

HW 8: Multithreaded Word Count Program

Instructor: Prof. Seokin Hong (seokin@knu.ac.kr)

Assigned: November 25, 2019

Due: 11:59pm December 2, 2019

The goal of this assignment is to extend the **“Word Count program (twordcount)”** so that it can count the number of words in a single file in parallel. The extended word count program will accept the number of threads on the command line and create that amount of threads for the parallel execution. In the `count_words()` function of the original `“twordcount3.c”`, each character is read from the file at a time, which leads to the significant performance degradation of the multithreaded program. Therefore, you need to extend the program so that it copies the contents of the file to a memory buffer and then read each character from the memory buffer.

In summary, extend the **“Word Count program (twordcount)”** to

1. create the arbitrary number of threads to count the number of words in a single file in parallel (the number of threads will be 1, 2, 4, and 8)
2. Copy contents of a file to a memory buffer and read each character from the memory buffer
3. Print the elapsed time on the screen at the end of the program execution

- Download a text file (200MB) with the following command and use the file when you run the extended word count program. (Please let me know if you have trouble in downloading this file)

```
$ wget http://155.230.28.247/wiki_200MB.xml ./
```

- A **skeleton C code** is available in the LMS. You can use any parts of the code for this assignment.

[Run]

```
$ ./hw8_2014123123 <the number of threads> ./wikipedia_20MB.txt
```

[Expected output]

```
seokin@compasslab1 $ ./hw8_2014123123 1 wikipedia_2GB.txt
[THREAD] the number of words : 326782021
[MAIN] total word count: 326782021
[MAIN] Elapsed: 11.833788 seconds
```

```
seokin@compasslab1 $ ./hw8_2014123123 2 wikipedia_2GB.txt
[THREAD] the number of words : 163442305
[THREAD] the number of words : 163339716
[MAIN] total word count: 326782021
[MAIN] Elapsed: 10.615140 seconds
```

```
seokin@compasslab1 $ ./hw8_2014123123 4 wikipedia_2GB.txt
[THREAD] the number of words : 81247461
[THREAD] the number of words : 81183647
[THREAD] the number of words : 82194844
[THREAD] the number of words : 82156069
[MAIN] total word count: 326782021
[MAIN] Elapsed: 7.517672 seconds
```

```
seokin@compasslab1 $ ./hw8_2014123123 8 wikipedia_2GB.txt
[THREAD] the number of words : 40599619
[THREAD] the number of words : 40647842
[THREAD] the number of words : 41206742
[THREAD] the number of words : 40988102
[THREAD] the number of words : 41494881
[THREAD] the number of words : 40661188
[THREAD] the number of words : 40597851
[THREAD] the number of words : 40585796
[MAIN] total word count: 326782021
[MAIN] Elapsed: 5.231806 seconds
```

Late Day Policy

All homeworks are due at 11:59pm on the due date. A grading penalty will be applied to late assignments. Any assignment turned in late will be penalized 25% per late day.

Plagiarism

No plagiarism will be tolerated. If the assignment is to be worked on your own, please respect it. If the instructor determines that there are substantial similarities exceeding the likelihood of such an event, he will call the two (or more) students to explain them and possibly to take an immediate test (or assignment, at the discretion of the instructor) to determine the student's abilities related to the offending work.