

## Lista zadań nr 5 (normalizacja)

Bazy Danych 2024

Pomóż zaprojektować bazę danych dla Wydziału Przystępczości Zorganizowanej (WPZ). Tajne służby opracowały już listę najbardziej zagrożonych miast oraz najbardziej zło-wieszczych procederów. Ustalono też ponad wszelką wątpliwość, że jeśli gang uprawia jakiś proceder to uprawia go we wszystkich miastach, w których jest obecny, a jeśli jest obecny w jakimś mieście to uprawia w nim wszystkie procedery, które uprawia gdziekolwiek.

W tej chwili baza składa się z relacji: **Mafia**(**Miasto**, **Gang**, **Proceder**, **Szef**, **Dochód**, **ROI**). Wiadomo też, że zachodzą zależności funkcyjne to:  $\text{Gang} \rightarrow \text{Szef}$ ,  $\text{Szef} \rightarrow \text{Dochód}$ ,  $\text{Proceder} \rightarrow \text{ROI}$ .

1. (0.5 pkt) Podaj definicję postaci BCNF. Czy powyższa baza jest w tej postaci? Dlaczego? Zaproponuj taki rozkład bazy do postaci BCNF, który jest odwracalny i zachowuje wszystkie zależności. Nie rozkładaj jakiegokolwiek tabeli, która już jest w postaci BCNF.
2. (0.5 pkt.) Czy mimo tego, że baza (po Twoich modyfikacjach) jest w postaci BCNF dostrzegasz w niej jeszcze jakąś oczywistą redundancję? Jaka jest jej przyczyna?
3. (0.5 pkt.) Czy istnieje odwracalny rozkład pozwalający się tej redundancji po-zbyć? Podaj go. Dlaczego jest on odwracalny?
4. (0.5 pkt) Napisz zapytanie algebry relacji zwracające niepustą odpowiedź wtedy i tylko wtedy gdy w relacji **Mafia** naruszona jest zależność funkcyjna  $\text{Proceder} \rightarrow \text{ROI}$ .
5. (1 pkt) Rozważamy relację  $S(F, M, R)$ . Zależność wielowartościowa  $F \twoheadrightarrow M$  za-chodzi wtw, gdy dla każdych dwóch krotek  $t_1, t_2 \in S$  takich, że  $\pi_F(t_1) = \pi_F(t_2)$  istnieje krotka  $t \in S$  taka, że:

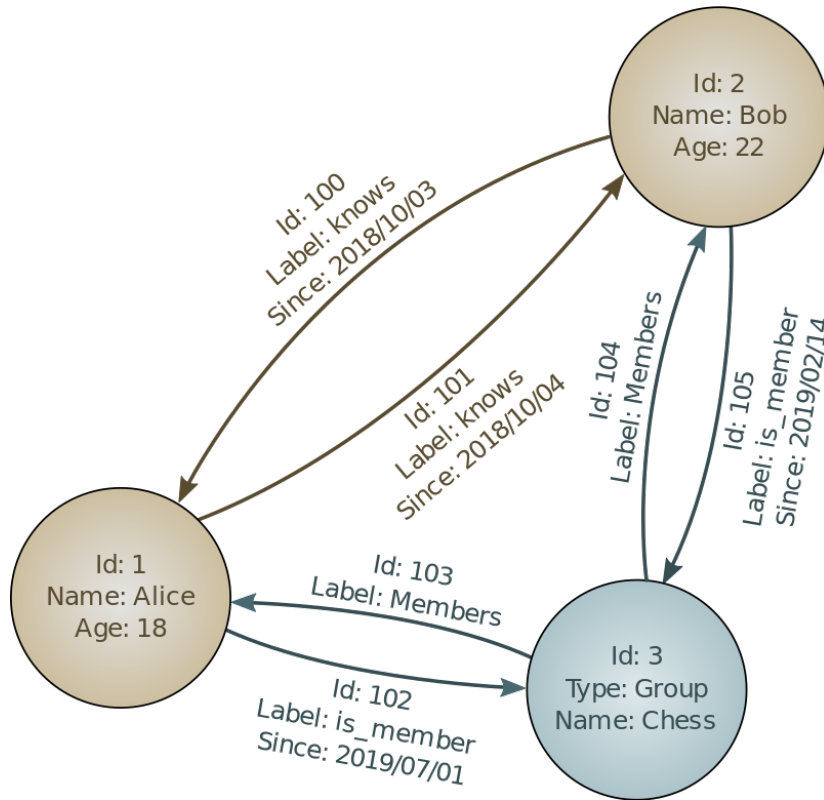
1.  $\pi_{FM}(t) = \pi_{FM}(t_1)$ ,
2.  $\pi_{FR}(t) = \pi_{FR}(t_2)$ .

Udowodnij lub pokaż kontrprzykład:

1. Jeśli  $F \twoheadrightarrow M$  to  $F \rightarrow M$ .
2. Jeśli  $F \rightarrow M$  to  $F \twoheadrightarrow M$ .

Jeśli w bazie danych wszystkie zależności wielowartościowe zależą od (nad)klucza to mówimy, że baza jest w 4. postaci normalnej (4NF).

6. (1 pkt) Jednym z najważniejszych obecnie modeli danych dla grafów są tzw. *prop-erty graphs* (przykład na rys. 1). Grafy w tym modelu składają się z wierzchoł-ków i krawędzi. Wierzchołki i krawędzie mogą posiadać wiele *własności*, które



Rysunek 1: Property graph

są parami klucz-wartość. Zaproponuj sensowny sposób przechowywania takich grafów w relacyjnej bazie danych.

7. (1 pkt) Tradycyjnie bardzo rzadko stosowało się postacie normalne powyżej BCNF (np. 4NF). Cała teoria normalizacji powyżej BCNF była uważana za odległą od praktyki.

W dużej mierze wynikało to z faktu, że sposoby wyliczania złączeń w bazach danych zakładały, że operator złączenia jest binarny (łączymy dwie tabele, wynik złączenia z kolejną itd.). W nowoczesnych rozwiązaniach czasami opłaca się posunąć normalizację aż do skrajności - każda tabela posiada co najwyżej jedną kolumnę, która nie jest częścią klucza (tzw. 6. postać normalna, 6NF). Jedną z zalet takiego podejścia jest chęć wykonywania złączeń wielu tabel jednocześnie (*worst-case optimal multi-way joins*).

Zmodyfikuj rozwiązanie poprzedniego zagadnienia tak aby property graphs przechowywać w 6NF.

8. (1 pkt) Rozważmy relację  $R$ , w której zachodzi zależność funkcyjna  $\xi = \alpha \rightarrow \beta$ , gdzie  $\alpha \cap \beta = \emptyset$  oraz  $\alpha$  i  $\beta$  są podzbiorami atrybutów  $R$ . Udowodnij, że rozkład relacji  $R$  wg  $\xi$ , na relacje  $R_1$  z atrybutami  $\alpha \cup \beta$  oraz  $R_2$  z atrybutami  $\text{attr}(R) \setminus \beta$  jest odwracalny.