

Name: KOUSTAV DAS

Entry:2015CSB1017

### Preprocessing Steps

1. We have parsed every word from all the documents.
2. We have removed if the word is a stop word. The list of stop words have been attached in the folder.
3. For each and every word we perform a stemming operation.
4. Then we have applied the higher and the lower thresholding. The values for thresholding has been taken from the user.
5. The keys recieved have been saved in 'dictionary.txt'.

### Query

1. The TDM has been generated.
2. While the TDM was being generated at that same time IDF was also generated.
3. The score for every document have been generated from TF-IDF.

### Regarding the functions

1.function [ diction, stop ] = getRepTerms( name )

This takes in the path of the directory and returns a map object for dictionary and stopwords

2. retRelDocs( str, name )

This takes in the input query string in the form of 'str'. 'name' contains the path of the directory.

### Example

1. retRelDocs('motion','corpus');

### Note

1. Please find the commenting in the code to understand the things in details.

### References

1. The stemming algorithm is being taken from <https://tartarus.org/martin/PorterStemmer/matlab.txt>