

**Zadanie 6 - MongoDB**

**20.12.2022**

**Matej Delinčák**

**Github classroom:** <https://github.com/FIIT-DBS/zadanie-pdt-mateju25>

## 1. Navrhnete dátový model (kolekcie a formát dokumentov) v MongoDB pre dataset tweetov, ktorý bude využívaný mobilnou aplikáciou, ktorá:

a. Zobrazovať tweety jednotlivých používateľov vo forme feedov

b. Zobrazovať jednotlivé tweety a ich retweets

Vytvorím dve kolekcie. Prvá bude obsahovať tweety a všetky dáta, ktoré môže obsahovať okrem autora. Na rozdiel od elasticu, bude autor v zvlášť kolekcií. Je to preto, lebo autor obsahuje polia (rôzne county), ktoré sa budú často updatovať, a kebyže mam autora embednutého v rámci tweetu, tak vzniká veľa duplicit. Tweet teda bude obsahovať len id. Taktiež zmena oproti elasticu, je aj v referenciách, kde som nechal len id parent tweetu a typ referencie.

Pretože podľa use-casov v úlohe tri, my tieto dáta stačia na dané dopyty. Na ostatných entitách typu links, annotations by sa dalo ešte vymyslieť, že by sa dali to zvlášť kolekcie, aby sa ušetrilo miesto. Ale len za podmienky, že by tieto dáta bolo treba len občas a nie pri každom dopyte.

### Kolekcia tweets

```
{
  "id": 1496912601997062148,
  "author_id": 1481371829046587394,
  "content": "Continuing the ukraine footage thread #ukraine #UkraineInvasion #UkraineWar https://t.co/y4uGtjghCe",
  "language": "en",
  "source": "Twitter for Android",
  "retweet_count": 0,
  "reply_count": 1,
  "like_count": 1,
  "quote_count": 0,
  "possibly_sensitive": false,
  "created_at": "2022-02-24T18:19:10+01:00",
  "hashtags": ["ukraine", "UkraineInvasion", "UkraineWar"],
  "links": [
    {
      "url": "https://twitter.com/mr_mercury_/status/1496912601997062148/video/1",
      "title": null,
      "description": null
    }
  ],
  "annotations": [
    {
      "type": "Place", "value": "ukraine", "probability": 0.777
    }
  ],
  "context_annotations": [
    {
      "domain": {
        "name": "Ongoing News Story",
        "description": "Ongoing News Stories like 'Brexit'"
      },
      "entity": {
        "name": "Russo-Ukrainian conflict",
        "description": null
      }
    },
    { ... }
  ],
  "conversation_references": [
    {
      "parent_id": 149691215098103800,
      "type": "replied_to"
    }
  ]
}
```

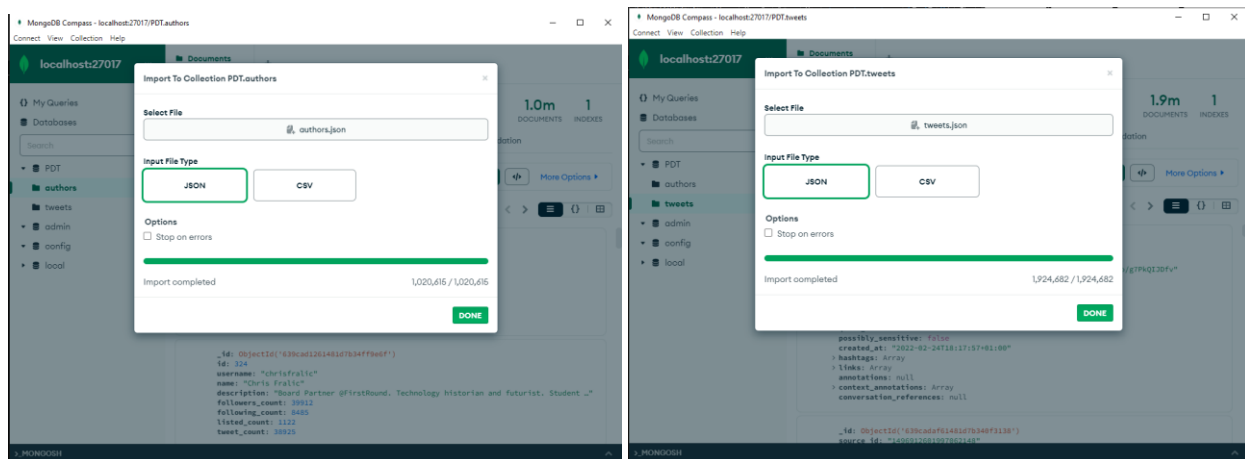
### Kolekcia authors

```
{
  "id": 1481371829046587394,
  "name": "rene",
  "username": "mr_mercury__",
  "description": "I'm a funny guy",
  "tweet_count": 112,
  "listed_count": 3,
  "followers_count": 53,
  "following_count": 18
}
```

## 2. Nainštalujte alebo využite využite online inštanciu MongoDB servera, do ktorého naimportujte všetky tweets (a s nimi spojené data – anotácie, referencie, odkazy a informácie o kontexte) zo dňa 24.02.2022

Pomocou podobného exportu ako pri elasticu som si vytvoril dva súbory. V jednom boli len tweets a s zagregovanými dátami bez autora. A v druhom len čisto autori. Obe exportné query príkladám v .txt súboroch.

Následne som si stiahol MongoDB a taktiež MongoDB Compass, pomocou ktorého som si importol ako tweets tak aj autorov. Ako vidno nižšie.



Ukážka tweetu z návrhu:

```
_id: ObjectId('639cb16f61481d7b342c8f83')
id: "1496912601997062148"
author_id: "1481371829046587394"
content: "Continuing the ukraine footage thread #ukraine #UkraineInvasion #Ukrai..."
language: "en"
source: "Twitter for Android"
retweet_count: 0
reply_count: 1
like_count: 1
quote_count: 0
possibly_sensitive: false
created_at: "2022-02-24T18:19:10+01:00"
hashtags: Array
  0: "ukraine"
  1: "UkraineInvasion"
  2: "UkraineWar"
links: Array
  0: Object
    url: "https://twitter.com/mr_mercury_/status/1496912601997062148/video/1"
    title: null
    description: null
annotations: Array
  0: Object
    type: "Place"
    value: "ukraine"
    probability: 0.777
context_annotations: Array
  0: Object
    domain: Object
      name: "Ongoing News Story"
      description: "Ongoing News Stories like 'Brexit'"
    entity: Object
      name: "Russo-Ukrainian conflict"
      description: null
  1: Object
    domain: Object
      name: "Ongoing News Story"
      description: "Ongoing News Stories like 'Brexit'"
    entity: Object
```

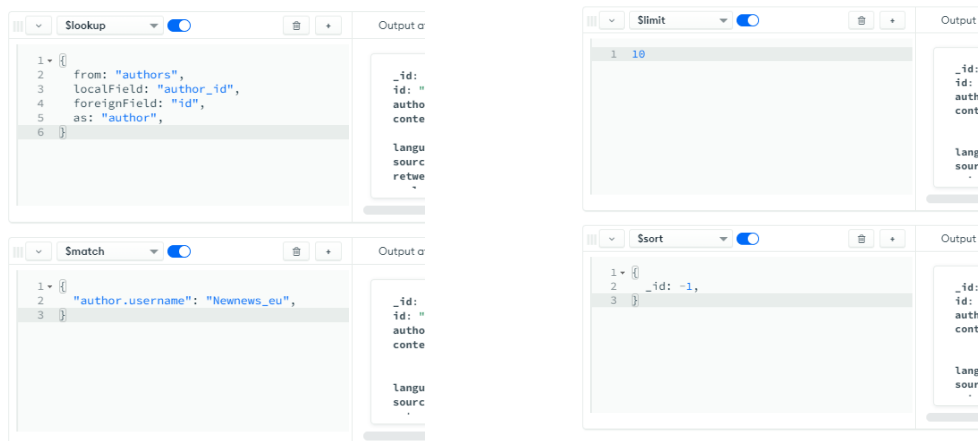
### 3. Napíšte dotaz, ktorý nad importovanou databázou:

#### a. Vypíše posledných 10 tweetov pre autora, ktorý má username Newnews\_eu

Najprv som sa pohral v MongoDB shelli. Potreboval som si k tweetom lookupnúť autora a na základe jeho username vyfiltrovať záznamy. Nakoniec som sortol podľa \_id a limitol 10. Keďže som „joinoval“ kolekcie, tak som potreboval urobiť agregáciu.

```
db.tweets.aggregate([
  { $lookup: {
    from: "authors",
    localField: "author_id",
    foreignField: "id",
    as: "author", }},
  { $match: { "author.username": "Newnews_eu" }},
  { $sort: { _id: -1 }},
  { $limit: 10 }])
```

Potom som túto query prepísal do MongoDB Compass cez pipeline, kde som spravil v podstate to isté.



Prvý dokument z 10.

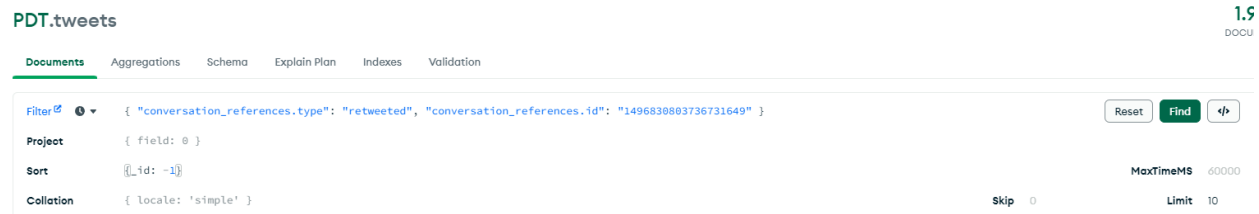
```
{
  "_id": "1496994513444913152",
  "author_id": "1496994513444913152",
  "content": "BREAKING | Clashes continue in Sum. #Ukraine #UkraineRussiaCrisis",
  "language": "en",
  "source": "Twitter for Android",
  "retweet_count": 3,
  "reply_count": 0,
  "like_count": 13,
  "quote_count": 0,
  "possibly_sensitive": false,
  "created_at": "2022-02-24T21:44:39+01:00",
  "hashtags": [
    "BREAKING",
    "Ukraine",
    "UkraineRussiaCrisis",
    "UkraineRussiaCrisis",
    "Russia",
    "RussiaUkraineConflict",
    "RussiaUkraineCrisis",
    "Putin",
    "nowarinukraine"
  ],
  "links": [
    {
      "url": "https://twitter.com/Newnews_eu/status/1496994513444913152/video/1",
      "title": null
    }
  ]
}
```

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

Na rozdiel od prvého dopytu som nepoužil agregáciu, ale iba find a na konci som pridal limit 10. Vnútri find sú filtre na typ vzťahu a na sortnutie podľa \_id.

```
db.tweets.find({
  'conversation_references.type': 'retweeted',
  'conversation_references.id': '1496830803736731649' })
.limit(10).sort({_id: -1})
```

Taktiež som si vyskúšal aj GUI v MongoDB Compass a vyzeralo to takto:



Prvý dokument z desiatich:

```
{
  "_id": ObjectId("639cb40b61481d7b3449bdb9"),
  "id": "1496958254294966274",
  "author_id": "1307745070259548168",
  "content": "RT @Hewnews_eu: 🚨#BREAKING | A father who sent his family to a safe zone bid farewell to his little girl and stayed behind to fight ...",
  "language": "en",
  "source": "Twitter for Android",
  "retweet count": 12704,
  "reply count": 0,
  "like count": 0,
  "quote count": 0,
  "possibly_sensitive": false,
  "created_at": "2022-02-24T21:20:34+01:00",
  "hashtags": [ "BREAKING" ],
  "links": null,
  "annotations": null,
  "context_annotations": null,
  "conversation_references": [ { id: '1496830803736731649', type: 'retweeted' } ]
},
```