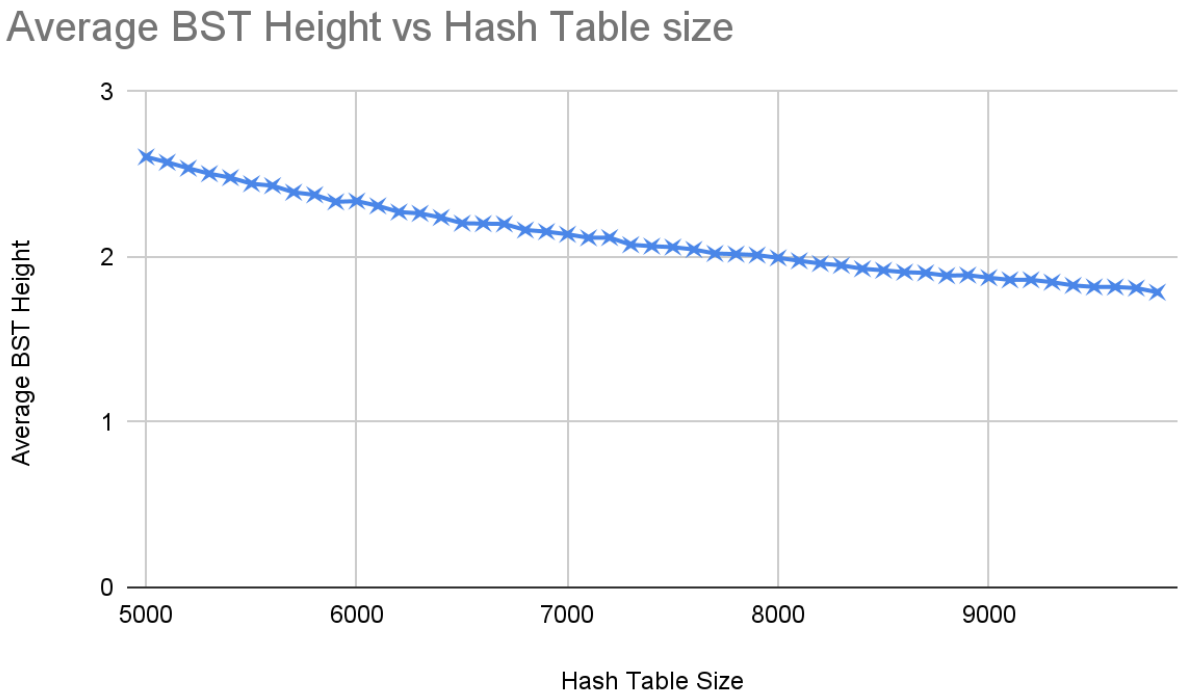
Assignment 7: The Great Firewall of Santa Cruz

WRITEUP.pdf

Nathan Ong

This program is a library that reads an input from stdin and scans for words that are in a premade list of "banned" words or a premade list of alternative words to use in place of other words. Depending on what is detected the program reacts accordingly with an appropriate message that describes the word crime that takes place and what words were detected that were part of the lists. The implementations included in this library include a bloom filter, binary search trees, a hash table, nodes, and bit vectors.

Listed below is a graph that compares the average binary search tree height to the selected hash table size. The Bloom filter size is its default size, 2^{20}, or 1048576.



The downward trend of the binary search tree height is evident just by looking at the graph. As the hash table's size increases, the average binary search tree height decreases. However, the rate at which the average height is decreasing is not uniform, with small variations between each point that vary. For example, in the charts below, the typical reduction in average tree height seems to be around 0.3 height units, but then you have parts like at hash table sizes 5900 and 6000, where the average tree height seems to increase ever so slightly before the pattern returns. This repeats at hash table sizes 7000 and 7100.

5600	2.430974
5700	2.391107
5800	2.37575
5900	2.333395
6000	2.337677
6100	2.309035
6200	2.271616
6300	2.263325
6400	2.238451

6800	2.162899
6900	2.152805
7000	2.136921
7100	2.11554
7200	2.116973
7300	2.073666
7400	2.064805

7100	2.11554
7200	2.116973
7300	2.073666
7400	2.064805
7500	2.059117
7600	2.04485
7700	2.020564
7800	2.016885
7900	2.011061
8000	1.994324

In the chart on the right, we can also see that after the hash table size exceeds 7300, the relation between the hash table size and the average binary search tree height changes. The average binary search tree height now reduces at a much slower and subtle rate, now reducing an average of around 0.1 height units compared to the previous typical reduction of 0.3 height units.

During testing, the average binary search tree height seems to be directly related to the hash table size, and as the size of the table increases, the drop in average height begins to level out eventually, settling around a height of 1, the lowest possible height value. This slow effect towards the average height leveling out is also present in the graph above, as the curve is beginning to curve more and more slightly as the size of the hash table increases. Changing the bloom filter size appears to have no effect on the average height of the tree.