Execution instructions:

After extracting the folder, you can run the **main.py** directly.

Note that 'easygui' library must be present in your system.

After running the file, please enter the query in the textbox. Later another textbox appears asking to enter the number of results needed for you.

Please enter a valid integer in this checkbox as no validations were written in it.

Next it displays the list of the results for the given query.

Explanation of the files present in the project have been mentioned below

Spider.ipynb

The spider/crawler is written in **spider.ipynb** jupyter notebook for its step by step execution process. Crawling of webpages, extracting the data from the web and downloading that data in done in this notebook. I have mentioned the number of pages as 4000, therefore it will take close to 3 hours to download the content into the local folder.

Tfidf.ipynb

In this notebook, the data has been preprocessed and the term frequency values and inverse document frequency values have been calculated and stored in the pickle files. This process will take close to 10 minutes for storing the values in the local directory.

Pagerank.ipynb

The pagerank formula is written in this notebook and all the values of the pages are then stored in the local file as mentioned the notebook.

Main.ipynb

This is the main file which consists of user interface and the search engine results.

NOTE- The main.ipynb and main.py consists of same code. In case the jupyter notebook is difficult to run, you can directly run the main.py file as mentioned in the beginning of the document.

IN CASE OF ANY FAILURE OR MISSING OF ANY FILES, PLEASE LET ME KNOW SO THAT I CAN LOOK AT IT AND RESOLVE IT