A Comparison of Practical Insertion Times for Scapegoat and Red-Black Trees

I. Introduction

While scapegoat trees and red-black trees are similar in their self-balancing nature, they differ slightly in their theoretical time complexities. Specifically, the worst-case time complexity of insertion for scapegoat trees is linear, while that of red-black trees is logarithmic. However, scapegoat trees maintain amortized logarithmic time complexity for insertion and additionally include an alpha value between 0.5 (inclusive) and 1 (exclusive) where values closer 1 result in less strict balancing—i.e. rebalances caused by insertion are less frequent—prioritizing insertion time over search and deletion times.

With this in mind, this project compares the practical insertion times for scapegoat trees and red-black trees.

II. Method

In total, 6 tree configurations were used: a red-black tree, a scapegoat tree with an alpha value of 0.5, another with an alpha value of 0.6, another of 0.7, another of 0.8, and another of 0.9. For each configuration, the times it took to insert 250,000 keys, 500,000 keys, and 1,000,000 keys were measured where the keys were random integers ranging from 0 (inclusive) to 1,000,000 (exclusive). For each configuration, this was repeated 30 times with a different set of random keys.

The project used a custom scapegoat class implemented in Python, following the original paper by Galperin and Rivest that introduces scapegoat trees as a guide, and a publicly-available red-black tree class implemented in Python courtesy of *Programiz*.[1]

III. Results

The table below shows the sample means for each configuration.

Mean Insertion Times (ms)			
Configuration	250,000 keys	500,000 keys	1,000,000 keys
Red-black tree	2192	3462	6811
Scapegoat tree, α = 0.5	6916	16048	38979
Scapegoat tree, α = 0.6	1251	2779	6499
Scapegoat tree, α = 0.7	1182	2655	6139
Scapegoat tree, α = 0.8	1169	2647	6136
Scapegoat tree, α = 0.9	1183	2659	6155

[1] "Red-Black Tree," *Programiz*, www.programiz.com/dsa/red-black-tree.

The mean insertion times for the red-black tree are significantly faster compared to those for the strictest scapegoat tree—i.e. the scapegoat tree with an alpha value of 0.5. However, the times for the red-black tree are surprisingly somewhat slower than those for the more relaxed scapegoat trees—i.e. those with higher alpha values such that rebalances caused by insertions are less frequent, making insertion times faster on average at the cost of making search and deletion times slower on average. Moreover, the mean insertion times for these more relaxed scapegoat trees are interestingly all extremely similar, despite their alpha values ranging from 0.6 to 0.9.

The table below shows the sample standard deviations for each configuration.

Standard Deviation of Insertion Times (ms)			
Configuration	250,000 keys	500,000 keys	1,000,000 keys
Red-black tree	260	247	268
Scapegoat tree, α = 0.5	279	393	865
Scapegoat tree, α = 0.6	8	16	33
Scapegoat tree, α = 0.7	17	27	54
Scapegoat tree, α = 0.8	20	50	99
Scapegoat tree, α = 0.9	26	52	102

Similarly, the insertion times for the red-black tree are more consistent than those for the strictest scapegoat tree. However, the more relaxed scapegoat trees are significantly more consistent in insertion times compared to the red-black tree. Additionally, it is notable that the standard deviations increase as the alpha value increases from 0.6 to 0.9, though not majorly.

IV. Discussion & Conclusion

The results suggest that relaxed scapegoat trees—i.e. ones with an alpha value of around 0.6 or greater—perform better than red-black trees in terms of average insertion time, despite having worst-case linear time complexity for insertion. Additionally, these relaxed scapegoat trees are significantly more consistent in their insertion times compared to red-black trees. Relaxed scapegoat trees, however, are theoretically slower on average in search and deletion times such that they are not universally preferable over red-black trees. However, in situations where insertions are more frequent and consistency in insertion time is valuable, the results suggest that relaxed scapegoat trees are preferable.

Additionally, the results suggest that relaxing a scapegoat tree beyond an alpha value of around 0.6 does not significantly improve insertion time and does in fact decrease consistency, suggesting a possible optimal alpha value of around 0.6 that improves insertion time and consistency, while limiting the tradeoff in search and deletion times.