elk_base_documents_template

November 1, 2023

```
[ ]: !pip install elasticsearch
```

```
Requirement already satisfied: elasticsearch in /opt/conda/lib/python3.11/site-
packages (8.10.1)
Requirement already satisfied: elastic-transport<9,>=8 in
/opt/conda/lib/python3.11/site-packages (from elasticsearch) (8.10.0)
Requirement already satisfied: urllib3<3,>=1.26.2 in
/opt/conda/lib/python3.11/site-packages (from elastic-
transport<9,>=8->elasticsearch) (2.0.7)
Requirement already satisfied: certifi in /opt/conda/lib/python3.11/site-
packages (from elastic-transport<9,>=8->elasticsearch) (2023.7.22)
```

```python
[ ]: from datetime import datetime
     from elasticsearch import Elasticsearch

     ELASTIC_PASSWORD = "p2iFCHUbC7ze1QoIMVw"

     es = Elasticsearch("http://elasticsearch:9200",
                        basic_auth=("elastic", ELASTIC_PASSWORD))

     es.info()
```

```
[ ]: ObjectApiResponse({'name': 'es-node', 'cluster_name': 'tdt4117-ir-data-cluster',
     'cluster_uuid': 'rAkxpaFNSfypbZhfl9rXuA', 'version': {'number': '8.4.2',
     'build_flavor': 'default', 'build_type': 'docker', 'build_hash':
     '89f8c6d8429db93b816403ee75e5c270b43a940a', 'build_date':
     '2022-09-14T16:26:04.382547801Z', 'build_snapshot': False, 'lucene_version':
     '9.3.0', 'minimum_wire_compatibility_version': '7.17.0',
     'minimum_index_compatibility_version': '7.0.0'}, 'tagline': 'You Know, for
     Search'})
```

```python
[ ]: # TODO import documents into elastic
     import os

     es.indices.delete(index="index", ignore_unavailable=True) # Delete the index if␣
     ↪it exists

     # Create the index
```

```python
es.indices.create(index='index')

collection = []
files = sorted(os.listdir("DataAssignment4"))
print(files)
corpus = []
for file in files:
    with open("DataAssignment4/" + file, "r") as f:
        corpus.append({"title": file, "content": f.read()})

# Index the documents
for i, doc in enumerate(corpus):
    es.index(index="index", id=i, body=doc)
```

```
['Text1.txt', 'Text2.txt', 'Text3.txt', 'Text4.txt', 'Text5.txt', 'Text6.txt']
```

```python
def pretty_print(query, res):
    print("Query: ", query)
    print("Number of results: ", res["hits"]["total"]["value"])
    print("Results: ")
    for hit in res["hits"]["hits"]:
        print("Score: ", hit["_score"])
        print("Title: ", hit["_source"]["title"])
        print("Content: ", hit["_source"]["content"][:100] + "...")
```

```python
# TODO Query "claim"
query = "claim"
res = es.search(index="index", query={"match": {"content": query}})
pretty_print(query, res)
```

```
Query:  claim
Number of results:  1
Results:
Score:  1.5510131
Title:  Text2.txt
Content:  One morning, when Gregor Samsa woke from troubled dreams, he found
himself transformed in his bed in…
```

```python
# TODO Query "claim*" (wildcard)
query = "claim*"
res = es.search(index="index", query={"wildcard": {"content": query}})
pretty_print(query, res)
```

```
Query:  claim*
Number of results:  2
Results:
Score:  1.0
Title:  Text2.txt
Content:  One morning, when Gregor Samsa woke from troubled dreams, he found
```

himself transformed in his bed in…
Score: 1.0
Title: Text6.txt
Content: But I must explain to you how all this mistaken idea of denouncing
pleasure and praising pain was bo…

```
[ ]: # TODO Query "claims of duty"
     query = "claims of duty"
     res = es.search(index="index", query={"match_phrase": {"content": query}})
     pretty_print(query, res)
```

Query: claims of duty
Number of results: 1
Results:
Score: 4.975612
Title: Text6.txt
Content: But I must explain to you how all this mistaken idea of denouncing
pleasure and praising pain was bo…

```
[ ]: # TODO Query "claims of duty" in an alternative way
     query = "claims of duty"
     res = es.search(index="index", query={"match": {"content": query}})
     pretty_print(query, res)
```

Query: claims of duty
Number of results: 6
Results:
Score: 5.417975
Title: Text6.txt
Content: But I must explain to you how all this mistaken idea of denouncing
pleasure and praising pain was bo…
Score: 0.1597732
Title: Text1.txt
Content: A wonderful serenity has taken possession of my entire soul, like
these sweet mornings of spring whi…
Score: 0.15703472
Title: Text4.txt
Content: Far far away, behind the word mountains, far from the countries
Vokalia and Consonantia, there live …
Score: 0.15641768
Title: Text5.txt
Content: The European languages are members of the same family. Their separate
existence is a myth. For scien…
Score: 0.15391785
Title: Text2.txt
Content: One morning, when Gregor Samsa woke from troubled dreams, he found
himself transformed in his bed in…
Score: 0.14406839
Title: Text3.txt

Content: The quick, brown fox jumps over a lazy dog. DJs flock by when MTV ax quiz prog. Junk MTV quiz graced…