

# CENG 313 - Homework 3

**Due date:** 25.04.2021 23:55

In this homework, you will write a code which counts the number of times a given string is used in txt files. You will be implementing threads to speed up the process and report the results.

As a first step, you should download at least 10 freely available books in .txt format from the Gutenberg project (<https://www.gutenberg.org/>). Rename the books you have downloaded to “book1.txt”, “book2.txt”, etc. in order to make it easier for testing (since you will all be downloading different books).

You are provided with the English stop-words in a file called “englishstopwords.txt”. Your task is to search each stop-word in that file among the books you have downloaded. You should be running your program with different number of threads (1,2,4, etc...) and observe the performance differences. You can use the library we used in lab to calculate the execution time.

You are free to implement your experimental setup the way you want but do not forget that it can effect your grade (the way you use your threads, the way you divide the task and the algorithm you use to search for the stop-words).

You are expected to provide a small report in .pdf format, about your findings. In that report, please provide and discuss your results: What is the result? Why did you end up with those results? What are the advantages and disadvantages of the method you used? What are the limitations? Feel free to comment on anything you observed during the experiment and their possible reasons.

In order to show that your program works properly and to inform the user, print some examples on the screen. Print the total execution time and some selected few words' frequencies. An example output should be like this:

**Example run:**

*frequency calculation started*

*frequency calculation ended*

*frequency of the word “again” in all books:*

*468 - 158 - 258 - 657 - 3125 - 31 - 356 - 326 - 624 - 6147 - 348*

*frequency of the word “during” in all books:*

*391 - 734 - 761 - 2148 - 3125 - 451 - 356 - 4238 - 45- 183 - 415*

*total execution time: 5.461s*

**IMPORTANT NOTES**

- You can also use other sources than Gutenberg, but be wary that they need to be long text sources.
- You can place your txt files in the same folder of your c code while testing in order not to deal with paths. I will be compiling them in a similar fashion.

**SUBMISSION RULES**

- Your code will be compiled using GCC on Linux. Indicate C version (e.g. C99) you used in the first line of your code as a comment. It can be written as follows:  

```
// This code compiled with C99
```
- You should submit your assignments through CMS until the due date. Your homework should be named as CENG322\_hw1\_studentID.zip  
e.g. “CENG322\_hw1\_123456789.zip” in the following structure.
- You can create 1 or more source code files but do not forget to include the necessary header file for linking them. If you send 2 or more .c files with no .h file (header file that does not work) the assignment will be evaluated as 0. So, make sure that your source code and header files work before submitting.
- Cheating will not be tolerated and will be evaluated as 0.
- Late submissions will not be allowed.
- Possible Grade reductions
  - Lack of Comment usage
  - Missing Controls
  - Unused/Dead Codes
  - Naming Conventions