REPORT

Advanced Algorithms Assignment – 3 - B-Tree

Rahul R Bharadwaj – 01FB15ECS366

To implement a B-Tree on disk, with algorithms given in CLRS textbook.

Problem Definition:

Implement with one of the options for each policy:

1. Insertion and deletion policy:
   -Logical deletion: mark the node as deleted get the new node for insertion: from the end of the array note that the deleted nodes become garbage.
   -Keep a free list of nodes get the new node for insertion from the free list on deletion, add the node to the free list.

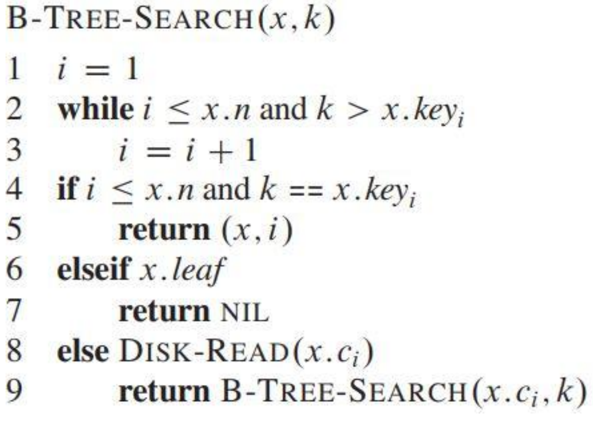
2. Fixed and varied sized records policy:
   -Records could be fixed size or variable size. If the records are of variable size, you require one more level of indirection.

3. Key policy:
   -Key could be a single field, could be a composite field (combination of fields).
   -Key could be of any type.
   -Key comparison (predicate) could be flexible.

4. Number of passes
   -You may implement a single pass or a two pass policy for insertion and deletion.

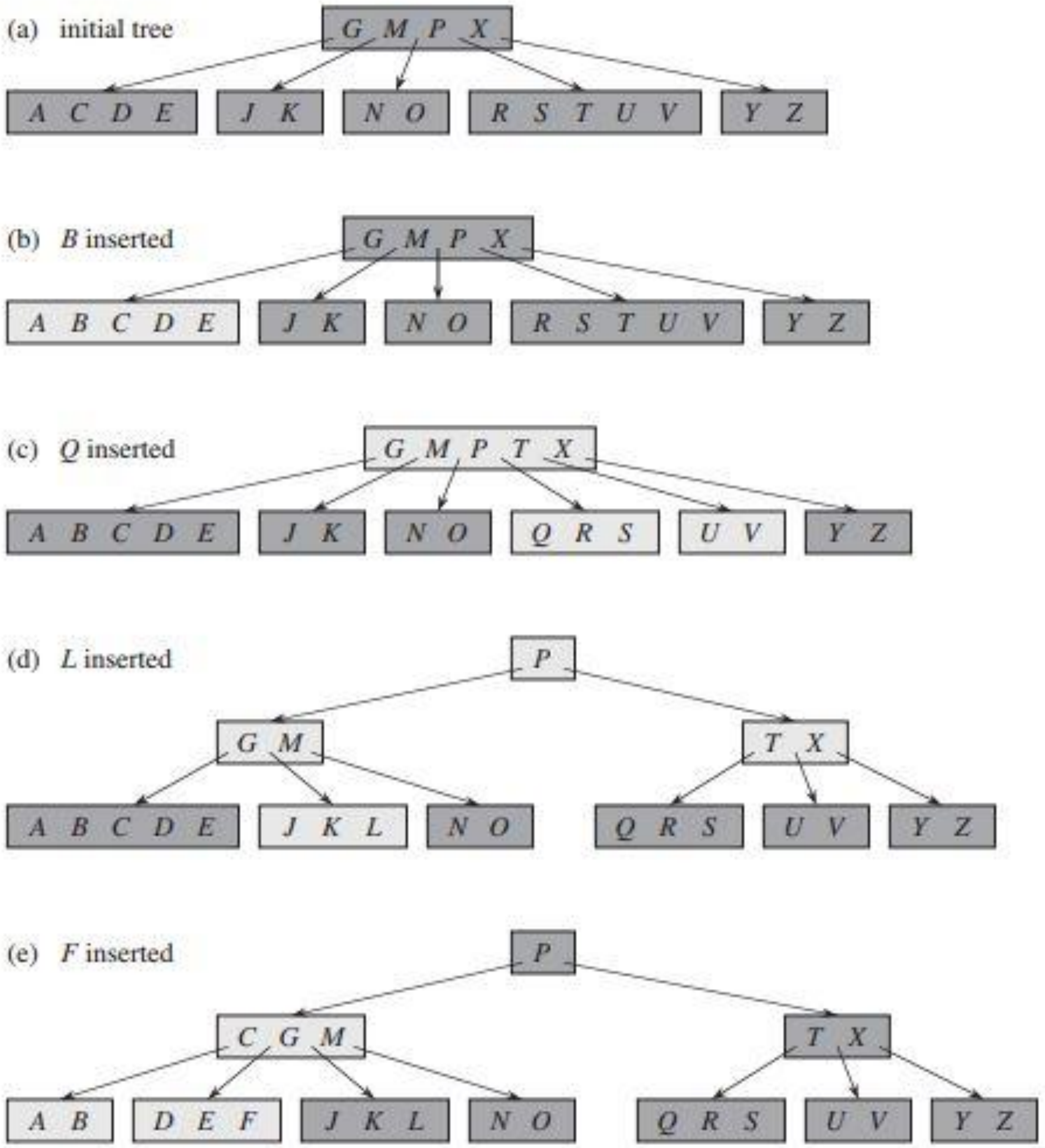
Approach:

For Search: The approach taken was similar to the algorithm given CLRS textbook. Array indices were not the same as the following algorithm uses 1 based indexing.

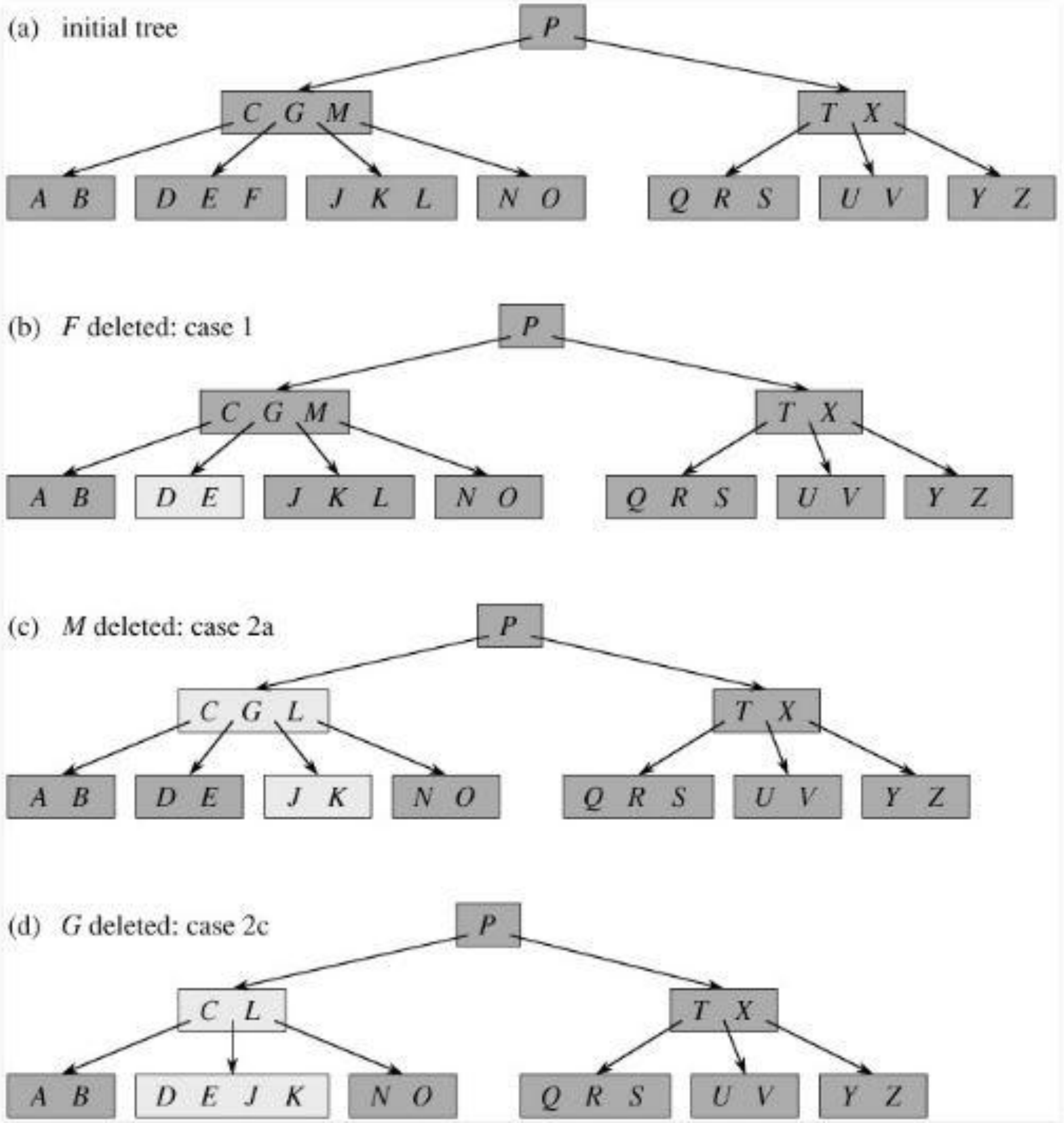


```
B-TREE-SEARCH($x, k$)
$i = 1$
while $i \le x.n$ and $k > x.key_i$
    $i = i + 1$
if $i \le x.n$ and $k == x.key_i$
    return $(x, i)$
elseif $x.leaf$
    return NIL
else DISK-READ($x.c_i$)
    return B-TREE-SEARCH($x.c_i, k$)
```

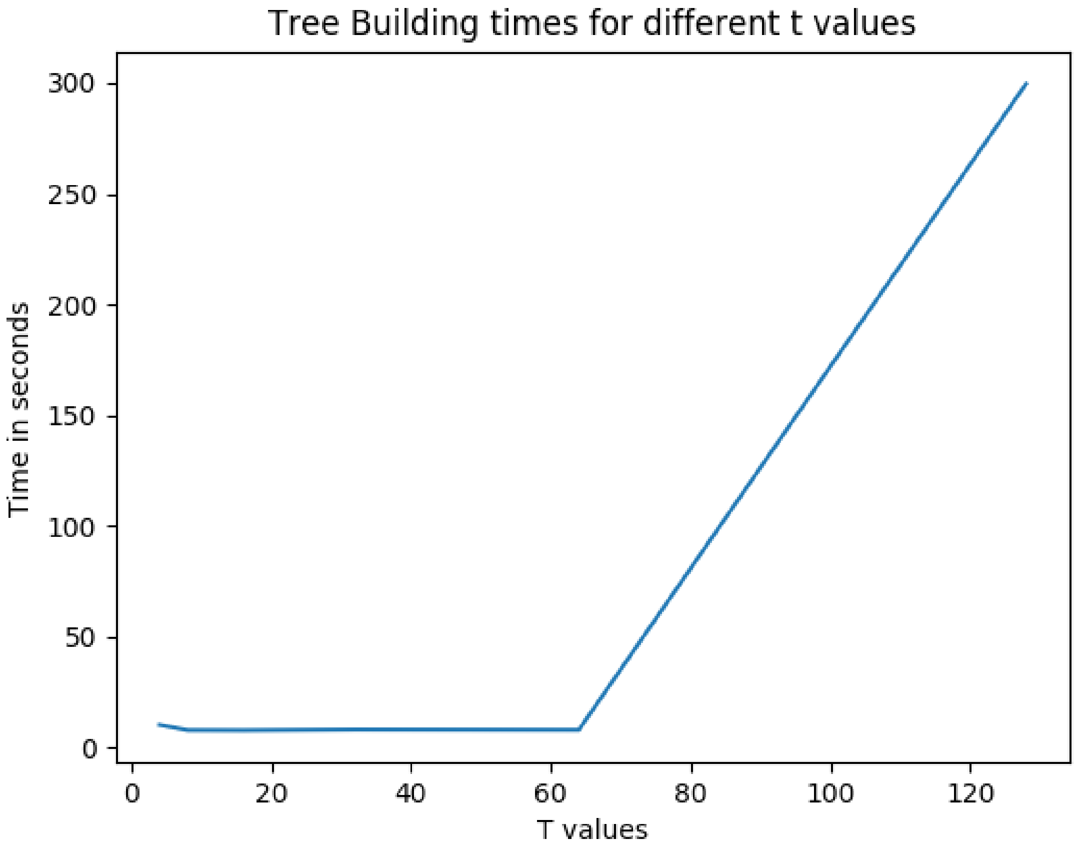
For Insertion: The approach is again based on algorithm given in CLRS textbook. The following image was taken from the same textbook.

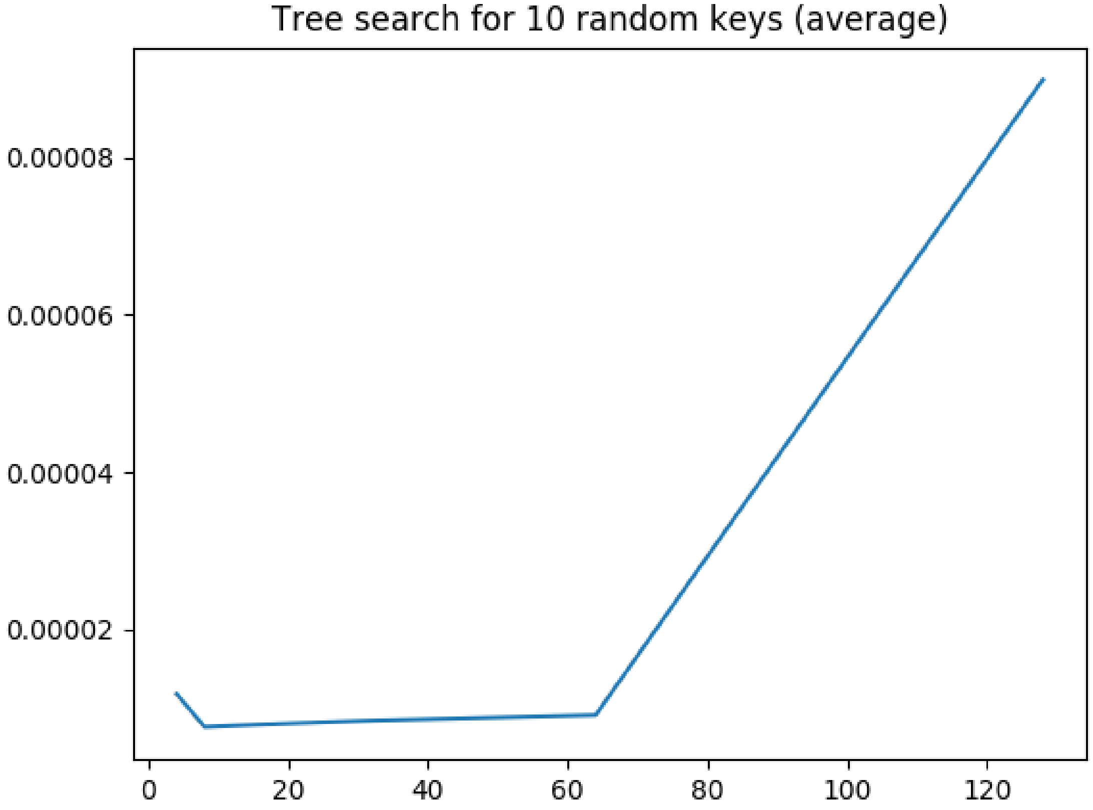


For Deletion: The approach is again based on algorithm given in CLRS textbook. The following image was taken from the same textbook.



Results and Graphs:





Result Analysis:

Above results indicate that a t value of a power of 2 between 20 to 40 seems ideal. One of those ideal values also indicate ‘32’.
We can argue our assumption, based on the following facts:

The goal of a b-tree is to minimize the number of disk accesses. If the file system cluster size is 4k, then the ideal size for the nodes is 4k. Also, the nodes should be properly aligned. A misaligned node will cause two clusters to be read, reducing performance.

Source: https://stackoverflow.com/questions/2678559/b-tree-node-sizing

Average Build times, for t=32

Problem Type	Average time, in seconds
Insert	0.097498
Build	8.044751

Acknowledgements:

I would like to thank our professors Mr. NS Kumar and Mr. Channa Bankapur, for providing valuable suggestions, important codes and this assignment in general. This assignment introduced me to the various ways in which modern file systems are implemented. The delete functions are a direct implementation based from geeksforgeeks.