**Mini Search Engine**

**Overview**

This is a from-scratch implementation of the search engine for a collection of 418 documents on environmental news from Kaggle.  
Link to dataset: <https://www.kaggle.com/amritvirsinghx/environmental-news-nlp-dataset>

The retrieved results are then compared using a standard realtime open source search engine like ‘ElasticSearch’.

This has been built as part of the course curriculum of Algorithms in Information Retrieval (UE17CS412).

Project Contributors:

-Sriya Samptur

-Bhavna Arora

-Lavanya L

-Abhishek M

Directory Structure:

Queries:

Queries1.json: Standard Queries

Queries2.json: Wildcard Queries

Queries3.json: Boolean Queries

Queries4.json: Combination Queries

1. Download and extract dataset from kaggle into as following directory structure in your current working directory  
 ./content/content/TelevisionNews/

2. Run all the cells in the AIR\_Assignment\_Team39.ipynb file after changing the path location

3. In the last cell of the AIR\_Assignment\_Team39.ipynb file, the user can pass the query as the parameter to the query\_search function. This function displays the top 10 results retrieved by the search engine.

4.

The folder ```Small``` contains a part of the dataset that is split and added to to test the model for 25 files.

```elsearch\_output<n>.json``` and ```mysearch\_output.json<n>.json``` {where n=1,2,3,4}

Contains results of the ```queries<n>.json``` after running on respective search engine

The ```Large``` folder contains the complete dataset.

```elsearch\_output<n>.json``` and ```mysearch\_output.json<n>.json``` {where n=1,2,3,4}

Contains results of the ```queries<n>.json``` after running on respective search engine

DUDE DONT WRITE THE TOY EXAMPLE PART, WRITE HOW SHOULD THEY USE ELASTIC SERACH NO. ONLY HOW TO USE THE TOOL?

CHANGE 2ND POINT 1ST IS PROPER

Now ok ???