If you want to run the program on windows system, you should compile all the code(because we can’t upload exe files) . The method to compile codes is given in paragraph 2.

If you want to modify code, you should compile four code files in all: main.cpp, porter2\_stemmer.cpp, insert\_index.cpp, search\_word.cpp. What's more, you can directly enter "make.bat" in your terminal (in Linux you should enter "make.sh" in your terminal) to complete all compilation processes.

In main.cpp, there are three output function: print\_posting\_list(to print all the index), print\_highfrequency\_word(you can enter a number, and the program will print all the word whose number of occurrences is bigger than your number) and search\_word(to realize the search function).

In the folder "txt\_source", there are the complete works of Shakespeare. Script1 to script17 are comedy texts; script18 to script 27 are history texts; script 28 to script 37 are tragedy texts; script38 to script 42 are poetry texts. Unfortunately, due to too many files, script38 is empty.

In the code:

Word stemmer function is from：

<https://github.com/smassung/porter2_stemmer>

HASH function is from：

<https://blog.csdn.net/qq_40342400/article/details/127232662?ops_request_misc=&request_id=&biz_id=102&utm_term=%E5%AD%97%E7%AC%A6%E4%B8%B2%E6%9F%A5%E6%89%BE%20%E5%93%88%E5%B8%8C%E5%87%BD%E6%95%B0&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduweb~default-1-127232662.142>^v100^pc\_search\_result\_base1&spm=1018.2226.3001.4187