Red-Black Tree Performance Validation Report

Introduction

This report presents the results of a performance validation conducted on a red-black tree set implementation. The primary objective was to confirm that the ‘insert’ operation adheres to the theoretical runtime complexity of Θ(n log n), a characteristic of balanced binary search trees.

Method

- The performance evaluation was automated through the program’s main.ml that measures the time taken to insert a series of distinct elements into an initially empty red-black tree. The experiment was designed to record the insertion times for varying sizes of the dataset, denoted by N.
- The program captured time measurements for inserting N distinct elements, starting with N = 1000 and exponentially increasing to N = 1,000,000. A total of 10 measurements were taken for each value of N to account for variability, with the median time reported to minimize the impact of outliers.
- The resulting data was outputted to a CSV file and then transferred to Google Sheets for analysis. Each entry in the CSV file included three columns: N (the number of elements), Time (the median time taken to insert N elements), and N log N (the product of N and the logarithm of N).

Results

The following table summarizes the data collected during the testing phase:

N	Time	N log N
1000	0.00030899	6907.76
10000	0.00364113	92103.4
100000	4.28E-02	1.15E+06
1000000	5.18E-01	1.38E+07

Using Google Sheets' LINEST function, a linear regression analysis was performed on the dataset with ‘N log N’ as the independent variable and ‘Time’ as the dependent variable. The analysis yielded a high r^2 value of 0.99999990289, which is exceptionally close to 1, indicating a very tight fit of the time data to the model ‘Time = a * (N log N) + b’.

R^2 Value

The r^2 value obtained from the linear regression suggests that the performance of the ‘insert’ operation closely follows the expected $\Theta(n \log n)$ runtime complexity. The high degree of correlation between ‘N log N’ and ‘Time’ supports the hypothesis that the implementation is efficient and likely free from significant performance bugs.

Challenges

The experiment was conducted with careful attention to mitigating factors that could introduce noise into the time measurements, such as system load and power settings. The main challenge encountered was ensuring the environment was consistent across all measurements, which was addressed by taking multiple samples and using the median value.

Conclusion

The experimental results validate the theoretical predictions for the performance of the red-black tree ‘insert’ operation. The r^2 value of nearly 1 confirms a strong linear relationship between ‘N log N’ and the insertion time, consistent with the expected logarithmic behavior of a well-balanced red-black tree.