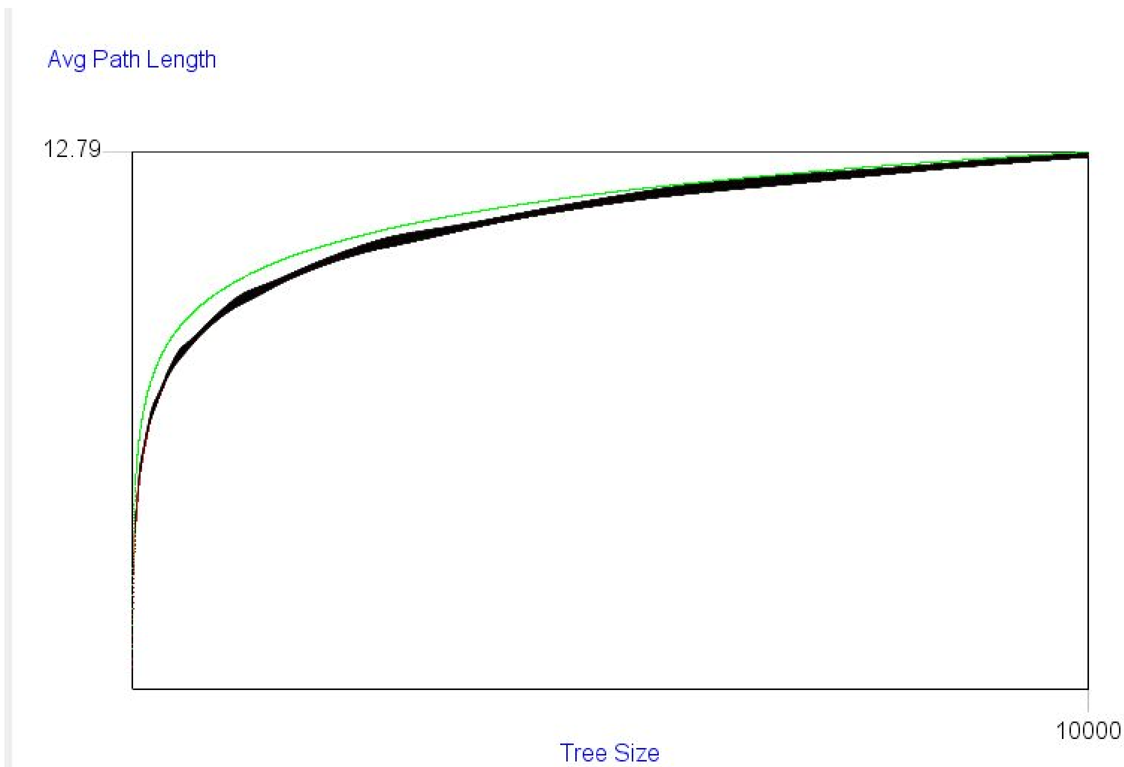
For all of the following graphs, the data was collected through a single run of 1000 trials per 10000 possible Tree Sizes. Three trials were done in total, one for each specific topic and all the data was printed to the text files, which I was not able to submit. If there is any doubt of data proof, I would be happy to provide the .txt files. The average of each specific data inquiry was calculated, as well as the standard deviation for each 1000 trial run. The results were plotted on each of the three graphs below. In every graph, the thin red line represents actual point values, the black is the standard deviation spread at every point, and the green line is the graph of log(N) - 0.5.

1. On Average Search Time

Average path length was calculated by creating a Red Black BST from random keys. After each creation, an array of the keys was stashed away, and they were systematically checked for their path lengths. The total distance of all path lengths was summed, and then divided by the tree size, this was then increased by one to receive an accurate metric of average path length to a random node.

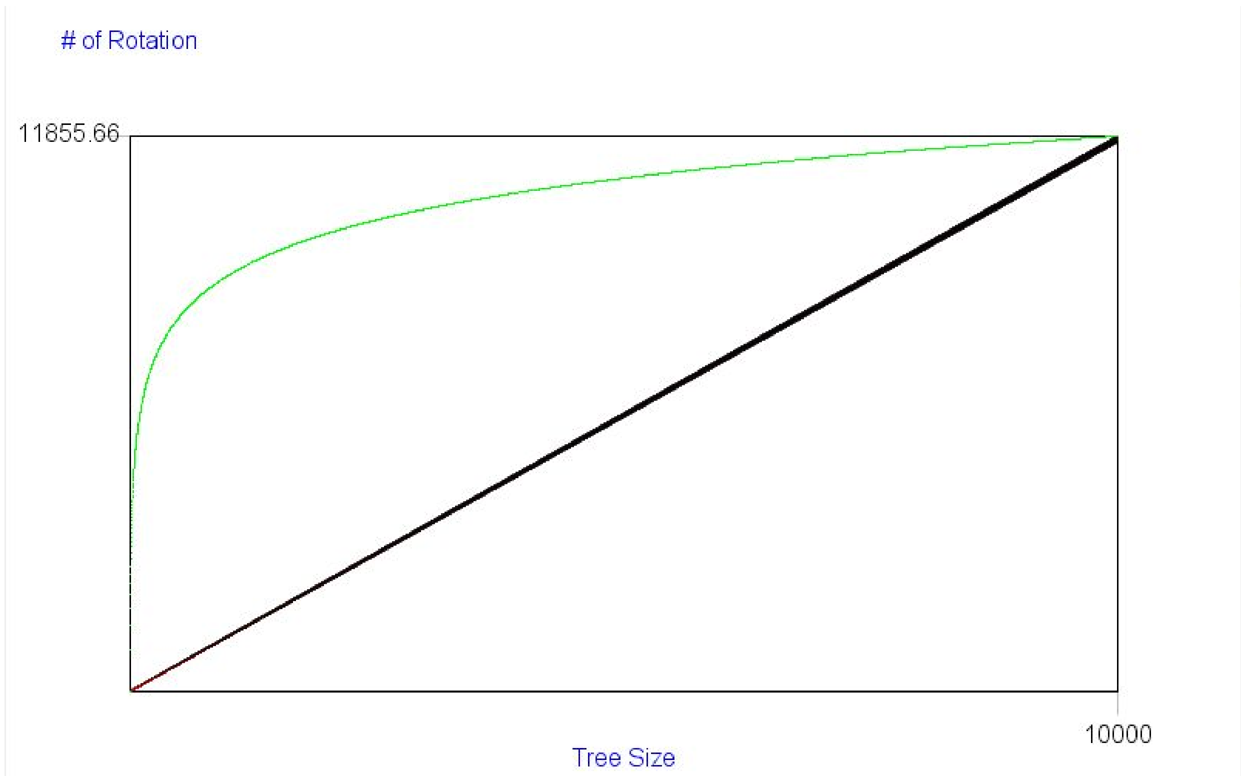
As far as the graph goes, the plot approached very close to 13 as the average path length, as predicted by the analysis shown in the book. My data was ever so slightly an undershoot of the predicted log(N) - 0.5 graph. However, as my data progressed with larger tree sizes, it approached very close to the log graph, with fluctuations in standard deviation all across the graph. The general trend looks pretty in line with the logistic representation.



2. On Rotations

Nothing previously suggested that the log(N) - 0.5 would be a good predictive measure for this particular assertion, and upon inspecting the data I think it is safe to say the relationship between # of Rotations and Tree Size is definitely linear. Unlike the other two relations, this graph is very distinctly linear, with the maximum number of rotations required sitting at an average of 11,855.66 at size = 10000 nodes. Standard deviation does not fluctuate very much across the board on this graph, and only slightly increases as the size gets larger.

It makes sense for this graph to be linear, as opposed to an exponential or logistic representation. Each tree is built from a bunch of smaller trees that have to use rotates in a Red-Black tree. At each tree size, there is a fixed range of potential rotations we are concerned with, and we can simply sum up values from this fixed range at each tree size until we hit our desired tree size.



3. On Tree Height

Tree height got no larger than an average of 17.34 even at a size of 10000. The top of the standard deviation predictions on this one kept pretty close in line with the expected graph of log(N) - 0.5. Due to what I can as many ranges of tree heights, the standard deviation on this graph for my data is much larger at every single point than the other two graphs. The general trend appears sensible, as each level of a BST, especially a Red-Black one produced a bounty of open potential Node location. It is not surprising to see this graph level off like the measure of average search length.