Leo Powers ID: 921661426
Github: powersleo CSC415 Operating Systems

Assignment 4 – Word Blast

Description:
This assignment is to write a C program that accepts arguments of a filename and a thread count and process the file using multithreading

Approach / What I Did:
My first step to resolve the assignment was to plan out my approach. I decided a linked list would be the best way to store the words as it grows dynamically and I know how to write my own. During the coding stage I copied a bubble sort algorithm from Geeks for Geeks so that I could focus on the multithreading aspect first and then try to solve the sorting part after. I initially used the pread and strtok functions but I found that I didn't understand the outputs they were giving me. After switching to the read function I found that it tended to work out better, as I could specify the buffer size and where I wanted the function to start.

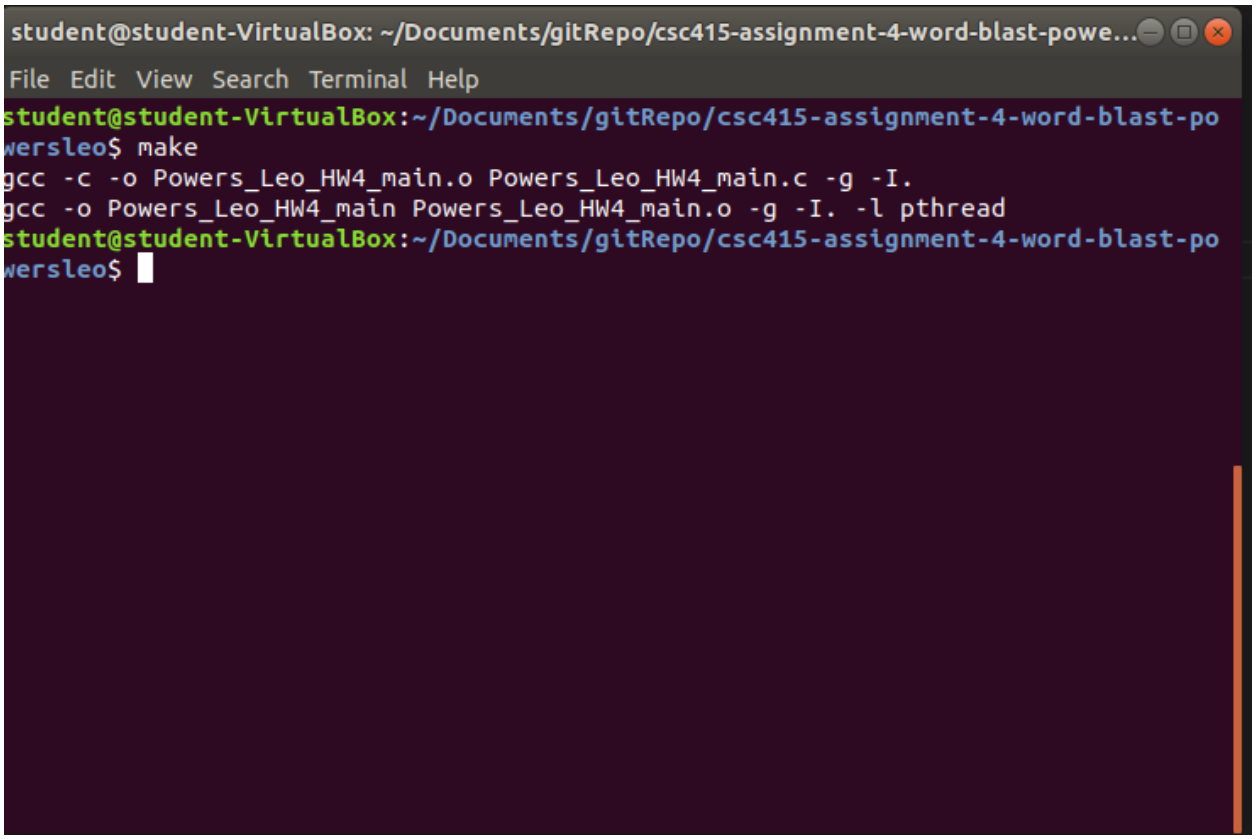
Issues and Resolutions:

My first issue I ran into was that the pread function wasn't reading the entire file. I tried to fix it but eventually switched over the read function and it ended up working better. Another issue was the strtok function was not working as I intended but the strtok_r seemed to give me the correct tokens without any finagling. Another issue was the strcmp function, I initially used that one and was very confused as to why my count was not the same as the readme file results but changing to the strcasecmp function fixed that.

Analysis:
At 1 thread my code ran at: 1.833 seconds
at 2 threads my code ran at:0.96 seconds
at 4 threads my code ran at 0.903 seconds
at 8 threads my code ran at 0.902 seconds

I think there was a major jump from 1 to 2 as the file was able to be read at a very fast rate. In 1 thread the reading took a greater amount of time to parse as only 1 thread was reading at a time. At 2 threads the file could be read at half the time, and words could be searched from the linked list. I think the times didn't really change at 4 and 8 threads both because the VM we used locked us to two threads, and because the bottle neck was likely inserting new nodes into the linked list. I think at that point the file was being read very quickly, it took a lot longer for the words to be inserted into the list. I think the way I sorted the linked list also reduced the time by a lot.

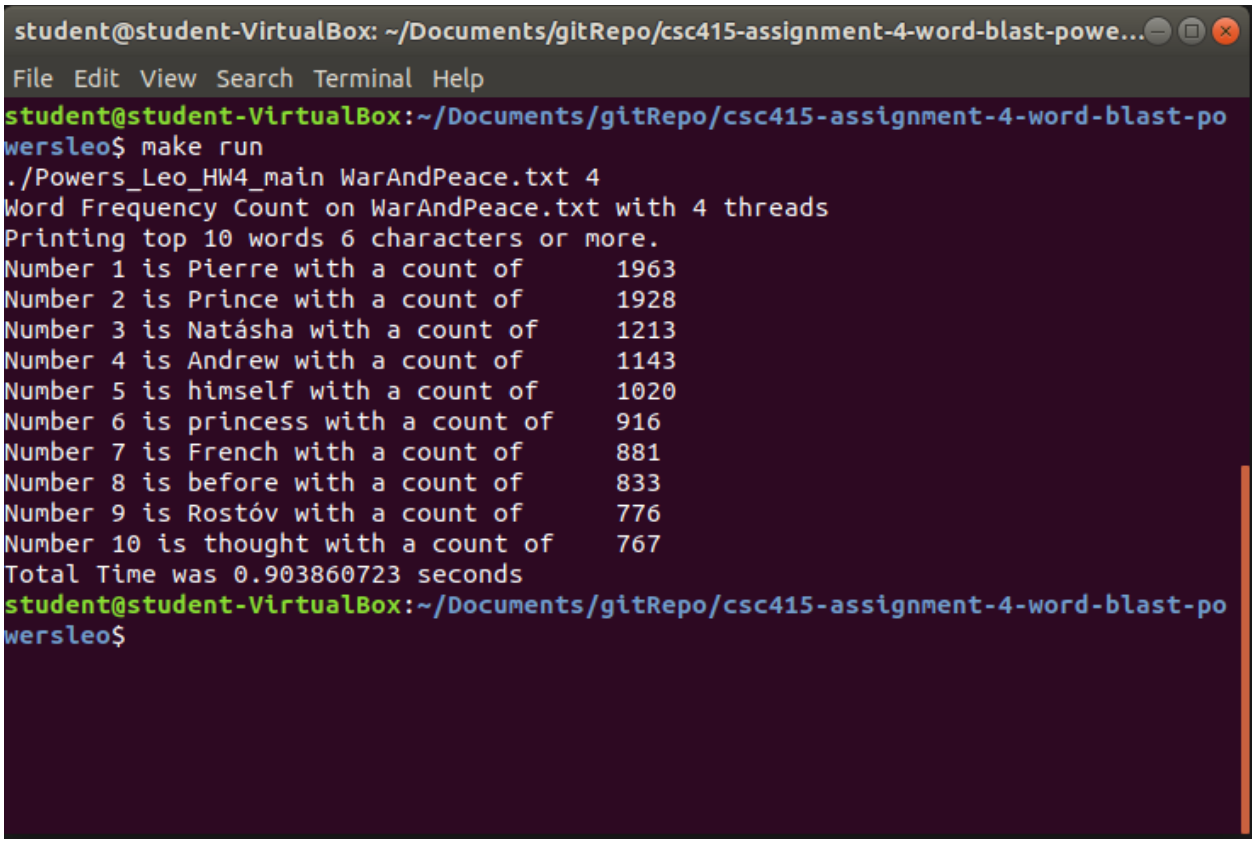
Screen shot of compilation:

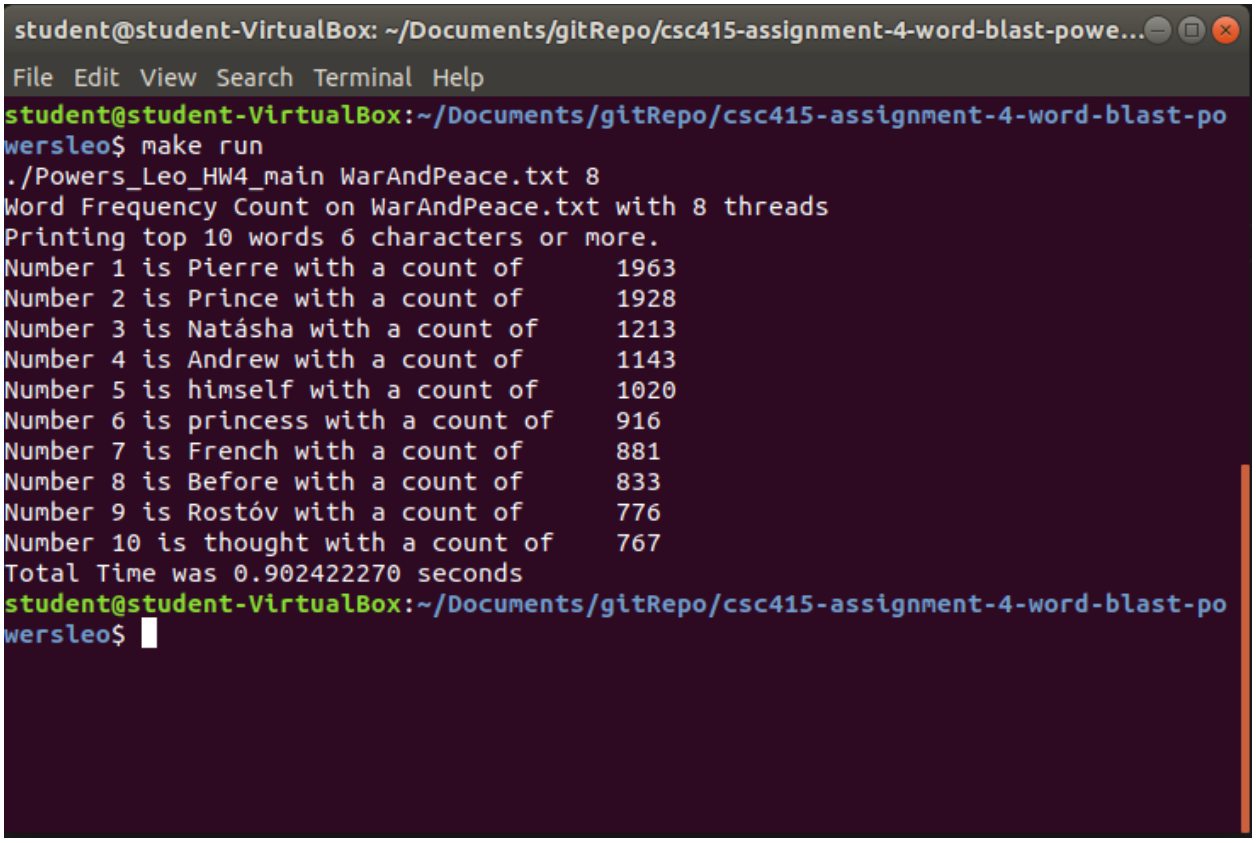


Screen shot(s) of the execution of the program:

```
student@student-VirtualBox: ~/Documents/gitRepo/csc415-assignment-4-word-blast-powe...
File  Edit  View  Search  Terminal  Help
student@student-VirtualBox:~/Documents/gitRepo/csc415-assignment-4-word-blast-po
wersleo$ make run
gcc -c -o Powers_Leo_HW4_main.o Powers_Leo_HW4_main.c -g -I.
gcc -o Powers_Leo_HW4_main Powers_Leo_HW4_main.o -g -I. -l pthread
./Powers_Leo_HW4_main WarAndPeace.txt 1
Word Frequency Count on WarAndPeace.txt with 1 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 1.833828838 seconds
student@student-VirtualBox:~/Documents/gitRepo/csc415-assignment-4-word-blast-po
wersleo$
```

```
student@student-VirtualBox: ~/Documents/gitRepo/csc415-assignment-4-word-blast-powe...
File  Edit  View  Search  Terminal  Help
student@student-VirtualBox:~/Documents/gitRepo/csc415-assignment-4-word-blast-po
wersleo$ make run
./Powers_Leo_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 0.963766269 seconds
student@student-VirtualBox:~/Documents/gitRepo/csc415-assignment-4-word-blast-po
wersleo$
```


student@student-VirtualBox: ~/Documents/gitRepo/csc415-assignment-4-word-blast-powe...
File Edit View Search Terminal Help
student@student-VirtualBox:~/Documents/gitRepo/csc415-assignment-4-word-blast-po
wersleo$ make run
./Powers_Leo_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 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 0.903860723 seconds
student@student-VirtualBox:~/Documents/gitRepo/csc415-assignment-4-word-blast-po
wersleo$


student@student-VirtualBox: ~/Documents/gitRepo/csc415-assignment-4-word-blast-powe...
File Edit View Search Terminal Help
student@student-VirtualBox:~/Documents/gitRepo/csc415-assignment-4-word-blast-po
wersleo$ make run
./Powers_Leo_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 0.902422270 seconds
student@student-VirtualBox:~/Documents/gitRepo/csc415-assignment-4-word-blast-po
wersleo$