Siarhei Pushkin ID:
Github: spushkin CSC415 Operating Systems

Assignment 4 - Word Blast

Description:

This assignment is to write a program to read War and Peace (a text copy is included). The program will count and tally each word that is six or more characters long. We have to do this using threads. Each thread will take a chunk of the file and process it, returning its results to the main, which tallies. Then, the main will print the ten, six, or more character words with the highest tallies, in order from highest to lowest, and their associated counts.

Approach:

My approach to the assignment began with a thorough understanding of the requirements. I planned how to use multithreading to divide the workload, enabling different file parts to be processed concurrently, thus improving performance. To ensure each thread worked correctly without causing data corruption, I used mutexes to handle shared data safely and avoid race conditions. I focused on using specific Linux file operations like "open," "close," "lseek," and "pread," as required by the assignment, while avoiding standard library functions for file handling. I emphasized error handling to manage potential issues like file opening failures or memory allocation errors. I organized my code into clear, modular functions, making it easy to read and maintain. This structured approach and careful planning helped me complete the assignment on time.

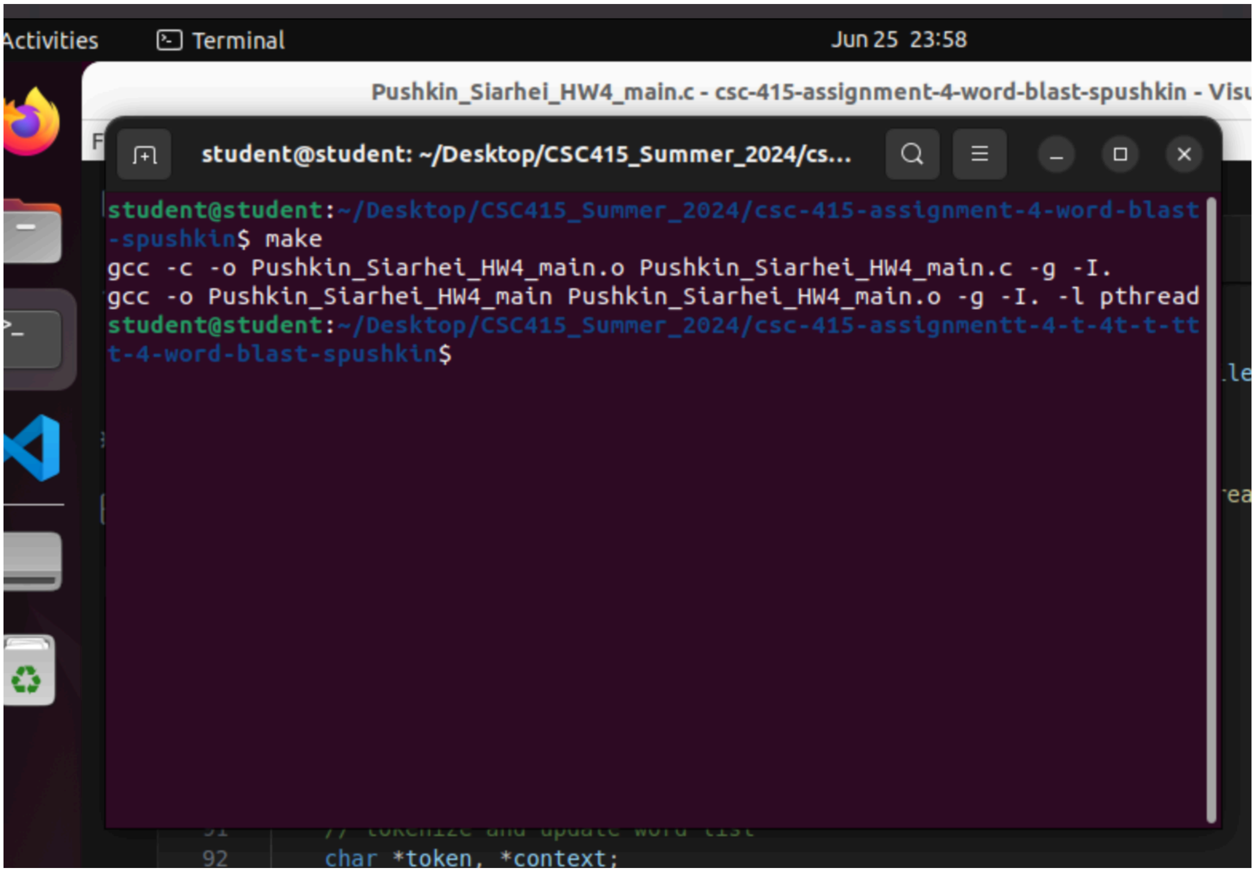
Issues and Resolutions:

During the assignment, I encountered several issues. One major challenge was ensuring thread safety when multiple threads accessed shared data, which I resolved by implementing mutexes to prevent race conditions. I spent a lot of time figuring out how to do it properly. Another problem was accurately dividing the file into segments for each thread without splitting words, which I addressed by adjusting the start and end points to align with delimiters.

Analysis:

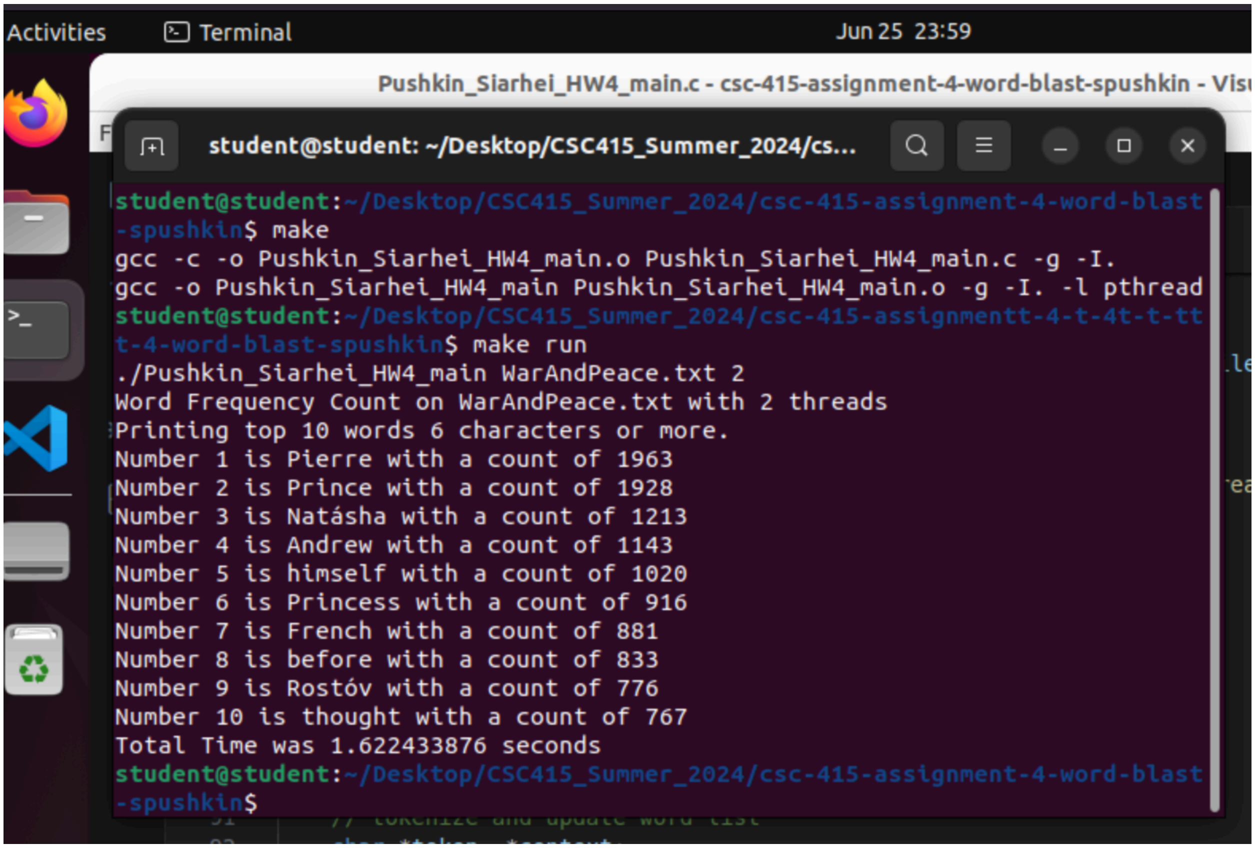
I got very interesting (and surprising for me) results while analyzing the outputs with various thread counts. With a single thread, the program took approximately 1.47 seconds to complete, serving as a baseline for comparison. Increasing the thread count to two resulted in a longer execution time of 1.63 seconds, likely due to the overhead that can be associated with the creation of the threads and synchronization. With four threads, the execution time slightly decreased to 1.57 seconds, indicating a better balance between parallel processing and overhead. Finally, with eight threads, the time was reduced to 1.54 seconds, showing better performance but also indicating diminishing returns due to the added complexity of managing more threads.

Screen shot of compilation:

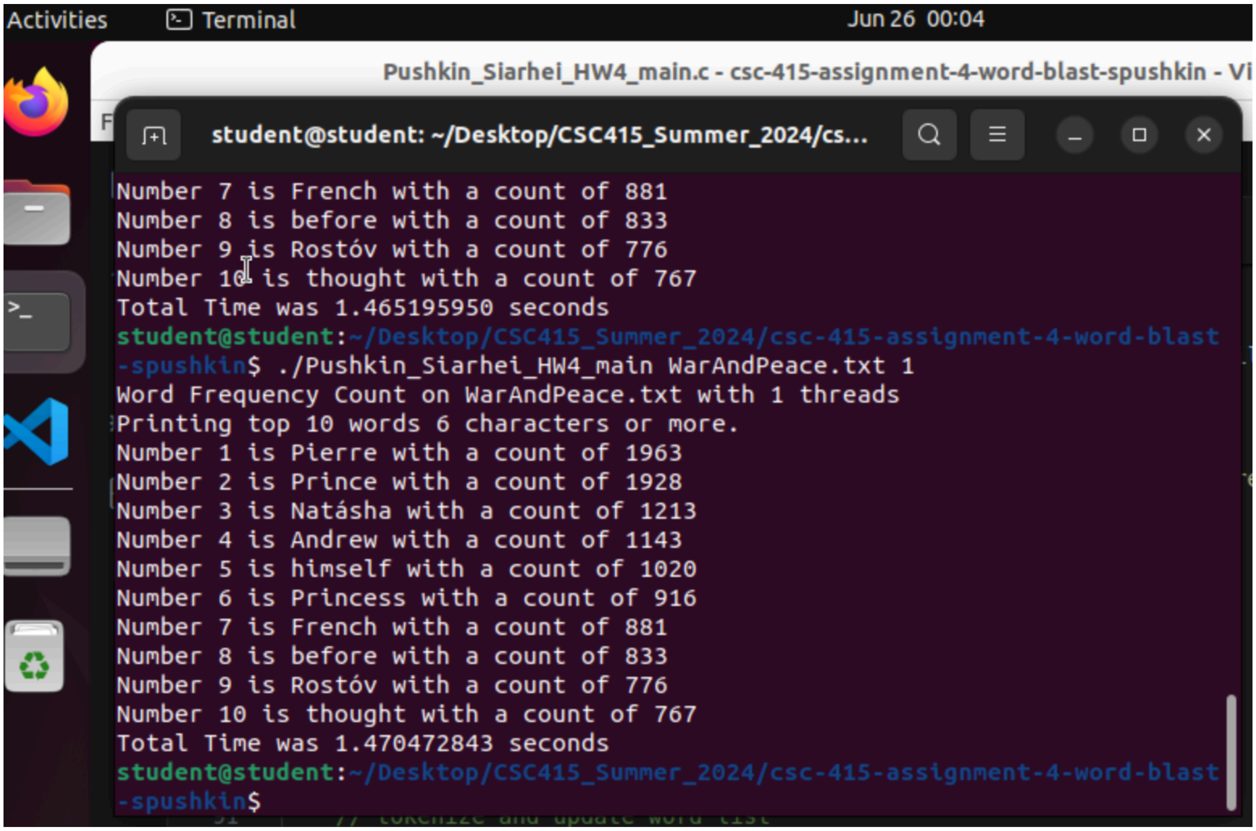

Activities
Terminal
Jun 25 23:58
Pushkin_Siarhei_HW4_main.c - csc-415-assignment-4-word-blast-spushkin - Visu
student@student: ~/Desktop/CSC415_Summer_2024/cs...
92
char *token, *context;

```
student@student:~/Desktop/CSC415_Summer_2024/csc-415-assignment-4-word-blast
-spushkin$ make
gcc -c -o Pushkin_Siarhei_HW4_main.o Pushkin_Siarhei_HW4_main.c -g -I.
gcc -o Pushkin_Siarhei_HW4_main Pushkin_Siarhei_HW4_main.o -g -I. -l pthread
student@student:~/Desktop/CSC415_Summer_2024/csc-415-assignmentt-4-t-4t-t-tt
t-4-word-blast-spushkin$
```

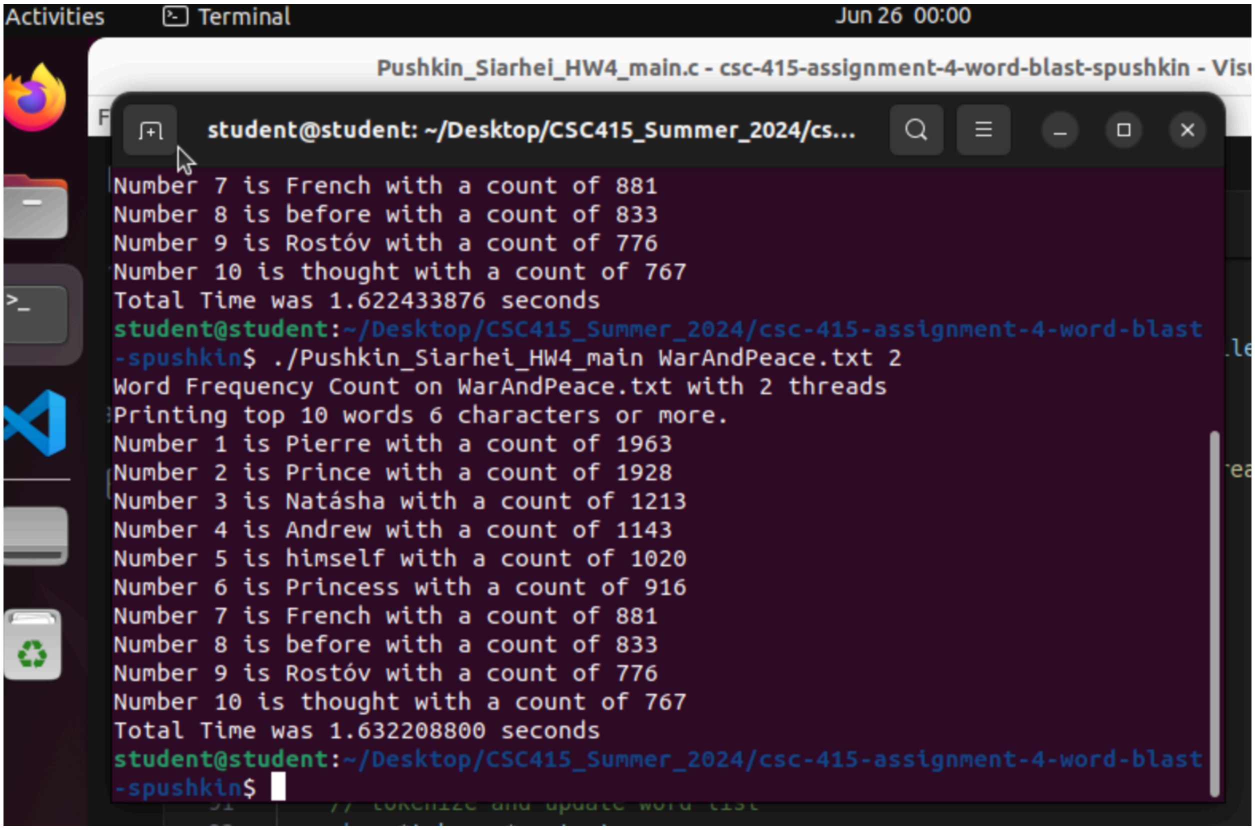
Screen shot(s) of the execution of the program:


Activities
Terminal
Jun 25 23:59
Pushkin_Siarhei_HW4_main.c - csc-415-assignment-4-word-blast-spushkin - Visu
student@student: ~/Desktop/CSC415_Summer_2024/cs...

```
student@student:~/Desktop/CSC415_Summer_2024/csc-415-assignment-4-word-blast
-spushkin$ make
gcc -c -o Pushkin_Siarhei_HW4_main.o Pushkin_Siarhei_HW4_main.c -g -I.
gcc -o Pushkin_Siarhei_HW4_main Pushkin_Siarhei_HW4_main.o -g -I. -l pthread
student@student:~/Desktop/CSC415_Summer_2024/csc-415-assignmentt-4-t-4t-t-tt
t-4-word-blast-spushkin$ make run
./Pushkin_Siarhei_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 1.622433876 seconds
student@student:~/Desktop/CSC415_Summer_2024/csc-415-assignment-4-word-blast
-spushkin$
```

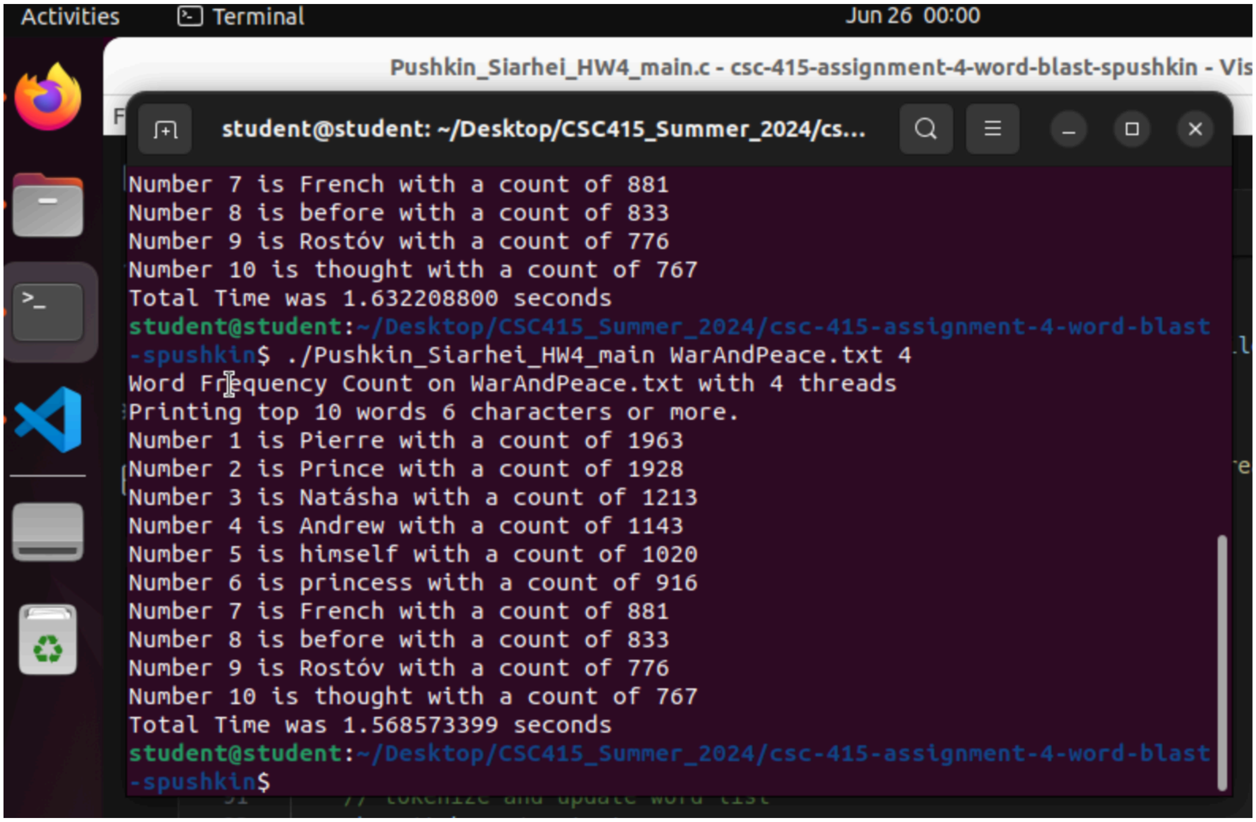

Activities
Terminal
Jun 26 00:04
Pushkin_Siarhei_HW4_main.c - csc-415-assignment-4-word-blast-spushkin - Vi
student@student: ~/Desktop/CSC415_Summer_2024/cs...

```
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 1.465195950 seconds
student@student:~/Desktop/CSC415_Summer_2024/csc-415-assignment-4-word-blast
-spushkin$ ./Pushkin_Siarhei_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 1.470472843 seconds
student@student:~/Desktop/CSC415_Summer_2024/csc-415-assignment-4-word-blast
-spushkin$
```

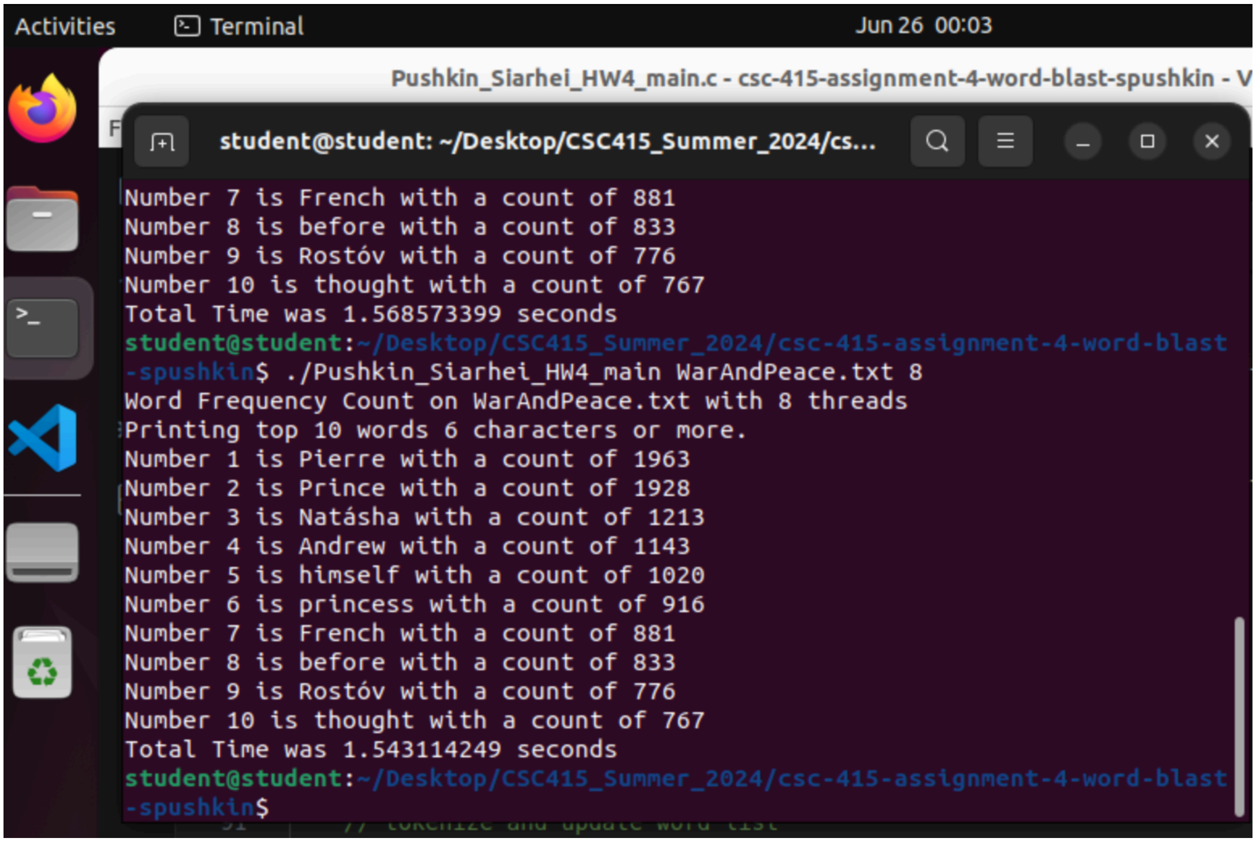


Activities
Terminal
Jun 26 00:00
Pushkin_Siarhei_HW4_main.c - csc-415-assignment-4-word-blast-spushkin - Visu
student@student: ~/Desktop/CSC415_Summer_2024/cs...

```
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 1.622433876 seconds
student@student:~/Desktop/CSC415_Summer_2024/csc-415-assignment-4-word-blast
-spushkin$ ./Pushkin_Siarhei_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 1.632208800 seconds
student@student:~/Desktop/CSC415_Summer_2024/csc-415-assignment-4-word-blast
-spushkin$ 
```


Activities
Terminal
Jun 26 00:00
Pushkin_Siarhei_HW4_main.c - csc-415-assignment-4-word-blast-spushkin - Vis
student@student: ~/Desktop/CSC415_Summer_2024/cs...

```
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 1.632208800 seconds
student@student:~/Desktop/CSC415_Summer_2024/csc-415-assignment-4-word-blast
-spushkin$ ./Pushkin_Siarhei_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 1.568573399 seconds
student@student:~/Desktop/CSC415_Summer_2024/csc-415-assignment-4-word-blast
-spushkin$
```



Activities
Terminal
Jun 26 00:03
Pushkin_Siarhei_HW4_main.c - csc-415-assignment-4-word-blast-spushkin - V
student@student: ~/Desktop/CSC415_Summer_2024/cs...

```
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 1.568573399 seconds
student@student:~/Desktop/CSC415_Summer_2024/csc-415-assignment-4-word-blast
-spushkin$ ./Pushkin_Siarhei_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 1.543114249 seconds
student@student:~/Desktop/CSC415_Summer_2024/csc-415-assignment-4-word-blast
-spushkin$
```