

Cognome Nome Matricola

Sistemi Informativi

14-02-2017

Sbarramento							
1 (4pt)	2 (4pt)	3 (6pt)	4 (5pt)	5 (4pt)	6 (5pt)	7 (5pt)	Somma (33pt)

Il compito ha la durata di 3 ore, per la compilazione attenersi alle seguenti istruzioni:

- Nome e cognome devono essere scritti in **STAMPATELLO**.
- É sola responsabilità dello studente scrivere in modo **LEGGIBILE**.
- Solo le soluzioni scritte all'interno dei riquadri verranno corrette.
- Non sono ammessi appunti o altro materiale, a chiunque copia verrà ritirato l'esame.
- Affinché l'esame venga corretto e quindi valutato, negli esercizi marchiat **con **** si dovrà raggiungere almeno 4 punti come somma totale.
- L'esame si ritiene superato se si raggiunge il punteggio di 18. La lode si ottiene con punteggio ≥ 31 .
- Qualsiasi soluzione ragionevole è accettata ma verranno premiate soluzioni complete in termini di sintassi e valutata positivamente la padronanza dei concetti presentati durante il corso.
- Sono ammesse le seguenti abbreviazioni nell'XML schema e nel DTD:
 - `<e ...> = <!ELEMENT ... >` e `<xs:element ... >`
 - `<a ...> = <!ATTLIST ... >` e `<xs:attribute ... >`
 - `<ct ...> = <xs:complexType ... >`
 - `<st ...> = <xs:simpleType ... >`

Esercizio 1 (4 punti). Date le Transazioni:

T1: r1(A); w1(A); w1(C); **T2:** r2(B); w2(B); w2(C); **T3:** r3(B); r3(C); w3(C);

E lo schedule:

S: r1(A); w1(A); r2(B); w2(B); r3(B); w1(C); w2(C); r3(C); w3(C)

2pt) ** Dimostrare se lo schedule è conflict serializable e, in tal caso, scrivere una corretta esecuzione seriale equivalente.

1pt) Se lo schedule precedente è conflict serializable modificare l'ordine delle operazioni creando uno schedule diverso che non lo sia, oppure viceversa. **La modifica deve essere minima**

1pt) Disegnare un diagramma generico sull'asse del tempo (timeline) che illustri un esempio di una serie di operazioni che provocano una lettura in ritardo per il caso di uno scheduler basato su timestamp

Sia dato il seguente XML contenuto nel file “pasticceria.xml”:

```
<?xml version="1.0" encoding="UTF-8"?><libro><ricette>
  <ricetta nome="R1" autore="Banderas"><!-- almeno una -->
    <tempo unita="minuti" durata="120" /><!-- secondi, minuti oppure ore, default minuti -->
    <r_ingredienti>
      <r_ingrediente iid="F1" >1</r_ingrediente>
      <r_ingrediente iid="A1" >0.01</r_ingrediente>
      <costo_totale valuta="euro">4</costo_totale><!-- euro oppure bitcoin, no default --!>
    </r_ingredienti>
  </ricetta>
  <ricetta nome="B2" autore="Rosita">
    <tempo unita="ore" durata="2" />
    <r_ingredienti>
      <r_ingrediente iid="F1" >5</r_ingrediente>
      <costo_totale valuta="euro">3</costo_totale>
    </r_ingredienti>
  </ricetta>
  <ricetta nome="aria fritta" autore="Trump">
    <tempo unita="ore" durata="0" />
    <r_ingredienti>
      <costo_totale valuta="euro">0</costo_totale>
    </r_ingredienti>
  </ricetta></ricette>
<ingredienti>
  <farina id="F1"><!-- 'id' vale come chiave -->
    <tipo>0</tipo>
    <unita>kg</unita><!-- kg oppure lt -->
    <produttore pid="001">Produttore1</produttore>
  </farina>
  <farina id="F2">
    <tipo>00</tipo>
    <unita>kg</unita>
    <produttore pid="002">Produttore2</produttore>
    <origine>Scozia</origine>
  </farina>
  <lievito id="L1">
    <nome>Lievito1</nome>
    <unita>kg</unita>
    <produttore pid="003">Produttore3</produttore>
    <origine>Italia</origine>
  </lievito>
  <altro id="A1"><nome>Cannella</nome>
    <unita>kg</unita>
    <produttore pid="005">Produttore5</produttore>
    <origine>India</origine>
  </altro>
</ingredienti>
<produttori>...</produttori></libro>
```

Esercizio 2 (4 punti). Rispondere alle seguenti domande

2pt) Scrivere un possibile schema DTD valido (escludendo il contenuto di <produttori>)

2pt) Scrivere il frammento di un possibile XML schema degli elementi dentro `<ingredienti>` considerando una chiave esportata a scelta verso `<produttori>`

Esercizio 3 (6 punti). Sull'XML precedente, scrivere le seguenti interrogazioni XQuery.

2pt) Restituire il nome delle ricette e il numero di ingredienti ordinate per costo.

```
<ricette><ricetta nome="xxx" ingredienti=".." costo=".." />...</ricette>
```

2pt) Restituire la somma dei costi divisi per bitcoin e euro

```
<costi><bit>0</bit><eur>7</eur></costi>
```

2pt) Trovare i produttori con almeno 2 ingredienti.

```
<produttori><produttore id="xxx" />...</produttori>
```

Esercizio 4 (5 punti). Siano dati i tre seguenti vettori di caratteristiche per gli oggetti *bird*, *plane* e *superman*, considerando le caratteristiche numerate da 1 a 14 dall'alto al basso:

1pt) ** Calcolare la Jaccard Distance tra i tre elementi

	BIRD	PLANE	SUPERMAN	
CARRIES PEOPLE		✓	✓	1
OFTEN FLIES IN GROUPS	✓	✓		2
CREATED IN 20 TH CENTURY		✓	✓	3
USES MAGNETIC NAVIGATION	✓	✓		4
ENTHUSIAST COMMUNITY OBSESSES OVER SMALL COLORATION DETAILS	✓	✓	✓	5
PREYED ON BY CATS	✓			6
OCCASIONAL MID-AIR SEX	✓	✓	✓	7
EATEN DURING SEASONAL FEASTS	✓			8
PROPELLED BY FLAPPING	✓			9
SOMETIMES LOSES ABILITY TO FLY, NEEDS TO SUNBATHE TO REGAIN IT	✓		✓	10
CAN TAKE A PUNCH		✓	✓	11
MATING BEHAVIOR OFTEN OBSERVED BY A HIDDEN DAVID ATTENBOROUGH	✓			12
CAPABLE OF INTENTIONALLY RELEASING POOP MID-FLIGHT	✓		✓	13
CHASES AND EATS BUGS	✓			14

xkcd/1792

2pt) Calcolare i valori di min-hash signature dei tre elementi utilizzando le seguenti permutazioni.

P1: 1, 3, 5, 7, 9, 11, 13, 2, 4, 6, 8, 10, 12, 14

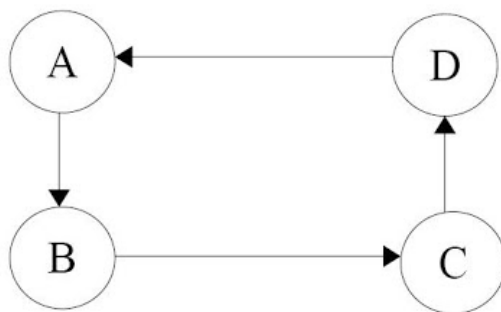
P2: 2, 4, 6, 8, 10, 12, 14, 1, 3, 5, 7, 9, 11, 13

P3: 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1

P4: 8, 7, 6, 5, 4, 3, 2, 1, 14, 13, 12, 11, 10, 9

2pt) **TEORIA:** Elencare le proprietà che deve rispettare un'operazione di merge valida per un algoritmo di **Entity Resolution**.

Esercizio 5 (4 punti). Calcolare pagerank per la rete che consiste di 3 nodi, A, B, C, D come in figura.



2pt) ** Calcolare il pagerank CON $\beta = 1$ e $\beta = 0$ considerando teleport set $\{A, B, C, D\}$.

2pt) Calcolare pagerank con $\beta = 0.8$ e teleport set $\{A, C\}$

Esercizio 6 (5 punti). Sia dato un database per la vendita di prodotti. Ogni cliente ha un nome e una città, ogni prodotto un nome, un codice produttore e una categoria. Una vendita è identificata da un utente, un prodotto, una data, un valore della vendita e una quantità di prodotto:

1pt) ** Definire fact table e dimension tables per questo scenario.

2pt) Ridefinire la fact table per un formal data cube che permetta le aggregazioni rispetto produttori e categorie di prodotto. Scrivere una query SQL per creare tale data cube.

2pt) Scrivere la query sul cubo definito per trovare il numero di prodotti venduti di categoria 'bellezza'

Esercizio 7 (5 punti). Si considerino 3 aziende Az1, Az2, Az3. Az1 vende prodotti di categoria “profumi” e “cosmetici”, Az2 vende “cosmetici” “biologici”, mentre Az3 “cibi” “biologici” . Ogni azienda ha un budget di 50\$ offerto dal motore di ricerca Altavista, e ogni *inserzione* costa 15\$.

Si consideri il seguente ordine di ricerche:

“profumi”; “biologici”; “profumi” ; “profumi” ; “cosmetici”; “biologici”; “cibi”

2pt) ** Si trovi l’assegnamento ottimo mostrando brevemente perchè è ottimo.

3pt) Si trovi l’assegnamento dato dall’algoritmo *balance greedy* mostrando brevemente il processo