

Assignment 2: Querying ES

Στο δεύτερο assignment δημιουργήσαμε ένα index στο elastic search βάση των δεδομένων που είχαμε. Αφού δημιουργήσαμε το mapping και ανεβάσαμε όσων το δυνατό περισσότερα δεδομένα, κάναμε ένα hybrid search για να βρίσκουμε αρχεία όμοια μεταξύ τους.

Πρώτα πήραμε τα embeddings από το μοντέλο "allenai/scibert_scivocab_uncased" το οποίο έχει εκπαιδευτεί σε επιστημονικά papers και κρίθηκε το καταλληλότερο. Στη συνέχεια κάναμε ένα hybrid search, δηλαδή πρώτα ένα lexical query (όμοιο abstract ή Title) και ένα semantic query (τα πιο κοντινά embeddings βάση cosine similarity) και τα συμψηφίσαμε.

To service έγινε exposed μέσω fastapi.

Για να ξεκινήσουμε το es και να σεταρουμε το kibana:

```
$docker network create elastic
```

```
$docker pull docker.elastic.co/elasticsearch/elasticsearch:7.17.22
```

```
$docker run --name es01-test --net elastic -p 127.0.0.1:9200:9200 -p 127.0.0.1:9300:9300 -e "discovery.type=single-node" docker.elastic.co/elasticsearch/elasticsearch:7.17.22
```

Σε ένα διαφορετικό terminal

```
$docker pull docker.elastic.co/kibana/kibana:7.17.22
```

```
$docker run --name kib01-test --net elastic -p 127.0.0.1:5601:5601 -e "ELASTICSEARCH_HOSTS=http://es01-test:9200" docker.elastic.co/kibana/kibana:7.17.22
```

Create index and launch the app

```
$python create_index.py
```

```
$python app.py
```

Send a request:

```
$python client.py
```