

Readme File

Methodology-

Preprocessing the data

Explanation of Preprocessing the Data-

In preprocessing of data, the steps include are-

- converting the data in lowercase
 - removing the stopwords
 - removing punctuation, apostrophes and other symbols
 - removing the one letter words
 - replacing numeric values with their corresponding text values
 - performing stemming on words in the data
1. Creating posting lists for the words in the data
 2. Converting Dataframe postings to a dictionary format
 3. Creating a set for all 467 documents