

# Bazy Danych 2024

## Problem list 2

- 1 (1 pt) Let's consider the relation  $R(A, B, C)$ . Write a relational algebra query and an DRC/TRC (in Polish: RRD/RRK) query that returns an empty result if and only if the attribute pair  $A, B$  is a key of the relation  $R$ .
- 2 (1 pt) Let's consider the relations  $R(A, B, C)$  and  $S(X, Z)$ , where the attribute  $A$  is a key in  $R$ . Write a relational algebra query and an DRC/TRC query that will return an empty result if and only if the attribute  $Z$  of the relation  $S$  is a foreign key referencing the attribute  $A$  of the relation  $R$ .
- 3 (1 pt) Given relations  $R, S$ , and  $T$  with schemas  $R = AB$ ,  $S = B_1B_2$ , and  $T = BC$ , analyze the meaning of the following queries and try to find a natural interpretation for the relations and queries in English/Polish. Consider whether they are domain-independent formulas. Write equivalent formulas in relational algebra whenever possible.
  1.  $\{a \mid (\exists b)(R(a, b) \wedge \neg((\exists a')a' > a \wedge (\exists b')(R(a', b'))))\}$
  2.  $\{a, b \mid (\forall c)(T(c, a) \vee T(c, b) \vee (\forall d)(\neg T(c, d)))\}$
- 4 (1 pt) Decide if the following equations hold. Provide a proof or a counterexample.
  - $R \bowtie S = S \bowtie R$
  - $(R \bowtie S) \bowtie T = R \bowtie (S \bowtie T)$
- 5 (2 pts, 0.5 pt per each item) The database consists of the following relations:
  - $F(idf, title, director, prodYear, duration)$  —  $idf$  is the key;  $title$  and other attributes do not have to be unique;  $duration$  denotes the duration of the movie and is given in minutes;
  - $S(idf, hall, date, time)$  — there is a screening of the movie with the given identifier in the specified hall and at the specified date and time;
  - $A(pseudo, firstname, lastname, nationality, yearBirth)$  — information about actors;  $pseudo$  is unique;
  - $R(pseudo, idf, character, salary)$  — information that the actor with the given pseudonym played the given character in the specified movie and received the specified salary for it;
  - $M(pseudo, year, minSalary)$  — information that the actor with the given pseudonym in the specified year negotiated the specified minimum salary for playing in a movie.

Write the following queries in DRC or TRC:

1. Provide data about actors (pseudonym, first name, last name, year of birth, nationality) who appeared in movies produced only in one year.
2. Provide full tuples of movies that are the newest films of their respective directors.
3. For each movie, find the actor who received the highest salary in that movie (was the highest paid from the movie's cast). In the resulting relation, provide the actor's pseudonym, `idf`, and salary.
4. Provide full tuples of actors who have never lowered their minimum salary (it could only increase in later years). The result is not affected by the years in which the actor did not specify a minimum salary.

**6 (1 pt)** Let's assume an interpretation of NULL values in which they denote *some value of the appropriate type*, meaning that we know such a value exists, but we don't know what it is. Under this assumption, it is convenient to represent NULLs using variables. For example, if we have a relation with attributes (`Name:String`, `Salary:Int`), and a tuple (`Joe`, `x`), this means that `Joe` has some salary that can be expressed as a value of type `Int`, but we don't know what that value is. We assume that each variable can appear in the database at most once (which closely approximates the idea of NULLs in SQL).

Let  $D$  be a relation with variables. We define  $\text{rep}(D)$  to be the set of all complete relations (i.e., relations without variables) represented by  $D$ , that is,

$$\{v(D) \mid v \text{ is a valuation of all variables in } D\}$$

We should think of  $\text{rep}(D)$  as the set of all complete relations represented by  $D$ . For example, if  $D$  contains only the tuple (`Joe`, `x`), then  $\text{rep}(D)$  contains all relations with exactly one tuple of the form (`Joe`, `n`), where  $x$  has been instantiated with an integer value  $n \in \text{Int}$ .

Of course, in order for this whole business with NULLs to make sense, we can only use such relational algebra expressions  $Q$  that for any relation  $D$ , there exists a relation (with variables)  $Q_D$  that *represents the result of  $Q$  on  $D$* , meaning that  $\text{rep}(Q_D) = Q(\text{rep}(D))$ , where  $Q(\text{rep}(D))$  denotes the image of the set  $\text{rep}(D)$  under  $Q$ .

Give an example of a relation  $D$  and a query  $Q$  involving a single selection, such that there is no representation of the result of  $Q$  on  $D$ , i.e., there is no relation (with variables)  $Q_D$  such that  $\text{rep}(Q_D) = Q(\text{rep}(D))$ . This means that we cannot use selection queries in this system.

*Hint: observe that (non)emptiness of a relation does not depend on the valuation of its variables*