

## Assignment 4 – Word Blast

### Description:

For this assignment, your program is 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. And again, you will only use Linux (not library) file functions, i.e. open, close, read, lseek, pread.

### Approach / What I Did:

I defined all the variables which are needed globally using `#define` keywords. Then I described structures that are required so that I don't need to write similar code again and again, and I can directly take the help of the structure defined above and access its elements. After that I wrote the function and logic for adding a word and processing the chunk, then comes the main function, where I've predefined the filename and thread count, but if it is given as an argument, it'll update the variables. Then I read the input file, counted the words, updated the counter in the meantime, and then printed out the words in decreasing order.

### Issues and Resolutions:

My first issue was in implementing the add a new word function, which is defined by adding a new word.

My Second issue was the string comparison, I was using loops.

My third issue was running the code in my Linux code environment, as it was not able to recognize `'pthread_create'` and `'pthread_join'`

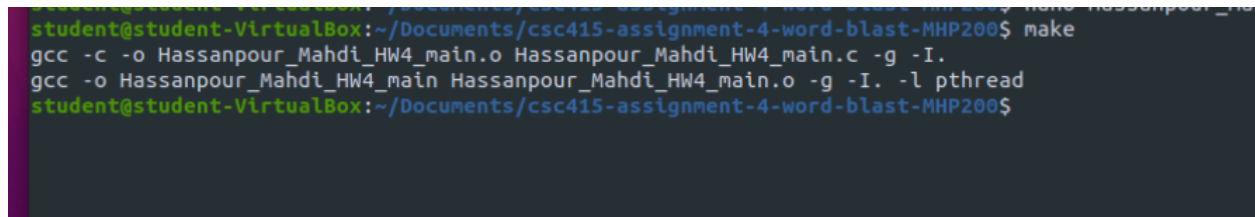
I resolved the above issues by comparing both the strings by using the library function `'strcasecmp'` inside a for loop, then applied conditional statements (if statements), and checked the size is within the limits.

I resolved my final issue by researching a lot on google, and finally was able to do it with the help of a stack overflow solution, which was to add a `-pthread` flag while executing the code in my terminal. That worked for me.

### Analysis:

As we increase the number of threads, the total execution time is decreasing, because the increase in the number of threads makes multiple processes concurrent, hence it divides the work into multiple threads.

### Compilation Screenshot:



```
student@student-VirtualBox:~/Documents/csc415-assignment-4-word-blast-MHP200$ make
gcc -c -o Hassanpour_Mahdi_HW4_main.o Hassanpour_Mahdi_HW4_main.c -g -I.
gcc -o Hassanpour_Mahdi_HW4_main Hassanpour_Mahdi_HW4_main.o -g -I. -l pthread
student@student-VirtualBox:~/Documents/csc415-assignment-4-word-blast-MHP200$
```

### Screenshot of the output:

```
make: *** [Makefile:60: run] Segmentation fault (core dumped)
student@student-VirtualBox:~/Documents/csc415-assignment-4-word-blast-MHP200$ make run RUNOPTIONS="WarAndPeace.txt 7"
./Hassanpour_Mahdi_HW4_main WarAndPeace.txt 7
```

```
Word Frequency Count on WarAndPeace.txt with 7 threads
Printing top 10 words 6 characters or more.
Number 1 is prince with a count of 1867
Number 2 is Pierre with a count of 1779
Number 3 is Natásha with a count of 1061
Number 4 is Andrew with a count of 1026
Number 5 is himself with a count of 1011
Number 6 is French with a count of 873
Number 7 is princess with a count of 869
Number 8 is before with a count of 821
Number 9 is thought with a count of 763
Number 10 is CHAPTER with a count of 732
Total Time Was 2.575211391 seconds
```

```
student@student-VirtualBox:~/Documents/csc415-assignment-4-word-blast-MHP200$
```

```
Total Time Was 2.933735243 seconds
```

```
student@student-VirtualBox:~/Documents/csc415-assignment-4-word-blast-MHP200$ make run RUNOPTIONS="WarAndPeace.txt 6"
./Hassanpour_Mahdi_HW4_main WarAndPeace.txt 6
```

```
Word Frequency Count on WarAndPeace.txt with 6 threads
Printing top 10 words 6 characters or more.
Number 1 is Prince with a count of 1867
Number 2 is Pierre with a count of 1779
Number 3 is Natásha with a count of 1061
Number 4 is Andrew with a count of 1026
Number 5 is himself with a count of 1011
Number 6 is French with a count of 873
Number 7 is Princess with a count of 869
Number 8 is before with a count of 821
Number 9 is thought with a count of 763
Number 10 is CHAPTER with a count of 732
Total Time Was 2.795686798 seconds
```

```
student@student-VirtualBox:~/Documents/csc415-assignment-4-word-blast-MHP200$ make run RUNOPTIONS="WarAndPeace.txt 8"
```

```
gcc: error: 3: No such file or directory
```

```
student@student-VirtualBox:~/Documents/csc415-assignment-4-word-blast-MHP200$ make run RUNOPTIONS="WarAndPeace.txt 1"
./Hassanpour_Mahdi_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 Prince with a count of 1867
Number 2 is Pierre with a count of 1779
Number 3 is Natásha with a count of 1061
Number 4 is Andrew with a count of 1026
Number 5 is himself with a count of 1011
Number 6 is French with a count of 873
Number 7 is Princess with a count of 869
Number 8 is before with a count of 821
Number 9 is thought with a count of 763
Number 10 is CHAPTER with a count of 732
Total Time Was 4.951870331 seconds
```

```
student@student-VirtualBox:~/Documents/csc415-assignment-4-word-blast-MHP200$ make run RUNOPTIONS="WarAndPeace.txt 2"
./Hassanpour_Mahdi_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 Prince with a count of 1867
Number 2 is Pierre with a count of 1779
Number 3 is Natásha with a count of 1061
Number 4 is Andrew with a count of 1026
Number 5 is himself with a count of 1011
Number 6 is French with a count of 873
Number 7 is Princess with a count of 869
Number 8 is before with a count of 821
Number 9 is thought with a count of 763
Number 10 is CHAPTER with a count of 732
Total Time Was 2.933735243 seconds
```

```
student@student-VirtualBox:~/Documents/csc415-assignment-4-word-blast-MHP200$ make run RUNOPTIONS="WarAndPeace.txt 6"
```