Mason Waters

Compare Three Trees Report

11/29/2019

After you perform your 500 (or more) searches:

What is the average number of comparisons it took for each tree?

Splay: 13.62

BST: 22.018

AVL: 14.243

What is the standard deviation of these averages?

Splay: 9.998

BST: 6.120

AVL: 9.896

Are these three averages statistically significant (i.e. are they statistically different from each other)? To answer this – you will have to do some statistics and look at p-values. You can use spreadsheet software to help you run a t-test or do it by hand: <https://www.wikihow.com/Assess-Statistical-Significance>



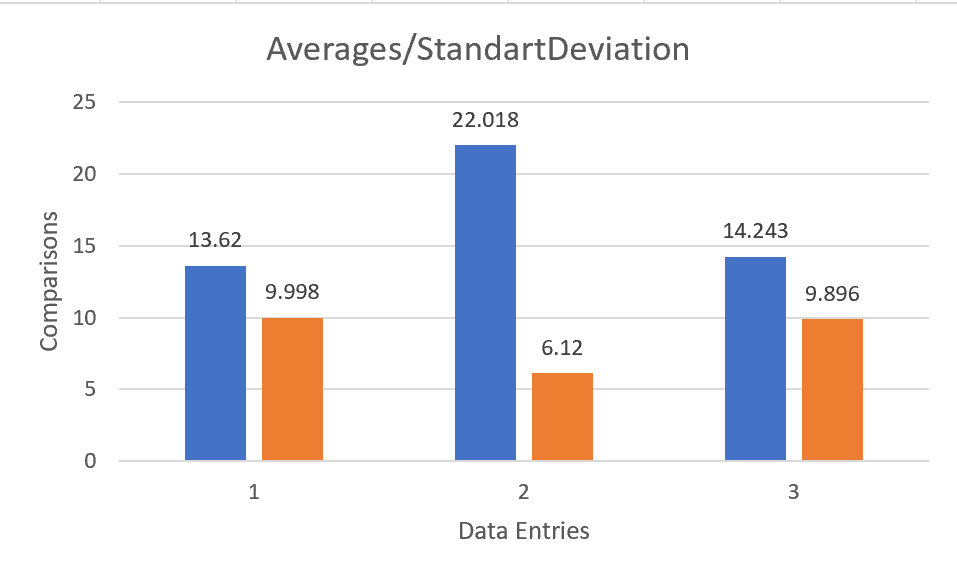
Write up your analysis:

So, to start off the report, what we did. We started with the Splay tree. We got some ideas on how to build it from the trees we made in class, the slides on Splay trees and the internet. It took a while to figure out, and we even encountered problems with the splay tree while working on other things. Then after splay tree was done, we built the Place class which wasn't too complicated because it was basically just getters and setters. We also decided to use the Place class as the class to create everything for the trees. So, it would create an object node that contained both the city state name from the file and then an array list of its associated zip codes. Lastly, the Trees class which we had do both the interaction with the user stuff, putting stuff from the file into the correct trees, and as well as getting our results. The class was originally just supposed to put stuff from the file into trees and ask the user questions/display information, but we also designed to put a method in to get us our results for the average and standard deviation because we decided it would be easier than creating a whole separate class for it (because we are lazy).

Now for what we found. The averages for each tree were Splay: 13.62, BST: 22.018, and AVL: 14.234. The standard deviations for each tree were Splay: 9.998, BST: 6.120, and AVL: 9.896. The TTest values are all in the image above. We made three comparisons to check the statistical significance between all the trees. What we found was that the AVL and BST trees are not statistically significant to each other (0.00), the AVL and Splay trees are statistically significant to each other(1.0587x10^-38), and the Splay and BST trees are not statistically significant to each other (0.00). The results could be impacted by the order in which things are searched for in really only one way. It would be the splay tree. Reason being is that if you search a splay tree, the searched for item gets splayed to the root. And if you search for it again and again after that, it will continue to be the root so the searching won’t take much. This would drive the average number of searches down.

Towards the bottom of the report you will notice two bar graphs. The top one is the averages/standard deviation bar graph. You can find this file in the zip file and it is named **“AveragesAndStandardDeviation**”. The averages are the blue bars and the Standard Deviations are the orange bars. The output file that has each place and number of comparisons associated with it is called **“CompareNumbers.txt”**. On each line of this file there is a City, State, and then number of comparisons for Splay, BST and AVL in that order. The bar graph on the bottom is associated with that file. It wasn’t necessary for this assignment, but I figured it would be a good visual.

The splay tree, which has a run time of O(log n) had a run time that associated with that pretty well. The splay trees average number of comparisons was the lowest of all the trees. The BST which has a run time of O(n) which also associates with these values well because it had the worst run time with the highest average number of comparisons. And finally, the AVL tree which has a run time of O(log n) also associated well because the average number of comparisons was similar to the Splay tree. I would say that all the actual results associated with the theoretical results really well.

In conclusion, all the trees performed about as well as we expected them to (according to the theoretical values). The BST was the worst tree with an average of 22.018 comparisons. The AVL and the Splay trees were the best running trees with similar averages of comparisons. But the overall best running tree was the splay tree with the average of 13.62 comparisons vs the AVL which was 14.243 comparisons.

