The application has three major files along with two minor files.

The files indexing.py and queryOperation.py contains the source code for the application. The file searchengine.py is written to call both the files. The files stopwords.dat and porter stemmer are the minor files and are used by the source code files.

After running the file searchengine.py two files, namely indexFile and titleIndexFile are created. indexFile is the file containg the index for words in the input documents. titleIndexFile contains the names of various files given as input and their id’s.

Index is created as follows

Term|docid:postings;docid:postings|tf|idf.

The retrieval is based on vector space model for information retrieval.

In indexing.py the lines from a file are given to a function which tokenises the terms and stores it in a list. Then another function stores this data in a file.

In queryOperation.py a query is taken as input. Words from it are tokenised and normalised. Then documents containing the terms are retrieved. The documents are ranked based on their tf-idf score.

The names of these documents are outputted in the descending order of their scores.

Major data-structures used are dictionary, defaultdict(list) from collections library, list, etc.