READ ME file

* Install python 2.7
* Unzip the contents of folder Main Project.

**PHASE1**

* To execute Phase1, Task1, go to folder, Main Project> Task1
* To run the 4 retrieval models execute following:
  + For BM25, execute file BM25.py
  + For Cosine Similarity, execute file cosinesimilarity.py
  + For tf.idf, execute tf-idf.py
  + For the Lucene code, follow the steps:
* Install Eclipse
* Copy the project folder for task1
* The input to the file is the raw data we obtained form hw3.
* Run the file
* It will prompt for the folder in which the results are to be saved and also the location of corpus
* Next the query is inputted and similarity score is obtained in the console
* Top 100 documents according to the score is obtained. These files are saved in a separate files which are in the folder Hw4 Task4. The names of these files are the query names.
* The indexer used is Lucene is simple analyzer.
* The jars required to run the file are provided in the project folder.
* To execute Phase1, Task2, go to folder, Main Project> Task1
* Run the code Pseudo\_Relevance.py
* To perform Task3(A) of Phase1 go to folder, Main Project> Task1
* Run BM25\_output\_with\_stopping
* For Task3(B), run file BM25\_output\_Stemmed

**PHASE2**

* Run file Pseudo\_Relevance\_With\_stopping.py.
* Go to folder Main Project> Task1
* Run eval.py
* To get evaluation result for different retrieval models, update value of file name at line number 25 of eval.py and choose one file name among the following:
  + BM25\_eval
  + BM25\_stopping\_eval
  + CosineSimilarity\_eval
  + Pseudo\_output\_eval
  + Pseudo\_output\_with\_stopping\_eval
  + Tf-idf\_eval