

Pokročilé databázové technológie

Zadanie 4 – Neo4j

Marek Adamovič

Obsah

1. otázka.....	3
2. otázka.....	4
3. otázka.....	5
4. otázka.....	6
5. otázka.....	7
6. otázka.....	8

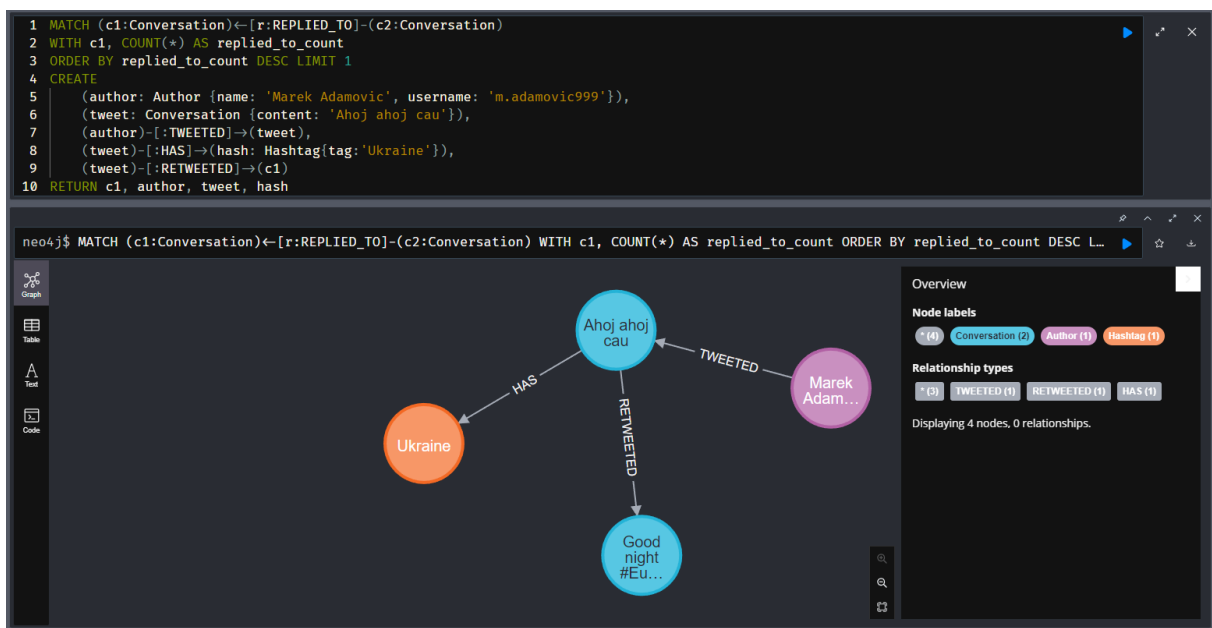
1. otázka

Otázka:

Vytvorte volaním iba jednej query nového Autora s Vaším 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.



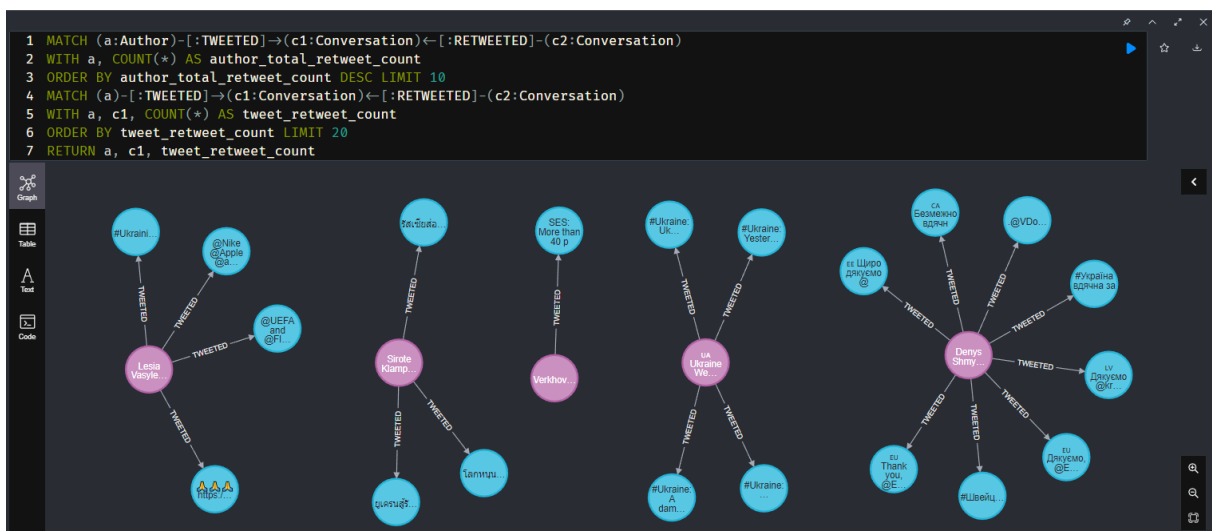
2. otázka

Otázka:

Vyhľadajte zlyhania influencerov. Vyhľadajte 20 najmenej retweetovaných 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, zráтали 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.



The image shows the results of a Cypher query in a Neo4j interface. The query is:

```
neo4j$ MATCH (a:Author)-[:TWEETED]->(c1:Conversation)-<[:RETWEETED]->(c2:Conversation) WITH a, COUNT(*) AS author_total_retweet_count ...
```

The results are displayed in a table with two columns: 'a' (author) and 'c1' (conversation). The 'tweet_retweet_count' column is also visible on the right.

a	c1	tweet_retweet_count
<pre>{ "identity": 2150224, "labels": ["Author"], "properties": { "tweet_count": 2056, "following_count": 310, "listed_count": 2879, "followers_count": 306162, "name": "Lesia Vasylenko", "description": "Ukrainian MP, @goloszmin, working mom of 3 lovely humans, lover of freedom, travel and all things green #Ukraine #geopolitics", "id": 1219232377605644289, "username": "lesiavasylenko" } }</pre>	<pre>{ "identity": 8357005, "labels": ["Conversation"], "properties": { "like_count": 5, "created_at": "2022-03-01T00:24:21", "language": "en", "source": "Twitter for iPhone", "id": 1498438954592387073, "reply_count": 0, "author_id": 1219232377605644289, "quote_count": 0, "content": "@UEFA and @FIFAWorldCup ban #Russia from ALL football matches and break their sponsorship deal with @GazpromFootball 🇺🇦 in January" } }</pre>	1

Started streaming 20 records in less than 1 ms and completed after 6942 ms.

3. otázka

Otázka:

Odporučte používateľovi (username: Marios59885699) followovanie ďalších autorov na základe zhody v retweetovaní rovnakých tweetov: Vyhľadajte 10 autorov, ktorí retweetli najviac tweetov rovnakých, 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.

```
1 MATCH (a1:Author {username: 'Marios59885699'})-[:TWEETED]-(c1:Conversation)-[:RETWEETED]-(c2:Conversation)
2 MATCH (a2:Author) -[:TWEETED]-(c3:Conversation)-[:RETWEETED]-(c2) WHERE a2.username <> 'Marios59885699'
3 WITH a2, COUNT(*) AS same_retweets
4 ORDER BY same_retweets DESC LIMIT 10
5 RETURN a2, same_retweets
```



The screenshot shows the Neo4j Cypher query interface. The query is as follows:

```
1 MATCH (a1:Author {username: 'Marios59885699'})-[:TWEETED]-(c1:Conversation)-[:RETWEETED]-(c2:Conversation)
2 MATCH (a2:Author) -[:TWEETED]-(c3:Conversation)-[:RETWEETED]-(c2) WHERE a2.username <> 'Marios59885699'
3 WITH a2, COUNT(*) AS same_retweets
4 ORDER BY same_retweets DESC LIMIT 10
5 RETURN a2, same_retweets
```

The results are displayed in a table view. The first result is for user 'Calibre Obscura' with 41 retweets. The table has two columns: 'a2' and 'same_retweets'.

a2	same_retweets
{ "identity": 4476676, "labels": ["Author"], "properties": { "tweet_count": 24900, "following_count": 675, "listed_count": 2075, "followers_count": 134951, "name": "Calibre Obscura", "description": "Arms Research & Occasional Effortposter. Interested in MENA/Asia NSAGs كالفير اوبسكورا Like/RT ≠ Approval. GB NATO SHILL", "id": 835552373615493120, "username": "CalibreObscura" } }	41

Started streaming 10 records in less than 1 ms and completed after 1878 ms.

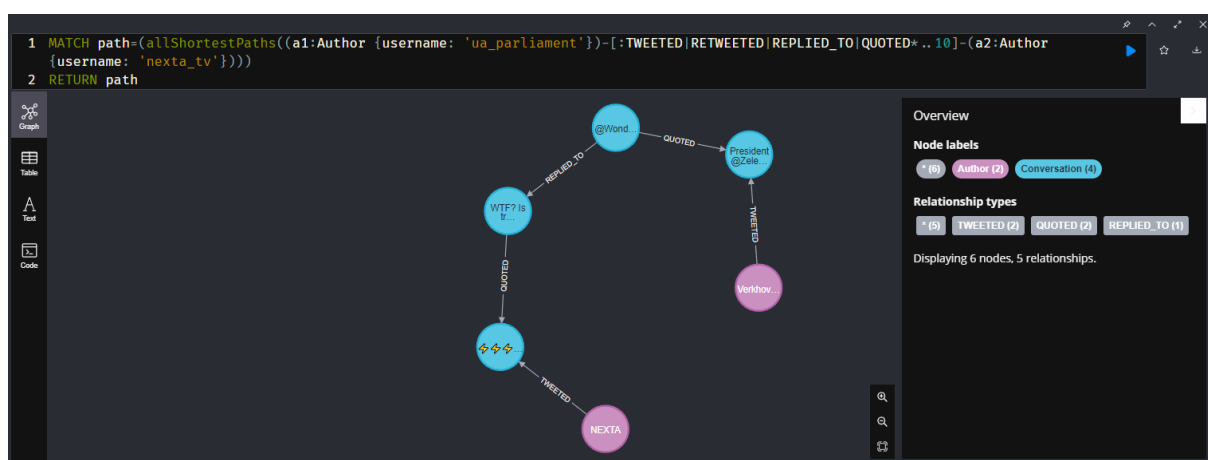
4. otázka

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“.



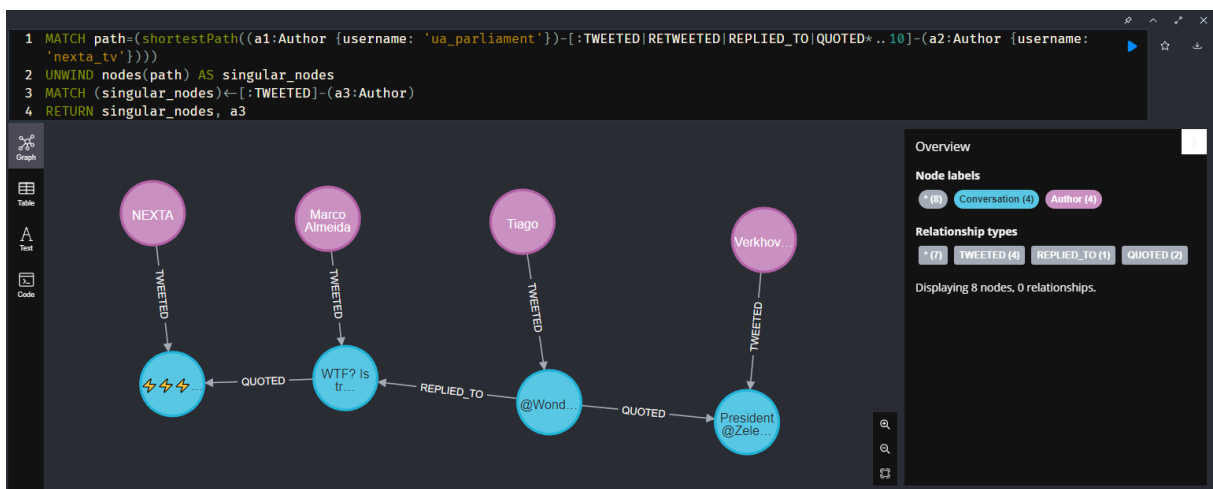
5. otázka

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.



6. otázka

Otázka:

Vypíšte 10 najpoužívanejších hashtagov v datasete aj s celkovým počtom použítí 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 použítí 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.

