COMS6111_Proj1_Group36
Prateek Gupta
Dhruvkumar Motwani[dgm2138]

List of files:
1. UpdateQuery.py
2. Run.py
3. Stopwords.txt
4. Transcript.pdf
5. README.pdf

Run instructions:

Following command can be used to run the program on clic machines:

Python run.py <bing_key> <precision> <query>

Internal design:
The program is mainly divided into two components run.py and UpdateQuery.py.

Run.py consists of the necessary code to run the program while UpdateQuery.py is used to refine the results obtained using tf-idf and rocchio's algorithm.

Run.py calls the bing api for each query and displays the top 10 results.

Stop.txt is consists of all the stop words used in UpdateQuery.py . It's used mainly to reduce the effect of common words which don't help in feedback loop of the algorithm.

UpdateQuery.py is used to refine the results obtained by calling bing API. Different values of alpha, beta and gamma were tried to refine the Rocchio's algorithm and suitable value is displayed in the program. The code in this program does various functions such as calculating the tf-idf score of the vocabulary, creating vocabulary, run rocchio's algorithm as well as return the tuples ranking calculated by the algorithm.

Bing Search Account Key : KVhu7dSuBucM2oUJ00/qMtKNKqtDNNmSwv8eZ6dk++g