Lista zadań nr 3
Bazy Danych 2022

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 na tej liście pozwalamy wyłącznie na atomy relacyjne oraz nie pozwalamy na używanie stałych.

Rozważmy bazę danych reprezentującą pewien graf skierowany o krawędziach zapisanych w relacji $E(S, T)$. 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 $P_4(x, y)$).

Na rozgrzewkę pokaż, jak przepisać zapytanie $P_7(x, y)$ używając wyłącznie perspektyw $P_2(x, y)$ i $P_3(x, y)$.

Z11. (1 pkt.) Pokaż, że nie istnieje takie zapytanie koniunkcyjne używające jako formuł atomowych wyłącznie perspektyw $P_3(x, y)$ i $P_4(x, y)$, które jest równoważne zapytaniu $P_5(x, y)$.

Z12. (1 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 $P_5(x, y)$.

Dziedziną bazy danych A, $\mathrm{dom}(A)$, nazwiemy zbiór wszystkich elementów zawartych w krotkach w relacjach z A. Homomorfizm pomiędzy bazami danych A i B to funkcja $h : \mathrm{dom}(A) \to \mathrm{dom}(B)$ spełniająca warunek: dla każdego symbolu relacji R i każdej krotki $(a_1, \ldots, a_n) \in \mathrm{dom}(A)^n$ (gdzie n jest liczbą atrybutów R) zachodzi

$$A \models R(a_1, \ldots, a_n) \implies B \models R(h(a_1), \ldots, h(a_n))$$

Z13. (1 pkt)

1. Pokaż, że istnieje graf G, taki, że dla dowolnego grafu G' istnieje homomorfizm z G' w G.
2. Udowodnij lub podaj kontrprzykład na stwierdzenie: dla dowolnych grafów G_1 i G_2 następujące warunki są równoważne:
   - istnieją homomorfizmy $h_1 : G_1 \to G_2$ oraz $h_2 : G_2 \to G_1$ (mówimy, że takie grafy są *homomorficznie równoważne*)

- istnieje izomorfizm $f : G_1 \to G_2$.

Z14. (1 pkt) Pokaż, że każde dwa grafy będące cyklami o parzystej długości są homomorficznie równoważne.

Z15. (1 pkt.) Mówimy, że formuła rachunku zdań jest w postaci 3CNF, gdy jest w koniunkcyjnej postaci normalnej, a dodatkowo każda klazula zawiera najwyżej 3 zmienne. Przykład formuły 3CNF: $(p \lor \neg q \lor r) \land (s \lor \neg r)$.

Rozważmy problem spełnialności formuł rachunku zdań zwany 3SAT: dla danej formuły w postaci 3CNF, rozstrzygnij czy istnieje wartościowanie ją spełniające.

Pokaż, że problem 3SAT można wyrazić jako problem istnienia homomorfizmu pomiędzy bazami danych, tzn. pokaż, że jesli potrafisz rozwiązywać problem istnienia homomorfizmu to potrafisz także rozwiązywać problem 3SAT.

Twoja konstrukcja powinna działać w czasie wielomianowym.

Z16. (2 pkt.) W zadaniu o nullach z poprzedniej listy wymagaliśmy aby dla Q i D istniała taka baza z nullami Q_D (wynik Q na D), że $\mathrm{rep}(Q_D) = Q(\mathrm{rep}(D))$. Przy takim wymaganiu okazało się, że nie można nawet zdefiniować wyniku zapytania z pojedynczą selekcją.

Spróbujmy więc rozwiązać problem inaczej. Nie będzie nam już zależało aby wynik zapytania reprezentował wszystkie możliwe zbiory krotek będące odpowiedzią na zapytanie. Załóżmy też dla uproszczenia, że D jest bazą z nullami o schemacie z pojedynczą relacja $T(A, B, C)$ i traktujmy D jako zbiór krotek w T.

Zdefiniujmy zbiór *pewnych odpowiedzi* na zapytanie Q na D następująco

$$\mathrm{certain}(Q, D) = \bigcap \{Q(I) | I \in \mathrm{rep}(D)\}$$

Są to krotki, które występują w każdej możliwej odpowiedzi na zapytanie Q.

a) (0.5 pkt.) Podaj sposób obliczenia dla podanej bazy z nullami D zbioru

$$\bigcap \{I | I \in \mathrm{rep}(D)\},$$

tzn. zbioru krotek obecnych w każdej bazie w $\mathrm{rep}(D)$.

b) (0.5 pkt.) Spróbujmy wymagać istnienia takiego Q_D aby $\mathrm{rep}(Q_D) = \mathrm{certain}(Q, D)$. Pokaż, że to nie wystarczy aby nasza semantyka miała sens. *Wskazówka:* znajdź bazę z nullami D oraz dwa zapytania Q' i Q'', które złożone zwracają inną odpowiedź na D niż gdy najpierw wyliczysz wynik Q'_D pierwszego zapytania i potem na nim obliczysz Q''.

Wskazówka 2: niech Q' będzie projekcją, patrz też na punkt powyżej...

Wniosek: wynik zapytania Q_D trzeba zdefiniować tak aby reprezentował coś więcej niż $\mathrm{certain}(Q, D)$.

c) (1pkt.) Powiemy, że bazy D' i D'' są *nierozróżnialne w języku zapytań* $\mathcal{L}$ gdy dla każdego zapytania $Q \in \mathcal{L}$

$$\text{certain}(Q, D') = \text{certain}(Q, D'')$$

Niech $\mathcal{L}$ będzie algebrą relacji z następującymi operatorami π, σ, ρ.

Oto nasza finalna definicja: mówimy, że odpowiedzią na zapytanie $Q \in \mathcal{L}$ na D jest takie Q_D, że $\text{rep}(Q_D)$ i $Q(\text{rep}(D))$ są nierozróżnialne w $\mathcal{L}$. Zauważ, że pozwala to mieć w Q_D jakieś dodatkowe krotki z nullami, byle zbiory pewnych odpowiedzi na zapytania w $\mathcal{L}$ się zgadzały.

Pokaż, że dla dowolnej bazy z nullami D istnieje odpowiedź na dowolne $Q \in \mathcal{L}$. Wskazówka: po prostu pokaż jak obliczać selekcję i projekcję na relacjach z nullami.