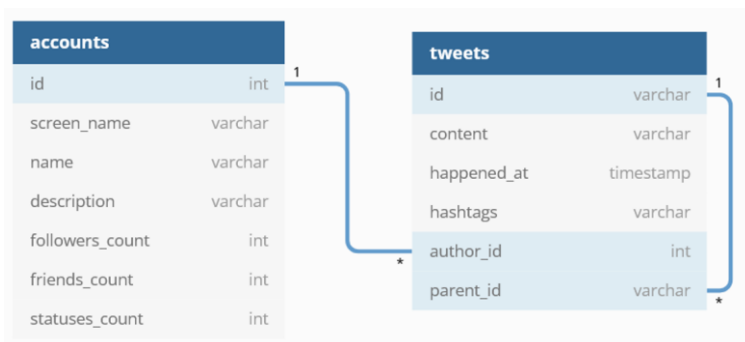


1. Pre znázornenie mongoDB schémy som využil online nástroj. Schéma je veľmi podobná tej v postgrese. Využíva referencing (v kolekcii tweets atribút author_id referencuje na hodnotu atribútu id v kolekcii accounts) a tiež embedding (kolekcia tweets obsahuje pole hashtags, v ktorom sú uložené hashtagy daného tweetu). Túto schému som zvolil z dôvodu, že mi príde jednoduchá a efektívna. Pri vytváraní nových záznamov tweetov sa pridá nový záznam, ktorý obsahuje údaje o tom, ktorý používateľ ho vytvoril (id). Je to lepšie ako vytvárať v kolekcii accounts pole id-čiek jednotlivých tweetov daného používateľa a to z dôvodu, že pri vkladaní nového záznamu v tomto prípade nemusíme zasahovať aj do záznamu o používateľovi. Pre hashtagy je nadbytočné vytvárať vlastnú kolekciu, pretože jediný dôležitý údaj je samotný hashtag, ktorý môžeme uložiť priamo do tweetu – takto vytvorený tweet už nemusíme nijako dodatočne upravovať (z pohľadu hashtagov) a tiež sa vyhneme vyhľadávaniu a spájaniu záznamov z dvoch kolekcií.



Názorná ukážka záznamu z kolekcie tweets:

```

{
  _id: new ObjectId("619e5dd4e1a75c02abe7de00"),
  id: '1034289685227614208',
  content: '#TrueMAGA 🦊🐱\n' +
    'PEDOPHILIA IS AN EPIDEMIC IN USA-WORLD: \n' +
    '\n' +
    'So called elites-many pedophile rings connected to Royals-Politicians-Hollywood. Our Justice Dept. @ AG Sessions uses padded gloves = #CrimeMonitor, only small fish-no big fish. \n' +
    '\n' +
    '#PedogateIsReal\n' +
    '#PedoPope \n' +
    '#JUSTICENOW! https://t.co/oScb63oA4y',
  happened_at: '2018-08-28T04:00:51.000Z',
  author_id: '553991829',
  parent_id: null,
  hashtags: [
    'TrueMAGA',
    'CrimeMonitor',
    'JUSTICENOW',
    'PedoPope',
    'PedogateIsReal'
  ]
}

```

Názorná ukážka záznamu z kolekcie accounts:

```

{
  _id: new ObjectId("619e851ef323bda557f1df4a"),
  id: '785',
  screen_name: 'kfury',
  name: 'Kevin Fox for Biden/Harris 2020 us 🦊🐱',
  description: null,
  followers_count: null,
  friends_count: null,
  statuses_count: null
}

```

2. Databázu mongo som už mal nainštalovanú v dockery ešte z minulého semestra. Pri importovaní datasetu z postgresu do mongo som využil nasledovné query:

Select údajov z tabuľky tweets:

```
SELECT t.id, t.content, t.happened_at, t.author_id, t.parent_id, ARRAY_AGG (h.value) hashtags
FROM tweets t
JOIN tweet_hashtags th ON th.tweet_id = t.id
JOIN hashtags h ON h.id = th.hashtag_id
WHERE t.compound < -0.5 OR t.compound > 0.5
GROUP BY t.id
OFFSET 0
LIMIT 5000;
```

OFFSET a LIMIT som použil pretože som si nebol istý, či mi počítač dokáže preniesť všetky záznamy naraz. Takto som záznamy rozdelil do viacerých skupín a do monga som záznamy vkladal po skupinách pomocou for-cyklu, ktorý posúva OFFSET (hodnota offsetu sa mení na základe iterátora v cykle).

Select údajov z tabuľky accounts:

```
SELECT a.*
FROM (SELECT DISTINCT a.id
FROM accounts a
JOIN tweets t ON t.author_id = a.id
WHERE t.compound < -0.5 OR t.compound > 0.5) accs
JOIN accounts a ON a.id = accs.id
OFFSET 0
LIMIT 10000;
```

Pozn.: vo finálnom zdrojovom kóde je volanie funkcie pre import záznamov z tabuliek tweets a accounts zakomentované, aby sa to pri opätovnom spúšťaní nepokúšalo importovať databázu znova.

3a.

```
{
  _id: new ObjectId("619e8700f323bda557f30f01"),
  id: '885448167168364544',
  screen_name: 'TomaszTrembowski',
  name: 'Tomasz Trembowski',
  description: 'Kontaktowy matros. \n' +
    'Zwolennik szerokiego dostępu do broni palnej. Jestem
chrześcijaninem.#IPPTV',
  followers_count: 272,
  friends_count: 424,
  statuses_count: 18158
}
1184554425479761920
{
  _id: new ObjectId("619e8865f323bda557f38a78"),
  id: '1184554425479761920',
  screen_name: 'ZawdzkiP',
  name: 'Zawdzki Pawel',
  description: 'srednie techn ,',
  followers_count: 170,
  friends_count: 90,
  statuses_count: 53729
}
Exit
PS C:\Users\Marek\Desktop\PDT\PDT-zadanie_4> node index
Connected to MongoDB
[
  {
    _id: new ObjectId("619e5fc2e1a75c02abe8908c"),
```

```

    id: '1247109014694932481',
    content: "RT @AgainstTideTV: This is a war, China's worse than Soviet
Russia. #BoycottChina and #MakeChinaPay #ChinaLiedPeopleDied - @SolomonYue
@GO...",
    happened_at: 2020-04-06T10:28:53.000Z,
    author_id: '885448167168364544',
    parent_id: '1246874043682299904',
    hashtagss: [ 'MakeChinaPay', 'BoycottChina', 'ChinaLiedPeopleDied' ],
    author: 'Tomasz Trembowski'
  },
  {
    _id: new ObjectId("619e5f5ae1a75c02abe86e1e"),
    content: "RT @AgainstTideTV: This is a war, China's worse than Soviet
Russia. #BoycottChina and #MakeChinaPay #ChinaLiedPeopleDied - @SolomonYue
@GO...",
    happened_at: 2020-04-05T21:11:07.000Z,
    author_id: '1184554425479761920',
    parent_id: '1246874043682299904',
    hashtagss: [ 'MakeChinaPay', 'BoycottChina', 'ChinaLiedPeopleDied' ],
    author: 'Zawdzki Pawel'
  }
]

```

Exit

PS C:\Users\Marek\Desktop\PDT\PDT-zadanie_4> node index

Connected to MongoDB

```

{
  _id: new ObjectId("619e863ff323bda557f2aefc"),
  id: '3003720760',
  screen_name: 'Marndin12',
  name: 'Martin Noakes',
  description: 'https://t.co/lIHnBiqfls\nhttps://t.co/HOA3RgtAc9...',
  followers_count: 1107,
  friends_count: 957,
  statuses_count: 18755,
  tweets: [
    {
      _id: new ObjectId("619e5dd4e1a75c02abe7de27"),
      id: '1219961881923899392',
      content: "Agenda 2030 explained in a 3 minute song, that neatly
outlines what REALLY could happen if it's forced upon us They will be
looking to poison us via vaccines because of this so called Chinese
coronavirus outbreak LIARS - https://t.co/8F3GRJBy3e Please Share #Agenda21
#Agenda2030",
      happened_at: 2020-01-22T12:35:53.000Z,
      author_id: '3003720760',
      parent_id: null,
      hashtagss: [Array]
    },
    {
      _id: new ObjectId("619e5dd4e1a75c02abe7de28"),
      id: '1219990024587968512',
      content: "Coronavirus will end up being biowar technology, in order
to facilitate the rolling out of mandatory vaccinations (laced with toxins
& more) - Here's a three minute futuristic song about agenda 2030 -
https://t.co/8F3GRJBy3e Please Share #Agenda21 #Agenda2030 #coronavirus",
      happened_at: 2020-01-22T14:27:42.000Z,
      author_id: '3003720760',
      parent_id: null,
      hashtagss: [Array]
    }
  ]
}

```

```

    _id: new ObjectId("619e5dd4e1a75c02abe7de29"),
    id: '1219995830184304641',
    content: 'They are moving fast now - Coronavirus is clearly a plan
along with 5G, Chemtrails, GM foods, Vaccinations, Radiation, Climate
Crisis & the manufactured bankers wars - Agenda 2030 described in a 3
minute futuristic song - https://t.co/8F3GRJBy3e #Agenda21 #Agenda2030
#coronavirus',
    happened_at: 2020-01-22T14:50:46.000Z,
    author_id: '3003720760',
    parent_id: null,
    hashtagss: [Array]
  },
  {
    _id: new ObjectId("619e5dd4e1a75c02abe7de2a"),
    id: '1220030824252428288',
    content: 'Coronavirus is clearly a biowar plan along with 5G,
Chemtrails, GM foods, Vaccinations, Radiation, Climate Crisis & the
manufactured bankers wars - Agenda 2030 The future describe in a 3 minute
song - https://t.co/8F3GRJBy3e #Agenda21 #Agenda2030 #coronavirus Plz RT
https://t.co/ZmYUsKS1PE',
    happened_at: 2020-01-22T17:09:50.000Z,
    author_id: '3003720760',
    parent_id: null,
    hashtagss: [Array]
  },
  {
    _id: new ObjectId("619e5dd4e1a75c02abe7de33"),
    id: '1220366860304162817',
    content: 'If anyone could help me spread this song about Agenda 2030
on other social media platforms, I would very much appreciate it -
https://t.co/8F3GRJBy3e Plz Share #Agenda21 #Agenda2030 #coronavirus
#bioweapon #chemtrails #5G #GMFood #GunControl',
    happened_at: 2020-01-23T15:25:07.000Z,
    author_id: '3003720760',
    parent_id: null,
    hashtagss: [Array]
  },
  {
    _id: new ObjectId("619e5dd4e1a75c02abe7de36"),
    id: '1220404632805134336',
    content: 'In China right now the elite are testing what will happen
when they lock us all down on a pack of lies to enable forced vaccinations
(agenda 21). People are literally fighting over food -
https://t.co/8F3GRJBy3e Plz Share #Agenda21 #Agenda2030 #coronavirus
#bioweapon #chemtrails',
    happened_at: 2020-01-23T17:55:13.000Z,
    author_id: '3003720760',
    parent_id: null,
    hashtagss: [Array]
  },
  {
    _id: new ObjectId("619e5dd4e1a75c02abe7dea7"),
    id: '1220672975965511681',
    content: '@Gyrobite @PSneeboer @OmroepON @marcelcrok @Clintel Hi Rob -
I have written quite a few controversial songs over the years,
& this one about Agenda 21 has been without doubt the most difficult to
spread https://t.co/8F3GRJBy3e Plz RT #Agenda21 #Agenda2030 #coronavirus
#bioweapon #chemtrails #5G #GMFood',
    happened_at: 2020-01-24T11:41:31.000Z,
    author_id: '3003720760',
    parent_id: null,
  }

```

```

    hashtagss: [Array]
  },
  {
    _id: new ObjectId("619e5dd4e1a75c02abe7dee5"),
    id: '1220730677685116930',
    content: '@mall1976retaf @praisin7 @LadyChesapeake @1USMarine4MAGA
@GelicaAF @TheAkmed @AzDerbyGirl @HarrietNix @NoLeftyEsty @WaydeDe
@EliLucal @Rabbit70414459 @TammyChabot @Adorable_Todd @FlashToso
@baldeguy56 @CWarrior17 @Brenro4 @howienudet @warrenwarmachil @lavo323
@NancyLucky @ClwtrBkAtty @QUIPZONE7 @GiacomoCeraolo @steviewheel1
@tehipisehaka @Mark_trumpiskag @CourageMill @_carolina_girl_ @Silentwoo
@sapiocrazy @BPNevada @Richard4m @HPot77 @zx225_skeeter @WilliamMillen5
@EYEFOLLOWHIM @TheCryptKeepel4 @AmericanUmbra @TakebackUSA2 @TAwakened1
@KaashifMohammell @JamesCali3 @DFBHarvard I have written quite a few
controversial songs over the years, & this one about Agenda 21 has
been without doubt the most difficult to spread around
https://t.co/8F3GRJBy3e Plz RT #Agenda21 #Agenda2030 #coronavirus
#bioweapon #chemtrails #5G #GMFood :-(',
    happened_at: 2020-01-24T15:30:48.000Z,
    author_id: '3003720760',
    parent_id: null,
    hashtagss: [Array]
  },
  {
    _id: new ObjectId("619e5dd4e1a75c02abe7df61"),
    id: '1221027188948066304',
    content: '@FULLDEFENCE I have placed a download link in the info
section of the YouTube vid - https://t.co/8F3GRJBy3e Please feel free to
download and use this song to help spread the word in anyway you can think
of. Radio & TV shows have my permission to play it #agenda21
#agenda2030 #coronavirus',
    happened_at: 2020-01-25T11:09:02.000Z,
    author_id: '3003720760',
    parent_id: null,
    hashtagss: [Array]
  },
  {
    _id: new ObjectId("619e5dd4e1a75c02abe7dfcc"),
    id: '1221419341972983808',
    content: '@SusanHavens10 @LaurenLindsayDj I have written quite a few
controversial songs over the years. This one's about what could
manifest if Agenda 21/2030 succeeds - https://t.co/8F3GRJBy3e Please feel
free to use this song to help spread the word #agenda21 #agenda2030
#coronavirus',
    happened_at: 2020-01-26T13:07:18.000Z,
    author_id: '3003720760',
    parent_id: null,
    hashtagss: [Array]
  }
]
}

```

3b. Ako bolo spomenuté na Slacku, ID napísané v zadaní nemá žiadne retweety. Použil som teda druhé, ktoré bolo napísané v komentároch

```
[
  {
    _id: new ObjectId("619e5fc2e1a75c02abe8908c"),
    id: '1247109014694932481',
    content: "RT @AgainstTideTV: This is a war, China's worse than Soviet Russia. #BoycottChina and #MakeChinaPay #ChinaLiedPeopleDied - @SolomonYue @GO..",
    happened_at: 2020-04-06T10:28:53.000Z,
    author_id: '885448167168364544',
    parent_id: '1246874043682299904',
    hashtagss: [ 'MakeChinaPay', 'BoycottChina', 'ChinaLiedPeopleDied' ],
    author: 'Tomasz Trembowski'
  },
  {
    _id: new ObjectId("619e5f5ae1a75c02abe86e1e"),
    id: '124690824957787393',
    content: "RT @AgainstTideTV: This is a war, China's worse than Soviet Russia. #BoycottChina and #MakeChinaPay #ChinaLiedPeopleDied - @SolomonYue @GO..",
    happened_at: 2020-04-05T21:11:07.000Z,
    author_id: '1184554425479761920',
    parent_id: '1246874043682299904',
    hashtagss: [ 'MakeChinaPay', 'BoycottChina', 'ChinaLiedPeopleDied' ],
    author: 'Zawdzki Pawel'
  }
]
```