

BSTs and AVL - Tree Analysis

Dr. Mark R. Floryan

August 20, 2019

1 SUMMARY

The goal of this homework is to write a report analyzing the use of a binary search tree versus an avl tree. For the first time, we will be asking to find a dataset on your own to use for your analysis.

You will perform an experiment by doing the following:

1. Find a large corpus of data. I would recommend a large piece of text. Perhaps a very long article or a an electronic book. You may share datasets with each other if you find good ones.
2. Take your file and insert every element (e.g., word from the book) into both a binary search tree and an avl tree. Use a timer to see how long it takes to insert all of the elements into the avl versus the bst.
3. Call the find method on every word in your dataset. Time how long this operation takes in each data structure total.
4. Print out the height of each tree.
5. Now, repeat the experiment, but this time don't use a real dataset. Use random numbers from 1-10000. Do you notice any differences?

6. Write a report summarizing and analyzing your findings. Which data structure was faster.
7. **FILES TO DOWNLOAD:** None
8. **FILES TO SUBMIT:** TreeAnalysis.pdf

1.1 REPORT

Summarize your experiment and your findings in a report. Make sure to adhere to these general guidelines:

- Your submission **MUST BE** a pdf document. You will receive a zero if it is not.
- Your document **MUST** be presented as if submitted to a professional publication outlet. You can use the [template](#) posted in the course repository or follow [Springer's guidelines for conference proceedings](#).
- You should write your report as if it is original novel research.
- The grammar / spelling / professionalism of this document should be sound.
- When possible, do not use the first person. Instead of "I ran the code 60 times", use "The code was executed 60 times..."

In addition to the general guidelines above, please follow the following rough outline for your paper:

- **Abstract:** Summarize the entire document in a single paragraph
- **Introduction:** Present the problem, and provide details regarding the two strategies you implemented.
- **Methods:** Describe your methodology for collecting data. How many method calls, how many executions, how you averaged things, etc.
- **Results:** Describe your results from your execution runs.
- **Conclusion:** Interpret your results. Which data structure was faster? Did the results change from a real dataset to a random one? If so, why? How different was the heights of the different trees? Why?

Lastly, your paper MUST contain the following things:

- A table (methods section) summarizing the experiments and how many execution runs were done in each group.
- At least one table (results section) summarizing the results of the experiment(s).
- Some kind of graph visualizing the results of the table from the previous bullet.