

Cognome Nome Matricola

Sistemi Informativi

15-06-2017

Sbarramento							
1 (4pt)	2 (4pt)	3 (6pt)	4 (6pt)	5 (4pt)	6 (4pt)	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(B); w1(C); **T2:** r2(A); w2(B); w2(C); **T3:** r3(A); r3(C); w3(C);

E lo schedule (incompleto):

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

2pt) ** Completare **S** in modo che risulti **conflict serializzabile** e con **w2(C)** come ultima operazione. Dimostrare la correttezza della soluzione.

1pt) Aggiornare lo schedule inserendo le operazioni per shared, exclusive e update lock.

1pt) Disegnare un diagramma generico sull'asse del tempo (timeline) che illustri un esempio di una serie di operazioni che provocano una scrittura 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="ISO8859-1" ?>
<menu>
  <food>
    <name>Belgian Waffles</name>
    <price curr="eur">5.95</price> <!-- 'eur' oppure 'dol' oppure 'pound' -->
    <serving>1</serving><!-- da 1 a 5 -->
    <calories>650</calories>
    <restrictions>
      <r incompatible="gf" />
    </restrictions>
  </food>
  <food>
    <name>Strawberry Belgian Waffles</name>
    <price curr="eur">7.95</price>
    <serving>1</serving>
    <calories>900</calories>
  </food>
  <food>
    <name>Cous Cous</name>
    <price curr="eur"></price>
    <serving>1</serving>
    <calories>1900</calories>
    <restrictions>
      <r compatible="hl" />
      <r incompatible="gf" />
    </restrictions>
  </food>
  <food>
    <name>Family Homestyle Breakfast</name>
    <price curr="eur"></price>
    <serving>3</serving>
    <calories>2950</calories>
    <restrictions>
      <r incompatible="vg" />
      <r incompatible="gf" />
      <r>Only on Sunday</r>
    </restrictions>
  </food>
  <dietary_restrictions>
    <restriction code="gf">Gluten Intollerance</restriction>
    <restriction code="vg">Vegan</restriction>
    <restriction code="hl">Halal</restriction>
  </dietary_restrictions>
</menu>
```

Esercizio 2 (4 punti). Rispondere alle seguenti domande

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

2pt) Scrivere il frammento di un possibile XML schema degli elementi `<price>` e `<restrictions>` includendo le chiavi esportate

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

2pt) Restituire i nomi e prezzo di tutti i cibi incompatibili con Gluten Intolerance.

```
<list><food name="name" price="price" />...</list>
```

2pt) Restituire i piatti con piú di una dietary restriction e il numero (solo incompatible) ordinati per serving

```
<foods><food name="name" serving="serving">num</food> ... </foods>
```

2pt) Trovare le dietary restriction con meno di 2 piatti.

```
<list><r name="name" />...</list>
```

Esercizio 4 (6 punti). Data la tabella con nome persona, colore preferito e città di provenienza.

2pt) ** Definisci in una frase il concetto di *Edit-Distance*. Calcolare Edit distance tra i nomi delle tuple 2, 4 e 5

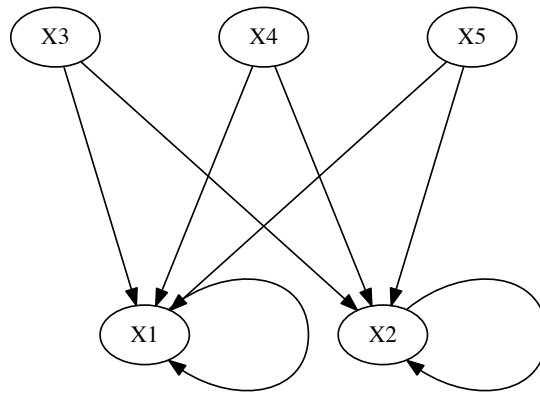
Id	Nome	Colore	Città
1	Alice	Rosso	TN
2	Bob	Rosso	MI
3	Carol	Blu	VR
4	Bilbo	Giallo	TN
5	Bobby	Verde	MI
6	Carl	Blu	VR
7	Alice	Rosso	TN

2pt) Applicare l'algoritmo di entity resolution utilizzando *Set-valued attributes*. Si considerano due tuple simili, se il valore della colonna *Nome* ha le stesse due lettere iniziali.

2pt) Considerare come funzione di distanza tra due tuple il numero di valori diversi (Es. tupla 1 e 2 distanza 3, tupla 1 e 7 distanza 1). Mostrare i passaggi di clustering gerarchico per ottenere 3 cluster.

Esplicitare eventuali assunzioni

Esercizio 5 (4 punti). Considerare la rete in figura, formata da 5 nodi, teleport set $\{X1, X3, X5\}$.



2pt) ** Calcolare il pagerank con $\beta = 1$ e $\beta = 0$.

2pt) Calcolare pagerank con $\beta = 0.8$ utilizzando le prime 2 iterazioni del metodo con relaxation

Esercizio 6 (4 punti). Sia dato un database per il noleggio di auto. Per ogni auto si ha marca, colore, e anno di immatricolazione. Per ogni cliente si ha professione, età e città. Si vuole analizzare per ogni cliente, l'auto noleggiata, la data, il numero di giorni, e costo totale.

2pt) Definire fact table e dimension tables per questo scenario.

2pt) Scrivere la query per definire il **formal data cube** con schema

`NCube(data, colore, marca, citta, numero_auto, somma_costo, massimo_costo, minimo_costo)`

Esercizio 7 (5 punti). Siano dati i seguenti elementi Aa, Bb, Cc, Dd, Ee , che appaiono nei basket:

$$B1 = \{Aa, Bb\}$$

$$B2 = \{Aa, Ee\}$$

$$B3 = \{Bb, Cc, Dd\}$$

$$B4 = \{Aa, Cc, Dd\}$$

$$B5 = \{Aa, Cc, Dd, Ee\}$$

$$B6 = \{Bb, Cc, Dd, Ee\}$$

$$B7 = \{Bb, Cc, Dd, Ee\}$$

$$B8 = \{Aa, Cc, Dd, Ee\}$$

1pt) ** Trovare 3 sottoinsiemi di elementi con supporto esattamente 3.

2pt) Calcolare 3 regole di associazione con supporto minimo = 3 e confidenza $\geq 50\%$.

2pt) Calcolare i valori di min-hash signature di $B2, B4, B6, eB8$ utilizzando le seguenti permutazioni.

P1: Aa, Bb, Cc, Dd, Ee , P2: Cc, Aa, Dd, Bb, Ee , P3: Ee, Cc, Dd, Aa, Bb