

Brent Claypool  
IEMS 308  
March 19 2021

#### HW 4

##### Methodology:

For this assignment, I wanted to use ElasticSearch as a way to do my QA system. I created an index called articles and loaded all the articles into it via:

```
documents_config = {
    'settings': {
        "analysis": {
            "analyzer": {
                "stop_stem_analyzer": {
                    "type": "custom",
                    "tokenizer": "standard",
                    "filter": [
                        "lowercase",
                        "stop",
                        "snowball"
                    ]
                }
            }
        }
    },
    "mappings": {
        "dynamic": "strict",
        "properties": {
            "articles_content": {"type": "text", "analyzer": "stop_stem_analyzer"}
        }
    }
}

index_name = 'articles'
es.indices.create(index=index_name,body=documents_config,ignore=400)

and:

for i in range(len(file_list)):
    #for fn in file_sample:
        #text = open('BI-articles' + '/' + fn, 'r',errors = 'ignore').read()
    text_cont = {"articles_content": open('BI-articles' + '/' + file_list[1], 'r',errors = 'ignore').read()}
    try:
        index_status= es.index(index=index_name,id=i,body=text_cont)
    except:
        print(f'Error loading document {i}.')
```

That code basically created 731 documents in the index and within all the documents, the content was the content of the articles.

To query the articles I used the following structure:

```
question_text = 'Who is the CEO of Facebook?'
query = {
  'query': {
    'match': {
      'document_text': 'question_text'
    }
  }
}
```

```
res = es.search(index='articles', body=query, size=5)
```

to return the top 5 search results of the question. Of course, the question is interchangeable. The user would have to alter the “question\_text =” line to enter the new query question.

Results:

For some reason my code would not get all the correct results despite writing my code as outlined in Elasticsearch’s documentation. Running my code would show that it performs well for CEO’s but not so well for the percentages questions.