****

**PDC PROJECT REPORT**

**WORD COUNT USING SHARED MEMORY SYSTEM**

Instructor:

Ms. Nausheen Shoaib

Sec:**C**

Group Members:

Muneeb Asif 18K-1163

Sameer Yousuf 18K-0158

# Objective:

Parallel programming is a type of computation in which many calculations or the execution of processes are carried out simultaneously. Large problems can often be divided into smaller ones, which can then be solved at the same time to obtain the result much quicker. There are many ways to achieve parallel programming, but we have implemented using Open MP. Our algorithm counts the frequency of each word present in the file.

# Introduction:

This project was developed in **Ubuntu Linux 18.04** and Windows.We have used Open MP to achieve parallelism using shared memory. The application was developed in C++ language and the library used for parallelism was OpenMP.

# Methodology:

For calculating the frequency of each word, we have created a data structure called TRIE. **TRIE** is an efficient information retrieval and a tree like data structure. Using TRIE, search complexities can be brought to optimal limit (key length). Using TRIE, we can search the key in O(M) time. Similarly using TRIE we can calculate each word frequency and stored in the variable occurrence mentioned in the code.

To achieve parallelism, main file is divided into 5 equal parts and **5 threads** are created and each thread is responsible to solve a part. Using #pragma omp section each thread is assigned a section to count the frequency of each word present in that part of the file.

# Comparison Graph:

|  |  |  |
| --- | --- | --- |
| ***FILE SIZES*** | ***OMP*** | ***SERIAL*** |
| 1mb | 0.176969 | 0.218143 |
| 2mb | 0.29094 | 0.442 |
| 6mb | 0.882503 | 1.41348 |
| 8 mb | 1.16376 | 1.6336 |
| 10 mb | 1.470838333 | 1.889133 |
| 12 mb | 1.87344 | 2.79989 |
| 16 mb | 2.23984 | 3.93328 |

# ****Conclusion**:**

The Time-Stamps recorded in the above graph from both the programs shows that Open MP is comparatively faster than Serial in terms of speed. The same file of 16mb took 2.2 seconds in Open MP whereas Serial Process took almost 4 seconds.

Therefore, we can say that calculating frequencies of all the words will be done more efficiently in Open MP than serial programming which concludes that Open MP will be a better choice for performing high order calculations in terms of speed and accuracy.