Project 1: Basic exact pattern matching

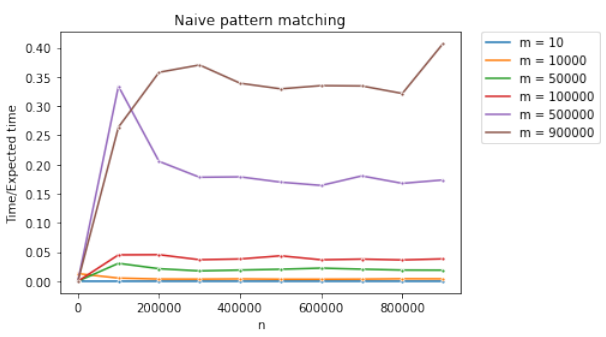
In this project, we have implemented a naive and a border search algorithm for finding patterns in strings.

Testing correctness

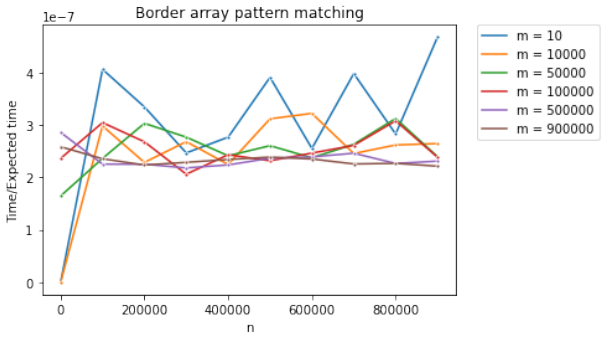
We tested the correctness of our programs by generating a test program that tests whether our programs can find matches in the beginning, end, at random positions, if it can find multiple hits, multiple overlapping hits, and how the program handle empty input files. The test program contains only 7 cases, so it might be too small though.

Testing running time

To verify the running time of our algorithms, we divided their actual running times with their expected running times (O(nm) and O(n+m) respectively), which results in a constant function. The data used for testing was FASTA and FASTQ files generated randomly. In the "worst" case we match the whole pattern at each position of the string, and the "best" case is when we don't find any match between the first character of the pattern and the string.



(a) Naive search algorithm



(b) Border search algorithm