

Kristina Wong, 76513468
Haoming Li, 20426226
Shengjie Xu, 10616769
Yirui Jiang, 64137163

1. Number of documents : 44536
2. Number of [unique] words(NOT including stop words) : 164569
3. sample index:
termID -> docID, tf-idf
1 = {1=0.0095, 2=0.0041, 4=0.0037, 5=0.0021 ...}
2 = {1=0.0024, 2=0.001, 4=0.0014, 5=3.0E-4 ...}
...
4. id_doc_index.txt (contains docID->doc) : 2,795 KB
id_term_index.txt (contains termID->term) : 2,858 KB
Index.txt (contains termID -> docID, tf-idf) : 53,925 KB
5. the time taken to create the index : 28.158 sec