**CPS842 Assignment 1 Report**

**Instructions to run the program:**

1. Make sure you have Python 3 installed and in the directory of the assignment folder.
2. Compile and run the file “test.py”, using either “terminal” or “command prompt” if on Windows
3. Enter a word to search or enter “HELP” to display options available.
4. Options to toggle for stemming and stop words are available in the “HELP” section which has the capabilities to turn on and off both options
5. The time taken for search along with relevant documents will be displayed in the terminal/command line.
6. Once the results for the search have appeared, you may enter a new word or exit.

**Algorithms and Data structures:**

The data structure that we used was a python dictionary for creating the dictionary of terms, as well as for the posting list. Since a Python Dictionary supports a “key-value” pairing it makes it a great fit for the solution to the search program. Furthermore, the python dictionary can also be easily exported to a JSON file, which can be easily read by a human and imported into a program. The time it takes to parse the data, search through it is efficient and finally insert new data into the dictionary which is the main reason for the choice. To implement the stemming feature, we used the Porter Stemming algorithm provided by the professor, which helps the program get rid of endings of the words to make the search more efficient and accurate. Furthermore, the program also has the feature to remove stop words from the search. The implementation of this was mainly checking if the toggle was turned on/off and based on that it runs a function which checks for the word in the posting list JSON file. Similarly, the same goes for the stemming words, where the search checks for the word as well as the toggle, and then check through the posting list file to find the word. For calculating the total time, the time library was imported which runs the time function to keep track of time for each search, and the search time is stored inside an array that is used to calculate the average time that is printed before the application is closed.