

Lista zadań nr 4

Bazy Danych 2018

1. (2 pkt.) Dana jest relacja o schemacie $R = MBKIID$ ze zbiorem zależności funkcyjnych $F = \{K \rightarrow D, A \rightarrow M, AK \rightarrow I, M \rightarrow B\}$.
 1. Znajdź wszystkie klucze relacji i uzasadnij, że są to wszystkie klucze;
 2. Czy relacja jest w BCNF?
 3. Rozważmy rozkład R na składowe $R_1 = AKD$, $R_2 = AKM$, $R_3 = AKI$ oraz $R_4 = AKB$. Wyznacz zbiory zależności funkcyjnych dla każdej składowej. Czy składowe są w BCNF?
2. (2 pkt.) Niech R będzie relacją, a F jej zbiorem zależności funkcyjnych. Rozważmy rozkład $R = R_1 \cup R_2$ i niech F_1 i F_2 będą rzutami F odpowiednio na R_1 i R_2 .
 - Pokaż, że dla każdego stanu r relacji R zachodzi zawieranie $r \subseteq \pi_{R_1}(r) \bowtie \pi_{R_2}(r)$ i nie zawsze zachodzi równość $r = \pi_{R_1}(r) \bowtie \pi_{R_2}(r)$
 - Pokaż, że zachodzi zawieranie $(F_1 \cup F_2)^+ \subseteq F^+$ i nie zawsze zachodzi równość $(F_1 \cup F_2)^+ = F^+$.
3. (2 pkt.) Rozważ relację $R = MBKIID$ z zależnościami $F = \{K \rightarrow D, A \rightarrow MB, AK \rightarrow I, M \rightarrow B\}$. Rozłóż tę relację na składowe w BCNF tak, by rozkład był odwracalny. Czy przedstawiony przez Ciebie rozkład zachowuje zależności?
4. (2 pkt.) Rozważmy relację $R = KNOUSGT$ z zależnościami $F = \{K \rightarrow N, KG \rightarrow SK, GS \rightarrow K, GN \rightarrow S, SGK \rightarrow N, KNU \rightarrow O, GU \rightarrow S\}$. Wyznacz pokrycie minimalne F , a następnie znajdź odwracalny i zachowujący zależności rozkład R do 3NF.
5. (2pkt) Korzystając z API Twittera zaimportowano 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 nagłówków:

```
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.