

# CS307- Fall 2020-2021

## Homework 4

### Memory Mapping & File Reading

by Baran Çimen

26336

My goal was to find the letter 'a' 's in the text file, loremipsum.txt. In order to read the files, I have used 3 different techniques; first one is written in C++ while the rest is written in C.

The first code is written with fstream library in C++. With the ifstream class and getline function, I read the file line by line and then looked at every character of the line. For every 'a' character in the line program increments a counter which was initially 0. Meanwhile, the program is indirectly calculates the runtime by getting the clock at the start of the program and at the end of the program.

The second code which is written with basic stdio.h library in C. I used fopen function to open the file and with the fgetc function in the while condition, I started a loop in order to get 128 character buffer. Just like in the first program, it increments a counter for every character 'a' and keeps track of the runtime.

The last program I have written used the mmap.h library in C in order to write the text file into the main memory so that it can be used later on. I used mmap function to map the file into the main memory and retrieved the pointer to the start of it. Again, with a counter the program calculates the number of occurrences of the letter 'a'. Just like all other programs, it keeps track of the runtime by using clock function.

## Time Efficiency

```
[cimenbaran@flow ~]$ gcc -o cout.out Source.c
[cimenbaran@flow ~]$ gcc -o cout1.out Source1.c
[cimenbaran@flow ~]$ g++ -o cpp.out Source.cpp
[cimenbaran@flow ~]$ ./cpp.out
19082160
Time: 3.32 seconds
[cimenbaran@flow ~]$ ./cout.out
19082160
Time: 1.660000 seconds
[cimenbaran@flow ~]$ ./cout1.out
19082160
Time: 1.380000 seconds
[cimenbaran@flow ~]$
```

The fastest program is the one with mman.h library.

(1.38 seconds)

The second fastest is the one with fopen and fgets.

(1.66 seconds)

The slowest is the C++ code with fstream library.

(3.32 seconds)

## Advantages and Disadvantages

Mman.h program has the fastest runtime but it uses too much memory since it writes all the text into the RAM. Also it has the longest code among the 3 codes I have provided

Fopen and fgets program still has a very fast runtime and only uses the buffer for the memory which is 128 bytes for my program. It is easier to write than the program that utilizes mman.h library, but a little bit longer than C++ program in terms of lines.

C++ code is 2.5 times slower than the fastest method although it is very easy to write it.

Overall, in my opinion the best option is the C code with fopen and fgets since it doesn't waste too much memory and still runs very fast.