

# Bazy Danych 2021

## Lista zadań nr 2

1. (2 pkt.) Udowodnij, że dowolne zapytanie algebry relacji z operatorami  $\pi, \rho, \sigma_F, \times, \setminus$  można zapisać używając rrk (lub rrd). Zakładamy, że formuły w warunkach selekcji  $F$  są kombinacjami boolowskimi atomów postaci  $R.k \oplus \text{const}$  lub  $R.k \oplus R'.k'$ , gdzie  $k$  i  $k'$  to odpowiednio kolumny pewnych relacji  $R$  i  $R'$ , a  $\oplus \in \{=, \neq, >\}$
2. (1 pkt.) Rozważmy relacje  $R(A, B, C)$  oraz  $S(X, Z)$ , przy czym atrybut  $A$  jest kluczem w  $R$ . Napisz zapytanie rrk lub rrd, które zwróci pusty wynik wtedy i tylko wtedy jeśli atrybut  $Z$  relacji  $S$  spełnia własność klucza obcego wskazującego na atrybut  $A$  relacji  $R$ .
3. (1 pkt.) Dane są relacje  $R, S$  i  $T$  o schematach  $R = AB, S = B_1B_2$  i  $T = BC$ . Przeanalizuj znaczenie poniższych zapytań i postaraj się znaleźć naturalną interpretację dla relacji i zapytań w języku polskim. Zastanów się, czy są to formuły niezależne od dziedziny. Zapisz równoważne im formuły w algebrze relacji zawsze jeśli to możliwe.

1.  $\{a \mid (\exists b)(R(a, b) \wedge \neg((\exists a')a' > a \wedge (\exists b')(R(a', b'))))\}$
2.  $\{a, b \mid (\forall c)(T(c, a) \vee T(c, b) \vee (\forall d)(\neg T(c, d)))\}$

W kolejnych zadaniach będziemy odwoływać się do bazy złożonej z relacji:

- $B(\text{osoba}, \text{bar})$ , czyli bywa osoba w barze,
- $P(\text{sok}, \text{bar})$ , czyli podają sok w barze,
- $L(\text{osoba}, \text{sok})$ , czyli lubi osoba sok.

W każdym z poniższych podpunktów wskaż, które zapytanie rrk lub rrd jest równoważne zapytaniu wyrażonemu w języku polskim (nie zawsze musi być dokładnie jedna poprawna odpowiedź, nie zawsze musi być poprawna odpowiedź pośród podanych):

4. (1 pkt.) Wypisz osoby bywające tylko w tych barach, w których podaje się (przynajmniej) jeden z ich ulubionych soków.
1.  $\{o \mid (\exists b)(B(o, b)) \wedge \neg(\exists b)(B(o, b) \wedge (\forall s)(P(s, b) \Rightarrow \neg L(o, s)))\}$
  2.  $\{o \mid (\exists b)(B(o, b)) \wedge \neg(\exists b)(B(o, b) \wedge (\forall s)(P(s, b) \Rightarrow L(o, s)))\}$
  3.  $\{o \mid (\exists b)(B(o, b)) \wedge (\forall b)(B(o, b) \Rightarrow (\exists s)(P(s, b) \wedge L(o, s)))\}$
  4.  $\{o \mid (\exists b)(B(o, b)) \wedge (\forall b)(B(o, b) \wedge (\exists s)(P(s, b) \wedge L(o, s)))\}$

5. (1 pkt.) Podaj osoby chodzące tylko do jednego baru.

1.  $\{o \mid (\exists b)(B(o, b) \wedge \neg(\exists b')(BYWA(o, b'))))\}$
2.  $\{o \mid (\exists b)(B(o, b))\}$
3.  $\{o \mid (\exists b)(B(o, b) \wedge \neg(\exists b')(b \neq b' \wedge BYWA(o, b'))))\}$
4.  $\{o \mid (\exists b)(B(o, b) \wedge \neg(\exists b', o')(b \neq b' \wedge o = o' \wedge BYWA(o', b'))))\}$

6. (2 pkt. - po 0.5 pkt. za podpunkt) Baza danych składa się z relacji:

- $F(idf, tytuł, reżyser, rokProd, czas)$  — idf jest kluczem; tytuł i inne atrybuty nie muszą być unikalne; czas oznacza czas trwania filmu i jest podany w minutach;
- $S(idf, sala, data, godz)$  — w podanej sali i terminie jest projekcja filmu o podanym identyfikatorze;
- $A(pseudo, imię, nazwisko, narodowość, rokUr)$  — informacje o aktorach; pseudonim jest unikalny;
- $R(pseudo, idf, postać, gaża)$  — informacja, że aktor o podanym pseudonimie grał w filmie daną postać i otrzymał za to podaną gażę.
- $M(pseudo, rok, minGaza)$  — informacja, że aktor o podanym pseudonimie w danym roku na podanym poziomie ustalił minimalną gażę za grę w filmie.

Zapisz poniższe zapytania w rrd lub rrk.

1. Podaj dane aktorów (pseudonim, imię, nazwisko, rok urodzenia, narodowość), którzy pojawili się w filmach produkowanych tylko w jednym roku (powiedzmy, że są to gwiazdy jednego sezonu).
2. Podaj pełne krotki filmów, które są najnowszymi filmami reżyserów.
3. Dla każdego filmu znajdź aktora, który dostał najwyższą gażę w tym filmie (został najlepiej opłacony z obsady filmu). W relacji wynikowej podaj pseudonim aktora, idf oraz gażę.
4. Podaj pełne krotki aktorów, którzy nigdy nie obniżyli swojej minimalnej gaży (w późniejszych latach mogła ona najwyżej rosnać). Na wynik nie wpływają lata, w których aktor nie podał minimalnej gaży.

7. (1 pkt.) Przyjmijmy taką interpretację wartości NULL, w której oznacza ona *jakąś wartość odpowiedniego typu*, tzn. wiemy, że taka wartość istnieje ale nie wiemy jaka ona jest. Przy takim założeniu wygodne jest zapisywanie NULLi za pomocą zmiennych tzn. jeśli w relacji o atrybutach ( $Imię:String$ ,  $Zarobki:Int$ ) jest krotka ( $Józek$ ,  $x$ ) to oznacza to, że Józek ma jakieś zarobki, które można wyrazić pewną wartością typu  $Int$ , ale nie wiemy jaką. Zakładamy, że każda zmienna może wystąpić w bazie danych co najwyżej jeden raz (co dość dobrze przybliża ideę NULLi z SQLa).

Niech  $D$  będzie relacją ze zmiennymi. Oznaczmy przez  $\text{rep}(D)$  następujący zbiór relacji

$$\{v(D) \mid v \text{ jest wartościowaniem wszystkich zmiennych z } D\}$$

O  $\text{rep}(D)$  należy myśleć, że jest zbiorem wszystkich *zupelnych* relacji (tj. relacji bez zmiennych) reprezentowanych przez  $D$ . Na przykład, jeśli  $D$  zawiera wyłącznie krotkę  $(\text{Józek}, x)$  to  $\text{rep}(D)$  zawiera wszystkie relacje z dokładnie jedną krotką postaci  $(\text{Józek}, n)$ , gdzie  $x$  została zwartościowana liczbą całkowitą  $n \in \text{Int}$ .

Oczywiście, żeby ta cała zabawa z NULLami miała sens możemy używać wyłącznie takich wyrażeń algebry relacji  $Q$ , że dla dowolnej relacji  $D$  istnieje relacja (ze zmiennymi)  $Q_D$  reprezentująca wynik  $Q$  na  $D$  tzn. taka, że  $\text{rep}(Q_D) = Q(\text{rep}(D))$ , przy czym przez  $Q(\text{rep}(D))$  oznaczamy obraz zbioru  $\text{rep}(D)$  przez  $Q$ .

Pokaż przykład relacji  $D$  i przykład zapytania  $Q$  będącego pojedynczą selekcją taką, że nie istnieje reprezentacja wyniku  $Q$  na  $D$  tj. nie istnieje relacja (ze zmiennymi)  $Q_D$ , taka że  $\text{rep}(Q_D) = Q(\text{rep}(D))$ . Oznacza to, że w tym systemie nie można używać zapytań z selekcją.

*Jeżeli zmiennych*

*Wskazówka: zauważ, że (nie)pustość relacji nie zależy od wartościowania*