BIT2008A/ITEC2000A: Multimedia Data Mangement
Fall 2022

Assignment 1

Relational Algebra

Question 1. We should include at least one candidate key in our superkey. To do so we have k options which is formally shown by $\binom{k}{1}$. In that superkey other non-prime attributes may or may not exist (two options). As a result we can show them by 2^{n-k}. Consequently, we have $\binom{k}{1} 2^{n-k}$ superkeys with only one prime attribute. We can use a combination of two prime attributes with arbitrary number of non-prime attributes. With the same logic they will sum up to $\binom{k}{2} 2^{n-k} = k\,(k-1)\,2^{n-k}$ options. We can continue the same process all the way to using all the prime attributes which has $\binom{k}{k} 2^{n-k} = 2^{n-k}$ options. The way that we counted different options do not have overlaps so we can sum them to have the solution.

$$\sum_{i=1}^{k}\left(\binom{k}{i} 2^{n-k}\right) = 2^{n-k}\sum_{i=1}^{k}\binom{k}{i} = 2^{n-k}\left(2^k - 1\right) = 2^n - 2^{n-k}$$

We can interpret the solution in another way. We can take any subset of attributes as a superkey except those with no prime-attributes. As a result, we have 2^n possible subsets and if we remove all of the prime attributes and compute their subsets we will have 2^{n-k} possibilities. Finally we can take undesirable subsets of attributes out of the total number of possibilities which leads to the same answer.

Question 2.

1) $\sigma_\theta(R \cup S) = \sigma_\theta(R) \cup \sigma_\theta(S)$ is always valid.

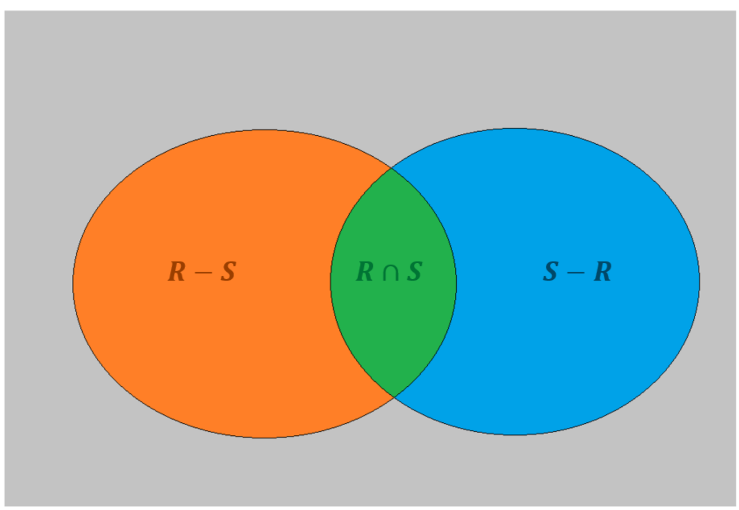

Figure 1

Let $X = \sigma_\theta(R \cup S)$, where θ is a set of conditions to apply to the argument of the operation. Without loosing the generality, we denote the all the records in the $R - S$ with A, $S - R$ with B, and $R \cap S$ with C. These records clearly do not overlap. So if the records in $S \cup R$ satisfying the condition set θ are $X = \{x_1, x_2, ..., x_i\}$, we can order them like $\{a_1, ..., a_j, b_1, ..., b_k, c_1, ..., c_l\}$. Let's denote $A' = \{a_1, ..., a_j\}$, $B' = \{b_1, ..., b_k\}$, $C' = \{c_1, ..., c_l\}$. Remember that we can shuffle the records however we want.

$$\forall x \in X, x \in (A' \cup B' \cup C')$$
$$x \in (A' \cup B') \cup (B' \cup C')$$
$$x \in \sigma_\theta(R) \cup \sigma_\theta(S)$$

2) $\Pi_L(\sigma_\theta(S)) = \sigma_\theta(\Pi_L(S))$ is not always valid.
Clearly if θ contains conditions on attributes not listed in L, then we cannot swap selection and projection operators.
3) $\sigma_\theta(R) - S = \sigma_\theta(R - S)$ is always valid.

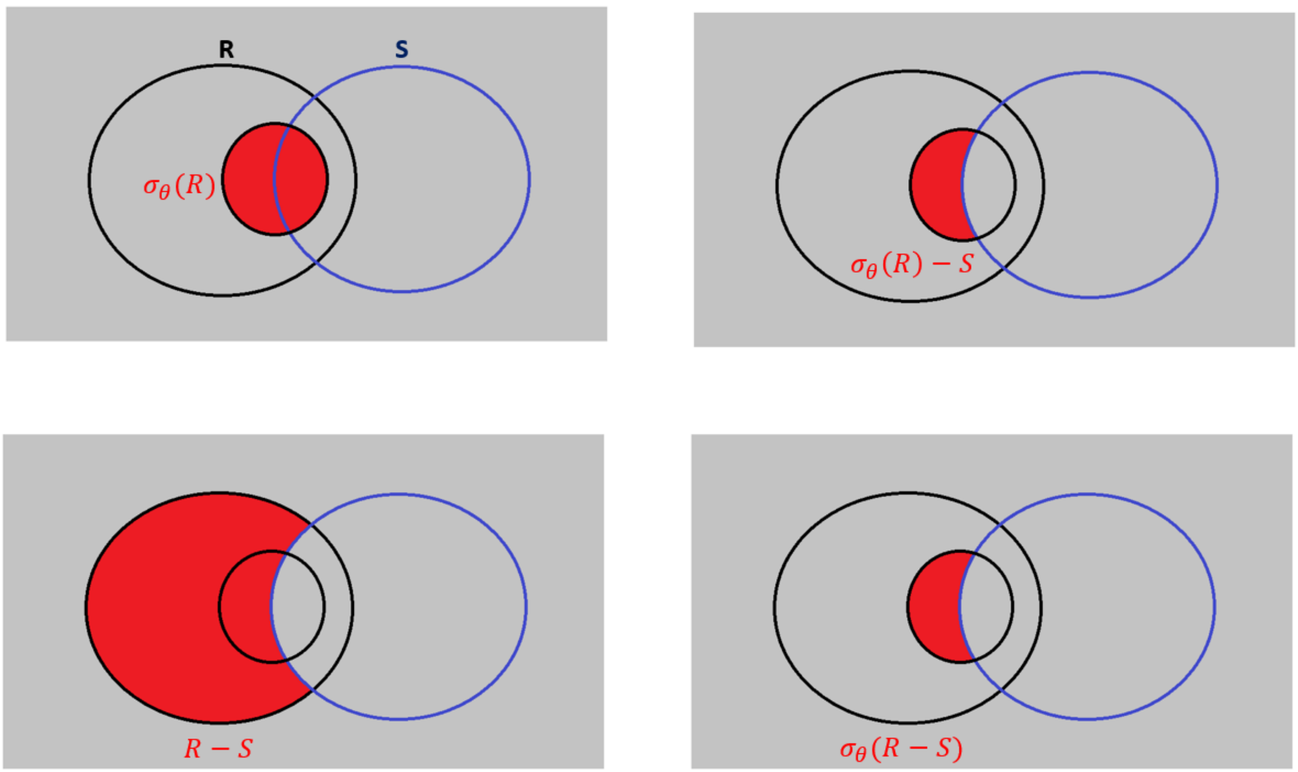

Figure 2

Question 3. Assume that we have manged to store the data of a transportation system as the relations below:

- Company(comp_id, name)
- Trip(trip_id, date, origin, dest, comp_id, duration)
- Passenger(pass_id, name)
- Pass_in_Trip(trip_id, pass_id,

Please write each of these queries using relational algebra. (15 points)

1) The date of every trip starting from "Ottawa".

$$\Pi_{date}\,(\sigma_{origin='Ottawa'}\,(Trip))$$

2) The list of every city that a passenger named "Alex" had a trip to.

$$\Pi_{dest}\,(Trip \bowtie \Pi_{trip\_id}\,(Pass\_in\_Trip \bowtie \Pi_{pass\_id}\,(\sigma_{name='Alex'}\,(Passenger))))$$

3) The ID of every passenger who has a trip after 2022-08-08, taking less than 12 minutes.

$$\Pi_{pass\_id}\,(Pass\_in\_Trip \bowtie \Pi_{trip\_id}\,(\sigma_{date>'2022-08-08' \wedge duration>12}\,(Trip)))$$

Question 4.

$$\Pi_{s\_name}(\sigma_{Producer.s\_city=P2.s\_city}(Producer \times \rho_{P2}(\Pi_{s\_city}(\sigma_{s\_id=8}(Producer)))))$$

This query returns the name of producers who are active in the same city as the producer with ID 8.