Lucene is the project folder.

Uses following libraries which has been included in the build path and is provided in the "lib" folder:

* lucene-analyzers-common-4.7.2
* lucene-core-4.7.2
* lucene-queryparser-4.7.2
* jsoup-1.9.2

No command line arguments to be passed.

just in case this fails can follow the below steps:

In Eclipse,

(right click on project) -> Build Path -> Configure Build Path and select the mentioned library from lib folder (add external jars) and the JavaSE-1.7

## Task 1:

* Folder for index creation is: “index\”
* Folder to get Corpus is: “Corpus\”
* Write query 1 by 1 and program will write the results into folder “Task1\” with the query.csv format

## Task 2:

* Reads the HW3 unigram inverted index “file.out” data structure and the document length (token count), “tokenPerDoc.out” of each doc
* Reads queries from the file “queries.txt”
* Writes the .csv results in “Task2\” folder with query.csv format

Format:

**query\_id, Q0, doc\_title, rank, CosineSim\_score, system\_name**

system\_name is Cosine\_Relevance

* Writes a .csv file “DocIDmapping” that shows the mapping of document title to docID

Task1 folder has the four tables generated from Task1

Task2 folder has the four tables generated from Task2

**IMPLEMENTATION** file contain the implementation details

**DISCUSSION** file contain the discussion on Top 5 results from each IR system.