

Slovenská technická univerzita v Bratislave

Fakulta informatiky a informačných technológií

MongoDB

Martin Schön , Bc.

AIS ID: 103121

E-mail: xschon@stuba.sk

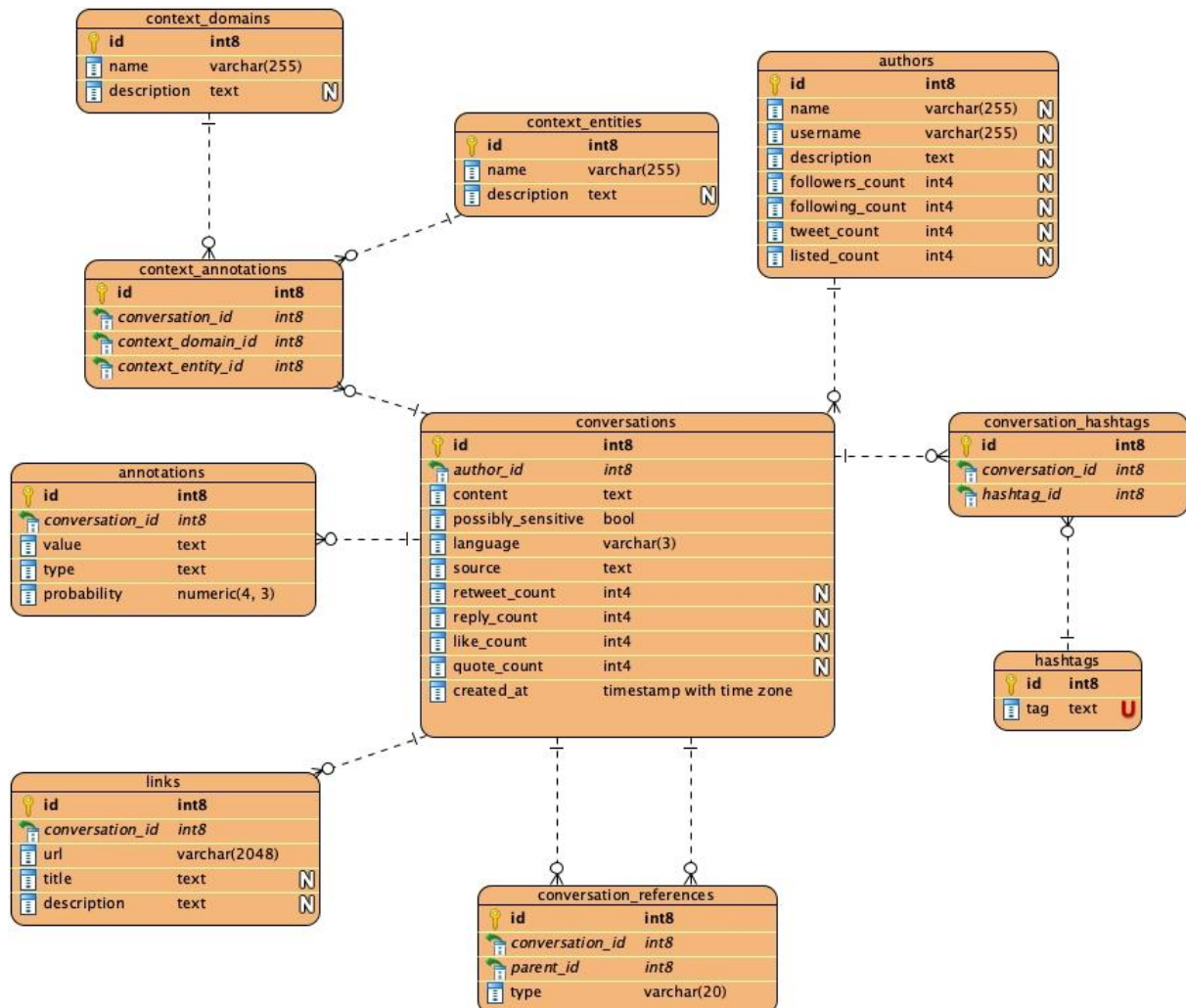
GitHub repozitár: <https://github.com/FIIT-DBS/zadanie-pdt-xSchon>

Predmet: Pokročilé databázové technológie

Zimný semester 2022/2023

1. Návrh dátového modelu

V tomto zadaní využívam denormalizované dáta na základe návrhu využívaného pri predchádzajúcich zadaniach v jazyku PostgreSQL.



Tieto dáta sú vhodné na zobrazovanie tweetov používateľov - je možné podľa autora dohľadať jednotlivé tweety, ale aj pre zobrazenie retweets - pomocou **conversation_references** tabuľky na nájdenie retweeted vzťahu.

Tieto dáta následne denormalizujem do json súboru, pomocou SQL query priloženej v súbore **denormalization_query.sql**. Tá rozbije jednotlivé tabuľky databázy a nad hlavnou tabuľkou - tabuľkou tweets (na obrázku **conversations**) vytvorí dopyt. Ten umožní načítať dáta z ostatných tabuliek ako jsony (prípadne ich polia) uložené ako embednuté v tabuľke tweets - teda dáta sú rovno dostupné v dokumente tweetu.

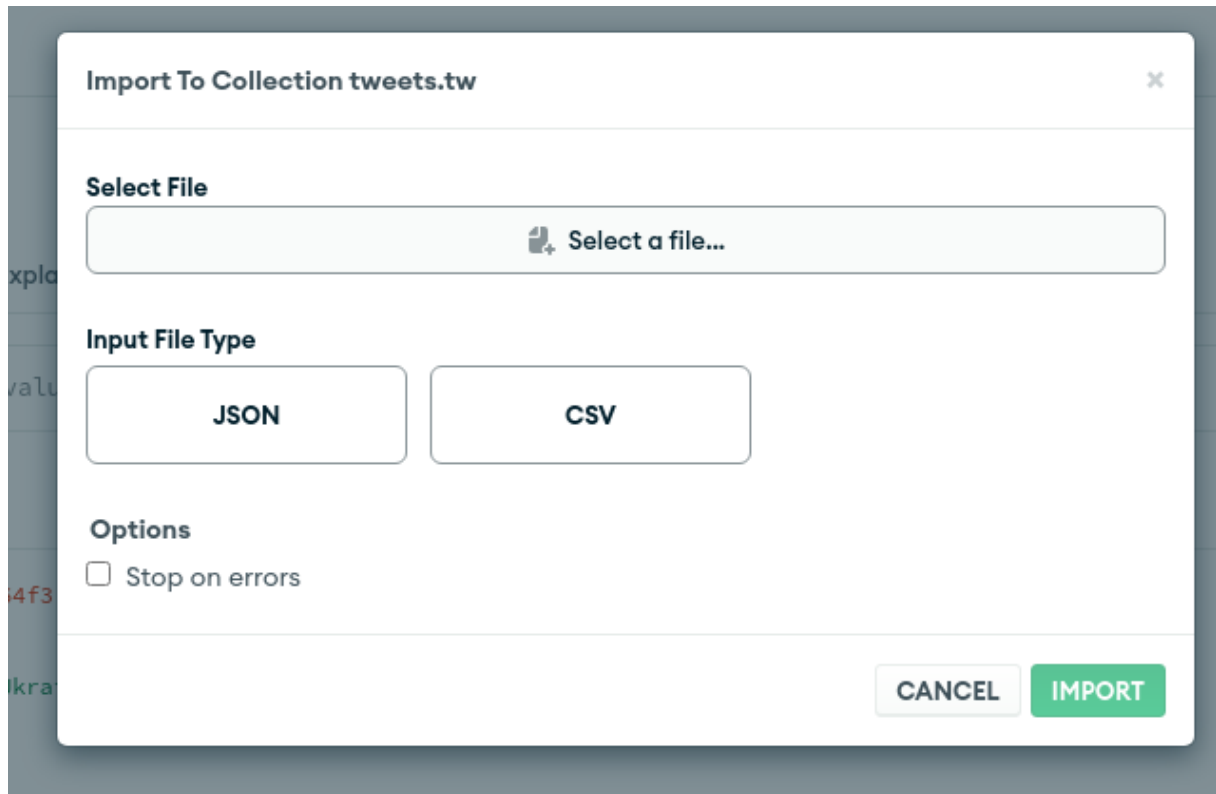
To je dobré, pretože vďaka tomu je možné pristúpiť k ostatným dátam (napríklad autorovi tweetu) veľmi rýchlo a jednoducho - v podstate je súčasťou informácie o konverzácií. Tabuľka tweets je už vopred pripravená a obsahuje dáta len z 24.02.2022, preto nie je potrebné riešiť v query túto podmienku.

Referencia potom v sebe drží aj všetky údaje o svojom tweete, takže vnárame všetky dáta. Príklad toho, ako vyzerá takýto tweet je priložený v prílohe **example.json**.

2. Import do MongoDB

MongoDB som lokálne rozbehal pomocou dockera - konkrétne som použil docker mongo image, ktorý mi umožňuje MongoDB spustiť a používať lokálne.

Dáta som následne získal z Postgres databázy za využitia COPY ... TO file príkazu a query spomenutej v bode 1. Takto vygenerované dáta som uložil v json formáte a následne som pomocou aplikácie [MongoDB compass](#) vložil JSON dáta do mojej databázy. JSON bolo potrebné očistiť a zmeniť \\ na \, spoločne s tým odstrániť symbol \n. To som dosiahol za pomoci sed Linuxového príkazu a dáta som jednoducho naimportoval do databázy.



3. Dotazy nad databázou

Dotazy nad databázou som takisto vykonával za pomoci rozhrania MongoDB compass. Tá mi umožnila filtrovať, sortovať a limitovať dáta pomocou GUI.

a) Vypíše posledných 10 tweetov pre autora, ktorý má username Newnews_eu

V tejto časti som vyfiltroval všetky dokumenty od užívateľa newnews_eu, zoradil ich podľa dátumu a vybral prvých 10. Vďaka embeddovaniu som jednoducho pristúpil k užívateľskému menu autora. Výsledok sa nachádza v súbore z06_03a.json



The screenshot shows the MongoDB Compass query builder interface. The Filter field contains the query `{ 'authors.username': 'Newnews_eu' }`. The Project field is set to `{ field: 0 }`. The Sort field is set to `{ created_at: -1 }`. The Collation field is set to `{ locale: 'simple' }`. Below the query fields, the MaxTimeMS is set to 60000 and the Limit is set to 10.

b) Vypíše posledných 10 retweetov pre tweet, ktorý má id 1496830803736731649.



The screenshot shows the MongoDB Compass query builder interface. The Filter field contains the query `{ $and: [{ 'conversation_references.json_agg.type': 'retweeted' }, { 'conversation_references.json_agg.conversations.id': '1496830803736731649' }] }`. The Project field is set to `{ field: 0 }`. The Sort field is set to `{ created_at: -1 }`. The Collation field is set to `{ locale: 'simple' }`. On the right side, there is a Skip button with a value of 0.

Tento dopyt fungoval veľmi podobne ako v bode a), avšak s tým rozdielom, že som musel použiť \$and operátor, aby som vybral iba retweety a potom tie, ktoré retweetujú určenú konverzáciu podľa id. Tie som zoradil podľa created_at (-1 pre zostupné) a limitoval 10. Výsledok sa nachádza v súbore z06_03b.json

Záver

Zadanie bolo vypracované samostatne, na OS Ubuntu 22.04 LTS. Pomocou príkazu `docker pull mongo:latest` som si nainštaloval najnovšiu verziu mongo dockeru aktuálnu ku dňu 24.12.2022. Priečinok obsahuje viacero príloh spomínaných v jednotlivých bodoch zadania. Všetky tieto súbory, rovnako ako aj dokumentácia sú dostupné na GitHubu projektu spomenutom na úvodnej strane. Prajem pekné prežitie Vianočných sviatkov.