

Corso di Laurea Magistrale in Bioingegneria
Insegnamento di INFORMATICA MEDICA - Primo Appello (27.01.2015) – Tema A

Dichiaro di essere, alla data odierna (barrare la casella di pertinenza):

- ☐ iscritto al Corso di Laurea Magistrale in Bioingegneria a Padova
☐ iscritto al Corso di Laurea Magistrale in _____ a Padova
☐ iscritto al Corso di Laurea Magistrale in Ingegneria Clinica a Trieste
☐ iscritto ad Informatica Medica come corso singolo

Dichiaro di essere consapevole che durante la prova **non è consentito l'uso di libri e appunti** e che il **possesso di apparecchi telefonici e/o fotografici o simili, anche se spenti, comporta l'esclusione**, nonché la denuncia agli organi accademici competenti

COGNOME

NOME

Firma

Tempo a disposizione: 2h 45 min (2h 25 min stimati + 20 min extra). Ritirarsi è possibile entro 75 min
La comprensione del testo fa parte dell'esame, non presentarsi alla cattedra a chiedere spiegazioni

Soluzione (in blu) e commenti del docente dopo aver corretto i compiti (in viola)

PARTE 1 –MATLAB e SQL [Totale 7/30, ~ 35 min]

3 RISPOSTE BREVI (Rispondere direttamente sul presente foglio) **1/30, Tempo max previsto: ~5 minuti**

R1. Assumendo che y sia un vettore di lunghezza divisibile per 4, già nel workspace assieme al vettore t dei tempi di campionamento a cui si riferisce, il codice Matlab

```
n=length(y);  
Ty=dct(y);  
Tyc=[Ty(1:3*n/4); zeros(n/4,1)];  
yc=idct(Tyc);  
plot(t,y, 'r--', t, yc, 'b-')
```

restituisce in figura

in rosso tratteggiato il segnale originale, e in blu continuo la sua versione decompressa ottenuta usando il 75% delle componenti DCT a più bassa frequenza (simula rapporto di compressione R 3:4)

R2. La sequenza di comandi Matlab

```
t=(1:1:100)';  
y=t*exp(t);  
plot(t,y)
```

provoca errore perché

alla riga 2 i vettori colonna t ed $\exp(t)$ non hanno dimensioni compatibili con la $*$ (per dare luogo al vettore y dei campioni di te^t andrebbero moltiplicati elemento per elemento usando la $.*$)

R3. In una struttura `SELECT ... FROM ... GROUP BY ... HAVING ...`, la presenza della clausola `HAVING` fa sì che

solo i raggruppamenti creati dalla `GROUP BY` che superano la condizione presente nella `HAVING` passano alla `SELECT` successiva

3 TEST (Rispondere direttamente sul presente foglio) **1/30, ~5 minuti (giusta 0.333/30, sbagliata -0.166/30, non data 0/30)**

T1-2-3. Considerata la relazione `DRG(codice_drg, descrizione, rimborso)`, la seguente query è formalmente corretta

```
SELECT *
FROM DRG
WHERE rimborso >1000;
```

Vero ☐

Falso ☐

```
SELECT codice_drg, rimborso
FROM DRG
GROUP BY rimborso
ORDER BY rimborso DESC;
```

Vero ☐

Falso ☐

```
SELECT COUNT(*)
FROM DRG
WHERE codice_drg >30
GROUP BY rimborso
HAVING COUNT(*)>1;
```

Vero ☐

Falso ☐

M1 (Matlab) 3/30, Tempo max previsto: ~15 minuti. (Rispondere direttamente **sul presente foglio**)

Scrivere in Matlab una function che riceva in input una stringa s e restituisca in uscita un digest in binario d calcolato con il semplice algoritmo di hash (quello che sfrutta la somma dei codici ascii dei caratteri che compongono s) discusso prima a lezione e poi nell'esercitazione 4. Prevedere che in input possa eventualmente essere passato un secondo argomento, k , corrispondente al numero di bit di cui si vuole composto il digest. Se l'utente non passasse in input il secondo argomento, la function dovrà calcolare un digest su 4 bit (in altri termini, 4 è il valore di default per la lunghezza del digest).

```
function d=hash(s,k)
```

```
somma=sum(double(s));
```

```
if nargin==1
```

```
    k=4;
```

```
end %if
```

```
resto=mod(somma, 2^k);
```

```
d=dec2bin(resto,k);
```

Errori riscontrati:

script e non function;

input nella function;

mod dentro il ciclo che calcola la somma;

...

Fattore di demerito (-10%):

codice inutilmente complicato (es. duplicazione di intere sezioni)

SQL 2/30, ~10 minuti (Rispondere direttamente **sul presente foglio**). Si riporti **l'output** prodotto dalla seguente query:

```
SELECT alb1.figlio AS colonna1, alb2.figlio AS colonna2
FROM albero alb1, albero alb2
WHERE alb1.genitore=alb2.genitore AND alb1.figlio<>alb2.figlio AND alb1.annif>alb2.annif;
```

fatta sulla seguente istanza della relazione ALBERO

genitore	annig	figlio	annif
Luca Rossi	55	Anna Rossi	21
Maria Pavan	54	Anna Rossi	21
Luca Rossi	55	Gianluca Rossi	18
Maria Pavan	54	Gianluca Rossi	18
Giorgio Rossi	80	Luca Rossi	55
Silvia Di Carlo	79	Maria Pavan	54
Enzo Pavan	82	Maria Pavan	54
Enzo Pavan	82	Franca Pavan	48
Silvia Di Carlo	79	Franca Pavan	48

colonna1 colonna2

Anna Rossi Gianluca Rossi

Anna Rossi Gianluca Rossi
Maria Pavan Franca Pavan
Maria Pavan Franca Pavan

Non richiesto: la query “aggancia” i fratelli (quelli che condividono un genitore), con il maggiore che andrà sulla colonna 1. Ogni coppia di fratelli è agganciata sia tramite il padre che tramite la madre. Pertanto avremo in output delle righe duplicate

PARTE 2 – TEORIA E ESERCIZI [Totale 23/30, ~ 110 min]

9 TEST (Rispondere direttamente **sul presente foglio**) 3/30, Tempo max previsto: ~10 minuti (giusta 0.333/30, sbagliata -0.166/30, non data 0/30)

Omissis...(vd "Riepilogo Infomed")

7 RISPOSTE BREVI (Rispondere direttamente **sul presente foglio**) 4/30, Tempo max previsto: ~10 minuti

Omissis...(vd "Riepilogo Infomed")

Rispondere negli spazi previsti in modo ordinato ed essenziale

D1 (Compressione). 3/30, Tempo max previsto: ~15 minuti

Illustrare **brevemente** perché si può far leva sull'irrelevanza percettiva per comprimere l'ingombro di un messaggio vocale all'interno di una segreteria telefonica digitale

Si nota innanzitutto che la specifica applicazione (segreteria telefonica, tipicamente da ascoltare in ambiente acusticamente rumoroso) necessita di **bassa fedeltà**: la qualità minima sufficiente è infatti quella che consente di riconoscere chi parla e comprendere il messaggio senza fatica. Molti dettagli del segnale originale, irrilevanti ai fini della comprensione del parlato da una memorizzazione nella segreteria, potranno quindi essere sacrificati senza perdita alcuna per l'applicazione stessa. Inoltre, è ben noto che il segnale "voce" è concentrato in una ristretta banda (~300-3000Hz). Componenti al di fuori di questa banda non sono di norma prodotte dal sistema fonatorio umano e, quindi, il sistema di archiviazione per segreteria telefonica può far leva anche su questa caratteristica del segnale rispetto ai generici segnali audio. Ad esempio, si potrebbe adottare una tecnica di codifica che considera, del segnale captato dal microfono, la sola banda voce e, al suo interno, le sole componenti a maggiore energia, identificate ad es. sfruttando la DCT monodimensionale.

Quesito non difficile, bastava pensarci con mentalità ingegneristica da 4° anno. Ad esempio:

- a) diversi studenti non hanno nemmeno considerato il fatto che è l'applicazione stessa a favorire l'uso di compressione anche spinta (non serve alta fedeltà)
- b) diversi studenti hanno pensato di trattare il segnale come se si facesse una registrazione HI-FI (banda a 20KHz), concentrandosi solo sulla banda uditiva ed ignorando che la banda voce è molto più stretta

E1. (Organizzazione Sanitaria) 3.5/30, Tempo max previsto: ~20 minuti Rispondere **nello spazio sotto**

Si ipotizzi che l'universo dei casi clinici sia riconducibile a soli 5 DRG, per i quali si ha, nello standard, $\{P_i\}$ e $\{D_i\}$ rispettivamente dati da $\{0.4, 0.3, 0.2, 0.07, 0.03\}$ e $\{7, 10, 10, 12, 18\}$ (D_i espresse in giorni). Le tariffe di rimborso $\{T_{2i}\}$ calcolate dal Ministero, in Euro, sono $\{700, 1000, 1200, 1440, 3600\}$. Si supponga che l'Azienda Ospedaliera X abbia avuto nel 2014 1000 ricoveri, tutti rimborsati con tariffa T2, con $\{p_j\}$ e $\{d_j\}$ rispettivamente dati da $\{0.4, 0.35, 0.15, 0.07, 0.03\}$ e $\{6, 10, 10, 12, 16\}$. Ipotizzando che l'Azienda Ospedaliera abbia costi medi giornalieri identici a quelli standard considerati al Ministero:

- a) Calcolare il valore in Euro del profitto o delle perdite che l'Azienda Ospedaliera ha avuto nel 2014;
- b) Riportare una stima (anche approssimata) dei valori di ICM e ICP e anche una breve interpretazione dei valori trovati per lo specifico caso in esame

a) Essendo i costi locali e quelli standard identici, confrontando le durate dei ricoveri si capisce che X avrà un profitto. Essendo profitto=ricavo-costo, è sufficiente concentrarsi sui due DRG in cui il ricavo sarà maggiore dei costi e, per la linearità di costi e ricavi, considerare che i casi singoli coincidano con il caso medio.

Per il DRG 1 (costo giornaliero 100 Euro), per ciascuno dei 400 casi si ha un profitto di 100 Euro, corrispondente al giorno di riduzione della durata della degenza rispetto allo standard.

Per il DRG 5 (costo giornaliero 200 Euro), per ciascuno dei 30 casi si ha un profitto di 400 Euro, corrispondente ai 2 giorni di riduzione della durata della degenza rispetto allo standard.

$$\text{Profitto} = 400 \times 100 + 30 \times 400 = 40000 + 12000 = 52000 \text{ Euro}$$

b)

$$\text{Dst} = [0.4, 0.3, 0.2, 0.07, 0.03] \times [7, 10, 10, 12, 18]T$$

$$\text{ICM} = ([0.4, 0.35, 0.15, 0.07, 0.03] \times [7, 10, 10, 12, 18]T) / \text{Dst} = 1$$

Si vede che ICM è esattamente 1 perché tutte le p_j sono uguali alle P_j , a parte p_2 e p_3 , che però danno somma uguale a P_2 e P_3 e corrispondono entrambe a ricoveri dello stesso numero di giorni (10)

$$\text{ICP} = ([0.4, 0.3, 0.2, 0.07, 0.03] \times [6, 10, 10, 12, 16]T) / \text{Dst}$$

Non so quanto ICP sia esattamente, ma so con certezza che è (di poco) minore di 1. Infatti il numeratore è minore del denominatore. Confrontando $\{D_j\}$ e $\{d_j\}$ si evince peraltro che la performance di X è migliore che nello standard

Se il docente non ha lasciato usare la calcolatrice, evidentemente non servivano particolari conti ... Era un segnale di non cacciarsi in strade faticose

E2 (Algebra relazionale e proprietà) 3.5/30, Tempo max previsto: ~15 minuti

Considerare il data base di un'azienda in uso da alcuni anni e fotografato il giorno 26.1.2015:

FATTURE(NFattura, Data, PIVA)
CLIENTI(PIVA, RagioneSociale, Città)
DETTAGLI(NFattura, CodiceArticolo, Quantità)
ARTICOLI(CodA, Descrizione, CostoUnitario)

con ovvio significato degli attributi e PIVA di FATTURE chiave esterna per CLIENTI e NFattura di DETTAGLI chiave esterna per FATTURE e CodiceArticolo di DETTAGLI chiave esterna per ARTICOLI (tutte con vincolo di integrità referenziale). Siano n_F , n_C , n_D ed n_A le cardinalità delle quattro tabelle. Con una **breve motivazione** (bastano 1-2 righe), dire se si può affermare qualcosa su (NB: $I > I$ indica il join):

a) cardinalità di $\pi_{PIVA} \{ CLIENTI I > I \text{ RIGHT } \sigma_{1/1/2014 < Data < 31/12/2014} (FATTURE) \}$

$\leq \min(n_C, n_F)$

Nella relazione prodotta dalla query trovo una colonna di PIVA diverse, tutte quelle già presenti nella scrematura della tabella FATTURE (nb: la join è solo "di confusione"). Queste, derivando da una selezione su FATTURE, non possono ovviamente essere più di n_F . Ovviamente le PIVA diverse nel DB sono al più n_C

b) cardinalità di $\sigma_{Quantità=Null} \{ ARTICOLI I > I \text{ LEFT } [p_{CodA} \leftarrow \text{CodiceArticolo} (DETTAGLI)] \}$

Le righe della relazione prodotta sono tante quanti sono gli articoli mai venduti (il valore Null su Quantità ci sarà solo nelle tuple risultanti da tuple di ARTICOLI non concatenabili a tuple di DETTAGLI)

$\leq n_A$

c) cardinalità di $\pi_{\text{CodiceArticolo}} \{ (DETTAGLI I > I \text{ ARTICOLI}) \}$

DETTAGLI I > I ARTICOLI realizza un prodotto cartesiano. Le righe della relazione prodotta dalla proiezione sono tante quanti sono gli articoli venduti almeno una volta (CodiceArticolo proviene da DETTAGLI). Sicuramente sono

$\leq n_A$

Sicuramente sono anche $\leq n_D$

La risposta è quindi

$\leq \min(n_A, n_D)$

Qualcuno ha riportato come cardinalità quantità tipo " $n_A - n_D$ " che si vedeva subito che non potevano andare bene, ad esempio perché potenzialmente **negative**

d) numero dato dalla somma dei due numeri b e c

Esattamente n_A

Il punto d) doveva servire come controllo (improbabile che venissero valori non conoscibili o senza interpretazione)

E3 (Progettazione di Data Base) 6/30, Tempo max previsto: ~40 minuti

Si consideri il problema di descrivere le vendite di libri in una rete internazionali di negozi. Ogni libro è identificato da un codice libro. Di ogni libro vanno censiti gli autori, mantenendo traccia dell'eventuale ordine (autore 1, autore 2, ...) in caso di opera a più firme. Ogni autore è identificato da un codice autore ed è descritto da nome e cognome. Per una piccola frazione di autori nel database sono anche censiti indirizzi web (come URL) di siti dedicati (eventualmente più siti per uno stesso autore). Ogni editore nel data base possiede un codice editore che lo identifica, nome e sito web. Si riporta inoltre città e nazione in cui l'editore ha la sede principale. Le vendite dei libri avvengono in negozi contraddistinti da un codice negozio univoco, nome, città, nazione (es. il negozio con nome "Feltrinelli Centro Giotto", situato a Padova in Italia, ha codice N6756). Per ogni libro va tenuta traccia di quante copie sono state vendute nei vari negozi della rete. Per semplicità:

- si consideri che i libri siano disponibili solo in lingua originale (la lingua di scrittura di ciascun libro va censita nel data base);
- si escluda l'esistenza di edizioni diverse della stessa opera (es. esiste una sola edizione, e ha codice libro A1456Y, de "I Promessi Sposi" di A. Manzoni, è edita in italiano dalla casa editrice "Rossi", con sede in Italia a Milano; esiste una sola edizione di "The Catcher in the Rye" di J.D. Salinger, è edita in inglese dalla casa editrice "Jones", con sede in Brasile a San Paolo, e ha codice libro F85R70);
- si assuma che non esistano nel data base due città con lo stesso nome (anche se in nazioni diverse).

Risolvere sul foglio protocollo, in modo leggibile ed ordinato (fa parte della valutazione), i seguenti punti:

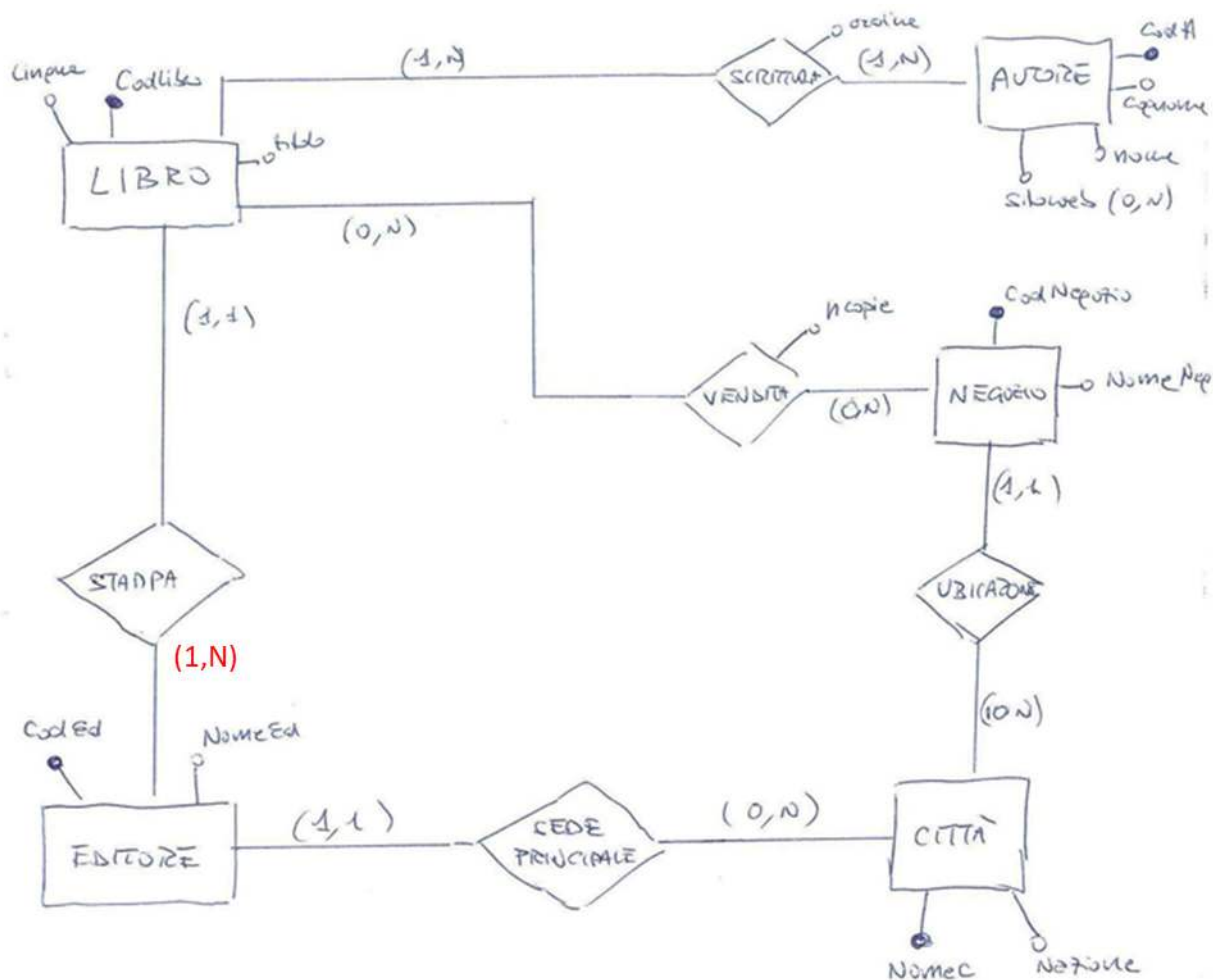
- A. eseguire la progettazione concettuale, disegnando il diagramma E-R, completo di cardinalità e analisi delle associazioni (nb: non introdurre altri attributi rispetto a quelli riportati nella descrizione sopra)
- B. tradurre il diagramma in tabelle in terza forma normale, complete dell'indicazione di chiavi primarie e di eventuali chiavi esterne

Errori frequenti.

Dimenticato campo "Titolo"

Entità CITTA' non riconosciuta

Alcuni errori nel diagramma sarebbero stati riconosciuti dalle osservando con spirito critico le tabelle da esso conseguenti



LIBRI (CodLibro, Titolo, Lingua, CodEd) CodEd FK per EDITORI

SCRITTURE (CodLibro, CodA, Ordine) CodLibro FK per LIBRI
CodA FK per AUTORI

AUTORI (CodA, Cognome, Nome)

SITI (URL, CodA) CodA FK per AUTORI

EDITORI (CodEd, NomeEd, NomeC) NomeC FK per CITTA'

VENDE (CodLibro, CodNegotio, Ncopie) CodLibro FK per LIBRI
CodNegotio FK per NEGOZI

NEGOZI (CodNegotio, NomeNeg, NomeC) NomeC FK per CITTA'

CITTA' (NomeC, Nazione)