Rajesh Sakhamuru
12/16/2020
CS 6200: Information Retrieval
Final Project Submission

README
Compiling and Running code:
Developed on Ubuntu 18.04 LTS

Project Dependencies:
- Python – 3.8.5
- Pandas – 1.1.0
- urllib3 – 1.25.10
- Flask – 1.1.2
- Requests – 2.24.0
- Elasticsearch – 7.10.1 (Along with compatible Elasticsearch server)

Project File Structure:
The file structure is very straightforward, all code is on a single “main.py” python file located in the 'src' directory. The ‘full_news_documents.csv’ file with all of the news article data is too large to store in a GitHub repository because it is larger than 100MB. So it is located at this link:
https://drive.google.com/file/d/1FtGrje9BA2zxAK3wV6YtANP3xF6x0N47/view?usp=sharing
Please download it and place it in the appropriate file location.

Running Project:

- Ensure dependencies are installed.
- Ensure file structure matches above image.
- Navigate to the 'src' folder containing ‘main.py’ in terminal/console
- Execute the command ‘python3 main.py’
- The program will ask for an administrator password in order to start the Elasticsearch server and again at the end when closing the program to stop the server.
    - Each time you have 20 seconds maximum to input your password.
- The first time the server is run, all documents are loaded into the Elasticsearch index which could take up to 5 minutes depending on the computer.
- The user interface can be accessed at http://127.0.0.1:5000/ while both the Flask and Elasticsearch server is running.