

Lista zadań nr 5 (normalizacja)

Pomóż zaprojektować bazę danych dla Wydziału Przestępczości 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: Gang \rightarrow Szef, Szef \rightarrow Dochód, Proceder \rightarrow ROI.

1. (1 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. (1 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. (1 pkt.) Czy istnieje odwracalny rozkład pozwalający się tej redundancji pozbyć? Podaj go. Dlaczego jest on odwracalny?
4. (1 pkt) Napisz zapytanie algebry relacji zwracające niepustą odpowiedź wtedy i tylko wtedy gdy w relacji Mafia naruszona jest zależność funkcyjna Proceder \rightarrow ROI.
5. (1 pkt) Korzystając z API Twittera wyeksportowano dane w postaci pliku csv o następującym nagłówku:

```
author_id, author_name, author_location, author_description, created_at, text,
tweet_id, hashtags, mentions, retweeted_tweet_id, in_reply_to_tweet_id
```

Przykładowa linia ilustrująca znaczenie nagłówka:

```
1, leroylovesusa, USA, "God. Country. Family. #InGodWeTrust", 2016-11-09,
"RT @w0tn0t: #ElectionDay #NeverHillary @TwitterMoments Voting Issues:
Some Trump Voters Reporting Ballots Switching To Clinton. https://t.col5",
2, ["ElectionDay", "NeverHillary"], ["w0tn0t", "twittermoments"], 3, 4
```

Powyższy tweet jest retweetem tweeta o tweet_id 3 i odpowiada na tweeta o tweet_id 4.

Zaprojektuj znormalizowany schemat bazy danych do przechowywania tweetów. Uzasadnij swoje założenia oraz swoje decyzje projektowe. Możesz przedstawić alternatywne wersje jeśli uważasz to za stosowne.

- 6. (2 pkt)** Rozważamy relację $S(F, M, R)$. Zależność wielowartościowa $F \twoheadrightarrow M$ zachodzi 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 $F \twoheadrightarrow M$.