EE488A

Advanced Programming Techniques for Electrical Engineering, Fall 2019

HW1

Due date: 10/04/2019 (11:59:59pm)

Overview: In this homework, you will implement simple text search engine for a small number of files in specific folder. The implementation may contain two main parts:

- *Bootstrapping:* Your program should make inverted index of words in all documents in the folder. This should be done when the program is started, and the program should store the generated indices in the memory.
- *Searching:* Your program should be able to search relevant documents and lines by the queried word.

Bootstrapping:

1. Your program should get a argument that specify the running folder when it started: **[your_program_name] [absolute_path_to__target_folder]**
2. Your program should parse all files in the folder into inverted index and store it in its own memory. (refer to "lecture 5" on KLMS) *Because we will only check text files, you don't have to consider binary files.*

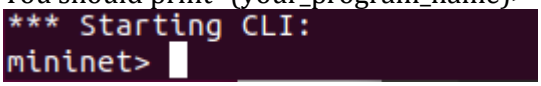
Notes:

- You can only use pointer arithmetics when parsing. In other words, you should not use string library.
- Generated inverted index should store line number with the file name.
- When parsing, you should deal with only alphabetic words. All white spaces and special characters should be considered as delimiter of words.

example)

text	**parsed words**
abc bcd	**[abc, bcd]**
abc"b cd"	**[abc, b, cd]**
abc[bcd]	**[abc, bcd]**
abc "bcd"	**[abc, bcd]**

Searching:

1. After Bootstrapping, you should print some messages that notify end of bootstrapping.
2. You should print "(your_program_name)>" to identify CLI like mininet.
   

3. In this part, you will implement searching word using inverted index which was created in Bootstrapping part. If you enter word for searching, documents(file names) and line numbers ,which describe location of this word, should be printed. Besides, your input/output form should follow 4) and 5) perfectly. If you do not follow below descriptions about input/output form, you will get 0 points.
4. Your input for searching world should follow below form exactly.
   search [word]
5. Your output for corresponded input should follow below form exactly.
   [filename1]: line #[line number1]
   [filename1]: line #[line number2]
   [filename2]: line #[line number3]
   [filename3]: line #[line number4]

Submission instructions: Use KAIST KLMS to submit your homeworks. Your submission should be one gzipped tar file whose name is YourStudentID_hw1.tar.gz. For example, if your student ID is 20191234, and it is for homework #1, please name the file as 20191234_hw1.tar.gz. If you do not follow this zip file name, you will lose some points.

Your zip file should contain three things;
1. **One PDF** file for document which explains your codes(hw1.pdf). This PDF file should show explanations for functions in code detailedly.
2. **C** files which satisfy detailed requirements for this homework.
3. **Readme** file should explain compile and execution procedures of your program detailedly for using other people. If, despite that using procedures in readme, your program is not executable, you will lose some points.
*Do not include Korean letters in any file name or directory name when you submit.

Important note: Files not following the above format will be heavily punished! (You will get 0p)

Test environment : We will test your codes in ubuntu 18.04 build-essential. Thus, using any external libraries is not allowed. If you use external libraries, you might lose most of the full credits.

Late submission policy : If you submit your homework after due, you will lose 20% for each late day(i.e. 80% of the full credit up to 24 hours late, 20% of the full credit up to 96 hours late)

Plagiarism
Discussions with other people are permitted and encouraged. However, when the time comes to write your solution, such discussions (except with course staff members) are no longer

appropriate: you must write down your own solutions independently. If you received any help, you must specify on the top of your written homework any individuals from whom you received help, and the nature of the help that you received. Do not, under any circumstances, copy another person's solution. We check all submissions for plagiarism and take any violations seriously.
*Plagiarism will get severely penalized If detected, 0 points for all assignments (both providers and consumers)