1. Building index file
   1. Fetch all files in the folder ./data/
   2. For all files do
      1. Fetch the tokens (all the individual words)
      2. Remove stop words (This is done based on a predefined stop words list which is included in the zip (filename : stopWords.txt))
      3. Perform stemming on tokens (performed with the help of library stemming.)
      4. Convert all tokens to lower case
      5. Map the tokens to the current document. If the token already has other documents associated with it, add the current document to the list.
      6. Write tokens to file for each document (filename = document name)
   3. Write all the indexes to file maps.txt
2. Processing input query
   1. Fetch all the words/tokens in the search query
   2. Perform stemming on the tokens
   3. Convert the tokens to lower case
3. Perform the Search operation
   1. Based on the tokens generated from query, fetch all the documents list for all the tokens
   2. If an ‘and’ operation is provided, intersect the previous and next document list
   3. If an ‘or’ operation is provided, union the previous and next document list
4. Assumptions
   1. Parenthesis are not considered for search query
5. Minimum requirements to run
   1. Python should be installed
   2. Packages to be present or installed
      1. os
      2. sys
      3. stemming – pip install stemming
      4. docx – pip install python-docx (python-docx.readthedocs.io/)
6. Basic setup
   1. Documents must be placed in folder “data”
   2. stopWords.txt file should be placed in the executing folder
   3. Output will be written to the folder “output”
   4. Output files written are:
      1. Tokens generated for all documents (6 files, 1 for each document)
      2. Index generated (maps.txt)
      3. Search result history – All search result query and output are saved (result.txt)
7. How to run:
   1. python IR.py “<<search query goes here – in quotes>>”