni nel trattamento e nella trasmissione delle informazioni
- Il termine **software** indica il livello logico (in contrapposizione con quello fisico dell'hardware), cioè l'insieme delle *istruzioni* che consentono all'hardware di svolgere i propri compiti

In sintesi (Rappresentazione)

- Con i COMPUTER, le macchine diventano strumenti per risolvere problemi basati sulle informazioni
  1. **Abbiamo bisogno di codificare e memorizzare opportunamente dati e informazioni**
  2. Abbiamo bisogno di impartire le giuste istruzioni (programmi o software) per risolvere correttamente i problemi dati
  3. Abbiamo bisogno di dotare queste macchine di una logica di funzionamento in grado di interpretare le nostre istruzioni ed eseguirle sui dati forniti

Rappresentazione dell'Informazione

- L'informazione è rappresentata dai dati che, a loro volta, sono espressi in forma di simboli

- La stessa informazione può essere codificata con simboli e modalità diverse:
  - 1963 -> simboli "0", "1", "2", …
  - MCMLXIII -> simboli della codifica romana
  - Millenovecentosessantatre -> rappresentazione testuale
  - …

Rappresentazione dell'Informazione nei Calcolatori (1/2)

- Consideriamo un alfabeto ridotto che contiene solo i simboli "0" e "1"
- Un **bit** (contrazione di **binary digit**) è un simbolo scelto sull'alfabeto {0, 1}
- Nei calcolatori ogni elemento (numeri, testo, audio, video, istruzioni …) viene rappresentato esclusivamente con sequenze di bit
- I dati (ma anche le istruzioni) vengono codificati con sequenze di bit

Rappresentazione dell'Informazione nei Calcolatori (2/2)

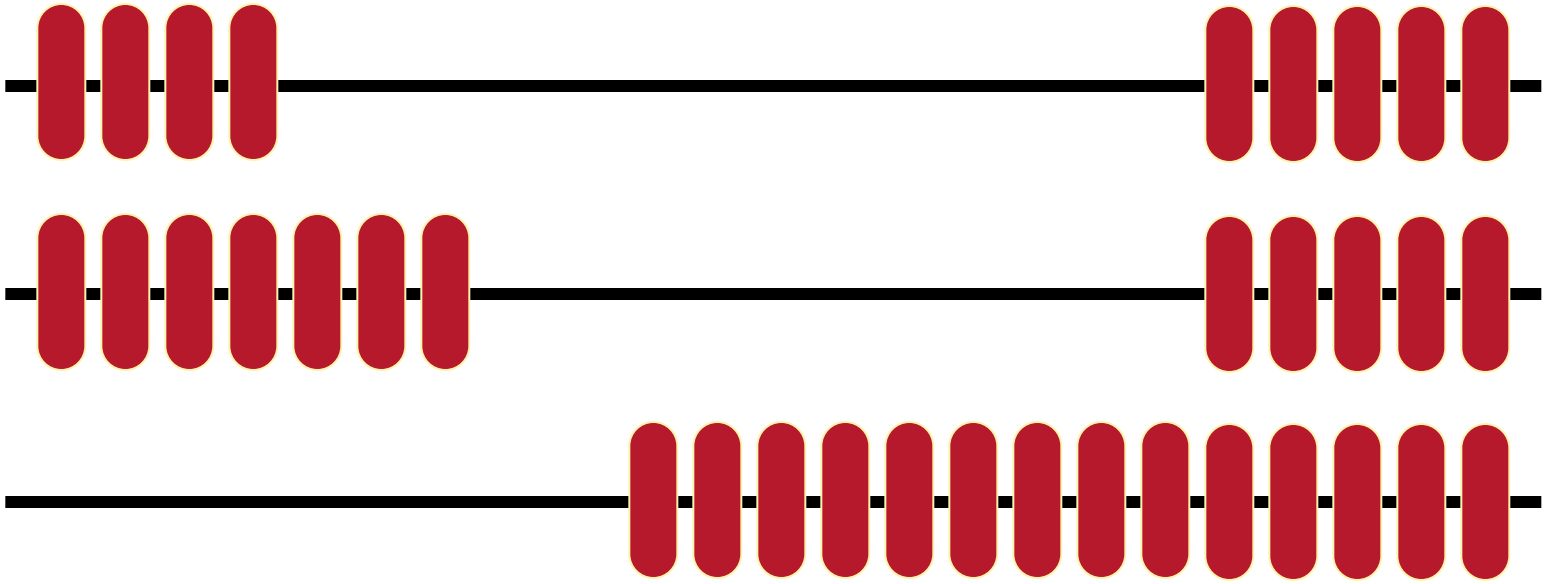
- Per elaborare, trasmettere e memorizzare l'informazione abbiamo bisogno di codificarla:
  - Livello alto
    - Linguaggio (numeri)
  - Livello basso
    - Alfabeto ({0,1} oppure {0,…, 9} oppure …)
    - Regole di codifica (binaria o decimale o …)
    - Supporto/Tecnologia (transistor di un computer o foglio di carta o …)
    - …
- La **codifica** dell'informazione è l'operazione con la quale l'informazione viene scritta su un supporto fisico
- La **decodifica** è, invece, l'operazione con la quale l'informazione viene letta dal supporto fisico

Perché codificare l'informazione

- Per trattare (memorizzare, elaborare, trasmettere, etc.) l'informazione:
  - Si codifica l'informazione disponibile scrivendo su un supporto fisico
  - Si manipola il supporto con opportune trasformazioni fisiche, ottenendo una versione modificata del supporto
  - Si decodifica l'informazione corrispondente al risultato dell'elaborazione leggendo dalla versione modificata del supporto

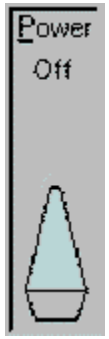
Vi ricordate la suddivisione delle caramelle?

Perché codificare l'informazione: il pallottoliere



Un esempio semplice: l'interruttore

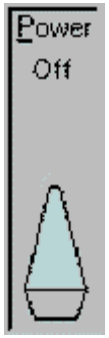
- Due sole possibilità:
  - Spento
  - Acceso
- L'informazione sullo stato dell'interruttore corrisponde dunque alla scelta fra due sole alternative



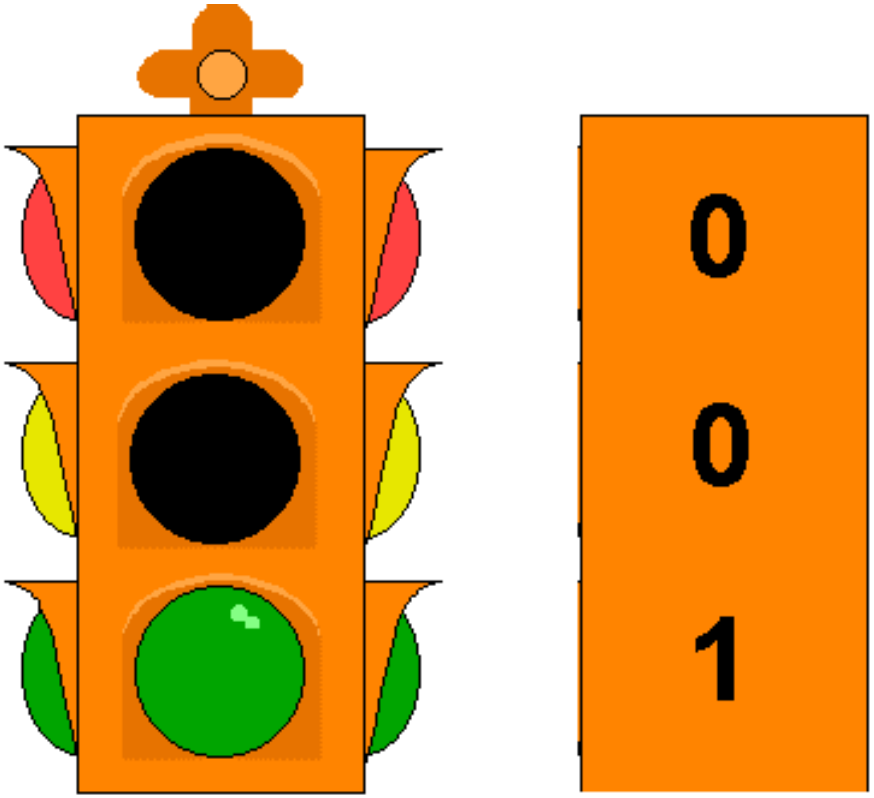
Codifica dell'Informazione in binario

- 1 bit rappresenta lo stato dell'interruttore

- Interruttore acceso: 1
- Interruttore spento: 0



Codifica di un'Informazione con più di due stati: Il Semaforo



Codifica di un'Informazione con più di due stati

- Nel nostro esempio, abbiamo deciso di utilizzare un bit per rappresentare lo stato di ciascuna delle lampadine del semaforo
- In questo modo, con 3 bit potremmo rappresentare tutti gli stati possibili del semaforo
- Nel caso del semaforo, le alternative disponibili sono comunque poche

Diverse Codifiche per le stesse informazioni

- Rappresentazione degli stati di un semaforo

3 stati

Stato	codifica
ROSSO	1 0 0
VERDE	0 0 1
GIALLO	0 1 0

2 stati

Stato	codifica
ROSSO	0 0
VERDE	0 1
GIALLO	1 0

Combinazioni di bit

Bit a disposizione	Le combinazioni	Il numero di combinazioni
1	0, 1	$2 = 2^1$
2	00, 01, 10, 11	$4 = 2^2$
3	000, 001, 010, 011, 100, 101, 110, 111	$8 = 2^3$
4	0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111,1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111	$16 = 2^4$
5	00000, 00001, 00010, …	$32 = 2^5$
…	…	…

Codifica dei caratteri

- Si, ma …
  - … è possibile applicare queste idee alla rappresentazione di informazione più complessa, ad esempio di un testo?
  - Un testo è rappresentato attraverso una successione di caratteri
  - Ogni carattere viene scelto all'interno di un insieme finito di simboli (alfabeto)

Codifica dei caratteri

- Con 8 bit, è possibile rappresentare la scelta fra 256 alternative diverse (2^8=256)
  - da 00000000… a 11111111
  - passando per tutte le combinazioni intermedie (00000001, 00000010, …)

Codifica dei caratteri

- Nel caso del testo, possiamo far corrispondere diverse combinazioni di 8 bit (otto cellette, ciascuna delle quali può contenere 0 o 1) a caratteri diversi

Ogni singolo CARATTERE viene codificato con una combinazione di 8 bit

Codifica dei caratteri

- Ad esempio:
  - 00000000 -> A
  - 00000001 -> B
  - 00000010 -> C
  - 00000011 -> D
  - 00000100 -> E
- …. e così via

Byte	Cod.	Char	Byte	Cod.	Char	Byte	Cod.	Char	Byte	Cod.	Char
00000000	0	Null	00100000	32	Spc	01000000	64	@	01100000	96	`
00000001	1	Start of heading	00100001	33	!	01000001	65	A	01100001	97	a
00000010	2	Start of text	00100010	34	"	01000010	66	B	01100010	98	b
00000011	3	End of text	00100011	35	#	01000011	67	C	01100011	99	c
00000100	4	End of transmit	00100100	36	$	01000100	68	D	01100100	100	d
00000101	5	Enquiry	00100101	37	%	01000101	69	E	01100101	101	e
00000110	6	Acknowledge	00100110	38	&	01000110	70	F	01100110	102	f
00000111	7	Audible bell	00100111	39	'	01000111	71	G	01100111	103	g
00001000	8	Backspace	00101000	40	(	01001000	72	H	01101000	104	h
00001001	9	Horizontal tab	00101001	41	)	01001001	73	I	01101001	105	i
00001010	10	Line feed	00101010	42	*	01001010	74	J	01101010	106	j
00001011	11	Vertical tab	00101011	43	+	01001011	75	K	01101011	107	k
00001100	12	Form Feed	00101100	44	,	01001100	76	L	01101100	108	l
00001101	13	Carriage return	00101101	45	-	01001101	77	M	01101101	109	m
00001110	14	Shift out	00101110	46	.	01001110	78	N	01101110	110	n
00001111	15	Shift in	00101111	47	/	01001111	79	O	01101111	111	o
00010000	16	Data link escape	00110000	48	0	01010000	80	P	01110000	112	p
00010001	17	Device control 1	00110001	49	1	01010001	81	Q	01110001	113	q
00010010	18	Device control 2	00110010	50	2	01010010	82	R	01110010	114	r
00010011	19	Device control 3	00110011	51	3	01010011	83	S	01110011	115	s
00010100	20	Device control 4	00110100	52	4	01010100	84	T	01110100	116	t
00010101	21	Neg. acknowledge	00110101	53	5	01010101	85	U	01110101	117	u
00010110	22	Synchronous idle	00110110	54	6	01010110	86	V	01110110	118	v
00010111	23	End trans. block	00110111	55	7	01010111	87	W	01110111	119	w
00011000	24	Cancel	00111000	56	8	01011000	88	X	01111000	120	x
00011001	25	End of medium	00111001	57	9	01011001	89	Y	01111001	121	y
00011010	26	Substitution	00111010	58	:	01011010	90	Z	01111010	122	z
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 dei caratteri

- Il risultato? Una parola (o più parole) sarà rappresentata dal computer come una successione di gruppi di 8 bit

O	G	G	I		P	I	O	V	E
01001111	01000111	01000111	01001001	00100000	01010000	01001001	01001111	01010110	01000101

Byte e Memorie

- Si definisce **byte** una sequenza di 8 bit
- La **memoria**
  - è la componente del calcolatore che ha lo scopo di memorizzare i dati nel tempo
  - La **memoria** è costituita da una collezione di dispositivi bistabili
- La **capacità**
  - è la quantità di dati che la memoria è in grado di memorizzare
  - è misurata in bit, byte o in multipli di essi

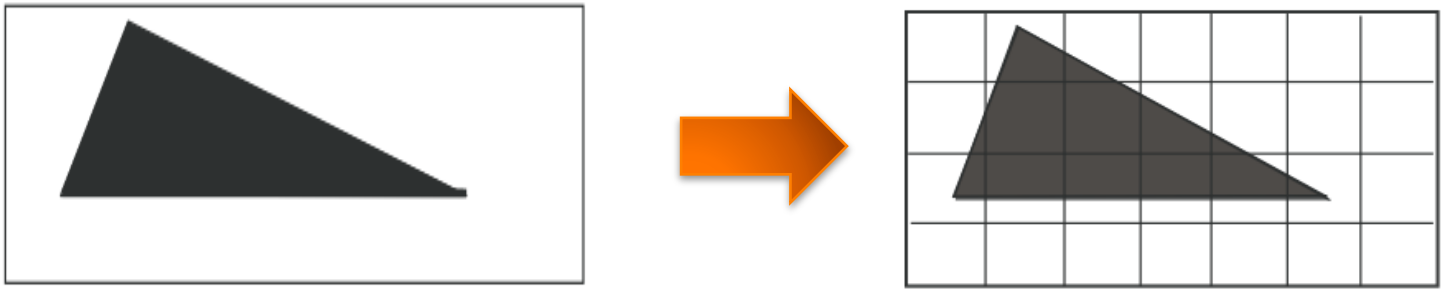
Analogia con la misura della Lunghezza

- L'unità di misura della lunghezza è il **metro**
- Il Sistema Internazionale (SI) definisce dei prefissi che rappresentano i multipli del **metro**
- **Kilo** è il prefisso che i metri per 10^3 (1.000)
- Due chilometri, ad esempio, equivalgono a 2.000 metri
- Gli stessi prefissi vengono utilizzati per i multipli del bit e del byte con la differenza che la base utilizzata è **2**
  - 1 Kilobit (Kb) corrisponde a 2^{10} bit
  - 1 Kilobyte (KB) corrisponde a 2^{10} byte

Multipli del bit e del byte

Prefisso	Sistema Binario	SI
Kilo	2^{10} (= 1.024)	10^{3} (= 1.000)
Mega	2^{20} (= 1.024 * 1.024)	10^{6} (= 1.000 * 1.000)
Giga	2^{30}	10^{9}
Tera	2^{40}	10^{12}
Peta	2^{50}	10^{15}
Exa	2^{60}	10^{18}
Zetta	2^{70}	10^{21}
Yotta	2^{80}	10^{24}

Rappresentazione e Codifica dei Dati Complessi: Le Immagini (1/3)



Rappresentazione Binaria:
- Alfabeto {0, 1}

Codifica:
-1 se è predominante il nero
-0 se è predominante il bianco

0_{22}	1_{23}	0_{24}	0_{25}	0_{26}	0_{27}	0_{28}
0_{15}	1_{16}	1_{17}	0_{18}	0_{19}	0_{20}	0_{21}
0_{8}	1_{9}	1_{10}	1_{11}	1_{12}	0_{13}	0_{14}
0_{1}	0_{2}	0_{3}	0_{4}	0_{5}	0_{6}	0_{7}

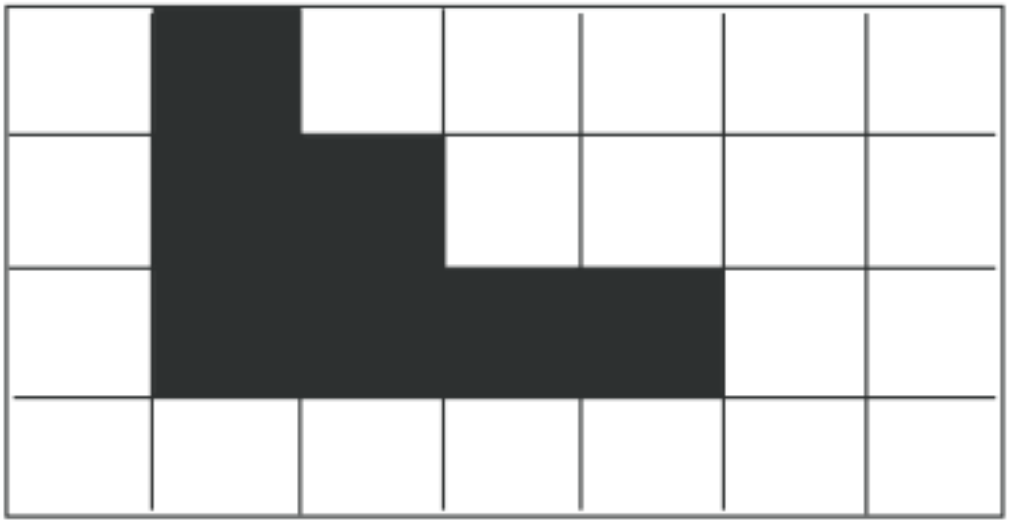
Rappresentazione:

0000000 0111100 0110000 0100000

Rappresentazione e Codifica dei Dati Complessi: Le Immagini (2/3)

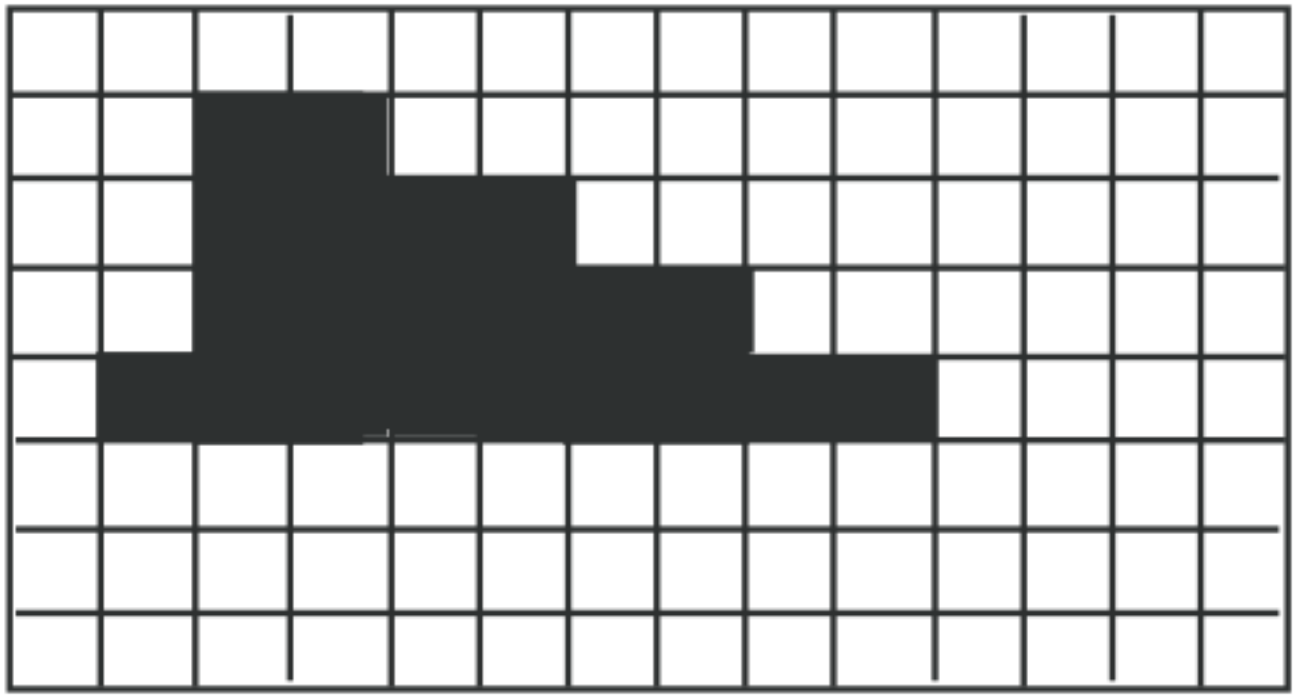
0000000 0111100 0110000 0100000

decodifica



Rappresentazione e Codifica dei Dati Complessi: Le Immagini (3/3)

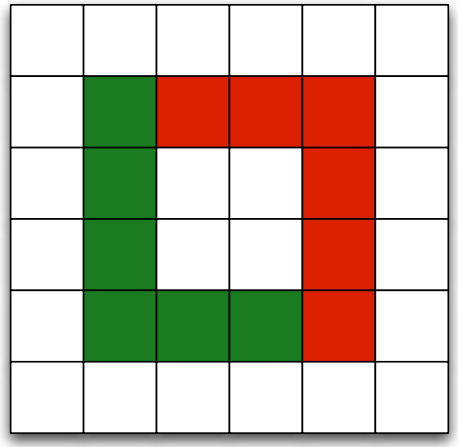
- Ottengo una rappresentazione più fedele se aumento il numero dei pixel:



Considerazioni

- La perdita di informazione è accettabile nella rappresentazione delle immagini (in alcuni casi) ma non è accettabile nella codifica dei numeri, dei testi, etc. (in questi casi l'informazione non deve essere soggetta ad errori e abbiamo bisogno di codifiche diverse

… E LA CODIFICA DELLE IMMAGINI A COLORI?



Codifica

Nero – 00
Bianco – 01
Rosso – 10
Verde – 11

01	01	01	01	01	01
01	11	10	10	10	01
01	11	01	01	10	01
01	11	01	01	10	01
01	11	11	11	10	01
01	01	01	01	01	01

010101010101 011110101001 011101011001 011101011001 011111111001 010101010101

Definizione di Computer

- Il computer è un **elaboratore elettronico digitale**:
  - **elaboratore**: macchina in grado di rappresentare ed elaborare dati in base ad una serie di istruzioni, ovvero un software, formulate e memorizzate in modo tale da poter essere eseguite automaticamente
  - **elettronico**: indica che il computer utilizza componenti elettronici
  - **digitale**: indica che il computer elabora e memorizza informazioni rappresentate mediante i due simboli (digit) della numerazione binaria: **0** e **1**. Con queste due cifre, usate in combinazioni diverse, si possono rappresentare tutti i dati (e.g. parole, numeri, immagini, filmati, etc.)

La velocità delle operazioni elettroniche

- Tutte le operazioni che un computer può operare sui bit (riconoscimento, conversione, trasferimento e confronto) avvengono mediante il flusso o l'inibizione della corrente elettrica all'interno dei circuiti nei *chip*
- La velocità della corrente elettrica è di circa 300.000 km al secondo per cui anche le operazioni sui bit sono dello stesso ordine di grandezza
- Le unità di tempo per le operazioni dei bit sono le seguenti:
  - Millisecondo – 1 millesimo di secondo
  - Microsecondo – 1 milionesimo di secondo
  - **Nanosecondo – 1 miliardesimo di secondo**
  - **Picosecondo – 1 millesimo di miliardesimo di secondo**

Lezione 3

Rappresentazione (Codifica Binaria di Numeri e Caratteri) ed Elaborazione delle Informazioni

La Codifica dei Numeri

- Obiettivo:
  - Codifica in binario dei numeri per **favorire l'elaborazione** da parte dei calcolatori
- Vincoli:
  - Codifica e decodifica devono essere definite in maniera tale da poter essere **compiute in maniera automatica** (algoritmo)
- Problema:
  - Deve essere possibile codificare tutti i numeri
    - 0, 1, 2, 3, …
    - -1, -2, -3, …
    - -12.4, -2.004, 0.56, 134.89, …
  - …in sequenze
    - 0000000, 000001, 000010, …

L'associazione tra un numero **decimale** e la sua codifica **binaria** deve essere univoca in maniera tale da poter effettuare la decodifica in maniera coerente (principio di non ambiguità)

Sistemi di Numerazione (1)

- Il nostro sistema di numerazione:
  - Utilizza una notazione **posizionale** ed è in base 10
  - L'alfabeto utilizzato è l'insieme dei simboli {0, 1, 2, …, 9}
  - Non è l'unico sistema possibile
- Essendo posizionale, il valore di una "sequenza" di simboli viene calcolata assegnando dei "pesi" ad ogni simbolo a seconda della sua posizione

Posizioni → 3 2 1 0

base

Stringa di simboli → $4523_{10} =$

$$4*10^3 + 5*10^2 + 2*10^1 + 3*10^0$$

migliaia centinaia decine unità

Rappresentazione dei numeri: Sistema Decimale

- 3251
  - 1 unità , 5 decine, 2 centinaia , 3 unità di migliaia
- 745814763
  - 3 unità, 6 decine, 7 centinaia, 4 unità di migliaia, 1 decina di migliaia, 8 centinaia di migliaia, 5 unità di milioni, 4 decine di milioni, 7 centinaia di milioni

Come conta "ET"

- Un Extra-Terrestre viene sulla Terra e ci dice che i re di Roma sono 13. Quante dita ha l'Extra-Terrestre?
  - Il 13 deve essere interpretato come una stringa di simboli
  - Non conosciamo la base della loro numerazione
  - Sappiamo che il loro sistema di numerazione è POSIZIONALE
  - Sappiamo che in decimale i re di Roma sono 7

$13_x = 1 * X^1 + 3 * X^0 = X + 3 = 7_{10}$

$\Rightarrow X = 7 - 3 = 4$

$\Rightarrow$ L'Extra-Terrestre conta in base 4 per cui (sfruttando l'esperienza del sistema decimale) possiamo dire che ha 4 dita (2 per mano)

$\Rightarrow$ L'Extra-Terrestre usa l'alfabeto {0, 1, 2, 3}

Sistemi di Numerazione Posizionale

- I re di Roma sono 7_{10}
  - Base 10, simboli {0, …, 9}

$$7_{10} = 7 * 10^0 = 7_{10}$$

- I re di Roma sono 13_4
  - Base 4, simboli {0, 1, 2, 3}

$$13_4 = 1 * 4^1 + 3 * 4^0 = 7_{10}$$

- I re di Roma sono 111_2
  - Base 2, simboli {0, 1}

$$111_2 = 1 * 2^2 + 1 * 2^1 + 1 * 2^0 = 7_{10}$$

Codifica e decodifica

(Da binario a decimale e viceversa)

$N_2 = 101010$

$N_{10} = 1 \times \boxed{2^5} + 0 \times 2^4 + 1 \times \boxed{2^3} + 0 \times 2^2 + 1 \times \boxed{2^1} + 0 \times 2^0$
$= 32 + 8 + 2 = 42$

$N_2 = 11011$

$N_{10} = 2^0 + 2^1 + 2^3 + 2^4 = 1 + 2 + 8 + 16 = 27$

Codifica e decodifica

(Da decimale a binario)

$N_{10} = 51$ $N_2 = ???$

51	2					
1	25	2				
	1	12	2			
		0	6	2		
			0	3	2	
				1	1	2
					1	0

$N_2 = 110011$

$$51 = 1*2^5+1*2^4+0*2^3+0*2^2+1*2^1+1*2^0$$

Limiti sui numeri rappresentabili: codifica binaria semplice

- 4 bit a disposizione
  - Possiamo rappresentare da 0000 a 1111, in decimale da 0 a 15
- 5 bit a disposizione
  - Possiamo rappresentare da 00000 a 11111, in decimale da 0 a 31
- …

- Con **k** bit a disposizione possiamo rappresentare numeri da **0** a $\mathbf{2^k - 1}$

Rappresentazione degli Interi: "Modulo e Segno"

N = 0,+1,-1,+2,-2,+3,-3,…

Come possiamo rappresentare il *segno* di un numero?

Aggiungiamo un ulteriore bit che poniamo a 1 se il numero è negativo!

Esempio

N_{10} = +14 N_2 = 01110

N_{10} = -14 N_2 = 11110

Limiti sui numeri rappresentabili: codifica binaria modulo e segno

- 4 bit a disposizione
  - Possiamo rappresentare da 0000 a 0111 e da 1000 a 1111, in decimale da 0 a 7 e da 0 a -7
- 5 bit a disposizione
  - Possiamo rappresentare da 00000 a 01111 e da 10000 a 11111, in decimale da 0 a 15 e da 0 a -15
- …

Con **k** bit a disposizione possiamo rappresentare numeri da **0** a $\mathbf{2^{k-1} - 1}$ e da $\mathbf{-(2^{k-1} - 1)}$ **a 0**

Esercizi

- Scrivere in binario semplice su 7 bit il numero 13_{10}
- Scrivere in modulo e segno su 7 bit il numero 13_{10}
- Scrivere in modulo e segno su 7 bit il numero -13_{10}
- Scrivere in modulo e segno su 5 bit il numero 17_{10}

Soluzione Esercizi

- Scrivere in binario semplice su 7 bit il numero 13_{10}
  - 0001101
- Scrivere in modulo e segno su 7 bit il numero 13_{10}
  - 0001101
- Scrivere in modulo e segno su 7 bit il numero -13_{10}
  - 1001101
- Scrivere in modulo e segno su 5 bit il numero 17_{10}
  - 10001, In modulo e segno è -1_{10}
  - RISPOSTA: Non è possibile. Ho bisogno di almeno 6 bit (010001)

Esercizi da svolgere …

- Scrivere in binario semplice su 7 bit il numero 11_{10}
- Scrivere in modulo e segno su 8 bit il numero 25_{10}
- Scrivere in modulo e segno su 7 bit il numero -12_{10}
- Scrivere in modulo e segno su 5 bit il numero 20_{10}

Codifica Binaria dei Caratteri (1/2)

- **ASCII** (**American Standard Code for Information Interchange**): ogni carattere è rappresentato da una sequenza di 7 bit (128 caratteri diversi possibili)
- **ASCII esteso**: ogni carattere è rappresentato da un byte (256 caratteri diversi possibili, comprende anche lettere greche, accentate …)
- **EBCDIC** (**Extended Binary Coded Decimal Interchange Code**): sviluppato ed impiegato da I.B.M., ogni carattere è rappresentato da un byte
- **UNICODE** (consorzio di produttori di HW e SW): ogni carattere è rappresentato da 16 (65.536 caratteri diversi possibili) o 21 bit (più di 2 milioni di caratteri)
  - Anche per Ideogrammi, Braille, Simboli Matematici, Simboli Chimici, etc.

In sintesi (Elaborazione)

- Con i COMPUTER, le macchine diventano strumenti per risolvere problemi basati sulle informazioni
  1. Abbiamo bisogno di codificare e memorizzare opportunamente dati e informazioni
  2. Abbiamo bisogno di impartire le giuste istruzioni (programmi o software) per risolvere correttamente i problemi dati
  3. **Abbiamo bisogno di dotare queste macchine di una logica di funzionamento in grado di interpretare le nostre istruzioni ed eseguirle sui dati forniti**

L'Architettura di Von Neumann: Cenni Storici

- Nel 1944 il matematico di origine ungherese John Von Neumann (1903-1956) iniziò a lavorare sul progetto ENIAC e poi sul progetto EDVAC
- Il suo compito era di stendere una bozza preliminare dell'architettura del nuovo calcolatore mettendo insieme anche le idee di Mauchly ed Eckert
- L'architettura definita prese il nome di Architettura di Von Neumann e divenne il riferimento per la **quasi totalità dei calcolatori** costruiti da allora ad oggi

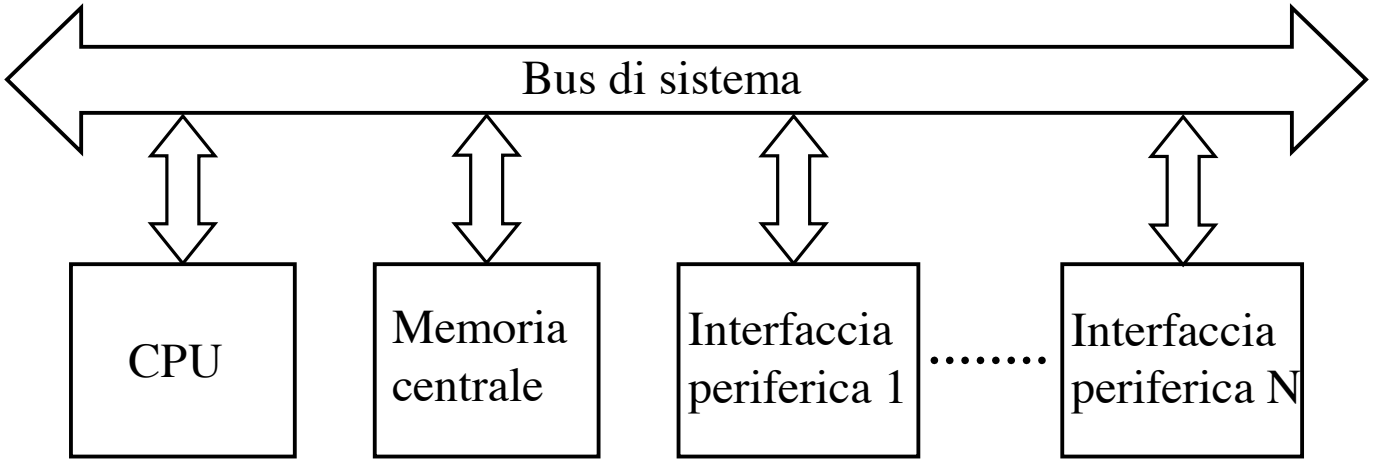
L'Architettura di Von Neumann: Caratteristiche Essenziali

- **Obiettivo**: realizzazione di un calcolatore universale (*general purpose*)

- Presenza di un dispositivo di memorizzazione in cui è possibile rappresentare sia **dati** che **istruzioni**
- Utilizzo dell'**aritmetica binaria** invece che quella decimale
- Separazione netta tra dispositivo di memorizzazione e dispositivo di elaborazione

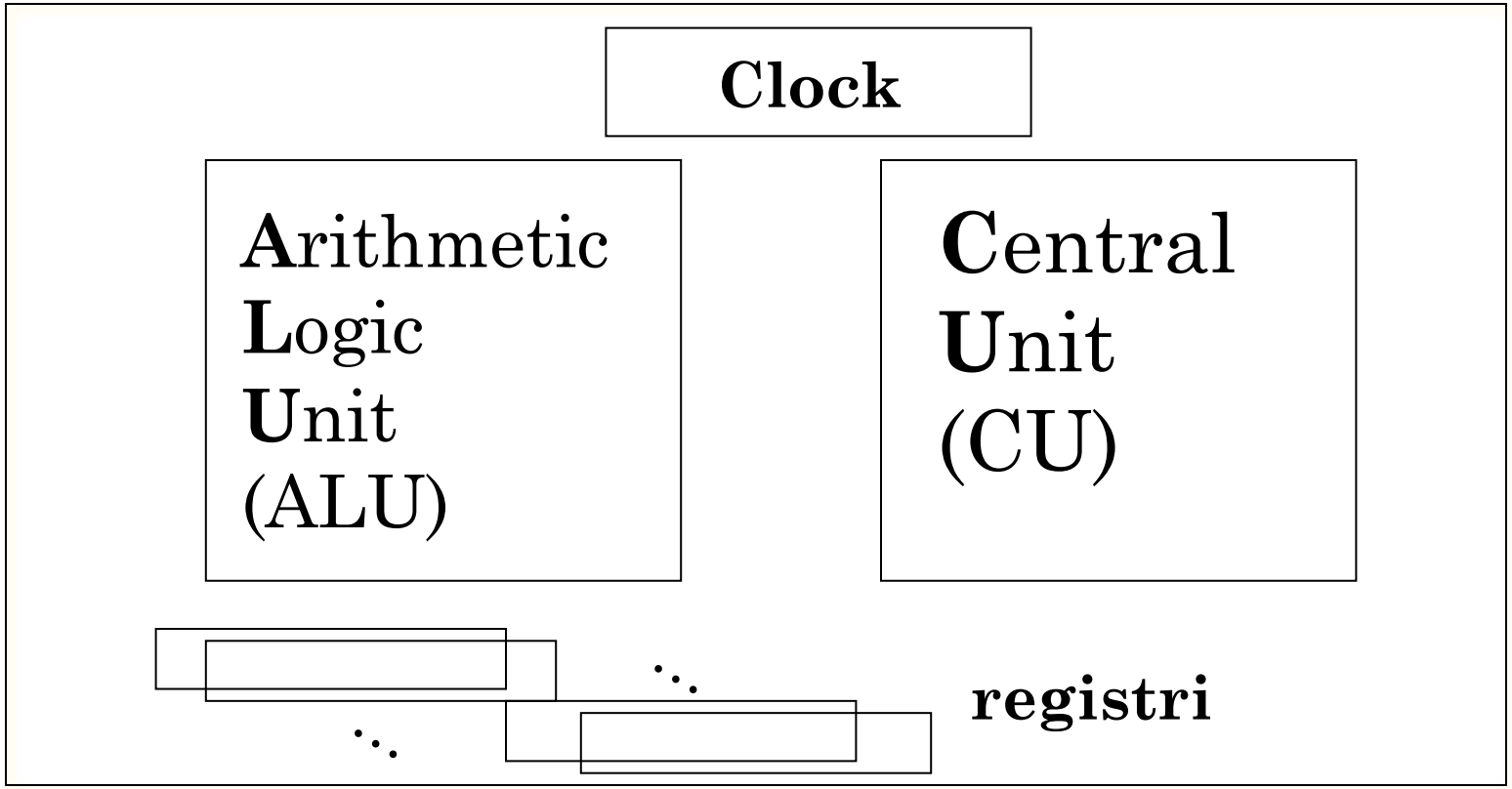
Componenti della Macchina di Von Neumann (1/2)

- E' il modello secondo il quale è organizzata la maggior parte dei moderni elaboratori



- L'unità centrale di elaborazione o **CPU** è costituita dai circuiti elettronici capaci di **leggere** (dalla memoria centrale), **decodificare** (interpretare) ed eseguire (impartendo gli opportuni comandi alle varie parti del sistema) le istruzioni di un programma, una alla volta

CPU (UN PO' PIÙ NEL DETTAGLIO)



Componenti della Macchina di Von Neumann (2/2)

- L'esecuzione delle **istruzioni** può comportare operazioni di elaborazione di dati (per esempio, operazioni aritmetiche) ovvero di trasferimento di dati (per esempio, dalla memoria centrale all'interfaccia di una periferica)
- La **memoria centrale** contiene le istruzioni di un programma e i dati necessari alla sua esecuzione

Periferiche

- Le **periferiche** sono le apparecchiature che consentono all'elaboratore di scambiare informazioni con il mondo esterno, mediante operazioni di ingresso (verso l'elaboratore) e uscita (verso l'esterno)
  - Vengono in realtà considerate appartenenti alla macchina di Von Neumann solo le interfacce di collegamento verso le periferiche, mentre le periferiche sono considerate componenti separati
  - E' da notare come nel modello di Von Neumann anche le memorie di massa siano incluse tra le periferiche, in quanto funzionalmente analoghe a queste ultime, dal punto di vista dell'interazione con l'elaboratore

BUS

- Il bus di sistema assicura la interconnessione tra gli elementi della macchina di Von Neumann: tutti i trasferimenti di dati avvengono attraverso il bus
- Il bus mette in collegamento logico i due elementi coinvolti nel trasferimento, in funzione dell'operazione da eseguire, mentre il collegamento fisico è sempre presente
- Le fasi di elaborazione si succedono in modo sincrono rispetto alla cadenza imposta da un orologio di sistema (**clock**): è l'unità di controllo, contenuta dentro la CPU, che durante ogni intervallo di tempo, coordina le attività che vengono svolte dentro la stessa CPU o negli altri elementi del sistema

Dati e Istruzioni

- Dati e istruzioni di programma sono codificate in **forma binaria**, cioè mediante sequenze finite di bit
- Una **istruzione** codificata si compone di due parti: il **codice operativo** e uno o più **operandi**:

CO	Op. 1	.....	Op. n

- Il **codice operativo** specifica, secondo una convenzione dipendente dalla specifica macchina, l'**istruzione da eseguire**. Per ogni macchina esistono tanti codici operativi differenti quante sono le istruzioni presenti nell'insieme (set) delle istruzioni che la macchina è in grado di interpretare ed eseguire
- Gli operandi contengono, ancora in una forma codificata dipendente dalla specifica macchina, le **informazioni necessarie a reperire i dati** sui quali l'istruzione deve operare
- Il **linguaggio macchina** è quindi strettamente legato alla architettura della macchina

Dati e Istruzioni in Memoria Centrale (sketch)

- Memoria

0	010000000010000		Istruzioni
1	010000000010001		
2	010000000010010		
3	010000000010011		
4	000000000010000		
……			
……			
15	110100000000000	halt	
16	000000000010001		Dati
17	000000000010010		
18	000000000010000		
……			
……			

Linguaggio Macchina

- Istruzioni codificate in binario:
  - **istruzioni aritmetico-logiche**
    (es. somma di due numeri, confronto tra due numeri …) prevedono indicazione dei dati su cui operare
  - **istruzioni di trasferimento dati**
    (es. da RAM a CPU, da CPU a RAM, input, output…) prevedono indicazione dei dati su cui operare
  - **istruzioni di controllo modificano il flusso di esecuzione** (altrimenti sequenziale) delle altre istruzioni in base ad eventi esterni (es. clic del mouse) o a risultati precedenti

La Memoria Centrale (1/2)

- La Memoria Centrale (comunemente detta RAM - Random Access Memory) è il dispositivo di memorizzazione con cui la CPU interagisce
  - Tramite il Bus
  - Per leggere e scrivere istruzioni e dati (rappresentati in binario)
- La RAM consiste in un insieme di unità elementari di memorizzazione dette Celle
  - Ogni Cella è costituita da un insieme (tipicamente 8) di dispositivi *bistabili* per memorizzare un dato o una istruzione
  - Ogni Cella ha un indirizzo di memoria anch'esso codificato in binario

La Memoria Centrale (2/2)

- La Memoria Centrale è **volatile**:
  - VOLATILE -> Richiede un flusso continuo di alimentazione elettrica per memorizza dati e istruzioni
  - CENTRALE -> Per indicare l'importante ruolo che essa gioca nell'ambito dell'Architettura di Von Neumann
- Altri tipi di memorie sono dette **persistenti**:
  - PERSISTENTE -> Mantiene i dati anche senza l'alimentazione elettrica

Tipologie di ACCESSO ALLE MEMORIE

- Accesso sequenziale
  - Prima di leggere una cella è necessario leggere tutte quelle che la precedono
- Accesso diretto
  - Dato l'indirizzo di una cella ne è possibile l'accesso immediato
- Accesso misto
  - Le celle sono organizzate in blocchi costituiti da più celle, per cui si ha accesso diretto ai blocchi e accesso sequenziale alle celle all'interno dei blocchi
- Accesso associativo
  - L'accesso ad una cella non è guidato dal suo indirizzo ma dal suo contenuto

RAM

- Random Access Memory –> Memoria ad Accesso Diretto
- 1, 2, 4, 8, 16 GB

ROM – Read Only Memory

- Ogni calcolatore dispone di una piccola memoria aggiuntiva denominata ROM (memoria di sola lettura)
  - Viene programmata al momento della sua produzione
  - Non può essere scritta in altri momenti
  - Contiene le istruzioni da eseguire all'avvio del computer (bootstrap) ed alcune funzioni per la diagnostica
- Esistono delle varianti della ROM come le EPROM-> Erasable Programmable Read-Only Memory
  - Memoria non volatile
  - Può essere modificata dal calcolatore con particolari procedure

Le Memorie di Massa (1/4)

- La Memoria Centrale (RAM)
  - è particolarmente adatta a supportare l'esecuzione dei programmi dati i suoi ridotti tempi d'accesso
  - non è adatta per archiviare dati su orizzonti temporali più lunghi data la sua volatilità
- Il calcolatore è dotato di altri dispositivi di memorizzazione chiamati Memorie di Massa (Memorie Secondarie)
  - La loro funzione principale è di garantire la persistenza dei dati
  - Capacità più elevate, costo per byte inferiore, tempi di accesso più lunghi delle memorie centrali

Le Memorie di Massa (2/4)

- Possono essere fisse (e.g. Hard Disk) o rimovibili (pen drive USB)

Le Memorie di Massa (3/4)

- Disaster Recovery
  - Insieme delle tecnologie atte a ripristinare sistemi, dati e infrastrutture a fronte di eventi dannosi
  - Es. BACKUP

Le Memorie di Massa (4/4)

- Le tecnologie più utilizzate per la realizzazione di memorie di massa sono:
  - Dispositivi Magnetici
  - Dispositivi Ottici
  - Memorie Flash

Gerarchie di Memoria

- Registri della CPU
- Cache di primo, secondo e terzo livello
- Memoria Centrale (RAM)
- Dischi interni
- Dischi esterni

Capacità, Velocità d'Accesso, Costo