Zhamilya Saparova

HW 2 - collection indexing using Elasticsearch

Firstly, I installed elasticsearch and created cloud elastic to connect to the elasticsearch server. Further, I created an index with the name “hello”. I wrote settings for the stemming with lowercase filter, “whitespace” tokenizer and snowball english analyzer. For the setting without stemming, I used only a lowercase filter and “whitespace” tokenizer.
Then, I recreated the index with proper settings. To index documents, I used parallel_bulk API. The results for setting with and without stemming are shown below:
With stemming:

```
CPU times: user 1min 19s, sys: 452 ms, total: 1min 19s
Wall time: 2min 17s
```

Without stemming:

```
CPU times: user 1min 21s, sys: 497 ms, total: 1min 22s
Wall time: 1min 38s
```

As a next step, I performed a query search. Firstly, I used the simplest query “match_all”, which takes nothing and returns all the documents. It took
With stemming:

```
CPU times: user 48.7 ms, sys: 4.08 ms, total: 52.8 ms
Wall time: 770 ms
```

Without stemming:

```
CPU times: user 131 ms, sys: 4.19 ms, total: 135 ms
Wall time: 5.35 s
```

Then, I randomly took the phrase “private boats” and passed it as a query. It showed top 20 results for this query with scores. It took
With stemming:

```
CPU times: user 100 ms, sys: 6.1 ms, total: 106 ms
Wall time: 3.02 s
```

Without stemming:

```
CPU times: user 144 ms, sys: 11.2 ms, total: 155 ms
Wall time: 10 s
```

Further, I ran test queries and received top 20 results for each query and estimated query execution time. Below, query index and query execution time of 10 queries are shown (overall, there are 100 queries).

```
0       0.04542422294616699
1       0.029604196548461914
2       0.03773689270019531
3       0.026151418685913086
4       0.022423505783081055
5       0.025595664978027344
6       0.03455328941345215
7       0.03414750099182129
8       0.024725914001464844
9       0.025181055068969727
```

I used ir_measures to format qrels and runs. My runs:

Zhamilya Saparova

```
{P@10:     0.22599999999999987,     AP:     0.0004083007253942143,     P@20:
0.16450000000000006}
TIME 0.15152812004089355
```

Creator’s runs:

```
{P@10:     0.3000000000000005,     AP:     0.010808749000016847,     P@20:
0.3000000000000005}
TIME 0.8022594451904297
```