DICo

Modulo informatica di base[1] — Linea 2

Mattia Monga

Dip. di Informatica e Comunicazione
Università degli Studi di Milano, Italia
mattia.monga@unimi.it

a.a. 2010/11

[1] © 2010 M. Monga. Creative Commons Attribuzione-Condividi allo stesso modo 2.5 Italia License.
http://creativecommons.org/licenses/by-sa/2.5/it/. Rielaborazione del materiale 2009/10 © S. Mascetti.

DICo

Lezione I: Introduzione

Informazioni sul corso

- Lezione frontale: mercoledí 10:30–12:30 (G24)
- Laboratorio: secondo calendario, venerdí 13:30–17:30 (307)
  - Due turni (suddivisione gestita dalla segreteria didattica)
- Esame:
  - Scritto con domande
  - Prova pratica per la parte di laboratorio (esercizio da risolvere al PC)
- Libro di testo *Fluency: Conoscere e usare l'informatica*, di L. Snyder, Pearson Education.
- https://mameli.docenti.dico.unimi.it/biotecno

Ricevimento

- Su appuntamento: mattia.monga@unimi.it
- Scrivere da indirizzo universitario (mailstudenti.unimi.it)
- Suggerimenti per scrivere una buona email (http://vigna.dsi.unimi.it/comescrivere.php)
- Ufficio 222, Via Comelico 39 (http://maps.google.com/maps?q=via+comelico+39+milan&ie=UTF8&hq=&hnear=Via+Comelico,+39,+20135+Milano,+Lombardia,+Italia&ei=Zx7RTLf7F87sOYe29MoM&ved=0CBUQ8gEwAA&z=16)

Informatica

ACM — Association for Computing Machinery

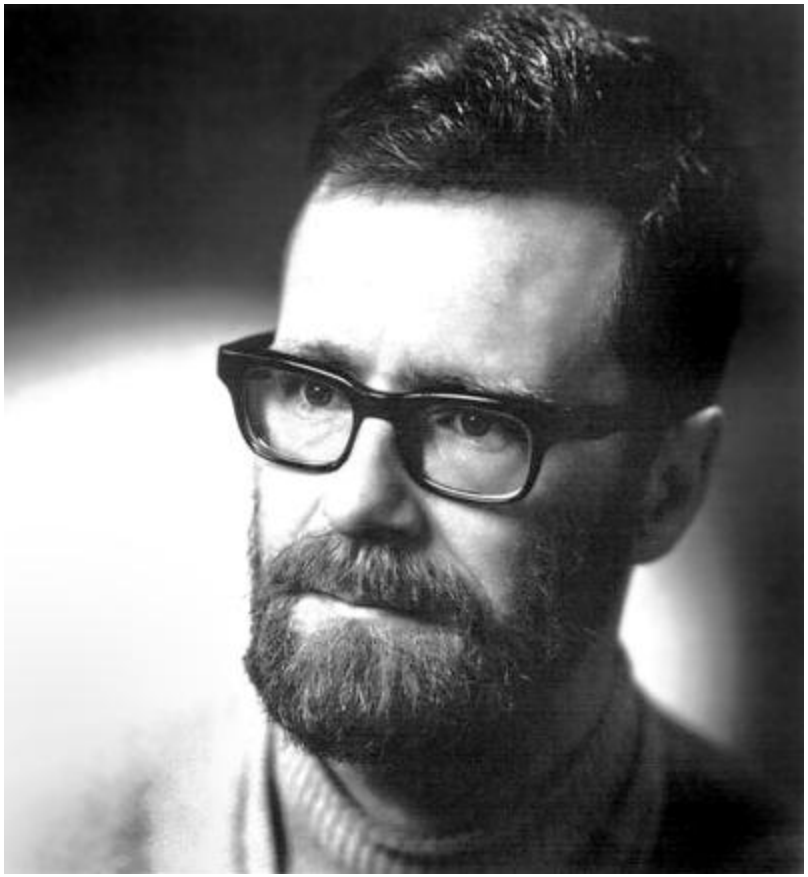
The discipline of computing is the systematic study of algorithmic processes that describe and transform information: their theory, analysis, design, efficiency, implementation, and application. The fundamental question underlying all computing is 'What can be (efficiently) automated?

- Non è l'uso delle applicazioni (applimatica?)

Scienza dei calcolatori?

Edsger Dijkstra

L'informatica non è la scienza dei calcolatori. Non piú di quanto l'astronomia sia la scienza dei telescopi o la chirurgia la scienza dei bisturi.



Perché serve in un corso di biotecnologie

- capire cosa c'è dietro alle applicazioni informatiche per non farsene dominare
- comprendere l'importanza dell'informatica nel campo delle ricerche biotecnologiche
- utilizzare un approccio informatico nella risoluzione di problemi in tutti i campi
  - Es.: Chimica $\rightsquigarrow$ hardware; DNA $\rightsquigarrow$ software

Esempio di problem solving informatico

DICo

Il signor Sbevazzo ha a disposizione due recipienti vuoti (una damigiana e un bottiglione), in grado di contenere rispettivamente 9 litri e 4 litri. Vuole riempire la damigiana con esattamente 6 litri di acqua, ma le uniche operazioni che sa fare sono le seguenti:

RIEMPI D	riempire la damigiana,
SVUOTA B	svuotare il bottiglione,
VERSA	versare il contenuto della damigiana nel bottiglione, fino a riempire il bottiglione o a svuotare la damigiana nel bottiglione.

Qual è la *piú breve* sequenza di queste operazioni che gli permette di raggiungere il suo obiettivo?

DICo

Lezione II: Sistema operativo e file system

Software

Un computer è sempre composto da

- hardware (hw): i dispositivi elettronici in grado di operare le elaborazioni di base (sostanzialmente cambiare un segnale elettrico in uno con caratteristiche diverse)
- software (sw) le istruzioni che combinano le elaborazioni di base (cosiddetto linguaggio macchina) per ottenere risultati ad un livello direttamente interpretabile dall'utente finale (per esempio cambiando il colore ai pixel dello schermo)

La caratteristica piú evidente del sw è che può essere cambiato facilmente: senza alterare i componenti elettronici della macchina.

Software di sistema e applicazioni

DICo

Sistema operativo

Il sistema operativo (so) (piú ampiamente il software di sistema) è sw che:

- facilita l'interazione con l'utente
- gestisce le risorse della macchina in maniera efficace (impedendo all'utente, p.es., di danneggiare il sistema)

Applicazioni

Le applicazioni sono programmi (sw) che fanno qualcosa di utile per l'utente del sistema:

- le si attiva (in gergo: lancia) *tramite* il sistema operativo
- nel periodo in cui sono attive trasformano il computer in qualcosa di diverso: p.es. una macchina per scrivere, che richiede abilità *tipografiche* molto piú che *informatiche*!

Domande

1. Che sistema operativo sta *girando* sulla macchina che stai usando? (Nome, versione, ecc.)
2. Conosci i nomi di altri sistemi operativi? Conosci sistemi operativi specializzati per i telefoni cellulari?
3. Indica nel sistema che stai usando almeno 5 applicazioni che sicuramente non fanno parte del sistema operativo (anche se magari sono state distribuite commercialmente con esso).
4. Sai indicare, invece, alcune funzionalità fornite direttamente dal sistema operativo?
5. Cosa pensi indichi il termine "applicazione web"?

Interfaccia

L'interfaccia è la modalità con la quale l'utente *dialoga* col computer

- chiedendo al so di attivare o disattivare un'applicazione
- cambiando la *configurazione* del sistema
- operando secondo la logica di una specifica applicazione

Tipi di interfaccia

I modi di interazione sono molti e molto diversi fra loro...

- scrivendo con la tastiera parole di comando
  - `notepad` $\rightsquigarrow$ lanciare il programma Notepad
- cliccando e muovendo il mouse su immagini sullo schermo, facendo riferimento ad una metafora (non sempre del tutto coerente...)
  - Doppio clic sull'icona di Notepad
  - Scelta di Notepad dal menú principale
  - Esegui (Run) notepad
  - Tasto destro del mouse, Apri con Notepad

Una metafora molto diffusa è quella della scrivania (*desktop*) sulla quale vengono poste immagini cliccabili di oggetti (*icone*) associate ad appl. o insiemi di dati (*file*) da manipolare con apposite applicazioni.

Domande

1. Sapresti spiegare cos'è una *finestra* in una modalità di interazione di tipo grafico? (Suggerimenti: cosa significa usare un'applicazione "a schermo intero"? Cos'è una "finestra di dialogo"?)
2. Descrivi le interfacce di dispositivi coi quali hai interagito (p.es. un lettore digitale di musica, un cellulare, un televisore digitale, ecc.)
3. Hai mai sentito parlare di *drag-n-drop*? Di cosa si tratta?
4. Hai mai sentito parlare di *copia e incolla*? Descrivi i passaggi logici di questo meccanismo.
5. Anche se in linea di principio ogni applicazione potrebbe avere una propria interfaccia, in realtà generalmente molti tratti sono comuni; identificane alcuni fra due applicazioni completamente diverse.

File

Ogni applicazione nel periodo in cui è attiva (si dice che è un processo in esecuzione) manipola dati. Spesso c'è l'esigenza di conservare i dati anche al termine dell'attività:

- per riprenderli (e ri-manipolarli)
- per condividerli con altre applicazioni
- per duplicarli (e avere cosí copie di sicurezza (backup)!

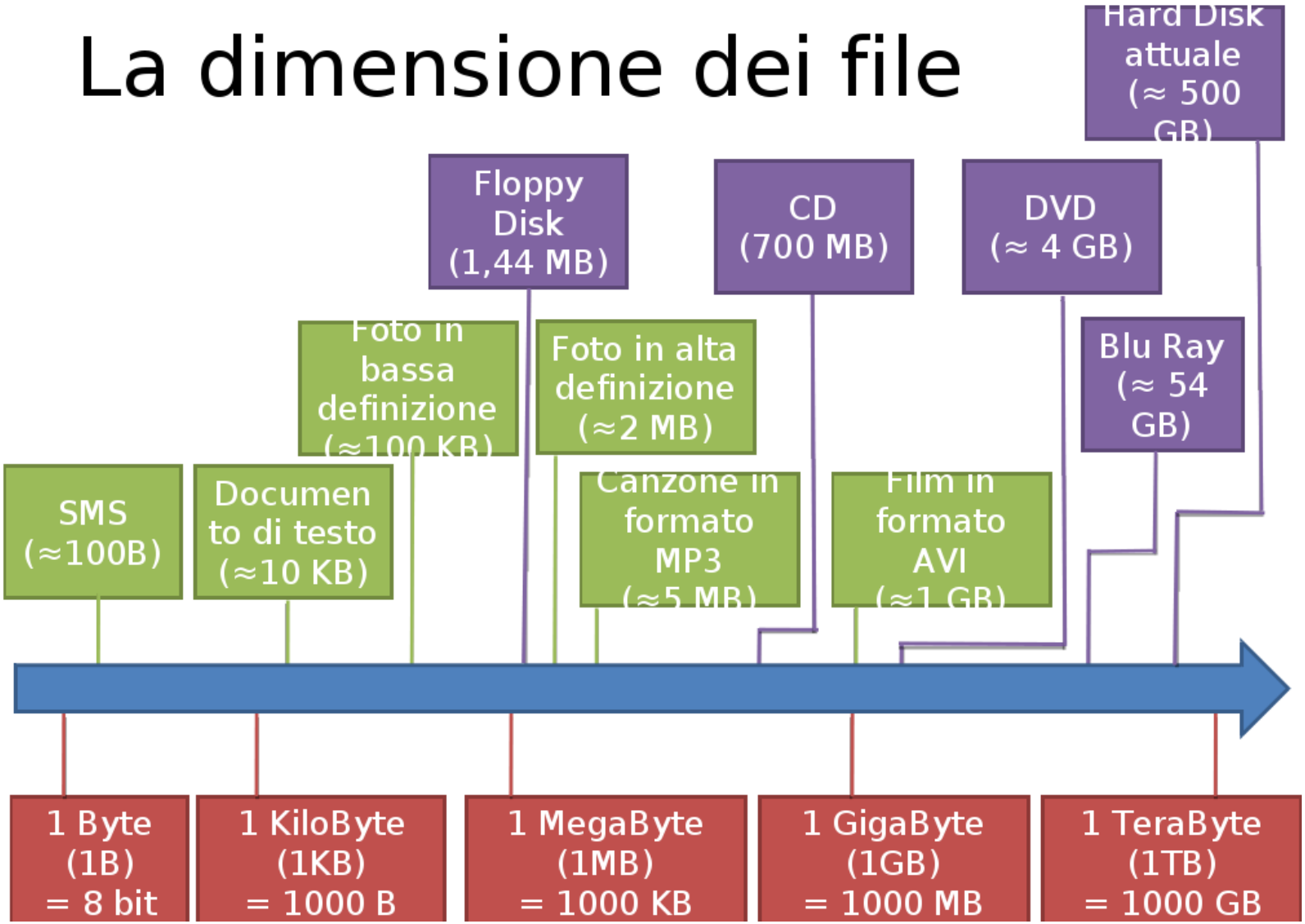
La modalità piú comune (praticamente onnipresente nei PC) è il file: una sequenza di dati codificati, identificata da un nome.

File system

Il file system è la parte di so che si occupa della gestione dei file. Alcune delle convenzioni piú diffuse:

- Al file si dà un nome con un suffisso, per riconoscere facilmente la tipologia di dati conservati e usare l'applicazione giusta per manipolarli
  - `lettera-alla-mamma.doc` $\rightsquigarrow$ probabilmente è un file che contiene dati manipolabili con MS Word (*estensione* del file)
  - La tipologia è evidenziata anche da un'icona che ricorda il programma piú frequentemente usato per la manipolazione
  - Vengono conservati anche dati (*metadati*) su: la data di ultima modifica, la dimensione, ecc. (vedi proprietà del file)

La dimensione dei file



Domande

- Identificare le proprietà di un file.
- Fare la copia di un file. A volte invece di una copia si preferisce creare un *collegamento* (link, scorciatoia, shortcut,...): che differenza c'è?
- Creare un file con Openoffice Writer e provare ad aprirlo con Notepad. Cosa succede?
- Cancellare il file creato. È possibile recuperarlo?
- Cos'è una cartella? Cosa può contenere?

Lezione III: Applicazioni in rete

Protocolli

Perché calcolatori potenzialmente diversi possano scambiare dati tramite una rete occorre un accordo convenzionale.

Protocollo

È la definizione formale di regole che devono essere rispettate per stabilire una comunicazione efficace

P.es.: il simbolo "é" deve essere codificato con i bit `11101001`, che vanno trasmessi a partire dal primo a sinistra; per richiedere la pagina web `http://www.example.com/ciao` occorre mandare esattamente i caratteri

```
1 GET␣/ciao␣HTTP/1.0
2
```

Applicazioni in rete

DICo

Per ogni applicazione in rete ci deve essere (almeno) un protocollo di comunicazione. Alcuni protocolli famosi (alcuni di questi li approfondiremo nella parte di teoria)

- TCP: il principale p. di trasporto (stabilisce come devono essere trasmessi i byte scambiati) delle reti Internet
- HTTP: regola la richiesta di pagine web
- HTTPS: idem, ma i dati scambiati vengono crittati
- SMTP: regola la spedizione di messaggi ad un server di posta elettronica
- IMAP: regola l'accesso ai messaggi di posta elettronica conservati su un server (un'alternativa è POP3, meno flessibile)

client è chi inizia una comunicazione, server è chi risponde.
Per evitare che ogni p. debba ridefinire tutte le regole di trasmissione, si usano a vicenda: si parla di pila di protocolli.
P.es.: tutti quelli citati usano TCP per trasmettere i dati.

Posta elettronica

- Un'applicazione per leggere, inviare, conservare messaggi di posta elettronica (p.es. Outlook, Eudora, TkRat, KMail, ...)
- È ormai molto diffuso l'accesso tramite *browser* (webmail): in questo caso la conservazione è compito del fornitore del servizio

Uso di un'applicazione per la posta elettronica

DICo

Generalmente occorre decidere se conservare la posta in

- locale (con la webmail non è sempre facile, però)
  - \+ consultazione anche senza connessione
  - − affidabile quanto lo è la propria macchina
- remoto
  - \+ accesso da qualunque macchina
  - \+ i backup sono responsabilità del fornitore del servizio
  - − privacy/sicurezza?
  - − spazio a disposizione generalmente piú limitato
- locale & remoto (sincronizzazione solo con IMAP)

Nei client occorre impostare i parametri per l'invio (server SMTP) e per la ricezione (server IMAP o POP3). Nella webmail di solito sono già impostati, ma può essere utile far *confluire* piú indirizzi in un'unica casella.

Messaggio di posta elettronica

Non è altro che un messaggio di solo testo (cioè solo caratteri) con intestazioni (header) predefiniti

```
1 From: Mattia Monga <mattia.monga@unimi.it>
2 To: Magnifico <rettore@unimi.it>
3 CC: Amplissimo <presidescienzemfn@unimi.it>
4 BCC: direttore@dico.unimi.it
5 Subject: piaggeria
6
7 Grazie per tutto quello che fai per noi!
8 --
9 MM
```

Il protocollo nel tempo è stato reso piú complicato per

- gestire file allegati (attachment)
  - Attenzione! Per interpretare un file occorre che il destinatario abbia un'applicazione opportuna. Solo il testo *puro* (e con caratteri americani...) è garantito dal protocollo di trasporto
- problemi di sicurezza

Esercizi

1. Accedete alla vostra casella di webmail
   `http://mailstudenti.unimi.it`
2. Spedite una mail a me con alcuni vostri colleghi in BCC, `Subject: [labinfo]`
3. Nella mail ricevuta riuscite a capire chi è il mittente?
4. Riuscite a capire chi sono i destinatari?
5. Mandate anche un messaggio con un (piccolo) allegato
6. Cosa significa *inoltrare* un messaggio?
7. Vengono inoltrati anche gli allegati
8. Che differenza c'è fra `Reply` e `Reply all`? Quando c'è qualche BCC?
9. Cosa si intende per spam?
10. Netiquette:
    `http://it.wikipedia.org/wiki/Netiquette`

Risorse web

In generale ogni risorsa accessibile via web ha un indirizzo detto URL

`http://homes.dico.unimi.it/~monga/lib/biotecno/infolab01.pdf`

`protocollo://server/risorsa`

Quindi se conosciamo l'url, basta chiedere al browser di visualizzarlo. Se non conosciamo l'url, possiamo cercarlo con un motore di ricerca (Google, Yahoo, Bing, ...)

Motori di ricerca

- Non tutto è indicizzato
- Può essere diverso cercare parole o frasi
- Anche l'ordine conta (es: Mike Siwek Lawyer MI)
- Affidabilità di una ricerca per parole chiave vs. affidabilità di un URL

Esercizi

1. Trovare il testo completo della canzone che contiene il verso “Io scopriremo solo vivendo”
2. Riuscite a trovare una foto di uno dei compagni che vi sta vicino?
3. Trovare informazioni su il famoso ingegnere del software “Michael Jackson”
4. Dovete scegliere se comprare una macchina fotografica digitale della Canon o della Nikon. Come impostereste la ricerca delle informazioni riguardo alle esperienze di acquisto? Come controllereste che il modello scelto è compatibile con la vostra piattaforma PC?

Wiki

Wiki

Pagine web modificate dai lettori, permettendo una collaborazione semplice e veloce (in hawaiano wiki wiki significa "rapido" oppure "molto veloce")

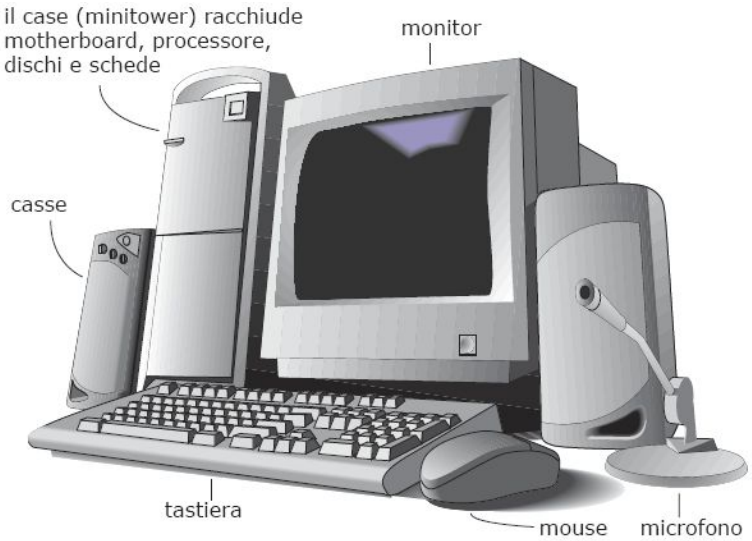
L'esempio piú famoso è Wikipedia (`wikipedia.org`), un'enciclopedia collaborativa.

Esercizi

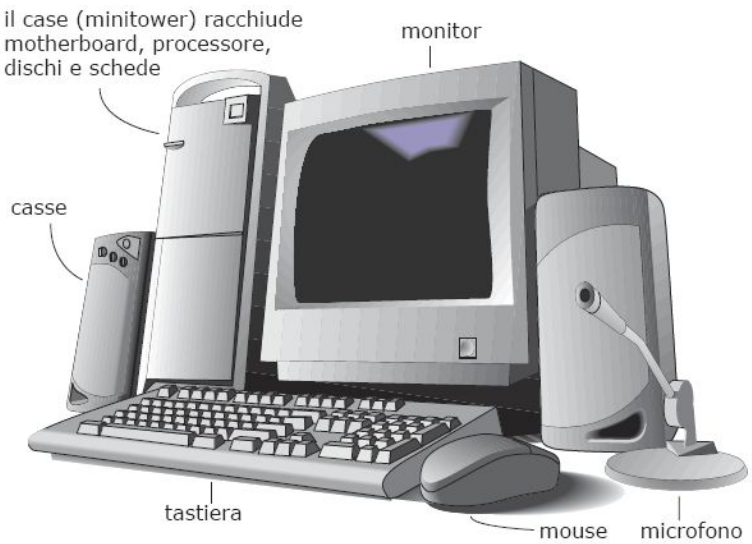
1. Trovate una pagina di Wikipedia che contiene un errore e correggetelo!
2. Guardate la cronologia di una pagina di Wikipedia e identificate i cambiamenti rispetto alla versione precedente
3. Come fareste a sincerarvi della veridicità delle informazioni contenute in http://it.wikipedia.org/wiki/Dna?
4. Quale grado di affidabilità dareste alle informazioni contenute in http://it.wikipedia.org/wiki/Mariastella_Gelmini? Come le controllereste?

Lezione IV: Gli strumenti di calcolo

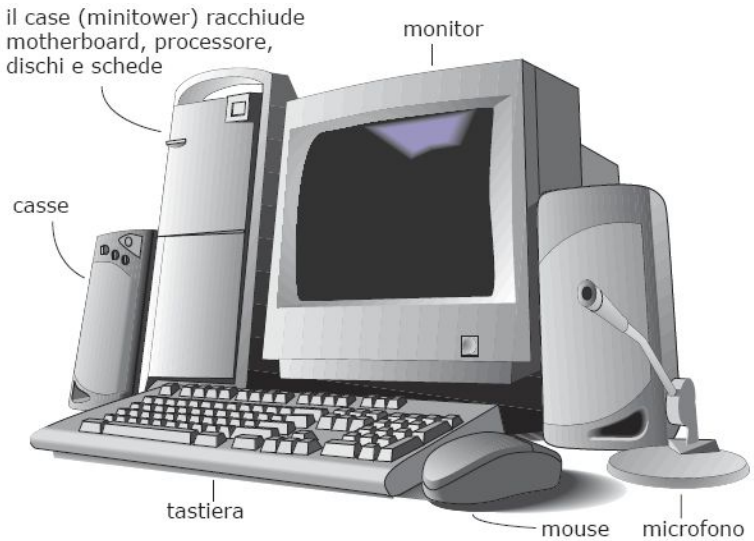
Computer


il case (minitower) racchiude
motherboard, processore,
dischi e schede
monitor
casse
tastiera
mouse
microfono

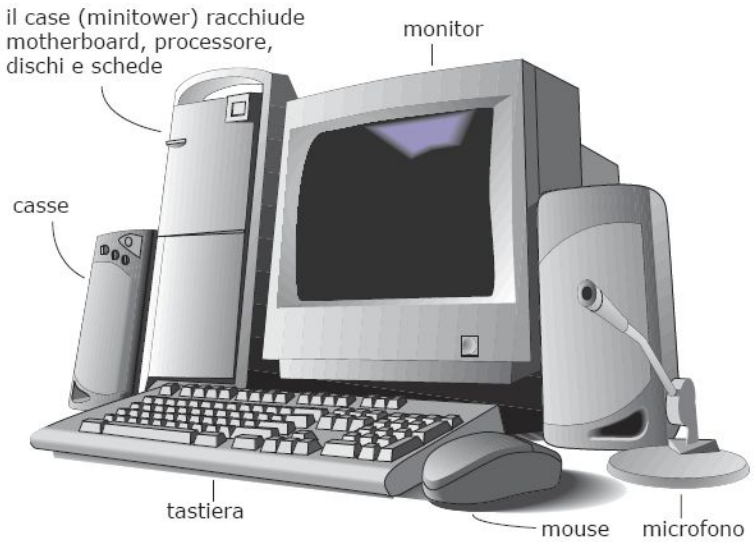
Computer

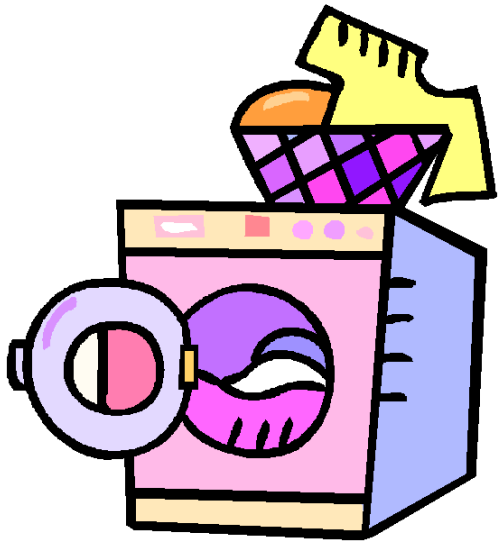

il case (minitower) racchiude
motherboard, processore,
dischi e schede
monitor
casse
tastiera
mouse
microfono

Computer


il case (minitower) racchiude
motherboard, processore,
dischi e schede
monitor
casse
tastiera
mouse
microfono

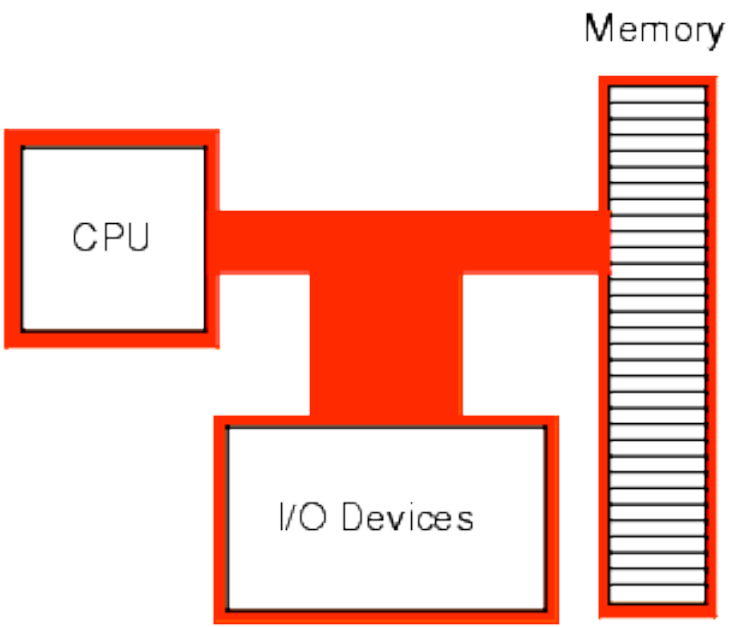
Computer


il case (minitower) racchiude
motherboard, processore,
dischi e schede
monitor
casse
tastiera
mouse
microfono



Un modello: la macchina di Von Neumann

Questi sistemi **molto** diversi fra loro possono essere *modellati* con uno schema molto semplice la macchina di Von Neumann

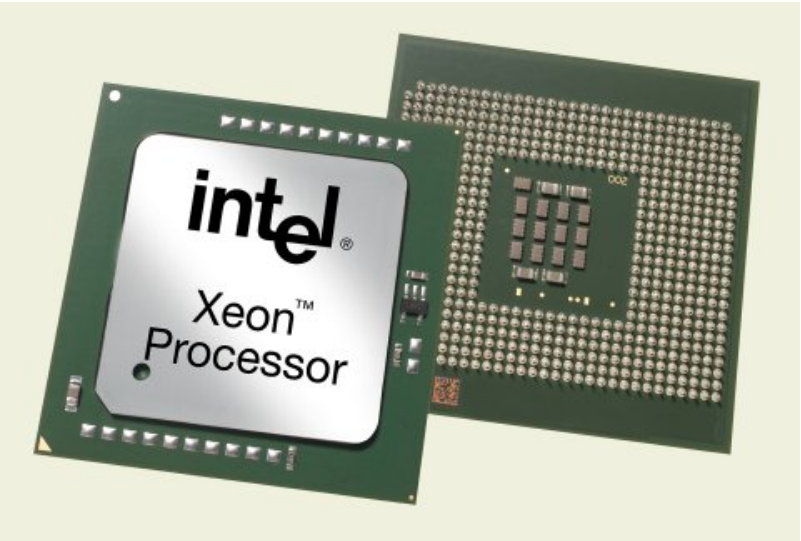


- un processore (CPU), in grado di operare elaborazioni di informazioni (segnali elettrici)
- una memoria, in grado conservare informazioni
- dispositivi periferici per trasdurre e attuare segnali elettrici verso il mondo esterno
- collegamenti diretti fra CPU e memoria e fra CPU e dispositivi (e in alcuni casi anche fra memoria e

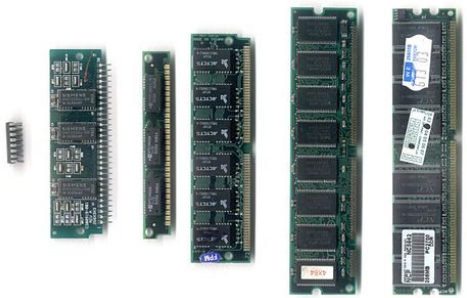
CPU

È in grado di eseguire istruzioni, in genere piuttosto semplici:

- p.es: attiva il segnale elettrico che simboleggia uno 0 in un certo componente elettronico
- piú comodo (e comprensibile) parlare metaforicamente: mette uno 0 in un *registro*
- somma il contenuto di due registri e mette il risultato nel terzo



Memoria



Memorizza segnali elettrici in componenti elettronici.

- È direttamente accessibile dal processore: ossia esistono istruzioni che possono riferirsi direttamente ai componenti della memoria (p.es. somma il contenuto di una *cella di memoria* con quello di un *registro*)
  - memoria centrale o primaria (RAM)
- Il segnale viene conservato solo fintanto che c'è corrente elettrica
  - memoria volatile (cioè non permanente)

Periferiche

Sono i dispositivi che permettono di interagire con il calcolatore

input forniscono dati da elaborare (trasdurre)

- tastiera
- mouse
- ...

output forniscono dati da interpretare nel mondo esterno (attuare)

- video
- speaker
- ...

input/output entrambe le cose

- touchscreen
- memorie di massa (permanenti!)
- schede di rete
- ...

Bus

Il bus sono collegamenti elettrici fra i componenti funzionali: sono ospitati sulla scheda madre

- Occorre stabilire le modalità con le quali si trasmette corrente elettrica sui *fili*
- p.es. PCI, SCSI, USB, Firewire, ...

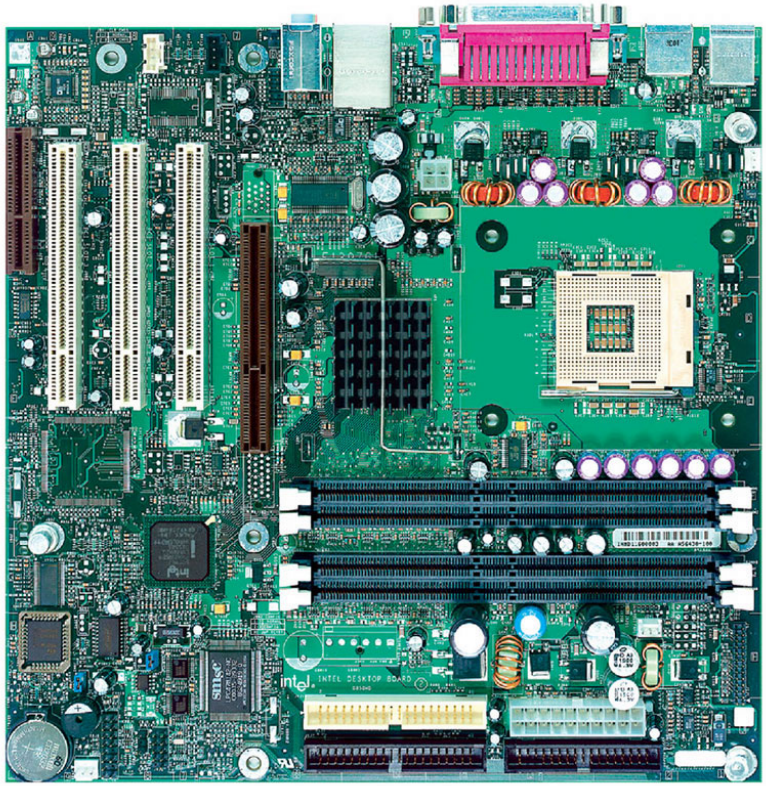


Figure 1.10. A computer motherboard.

I circuiti elettronici del processore

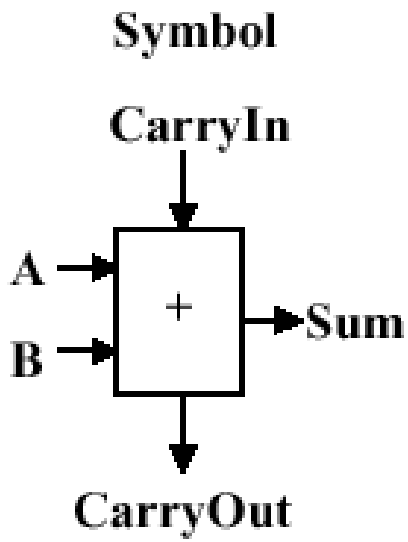
- Circuiti elettronici bistabili: due stati possibili, che convenzionalmente indichiamo con i simboli 0 e 1
- Non è una limitazione: qualsiasi informazione può essere rappresentata con una sequenza abbastanza lunga di simboli binari (rappresentazione digitale binaria)
- È molto facile definire le tabelle delle operazioni

	0	1
0	0	1
1	1	0^*

I circuiti elettronici del processore

- Circuiti elettronici bistabili: due stati possibili, che convenzionalmente indichiamo con i simboli 0 e 1
- Non è una limitazione: qualsiasi informazione può essere rappresentata con una sequenza abbastanza lunga di simboli binari (rappresentazione digitale binaria)
- È molto facile definire le tabelle delle operazioni

	0	1
0	0	1
1	1	0*



Definition

A	B	CarryIn	CarryOut	Sum
0	0	0	0	0
0	0	1	0	1
0	1	0	0	1
0	1	1	1	0
1	0	0	0	1
1	0	1	1	0
1	1	0	1	0
1	1	1	1	1

Ciclo di esecuzione

Le macchine del tipo Von Neumann funzionano semplicemente ripetendo ciclicamente queste tre operazioni

1. Fetch preleva una parola dalla memoria
2. Decode ogni sequenza di bit (parola) rappresenta un'operazione (e probabilmente i suoi operandi), quindi il processore decide quale istruzione deve essere attivata
3. Execute attiva l'istruzione, ottenendo i segnali elettrici risultanti

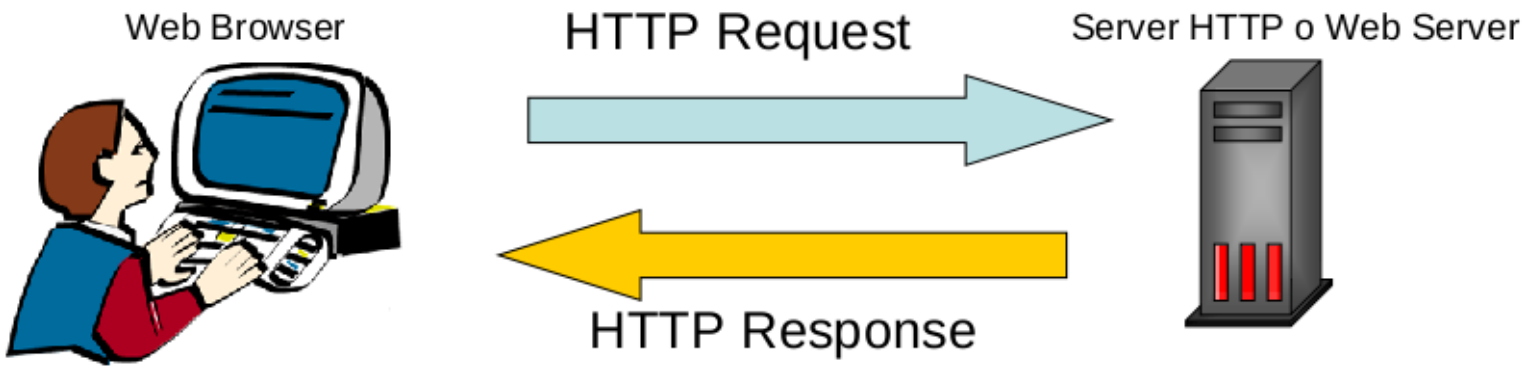
Queste operazioni vengono eseguite miliardi di volte al secondo (1GHz $= 10^9$ cicli al secondo).
L'insieme delle *parole di memoria* che vengono prelevate dal processore, costituiscono la parte variabile che determina il funzionamento di alto livello del sistema (software)

DICo

Lezione V: Testi formattati e collegamenti ipertestuali

Web browser e web server

Web browser e web server sono due programmi che comunicano (generalmente tramite Internet) utilizzando il protocollo HTTP.



Web browser

È il client della comunicazione e svolge principalmente 2 funzioni:

1. Richiede al web server (via HTTP) le risorse (tipicamente *file*), da visualizzare sullo schermo dell'utente del web browser
2. Interpreta le risorse ricevute, che generalmente contengono dati nel formato HTML, un linguaggio di *markup* che serve a descrivere testi contenenti elementi multimediali e collegamenti ipertestuali

Il browser è un programma molto complicato e c'è molto da scoprire! (vedi per esempio: http://www.20thingsilearned.com/)

HTML

DICo

HTML è un linguaggio formale che serve a descrivere ipertesti. Tale descrizione viene poi interpretata dal browser fornendo come risultato una pagina web.

- Faremo qualche esperimento con XHTML, una versione recente di HTML
- Utile a capire:
  - cosa vuole dire scrivere istruzioni che possano essere interpretate automaticamente da un computer
  - distinguere codice sorgente e visualizzazione
  - come funzionano in linea di principio i programmi di formattazione del testo

Esempio

```
1 <html>
2   <head>
3     <title>Esempio</title>
4   </head>
5   <body>
6     <p> Questo &egrave; un semplice esempio della
7     struttura base di un documento HTML. </p>
8   </body>
9 </html>
```

Terminologia

tag un marcatore, delimitato da < e > per distinguerlo dai caratteri del testo

elemento una coppia (non sempre!) di marcatori e il testo marcato. Esempio: <**p**>bla bla bla</**p**>, ma anche <**img src**="mattia.jpg" **alt**="Mattia" /> (si noti il carattere / verso la fine del marcatore per segnalare che non va cercato un marcatore finale.

attributi qualificazioni di un elemento: nell'esempio precedente src e alt sono attributi (il cui valore è fra virgolette)

entità un marcatore che denota un simbolo: per esempio È $\rightsquigarrow$ È.

Collegamenti

- Collegamenti con altre pagine web

  <**a href**="url">testo cliccabile</**a**>
- Collegamenti all'interno di una pagina
  1. <**a id**="#nome">punto a cui collegarsi</**a**>
  2. <**a href**="#nome">testo cliccabile</**a**>
- Collegamenti verso punti specifici di altre pagine web

  <**a href**="url#nome">testo cliccabile</**a**> (#nome deve esistere!)

Conformità alle regole

(X)HTML è un linguaggio formale: il significato di una frase dipende solo dalla sua *forma*. Due tipi di conformità; un testo HTML deve essere:

Ben formato

Gli elementi si *aprono* e *chiudono* e non si *accavallano*

```
1 <pippo>abc</pippo>
```

```
1 <h1>abc
```

```
1 <p><h1></p></h1>
```

Valido

I tag sono quelli previsti e sono *combinati* nel modo giusto.

Validità di un documento (cont.)

Come si fa ad accertarsi che un documento HTML sia valido?

- Esistono tante variazioni alle regole (HTML 3.2, HTML 4.01, XHTML 1.0, . . .)
- I browser cercano di interpretare anche documenti non del tutto validi (ma lo fanno in maniera imprevedibile!)

`http://validator.w3.org/`

Style sheet

- Separare struttura e contenuto e apparenza (stile) del documento

struttura	organizzazione logica del documento
contenuto	testo, immagini, ecc.
apparenza	spaziatura, margini, attributi di formattazione, layout

- Pensare allo stile comunicativo di una serie di pagine, che richiedono stile uniforme, e non solo alla singola pagina
- Lo stesso contenuto può apparire con stili differenti

Stile degli elementi

- Per ogni elemento è possibile specificare lo stile col quale visualizzarlo
- Viene usato un nuovo linguaggio: CSS (Cascading Style Sheets)
- http://www.w3.org/TR/REC-CSS2/

```
<style>
 h1 { /* elemento h1 */
      font-family: sans-serif;
      font-size: xx-large;
      text-align: center;
 }
 ul li { /* elemento li "figlio" di ul */
      font-size: medium;
      list-style-type: square;
 }
 </style>
```

Esercizio

- Scrivere un piccolo curriculum in XHTML. Guide utili:
  - Tutorial XHTML http://xhtml.html.it/guide/leggi/52/guida-xhtml/
  - Tutorial CSS http://css.html.it/guide/leggi/2/guida-css-di-base/
- Validare le pagine prodotte con validator.w3.org

Lezione VI: Rappresentazione dell'informazione

Informazione

L'informatica studia la rappresentazione e elaborazione dell'informazione. Per ora abbiamo solo accennato a come viene rappresentata l'informazione nei calcolatori:

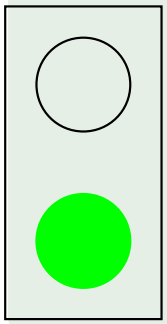
- Rappresentazione "numerica" (codifica digitale)
- In realtà i "numeri" sono solo un *alfabeto* di cifre usate per il loro valore simbolico (naturalmente è possibile codificare numeri, ma qualsiasi informazione viene rappresentata con il medesimo meccanismo)

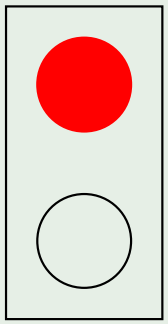
La codifica digitale rende l'elaborazione una pura manipolazione simbolica: da una sequenza di simboli di partenza si produce una nuova sequenza di simboli.

Esempio

Informazione

Un semaforo può essere $\{\texttt{rosso}, \texttt{verde}\}$

 $= 0$

 $= 1$

Elaborazione algoritmica

Se il semaforo è verde `attraversa` $(= 1)$,
altrimenti `non attraversare` $(= 0)$

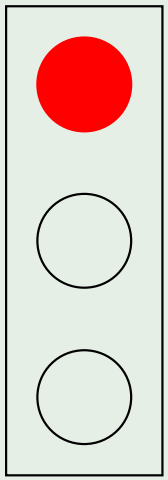
$0 \longrightarrow\!\!\!\!\!\rightarrow 1$

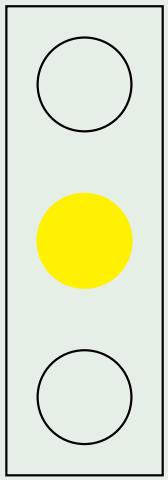
$1 \longrightarrow\!\!\!\!\!\rightarrow 0$

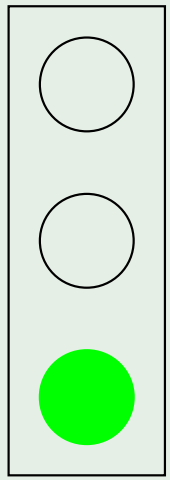
Esempio

Informazione

Un semaforo può essere {`rosso`, `verde`, `giallo`}.

 = 00

 = 01

 = 11

Elaborazione algoritmica

Se il semaforo è verde o giallo `attraversa` (= 1), altrimenti `non attraversare` (= 0)

$$00 \longrightarrow\!\!\!\!\!\rightarrow 0$$
$$01 \longrightarrow\!\!\!\!\!\rightarrow 1$$
$$11 \longrightarrow\!\!\!\!\!\rightarrow 1$$

Rappresentazione binaria dell'informazione

Le componenti hardware sono spesso *bistabili:* Rappresentare l'informazione con solo due simboli (rappr. binaria)
0 1
00 01 10 11
000 001 010 011 100 101 110 111

0000.0001.0010.0011.0100.0101.0110.0111.1000.1001.1010.1011.1100.1101.1110.1111
0 1 2 3 4 5 6 7 8 9 A B C D E F

- 1 cifra binaria viene detta bit
- 1 bit è la quantità di informazione che permette di discernere fra due alternative
- con n bit è possibile distinguere 2^n informazioni diverse, altrimenti equiprobabili
- Le stringhe di bit vengono spesso scritte in maniera piú compatta usando i 16 simboli 0123456789ABCDEF, ognuno dei quali rappresenta un gruppo di 4 bit (notazione esadecimale)

Esempi

- Quanti bit servono per codificare le sette note musicali?
- Quanti bit servono per distinguere gli abitanti del pianeta (circa 7 miliardi di persone)?
- Quanti bit servono per codificare i numeri da 0 a 255?
- Supponiamo che i nanetti di Biancaneve dicano sempre e solo la verità. Supponendo che inizialmente non abbia alcun sospetto specifico, quante domande sí/no dovrà fare Biancaneve per scoprire chi, fra i nanetti, s'è mangiato la torta?

Codifica binaria di numeri

La rappresentazione abituale dei numeri usa dieci simboli, secondo una notazione posizionale (numerazione decimale)

$$1035 = 1 \times 10^3 + 0 \times 10^2 + 3 \times 10^1 + 5 \times 10^0$$

Possiamo fare la stessa cosa con qualsiasi numero di simboli, per esempio 2 (numerazione binaria)

$$(1035)_{10} = (10000001011)_2$$

Trasformare da decimale a binario

Dato il numero che vogliamo convertire, dobbiamo:

- dividere per due e memorizzare il resto (il resto sarà 0 o 1)
- se il quoziente/quoto (il risultato della divisione precedente) non è uguale a zero, dividerlo per due e memorizzare il resto
- ripetere fintantoché il quoziente non diventa zero

Come rappresentare i numeri non interi? Come rappresentare i numeri negativi?

Codifica dei caratteri

ASCII	0000 0000	0000 0001	0000 0010	0000 0011	0000 0100	0000 0101	0000 0110	0000 0111	0000 1000	0000 1001	0000 1010	0000 1011	0000 1100	0000 1101	0000 1110	0000 1111
0000	$^{N}_{U}$	$^{S}_{H}$	$^{S}_{X}$	$^{E}_{X}$	$^{E}_{T}$	$^{E}_{Q}$	$^{A}_{K}$	$^{B}_{L}$	$^{B}_{S}$	$^{H}_{T}$	$^{L}_{F}$	$^{Y}_{T}$	$^{F}_{F}$	$^{C}_{R}$	$^{S}_{0}$	$^{S}_{I}$
0001	$^{D}_{L}$	$^{D}_{1}$	$^{D}_{2}$	$^{D}_{3}$	$^{D}_{4}$	$^{N}_{K}$	$^{S}_{Y}$	$^{E}_{\Sigma}$	$^{C}_{N}$	$^{E}_{M}$	$^{S}_{B}$	$^{E}_{C}$	$^{F}_{S}$	$^{G}_{S}$	$^{R}_{S}$	$^{U}_{S}$
0010		!	"	#	$	%	&	'	(	)	*	+	,	-	.	/
0011	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
0100	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
0101	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
0110	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
0111	p	q	r	s	t	u	v	w	x	y	z	{	|	}	~	$^{D}_{T}$
1000	$^{8}_{0}$	$^{8}_{1}$	$^{8}_{2}$	$^{8}_{3}$	$^{I}_{N}$	$^{N}_{L}$	$^{S}_{S}$	$^{E}_{S}$	$^{H}_{S}$	$^{H}_{J}$	$^{Y}_{S}$	$^{P}_{D}$	$^{P}_{V}$	$^{R}_{I}$	$^{S}_{2}$	$^{S}_{3}$
1001	$^{D}_{C}$	$^{P}_{1}$	$^{P}_{Z}$	$^{S}_{E}$	$^{C}_{C}$	$^{M}_{M}$	$^{S}_{P}$	$^{E}_{P}$	$^{Q}_{8}$	$^{Q}_{Q}$	$^{Q}_{A}$	$^{C}_{S}$	$^{S}_{T}$	$^{O}_{S}$	$^{P}_{M}$	$^{A}_{P}$
1010	$^{A}_{0}$	¡	¢	£		¥	¦	§	¨	©	♀	«	¬	-	®	¯
1011	°	±	²	³	´	µ	¶	·	¸	¹	♂	»	¼	½	¾	¿
1100	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
1101	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	β
1110	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
1111	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

Figura 6.6 Extended ASCII, l'American Standard Code for Information Interchange.

File di testo

- Un file di testo contiene solo caratteri (ASCII, ASCII esteso, UNICODE, . . .)
- Un file prodotto da un word processor, generalmente **non** è un file di testo (testo formattato)
- (Un file HTML **è** un file di testo, che viene interpretato dal browser per visualizzare un testo formattato)

Codifica di colori

Uno dei metodi piú usati è quello cosiddetto RGB (Red, Green, Blue).

- Ogni colore è la composizione di 3 livelli 0–255: uno si riferisce al rosso, uno al verde, uno al blu;
- Il colore finale è ottenuto secondo il "modello additivo" dei colori primari (p.es. rosso + verde = giallo)
- Ogni livello è codificato indipendentemente come un numero (< 256 quindi bastano 8 bit)
- Gli 8 bit vengono generalmente scritti come due cifre esadecimali ($204 \rightsquigarrow$ CC)

Esempi: http://www.visibone.com/colorlab/

Lezione VII: Fogli elettronici

Fogli di calcolo

I fogli di calcolo (spreadsheet) sono una delle applicazioni piú antiche e di maggior successo dei calcolatori.

- Il primo foglio di calcolo simile a quelli che usiamo ora è stato VisiCalc (anni '70–'80), poi Lotus 1-2-3, Excel, Gnumeric, OpenOffice Calc, ...
- Al giorno d'oggi è lo strumento di calcolo personale piú utilizzato (spesso impropriamente: come vedremo in molte situazioni un database manager è molto piú appropriato), soprattutto da chi non ha una formazione informatica

Le idea di base (comuni a tutti i fogli di calcolo)

- Lo spazio di lavoro rettangolare è diviso in celle, ciascuna contrassegnata dalla colonna (tramite una lettera) e dalla riga (tramite un numero) cui appartiene (es. A6, prima colonna da sinistra, sesta riga dall'alto)
- le celle possono contenere dati o formule: in quest'ultimo caso verrà visualizzato il risultato della formula (aggiornato ogni volta che i dati coinvolti cambiano)
  - in A2: `mamma`, in B2: 3
  - in C2: `=concatena(A2;B2)`, visualizza `mamma3`
- le celle hanno un formato di visualizzazione
  - 3 visualizzato con formato Data, diventa `martedí 2 gennaio 1990` (ma sempre 3 rimane in eventuali calcoli)
- Una cella vuota è come se contenesse il dato 0

Sono possibili le consuete operazioni aritmetiche con i simboli infissi `+ - * /` (es. `=B2*B2+33`) e centinaia di formule matematiche, statistiche, di elaborazione testuale, ... (es. `=max(A2;A4;B5)`)

Alcune comodità nell'utilizzo

- Le formule possono fare riferimento a intervalli rettangolari di celle: basta scrivere la cella in alto a sx e quella in basso a dx separate da : (es. `=max(A4:B9)`)
- Le formule possono essere copiate mantenendo i riferimenti relativi
  - in C2: `=concatena(A2;B2)`
  - copiamo la formula in D5, quindi operando uno spostamento relativo di 1 colonna (da C a D) e 3 righe (da 2 a 5)
  - in D5 risulterà: `=concatena(B5;C5)`
- È possibile indicare un riferimento (ad una colonna o riga) assoluto con il simbolo $ (es. `=concatena($A2;B$2)`, in questo caso la copia in D5 risulterà in `=concatena($A5;C$2)`)

Esercizio

http://homes.dico.unimi.it/~monga/lib/biotecno/esexcel.html

DICo

Lezione VIII: Rappresentazione dell'informazione (2)

Immagini digitali

Un'immagine digitale è una rappresentazione (numerica!) di una figura bidimensionale

formato bitmap o raster : immagine divisa in punti (pixel, *picture element*). Il codice rappresenta il colore del punto (p.es. in RGB: Light Orange-Yellow= $(255_{10},204_{10},102_{10}) = FFCC66_{16} = 16764006_{10} = 111111111100110001100110_2$)

formato vettoriale : immagine è codificata come un insieme di linee e figure geometriche (p.es. retta((0,0);(0;3)); cerchio((0,1),4) "retta" e "cerchio" sono come *comandi* di un linguaggio di descrizione e saranno anch'essi codificati numericamente)

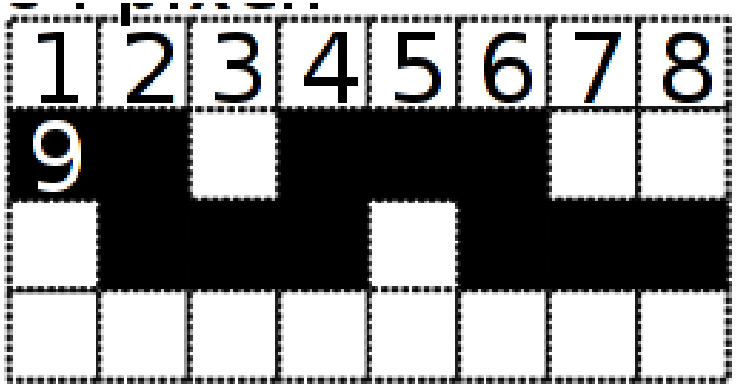
Risoluzione

Si chiama risoluzione il numero di pixel disponibili per codificare un'immagine

- spesso dato come *geometria* della figura (1024×768)
- nelle fotocamere in megapixel
- non è un buon indicatore della qualità di uno schermo o di una stampa: si preferisce la densità di pixel (pixel per inch, PPI, dot per inch, DPI) ossia il numero di pixel per unità di superficie (es. netbook con schermo da $10{,}1''$ $1024 \times 768 \rightsquigarrow 127\text{PPI}$)

Figura 1.2 Un ingrandimento della matrice di bit visualizzata sul monitor e i bit corrispondenti per ogni pixel.

Codifica raster

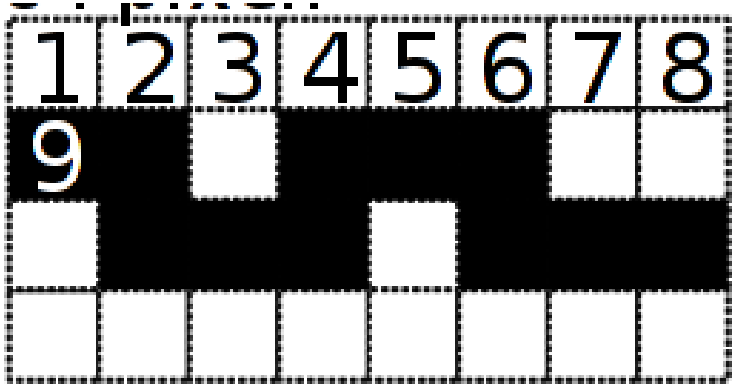


- L'immagine viene divisa in una griglia di pixel e numeriamo i pixel
- Per ogni pixel mettiamo 0/1

Se volessi dei colori, 1 bit non è piú sufficiente:

- 65535 colori $\rightsquigarrow$ 16 bit
- RGB (24 bit) $\rightsquigarrow$ 16777216 colori
- il numero di bit è detto profondità del colore

Codifica raster



- L'immagine viene divisa in una griglia di pixel e numeriamo i pixel
- Per ogni pixel mettiamo 0/1
- 0000 0000 1101 1100 0111 0111 0000 0000

Se volessi dei colori, 1 bit non è piú sufficiente:

- 65535 colori $\rightsquigarrow$ 16 bit
- RGB (24 bit) $\rightsquigarrow$ 16777216 colori
- il numero di bit è detto profondità del colore

Occupazione di un'immagine

- Una fotografia digitale da 5 megapixel
- Per ogni pixel, 24 bit $=$ 3 byte
- $5 \times 10^6 \times 3 = 15 \cdot 10^6$ 15MB

In realtà generalmente occupano molto meno spazio sulla memoria di massa. Come mai?

Codifiche compresse

Immaginate di dover codificare un testo: “Trentatré trentini entrarono in Trento trotterellando”. Possiamo codificare ogni carattere: 8 bit per 57 caratteri.

- Potremmo codificare in maniera diversa le parti che si ripetono (per esempio “rent”): in questo modo i bit necessari potrebbero ridursi
  - compressioni precise (senza perdita di informazione, lossless: viene codificata tutta l’informazione sfruttando proprietà statistiche, per esempio ripetizioni.
  - compressioni imprecise (con perdita di informazione, lossy: viene codificata soltanto parte dell’informazione (non è quindi possibile risalire all’informazione *originale*) sfruttando proprietà nella fruizione, per esempio un pixel rosa vicino a tanti rossi potrebbe essere praticamente invisibile per un occhio umano.

Codifiche compresse

I formati JPEG e MPEG (video) sono formati lossy. Il formato ZIP è un formato compresso (generico) lossless.

- .bmp 3MB
- .jpg 176KB
- .zip 1.4MB

Formati vettoriali

Diagrammi, planimetrie, disegni tecnici in generale vengono memorizzati in formato vettoriale

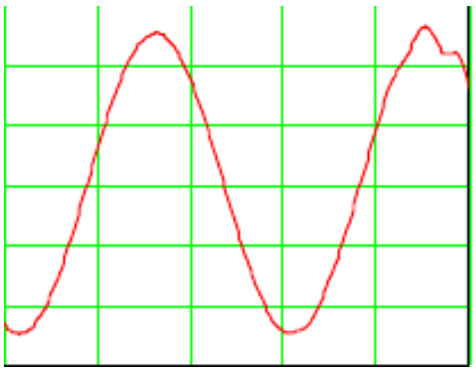
+ se e semplice, molto meno spazio
+ si può ingrandire mantenendo la stessa qualità
- non si adattano a rappresentare immagini generiche, come le fotografie

Rappresentazione digitale di segnali continui

Il primo passaggio nell'operazione di digitalizzazione consiste nell'individuare come rendere discreta un fenomeno che è continua

- La realtà che rappresentiamo in un'immagine non è divisa in pixel
- Un suono è una grandezza continua, p.es. la frequenza di vibrazione di una corda

(In matematica un insieme totalmente ordinato è continuo se tra due elementi ne esiste sempre uno compreso. Un insieme non continuo è detto discreto)



Livelli di digitalizzazione (profondità) e frequenza di digitalizzazione

Il problema del campionamento

- quanto frequente deve essere la campionatura? (40.000 Hz (Hz=1/s)
- quanti bit vogliamo usare per ogni campione? (CD audio, 2 byte)
- quanti sono i campioni? (la durata dell'audio (in secondi) per la frequenza)
- che dimensione occupa un file audio?

Un'ora di musica

- 60s per ogni minuto
- 44.100 campioni da 16 bit ciascuno
- 60s $\times$ 44.100 (1/s) $\times$ 16 (bit) $\times$ 2
  $= 84672000 bit = 10584000 byte = 10.5$MB

Un'ora di musica: 635 MB! Formati con compressione lossy MP3, OGG...

Lezione IX: Fogli elettronici (2)

Aiutiamo Babbo Natale!

Babbo Natale riceve molte richieste di regali e per non dimenticarsene le fa memorizzare in un file di "comma separated values" (CSV), un formato testuale molto utile per conservare tabelle di dati: essendo un formato *solo testo* è molto facile utilizzarlo su sistemi anche molto diversi.

```
"Provincia","Nome","Anno"
"TO","Bambola",2009
....
```

Tutti gli spreadsheet importano facilmente i csv.

Babbo Natale (cont.)

DICo

Babbo Natale si è reso conto che nel 2009 ha ricevuto poche richieste di libri. Per aumentarle, durante il 2010 ha incaricato la Befana di svolgere una campagna pubblicitaria per aumentare l'interesse dei bambini verso i libri. Tale campagna è stata svolta nella provincia di Milano.
Dopo aver ricevuto le richieste dei regali per il 2010, vuole vedere se tale campagna ha avuto il risultato atteso (aumento delle richieste di libri nella provincia di Milano)

- A tale scopo, confronta come è variata la percentuale di richieste di libri (rispetto al totale delle richieste) tra il 2009 e il 2010 nella provincia di Milano
- Come fareste per convincere Babbo Natale che l'analisi è convincente, ossia che il risultato ottenuto a Milano è proprio *causato* dalla campagna pubblicitaria delle Befana e non, per esempio, da qualche altra influenza a livello nazionale?

Babbo Natale (cont.)

Sfortunatamente gli elfi addetti alla lettura delle letterine e all'inserimento delle richieste nella base di dati non hanno catalogato i giochi in base alla categoria.

- Dunque non troverai nella base di dati un attributo con alcuni valori impostati a "libro"
- Per svolgere l'esercizio devi cercare, all'interno della descrizione della richiesta, se compare la parola "libro"

Suggerimenti

- Trova la funzione che cerca una parola dentro un testo
- Utilizza la funzione `VAL.ERR` per distinguere quando la parola viene trovata o non viene trovata.
- Per combinare i valori logici (vero, falso) sono utili le funzioni *logiche*:
  - `NON` x è vero se e solo se x è falso
  - `E` x,y è vero se e solo se sia x che y sono veri
  - `O` x,y è vero se e solo se almeno uno fra x e y è vero

commenti

- Questo esercizio mette in evidenza un problema molto comune: la cattiva organizzazione dell'informazione può rendere piú complicata del necessario il trattamento automatico dell'informazione.
- In effetti, se fosse stata prevista l'informazione "tipologia di regalo" associata ad ogni dono, il problema si sarebbe potuto risolvere in maniera molto piú semplice.
- La soluzione adottata in questo esercizio può andare bene in molti casi, ma per alcuni nomi di giochi potrebbe portare ad errori. Riesci a trovare un esempio di nome di gioco per il quale la soluzione proposta non è corretta?

Lezione X: Basi di dati

Dati in forma tabellare

Conservare i dati in forma tabellare è molto comune ed intuitivo: ciò spiega il successo dei fogli di calcolo.
Nei fogli di calcolo, però:

- Ogni cella è trattata singolarmente
- Il formato dei dati è descritto "per cella"

Per organizzare grandi insiemi di dati servono forme piú strutturate.

Le basi di dati

Le basi di dati (o database, DB) servono per organizzare, memorizzare e facilitare l'elaborazione automatica dell'informazione.
Il tipo di basi di dati piú comune è quello basato su dati in forma tabellare (basi di dati relazionali, proposte per la prima volta da E. Codd negli anni '70).
Un DB è gestito tramite un "Database management system" (DBMS).

Basi di dati relazionali

Nei DB relazionali si conservano:

Entità

Le entità denotano i dati da memorizzare: una persona, un'automobile, una lettera a Babbo Natale, ecc.

Attributi

Ogni entità è caratterizzata dagli attributi che si vogliono conservare: per una persona potrebbero essere il nome, il cognome, il codice fiscale, ecc.

Inoltre le entità possono essere messe in relazione fra loro (p.es. Una persona possiede un'automobile): per capire come dobbiamo esaminare meglio come vengono memorizzate le entità.

Esempio

Entità automobile:

modello	“Fiat Panda NP”
colore	“Giallo”
targa	“AB768BA”
immatricolazione	“1/4/2003”
passeggeri	4

Ogni automobile registrata nel database è detta record del DB.
In forma tabellare gli attributi sarebbero probabilmente le colonne, e ogni record una riga.
Gli attributi sono *atomici:* se volessi considerare il mese e il giorno di immatricolazione separatamente dovrei avere altri attributi.

Dati e metadati

La struttura di ogni record è predefinita e fissa: è decisa dal progettista del DB in funzione dell'uso dei dati previsto.

- Ogni record ha un insieme fisso (tupla) di attributi: automobile $\rightsquigarrow \langle$ modello, colore, targa, immatricolazione, passeggeri $\rangle$
- Per ogni attributo è definito il *dominio:* modello $\rightsquigarrow$ stringa di max 30 caratteri, targa $\rightsquigarrow$ stringa di esattamente 7 caratteri, passeggeri $\rightsquigarrow$ numero intero positivo

Queste sono *informazioni sui dati:* metadati.

Esemplari di dati

I metadati definiscono il *tipo* di dato che verrà conservato nel database. Ciascun record dovrà essere un esemplare conforme al tipo definito dai metadati.
In gergo si dice istanza (instance) del tipo definito dall'entità.
Es: il record della Panda è un'*istanza* dell'entità automobile.

Chiave

Esiste un insieme di attributi in grado di identificare univocamente un record?
Se c'è viene detto chiave candidata e normalmente sono piú d'uno. Se ne sceglie uno: chiave primaria. Generalmente è opportuno scegliere insiemi di attributi il piú piccoli possibile, spesso con un solo attributo.
Nel caso dell'automobile potrebbe essere la targa.

Relazione

Una relazione può essere memorizzata con una tabella che associ le chiavi primarie delle entità coinvolte:

- Entità automobile (chiave: targa)
- Entità persona (chiave: codice-fiscale)
- Relazione appartiene-a $\langle$ automobile.targa, persona.codice-fiscale $\rangle$

Generalmente si hanno relazioni *binarie* e possono esserci relazioni 1 a 1, 1 a molti, molti a molti.

Operazioni su dati tabellari

Un insieme di tre semplici operazioni è sufficiente per estrarre i dati da un DB senza limitazioni ai criteri di estrazione.

selezione Un sottoinsieme di righe che rispettano un determinato criterio

proiezione Un sottoinsieme di colonne

prodotto cartesiano La tabella ottenuta combinando tutte le righe della prima con tutte le righe della seconda

Selezione

Superficie	Piano	Stanze	Indirizzo	Box
100	4	4	Via Rossi, 5	No
50	4	2	Via Verdi, 4	Sì
150	8	7	Via Bianchi, 9	Sì

criterio di selezione: Piano $=$ 4

Superficie	Piano	Stanze	Indirizzo	Box
100	4	4	Via Rossi, 5	No
50	4	2	Via Verdi, 4	Sì

I criteri di selezione possono essere combinati con le solite funzioni logiche

- **not** trasforma una condizione vera in una falsa, e viceversa
- **and** vera solo se tutte le condizioni sono vere
- **or** vera se almeno una condizione è vera

not (Piano $=$ 8) and ((stanze $>$ 2) or (superficie $\geq$ 75))

Proiezione

Superficie	Piano	Stanze	Indirizzo	Box
100	4	4	Via Rossi, 5	No
50	4	2	Via Verdi, 4	Sì
150	8	7	Via Bianchi, 9	Sì

proiezione: indirizzo, box

Indirizzo	Box
Via Rossi, 5	No
Via Verdi, 4	Sì
Via Bianchi, 9	Sì

Prodotto cartesiano

Si applica a due tabelle ed è analogo al prodotto cartesiano fra insiemi

$$\{a,b\} \times \{c,d,e\} = \{\{a,c\},\{a,d\},\{a,e\},\{b,c\},\{b,d\},\{b,e\}\}$$

Nome	Anni	Genere
Andrea	2	Maschio
Luca	3	Maschio
Anna	4	Femmina

Gioco	Colore	GiocoDa
Macchinina	blu	Maschio
Bambola	rosa	Femmina

Nome	Anni	Genere	Gioco	Colore	GiocoDa
Andrea	2	Maschio	Macchinina	blu	Maschio
Andrea	2	Maschio	Bambola	rosa	Femmina
Luca	3	Maschio	Macchinina	blu	Maschio
Luca	3	Maschio	Bambola	rosa	Femmina
Anna	4	Femmina	Macchinina	blu	Maschio
Anna	4	Femmina	Bambola	rosa	Femmina

Join

È frequentissimo fare il prodotto e poi una selezione.

join naturale

Si moltiplicano due tabelle A e B e si selezionano solo le righe per le quali un attributo (specificato) di A ha valore uguale ad un attributo (specificato) di B

Join (cont.)

Nome	Anni	Genere	Gioco	Colore	GiocoDa
Andrea	2	Maschio	Macchinina	blu	Maschio
Andrea	2	Maschio	Bambola	rosa	Femmina
Luca	3	Maschio	Macchinina	blu	Maschio
Luca	3	Maschio	Bambola	rosa	Femmina
Anna	4	Femmina	Macchinina	blu	Maschio
Anna	4	Femmina	Bambola	rosa	Femmina

join con Bambini.genere = Regalo.GiocoDa

Nome	Anni	Genere	Gioco	Colore	GiocoDa
Andrea	2	Maschio	Macchinina	blu	Maschio
Luca	3	Maschio	Macchinina	blu	Maschio
Anna	4	Femmina	Bambola	rosa	Femmina

DICo

Lezione XI: Basi di dati (cont.)

Database relazionali

- I dati sono conservati in forma tabellare: ogni riga rappresenta una delle istanze (record) di un'entità da memorizzare, le colonne sono gli attributi memorizzati;
- Un insieme di attributi in grado di identificare univocamente un record si dice chiave.
- I dati possono essere identificati utilizzando tre semplici operazioni algebriche
  1. selezione
  2. proiezione
  3. prodotto cartesiano
- A queste generalmente si aggiunge il join, una selezione di un prodotto cartesiano in cui si prendono in considerazione solo i record correlati da una qualche chiave comune.

Progettazione di DB

Quando si progetta un database relazionale, occorre definire le entità da memorizzare e quali attributi conservare.
Un obiettivo importante è quello di evitare ridondanze nei dati:

- le stesse informazioni vengono memorizzate piú volte: spreco di spazio di memorizzazione
- ogni volta che un dato viene cambiato occorre aggiornare tutte le copie: incoerenza dei dati

Ridondanza

L'Università mantiene una tabella degli studenti iscritti:
`(matricola, nome, cognome, indirizzo, corso)`
La biblioteca tiene un elenco degli studenti e dei libri in prestito
`(matricola, nome, cognome, indirizzo, tessera)`
`(matricola, libro)`
Che succede se uno studente cambia indirizzo? Occorre assicurarsi che venga cambiato nelle due tabelle!
Invece:
`(matricola, corso)`
`(matricola, nome, cognome, indirizzo)`
`(matricola, tessera)`
`(matricola, libro)`
A questo punto il join diventa molto utile!

SQL

I DB vengono interrogati (query) per conoscere i dati che rispettano determinati criteri.
SQL (Structured Query Language) è un linguaggio formale per l'interrogazione di DB relazionali, molto diffuso (dagli anni '70 del '900).

SELECT attributi **FROM** prodotto cartesiano di tabelle **WHERE** condizione;

In pratica con **SELECT** si fa la proiezione (sic!) e con **WHERE** la selezione **FROM** un insieme ottenuto come prodotto cartesiano di tabelle.

SELECT nome, cognome **FROM** studente **WHERE** matricola $>$ 50000;

SQL esempi

Tabella Docenti: (identificativo, nome, cognome)
Tabella Esami: (nomeEsame, docente)

```
1 SELECT nome, cognome, nomeEsame
2 FROM Docenti, Esami
3 WHERE Docenti.identificativo = Esami.docente;
```

Esempi

Tabella World: (name, region, area, population, gdp)
Esempio: (“Afghanistan”, “South Asia”, 652225, 26000000)

```
1  SELECT population FROM World
2    WHERE name = 'France';
3
4  SELECT name, population/area FROM World
5    WHERE area > 5000000;
6
7  SELECT name , region
8    FROM World
9    WHERE population < 2000000
10     AND gdp > 5000000000;
11
12 SELECT name, population FROM World
13   WHERE name = 'Ireland' OR name = 'Iceland' OR name = 'Denmark');
14
15 SELECT name FROM World
16   WHERE name LIKE 'D%';
```

Esempi

Tabella PingPongWinners: (games, color, who, country)
Esempio: (1988, “gold”, “Yoo Nam-Kyu”, “KOR”)
Tabella Country: (id, name)
Esempio: (“ALG”, “Algeria”)

```
1 SELECT who, Country.name
2   FROM PingPongWinners JOIN Country
3           ON (PingPongWinners.country=Country.id)
4  WHERE games = 2000;
5
6 SELECT who, Country.name
7   FROM PingPongWinners, Country
8  WHERE PingPongWinners.country=Country.id AND games = 2000;
```

Lezione XII: Database

Acquisire dimestichezza con un DBMS

- Utilizzeremo il DBMS OpenOffice Base
- Scaricate e aprite il file `EsDB.odb`
- Esplorate l'applicazione Base di OpenOffice (o LibreOffice: la nuova versione si chiamerà cosí) e cercate di capire quali funzionalità rende disponibili
  - modificate tutto quello che volete, non temete di fare danni;
  - per ricominciare da zero, ri-scaricate il file d'esempio

Modifica dei dati

- Aggiungete un nuovo studente tra gli studenti
- Modificate il cognome di un docente
- Rimuovete un esame sostenuto

Lo schema del DB

- Trovate il comando per visualizzare le relazioni tra tabelle studia bene le relazioni
- Cosa vuol dire il simbolo a forma di chiave?
- Provate a guardare nuovamente il contenuto delle tabelle tenendo al contempo sott'occhio le relazioni: cosa si può dire sulla struttura della base di dati?

Interrogazioni

- Trovare il nome dei corsi tenuti da ogni docente
- Trovare i voti ottenuti dagli studenti con cognome `Bianchi`
- Trovare il nome del corso degli esami sostenuti dagli studenti con cognome `Bianchi`
- Trovare il nome dei docenti, il nome del corso e il voto d'esame degli studenti con cognome `Bianchi`
- Calcolare la media dei voti degli studenti con cognome `Bianchi`

Lezione XIII: Sicurezza informatica

Cos'è la sicurezza informatica

- Confidenzialità
- Integrità
  - . . . dei dati (integrità)
  - . . . dell'origine (autenticazione)
- Disponibilità

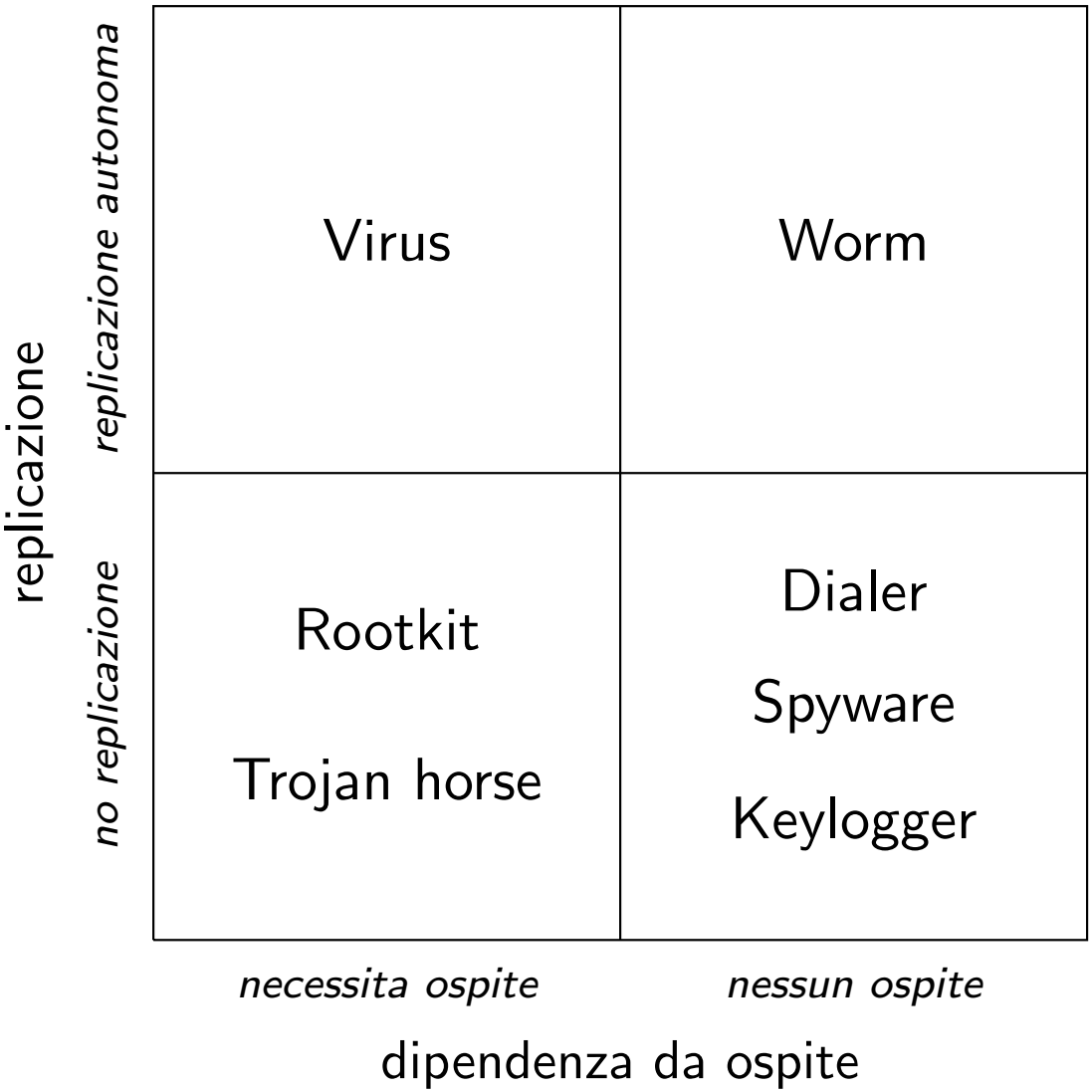
C'è davvero bisogno di parlarne?

Il software dannoso per le nostre attività è sempre piú diffuso, tanto da meritare un nome

Malware

sequenza di codice progettata per danneggiare intenzionalmente un sistema, i dati che contiene o comunque alterare il suo normale funzionamento, all'insaputa dell'utente

Tipologie di malware



Trojan horse

- funzionalità maligne cammuffate tra altre benigne
- propagazione manuale: diffusione di applicazioni con funzionalità "secondarie" o inserimento di nuove funzionalità in applicazioni esistenti
- rientrano in questa categoria *adware* e *spyware*

Tipologie di malware
Trojan horse & backdoor

DICo

Trojan horse

- funzionalità maligne cammuffate tra altre benigne
- propagazione manuale: diffusione di applicazioni con funzionalità "secondarie" o inserimento di nuove funzionalità in applicazioni esistenti
- rientrano in questa categoria *adware* e *spyware*

Backdoor

- per assicurare l'accesso ad un sistema compromesso
- rientrano in questa categoria i *RAT* (*Remote Access Trojan*)

Tipologie di malware
Trojan horse & backdoor

Trojan horse

- funzionalità maligne cammuffate tra altre benigne
- propagazione manuale: diffusione di applicazioni con funzionalità "secondarie" o inserimento di nuove funzionalità in applicazioni esistenti
- rientrano in questa categoria *adware* e *spyware*

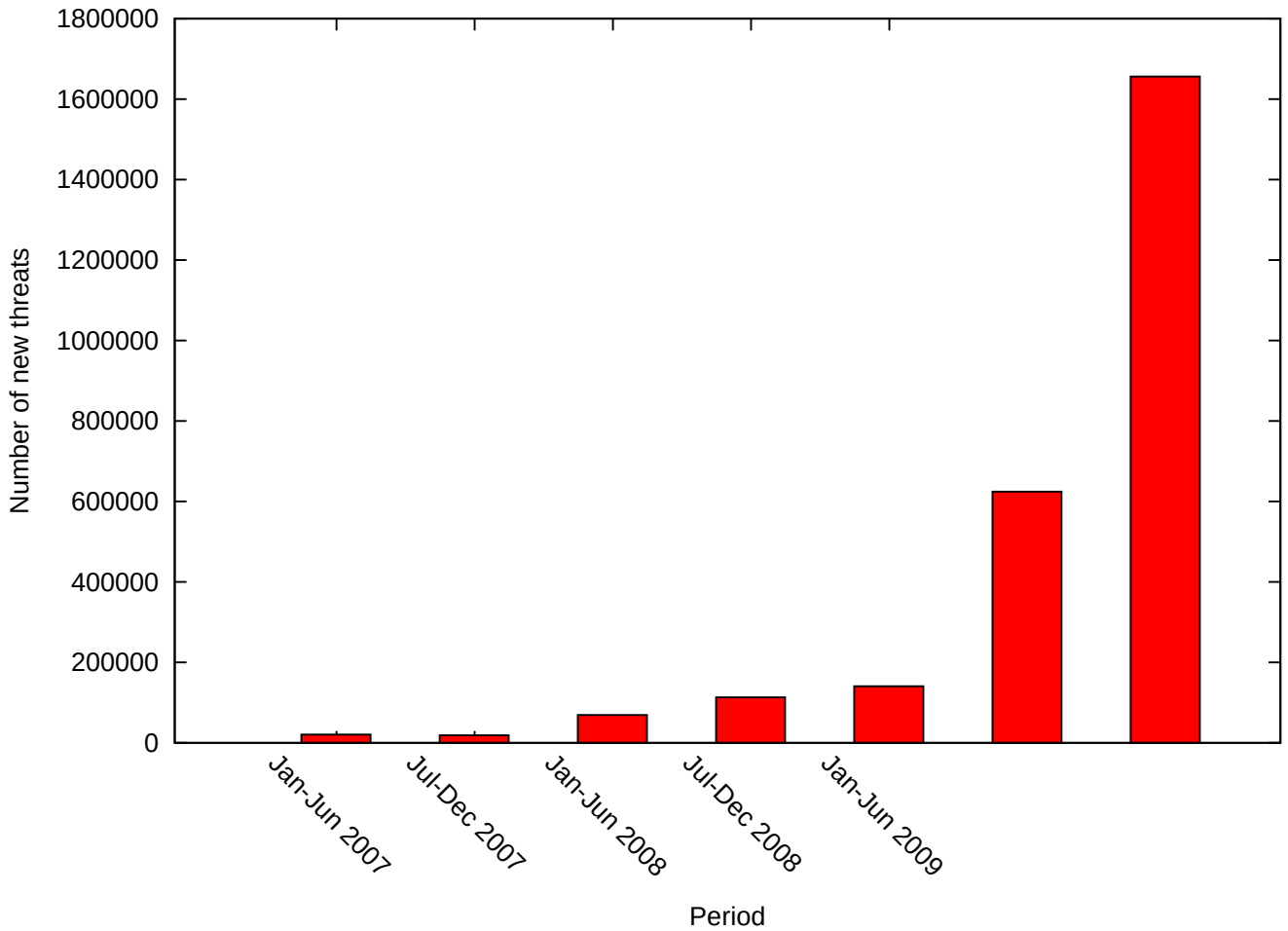
Backdoor

- per assicurare l'accesso ad un sistema compromesso
- rientrano in questa categoria i *RAT* (*Remote Access Trojan*)

Rootkit

- strumenti utilizzati per mantenere l'accesso ad un sistema compromesso senza fare nascere sospetti
- utilizzati per nascondere file, processi, connessioni di rete, ...
- sia a livello kernel che a livello utente

Nuove minacce



Fonte: Symantec

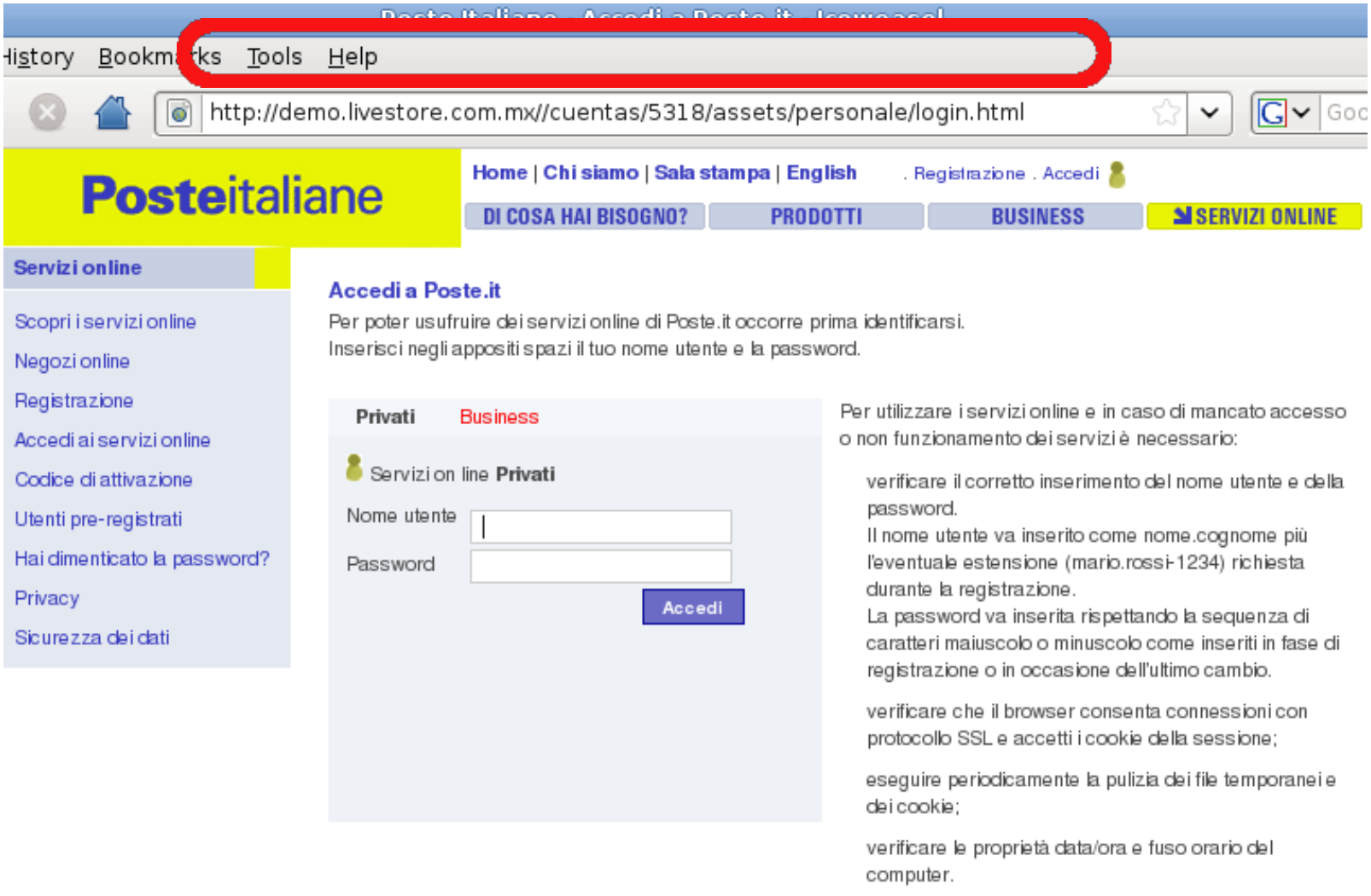
DICo

Malware & underground economy

Phishing

Poste Italiane - Accedi a Poste.it - Iceweasel
History Bookmarks Tools Help
http://demo.livestore.com.mx//cuentas/5318/assets/personale/login.html
Posteitaliane
Home | Chi siamo | Sala stampa | English
. Registrazione . Accedi
DI COSA HAI BISOGNO?
PRODOTTI
BUSINESS
SERVIZI ONLINE
Servizi online
Scopri i servizi online
Negozi online
Registrazione
Accedi ai servizi online
Codice di attivazione
Utenti pre-registrati
Hai dimenticato la password?
Privacy
Sicurezza dei dati
Accedi a Poste.it
Per poter usufruire dei servizi online di Poste.it occorre prima identificarsi.
Inserisci negli appositi spazi il tuo nome utente e la password.
Privati
Business
Servizi on line Privati
Nome utente
Password
Accedi
Per utilizzare i servizi online e in caso di mancato accesso o non funzionamento dei servizi è necessario:
verificare il corretto inserimento del nome utente e della password.
Il nome utente va inserito come nome.cognome più l'eventuale estensione (mario.rossi-1234) richiesta durante la registrazione.
La password va inserita rispettando la sequenza di caratteri maiuscolo o minuscolo come inseriti in fase di registrazione o in occasione dell'ultimo cambio.
verificare che il browser consenta connessioni con protocollo SSL e accetti i cookie della sessione;
eseguire periodicamente la pulizia dei file temporanei e dei cookie;
verificare le proprietà data/ora e fuso orario del computer.

Phishing

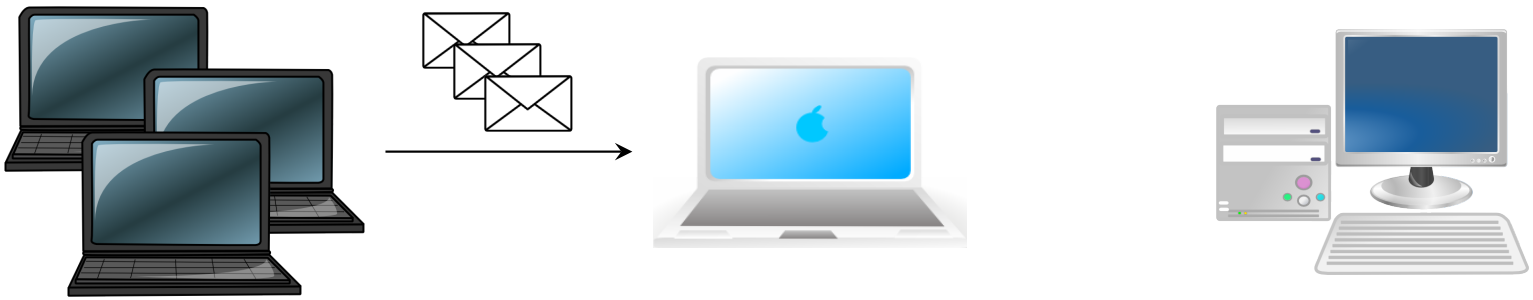

History Bookmarks Tools Help
http://demo.livestore.com.mx//cuentas/5318/assets/personale/login.html
Posteitaliane
Home | Chi siamo | Sala stampa | English
. Registrazione . Accedi
DI COSA HAI BISOGNO?
PRODOTTI
BUSINESS
SERVIZI ONLINE
Servizi online
Scopri i servizi online
Negozi online
Registrazione
Accedi ai servizi online
Codice di attivazione
Utenti pre-registrati
Hai dimenticato la password?
Privacy
Sicurezza dei dati
Accedi a Poste.it
Per poter usufruire dei servizi online di Poste.it occorre prima identificarsi.
Inserisci negli appositi spazi il tuo nome utente e la password.
Privati
Business
Servizi on line Privati
Nome utente
Password
Accedi
Per utilizzare i servizi online e in caso di mancato accesso o non funzionamento dei servizi è necessario:
verificare il corretto inserimento del nome utente e della password.
Il nome utente va inserito come nome.cognome più l'eventuale estensione (mario.rossi-1234) richiesta durante la registrazione.
La password va inserita rispettando la sequenza di caratteri maiuscolo o minuscolo come inseriti in fase di registrazione o in occasione dell'ultimo cambio.
verificare che il browser consenta connessioni con protocollo SSL e accetti i cookie della sessione;
eseguire periodicamente la pulizia dei file temporanei e dei cookie;
verificare le proprietà data/ora e fuso orario del computer.

Scam

Canadian Healthcare
Casa
Bestseller
Tutti i prodotti
FAQ
Contattaci
FREE VIAGRA
For Order more than $300
12 VIAGRA PILLS
For Other Orders
4 VIAGRA PILLS
VIAGRA
Cialis
$69.99
+ 4 Free pills
10 pills x 100 mg
10 pills x 20 mg
VIAGRA + CIALIS
Viagra + Cialis
€59.62
Cialis
€169.00
FAQ
Last viewed items
Denmark
Home
Bestsellers
All products
FAQ
Contact us
USD EUR GBP
CAD AUD CHF
Pharma Bonus
Your cart: $0.00 (0 items)
Proceed to Checkout
Canadian Pharmacy
#1 Internet Online Drugstore
Special Offer
Free Viagra samples
4 pills for every order
12 pills for order >$300
Product list
For Order more than $300:
12 VIAGRA PILLS
FREE
For other Orders:
4 VIAGRA PILLS
Bestsellers
Erectile Dysfunction
Male Enhancement
Viagra + Cialis
$69.99
10 x Viagra 100 mg
10 x Cialis 20 mg
ORDER NOW
Cialis
$198.40
60 pills 20 mg
+4 Free pills
ORDER NOW
Viagra
$230.12
120 pills 100 mg
+ 4 free pills
+ free delivery
ORDER NOW
Search by name:
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Search:

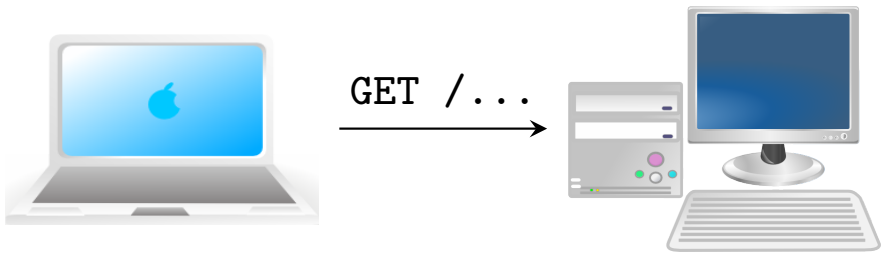
Come funziona?

1. campagna di spam



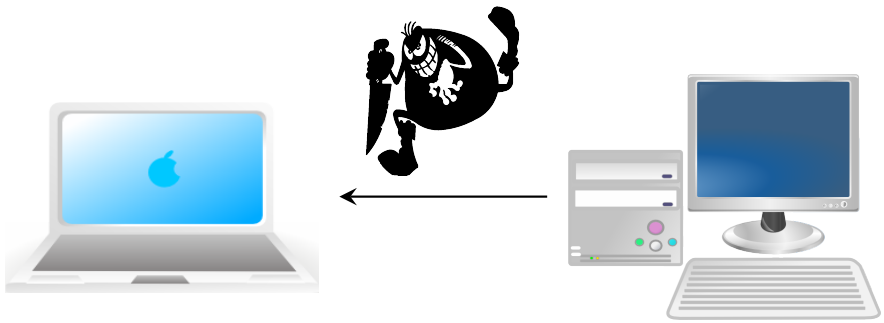
Come funziona?

1. campagna di spam
2. social engineering



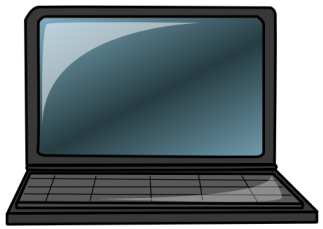
Come funziona?

1. campagna di spam
2. social engineering
3. furto credenziali & malware



Come funziona?

1. campagna di spam
2. social engineering
3. furto credenziali & malware
4. infezione macchine



Underground economy
Vendita informazioni rubate

DICo

Goods & services	**Percentage**	**Range of prices**
Bank accounts	22%	\$10-\$1000
Credit cards	13%	\$0.40-\$20
Full identities	9%	\$1-\$15
Online auction site accounts	7%	\$1-\$8
Scams	7%	\$2.50-\$50/week (hosting)
Mailers	6%	\$1-\$10
Email addresses	5%	\$0.83/MB-\$10/MB
Email passwords	5%	\$4-\$30
Drop (request or offer)	5%	10%-20% of drop amount
Proxies	5%	\$1.50-\$30

Fonte: Symantec

Underground economy
Furto credenziali – Portata del fenomeno

- Università di Mannheim – Limbo & ZeuS
- $\sim$ 70 dropzone
- **33 GB** di dati
- 11000 account bancari, 150000 account mail

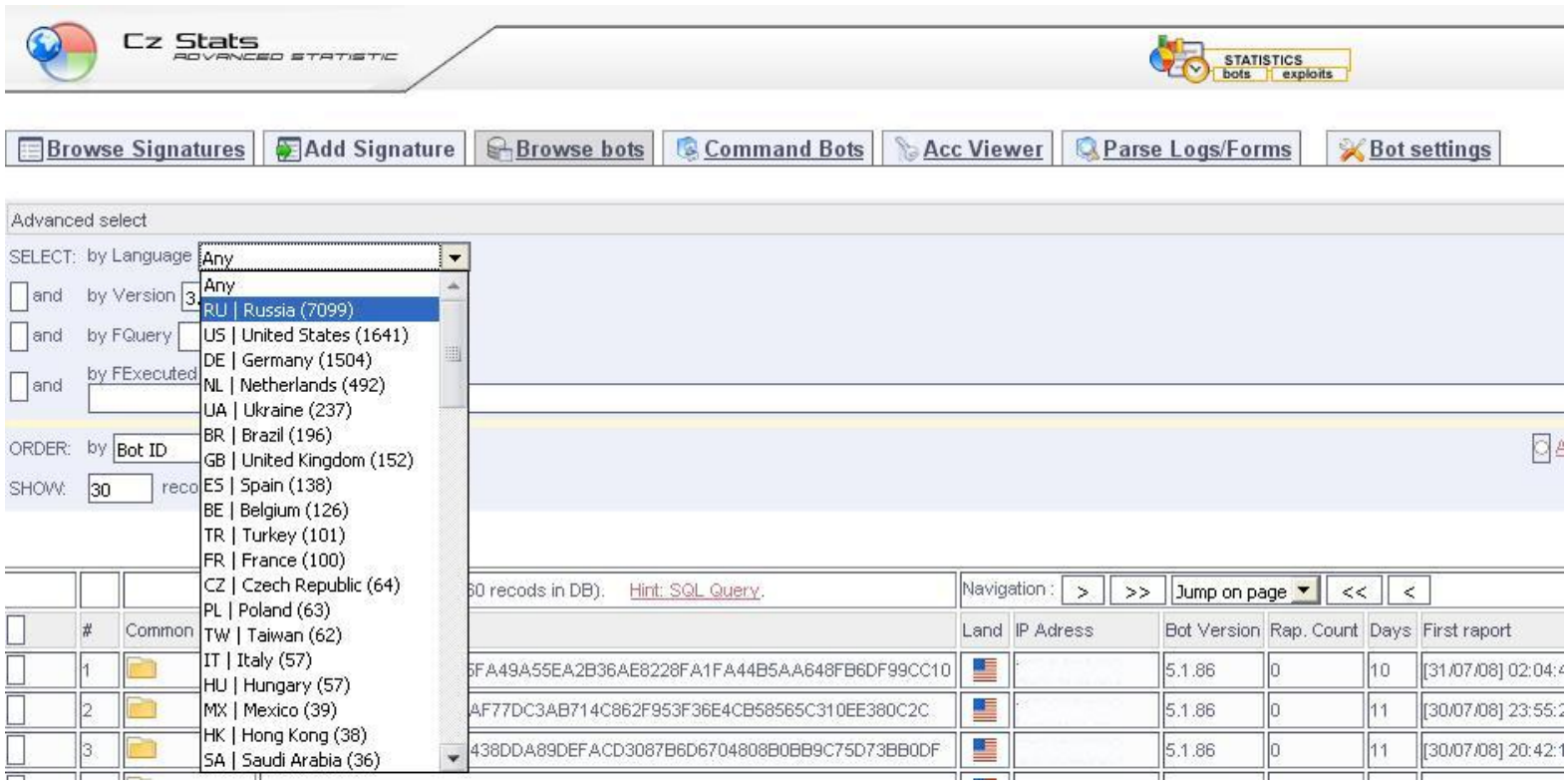
Dropzone	**# Machines**	**Data amount**	**Country**
webpinkXXX.cn	26,150	1.5 GB	China
coXXX-google.cn	12,460	1.2 GB	Malaysia
77.XXX.159.202	10,394	503 MB	Russia
finXXXonline.com	6,932	438 MB	Estonia
Other	108,122	24.4 GB	
Total	164,058	28.0 GB	

Fonte: Learning More About the Underground Economy – T. Holz, M. Engelberth, F. Freiling, 2008

Underground economy

"Malware as a service"

- Bot in affitto ($\sim$ \$1000-\$2000/mese)
- MPACK: exploit toolkit a $\sim$ \$1000



CAPTCHA?

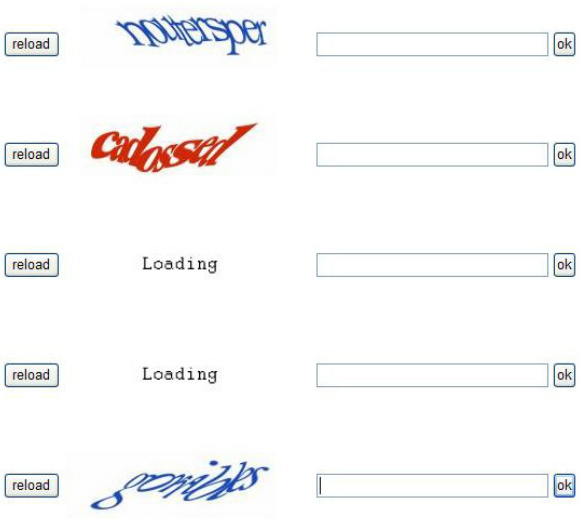
- OCR, Fuzzy OCR, . . .

CAPTCHA?

- OCR, Fuzzy OCR, ...
- Soluzioni migliori?

CAPTCHA?

- OCR, Fuzzy OCR, ...
- Soluzioni migliori?
- "Human computation"!

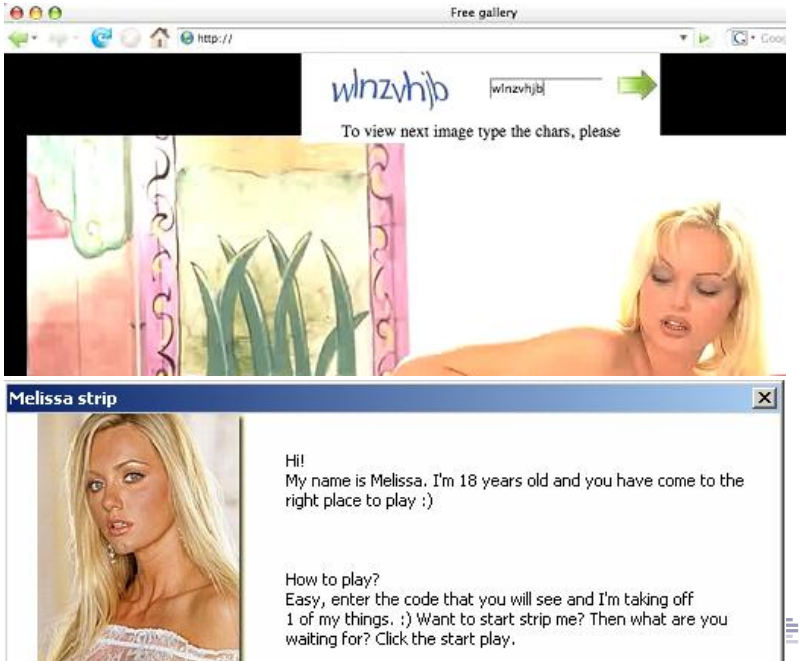


> 100K captcha al giorno, $1.5-$8 per 1000 captcha

Underground economy
The spam business

CAPTCHA?

- OCR, Fuzzy OCR, ...
- Soluzioni migliori?
- "Human computation"!



Funzionalità del malware

Non solo spam...

- mass-mailing (Sobig)
- HTTP/SOCKS proxy (MyDoom)
- SMTP proxy (Taripox, Happy99)
- diffusione tramite IM (Peacomm)

Funzionalità del malware
Furto credenziali

DICo

Obiettivi

- credenziali account bancari/posta
- serial prodotti commerciali (Agobot)
- e-mail harvesting

Come?

- installazione plugin BHO/XUL
- modifica `C:\Windows\...\etc\hosts`
- keylogging
- screen grabbing

Funzionalità del malware
Click fraud

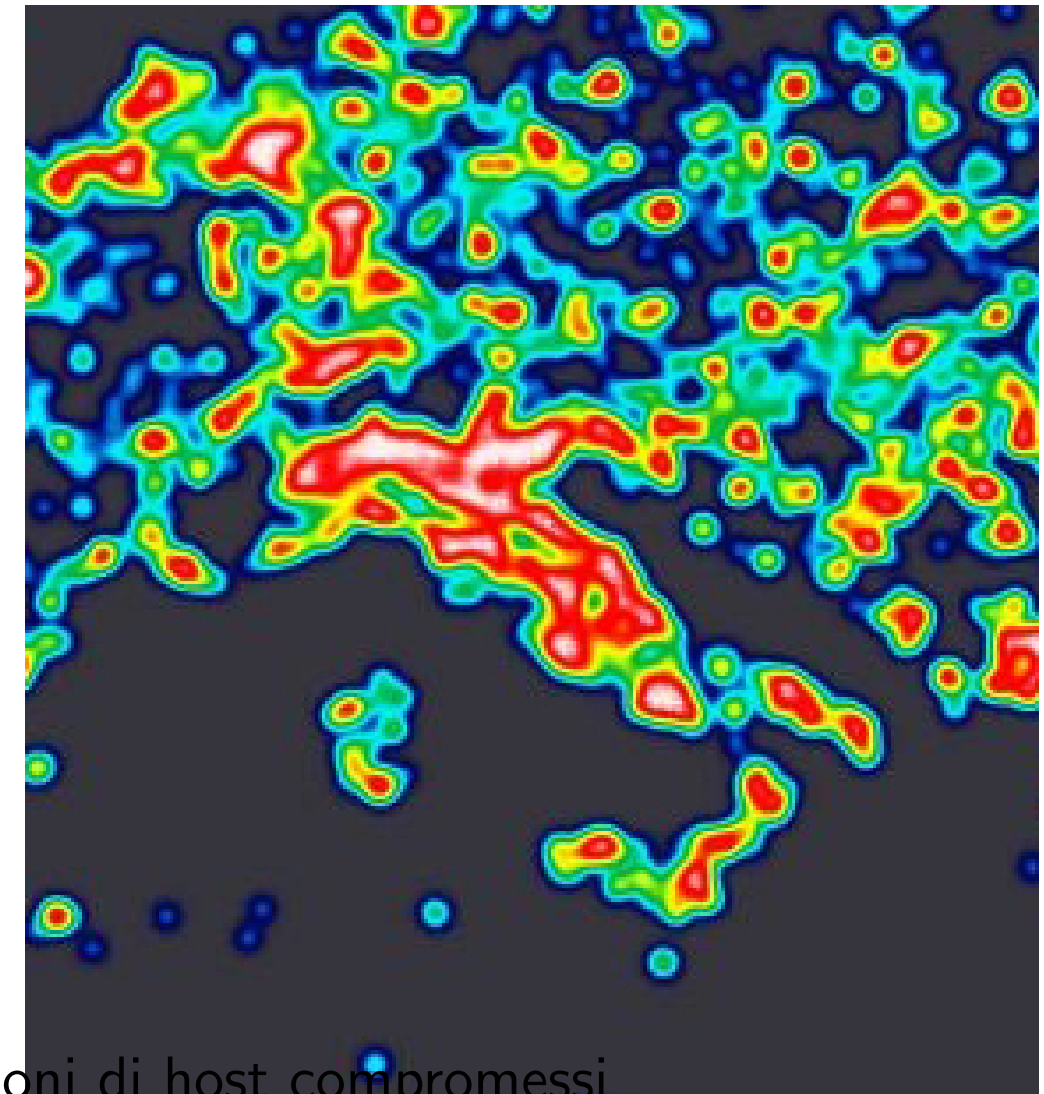
- *Google*: 10% dei "click" sono fraudolenti ($\sim$ \$1B)
- Clickbot.A ($\sim$ 50k host infetti)
- molti "clickbot" commerciali
- ClickJacking

IP	Country	Time	Clicks	Version	Manage
		03:30:05	0	v0.007	Block
		03:30:04	Holded	v0.007	Allow
		03:30:04	8	v0.007	Block

1.7 milioni di host compromessi

Fonte: Team Cymru

Botnet
Conficker (29/01/2009)



1.7 milioni di host compromessi

Fonte: Team Cymru

Nome	Dimensione	Capacità di spam
Conficker	9.000.000	10G/giorno
Kraken	495.000	9G/giorno
Srizbi	450.000	60G/giorno
Rustock	150.000	30G/giorno
Cutwail	125.000	16G/giorno
Storm	$>$ 1.000.000	3G/giorno
Grum	50.000	2G/giorno
Mega-D	35.000	10G/giorno

Analisi di 10 giorni di traffico di rete generato da Torpig:

Unique IP Count	1.148.264
Unique Torpig keys (machines)	180.835
POP accounts	415.206
Email addresses	1.235.122
Passwords	411.039
Unique credit cards	875
Unique ATM pins	141
Unique social security numbers	21

Tecniche di propagazione

Propagation mechanisms	**Percentage**
File sharing executables	40%
File transfer/email attachment	32%
File transfer/CIFS	28%
File sharing/P2P	19%
Remotely exploitable vulnerability	17%
SQL	3%
Back door/Kuang2	3%
Back door/SubSeven	3%
File transfer/embedded HTTP URI/Yahoo! Messenger	2%
Web	1%

Fonte: Symantec, 2007

Situazione malware detector

Rank	Detected	Missed	Product
1	91%	178	Sophos
2	91%	179	AntiVir
3	90%	194	Microsoft
4	90%	195	AVG
5	90%	202	Ikarus
6	89%	213	BitDefender
7	88%	241	Norman
8	88%	247	TrendMicro
9	87%	259	Kaspersky
10	87%	268	F-Secure

Fonte: SRI International + VirusTotal, campione di 2064 malware

Lezione XIV: Esercizi

Biblioteca

- Prendete il file `EsDB2.odb`
- Scopo dell'esercizio è anche quello di lasciarvi capire, senza spiegazione, come è stata strutturata l'informazione.
- Due note (da meditare dopo aver osservato le relazioni della base di dati):
  1. le copie fisiche dei libri sono memorizzate nella tabella `volumi`.
  2. per come è organizzata la base di dati, ogni libro può avere un solo autore.

Interrogazioni

- Trovare il titolo dei libri il cui autore `Amado`
- Trovare il titolo dei libri in prestito all'utente `Annoni`.
- Trovare lo stato di conservazione e il codice biblioteca dei libri editi dalla Feltrinelli in stato di conservazione mediocre o fortemente rovinato
- Trovare, per ogni utente, il numero di libri che ha attualmente in prestito.

Per non dimenticare i fogli elettronici

- Caricare in un fogli di calcolo il file `birthday.cvs`
- Contiene il numero di nascite per giorno (nel 1978)
- Ottenere un grafico che abbia in ascisse i giorni della settimana (lunedí, martedí,. . .) e in ordinate il numero di nascite.

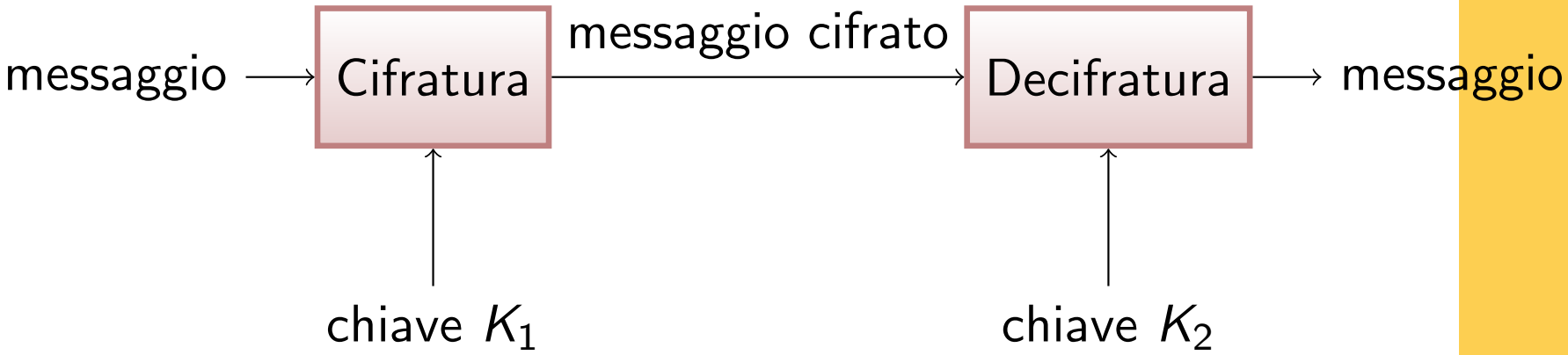
Lezione XV: Sicurezza informatica

Cos'è la sicurezza informatica

- Confidenzialità
- Integrità
  - . . . dei dati (integrità)
  - . . . dell'origine (autenticazione)
- Disponibilità

Confidenzialità e integrità possono essere perseguite utilizzando la crittografia.

Crittografia



Se $K_1 = K_2$ si parla di crittografia a chiave simmetrica.
Se $K_1 \neq K_2$ si parla di crittografia a chiave asimmetrica.

Crittografia a chiave simmetrica

Esempio (molto semplice, ma assai debole: serve a chiarire il funzionamento). Funzione logica XOR ($\oplus$)

a	b	$a \oplus b$
0	0	0
0	1	1
1	0	1
1	1	0

Applicato bit a bit:

$0101 \oplus 1001 = 1100$

Vale la seguente proprietà $(m \oplus k) \oplus k = m$

$(0101 \oplus 1001) \oplus 1001 = 0101$

Perché funziona?

- Cifratura: "testo in chiaro" XOR "chiave" = testo cifrato
- Decifratura: "testo cifrato" XOR "chiave" = testo in chiaro

Supponiamo che l'attaccante ottenga un bit di informazione cifrata e non conosca la chiave:

- se la chiave fosse 1, allora il testo in chiaro sarebbe 0
- se la chiave fosse 0, allora il testo in chiaro sarebbe 1
  - la chiave ha il 50% di probabilità di essere 0 e il 50% di essere 1
  - Dunque dato il testo cifrato, l'attaccante sa che il testo in chiaro è 1 al 50% e 0 al 50%
  - Quindi non ha abbastanza informazione per decidere qual era il messaggio!

L'idea di fondo della crittografia

- Cifrare è *facile*
- Decifrare è *facile se si conosce la chiave*
- Decifrare è *molto difficile* se non si conosce la chiave

"Molto difficile" significa che occorre molte risorse di calcolo: se per esempio un calcolo necessitasse di 2^n passi con n i bit del messaggio di quale KB sarebbe già indecifrabile (il numero di secondi stimato dall'inizio dell'universo è circa 2^{45}).

Problemi della cifratura simmetrica

La cifratura simmetrica è molto efficace per garantire la confidenzialità e l'integrità di un messaggio, ma
Necessita uno scambio di chiave precedente alla comunicazione cifrata
Lo scambio di chiave deve essere effettuato su un canale sicuro.

Diffie-Hellman

DICo

Diffie e Hellman hanno inventato (1976) un meccanismo per evitare la necessità del canale sicuro

- Alice e Bob si accordano su un numero primo $p = 23$ e una base $g = 5$ (possono essere resi pubblici).
- Alice sceglie *segretamente* $a = 6$, e manda a Bob $A = g^a \mathtt{mod} p$ $(A = 5^6 mod 23 = 8)$
- Bob sceglie *segretamente* $b = 15$, e manda ad Alice $B = g^b \mathtt{mod} p$ $(B = 5^{15} mod 23 = 19)$
- Alic calcola $s = B^a \mathtt{mod} p$ $(s = 19^6 \mathtt{mod} 23 = 2)$
- Bob calcola $s = A^b \mathtt{mod} p$ $(s = 8^{15} \mathtt{mod} 23 = 2)$

Alice e Bob hanno in mano lo stesso numero (senza essersi accordati) che possono usare come chiave!

Crittografia a chiave asimmetrica

$K_1 \neq K_2$ e possiamo usare questa proprietà per ottenere la cosiddetta crittografia a *chiave pubblica* (!)
Alice ha due chiavi α e a di cui la prima (α) è segreta e la seconda è pubblica.
Anche Bob ha due chiavi β e b, di cui solo b è pubblica.
Le coppie di chiavi sono tali che ciò che è cifrato con una chiave può essere decifrato solo con l'altra.

Cosa si può ottenere?

- Alice cifra un messaggio con la chiave b (che è pubblica!). Chi può decifrarlo?
- Alice cifra un messaggio con la chiave α (che è privata!). Chi può decifrarlo?

Rimane però il problema di assicurarsi che a sia davvero la chiave pubblica la cui corrispondente privata è controllata solo da Alice.

Cosa si può ottenere?

- Alice cifra un messaggio con la chiave b (che è pubblica!). Chi può decifrarlo?
  - Solo Bob che conosce β
- Alice cifra un messaggio con la chiave α (che è privata!). Chi può decifrarlo?

Rimane però il problema di assicurarsi che a sia davvero la chiave pubblica la cui corrispondente privata è controllata solo da Alice.

Cosa si può ottenere?

- Alice cifra un messaggio con la chiave b (che è pubblica!). Chi può decifrarlo?
  - Solo Bob che conosce β
- Alice cifra un messaggio con la chiave α (che è privata!). Chi può decifrarlo?
  - Tutti, perché la chiave a è pubblica

Rimane però il problema di assicurarsi che a sia davvero la chiave pubblica la cui corrispondente privata è controllata solo da Alice.

Cosa si può ottenere?

- Alice cifra un messaggio con la chiave b (che è pubblica!). Chi può decifrarlo?
  - Solo Bob che conosce β
- Alice cifra un messaggio con la chiave α (che è privata!). Chi può decifrarlo?
  - Tutti, perché la chiave a è pubblica
  - Ma solo Alice può averlo cifrato: firma digitale

Rimane però il problema di assicurarsi che a sia davvero la chiave pubblica la cui corrispondente privata è controllata solo da Alice.

Lezione XVI: Esercizi

Esercizio Magazzino

"Publiweb & TV" effettua numerose campagne pubblicitarie per far conoscere il proprio marchio. L'azienda dispone di un magazzino che stocca cartelloni pubblicitari, striscioni e gadget di vario genere. Poiché la produzione di tale materiale è mirata in base alle esigenze la "Publiweb & TV" gestisce il magazzino in modo molto accurato, al fine di non rimanere senza articoli pubblicitari, ma senza nemmeno prendere un numero eccessivo di pezzi che magari rimarrebbero inutilizzati.

- La colonna "Valore" è calcolata moltiplicando il "Prezzo unitario" per la "Giacenza". Alle colonne "Prezzo unitario" e "Valore" va assegnato il corretto formato di Valuta in Euro.
- In "Avviso scorta" si faccia scrivere: "OK" se il valore della giacenza è compreso tra il minimo e il massimo consentito, "Scorta eccessiva" se il valore eccede la "Scorta Massima" e "Sotto scorta" se la è sotto il valore di "Scorta minima" consentita.

Frutta

DICo

La ditta TantaFrutta produce frutta e la trasforma in succhi di frutta. Nel 2009, a fronte di una produzione particolarmente abbondante di arance, decide di regalare succo di arancia ad alcuni paesi del terzo mondo. Per la spedizione il prodotto deve essere confezionato in contenitori da 100 litri, ognuno del costo di 100EUR. Occorre misurare quanto è il costo per l'acquisto dei contenitori. A tale scopo, ti viene fornita la base di dati in cui sono memorizzati i risultati di produzione delle varie aziende agricole appartenenti alla ditta TantaFrutta.

Estrai i dati che ti servono dalla base di dati, calcola la sovrapproduzione (vedi sotto come calcolarla) e, considerando che per ogni chilo di arance si produce mezzo litro di succo d'arancia, calcola quanto sarà il costo totale per l'acquisto dei contenitori.

Modalità d'esame

Tre appelli nella sessione invernale. Le date potranno cambiare nel caso non ci fosse la disponibilità delle aule di laboratorio (notizie aggiornate sul sito).

1. 15 febbraio
2. 22 febbraio
3. 3 marzo

Per questa sessione non ci sarà "salto d'appello", ma verrà introdotto in seguito in caso di abusi.

Iscrizioni

Per ogni appello occorre iscriversi entro precisi temporali

1. 15 febbraio: entro il 9 febbraio
2. 22 febbraio: entro il 16 febbraio
3. 3 marzo: entro il 23 febbraio

Su questi vincoli non ci saranno eccezioni.

Come iscriversi

1. creare un file di testo (con estensione `.txt`) e salvarlo col nome `CognomeNomeMatricola.txt`
2. collegarsi al sito `http://upload.dico.unimi.it/`
3. Accedere all'appello desiderato e inserisci nome utente e password della posta di ateneo seleziona (`@studenti.unimi.it`)
4. caricare il file `CognomeNomeMatricola.txt`
5. è possibile controllare l'avvenuta iscrizione con la voce "Verifica i file che hai già sottoposto"

L'esame

L'esame consiste in:

- una domanda aperta (6 punti)
- dodici domande a risposta multipla (12 punti)
- Un esercizio da svolgere sul computer (es. basi di dati o foglio di calcolo, 12 punti)

Alle domande occorre rispondere nella prima mezz'ora: eventuale tempo risparmiato può essere usato per l'esercizio, che dovrà essere completato nell'ora seguente.
Per tutto il tempo dell'esame (1h30') non è possibile uscire dall'aula

Cosa portare all'esame

- Un documento di identificazione
- Una penna (ed eventualmente bianchetto)
- Un orologio (non il cellulare o altri dispositivi che abbiano altre funzioni)

Non si può usare nient'altro!
Comunicazioni, copiature, sguardi sul monitor del vicino comportano l'esclusione dalla valutazione.