Lista zadań nr 2 6.03.2018

Bazy Danych 2018

1. (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 bezpieczne. Zapisz równoważne im formuły w algebrze relacji.

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

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

- *B(osoba,bar)*, czyli bywa osoba w barze,
- *P(sok,bar)*, czyli podają sok w barze,
- *L(osoba,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):

2. (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. $\{o \mid (\exists b)(B(o,b)) \wedge (\forall b)(B(o,b) \wedge (\exists s)(P(s,b) \wedge L(o,s)))\}$

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

4. (3 pkt. - po 0.5 pkt. za podpunkt) Baza danych składa się z relacji:

- `F(idf,tytul,rezyser,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(pesudo,imie, nazwisko,narodowość,rokUr)` — informacje o aktorach; pseudonim jest unikalny;
- `R(pseudo,idf,postac,gaza)` — 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 sale, w których odbyła się projekcja każdego filmu reżysera "Olańskiego". Załóż, że w każdej sali jest jakiś seans.
5. Podaj pełne krotki aktorów, którzy nigdy nie obniżyli swojej minimalnej gaży (w późniejszych latach mogła ona najwyżej rosnąć). Na wynik nie wpływają lata, w których aktor nie podał minimalnej gaży.
6. Podaj tytuły filmów, w których zagrał ktoś, kto nie grał w filmie "Roll".

5. (1pkt.) 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, 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.

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

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

O $\mathrm{rep}(D)$ należy myśleć, że jest zbiorem wszystkich *zupełnych* relacji (tj. relacji bez zmiennych) reprezentowanych przez D. Na przykład, jeśli D zawiera wyłącznie krotkę `(Józek, x)` to $\mathrm{rep}(D)$ zawiera wszystkie relacje z dokładnie jedną krotką postaci `(Józek, n)`, gdzie x została zwartościowana liczbą całkowitą $n \in$ `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 $\mathrm{rep}(Q_D) = Q(\mathrm{rep}(D))$, przy czym przez $Q(\mathrm{rep}(D))$ oznaczamy obraz zbioru $\mathrm{rep}(D)$ 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 $\mathrm{rep}(Q_D) = Q(\mathrm{rep}(D))$. Oznacza to, że w tym systemie nie można używać zapytań z selekcją.

Wskazówka: zauważ, że (nie)pustość relacji nie zależy od wartościowania jej zmiennych.

6. (4 pkt.) Zapytania koniunkcyjne to zapytania rrd zbudowane z formuł atomowych (np. $R(x, y)$) oraz koniunkcji i kwantyfikatorów egzystencjalnych. W ogólności formułami atomowymi mogą być również równości i nierówności między stałymi i zmiennymi (np. $x = 5$, $x \neq y$, $x < z$) ale w tym zadaniu pozwalamy wyłącznie na atomy relacyjne oraz nie pozwalamy na używanie stałych.

Rozważmy bazę danych reprezentującą pewien graf skierowany o krawędziach zapisanych w relacji $E(\mathrm{start}, \mathrm{end})$. Niestety w naszych zapytaniach nie możemy używać relacji E. W zamian mamy dostęp do relacji $P_i(x, y)$ dla pewnych $i > 1$. Relacja $P_i(x, y)$ zawiera pary wierzchołków połączone ścieżką długości i, np. $P_2(x, z)$ mogłaby by zdefiniowana jako $(\exists y)E(x, y) \wedge E(y, z)$. Odpowiada to sytuacji, w której np. ze względów bezpieczeństwa dostęp do bazy danych mamy wyłącznie za pomocą zestawu perspektyw (widoków), a dostęp do oryginalnych relacji jest zablokowany.

Jeśli chcemy wyliczyć odpowiedzi na jakieś zapytanie ψ używające relacji E możemy spróbować zmodyfikować (przepisać) ψ tak aby zamiast E wykorzystać symbole dostępnych perspektyw. Np. jeśli mamy wyłącznie dostęp do perspektywy $P_2(x, y)$, a chcemy zapisać zapytanie $P_4(x, z) = (\exists y_1, y_2, y_3)E(x, y_1) \wedge E(y_1, y_2) \wedge E(y_2, y_3) \wedge E(y_3, z)$ możemy to zrobić tak: $P_4'(x, z) = (\exists y)P_2(x, y) \wedge P_2(y, z)$ (zauważ, że P_4' też jest zapytaniem koniunkcyjnym i jest równoważne $P_4(x, y)$).

a) (0 pkt.) Pokaż, jak przepisać zapytanie $P_5(x, y)$ używając wyłącznie perspektyw $P_2(x, y)$ i $P_3(x, y)$.

b) (2 pkt.) Pokaż, że nie istnieje zapytanie koniunkcyjne używające wyłącznie perspektyw $P_3(x, y)$ i $P_4(x, y)$, które jest równoważne zapytaniu $P_5(x, y)$.

Wskazówka: rozważ dwa grafy, z których jeden jest ścieżką długości 5, drugi takiej ścieżki nie zawiera, a przy tym trudno go od tego pierwszego odróżnić zapytaniem pożądanej postaci.

c) (2 pkt.) Napisz zapytanie rrd (dozwolone $\exists, \forall$ i wszystkie spójniki boolowskie), które korzysta wyłącznie z perspektyw $P_3(x, y)$ i $P_4(x, y)$ i jest równoważne zapytaniu $P_5(x, y)$.