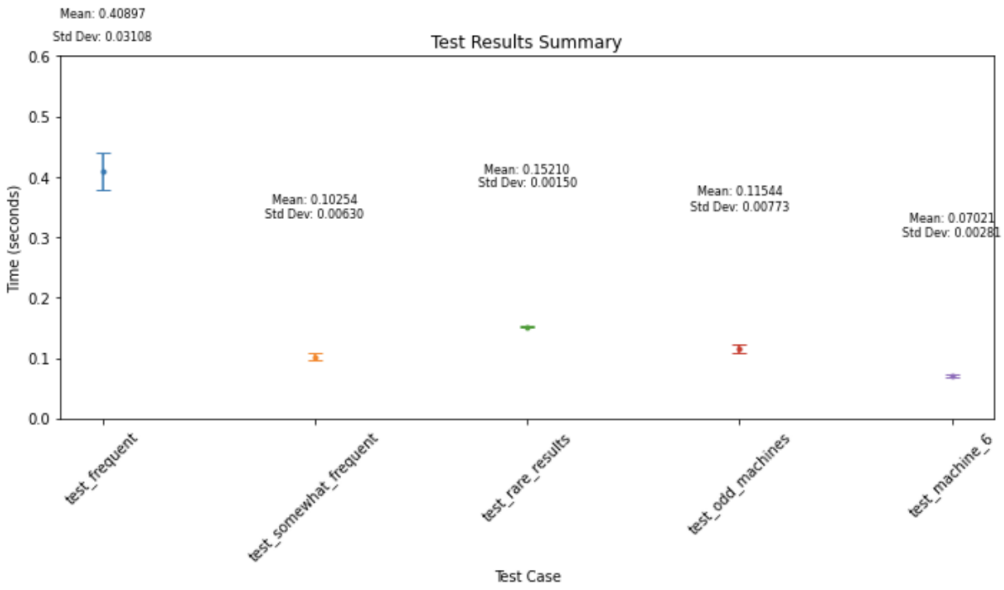
Georges Durand and Akshay Ghosh
Professor Gupta and Professor Ganesan
CS 425
10 September 2023

CS 425 MP 1 Report

Design: For this MP, we decided to use TCP sockets to transmit our data reliably and efficiently. We initially were using UDP, but then realized that reliability was much more important for our application than the speed of the transmission. We start the TCP receivers on each of the target machines, which receive packets from the client (the machine that requests the grep query). The results are then sent back to the sender and printed to standard output. We are also using threading to run the grep results over every machine in parallel as opposed to sequentially, which gives us a nice speed boost.

Testing: We are using Pytest to automate the testing of our code. On each machine, we generate a 500,000 line long log file with periodic occurrences of several different keywords based on rarity (i.e. “rare”, “somewhat frequent”, etc.) based on the number of the virtual machine (1-10). The results of the test queries are then validated against the known keyword placements.



Above are the results of running each individual test 100 times across four machines (5, 6, 7, and 8) on 60+ MB log files. Overall, our queries run relatively fast across all types of patterns and scale well when running over multiple machines due our use of threading.