CS231: Project 6

Word Frequency

Parth Parth | CS231L-A | Dr. Alan Harper

**Abstract**

In the project, a Generic Binary Search Tree has been created and implemented. This is specially useful for the use case of this project as it allows us to store and manipulate data in a more efficient way than a list. This data structure has then been used to analyze Reddit comments from years 2008 – 2015. The results have then been tabulated and analyzed.

The analysis is as follows:

**Extension 1**

In this extension, I have implemented a *remove(K key)* method that takes in a key of type *K* and removes and returns the node where the key was found if it exists, otherwise it returns null. It uses the following logic:

1. If node has no children, just remove the reference to it in the parent.
2. If it has only one child (or subtree), replace the reference to that node with a reference to the subtree root in the parent
3. If node has both children
   1. Find minimum in right subtree
   2. Copy the value to the target node
   3. Remove the minimum in the right subtree

The following image shows a demonstration:

Text

Description automatically generated

**Extension 1**

In this extension, I have edited the *toString()* method to return a *String* that contains a breadth-first traversal of the tree i.e. from left to right, level-by-level.

The following image demonstrates this:

Text

Description automatically generated

The first two lines are the inorder traversal of the BST before and after removing the value 6 from the tree.

**References/Acknowledgements**

The algorithms for the remove method for extension 1 and the level-by-level traversal that I have employed were covered by Dr. Al Madi in class.