

Cognome e Nome:

Matricola:

(Scrivere solo nello spazio bianco. Se necessario, usare il retro del foglio.

Non sono ammessi elaborati su fogli diversi.)

Esercizio 1 (punti 5 su 30)

Data una tabella con 6 tuple. Si supponga che su un attributo X valgano le seguenti relazioni:

$$t1[X] = t2[X] = t3[X] \neq t4[X] = t5[X] \approx t6[X] \neq t1[X]$$

$$t1[Y] \neq t2[Y] = t3[Y] \approx t4[Y] = t5[Y] = t6[Y] \neq t1[Y]$$

dove il simbolo \approx indica che è possibile trovare una soglia ragionevole affinché due tuple siano simili, mentre ciò non vale per il simbolo \neq .

a) Dire se esiste una FD e/o una RFD $X \rightarrow Y$. Qualora valga una RFD che rilassi sull'extent, indicare il relativo g3 error.

b) Modificare le relazioni di cui sopra in modo che possa valere una FD.

Esercizio 2 (punti 5 su 30)

Disegnare la struttura a lattice per l'estrazione di dipendenze funzionali da un dataset con attributi A, B, C, D, E, indicando il numero totale di dipendenze candidate.

Esercizio 3 (punti 5 su 30)

Dati i seguenti due schemi di database:

// DB1 Matricole pari

Studente(Matricola, Codice Fiscale, Cognome, Nome, CFU superati)

Esami-2021(Matricola-Studente, Corso, Docente, CFU, Data, Voto, Lode)

// DB2 matricole dispari

Studente(Matricola, Cognome, Nome, Anno-Iscrizione, Corso-Studi, CFU Superati)

Esame-2022(Matricola-Studente, Corso, Docente, CFU, Data, Voto, Lode)

Integrare i due schemi tramite una vista ed utilizzando l'approccio GAV.

Esercizio 4 (punti 6 su 30)

Scrivere le funzioni Map e Reduce di un programma distribuito che legga in input il database integrato dell'esercizio 3 e dica per ogni corso il numero di studenti che hanno superato il relativo esame nel biennio 2021-2022.

Esercizio 5 (punti 6 su 30)

Data la seguente signature matrix:

| Shingle | S ₁ | S ₂ | S ₃ | S ₄ |
|---------|----------------|----------------|----------------|----------------|
| 0 | 1 | 0 | 1 | 0 |
| 1 | 0 | 0 | 1 | 0 |
| 2 | 1 | 0 | 1 | 1 |
| 3 | 1 | 0 | 0 | 1 |
| 4 | 0 | 1 | 0 | 0 |
| 5 | 0 | 1 | 1 | 1 |

- a) (1 punto) Calcolare la similarità di Jaccard tra ogni coppia di colonne;
- b) (3 punti) Calcolare la signature di ogni colonna usando le seguenti 3 funzioni hash:

$$h1(x)=(2x^4 + 2) \bmod 6; h2(x)=(3x^2 + 6) \bmod 6; h3(x)=(x^3 + 9) \bmod 6.$$

Mostrare l'evoluzione della matrice delle signature di minhash, simulando l'esecuzione dell'algoritmo per il loro calcolo. Inoltre, calcolare le similarità di Jaccard tra tutte le coppie di signature di minhash.

- c) (2 punti) Sulla base delle similarità di Jaccard calcolate al punto b), per ogni coppia di colonne calcolare la probabilità che la coppia venga selezionata per il confronto, ipotizzando che la matrice delle signature sia partizionata in 3 bande di 2 righe ciascuna.

Esercizio 6 (punti 3 su 30)

Data la seguente confusion matrix:

| | Predicted Negative | Predicted Positive |
|---------------|--------------------|--------------------|
| True Negative | 141 | 20 |
| True Positive | 40 | 49 |

Calcolare il True Positive Rate, la Precision, la Recall e l'F1-measure.