

DMD HW6

Nikolay Troshkov

September 2015

Question 1

- (a) Assuming uniform distribution, there are approximately

$$\frac{60,000}{90} = 666.(6)$$

tuples for each year. In the database we only have years starting from 1925, so we have no tuples for the year 1924. Therefore, we have $666.(6) \approx 666.6667$ tuples as a result.

- (b) Considering the prior calculations, we have $666.(6)$ tuples for the year 1999. Also, we need to calculate expected value for the number of tuples where title = "Fight Club". Having 60,000 tuples with 40,000 unique titles among them, we have

$$\frac{60,000}{40,000} = 1.5$$

tuples with title = "Fight Club".

According to the inclusion-exclusion principle we have to exclude the amount of tuples where title = "Fight Club" and year = 1999 from the union of results of the queries with two given conditions. Having 1.5 tuples for "Fight Club" and 90 distinct years, we have

$$\frac{1.5}{90} = 0.01(6)$$

tuples where title = "Fight Club" and year = 1999. Therefore, as a result of the original query we have

$$666.(6) + 1.5 - 0.01(6) = 668.15$$

tuples.

- (c) Considering uniform distribution, there are approximately

$$\frac{1,300,000}{5} = 260,000$$

reviews for each distinct rating. Therefore, there are approximately

$$1,300,000 - 260,000 = 1,040,000$$

tuples in the result of the query.

(d) There are 40,000 unique movie titles, therefore the query will result in 40,000 tuples.

(e) Assuming that the relation Reviews contains only reviews of the movies present in the relation Movies and considering grouping by movie title, the result of the query will consist of 33,000 tuples.

Question 2

2.1

(a)

- i Yes, because there are different values of B for the same value of A.
- ii No, because equal values on the right side of a FD are allowed for different values on the left side.
- iii Yes, because of the same reason as in the first case.
- iv No, because all combinations of the left side attributes are unique.
- v No, all the values of the B are unique.

(b) No, as long as there can be any set of data for each attribute and our assumptions can be proven wrong.

2.2

(a) The minimum cover is (b).

(b)

- i Yes. Transitivity from (1) to (2).
- ii Yes. From (1), (3) and (4).
- iii No.
- iv No.

(c) False.

$$\begin{aligned}(Q \rightarrow Q \wedge Q \rightarrow R) &\rightarrow QR \\ Q \rightarrow QR\end{aligned}$$

(d) True.

$$(PS \rightarrow PS \wedge PS \rightarrow P \wedge P \rightarrow Q \wedge Q \rightarrow R \wedge PS \rightarrow TRV) \rightarrow PQRSTV \quad (1)$$

$$(PQRSTV \rightarrow QT \wedge QT \rightarrow UR) \rightarrow PQRSTUV \quad (2)$$

$$(1) \wedge (2) \rightarrow (PS \rightarrow PQRSTUV)$$

Discussed with

- Artem Kondyukov
- Vladimir Marochko