

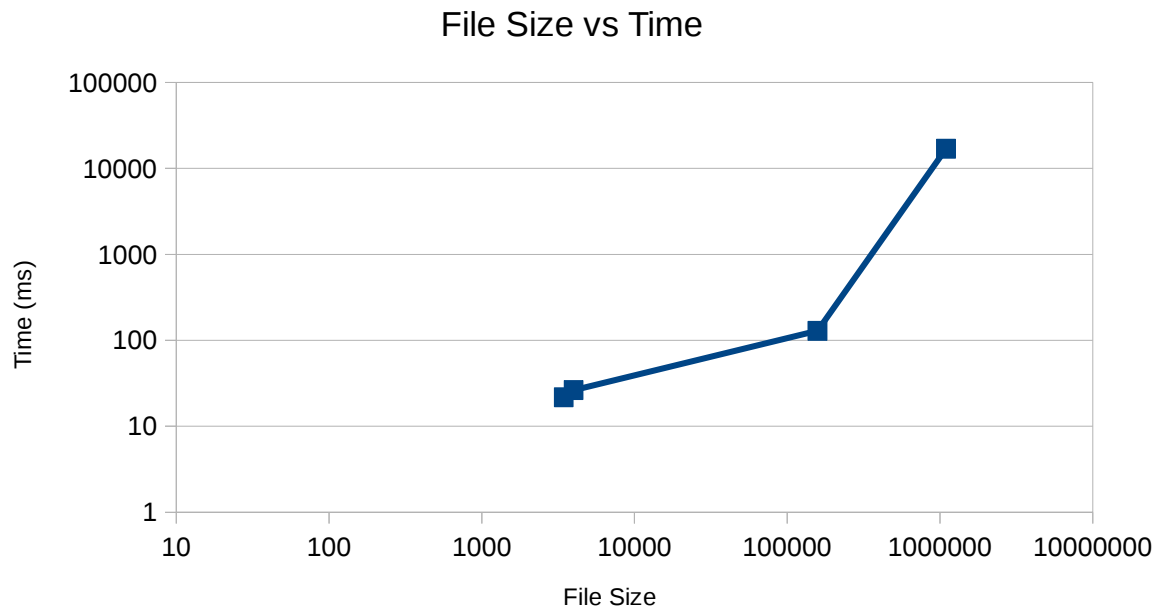
Project 2: McCreight's Suffix Tree Construction

System Configuration:

- 15.6 GB memory
- 2.6 GHz processor (x8)

Construction Time Complexity and Justification

Our time complexity was computed by running each input file 3 times and taking the average. The resulting output graph is as follows:

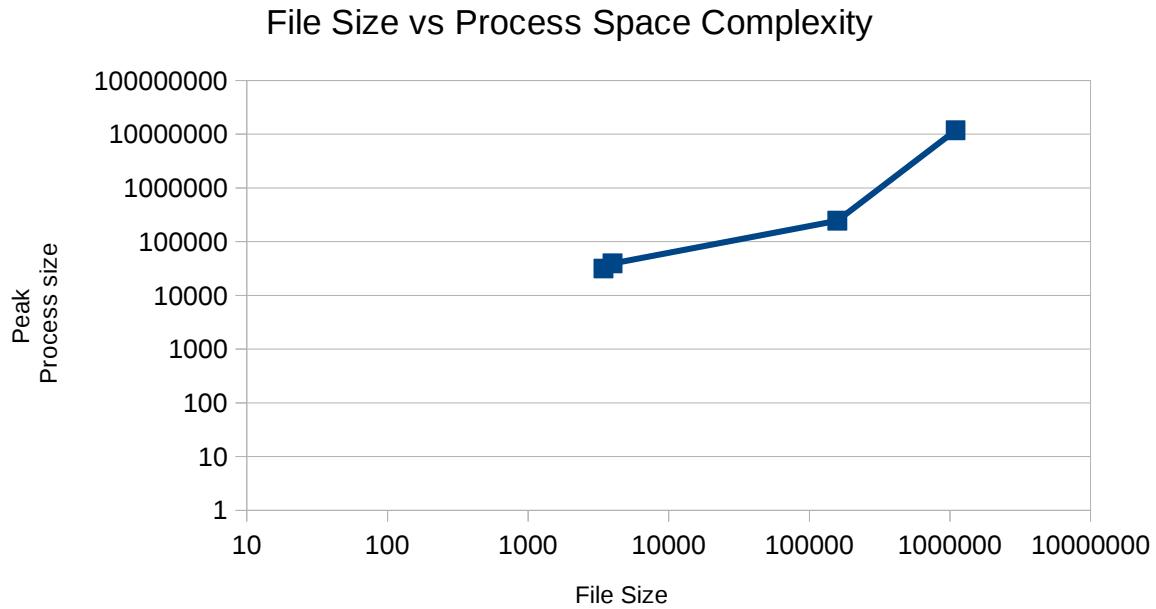


Graph 1: Input file size in bytes vs runtime in milliseconds.

We observed scaling that was close to linear for input files smaller than the tomato chloroplast gene, however, for larger files, the time complexity began to scale more poorly. We believe this is due to hardware limitations, as more resources were demanded.

Implementation Constant

To estimate the implementation constant, we measured the relative amount of memory required for each input file, and converted it to a size in bytes based on our system's specifications. We then plotted these points to a graph:



Graph 2: Input file size in bytes vs. Peak memory usage in bytes.

From Graph [2], we calculated the slope to estimate the implementation constant: 10.7.

Longest Repeating Section

During our depth-first search, we found the internal node with the largest depth, and saved a reference. Upon completion, we found the children id numbers for the deepest internal node, which are related to the starting position of the occurrences.

Banana

Locations: 4, 2

Length: 3

Mississippi

Locations: 5, 2

Length: 4

Opsin Human

Locations: 1818, 1811

Length: 18

Opsin Mouse

Locations: 2840, 2839

Length: 14

BRCA2

Locations: 7252, 9237

Length: 14

Tomato Chloroplast

Locations: 88151, 88131

Length: 48