Abbas Mahdavi ID: 918345420
Github: AbbasMahdavi021 CSC415 -03 Operating Systems

Assignment 4 – Word Blast

Description:
This assignment is to write a C program that reads a text file "WarAndPeace.txt". The program counts and tally each of the words that are 6 or more characters long. This program only uses Linux file functions such as open, close, read, lseek, and pread etc.
The program splits the file into equally sized blocks, depending on the number of provided threads. And so, the time it takes to complete the program will vary, depending on the number of threads used.

Approach / What I Did:
For this assignment, I first created two structs, word, and ThreadArgs, to help me initialize the word and number of frequencies, as well as the variables needed for thread creations.
I created two helper functions as well. One will parse the given word, and store in a word list, and increment the list as well as the frequency of the word. The other will take advantage of thread buffer to allocate the file, and use the wordProcess function to tokenize the words.

The main will use these helper functions to open the file, divide the threads in different blocks, and allocate the file and words in a unique structure. The built in functions strtok_r, pthread_mutex_lock, pthread_mutex_unlock, pthread_mutex_init, pthread_create and pthread_join, are used. Each thread process gets their own unique id and exatly one process, where no thread can exist outside a process. An argument list is used to store the ThreadArgs variables, depending on the threadCount. All threads are joined back together in the main thread when the word counting is done. But before this, the top 10 words are displayed using a nested for loop, and are printed to the console. The number of threads used changes the for loop. At the end, I freed the elements in the list, and closed the file.

Issues and Resolutions:

- One of the biggest problems I had when programming this assignment was creating a helper function for file processing, and threat creation. I decided to use pread, which made thinks complicated for me.

  I had trouble separating a portion of the file to be used in a buffer, and the allocation steps were confusing. I managed to better understand the process from the website below, which cleared things up for: Source: https://linux.die.net/man/2/pread
  I was still having issue with the file descriptor (id), and properly using the buffer and block size. A classmate pointed out my issue to me, and I managed to now do it properly.
  They also noted that I should null terminate the buffer:
  threadBuffer[line->blockSize] = '\0';

- Another issue I was having was getting NULL errors, when parsing the words into the word list. I was unsure of why this was occurring.

I used an if statement where if word, or the word in the list is null, it would continue. And this fixed my issue, as a way around the problem.

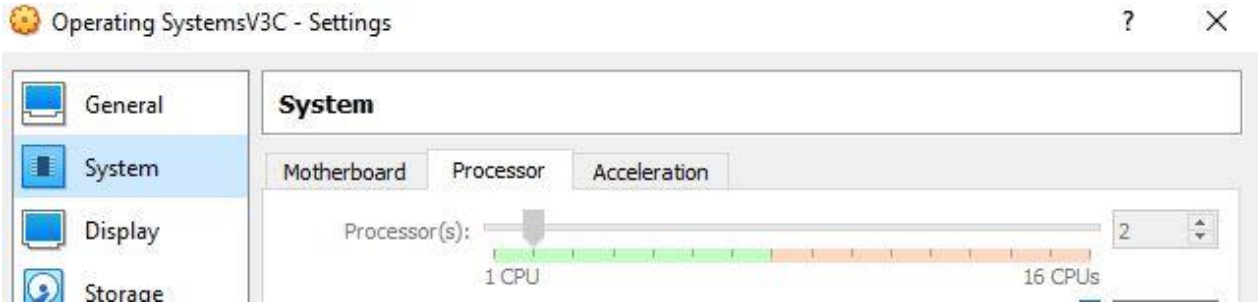
Analysis:

The output images below are the runs of the program with RUNOPTIONS of different threads. My runtime is fairly similar to the provide output. Although there is minor difference. The screen shots show the top 10 words, that are longer than 6 characters, in the text file WarAndPeace. Since the file is being split into different chunks, some words might be split into half, and so the count of certain words might differ by 1 or more. The provided timer helps us understand how the usage of different threads help the program, as the word is divided into different parts, where different threads can take care of different part of the process.

When the file is read using a single thread, the OS does not need to call the file in different chucks, as there is only 1 thread being used. Because of this the runtime is longer, than any other number of threads.

If we simple write make run in the cmd line, the program will run with 2 threads. This is because the Virtual Machine is limited to 2 threads by default. Running the program with RUNOPTIONS of 2 threads will results in the same output, and roughly same runtime. This run time is almost double as fast (1.28 seconds faster) as the runtime with 1 thread.

Running the program with 4 threads, gives a slightly better run time, but 8 threads fluctuate around the same runtime.
A picture below shows that the VM only having 2 cpu's, when looking at the VM system settings:

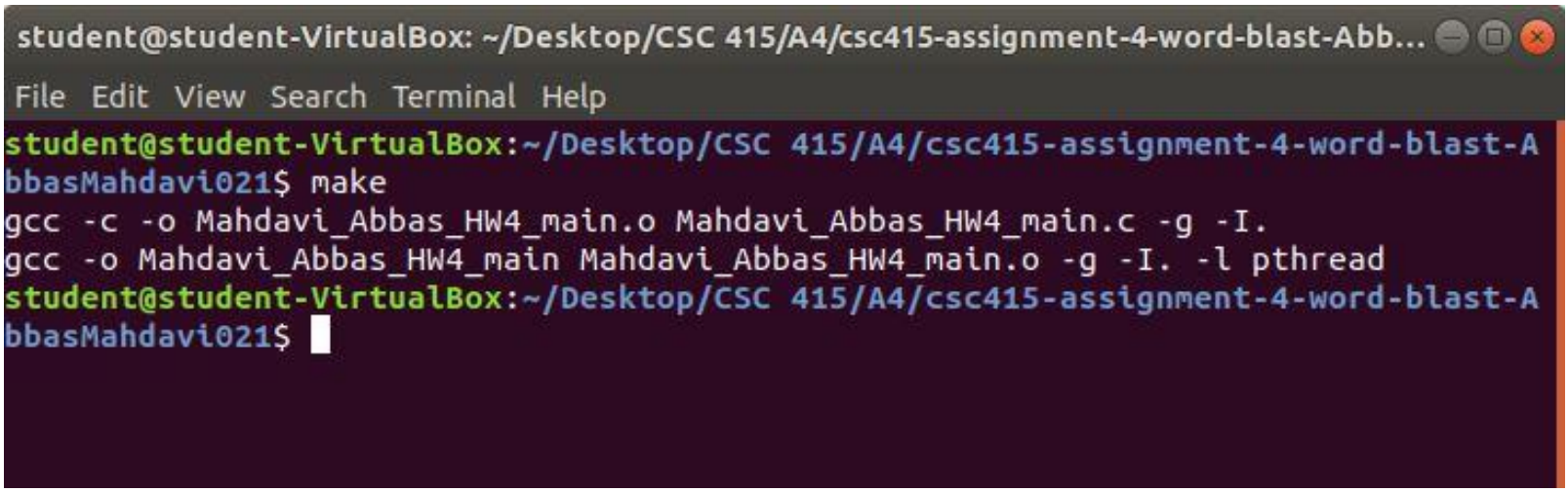


This shows us that because the VM is limited to 2 threads/cpus, the amount of threads used higher than 2 does not make a difference. We see this when run time of 4 threads is similar to 2 threads, where as 2 threads, almost doubles the 1 thread usage.

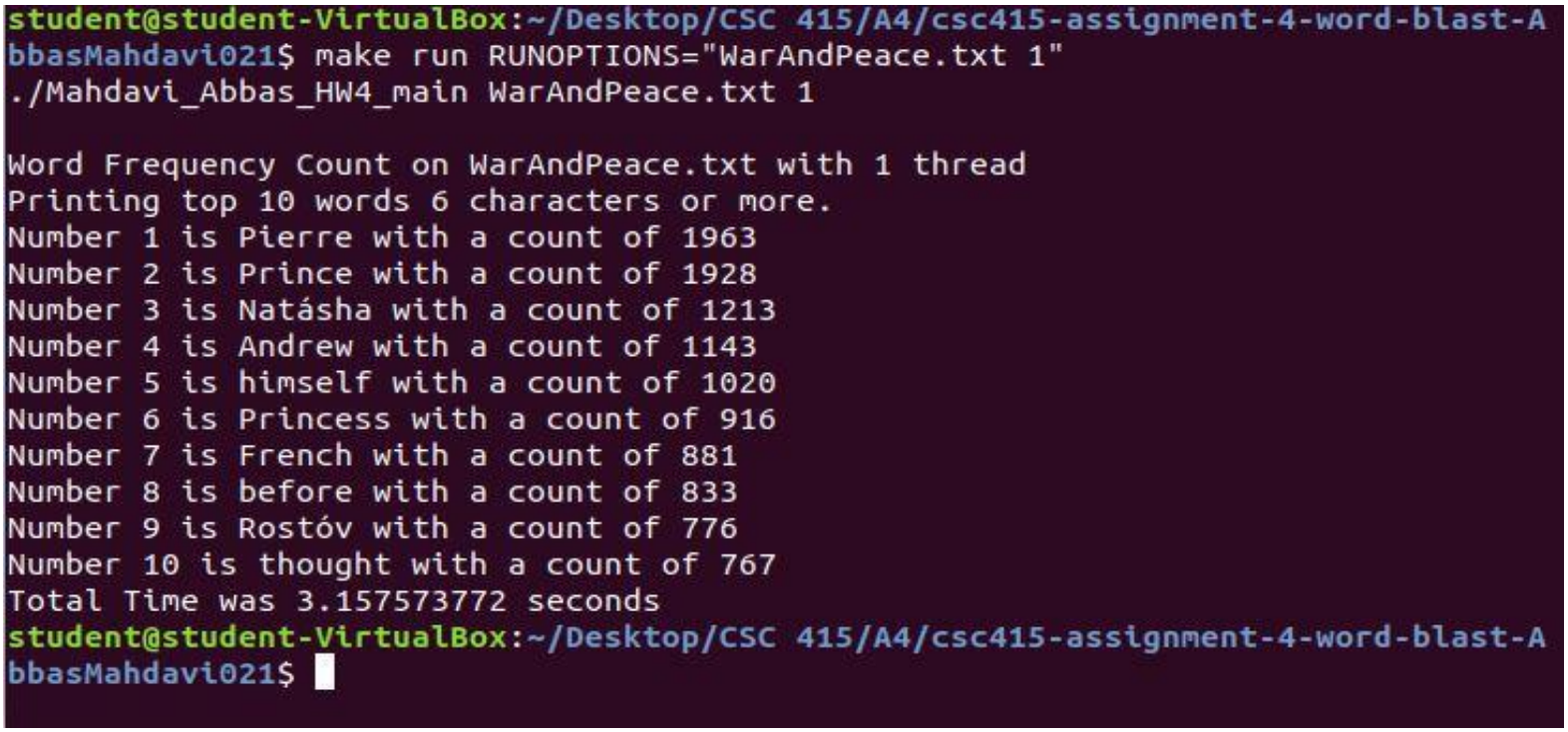
Abbas Mahdavi
Github: AbbasMahdavi021

ID: 918345420

Screen shot of compilation:



Screen shot(s) of the execution of the program:



```
student@student-VirtualBox:~/Desktop/CSC 415/A4/csc415-assignment-4-word-blast-A
bbasMahdavi021$ make run RUNOPTIONS="WarAndPeace.txt 2"
./Mahdavi_Abbas_HW4_main WarAndPeace.txt 2

Word Frequency Count on WarAndPeace.txt with 2 threads
Printing top 10 words 6 characters or more.
Number 1 is Pierre with a count of 1963
Number 2 is Prince with a count of 1928
Number 3 is Natásha with a count of 1213
Number 4 is Andrew with a count of 1143
Number 5 is himself with a count of 1020
Number 6 is Princess with a count of 916
Number 7 is French with a count of 881
Number 8 is before with a count of 833
Number 9 is Rostóv with a count of 776
Number 10 is thought with a count of 767
Total Time was 1.877562082 seconds
student@student-VirtualBox:~/Desktop/CSC 415/A4/csc415-assignment-4-word-blast-A
bbasMahdavi021$
```

```
student@student-VirtualBox:~/Desktop/CSC 415/A4/csc415-assignment-4-word-blast-A
bbasMahdavi021$ make run RUNOPTIONS="WarAndPeace.txt 4"
./Mahdavi_Abbas_HW4_main WarAndPeace.txt 4

Word Frequency Count on WarAndPeace.txt with 4 threads
Printing top 10 words 6 characters or more.
Number 1 is Pierre with a count of 1963
Number 2 is Prince with a count of 1928
Number 3 is Natásha with a count of 1212
Number 4 is Andrew with a count of 1143
Number 5 is himself with a count of 1020
Number 6 is princess with a count of 916
Number 7 is French with a count of 881
Number 8 is before with a count of 833
Number 9 is Rostóv with a count of 776
Number 10 is thought with a count of 767
Total Time was 1.676278580 seconds
student@student-VirtualBox:~/Desktop/CSC 415/A4/csc415-assignment-4-word-blast-A
bbasMahdavi021$
```

```
student@student-VirtualBox:~/Desktop/CSC 415/A4/csc415-assignment-4-word-blast-A
bbasMahdavi021$ make run RUNOPTIONS="WarAndPeace.txt 8"
./Mahdavi_Abbas_HW4_main WarAndPeace.txt 8

Word Frequency Count on WarAndPeace.txt with 8 threads
Printing top 10 words 6 characters or more.
Number 1 is Pierre with a count of 1963
Number 2 is Prince with a count of 1928
Number 3 is Natásha with a count of 1213
Number 4 is Andrew with a count of 1143
Number 5 is himself with a count of 1020
Number 6 is princess with a count of 916
Number 7 is French with a count of 881
Number 8 is before with a count of 833
Number 9 is Rostóv with a count of 776
Number 10 is thought with a count of 767
Total Time was 1.736279726 seconds
student@student-VirtualBox:~/Desktop/CSC 415/A4/csc415-assignment-4-word-blast-A
bbasMahdavi021$
```