Anmol Burmy ID: 921143454
Github: Burmy CSC415 Operating Systems

Assignment 4 - Word Blast

Description:

This assignment is to write a C program to read War and Peace (a text copy is included with this assignment) and it is to count and tally each of the words that are 6 or more characters long. We are only using Linux (not library) file functions, i.e. open, close, read, lseek, pread.

Approach / What I Did:

Firstly, I read all the steps and instructions carefully to fully understand the assignment. I started the assignment by firstly editing the "Makefile". I added my first name and last name. After that I started working on a program written in C. I started with creating a data structure that would store each word and the amount of times the words occur. I also created two functions, readFile function will scan through each chunk of the file. I created multi thread processes. Each thread would take a chunk of the text file then parse it, making sure the token is 6 or more characters and then pass it to the addWords function which will save or update words in an array of "word" structs. After threading, the main will print the ten, 6 or more characters, words with the highest tallies, in order highest to lowest, and their associated counts. i.e. The top ten words and the number of times that word appears in the text. After the process completes, cleanup is also performed to free all the memory for further usage.

Issues and Resolutions:

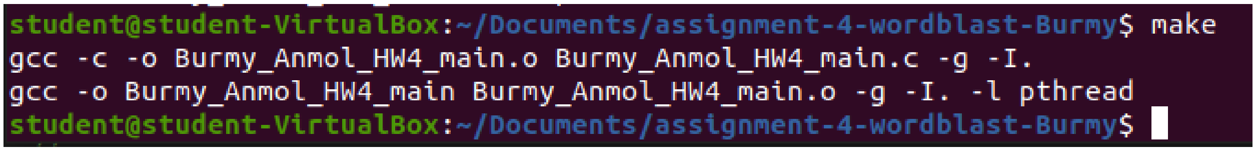
One of the issues I came across was to choose what data structure to use that would require less time to do but also implement the program correctly. First I was trying out using a hashmap but figured it was going to take a lot of time to implement and I figured I would not have enough time to finish the assignment. Then after consulting with my classmates, I found out that just using global array of structures would be the best and easiest way to do the assignment.

Analysis:

From the 4 different execution outputs, I noticed that as the number of threads are increased the time it takes to compile the program also increases. I believe that the time is mainly based on the loop operations when trying to add a new frequency to a word. I also noticed that compared to 1 thread, and all the other threads had only had a difference of 0.01 sec between them. Multiple threads provides more economy and scalability as the program divides the file into number of threads and performs the operations in parallel

Anmol Burmy ID: 921143454
Github: Burmy CSC415 Operating Systems

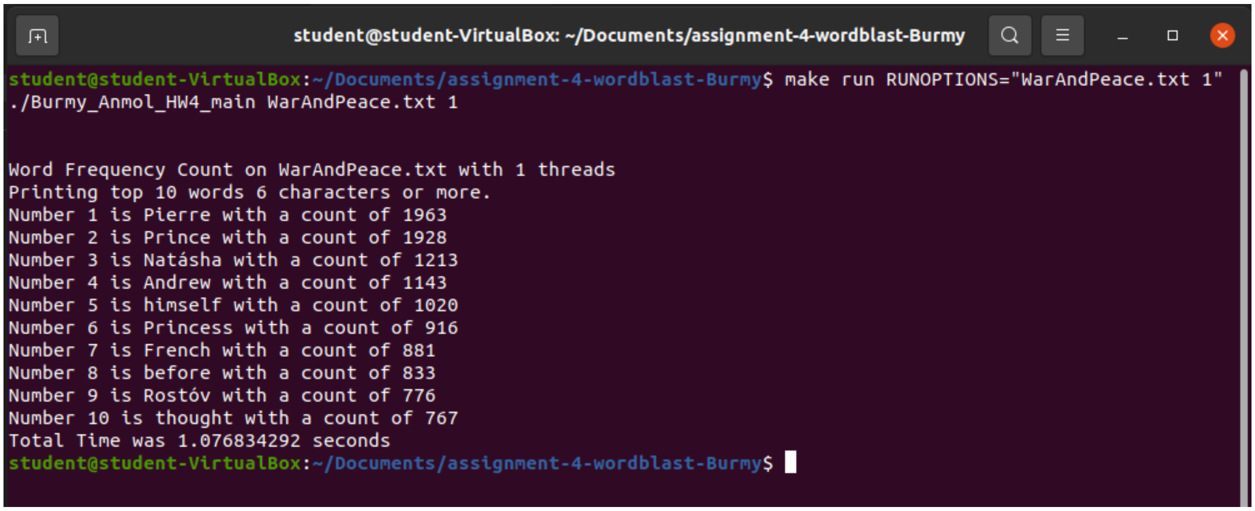
Screenshot of compilation:



```
student@student-VirtualBox:~/Documents/assignment-4-wordblast-Burmy$ make
gcc -c -o Burmy_Anmol_HW4_main.o Burmy_Anmol_HW4_main.c -g -I.
gcc -o Burmy_Anmol_HW4_main Burmy_Anmol_HW4_main.o -g -I. -l pthread
student@student-VirtualBox:~/Documents/assignment-4-wordblast-Burmy$ 
```

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

1 Thread -

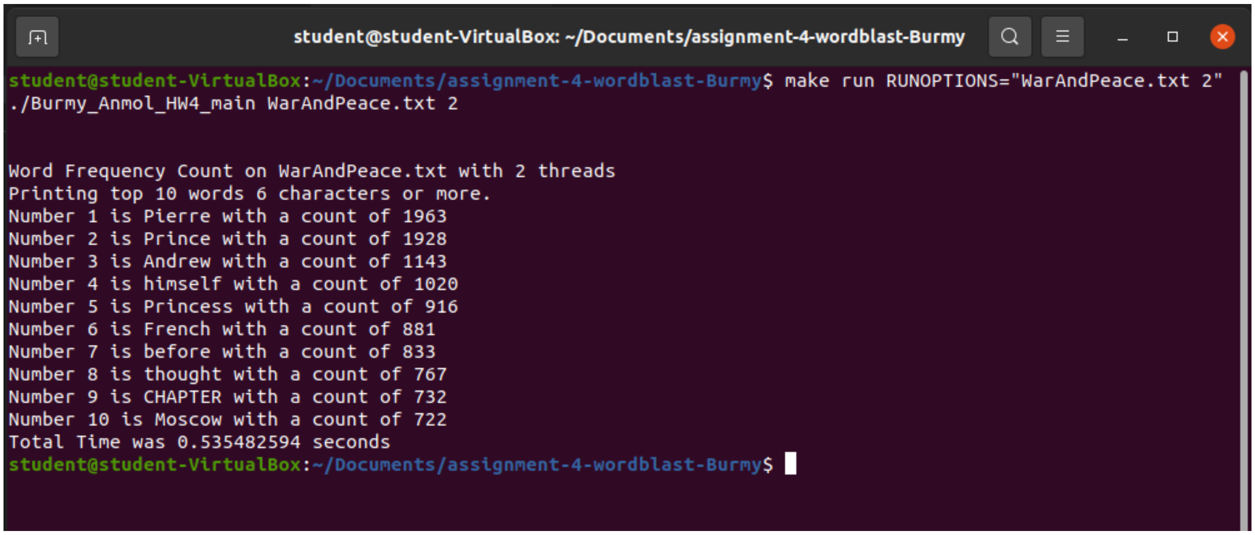


student@student-VirtualBox: ~/Documents/assignment-4-wordblast-Burmy

```
student@student-VirtualBox:~/Documents/assignment-4-wordblast-Burmy$ make run RUNOPTIONS="WarAndPeace.txt 1"
./Burmy_Anmol_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 1.076834292 seconds
student@student-VirtualBox:~/Documents/assignment-4-wordblast-Burmy$ 
```

2 Threads -



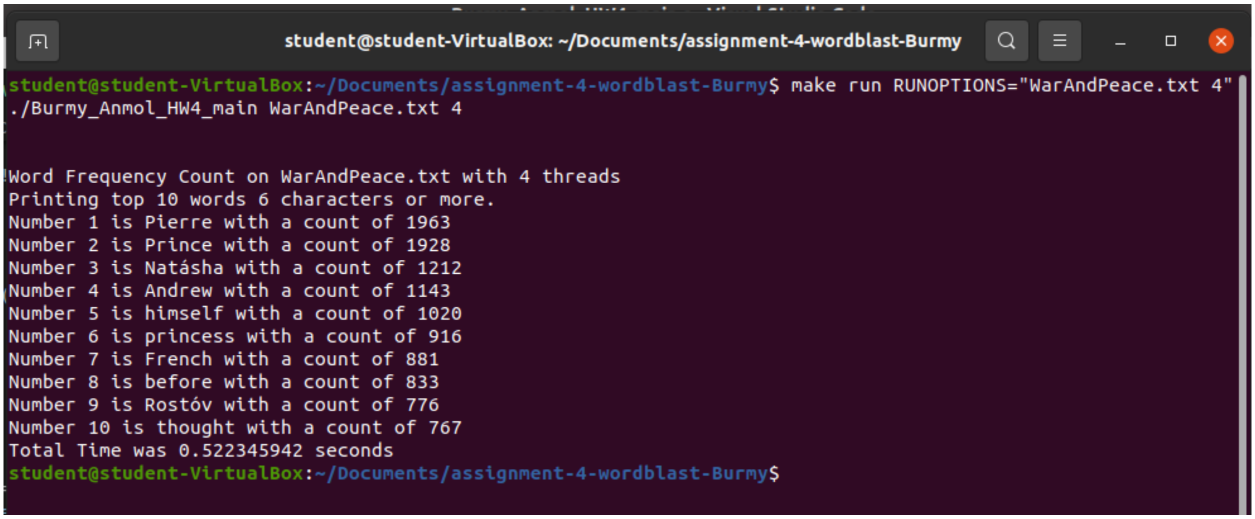
student@student-VirtualBox: ~/Documents/assignment-4-wordblast-Burmy

```
student@student-VirtualBox:~/Documents/assignment-4-wordblast-Burmy$ make run RUNOPTIONS="WarAndPeace.txt 2"
./Burmy_Anmol_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 Andrew with a count of 1143
Number 4 is himself with a count of 1020
Number 5 is Princess with a count of 916
Number 6 is French with a count of 881
Number 7 is before with a count of 833
Number 8 is thought with a count of 767
Number 9 is CHAPTER with a count of 732
Number 10 is Moscow with a count of 722
Total Time was 0.535482594 seconds
student@student-VirtualBox:~/Documents/assignment-4-wordblast-Burmy$ 
```

Anmol Burmy ID: 921143454
Github: Burmy CSC415 Operating Systems

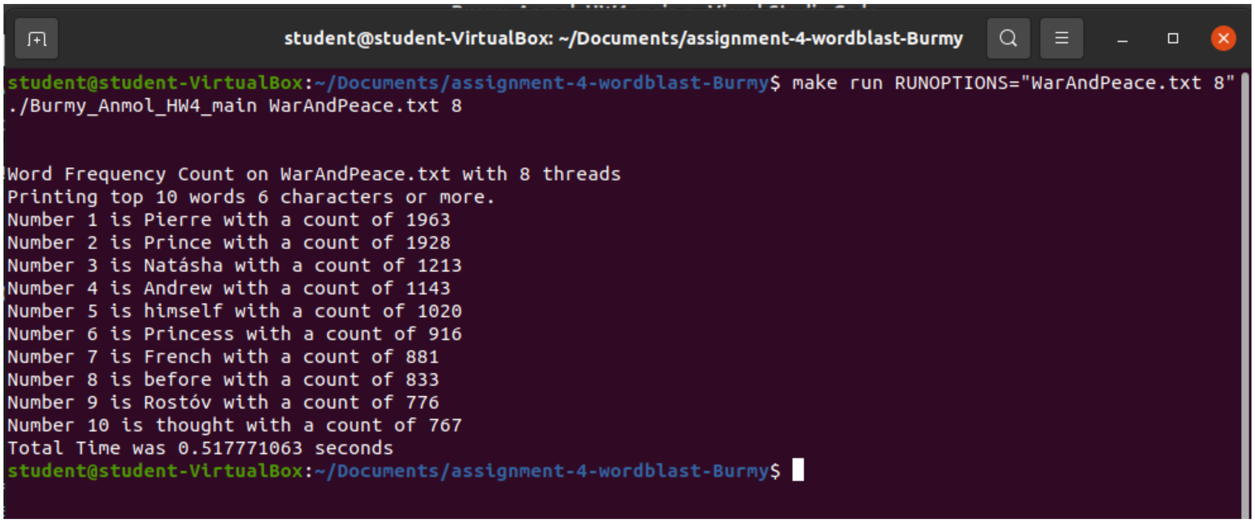
4 Threads -


student@student-VirtualBox: ~/Documents/assignment-4-wordblast-Burmy

```
student@student-VirtualBox:~/Documents/assignment-4-wordblast-Burmy$ make run RUNOPTIONS="WarAndPeace.txt 4"
./Burmy_Anmol_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 0.522345942 seconds
student@student-VirtualBox:~/Documents/assignment-4-wordblast-Burmy$
```

8 Threads -


student@student-VirtualBox: ~/Documents/assignment-4-wordblast-Burmy

```
student@student-VirtualBox:~/Documents/assignment-4-wordblast-Burmy$ make run RUNOPTIONS="WarAndPeace.txt 8"
./Burmy_Anmol_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 0.517771063 seconds
student@student-VirtualBox:~/Documents/assignment-4-wordblast-Burmy$
```