

Bazy Danych 2022

Lista zadań nr 2

Z5 (2 pkt.) Rozważamy bazy danych o schemacie $E(A, B)$, gdzie E jest symbolem relacji binarnej, a A i B są jej atrybutami. Ustalmy pewne dowolne zapytanie rrd φ , bez zmiennych wolnych, odwołujące się wyłącznie do relacji E . Możesz założyć, że φ ma postać

$$\exists x_1 \forall x_2 \exists x_3 \dots \forall x_n \psi(x_1, \dots, x_n)$$

gdzie $\psi(x_1, \dots, x_n)$ jest zapytaniem rrd nie zawierającym kwantyfikatorów.

Rozważmy następujący problem.

Wejście: baza danych D o schemacie $E(A, B)$.

Wyjście: prawda jeśli w D spełnione jest φ , wpp. fałsz.

Podaj wielomianowy algorytm rozwiązujący powyższy problem. Czy coś by się zmieniło gdyby zapytanie φ nie było ustalone lecz podane na wejściu?

Wskazówka: sformułowanie „ustalmy zapytanie” pozwala Ci założyć, że istnieje pewna stała c taka, że rozmiar φ nie przekracza c , a w szczególności $n < c$.

Z6 (1 pkt.) Rozważmy relację $R(A, B, C)$. Napisz zapytanie algebry relacji oraz zapytanie rrd/rrk, które zwróci pusty wynik wtedy i tylko wtedy gdy para atrybutów A, B jest kluczem relacji R .

Z7 (1 pkt.) Rozważmy relacje $R(A, B, C)$ oraz $S(X, Z)$, przy czym atrybut A jest kluczem w R . Napisz zapytanie algebry relacji oraz zapytanie rrd/rrk, które zwróci pusty wynik wtedy i tylko wtedy gdy atrybut Z relacji S jest kluczem obcym wskazującym na atrybut A relacji R .

Z8 (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.

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

Z9 (2 pkt. - po 0.5 pkt. za podpunkt) Baza danych składa się z relacji:

- $F(\text{idf}, \text{tytuł}, \text{reżyser}, \text{rokProd}, \text{czas})$ — idf jest kluczem; tytuł i inne atrybuty nie muszą być unikalne; czas oznacza czas trwania filmu i jest podany w minutach;
- $S(\text{idf}, \text{sala}, \text{data}, \text{godz})$ — w podanej sali i terminie jest projekcja filmu o podanym identyfikatorze;

- $A(\text{pesudo}, \text{imie}, \text{nazwisko}, \text{narodowość}, \text{rokUr})$ — informacje o aktorach; pseudonim jest unikalny;
- $R(\text{pseudo}, \text{idf}, \text{postac}, \text{gaza})$ — informacja, że aktor o podanym pseudonimie grał w filmie daną postacią i otrzymał za to podaną gażę.
- $M(\text{pseudo}, \text{rok}, \text{minGaza})$ — 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.
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.

Z10 (1 pkt.) Przyjmijmy taką interpretacją 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 , \mathbf{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 *zupelných* relacji (tj. relacji bez zmiennych) reprezentowanych przez D . Na przykład, jeśli D zawiera wyłącznie krotkę (Józek , \mathbf{x}) to $\text{rep}(D)$ zawiera wszystkie relacje z dokładnie jedną krotką postaci (Józek , \mathbf{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.

Wskaźówka: zauważ, że (nie)pułstość relacji nie zależy od wartościowania jej zmiennych.