**Average Case Analysis of BST Operations**

**RESULT**

If we insert *n* random elements into an initially empty BST, then the average path length from the root to a node is O(log n)

* Note that the BST is formed by insertions only. Obviously the tree so formed need not be complete.
* We shall assume that all orders of n inserted elements are equally likely. This means that any of the n! permutations is equally likely to be the sequence of keys inserted.

|  |  |  |
| --- | --- | --- |
| Let |  |  |
| P(n) | = | average path length in a BST with n nodes |
|  |  | (average number of nodes on the path from the root |
|  |  | to a node, not necessarily a leaf) |
|  |  |  |
| Let |  |  |
| a | = | first element inserted. This will be the root of |
|  |  | the BST. Also this is equally likely to be the first, second |
|  |  | . . . , *i*th, . . . , or nth in the sorted order of the |
|  |  | n elements. |

Note that P(0) = 0 and P(1) = 1. Consider a fixed *i*, 0 $ \leq$*i$ \leq$n* - 1. If *i* elements are less than a, the BST will look like in Figure [4.16](http://lcm.csa.iisc.ernet.in/dsa/node92.html#fig:bst1).   
 

|  |
| --- |
| A typical binary search tree with *n* elements |
| \begin{figure}\centerline{\psfig{figure=figures/Fbst1.ps}}\end{figure} |

$ \mbox{$\bullet$}$

Since all orders for the *i* small elements are equally likely and likewise for the (*n* - *i* - 1) larger elements,

Average path length in the

* left subtree = P(*i*)
* right subtree = P(*n* - *i* - 1)

$ \mbox{$\bullet$}$

For a fixed *i*, let us compute the average path length of the above tree.

Number of probes if the element a is being sought = 1

Average number of probes if an element from the LST is sought = 1+P(*i*)

Average number of probes if an element from the RST is sought = 1 + P(*n* - *i* - 1)

Probability of seeking any of the n elements = $ {\frac{1}{n}}$

Thus, average path length for a fixed *i*

|  |  |  |  |
| --- | --- | --- | --- |
|  | = | $\displaystyle {\frac{1}{n}}$$\displaystyle \left\{\vphantom{1 + i (1 + P(i)) + (n-i-1) (1 + P(n-i-1))}\right.$1 + *i*(1 + *P*(*i*)) + (*n* - *i* - 1)(1 + *P*(*n* - *i* - 1))$\displaystyle \left.\vphantom{1 + i (1 + P(i)) + (n-i-1) (1 + P(n-i-1))}\right\}$ |  |
|  | = | 1 + $\displaystyle {\frac{i}{n}}$*P*(*i*) + $\displaystyle {\frac{n-i-1}{n}}$*P*(*n* - *i* - 1) |  |
|  | = | *P*(*n*, *i*),   say. |  |

$ \mbox{$\bullet$}$

Observe that *P*(*n*) is given by

*P*(*n*) = $\displaystyle \sum_{i=0}^{n-1}$  Prob{LST has $i$ nodes }*P*(*n*, *i*)

Since the probability that the LST has *i* elements which is the same as the probability that *a* is the (*i* + 1th element (where *i* = 0, 1,..., *n* - 1) = $ {\frac{1}{n}}$, we have

|  |  |  |  |
| --- | --- | --- | --- |
| *P*(*n*) | = | $\displaystyle {\frac{1}{n}}$$\displaystyle \sum_{i=0}^{n-1}$*P*(*n*, *i*) |  |
|  |  |  |  |
|  | = | $\displaystyle {\frac{1}{n}}$$\displaystyle \left\{\vphantom{\sum_{i=0}^{n-1} \space \left[1 + \frac{i}{n} P(i) +\frac{n-i-1}{n} P(n-i-1\right]}\right.$$\displaystyle \sum_{i=0}^{n-1}$$\displaystyle \left[\vphantom{1 + \frac{i}{n} P(i) +\frac{n-i-1}{n} P(n-i-1}\right.$1 + $\displaystyle {\frac{i}{n}}$*P*(*i*) + $\displaystyle {\frac{n-i-1}{n}}$*P*(*n* - *i* - 1$\displaystyle \left.\vphantom{1 + \frac{i}{n} P(i) +\frac{n-i-1}{n} P(n-i-1}\right]$$\displaystyle \left.\vphantom{\sum_{i=0}^{n-1} \space \left[1 + \frac{i}{n} P(i) +\frac{n-i-1}{n} P(n-i-1\right]}\right\}$ |  |
|  |  |  |  |
|  | = | 1 + $\displaystyle {\frac{1}{n^2}}$$\displaystyle \sum_{i=0}^{n-1}$$\displaystyle \left[\vphantom{iP(i) +(n-i-1)P(n-i-1}\right.$*iP*(*i*) + (*n* - *i* - 1)*P*(*n* - *i* - 1$\displaystyle \left.\vphantom{iP(i) +(n-i-1)P(n-i-1}\right]$ |  |
|  |  |  |  |
|  | = | 1 + $\displaystyle {\frac{2}{n^2}}$$\displaystyle \sum_{i=0}^{n-1}$ *iP*(*i*) |  |

since

$\displaystyle \sum_{i=0}^{n-1}$*iP*(*i*) = $\displaystyle \sum_{i=0}^{n-1}$ (*n* - *i* - 1)*P*(*n* - *i* - 1)

$ \mbox{$\bullet$}$

Thus the average path length in a BST satisfies the recurrence:

*P*(*n*) = 1 + $\displaystyle {\frac{2}{n^2}}$$\displaystyle \sum_{i=0}^{n-1}$ *iP*(*i*)

$ \mbox{$\bullet$}$

We shall show that *P*(*n*) $ \leq$ 1 + 4log *n*, by Induction.

**Basis:**

*P*(1) is known to be 1. Also the RHS = 1 for n = 1

**Induction:**

Let the result be true $ \forall$ *i* < *n*. We shall show that the above is true for *i* = *n*.

Consider

|  |  |  |  |
| --- | --- | --- | --- |
| *P*(*n*) | $\displaystyle \leq$ | 1 + $\displaystyle {\frac{2}{n^2}}$$\displaystyle \sum_{i=1}^{n-1}$ *i*(1 + 4log *i*) |  |
|  |  |  |  |
|  | = | 1 + $\displaystyle {\frac{2}{n^2}}$$\displaystyle \sum_{i=1}^{n-1}$ 4*i*log*i* + $\displaystyle {\frac{2}{n^2}}$$\displaystyle \sum_{i=0}^{n-1}$ *i* |  |
|  |  |  |  |
|  | $\displaystyle \leq$ | 1 + $\displaystyle {\frac{2}{n^2}}$$\displaystyle \sum_{i=1}^{n-1}$ 4*i*log*i* + $\displaystyle {\frac{2}{n^2}}$tex2html\_image\_mark>#tex2html\_wrap\_indisplay25297#$\displaystyle {\frac{n^2}{2}}$$\displaystyle \left.\vphantom{\frac{n^2}{2}}\right)$  since $\displaystyle \sum_{i=1}^{n-1}$ *i*$\displaystyle \leq$$\displaystyle {\frac{n^2}{2}}$ |  |

Thus    
*P*(*n*) $\displaystyle \leq$ 2 + $\displaystyle {\frac{8}{n^2}}$$\displaystyle \sum_{i=1}^{n-1}$ *i*log *i*   
Now

|  |  |  |  |
| --- | --- | --- | --- |
| $\displaystyle \sum_{i=1}^{n-1}$*i*log *i* | = | $\displaystyle \sum_{i=1}^{\lceil \frac{n}{2}\rceil -1}$*i*log*i* + $\displaystyle \sum_{i=\lceil \frac{n}{2}\rceil}^{n-1}$ *i*log *i* |  |
|  |  |  |  |
|  | $\displaystyle \leq$ | $\displaystyle \sum_{i=1}^{\lceil \frac{n}{2}\rceil -1}$*i*log$\displaystyle {\frac{n}{2}}$ + $\displaystyle \sum_{i=\lceil \frac{n}{2}\rceil}^{n-1}$ *i*log *n* |  |
|  |  |  |  |
|  | $\displaystyle \leq$ | $\displaystyle {\frac{n^2}{8}}$log$\displaystyle {\frac{n}{2}}$ + $\displaystyle {\frac{3n^2}{8}}$log *n* |  |
|  |  |  |  |
|  | = | $\displaystyle {\frac{n^2}{2}}$log*n* - $\displaystyle {\frac{n^2}{8}}$ |  |

Therefore    
*P*(*n*) $\displaystyle \leq$ 2 + $\displaystyle {\frac{8}{n^2}}$$\displaystyle \left(\vphantom{\frac{n^2}{2} \log - \frac{n^2}{8} }\right.$$\displaystyle {\frac{n^2}{2}}$log - $\displaystyle {\frac{n^2}{8}}$$\displaystyle \left.\vphantom{\frac{n^2}{2} \log - \frac{n^2}{8} }\right)$ = 1 + 4log *n*   
A more careful analysis can be done and it can be shown that

*P*(*n*) $\displaystyle \leq$ 1 + 1.4log *n*