Analysis Graphs

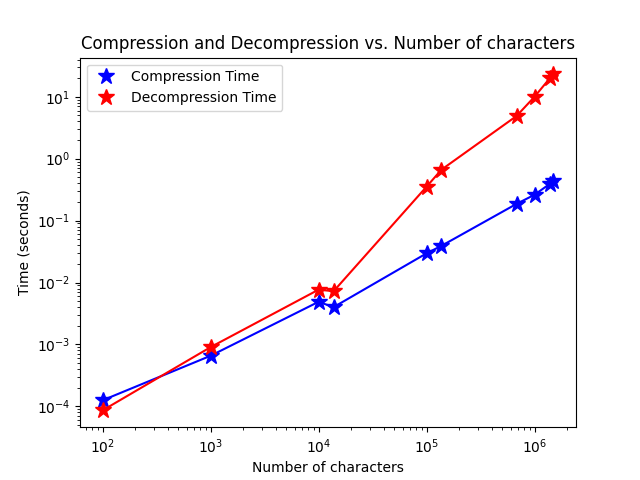


Figure : Time to compress and decompress text as a function of the number of characters in the file. Compression shows a linear relationship over time while decompression shows exponential. Compression is expected to be linear O(N) while decompression is expected to be O(NlogN). For decompression the difference between expectation versus output may be due to data loaded into memory, overhead or some other factor of code implementation.

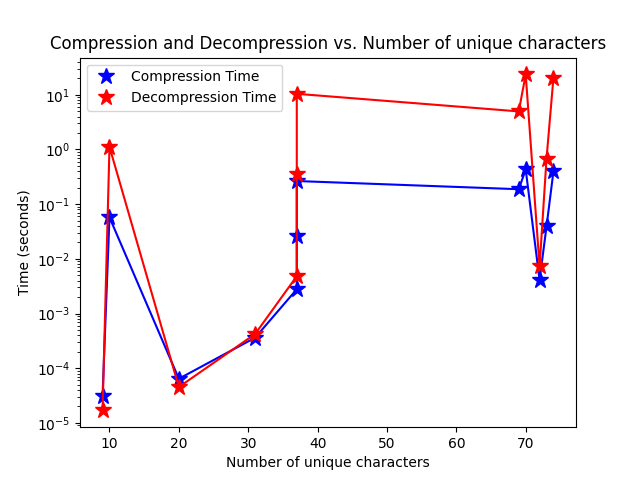


Figure : Time to compress and decompress as a function of the number of unique characters (which correlates with number of nodes in the Huffman tree). There is a general linear relationship between 0 and 37. The sharp increase in time is due to the fact that all the larger files have the same amount of characters as some of the smaller files. Therefore the sharp increase isn’t due to the number of unique characters but the number of characters in general.

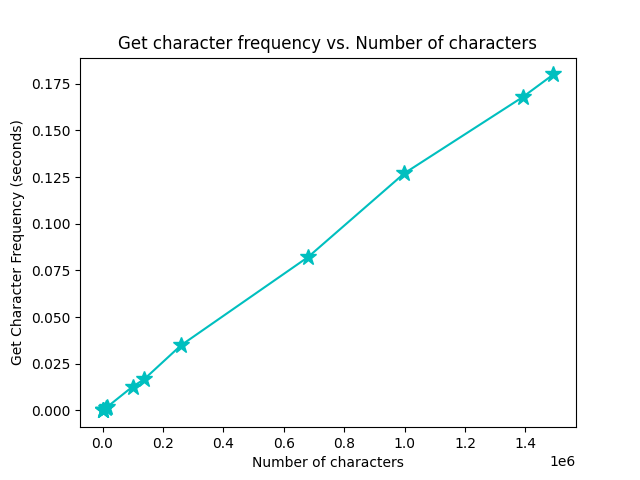


Figure : Time to obtain frequency of each unique character as a function of number of characters in the text file. This is an expected linear relationship and is a sanity check to make sure operations work as expected.

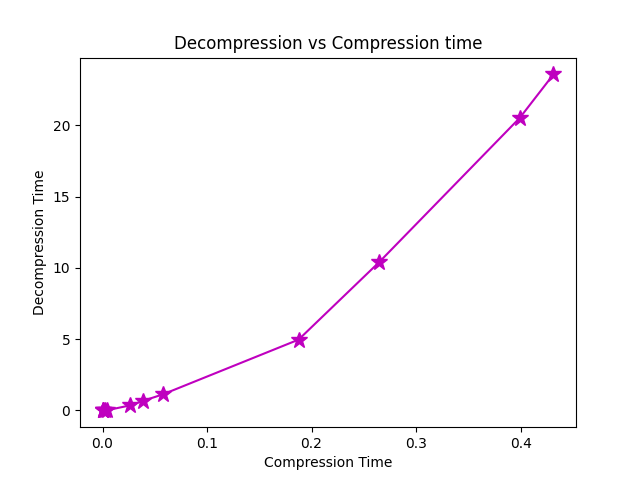


Figure : Decompression time as a function of compression time. This is a sanity check as the operation should be linearly related.

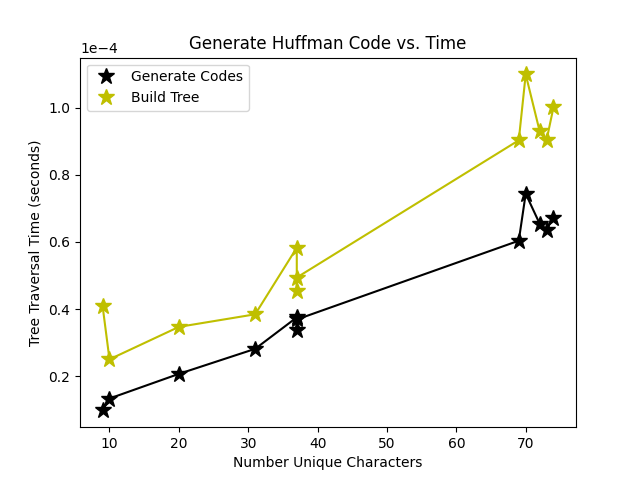


Figure : Time to generate Huffman codes (perform inorder traversal of tree) as a function of time.

Average Bits/Char

Mean=4.306

std=1.090

range=2.255-5.404

Compression (Size Reduction)

Mean=2.020

std=0.689

range=1.481-3.548