

## PROJECT – 1A

### Design Description

In my implementation I have used 3 IO bounded threads and 1 CPU bounded thread for finding the largest 'k' occurrences of a word.

Initially I will split the file in to three parts depending on the number of lines presented in the given input file. I used "**wc -l file\_name**" Linux command to get the number of lines in the file and used split command to split the input files in to three files (**split -l number\_of\_lines file\_name**). Split command will create three file naming "**split\_fileaa**", "**split\_fileab**", and "**split\_fileac**" respectively. Mentioned files are memory mapped separately and they are processed concurrently with three IO threads.

In a file we may have several number of valid words depending on the input. We need a data structure to efficiently store the word and its occurrences. I have implemented **Chain-Hashing Algorithm** in C. It has 26 buckets. Each bucket corresponds to individual alphabet like "Buck 1 - 'A' ". Each bucket will maintain a linked list of words for a given starting alphabet. For example bucket 1 will have a reference to a linked list whose words are starting with 'A' and its corresponding occurrences.

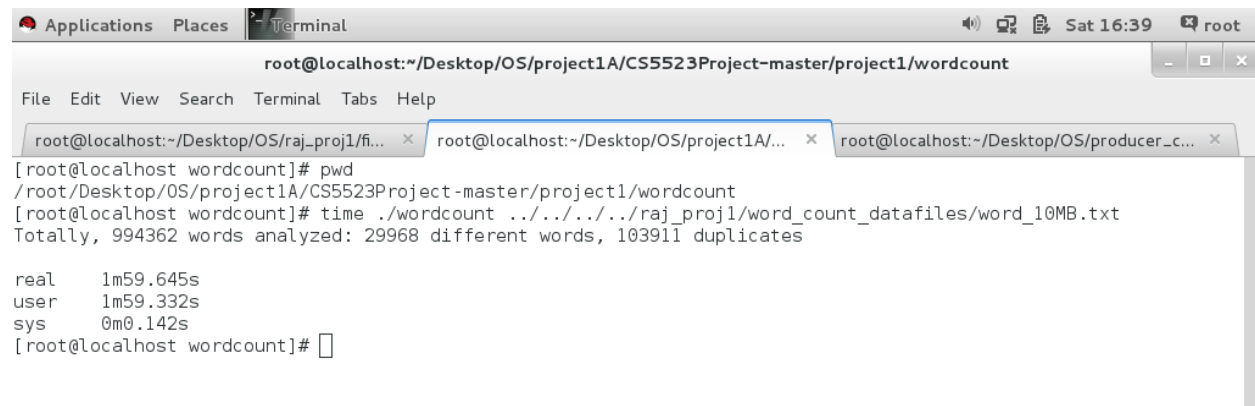
After bookkeeping the list of words and occurrences using hashing, we need an efficient sorting algorithm to display the largest 'k' occurrences in the list. I have implemented **Min-Heap Algorithm** in c to find the largest 'k' occurrences. One CPU thread will perform sorting in the hash structure and finally stores it in the Min-Heap.

The 3 CPU bound thread processes the file concurrently, in the meantime 1 CPU bound thread will wait for the processing for files to complete. After that, it will start running to provide the top most occurrences.

### Performance Comparison

Performance is considerably improved with multithreading compared single threading. Below are the screen shorts of same program with both single and multithreading.

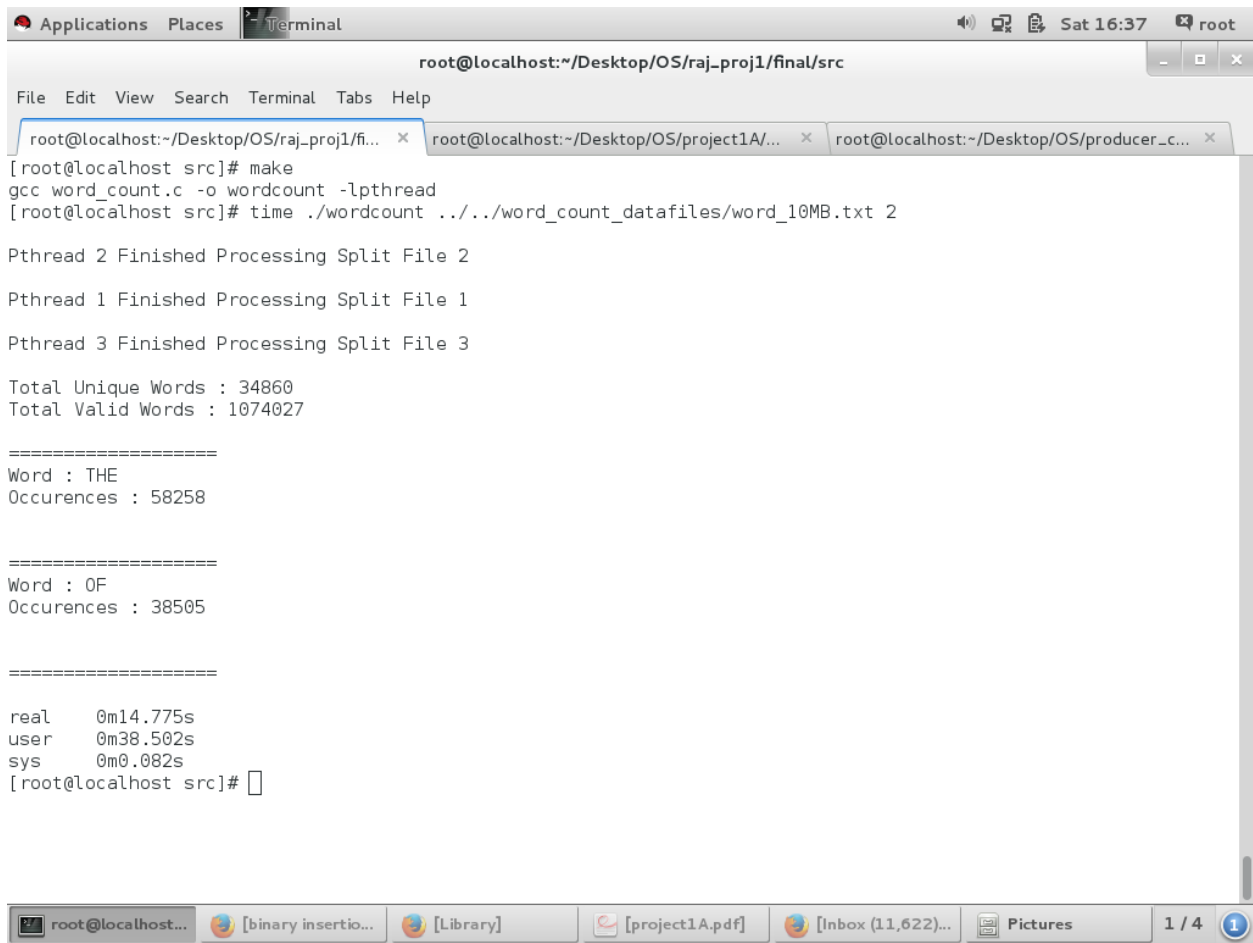
#### Single Threaded Program



```
root@localhost:~/Desktop/OS/project1A/CS5523Project-master/project1/wordcount
File Edit View Search Terminal Tabs Help
root@localhost:~/Desktop/OS/raj_proj1/fi... x root@localhost:~/Desktop/OS/project1A/... x root@localhost:~/Desktop/OS/producer_c... x
[root@localhost wordcount]# pwd
/root/Desktop/OS/project1A/CS5523Project-master/project1/wordcount
[root@localhost wordcount]# time ./wordcount ../../../../raj_proj1/word_count_datafiles/word_10MB.txt
Totally, 994362 words analyzed: 29968 different words, 103911 duplicates

real    1m59.645s
user    1m59.332s
sys      0m0.142s
[root@localhost wordcount]#
```

## Multithreaded Program



The screenshot shows a terminal window titled "Terminal" with the path "root@localhost:~/Desktop/OS/raj\_proj1/final/src". The terminal displays the following output:

```
root@localhost:~/Desktop/OS/raj_proj1/fi... x root@localhost:~/Desktop/OS/project1A/... x root@localhost:~/Desktop/OS/producer_c... x
[root@localhost src]# make
gcc word_count.c -o wordcount -lpthread
[root@localhost src]# time ./wordcount ../../word_count_datafiles/word_10MB.txt 2

Pthread 2 Finished Processing Split File 2
Pthread 1 Finished Processing Split File 1
Pthread 3 Finished Processing Split File 3

Total Unique Words : 34860
Total Valid Words : 1074027

=====
Word : THE
Occurences : 58258

=====
Word : OF
Occurences : 38505

=====

real    0m14.775s
user    0m38.502s
sys     0m0.082s
[root@localhost src]#
```

The terminal window has a menu bar with "File", "Edit", "View", "Search", "Terminal", "Tabs", and "Help". The status bar at the bottom shows the current directory and several open files: "root@localhost...", "[binary insertio...", "[Library]", "[project1A.pdf]", "[Inbox (11,622)...", and "Pictures". The page number "1 / 4" is also visible.

The above outputs may vary depending on the specifications of the computer. But multithreaded program will always have minimum execution time when compared to single threaded program.

The above input is observed in the system with following specifications:

### [View basic information about your computer](#)

#### Windows edition

---

Windows 7 Home Premium


Copyright © 2009 Microsoft Corporation. All rights reserved.

Service Pack 1

[Get more features with a new edition of Windows 7](#)

#### System

---

Manufacturer:	Lenovo
Model:	Lenovo Win7 PC
Rating:	 <a href="#">Windows Experience Index</a>
Processor:	Intel(R) Core(TM) i5 CPU M 480 @ 2.67GHz 2.67 GHz
Installed memory (RAM):	4.00 GB (3.87 GB usable)
System type:	64-bit Operating System
Pen and Touch:	No Pen or Touch Input is available for this Display