Pokročilé databázové technológie

**Zadanie č. 6 – Neo4j**

**Do kedy?**

Keďže všetci robíte radi s predstihom, deadline je 31.12.2021 o 23:59.

**Čo odovzdať?**

Dotazy v separátnom textovom editore + dotaz a screenshot výsledku v pdf.

**Hodnotenie**

Úlohy spolu za 10 bodov rozdelené nasledovne:

* 1,2,3,4,5 a 6 sú za jeden bod,
* 7, 8 za 2 body.
* Bonus za 2 extra body.

**Zadanie – Neo4j**

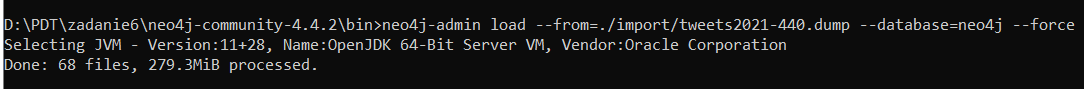
1. Zoberte si nami vytvorený dataset z vašich tweetov:

https://drive.google.com/file/d/1kufTAFP5DH8hiCT-BRLHqAicoulJ\_0Qi/view?usp=sharing

a importujte ho cez neo4j-admin

Riešenie:

neo4j-admin load --from=./import/tweets2021-440.dump --database=neo4j –force



Obrázok, na ktorom je text, monitor, snímka obrazovky, obrazovka

Automaticky generovaný popis

2. Vypíšte 5 Accountov s najvyšším množstvom followerov. Nezoraďujte Accounty podľa poľa

followers\_count, hodnota je prevzatá z Twitteru a nezodpovedá našim vzťahom v datasete.

Zaujímajú nás followujúce Accounty v našom datasete cez vzťah FOLLOWS.

Riešenie:

Obrázok, na ktorom je text, snímka obrazovky, monitor

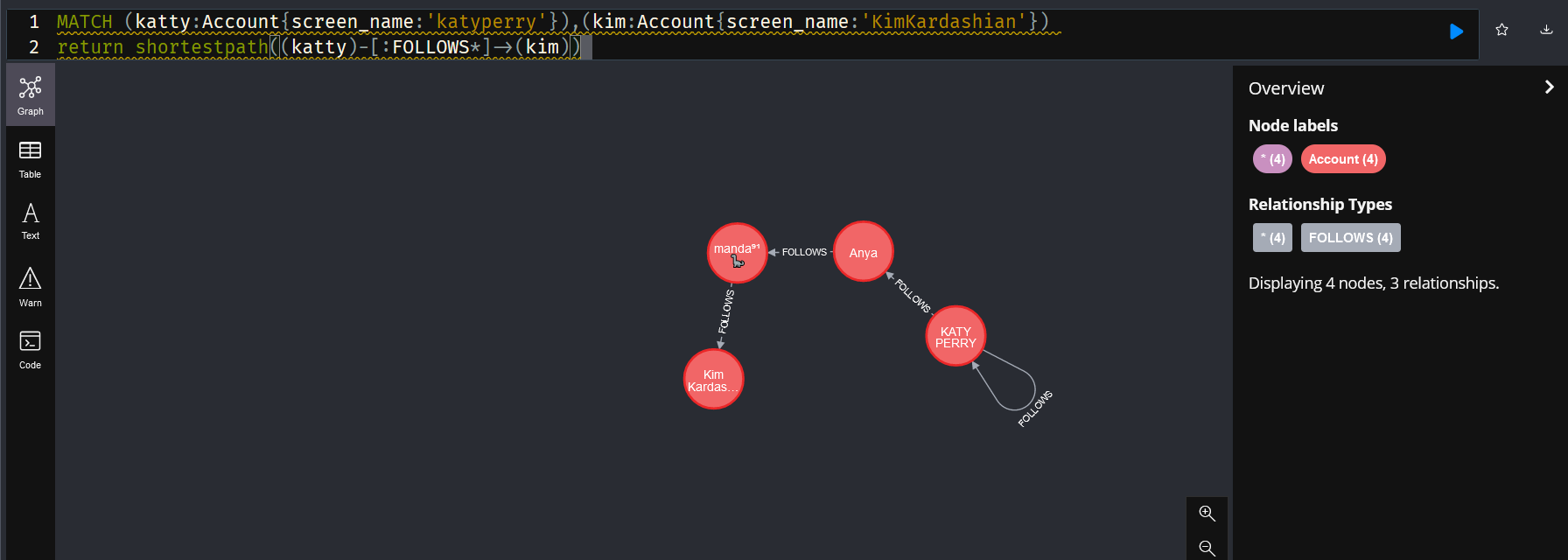
Automaticky generovaný popis

3. Nájdite najkratšie cesty medzi Katy Perry {screen\_name: ‘katyperry‘} a Kim Kardashian

{screen\_name: ‘KimKardashian‘} cez vzťah FOLLOWS. Všetky cesty, kde Kim Kardashian

followuje niekoho, kto followuje niekoho, kto..., kto followuje Katy Perry.

Riešenie:



4. Vyhľadajte neúspešné tweety influencerov. Vyhľadajte 10 najmenej retweetovanych tweetov

od Accountov, ktoré sú na prvých 10 miestach v celkovom počte retweetov.

Riešenie:

Obrázok, na ktorom je text, snímka obrazovky, monitor, čierne

Automaticky generovaný popis

Aj s vizualizáciou vzťahov:

Obrázok, na ktorom je text, snímka obrazovky

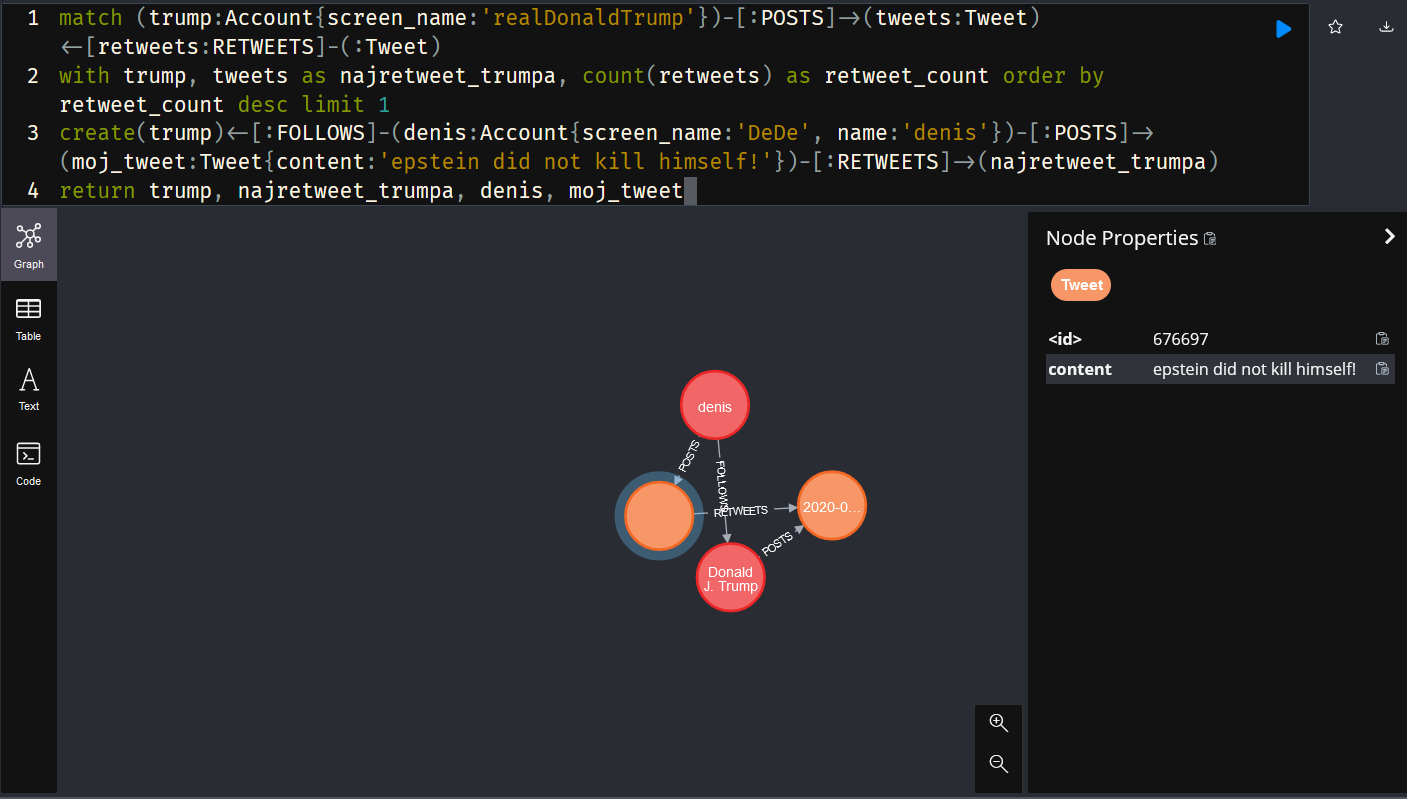
Automaticky generovaný popis

5. Vytvorte volaním iba jednej query nový Account s Vašim menom, ktorý bude followovať

Donalda Trumpa {screen\_name:"realDonaldTrump"} a v tom istom volaní vytvorte tweet,

ktorý bude retweetom Donaldovho najretweetovanejšieho tweetu.

Riešenie:



6. Odporučte používateľovi {screen\_name:"777stl"} followovanie ďalších Accountov, na základe

followovania rovnakých Accountov: Vyhľadajte 10 Accountov, ktoré followujú najviac

rovnakých Accountov ako náš používateľ, ale náš používateľ ich ešte nefollowuje.

Riešenie:

Obrázok, na ktorom je text, monitor, snímka obrazovky, obrazovka

Automaticky generovaný popis

Aj s vizualizáciou vzťahov:

Obrázok, na ktorom je text, snímka obrazovky

Automaticky generovaný popis

7. Odporučte používateľovi {screen\_name:"DaynerWilson"} followovanie ďalších Accountov na

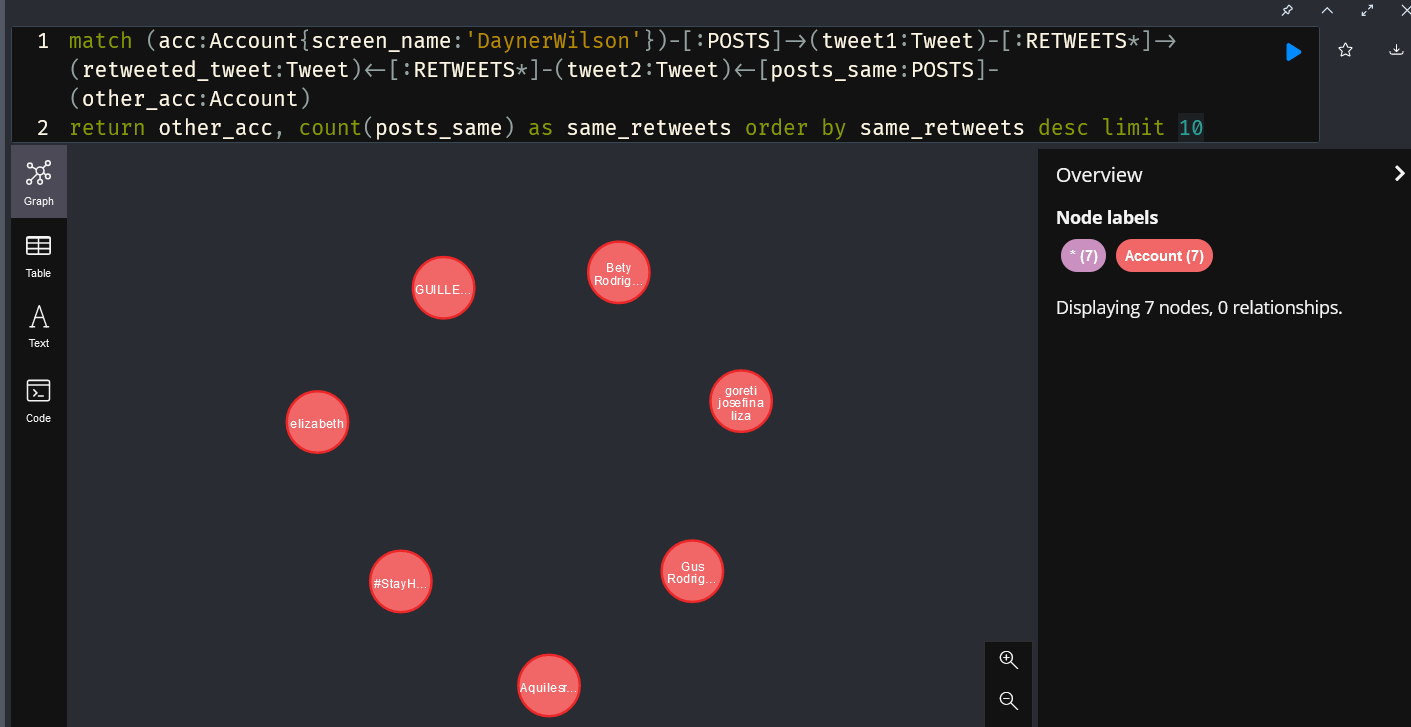
základe zhody v retweetovaní rovnakých tweetov: Vyhľadajte 10 accountov, ktoré retweetli

najviac tých istých tweetov, ako náš používateľ. Ak tweet ktorý retweetujeme, je už tiež

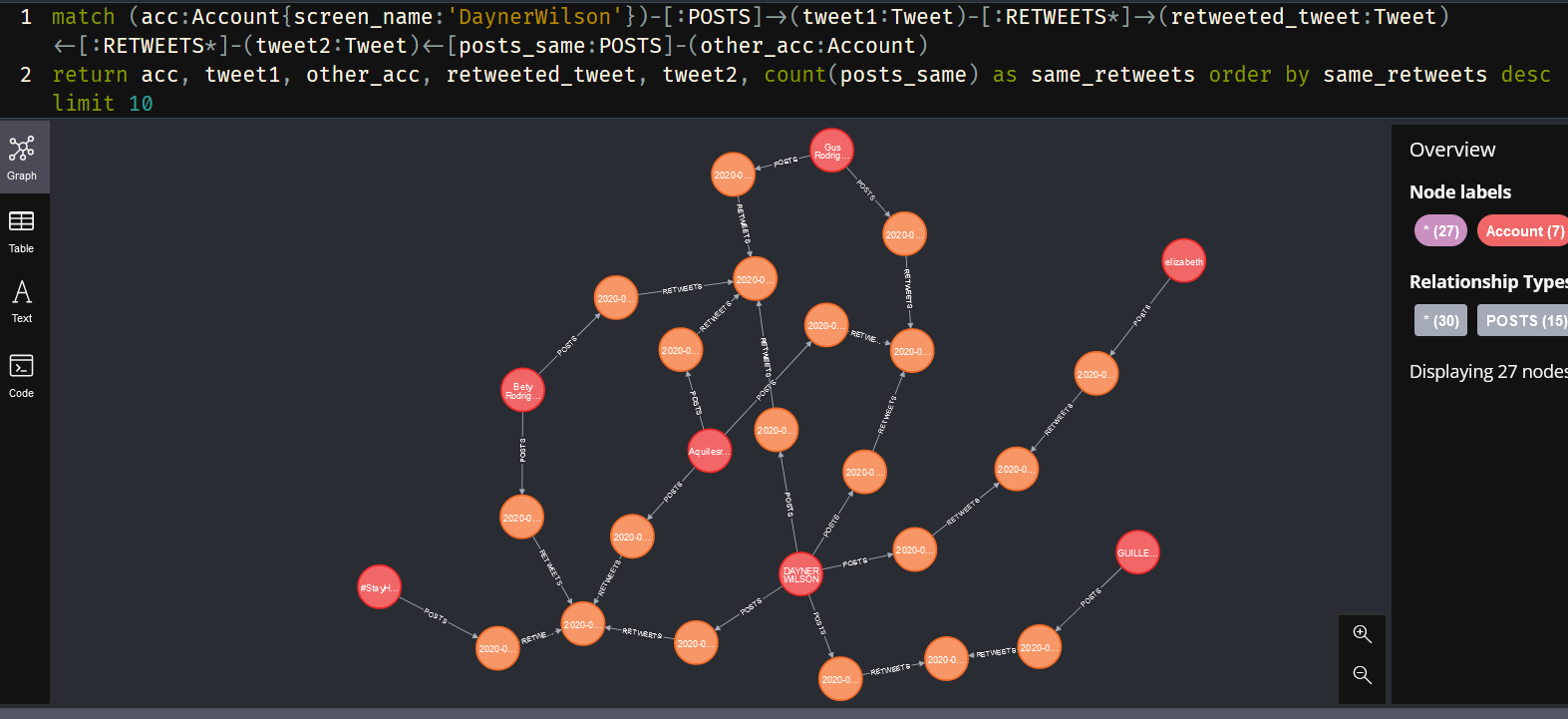
retweetom, rátajte za zhodu aj retweetovanie jeho parent tweetu – retweetovanie teda

zohľadňujte rekurzívne.

Riešenie:



Aj s vizualizáciou vzťahov:



8. Vyhľadajte 5 tweetov ostatných Accountov, ktoré do hĺbky 5 followujú account, ktorý napísal

tweet {id: "1289380305728503808"}, ktoré síce nie sú retweetom vybraného tweetu, ale

napriek tomu majú čo najviac rovnakých slov v poli content zhodných s vybraným tweetom

(stačí rozdeliť content na slová cez split(tweet.content, " "). Account, ktorý followuje

Account, ktorý followuje nami vybraný Account rozumieme hĺbkou 2. Odporúčam pozrieť si

procedúry v knižnici APOC pracujúce s collections, ale nie je to podmienkou na zvládnutie

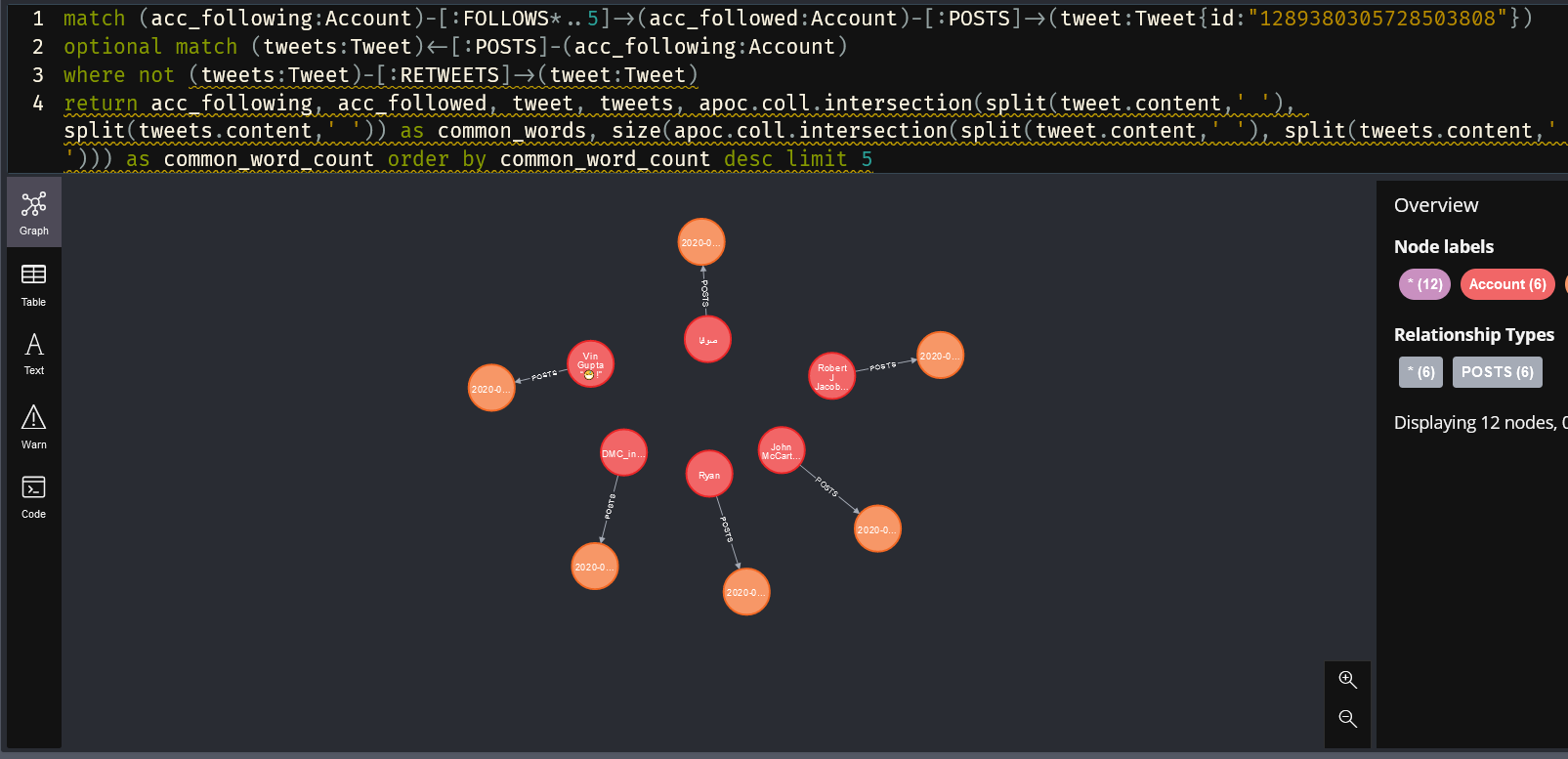
úlohy.

Riešenie:

Obrázok, na ktorom je text

Automaticky generovaný popis

Aj s vizualizáciou vzťahov:



BONUS: Nájdite najkratšie cesty medzi Katy Perry (katyperry) a Donaldom Trumpom cez vzťah

RETWEETS (a tým pádom aj POST). Všetky cesty, kde Katy Perry retweetla post Accountu, ktorý

retweetol post Accountu, ktorý..., ktorý retweetol post Donalda Trumpa.

**(Account) -> [POSTS] -> (Tweet) – [RETWEETS] -> (Tweet) <- [POSTS] – (Account)** – [POSTS] ->

(Tweet) – [RETWEETS] -> (Tweet) <- [POSTS] - (Account) -> …

Riešenie:

