Gowtham Kommineni

[gkommine@uncc.edu](mailto:gkommine@uncc.edu)

**Execution Instructions:**

Environment: Cloudera

Create input folder and move all the input files into it

1. DocWordCount:

a. Give input and output paths as arguments.

b. hadoop jar <jarFileName> DocWordCount <input> <output>

2. TermFrequency

a. Give input and output paths as arguments

b. hadoop jar <jarFileName> DocWordCount <input> <output>

3. TFIDF job

a. Give input and output paths as arguments

b. h*adoop jar <*jarFileName*> TFIDF <input> <output>*

*Note: Since we use chaining of TFIDF with TermFrequency an output file with the name “intermediate”. So, please delete this folder every time you run the jar.*

4. Search

Run the TFIDF job before running the Search job.

Provide the output folder of TFIDF as input, output path and a search query as arguments.

h*adoop jar <*jarFileName*> Search <tfidf\_output\_patht> <output> <search\_query>*

5. Rank

Run the Search algorithm before Rank.

hadoop jar <jarFileName> Rank <Search\_ouput\_path> <output>