

Midterm Review

CompSci 316 Fall 2022

Discussion - 5

9/30/2022

Roadmap

- Practice problem True/False
- Practice problem on FD, BCNF
- Practice problem on SQL, RA
- Practice problem on E/R
- Q/A Session

Sample Question 1 - True/False

For each of the following questions, please answer “True” or “False”

- a. Consider relation $R(A,B,C)$. $\{A\}$ and $\{A,C\}$ cannot both be keys of R at the same time.
- b. Given two relations $R(A,B)$ and $S(C,D)$ both without any nulls, the following equality holds:

$$R - \pi_{A,B} [R \bowtie_{B=C} S] = \pi_{A,B} [R \bowtie_{B \neq C} S]$$

- c. Consider relations $R(A,B)$, $S(B,C)$, and $T(C,A)$ and each relation has n tuples ($|R| = |S| = |T| = n$). The number of tuples in $\pi_{A,B,C} (R \bowtie S \bowtie T)$ can be more than n . (assume natural join)
- d. A relation R (using set semantic) has at least one superkey.
- e. A left outer join can always be written using a right outer join.

Answers to True/False

- a. T. {A,C} would be a superkey. Key must be minimal.
- b. F. Try R(1,1), R(1,2), S(1,3), and S(3,4).
- c. T
When you join $R \bowtie S \bowtie T$ to get (A, B, C), you do not have the guarantee that a combination (A, B) can not combine with multiple values of C. For instance, try R(0, 0), R(0, 1), R(1, 0).
S(0, 0), S(0, 1), S(1, 0).
T(0, 0), T(0, 1), T(1, 1).
 $R \bowtie S \bowtie T = \{(0,0,0), (0, 1, 0), (1, 0, 1), (0, 0, 1)\}$
- d. T. Set of all attributes is a superkey.
- e. T. Just reverse the relations.

Sample Question 2 - FD, closure, BCNF

Consider a relation $R(A, B, C, D, E, F)$ with functional dependencies:

- $A \rightarrow C$
- $B \rightarrow C$
- $AF \rightarrow D$
- $ACD \rightarrow E$
- $CD \rightarrow A$

Find all keys from this relation. Give a brief explanation.

Using your answer, find a superkey from this relation that is neither a key nor ABCDEF.

Answer 2:

Part 1: ABF and BDF

First, notice that B and F don't appear in any right side, so they must be in any key. Thus, we might as well assume B and F are in any closure we care about, and remove them from left sides. And since $B \rightarrow C$ then we can remove C as well. That leaves us with $A \rightarrow D$, $D \rightarrow A$ and $AD \rightarrow E$. To get A, D, E in a closure, we can start with either A alone, or D. Thus, the only keys are ABF and BDF.

Part 2: Any of ABCF or ABDF or ABEF or BCDF or BDEF or ABCDF or ABCEF or ABDEF or BCDEF

There are ten superkeys other than the nine keys above. These are the proper supersets of BDF or ABF, namely ABCDEF and all above. Thus, there are two keys, twelve superkeys, and nine superkeys that are not keys nor ABCDEF.

Sample Question 3 - SQL

Consider the schema:

- Professor (pid, name)
- Teaches (pid, cid, sem, year)
- Course (cid, title, dept) : dept is a new attribute denoting the department of the course, like CompSci stands for CS courses.

Write a SQL query to output pid and names of professors who never taught a course before the year 2007.

You may have to write such queries on paper - so practice now!
(5-10 mins)

Solution 3: option 1, 2

```
SELECT P.pid, P.name  
FROM Professor P left outer join Teaches T on P.pid = T.pid  
GROUP BY P.pid, P.name  
HAVING MIN(year) >= 2007 or COUNT(year) = 0;
```

```
SELECT P.pid, P.name  
FROM Professor P  
WHERE NOT EXISTS (  
  SELECT *  
  FROM Teaches T  
  WHERE T.pid = P.pid AND T.year < 2007  
)
```

Consider the schema:

- **Professor (pid, name)**
- **Teaches (pid, cid, sem, year)**
- **Course (cid, title, dept) : dept is a new attribute denoting the department of the course, like CompSci stands for CS courses.**

Output pid and names of professors who never taught a course before the year 2007.

Solution 3 - Option 3, 4

```
SELECT P.pid, P.name  
FROM Professor P  
WHERE P.pid NOT IN (  
  SELECT pid FROM Teaches WHERE year < 2007  
)
```

```
(SELECT pid, name FROM Professor P)  
EXCEPT  
(SELECT P.pid, P.name FROM Professor P, Teaches T WHERE P.pid = T.pid AND year < 2007);
```

Consider the schema:

- Professor (pid, name)
- Teaches (pid, cid, sem, year)
- Course (cid, title, dept) : dept is a new attribute denoting the department of the course, like CompSci stands for CS courses.

Output pid and names of professors who never taught a course before the year 2007.

Sample Question 4 - RA

Consider the schema:

- Professor (pid, name)
- Teaches (pid, cid, sem, year)
- Course (cid, title, dept) : dept is a new attribute denoting the department of the course, like CompSci stands for CS courses.

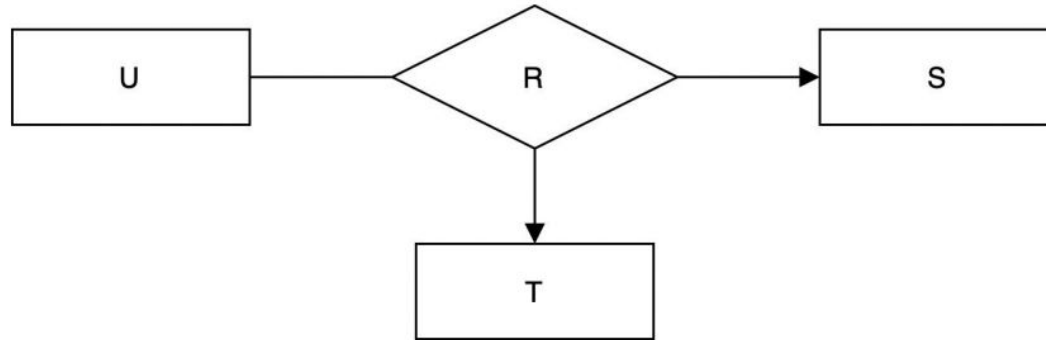
Write a RA query to output pid and names of professors who never taught a course before the year 2007.

Remember to use the same schema on both sides of difference (-)

Solution 4 - RA

```
Professor \diff \project_{pid, name} \select_{Teaches.year < 2007} (Professor  
\join_{Professor.pid = Teaches.pid} Teaches)
```

Question 5 ER Diagram



Suppose u, r, t, s be the number of tuples in U, R, T, S respectively. Are the following possible combinations of values for u, r, t, s ? Write “Yes” (for valid), and “No” (for invalid).

(a) $u = 4, s = 3, t = 2, r = 1$

(b) $u = 50, s = 1, t = 1, r = 100$

Solution Q5

- a. Yes
- b. No

Remember to read constraint from each arrow separately

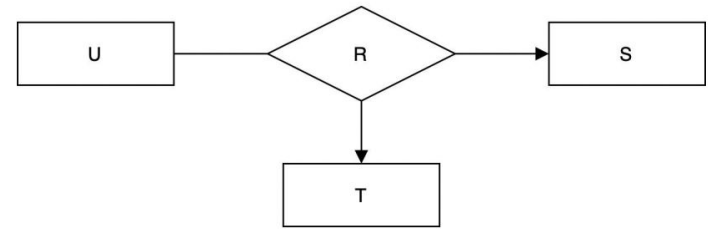
For every combination of U & S, there can be at most one value of T

For every combination of U & T, there can be at most one value of S

Since U and S together determine at most one T, there cannot be two tuples in R that agree on both U and S. Thus, $r \leq u * s$.

Likewise, because U and T together determine at most one S, there cannot be two tuples in R that agree on both U and T. So $r \leq u * t$.

This is satisfied by (a) not by (b)



(a) $u = 4, s = 3, t = 2, r = 1$

(b) $u = 50, s = 1, t = 1, r = 100$