Below are the description of the java files and their respective output files

* JAVA FILE- DocWordCount.java

This program attaches the name of the file where a particular word is present to that word separated by ‘#####’ delimeter.

OUTPUT FILE: DocWordCount.java

Example: word#####filename count

Arguments to be given: path for folder to stored input file and output file

* JAVA FILE- TermFrequency.java

This program calculates term frequency of each word

OUTPUT FILE-  TermFrequency.java

Example: Hadoop#####file1.txt 1000

Arguments to be given: path for folder to stored input file and output file

* JAVA FILE- TFIDF.java

In this program, first the term frequency is calculated which is given as input further to calculate TDIFD

OUTPUT FILE- TFIDF.java

Example: yellow#####file2.txt 0.30102999566

Arguments to be given: path for folder to stored input file, intermediate file and output file

* JAVA FILE- Search.java

In this program a query is given as input and a list if files containing that word is the output.

OUTPUT FILE: a. query1: “computer science” output file: query1.out  
 b. query2: “data analysis”, output file: query2.out

Example: lcet10.txt 2.4769516218371552

Arguments to be given: path for folder to stored input file (here input file is the output file of TFIDF.java program) and output file