Bilkent University
Department of Computer Engineering

CS 342 - Database Systems

Project 1 Report

Ayşe Kelleci, 21902532, Section 2
Çağatay Şafak, 21902730, Section 2

Instructor: Dr. İbrahim Körpeoğlu

21st of March 2023

This report is submitted to the Department of Computer Engineering of Bilkent University in partial fulfillment of the requirements of the Operating Systems course CS 342.

To find the occurrences of words in a series of input files, then printing the top K words amongst them to another output file, we have developed two different programs: proctopk and threadtopk. While proctopk uses multiple child processes to do it, threadtopk uses multiple POSIX threads (Pthreads). We compared the execution times of these programs for different values of K and N. At the end, we have seen that threads are generally a faster option compared to child processes, by their nature. Here are the results.

Proctopk

Execution time(s) for proctopk program

	N=3	N=5	N=7	N=10
K=3	0.006	0.015	0.016	0.023
K=5	0.005	0.012	0.021	0.022
K=10	0.008	0.010	0.020	0.024
K=20	0.125	0.135	0.137	0.144

Table 1: Execution times for proctopk.

As it can be seen, for the small values of K, the difference is not very observable, however, as the value of K increases, the time it takes to execute increases exponentially. This effect can also be seen in Figure 2. This is because the program has a time complexity of $O(K^2)$ in many code blocks.

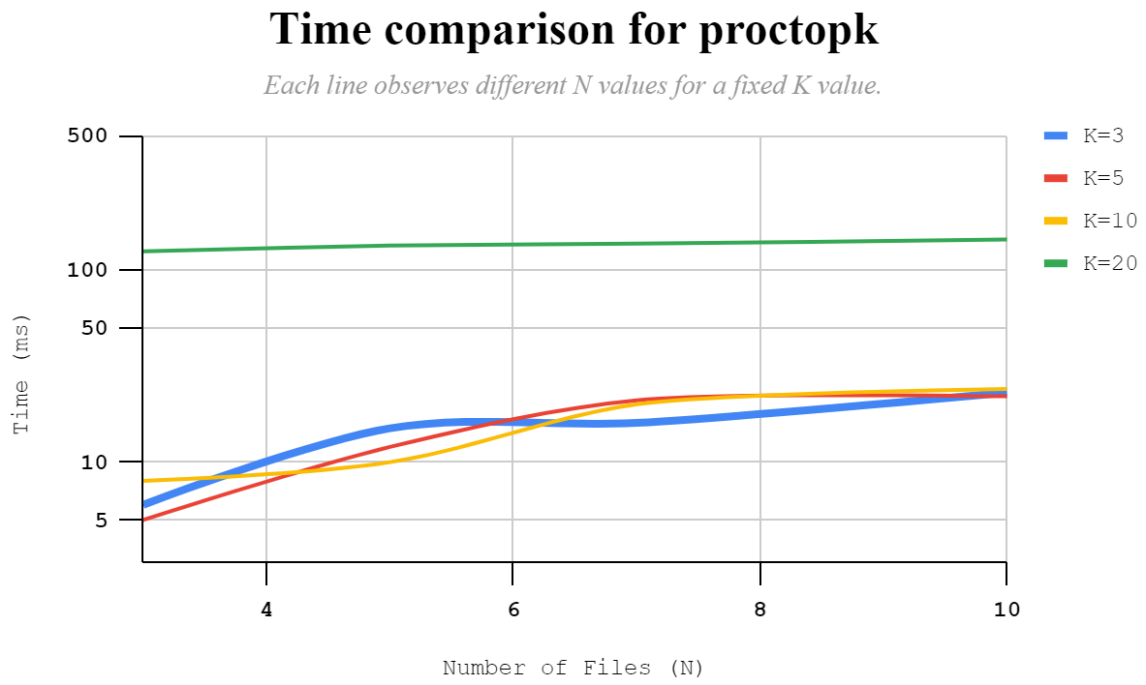

Fig. 1: Execution time comparison for proctopk for fixed K value.

In this part, we measure the execution time of the proctopk program for different fixed K values by changing the number of files (N). The execution time did not change too much when increasing the number of files. There are slight changes between time execution of

each line. It shows us, changing the number of files (it also determines the number of child processes) changes the execution time linearly.

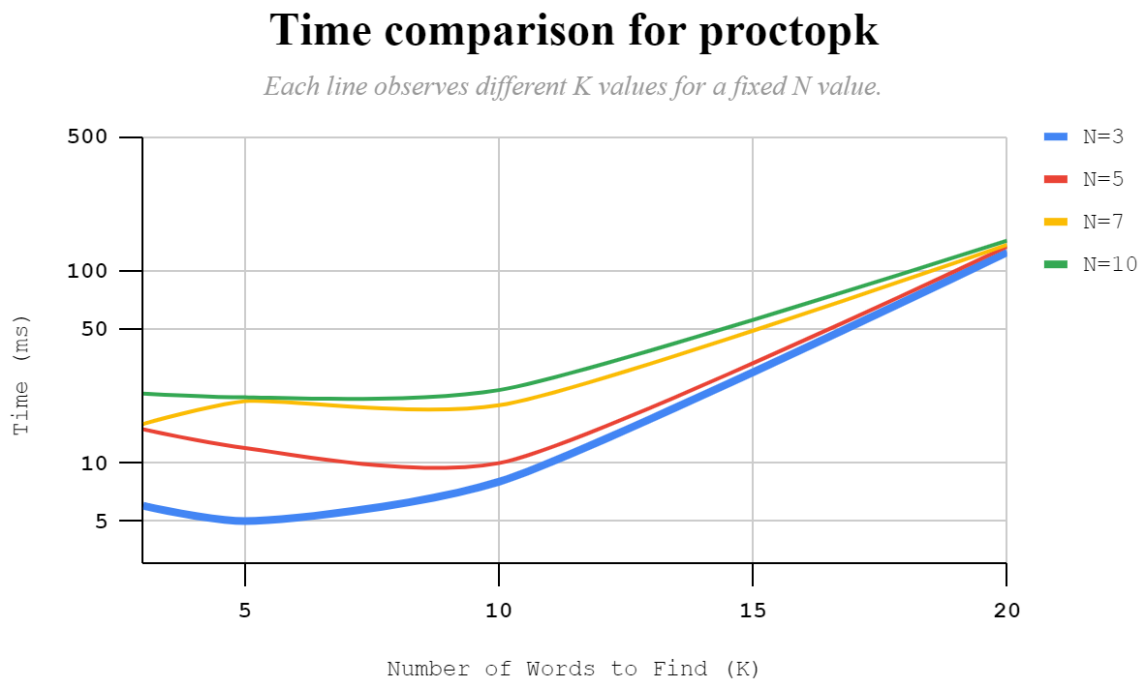

Fig. 2: Execution time comparison for proctopk for fixed N value.

In this part, the program execution time was measured for the constant N value and for the various K values for the proctopk program. When K value is increasing, the execution time is also increasing. Especially from 10 to 20, it increases a lot.

Threadtopk

Execution time(s) for threadtopk program

	N=3	N=5	N=7	N=10
K=3	0.006	0.019	0.024	0.023
K=5	0.006	0.009	0.018	0.022
K=10	0.006	0.010	0.022	0.024
K=20	0.006	0.013	0.023	0.023

Table 2: Execution times for threadtopk.

As it can be seen, the variance in the variable K does not affect the execution time mainly. The main factor is number of files, i.e., number of threads. This is because of the natural advantage of using threads in lieu of child processes, they are faster in general.

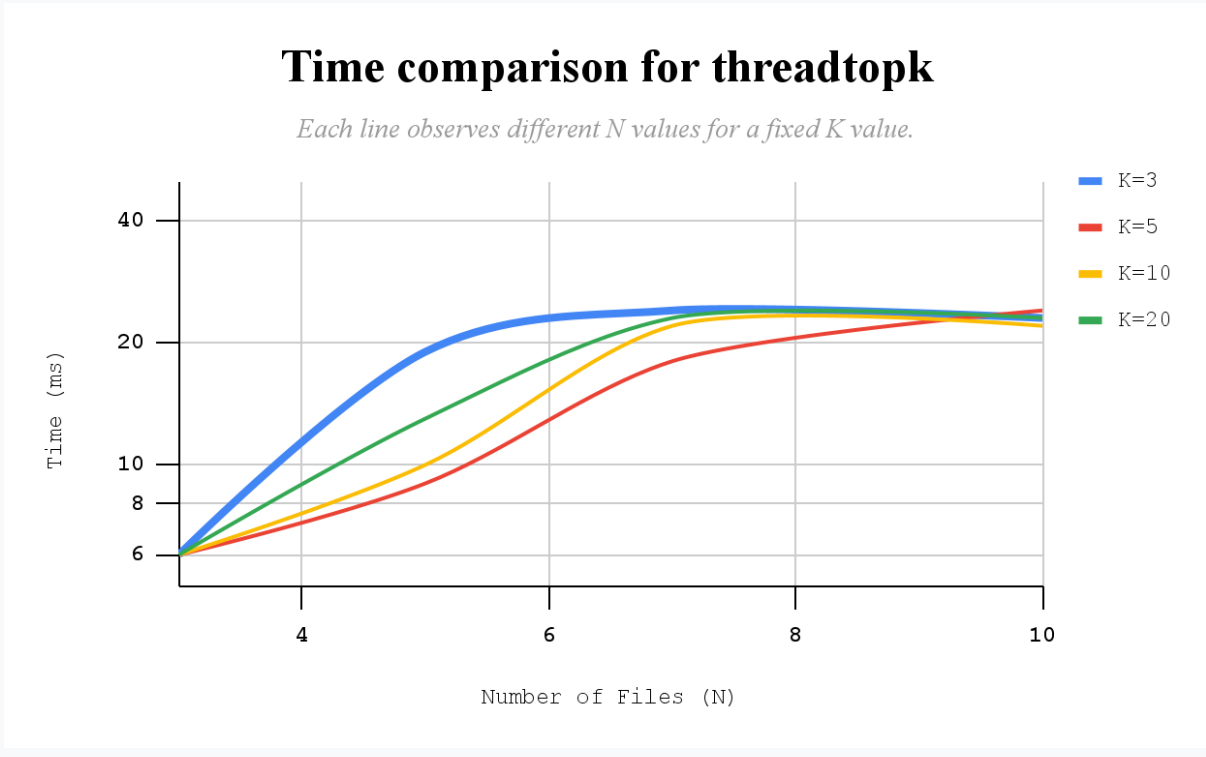

Fig. 3: Execution time comparison for threadtopk for fixed K value.

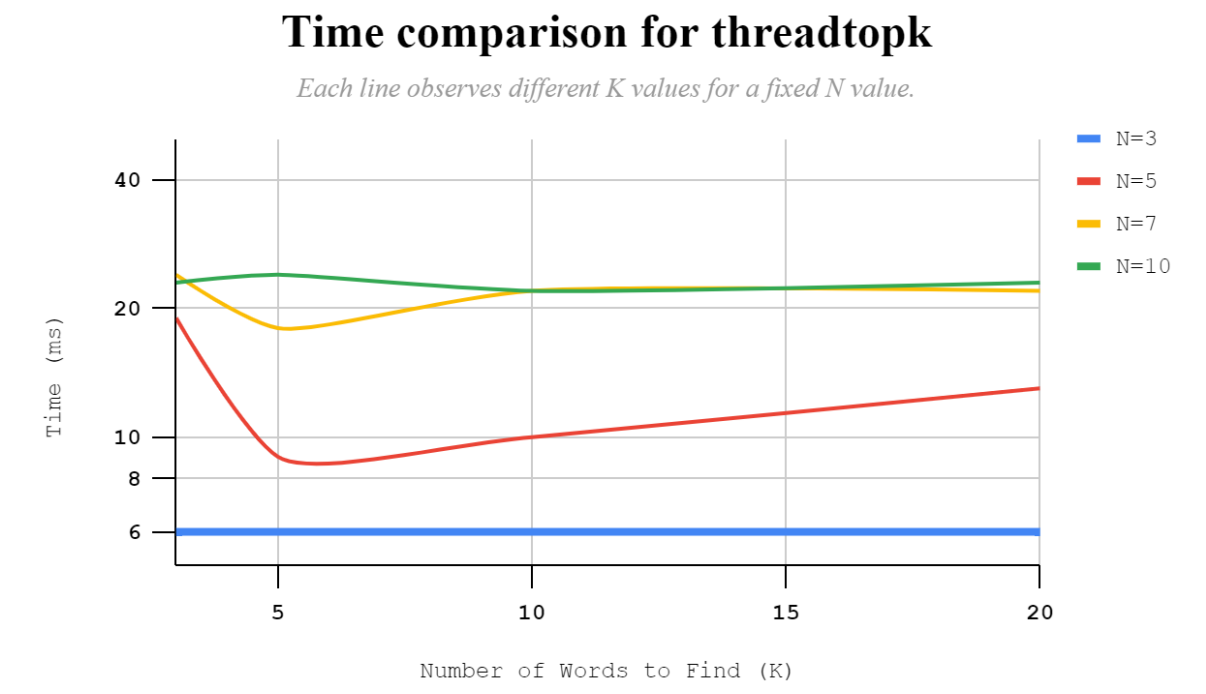

Fig. 4: Execution time comparison for threadtopk for fixed N value.

In this part, the program execution time was measured for the constant N value and for the various K values for the threadtopk program. When the K value is increasing, the execution time is increasing very slightly.