Alper Canberk Balcı

Cmpe493 Project1 Report

1. Describe the steps you have performed for data preprocessing.  
   1- The main.py program takes the data and reads it in the main function  
   2- It strips the useless line at the top.  
   3- It finds the id of the document with NEWID.

4- It takes the title and body, and preprocess them with case folding, punctuation removal, and stop word removal. Then, it makes them lists of words  
5- It fills the inverted index

1. Describe the data structures (hash, b-tree, linked list etc.) that you used for representing the inverted index (i.e., the dictionary and the postings lists).  
   I used simple dictionary and postings lists, with python’s list.
2. Provide a screenshot of running the indexing module of your system.  
   Text

   Description automatically generated  
   it continues to read the 21 files, I made these prints to console to check whether it runs smoothly.  
   Text

   Description automatically generated  
   It provides an output to the console, in addition to creating the json file of inverted index.
3. Provide four screenshots of running your system for each of the four types of queries.  
   Graphical user interface

   Description automatically generated with medium confidence  
   A screenshot of a computer

   Description automatically generated with medium confidence  
   A screenshot of a computer

   Description automatically generated with medium confidence  
   Graphical user interface, text

   Description automatically generated  
   