

elk-base-documents-template

November 2, 2023

```
[15]: %pip install elasticsearch
```

```
Requirement already satisfied: elasticsearch in c:\users\andre\appdata\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\local-packages\python310\site-packages (8.10.1)
```

```
Requirement already satisfied: elastic-transport<9,>=8 in c:\users\andre\appdata\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\local-packages\python310\site-packages (from elasticsearch) (8.10.0)
```

```
Requirement already satisfied: urllib3<3,>=1.26.2 in c:\users\andre\appdata\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\local-packages\python310\site-packages (from elastic-transport<9,>=8->elasticsearch) (1.26.13)
```

```
Requirement already satisfied: certifi in c:\users\andre\appdata\local\packages\pythonsoftwarefoundation.python.3.10_qbz5n2kfra8p0\localcache\local-packages\python310\site-packages (from elastic-transport<9,>=8->elasticsearch) (2022.12.7)
```

Note: you may need to restart the kernel to use updated packages.

[notice] A new release of pip is available: 23.2.1 -> 23.3.1

[notice] To update, run: C:\Users\andre\AppData\Local\Microsoft\WindowsApps\PythonSoftwareFoundation.Python.3.10_qbz5n2kfra8p0\python.exe -m pip install --upgrade pip

```
[16]: from datetime import datetime
      from elasticsearch import Elasticsearch
      from indexer import CustomIndexer

      ELASTIC_PASSWORD = "p2iFCHUbC7ze1QoIMVw"

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

      es.info()
```

```
[16]: ObjectApiResponse({'name': 'es-node', 'cluster_name': 'tdt4117-ir-data-cluster',
'cluster_uuid': 'i6jKZRxDJRJ212GIEPZpLWA', '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'})
```

```
[17]: # Import documents into elastic
index_name = "k"
corpus = []
for i in range(1,7):
    with open(f"./DataAssignment4/Text{i}.txt") as f:
        title = f"Document {i-1}"
        content = f.read()
        corpus.append({'title': title, 'content': content})
if not es.indices.exists(index=index_name):
    es.indices.create(index=index_name)
for i, doc in enumerate(corpus):
    es.index(index=index_name, id=i, body=doc)

# Initiate our own indexer:
indexer = CustomIndexer([doc['content'] for doc in corpus])
```

```
100%|      | 6/6 [00:00<00:00, 56.08it/s]
100%|      | 6/6 [00:00<00:00, 2997.72it/s]
```

```
[18]: def print_es(results):
    print("ElasticSearch:")
    print(f"Results ({len(results['hits']['hits'])}):")
    for hit in results['hits']['hits']:
        content = hit['_source']['content'][:100].replace("\n", "")
        print(f"ID: {hit['_id']}, Score: {hit['_score']}, Content: {content}")
def print_custom(results):
    print("Custom (ours):")
    if results:
        print(f"Results ({len(results)}):")
        for res in results:
            content = res['content'][:100].replace("\n", "")
            print(f"ID: {res['id']}, Score: {res['score']:.4f}, Content: {content}")
    else:
        print("No results...")
```

```
[19]: # Query "claim"
query = "claim"
es_query = {
    "query": {
        "match": {
```

```

        "content": query
    }
}
}
print_es(es.search(index=index_name, body=es_query))
print()
print_custom(indexer.search(query))

```

ElasticSearch:

Results (1):

ID: 1, Score: 1.5535183, Content: One morning, when Gregor Samsa woke from troubled dreams, he found himself transformed in his bed in

Custom (ours):

Results (2):

ID: 5, Score: 4.7549, Content: But I must explain to you how all this mistaken idea of denouncing pleasure and praising pain was bo

ID: 1, Score: 1.5850, Content: One morning, when Gregor Samsa woke from troubled dreams, he found himself transformed in his bed in

```

[20]: # Query "claim*"
query = "claim*"

es_query = {
    "query": {
        "wildcard": {
            "content": {
                "value": query
            }
        }
    }
}

print_es(es.search(index=index_name, body=es_query))
print()
print_custom(indexer.search(query))

```

ElasticSearch:

Results (2):

ID: 1, Score: 1.0, Content: One morning, when Gregor Samsa woke from troubled dreams, he found himself transformed in his bed in

ID: 5, Score: 1.0, Content: But I must explain to you how all this mistaken idea of denouncing pleasure and praising pain was bo

Custom (ours):

Results (2):

ID: 5, Score: 4.7549, Content: But I must explain to you how all this mistaken idea of denouncing pleasure and praising pain was bo

ID: 1, Score: 1.5850, Content: One morning, when Gregor Samsa woke from troubled

dreams, he found himself transformed in his bed in

```
[21]: # Query "claims of duty"
query = "claims of duty"
es_query = {
    "query": {
        "match": {
            "content": query
        }
    }
}
print_es(es.search(index=index_name, body=es_query))
print()
print_custom(indexer.search(query))
```

ElasticSearch:

Results (6):

ID: 5, Score: 5.4214234, Content: But I must explain to you how all this mistaken idea of denouncing pleasure and praising pain was bo
ID: 0, Score: 0.15978271, Content: A wonderful serenity has taken possession of my entire soul, like these sweet mornings of spring whi
ID: 3, Score: 0.15705192, Content: Far far away, behind the word mountains, far from the countries Vokalia and Consonantia, there live
ID: 4, Score: 0.15643656, Content: The European languages are members of the same family. Their separate existence is a myth. For scien
ID: 1, Score: 0.15394345, Content: One morning, when Gregor Samsa woke from troubled dreams, he found himself transformed in his bed in
ID: 2, Score: 0.14411825, Content: The quick, brown fox jumps over a lazy dog.
DJs flock by when MTV ax quiz prog. Junk MTV quiz graced

Custom (ours):

Results (2):

ID: 5, Score: 20.2647, Content: But I must explain to you how all this mistaken idea of denouncing pleasure and praising pain was bo
ID: 1, Score: 1.5850, Content: One morning, when Gregor Samsa woke from troubled dreams, he found himself transformed in his bed in

```
[22]: # Query "claims of duty" in an alternative way
# The alternative way uses match_phrase which returns only documents that
#   ↳ contain the query exactly as it is written. Which in this case turned out to
#   ↳ only be document 5
query = "claims of duty"
es_query = {
    "query": {
        "match_phrase": {
            "content": query
        }
    }
}
```

```
    }  
}  
print_es(es.search(index=index_name, body=es_query))  
print()  
print_custom(indexer.search(query))
```

ElasticSearch:

Results (1):

ID: 5, Score: 4.979805, Content: But I must explain to you how all this mistaken idea of denouncing pleasure and praising pain was bo

Custom (ours):

Results (2):

ID: 5, Score: 20.2647, Content: But I must explain to you how all this mistaken idea of denouncing pleasure and praising pain was bo

ID: 1, Score: 1.5850, Content: One morning, when Gregor Samsa woke from troubled dreams, he found himself transformed in his bed in