CS2102 Tutorial 1: Relational Model   Wk 3, Sem 2, 2021/22

Questions to be presented: 1, 2, 3(a)-(d), 4(a)-(d)

1. Consider the following relation instance r of the relational schema $R(A, B, C, D)$.

R

A	B	C	D
0	0	0	1
2	1	2	0
1	1	2	0
0	0	1	2

   (a) Assuming that r is a *valid* relation instance of R, write down all the *possible* superkeys of R.

   (b) Additionally, suppose that it is also known that $\{A, C\}$ is *definitely* a superkey of R. Based on the additional information, write down all the *possible* candidate keys of R. Which of these (*if any*) must be the candidate key of R?

2. Consider a relational database consisting of two relations with schema $R(A, B)$ and $S(W, X, Y, Z)$ such that A is the primary key of R and W is the primary key of S (*for future use, we denote this with* $R(\underline{A}, B)$ *and* $S(\underline{W}, X, Y, Z)$ *where the attributes being underlined are parts of primary key*). Let r and s be the current instances of R and S, respectively, as shown below.

R

A	B
3	0
2	1
1	1
0	0

S

W	X	Y	Z
0	4	0	NULL
1	NULL	2	NULL
2	1	2	NULL
3	0	1	NULL

Based on the current database instance above, write down all the *possible* foreign keys in S that refer to attribute A in R.

W, Y, Z

They include values which appear in the primary key of R or are null values

X cannot be a foreign key as the value 4 cannot be found in attribute A

3. Two queries Q_1 and Q_2 on a relational database with schema D are defined to be **equivalent queries** (denoted by $Q_1 \equiv Q_2$) if for *every* valid instance d of D, both Q_1 and Q_2 always compute the same results on d.

   Consider a database with the following relational schema: $R(\underline{A}, C)$, $S(\underline{A}, D)$, and $T(\underline{X}, Y)$, with primary key attributes underlined. Assume all the attributes have integer domain. For each of the following pairs of queries Q_1 and Q_2, state whether or not $Q_1 \equiv Q_2$.

   (a) $Q_1 = \pi_A(\sigma_{A<10}(R))$ **and** $Q_2 = \sigma_{A<10}(\pi_A(R))$

   (b) $Q_1 = \pi_A(\sigma_{C<10}(R))$ **and** $Q_2 = \sigma_{C<10}(\pi_A(R))$

   (c) $Q_1 = \pi_{D,Y}(S \times T)$ **and** $Q_2 = \pi_D(S) \times \pi_Y(T)$

   (d) $Q_1 = \pi_{D,Y}(S \times T)$ **and** $Q_2 = \pi_{D,Y}(T \times S)$

   (e) $Q_1 = (R \times \pi_D(S)) \times T$ **and** $Q_2 = R \times (\pi_D(S) \times T)$

   (f) $Q_1 = \pi_A(R \cup S)$ **and** $Q_2 = \pi_A(R) \cup \pi_A(S)$

   (g) $Q_1 = \pi_A(R - S)$ **and** $Q_2 = \pi_A(R) - \pi_A(S)$

4. Consider the following relational database schema discussed in lecture. We further add the primary keys of each relation in underline.

   - `Pizza(pizza)`: *All the pizzas of interest.*
   - `Customers(cname, area)`: *The name and location of each customer.*
   - `Restaurants(rname, area)`: *The name and location of each restaurant.*
   - `Contains(pizza, ingredient)`: *The ingredients used in each pizza.*
   - `Sells(rname, pizza, price)`: *Pizzas sold by restaurants and the prices.*
   - `Likes(cname, pizza)`: *Pizzas that customers like.*

   Additionally, we have the following foreign key constraints on the database schema:

   - `Contains.pizza` is a foreign key that refers to `Pizzas.pizza`
   - `Sells.rname` is a foreign key that refers to `Restaurants.rname`
   - `Sells.pizza` is a foreign key that refers to `Pizzas.pizza`
   - `Likes.cname` is a foreign key that refers to `Customers.cname`
   - `Likes.pizza` is a foreign key that refers to `Pizzas.pizza`

   Answer each of the following queries using relational algebra.

   (a) Find all pizzas that Alice likes but is not liked by Bob[1].

   (b) Find all customer-restaurant pairs (C, R) where C and R both located in the same area and C likes some pizza that is sold by R.

   (c) Suppose the relation `Likes` contains all information about all customers. In other words, if the pair (`cname`, `pizza`) is not in the relation `likes`, it means that the customer `cname` *dislikes* the pizza `pizza`. Write a relational algebra expression to find for all customers, the pizza that they dislike. The result should be of the form (`cname`, `pizza`).

   (d) Consider having a relation `Dislikes(cname, pizza)` that are created based on the query from Part (B). Find all customer pairs $(C1, C2)$ such that $C1$ likes some pizza that $C2$ does not like.

   (e) Find all customer pairs $(C1, C2)$ such that $C1 < C2$ and they like *exactly* the same pizzas. You may assume that you have a relation `LikesDislikes(C1,C2)` that are created based on the query from Part (C). *Exclude* pairs of customers who do not like any pizza.

   (f) For each restaurant, find the price of the most expensive pizzas sold by that restaurant. Exclude restaurants that do not sell any pizza.

   (g) Find all customer-pizza pairs (C, P) where the pizza P sold by some restaurant that is located in the same area as that of the customer C. Include customers whose associated set of pizzas is empty.

[1] The intention is to state it as *"all pizzas that Alice likes but Bob does not like"*, however, this may indicate dislike which we have not discussed the underlying assumption yet.

5. Consider the following relational algebra query expressed on the database schema in Question 4.

$$
\begin{aligned}
R_1 &:= \pi_{\texttt{pizza}}(\sigma_{\texttt{cname='Maggie'}}(\texttt{Likes}))\\
R_2 &:= \pi_{\texttt{rname}}(\texttt{Sells}) \times R_1\\
R_3 &:= \pi_{\texttt{rname}}(R_2 - \pi_{\texttt{rname,pizza}}(\texttt{Sells}))\\
R_4 &:= \pi_{\texttt{rname}}(\texttt{Sells}) \times R_3\\
R_5 &:= \pi_{\texttt{pizza}}(\sigma_{\texttt{cname='Ralph'}}(\texttt{Likes}))\\
R_6 &:= \pi_{\texttt{rname}}(\sigma_{\texttt{pizza5=pizza}}((\texttt{Sells} \times \rho_{\texttt{pizza}\leftarrow\texttt{pizza5}}(R_5))))\\
R_7 &:= R_4 - R_6
\end{aligned}
$$

For each of the relational algebra expression R_i, write down a concise English sentence to precisely describe the information retrieved by R_i.