

Steps to run the Program:

Note: some libraries and packages might needed to be installed

How to start searching

1.run search.py

How to create crawl

1.run run_multithreaded.py

How to preprocess

1.run m_invert.py

2.run run_page_rank.py

Description of the files :

1. Crawled.txt ==> crawled pages
2. Queue.txt ==> pages to be crawled
3. code_from_url_dict.pickle , code_from_url_dict.pickle ==> mapping of url to docID
4. Dictionary.pickle ==> the dictionary
5. code_length_dict.pickle ==> length of all the documents
6. page_rank.pickle ==> page rank score
7. posting_list.pickle ==> the posting list
8. Ryerson/pages ==> the webpages contents