

README

Run the files in the following order:

- shingling.py
- vectorize.py
- minimumHashing.py
- lsh.py
- frun.py

or running frun.py would be enough as all the functions are called in this file.

After running frun.py in terminal a statement saying “enter the query:” will be printed and underneath which user need to enter the query that he wants to search for. Basically, the query should be picked up from any line of the dataset with enough no. of characters in it. Then you need to input shingle size after a statement appears demanding it. Then input the band size and similarity threshold in order once the statements appear for it. The results of tuples containing 0 in them are displayed, As query is stored in the 1st line of the dataset temporarily.

All the documents are stored in the dataset text file.

TEAM MEMBERS:

- 1.CHINTHAKINDI SIDHARTHA RAMAN (2018A7PS0283H)
- 2.POOJARI VENKATA SHREESH (2018A7PS0301H)
- 3.BEJANKI ADITYA (2018A7PS0282H)
- 4.DINESH REDDY NANDYALA (2018A7PS0431H)