Informatica per Statistica Riassunto della lezione del 01/10/2012

Igor Melatti

Qualche informazione generale sul corso

- Dove potete trovare materiale: twiki.di.uniroma1.it/twiki/view/Info4stat/WebHome
 - ci saranno anche informazioni sugli esami ed in generale sull'intero corso
- Docenti: Igor Melatti (titolare) ed Ivano Salvo
- Libri di testo:
 - meramente introduttivo: Laganà, Loghi, Romani, "Informatica Concetti e sperimentazione", Apogeo
 - algoritmi: Cormen, Leiserson, Rivest, Stein, "Introduzione agli algoritmi", McGraw-Hill
 - linguaggio C: Deitel, Deitel, "C Corso completo di programmazione", Apogeo
 - basi di dati: P. Atzeni, S. Ceri, S. Paraboschi, R. Torlone: "Basi di dati - Modelli e linguaggi di interrogazione", McGraw-Hill
 - per ogni lezione verrà fornito il relativo riassunto, disponibile all'indirizzo dato sopra
- Regole per gli esami:
 - ci sono 6 appelli d'esame nel corso dell'anno, più due "straordinari"
 - * uno a gennaio e uno a febbraio (subito dopo la fine del corso), uno a giugno, uno a luglio e uno a settembre nel 2012, e uno a gennaio 2013
 - * i due appelli straordinari (verso aprile il primo e verso novembre il secondo) sono riservati a studenti lavoratori o fuoricorso
 - * per partecipare agli appelli straordinari, occorre dare una prova di essere studente lavoratore (portando il contratto di lavoro) o studente fuoricorso (stampa della carriera universitaria)

- ciascun appello d'esame è composto da uno scritto e da un orale
 - * sono possibili eccezioni a discrezione del docente in caso di appello con pochissimi iscritti
- è possibile partecipare a qualsiasi numero di esami (scritti e/o orali) nel corso dell'anno accademico
- per superare l'esame scritto occorre aver preso almeno 18
 - * fa fede il voto dell'ultimo esame scritto sostenuto
 - * nota bene: questo significa che non superare un esame scritto invalida eventuali esami scritti superati in precedenza
- il voto dell'esame scritto sarà al massimo 25
 - * sono possibili eccezioni a discrezione del docente
- chiunque abbia superato lo scritto può verbalizzare direttamente il voto dello scritto stesso, senza un esame orale
 - * sono possibili eccezioni a discrezione del docente (in caso di dubbio di copiatura)
- chiunque abbia superato lo scritto può richiedere un esame orale per alzare il voto
 - * l'orale sostanzialmente neutralizza il voto dello scritto: chi parte da 18 può arrivare a 30 e lode, e chi parte da 25 può essere bocciato
- chi venga bocciato all'orale dovrà nuovamente superare un esame scritto
- l'orale può essere anche sostenuto (a discrezione dello studente) in un appello diverso da quello in cui si è superato lo scritto
- occorre sempre iscriversi su Infostud, sia per gli esami scritti che per quelli orali
 - * se si fa l'esame orale nello stesso appello in cui si fa lo scritto, basta iscriversi una volta sola (prima dello scritto)
- le domande (scritte o orali) verteranno sempre sull'intero programma del corso
- Obiettivo di questo corso: imparare alcuni rudimenti di informatica
 - cenni di architettura di elaboratori e sistemi operativi (solamente la prima settimana)
 - introduzione ad alcuni semplici algoritmi, e loro implementazione in linguaggio C
 - comprensione e modifica di semplici programmi in C
 - introduzione alla progettazione di semplici basi di dati

- Organizzazione del corso
 - 36 lezioni in tutto, se non ne salta nessuna
 - ogni settimana 3 lezioni: una in laboratorio e due in aula
 - all'incirca 8 lezioni (verso la fine del corso) sono per le basi di dati,
 le altre per algoritmi e linguaggio C
 - le lezioni su algoritmi e linguaggio C saranno disseminate di esercizi, da risolvere durante le ore di laboratorio

Qualche informazione generale sull'informatica

- Cominciamo col parlare in generale di informatica prima, e di computer poi
- Informatica: definizioni (Wikipedia, http://it.wikipedia.org/wiki/Informatica)
 - scienza interdisciplinare che riguarda tutti gli aspetti del trattamento dell'informazione mediante procedure automatizzabili (vecchia definizione)
 - studio dei fondamenti teorici dell'informazione, della sua computazione a livello logico e delle tecniche pratiche per la loro implementazione e applicazione in sistemi elettronici automatizzati detti quindi sistemi informatici
 - dal francese: inform(ation electronique ou autom)atique (1962)
 - in parole povere, l'informatica si occupa dei vari metodi di acquisire informazioni, elaborarle e poi restituire il risultato, usando mezzi automatici
 - * per mezzi automatici, si intende il fatto che non è l'essere umano ad "eseguire" tali metodi
 - * al contrario, l'essere umano "insegna" ad un congegno elettronico come si esegue il compito da eseguire (ovvero, programma un computer), ed è poi il congegno ad eseguire il compito stesso
 - * qui il concetto di "programmazione", come anche quello di "computer", è da intendere in modo molto ampio, ci ritorneremo
 - quindi, l'informatica si può occupare di una miriade di cose:
 - * prendere due interi e sommarli
 - * prendere un sistema di equazioni differenziali e approssimarne la soluzione
 - * trasferire dati da una posizione ad un'altra (es: Internet)
 - * memorizzare dati complessi e ritrovarli quando richiesto (databases)

- * acquisire misurazioni e decidere quale comando dare ad un sistema fisico (es.: pilota automatico di un aereo)
- * analizzare sequenze lunghissime per trovare sottosequenze ricorrenti (es.: decodifica del genoma umano)

* ...

- lo strumento base per fare tutto questo (il "congegno" di cui sopra)
 è ciò che comunemente viene chiamato "computer"
 - * elaboratore, in italiano, sottintende un processo prossimo all'intelligenza umana
 - * ordinateur, in francese, a sottolineare le sue capacità di organizzare i dati (oggi le informazioni)
 - * computer, in inglese, letteralmente calcolatore, in diretta discendenza delle calcolatrici, prima meccaniche, poi elettromeccaniche, poi elettroniche (analogo all'altro termine italiano, calcolatore).

• E vediamoli questi computer

- lasciando da parte le varie dispute su quale sia stato il primo computer, si cominciarono a costruire i primi veri calcolatori intorno agli anni 40 del Novecento
- i primi computer occupavano intere stanze ed erano migliaia di volte meno potenti di un normale portatile odierno, e anche meno potenti di uno smartphone d'oggidì (Fig. 1)
- esistono ancora computer che occupano intere stanze, ma oggi come allora se li possono permettere solo università, governi o grandi aziende
 - * si tratta di computer composti da tanti altri computer che collaborano tra loro: sono i *cluster* (Fig. 2)
 - * usati per alcuni calcoli molto lunghi e complicati, come ad esempio per la ricerca genetica e per le previsioni del tempo
 - * ovviamente, il confronto di potenza di calcolo con gli antenati degli anni quaranta è impietoso
- ma i computer di cui ci occupiamo qui sono più semplici (almeno da capire)
- intenderemo per computer una macchina elettronica che può essere programmata per svolgere diversi compiti
- esistono svariati tipi di computer, che possono essere categorizzati per dimensioni e utilizzo
 - * vedere http://en.wikipedia.org/wiki/List_of_computer_size_categories
- semplificando un po', si possono distinguere le seguenti categorie:

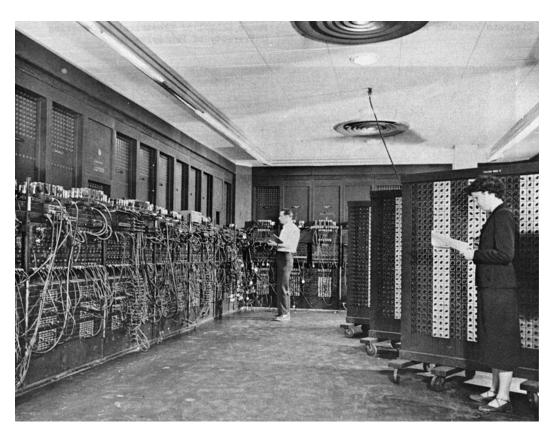


Figure 1: ENIAC, uno dei primi veri computer (1946, da Wikipedia)

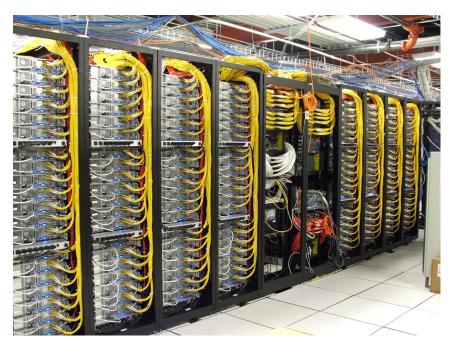


Figure 2: Emulab, cluster di circa 300 computer della University of Utah (2008)

- 1. supercomputer: quelli di cui sopra, sono tipicamente dei cluster impiegati per calcoli lunghi e complessi (e parallelizzabili)
- 2. server; sono abbastanza simili ai computer desktop "fissi" (vedere più avanti), ma:
 - * hanno molta più memoria RAM (anche centinaia di GBs)
 - * hanno maggiore potenza di calcolo a livello di velocità di CPU (ma su questo la differenza si è assottigliata)
 - * hanno maggior numero di CPU
 - * hanno molta più memoria su disco fisso
 - * hanno tipicamente vari sistemi di backup automatici (anche su nastri magnetici!)
 - * sono solitamente usati da grandi aziende per fornire servizi; ad esempio le banche o le poste
 - * un insieme di server viene talvolta chiamato "mainframe" (termine che fino agli anni 80 indicava i computer usati dalle aziende)
- 3. microcomputer, termine un po' vecchiotto, designa i computer normalmente usati; sono quelli introdotti qui di seguito
- i computer più diffusi sono i "personal computer" (PC), talvolta chiamati microcomputer, sono quelli che ci si può comprare con meno di un migliaio di euro





Figure 3: Un desktop (sinistra) e un laptop (destra)





Figure 4: Un netbook (sinistra) e un altro tablet (destra)

- * se li può quindi permettere anche una singola persona, da usare per sé stesso (da cui il nome "personal")
- * sono tipicamente quelli che trovate anche nei laboratori
- * ci sono computer "portatili" (laptop o notebook) e computer "fissi" (desktop, Fig. 3)
- * i portatili possono essere veramente piccoli, e venir detti subnotebook o netbook (si può far rientrare rientrare in questa categoria l'iPad; Fig. 4)
- * anche alcuni dei moderni telefonini (non a caso detti *smartphone*, come l'iPhone o prodotti analoghi) sono praticamente dei computer

L'Hardware (ed il software) di un (micro)computer

• Come descritto nel seguito, un computer è il risultato dell'unione di diverse componenti che, comunicando tra loro, contribuiscono al suo fun-

zionamento

- Ad oggi, ci sono due modi di costruire PC che hanno maggiore successo commerciale
 - 1. secondo un'architettura "aperta", un tempo chiamata IBM-compatibile
 - i vari pezzi che compongono il computer possono essere prodotti da diverse aziende
 - speciali accordi (standard) su alcune modalità di costruzione (in particolare della parte di comunicazione tra componenti) permettono di mettere assieme (assemblare) il tutto affinché funzioni nella maniera corretta
 - ci sono aziende che, pur non producendo pezzi particolari, sono specializzate proprio nell'assemblaggio
 - 2. secondo un'architettura "chiusa", prodotta dalla Apple
 - i vari pezzi che compongono il computer sono tutti prodotti dalla Apple, che al massimo subappalta
 - fa eccezione il microprocessore (vedere più avanti)
 - tra l'altro, Apple e i suoi molti sostenitori rifiutano il nome di PC per i loro computer
 - il termine PC viene infatti spesso usato per riferirsi solo a quelli dell'architettura aperta
 - comunque, i computer della Apple sono dei microcomputer
- Ma per funzionare, come vedremo, un computer ha bisogno di software e non solo di hardware
 - semplificando, l'hardware è la parte che si tocca con mano
 - * il coperchio del *case*, la tastiera, il monitor, la stampante; ma anche quello che c'è "dentro", e che vediamo tra poco
 - il software sono i vari programmi che dicono al computer cosa deve fare
 - * programmi per scrivere testi e poi stamparli, per fare conti...
- Il software di base viene chiamato sistema operativo
 - ovvero, è il programma che viene eseguito non appena (o quasi) accendete il computer
 - nel caso degli IBM-compatibili ci sono sostanzialmente 2 possibili scelte:
 - \ast i sistemi operativi della famiglia Windows (XP, 2003, Vista, 7, 8...)



Figure 5: Cosa c'è dentro un computer

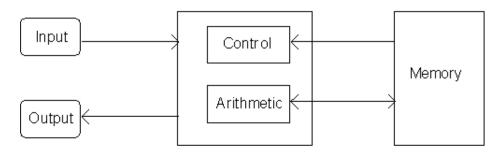
- * i sistemi operativi della famiglia Linux (Ubuntu, Fedora, Mandriva...)
- nel caso degli Apple ci sono il Mac OS X e le sue varie evoluzioni (es. Leopard)
- Da un punto di vista hardware, i computer sono tipicamente costruiti più o meno tutti nella stessa maniera, perché seguono l'architettura di Von Neumann (Fig. 6)
 - esistono alcune eccezioni per compiti speciali, ma sorvoliamo
 - questo concetto risale agli anni 40, è stato in seguito reso più efficiente (architetture di Harvard) ma non superato
- Si immagini di dover svolgere un qualche compito complicato, per il quale vengano date delle istruzioni ben precise
- $\bullet\,$ Ad esempio, trovare una particolare sequenza (sonda) di nucleotidi in un genoma umano
 - -genoma: ...TCGTTCCGAAAGTT...
 - sonda: GTT
 - risultato: 2 corrispondenze, al posto 3 e al posto 12
- Le istruzioni (algoritmo) saranno una cosa di questo tipo
 - 1. comincia dal primo nucleotide del genoma
 - 2. considera il primo nucleotide della sonda

- 3. se il nucleotide attuale della sonda è uguale a quello attuale del genoma
 - (a) confronta tutti i seguenti nucleotidi della sonda con quelli seguenti del genoma
 - (b) se c'è corrispondenza, memorizza la posizione di inizio
- 4. passa al nucleotide successivo del genoma, e torna al passo 2, altrimenti prosegui col passo 5
- 5. restituisci in uscita le posizioni memorizzate al passo 3b
- Per eseguire tali istruzioni, la cosa migliore è mettersi su un grande tavolo, con un foglio recante le istruzioni stesse e altri fogli per appunti intermedi (memoria)
- Prendere il genoma e la sonda che vengono forniti dall'esterno (input)
- Eseguire di volta in volta le istruzioni di cui sopra, eventualmente servendosi di una calcolatrice per qualche conto (controllo e unità aritmeticologica)
 - in questo esempio al massimo si somma 1 e si confrontano lettere,
 ma per altri problemi la situazione potrebbe essere più complicata...
- Dare il risultato all'esterno (output)
- Da notare che ciò implica la capacità di accedere a proprio piacimento alla memoria, ai dati in input o in output (bus di comunicazione, unico non esplicitamente indicato in Fig. 6)

Controllo ed unità aritmetico-logica: la CPU

- Grazie al bus, preleva le istruzioni dalla memoria e le esegue, eventualmente usando la ALU (Arithmetic-Logic Unit)
- \bullet Detta anche microprocessore per le piccole dimensioni (come un francobollo)
 - volendo essere pedanti, il microprocessore è l'implementazione più frequente della CPU
 - per la precisione è l'implementazione che mette tutto su una piattaforma di silicio
 - ovvero, la CPU è un concetto più astratto, il microprocessore è il blocchetto concreto
- Sono microprocessori quelli prodotti dalla Intel (i famosi Pentium), dalla AMD (compatibile con l'Intel) e dalla Apple (con IBM e Motorola, ma oramai usano i Pentium pure loro)

Central Processing Unit



von Neumann Architecture (Adapted from Aspray, 1990a)

Figure 6: L'architettura di Von Neumann

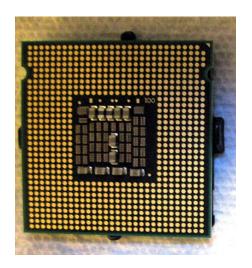


Figure 7: Un microprocessore



Figure 8: Alcune memorie RAM

- La potenza di un processore è tipicamente misurata col suo ciclo di clock
 - l'unità di misura è il GHz (GigaHertz), i processori moderni sono sui 3 GHz
- Fino a qualche anno fa, grazie a continue migliorie tecnologiche si raddoppiavano i GHz ogni 2 anni (*Legge di Moore*)
- Oggi si preferisce costruire computer con almeno 2 microprocessori e distribuire tra di loro le computazioni

Memoria

- Ci sono 2 tipi di memoria in un computer di cui l'utente si accorge
 - quella secondaria o di massa, ovvero la memoria su disco (hard disk)
 - quella primaria o centrale, ovvero RAM (Random Access Memory)
 - $-\,$ solo la RAM fa parte della memoria com'è intesa nell'architettura di Von Neumann
 - il disco è a tutti gli effetti input/output, quindi ci ritorneremo dopo
 - è importante sottolineare tuttavia la differenza fondamentale tra questi due tipi di memoria
 - -la RAM è volatile: spegnere il computer vuol dire perdere il contenuto della RAM
 - il disco è permanente: i dati restano lì anche dopo lo spegnimento del computer
 - inoltre, la RAM ha una capacità di memorizzazione di pochi GB, il disco di centinaia di GB
 - ma leggere e scrivere sulla RAM è molto più veloce



Figure 9: Un hard disk

Input/output

- Sono di solo output: monitor, stampante, masterizzatore CD/DVD
- Sono di solo input: tastiera, mouse, lettore CD/DVD, scanner
- Sono sia di input che di output: hard disk (Fig. 9), scheda di rete, lettore-masterizzatore CD/DVD
 - il disco (detto hard per distinguerlo dagli ormai obsoleti floppy disk)
 è di output nel momento in cui vi si memorizzano dati di interesse permanente
 - è di input nel momento in cui vi si prelevano dati precedentemente memorizzati
 - un disco fisso può tipicamente memorizzare 100 GB di informazione, all'incirca l'equivalente di 100 film
 - giocano lo stesso ruolo dell'hard disk:
 - * i floppy disk (Fig. 10), oramai obsoleti, però i desktop hanno tuttora il lettore/scrittore di floppy disk
 - * avevano una capienza di 1.5 MB circa
 - * oggi ci sono i CD (700 MB circa, Fig. 11)
 - * i DVD (7 GB circa, ad occhio non distinguibile dal CD)
 - * le chiavette USB (tra 1 GB e 8 GB, di solito; Fig. 13); molto comode perché usabili senza particolari dispositivi
 - * invece, per CD e DVD servono lettori (e scrittori, o masterizzatori) particolari (Fig. 12)
 - * d'altronde, CD e DVD sono più economici di una chiavetta USB, anche se questa differenza va assottigliandosi
 - infine, è di input/output la scheda di rete (Fig. 14), quella che permette di usare Internet

Il BUS: la scheda madre

- La scheda madre è una piattaforma grande quanto un foglio di giornale sulla quale si posizionano:
 - il microprocessore



Figure 10: Un floppy disk (dischetto o disquette)



Figure 11: Un DVD



Figure 12: Un lettore CD interno



Figure 13: Una chiavetta USB



Figure 14: Una scheda di rete interna



Figure 15: Una scheda madre





Figure 16: Sin: una porta parallela (tipicamente per stampanti); Dx: una porta seriale (tipicamente per monitor)





Figure 17: Sin: porta PS2 (per tastiera e mouse); Dx: connettore USB (qualsiasi dispositivo USB)

- la RAM
- i connettori con le varie unità di input/output
 - * alcune unità di input/output possono essere connesse direttamente alla scheda madre "da dentro"
 - * sono i cosiddetti apparecchi "interni" (ad esempio un lettore floppy o un lettore CD/DVD)
 - * se invece vengono comprati a parte e poi connessi al computer, sono apparecchi "esterni"
 - * la connessione avviene tramite questi connettori (com'era prevedibile)
 - * i connettori USB si sono diffusi negli ultimi 10 anni, è un'unificazione dei tipi di connettori
 - * esistono mouse USB, tastiere USB, stampanti USB...
- un canale che permetta a tutte queste cose di comunicare tra loro (il bus dell'architettura di Von Neumann)
 - * o meglio, un canale che permette alla CPU di controllare il tutto