INTRODUZIONE

Il file nasce come integrazione delle sbobine tratte dalle lezioni dei professori Nicola Mazzocca e Antonio Maria Rinaldi nel corso del primo semestre, dei files forniti dal professore Mazzocca nonché delle slides lasciateci da entrambi. Nella cartella *Materiale utile* potete trovare altri files forniti da loro oppure elaborati insieme a loro che vi aiuteranno nella stesura dei vostri elaborati. Alcuni argomenti della parte di *Basi di dati* non sono stati trattati a lezione (in particolare Sistemi Informativi Direzionali SID, Access e PowerBI) nonostante siano fondamentali per il superamento dell'esame, motivo per il quale vi consigliamo caldamente di consultare le slides fornite dal docente per questi ultimi. Per quanto riguarda le appendici, sono frutto della trascrizione di diapositive fornite dal professore Mazzocca ma a nostro parere inutili ai fini dell'esame – le alleghiamo per completezza. Come per statistica, anche in questo caso il nostro lavoro potrebbe risultare a tratti eccessivamente approfondito rispetto al necessario (SOLO per la parte di AI, ML e database): se vi sembra che sia così ed avete seguito il corso, solo le slides dovrebbero bastare. In bocca al lupo!

INDICE

INFORMAZIONI E ALGORITMI

Lezione del 6/10/2022

N.B. Per approfondire il discorso su sistemi informatici/informativi e basi di dati, consultare Appendice 1

Domanda del primo elaborato:

1. cos'è un sistema informatico?
2. cos'è uno informativo? Qual è la differenza tra i due?
3. quali sono i sistemi informativi che operano in ambito sanitario?
4. quali soni le applicazioni dei sistemi informatici al mondo sanitario?
5. La rappresentazione dell'informazione in termini binari nei sistemi informatici (conversione A/D).
6. che cos'è un algoritmo? Costrutti forti. Esempi elementari.
7. trasformazione dei dati da analogici (continui) a digitale.
8. operatori logici *and* e *or*; costrutti linguistici *if*, *then*, *else*.

Il **sistema informatico** è un **sistema fisico** che consente di elaborare delle informazioni per mezzo di circuiti elettronici (es. computer, che ha una sua organizzazione). Quando si parla di "trattamento dell'informazioni", si intende invece qualcosa di più globale – l'insieme di **persone**, **strutture** e **tecnologie** dà il **sistema informativo**, costituito da una propria organizzazione, essendo un complesso di più azioni collegabili tra loro. Nella sanità ciò è importante perché l'organizzazione – ad es. del sistema di prenotazione delle visite o dei livelli assistenziali minimi – richiede un sistema informativo, non informatico. La sanità è delicata perché nei sistemi informativi vanno considerati anche **imputabilità** e **privacy**. I sistemi informativi si parlano tra di loro.

Esempio. "Gestione del policlinico" – "prof. di statistica" – "gestione della sanità globale"

Quando si vogliono analizzare dei dati oncologici, si avrà una serie di pazienti da classificare in base ai tipi di patologie (ad esempio per un'analisi in base ad un particolare tipo di neoplasia) con il sistema di "*gestione del policlinico*". Il problema del sistema informatico è conservare queste informazioni, ma questo avviene agevolmente grazie alla stesura della **cartella clinica** (un insieme di informazioni che consentono di caratterizzare un paziente in un certo numero di persone *k*). Anche se i pazienti sono tanti (es. studio dell'incidenza dell'inquinamento sulle patologie neoplastiche), servirebbe avere *k* dati dal policlinico ed *m* dati da altre fonti (altrimenti sarebbero troppo poche). L'oggetto dello statistico non sono gli affetti ma i pazienti su cui applicare l'analisi, che si occupa di sintetizzare i dati.

Non tutti possono dare informazioni, ma bisogna controllare gli **accessi** (riconoscimento degli individui che possono inserire informazioni) ed avere un **database** che consenta di gestire gli archivi di informazioni; serve anche un sistema di **imputabilità**, permettendo di tracciare chi ha inserito tale informazione (solo così si possono distinguere le informazioni utili) – molti sistemi hanno algoritmi (elaborazioni) su archivi che permettono di capire chi agisce su determinati dati e come li modifica (vietando l'accesso a chi non è ammesso). Dopo questi primi tre punti (accessi, database, tracciabilità/imputabilità), manca la possibilità di fare delle estrazioni ed elaborazioni dei dati tramite **algoritmi** (applicazioni), che prendono la base dati e fanno delle **inferenze** su questi ultimi. I sistemi informatici che supportano quelli informativi hanno un obiettivo.

Per realizzare le informazioni provenienti da sistemi informativi, vanno stabilite delle regole di **interazioni** – es. la regione Campania chiede a tutti i suoi ospedali di inviare dati sul tempo medio da aspettare per avere una mammografia in un formato dati che è quello che essa stessa chiede così da poterli inviare al Ministero. Questo sistema non cura (non si basa sulla persona singola) ma dà delle **linee di tendenza**.

"*Informativo*" è un aggettivo per "*sistema*", mentre "*informatico*" è vicino al **trattamento dell'informazione** con metodi automatici. Un sistema informativo non necessariamente include o si basa su un sistema informatico, ad esempio un computer con la sua organizzazione.

I sistemi informativi che operano in ambito sanitario prevedono la gestione delle prenotazioni (CUP), quello che consente di gestire i sistemi assistenziali minimi o quello che permette di creare il registro dei malati oncologici – ma ve ne sono molti altri ancora.

https://sinfonia.regione.campania.it/preview/cup & https://app.diagrams.net/

Con un problema ben definito si può usare l'intelligenza artificiale: dopo aver caratterizzato molti pazienti (età, sesso, localizzazione, ecc…) e somministrando loro una cura dopo aver sequenziato anche il loro DNA, sarebbe possibile servirsene per fare una medicina di precisione (se tutti avessero caratteristiche molto simili) e trovare una soluzione facendo determinate analisi – più persone implica avere più informazioni.

Serve un numero finito di campioni visto che qualsiasi computer ha un numero finito di celle di memoria. Fare una misura ad 1 Hz significa prendere un campione ogni secondo. Per prendere il ritmo cardiaco sarebbe meglio usare una frequenza di 1000 Hz (1 KHz) per avere mille campioni al secondo. Un conto è il numero di campioni, mentre l'altro è la precisione con cui vengono presi: normalmente precisione e velocità non vanno molto d'accordo. I campioni così presi devono poi essere inseriti nel database per essere poi elaborati.

Immaginando di avere un'onda quadra, se non si prende un numero sufficiente di campioni si rischia di avere perdita di informazioni riguardanti la rapida variazione – es. l'aritmia (un evento appuntito, *spike*, che sale e scende), mentre ciò non accadrebbe per l'onda diastolica (che è molto alta). Si pensi poi ad un'immagine radiologica: essa è formata da punti infiniti, per cui ci si mette sopra una **griglia** (quanto fitta dipende dal campionamento che si vuole fare) – es. 1000 campioni per ogni lato equivalgono a 2^{20} punti (circa un milione – 2^{10} x 2^{10}), cioè pixel. Un pixel si rappresenta con un colore o in toni di grigio: se il bianco viene indicato con 0 ed il nero con 255, significa che ci sono 256 elementi, per cui sono necessari 8 bit per poter analizzare questi dati ($\log_2 256$); se si vogliono usare i colori sono necessari più bit (es. 16 per poter identificare 2^{16} colori), così che l'immagine diventa più grande e si possono valutare 64000 colori diversi.

N.B. L'informazione minima che il calcolatore può utilizzare è $\mathbf{2^8}$ (il numeratore deve essere un multiplo di 8), perché esso è il **quanto di informazione minima** (corrisponde ad un **byte**) – questo perché con 256 è possibile memorizzare tutti i caratteri possibili della tastiera (che sono meno di 256 ma più di 128).

N.B. Un MHz = 1.000.000 Hz, mentre un GHz = 1.000.000.000 Hz

Presa una cella di memoria, essa riesce a memorizzare solo 0 e 1 (gli unici valori ascrivibili nei registri). Se si devono rappresentare 4 dati, bastano due bit; se ce ne hanno 8, ne bastano solo tre. La formula da applicare è:

$$\log_2 n.di\ dati = n.di\ bit$$

N.B. Più risoluzione, più memoria, meno precisione.

Esercizio n.1

Si immagini di avere 16 oggetti da codificare. Quanti bit sono necessari servendosi di un sistema binario? E quanti per un sistema ternario (con valori 0, 1 e X)?

Basterebbero 4 bit nel sistema binario e 3 bit in quello ternario, in quanto:

- $2^4 = 16$, dunque non avanzano coppie non sfruttate;
- $3^3 = 27$, per cui ci saranno 11 combinazioni non utilizzate (ne servono solo 16).

Per rendere il concetto più immediato, si immagini ancora di voler codificare solo tre valori:

- con il sistema binario basterebbero 2 bit (2^2) ed una coppia non verrebbe utilizzata;
- con il sistema ternario ne basterebbe uno solo.

Valore	Binario	Ternario
A	00	0
B	01	1
C	10	X

Esercizio n.2

Campionare un segnale con una frequenza di 100 MHz significa ottenere 100.000.000 campioni al secondo – i bytes dipendono dalla precisione del campionamento: più bytes per pixel ci sono, più precisa sarà l'informazione (ma questo, nel caso di un'immagine, con la possibilità di vedere meno colori).

Esercizio n.3

Si immagini di voler fare una foto dall'alto ad una campagna che misuri 1024 m per 1024 m. Per ogni lato, per poter avere una risoluzione di 1 m, si dovrebbero ottenere 1024 campioni, mentre per una risoluzione di 0,5 m ne servirebbero 2048. Nel primo caso, il totale dei campioni sarebbe uguale a $1024 \cdot 1024 = 2^{10} \cdot 2^{10} = 2^{20} =$ un milione di pixel. Qualora si avesse a disposizione 1 Mb di memoria, significherebbe avere un byte per pixel, ovvero 2^8 valori per un totale di 256 differenti colori (poco preciso); invece, se si avessero a disposizione 2 Mb di memoria, ciò implicherebbe avere a disposizione 2 bytes per pixel per un totale di $2^{16} = 64.000$ valori, potendo avere una maggior precisione. In parole povere, la precisione è direttamente proporzionale ai bytes (memoria) ma inversamente proporzionale alla risoluzione (ecco perché con la TAC si ottiene un'immagine alla volta).

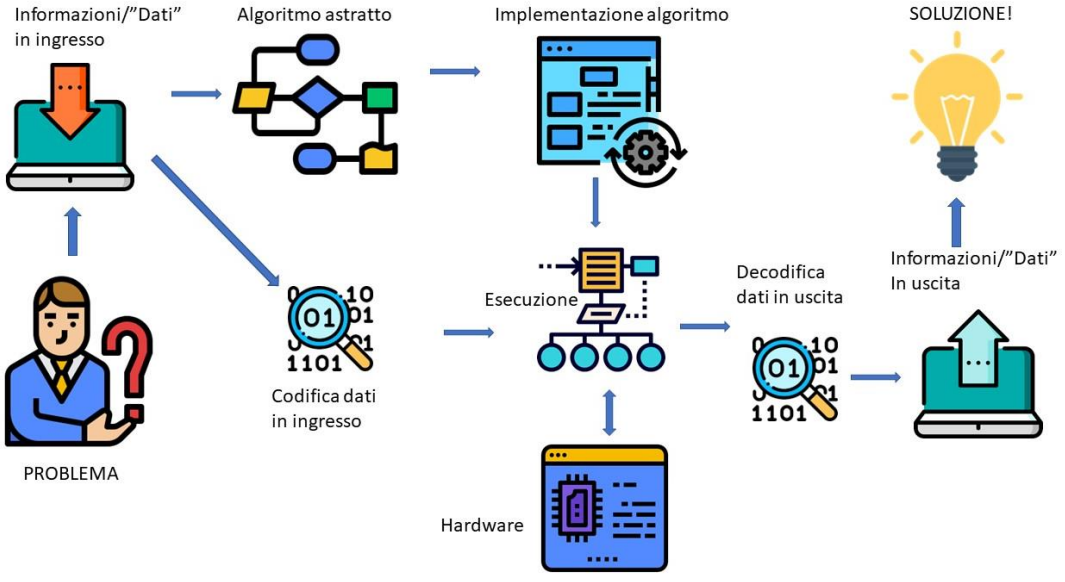
Lezione del 7/10/2022

Per risolvere un problema, si deve definire un **algoritmo**: esso è un insieme di regole che identificano dei passi non ambigui (chi lo legge deve poterlo interpretare in modo univoco) e in numero finito che se svolti in modo corretto consentono di ottenere un risultato – es. la preparazione di un dolce: c'è un insieme di passi finito che, svolti in modo corretto, consentono di ottenere l'oggetto della ricetta, la quale deve essere scritta in modo tale che l'esecutore dell'algoritmo sia in grado di interpretarla (se non è intellegibile, l'algoritmo esiste ma non è utile). Il **linguaggio di programmazione** nasce per codificare gli algoritmi e renderli leggibili da una macchina che deve utilizzarlo: se si volesse costruire una macchina in grado di rilevare e comunicare presenza/assenza di tensione, bisognerebbe convertire questo concetto del linguaggio naturale in uno informatico per consentire alla macchina di poterlo elaborare (es. 0 per assenza e 1 per presenza di tensione).

Quando si prende un sistema informatico (es. computer), bisogna sapere che le sue parti importanti sono:

- **memoria** per conservare i dati e gli algoritmi; in poche parole, un oggetto passivo dedito alla conservazione (es. foglio della ricetta);
- un **processore CPU**, cioè un soggetto attivo che prende i dati dalla memoria e li elabora compiendo delle operazioni dopo aver compreso le informazioni fornite dall'algoritmo;
- **unità di ingresso** e **uscita**.

Il vantaggio del sistema informatico è che, cambiando gli algoritmi, la macchina è in grado di compiere azioni differenti pur rimanendo sempre la stessa – scrivendo istruzioni diverse, essa può assumere qualsiasi tipo di comportamento. Devono però necessariamente essere utilizzate diverse **periferiche**: la loro standardizzazione rappresenta la potenza del sistema dato che queste si collegano al sistema e consentono ad un dispositivo qualsiasi di prendere i dati da una moltitudine di oggetti (le periferiche, appunto) e di elaborarli servendosi di vari algoritmi. In ambito medico, comprare periferiche differenti significa collegarle ad una stessa macchina per ottenere diverse informazioni dal corpo umano utili per varie elaborazioni - es. il tecnico di laboratorio usa l'ecografia per ottenere informazioni e poi le inserisce nella cartella clinica digitale; l'algoritmo ottiene le immagini e le conserva per permetterne future ulteriori utilizzazioni.

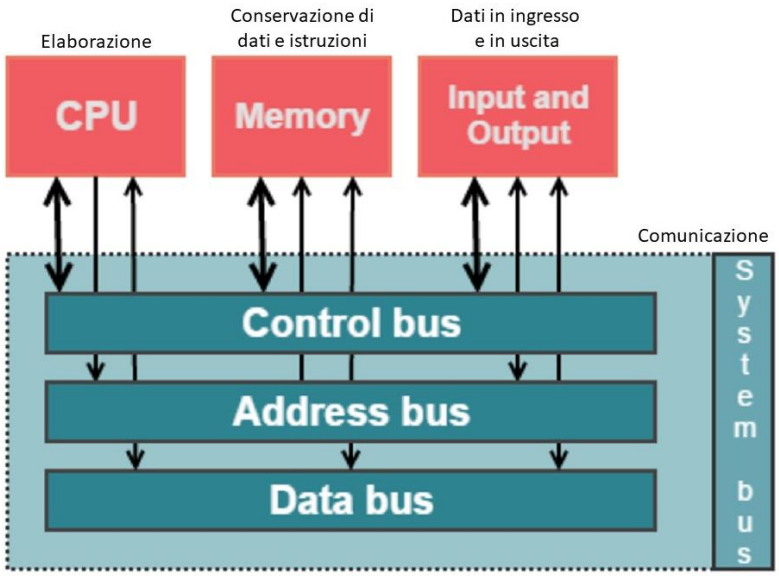


Per poter lavorare con la macchina bisogna codificare le informazioni sottoforma di bit per poi implementare l'algoritmo (il dominio degli ingegneri e dei programmatori); segue poi l'esecuzione che avviene su un certo **hardware** e, alla fine, poiché in uscita si ottengono ancora bit, queste informazioni devono essere decodificate per avere dei dati intellegibili.

Per essere potente, un computer dovrebbe avere una memoria grande e veloce (da cui sia possibile prendere subito le informazioni), ma solitamente quelle veloci sono anche piccole mentre quelle grandi sono lente – così come una valigia più grande sarà anche più lenta. Esiste una gerarchia di memorie fatta di gradi di velocità. Il nome di un processore è costituito da diverse sigle, un'indicazione della capacità del processore di fare elaborazioni rispondendo al sistema: per fare questo, esso riceve un segnale chiamato ***clock*** (ogni volta che ne arriva uno, esegue un'operazione). A parità di *clock*, bisogna vedere anche le operazioni che si possono eseguire: se il clock è veloce ma le operazioni sono più elementari, sarebbe meglio averne uno più lento che permetta un'elaborazione più complessa.

Es. I processori di marca Intel sono identificati da una sigla, ad esempio *i9*, dove "9" indica la generazione di quest'ultimo – nell'ambito della stessa marca, un numero maggiore indica anche una maggiore velocità, per cui *i9* sarà più veloce di *i7*. A parità di oggetto e di marca, sarà migliore quello con i clock più veloci): è possibile confrontare processori nell'ambito della stessa marca per stabilire qual è il migliore, ma tra marche diverse è difficile, perché potrebbero essere costruiti per compiere operazioni diverse. L'evoluzione con i numeri non implica solo un aumento di velocità ma anche, per esempio, di operazioni possibili.

N.B. Ci sono dei sistemi fisici che collegano CPU, memoria e input / output tra di loro.



Se aumenta la velocità della CPU, bisogna aumentare anche quella della memoria – altrimenti quest'ultima rischierebbe di rallentare il sistema. Esistono tre tipi di memorie in un computer:

- **cache** = piccola ma veloce;
- **RAM** (elettronica) = un po' più lenta;
- **hard-disk** = più lenta ma più capace.

La memoria è il serbatoio del processore: per stargli dietro ed essere veloce, deve essere anche abbastanza grande. Essa contiene dati e istruzioni che il processore deve utilizzare per lavorare, per cui se fosse troppo piccola rispetto alla velocità del processore, quest'ultimo sarebbe costretto a fermarsi (avrebbe già elaborato tutte le informazioni al suo interno). Ecco perché salvare immagini meno fitte significa risparmiare memoria.

La maggior parte delle volte, per salvare i dati si usa la cache; se i dati sono più di quelli che essa può contenere e non possono (per limite fisico) trovarsi al suo interno, il dispositivo comincia ad utilizzare e riempire la RAM, ancora abbastanza veloce – motivo per cui il passaggio tra la cache e quest'ultima è praticamente impercettibile e non rallenta il sistema. Tuttavia, se un'informazione è ancora troppo grande per essere contenuta all'interno della singola RAM, parte dei dati contenuti in questa dovranno per forza essere localizzati nell'hard-disk, ma interpellarlo richiede molto più tempo. Se si volesse elaborare l'immagine di un'ecografia, per esempio, si dovrebbero lasciare le altre immagini raccolte (che non necessitano di essere immediatamente processate) sull'hard-disk, mentre quella che deve essere elaborata per dare una diagnosi dovrà quantomeno essere presente solo sulla RAM (altrimenti il computer dovrebbe fare costantemente avanti e indietro tra le due memorie e ciò rallenterebbe il tutto).