**Title**: Elasticsearch PA
**Author**: Daniel Hefel
**Description**: This project implements Elasticsearch to search over a corpus. There are options to search using fasttext, sbert, custom analyzer, or the standard bm25 default
**Dependencies**: The following are required: zmq, elasticsearch, elasticsearch-dsl, flask, numpy, sentence-transformers

**Build Instructions**: To build, run load_es_index.py with the proper parameters of what corpus to intake and the index name. Also choose to comment out either the custom or standard analyzer. To use elastic search, make sure to start the server by navigating to the elasticsearch folder and running
$ ./bin/elasticsearch

To use sbert or fasttext, start the encoder for these. See the scripts.sh file for what to run.

**Run Instructions**: Run hw5.py and open in a browser after building the index and starting sbert, fasttext, and elasticsearch.

**Testing**: Various terms were tested to ensure that results were returning relevant documents. Terms included titles, descriptions, and narrations of different TREC topics. different search types including matching on title, content, ID number, etc were also tested. Edge cases like terms not in the corpus were also used.

Evaluate.py results:

| Topic 321 | Title | Description | Narrative |
|---|---|---|---|
| BM25 + default | 0.812653045383133 | 0.6844642384164166 | 0.6263632062847393 |
| BM25 + custom | 0.6311881909557763 | 0.7770219357857179 | 0.5842377548499134 |
| Fasttext | 0.6871475159848289 | 0.634882046762938 | 0.5295086944048898 |
| SBERT | 0.713415424360839 | 0.6509074618618625 | 0.6901038027433183 |

| Topic 336 | Title | Description | Narrative |
|---|---|---|---|
| BM25 + default | 0.8226581712710413 | 0.35763246749526395 | 0.4304608796426586 |
| BM25 + custom | 0.8445376759218184 | 0.43067655807339306 | 0.36830651802596664 |
| Fasttext | 0.4709309686778958 | 0.40903342763899403 | 0.37039702239779687 |
| SBERT | 0.6797356296976694 | 0.41616222859754975 | 0.3941776328249064 |

| Topic 341 | Title | Description | Narrative |
|---|---|---|---|
| BM25 + default | 0.7616615178017226 | 0.5926257741885227 | 0.8960044372945338 |
| BM25 + custom | 0.806453410916153 | 0.8163180703422163 | 0.9292439937028296 |
| Fasttext | 0.805321446963456 | 0.6192345165088121 | 0.7520325050094444 |
| SBERT | 0.7802548138621055 | 0.6878859381126013 | 0.7628588645473199 |

| Topic 347 | Title | Description | Narrative |
|---|---|---|---|

| | | | |
|---|---|---|---|
| BM25 + default | 0.34975904391718976 | 0.46803918329182936 | 0.3229311738182005 |
| BM25 + custom | 0.40815580399525303 | 0.432356564539696 | 0.5027959742923621 |
| Fasttext | 0.3973743676223578 | 0.26536829375182097 | 0.2816707592267034 |
| SBERT | 0.3394000146194867 | 0.295852030825877 | 0.3642313813719837 |

| Topic 350 | Title | Description | Narrative |
|---|---|---|---|
| BM25 + default | 0.0 | 0.0 | 0.3562071871080222 |
| BM25 + custom | 0.0 | 0.0 | 0.6309297535714575 |
| Fasttext | 0.0 | 0.0 | 0.2890648263178879 |
| SBERT | 0.0 | 0.0 | 0.6309297535714575 |