READ ME file

**TASK1**

* Install Eclipse
* Copy the project folder workspace for task1
* Open the folder HW4>src
* Run the .java file HW4 in src.
* It will prompt for the folder in which the results are to be saved. Give the location of index folder in workspace>HW4>index
* It will next prompt for the parsed corpus location. Give the location of the ParsedCorpus\_temp folder in workspace>HW4>ParsedCorpus\_temp.
* Next there will be the prompt for the query. Type the first query.
* Top 100 documents according to the score is obtained. These files are saved.
* Again go into the code and change value of j=2
* Again run the .java file with j=2 and repeat the process with the second query.
* Top 100 documents for this query will also be saved.
* Similarly do the same for all the queries.
* Results for the 4 queries are saved in Q1, Q2, Q3 and Q4.

**TASK2**

* Install python 2.7
* Copy the entire project folder HW4
* Run the main file, Task2
* The inputs for this task are parsed corpus and the queries from query.
* We read each file from the “Parsed\_temp” and their weights are calculated using tf-idf formula. Next the queries are taken one by one from the query file and weights are calculated for each query term also.
* Next we use the cosine similarity formula to calculate the score for each document.
* Then the documents are sorted according to the ranks obtained and top 100 results are sorted out and stored in text files (Name of text file is in the following format “top100forquery[i] i=0,1, 2,”).
* The format of entries in these text files are as follows:
* *query\_id Q0 doc\_id rank CosineSim\_score system\_name*

e.g.: 1 Q0 Globalwarming 1 0.176592821601 Sugandha