Pokročilé databázové technológie

Zadanie 4 – Neo4j

Marek Adamovič

Cvičenie: Piatok 11:00 – 12:40

Ing. Ján Balažia, PhD. 2022/2023

Obsah

[1. otázka 3](#_Toc119676776)

[2. otázka 4](#_Toc119676777)

[3. otázka 5](#_Toc119676778)

[4. otázka 6](#_Toc119676779)

[5. otázka 7](#_Toc119676780)

[6. otázka 8](#_Toc119676781)

# otázka

**Otázka:**

Vytvorte volaním iba jednej query nového Autora s Vašim menom (vyplňte minimálne name a username) a tým istým volaním vytvorte tweet ktorý bude obsahovať aspoň jeden vami vybraný Hashtag (vzťah :HAS) a ktorý bude retweetovať najretweetovanejší tweet Vladimíra Zelenského{username:"ZelenskyyUa"}.

**Odpoveď:**

Najskôr sme si našli tweet s najväčším počtom vzťahov REPLIED\_TO a uložili ho do premennej c1. Následne sme vytvorili autora, tweet a jednotlivé vzťahy, ktoré pospájali všetky tieto prvky. Na výslednom grafe sme si overili správnosť riešenia.

**Obrázok, na ktorom je text, monitor, snímka obrazovky, vnútri

Automaticky generovaný popis**

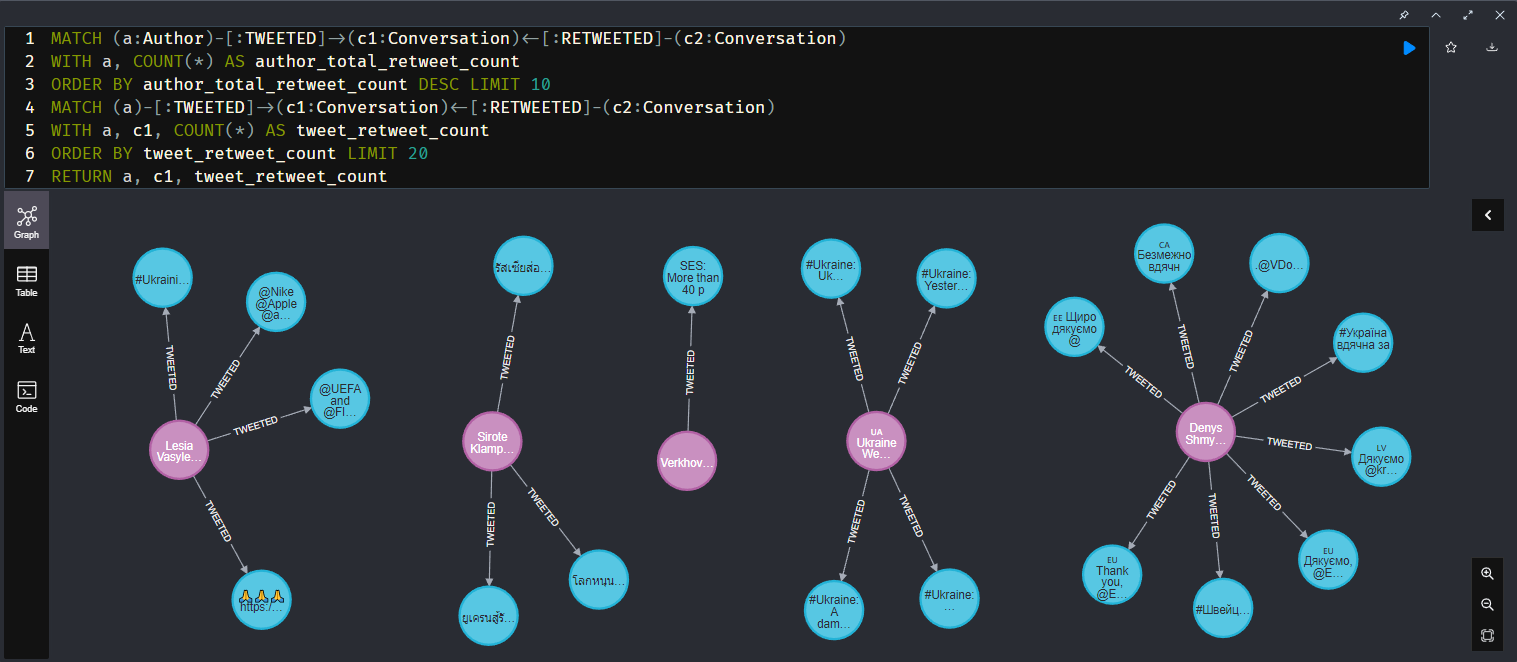
# otázka

**Otázka:**

Vyhľadajte zlyhania influencerov. Vyhľadajte 20 najmenej retweetovanych tweetov od Accountov, ktoré sú na prvých 10 miestach v celkovom počte retweetov. Aj keď taký príklad v datasete nie je, počítajte aj s prípadom, ak by niektorý tweet bol retweetnutý 0 krát.

**Odpoveď:**

Najskôr sme si vyhľadali všetky retweetnuté tweety, ktoré sme groupli (pomocou WITH) podľa autorov a spočítali zvyšok uzlov ako celkový počet retweetov. Podľa tohto počtu sme zobrali len 10 najúspešnejších autorov. V ďalšom kroku sme našli tweety týchto úspešných autorov, zrátali sme počet ich retweetov pomocou WITH + COUNT, zoradili ich podľa počtu retweetov a vo výsledku sme nechali len tie najmenej úspešné. Na druhom obrázku môžeme vidieť najmenej úspešný z týchto tweetov.



Obrázok, na ktorom je text, snímka obrazovky, monitor

Automaticky generovaný popis

# otázka

**Otázka:**

Odporučte používateľovi (username: Marios59885699) followovanie ďalších autorov na základe zhody v retweetovaní rovnakých tweetov: Vyhľadajte 10 autorov, ktorí retweetli najviac tweetov rovnakych, ako používateľ Marios59885699. Počítajú sa aj retweety tweetov, ktoré retweetujú rovnaký tweet.

**Odpoveď:**

V prvom kroku sme si vyhľadali (re)tweety od usera Marios59885699. Následne vyberáme všetkých ostatných autorov, ktorí majú rovnaké retweety. Tieto rovnaké retweety spočítame a vrátime 10 autorov, ktorí ich majú najväčší počet. Ako kontrolu sme vykonali túto query aj bez filtru (WHERE) na riadku 2, čo nám na prvom mieste vrátilo samotného usera Marios59885699, keďže ten má, pochopiteľne, najviac zhodných retweetov so sebou samým.

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

Automaticky generovaný popis

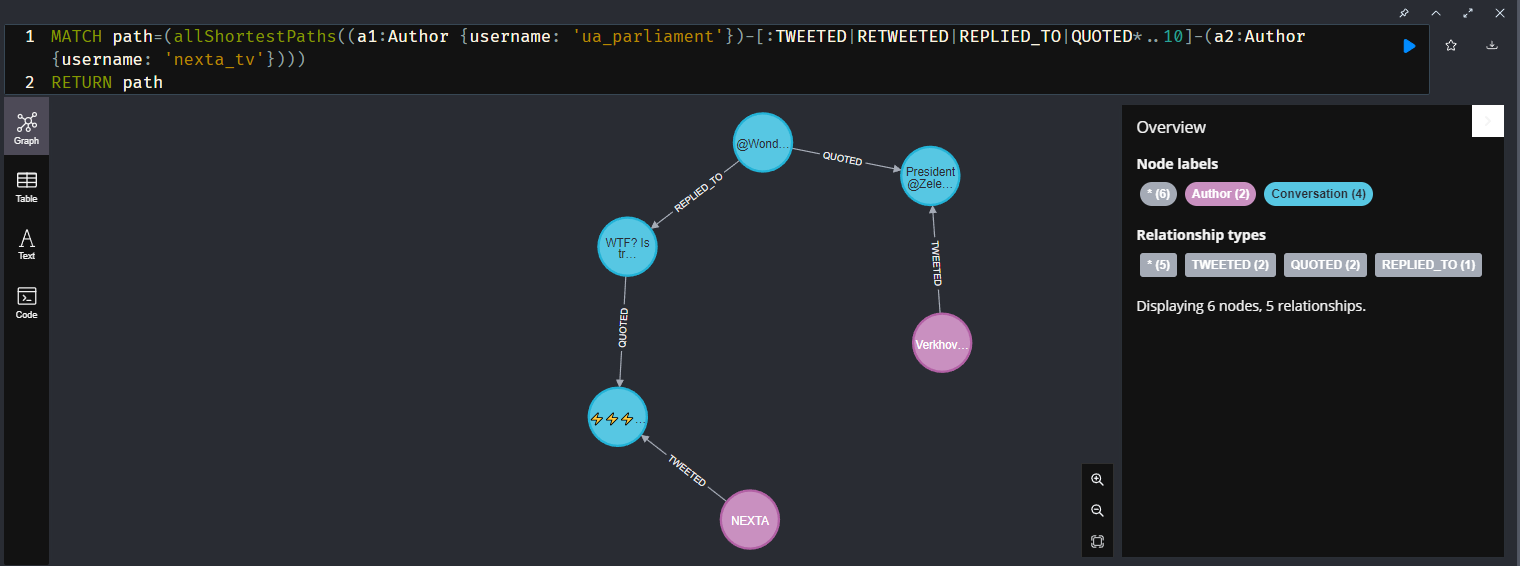
# otázka

**Otázka:**

Nájdite najkratšie cesty medzi Ukrajinským parlamentom (username: “ua\_parliament”) a NextaTV (username: “nexta\_tv”) cez vzťahy TWEETED, RETWEETED, REPLIED\_TO a QUOTED. Hľadajte do hĺbky maximálne 10. (hint: allShortestPaths)

**Odpoveď:**

Existuje len jedna cesta dĺžky 5, ktorá je najkratším existujúcim spojením medzi „ua\_parliament“ a „nexta\_tv“.



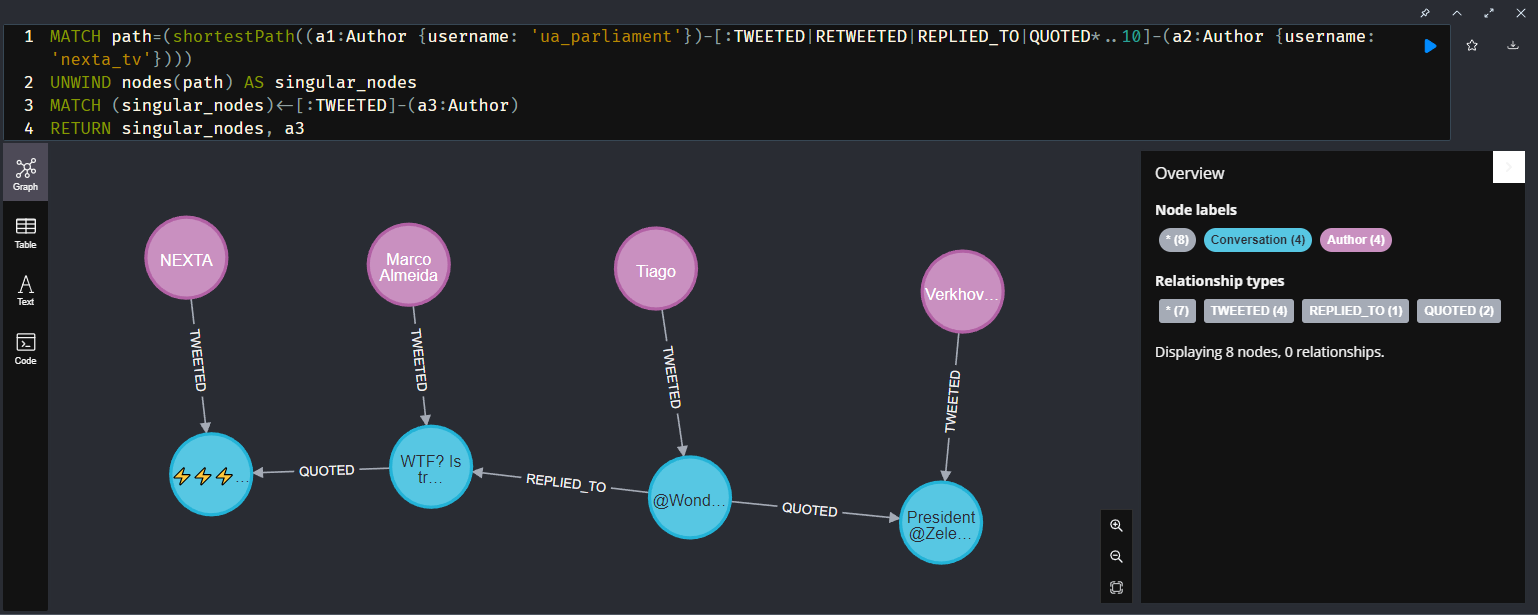
# otázka

**Otázka:**

Upravte query z predchádzajúcej úlohy, aby vrátila iba nájdenú najkratšiu cestu a jednotlivé uzly typu Conversation v nej aj spolu s autorom, ktorý ich tweetol. (hint: UNWIND)

**Odpoveď:**

Zmenili sme funkciu z allShortestPaths na shortestPath (aby sme v každom prípade dostali len jednu najkratšiu cestu). Následne sme výsledok z tejto query, ktorý bol typu path, pretypovali na pole s uzlami a pomocou UNWIND operátoru sme z nich spravili samostatné uzly, ku ktorým sme už ľahko našli autorov, ktorí ich tweetli.

****

# otázka

**Otázka:**

Vypíšte 10 najpoužívanejších hashtagov v datasete aj s celkovým počtom použití a autorom, ktorý daný hashtag najviac krát použil. (hint: collect)

**Odpoveď:**

Najskôr sme si našli 10 najviac používaných hashtagov. Následne sme našli spojenie týchto hashtagov s autormi (cez konverzácie, teda tweety). Spočítali sme koľko použití mali jednotlivé hashtagy a následne pomocou funkcie COLLECT sme sa dostali k informácií, ktorí autori ich používali. Funkcia MAX nám zabezpečuje, že vyberieme autora, ktorý má týchto hashtagov najviac.

