ANALYSIS REPORT FOR ASSIGNMENT-3

NAME- MAREESH KUMAR ISSAR
NET ID- mi251
COURSE- 16:332:573 DATA STRUCTURE AND ALGORITHMS
DATE OF SUBMISSION- 2ND APRIL,2019

ANSWER 1

TWO-THREE TREES BASIC SYMBOL TABLE API

The program Two_Three_Tree.cpp implements the basic symbol table API for the two-three trees according to the required specifications.

ANSWER 2

AVERAGE PATH LENGTH ESTIMATION

NOTE: For this answer, I am using the symbol table API created for answer1. Also, 100 trials were performed for each value of N, where N is the number of keys in the tree. The program generates two output files avg_sorted_black.txt and avg_random_black.txt containing the desired values.

The data collected from the average path length estimation implementation program is given in Table 1 (along with its plot in Figure:1 and Figure:2). While calculating the internal path length I have taken only black links into consideration for computing values in Table 1.

Formula for internal path length(ipl)

$$ipl(t) = ipl(t_L) + ipl(t_R) + s$$

Where t_L, t_R are the left and right subtrees of tree t, s is the size of t.

Since, I am considering only black links for computing the internal path length, therefore I subtract the size of a red node, whenever I encounter a red node in the calculation of internal path length.

The average path length for a particular tree size is given by:-

$$avg\ path\, length = ipl\, of\ the\ tree / tree\, size$$

The final value for avg path length was calculated by averaging the avg path length over the number of iterations (100).

N	AVERAGE PATH LENGTH (RANDOM INSERTION)	AVERAGE PATH LENGTH (SORTED INSERTION)
1	1	1
3	1.33333	1.33333
7	1.91857	2.28571
15	2.59533	4.26667
31	3.31871	8.25806
63	4.15127	16.254
127	4.98323	32.252
255	5.74259	64.251
511	6.56192	128.25
1023	7.34746	256.25
2047	8.21354	512.25
4095	9.05874	1024.25
8191	9.91741	2048.25

Table:1

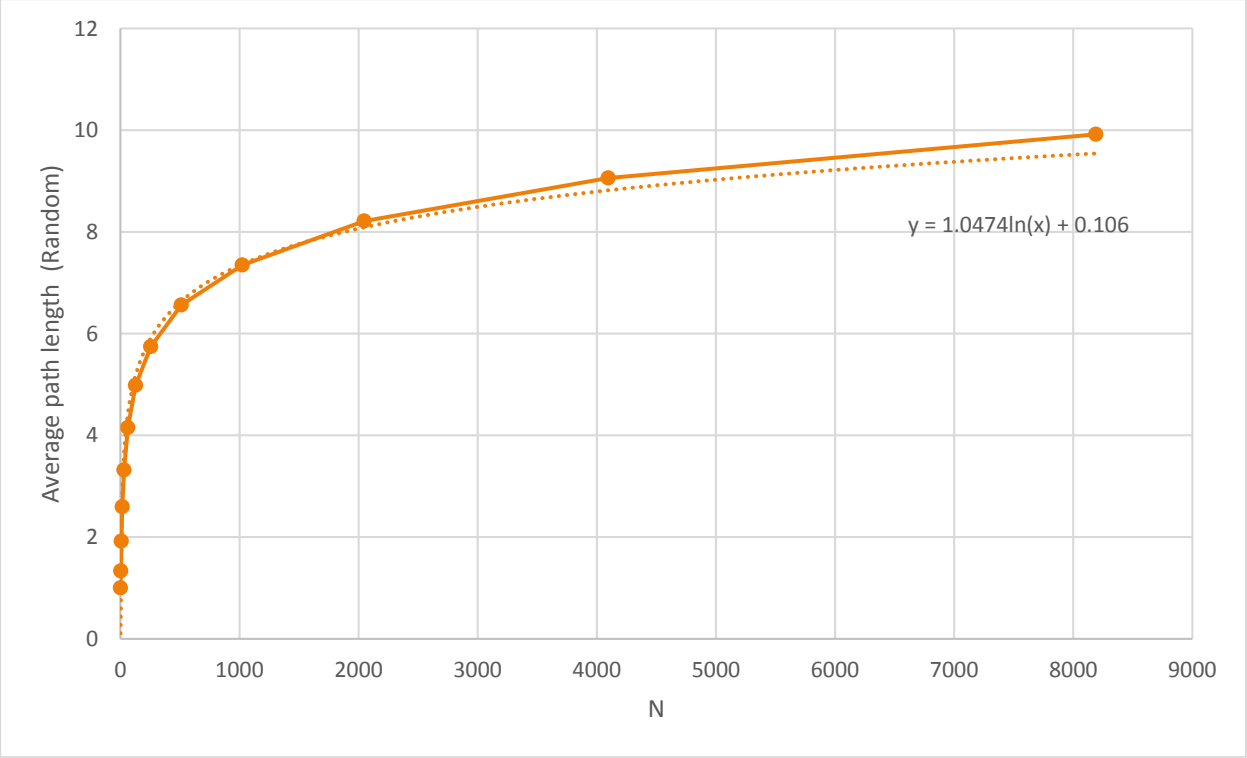

Figure:1

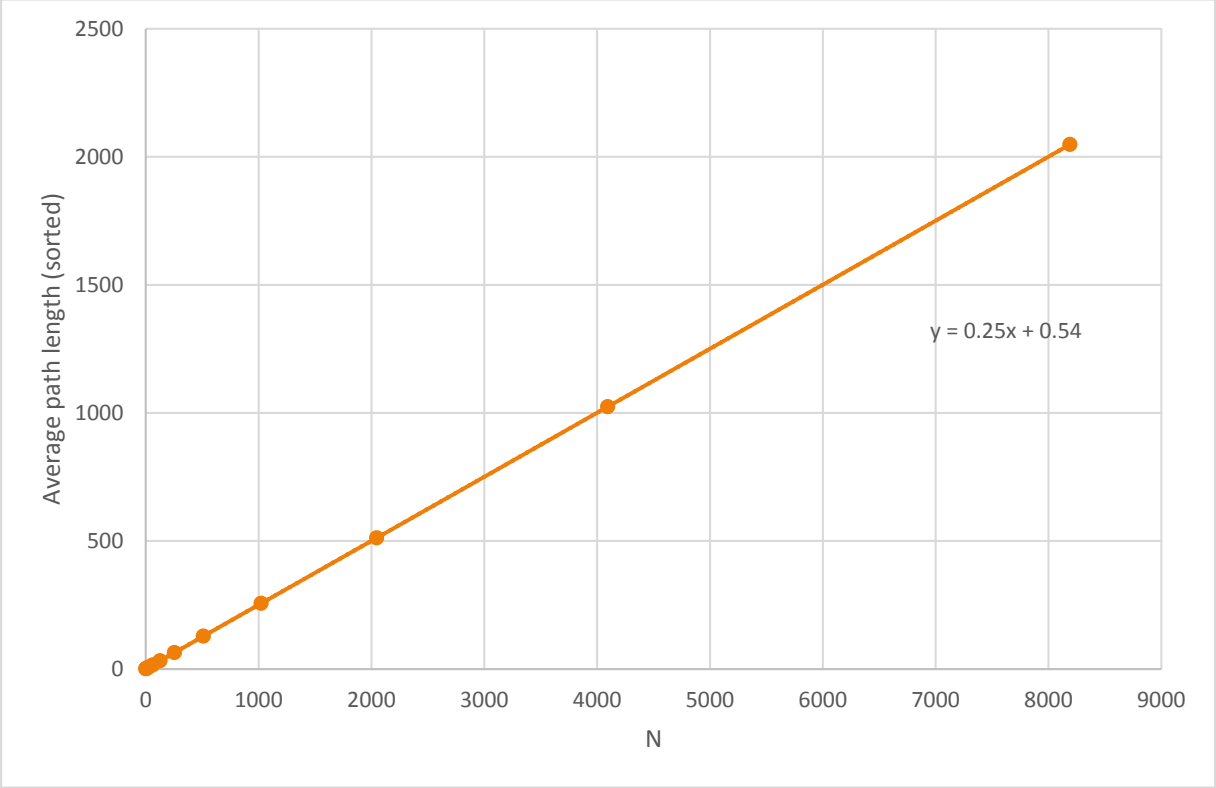

Figure:2

Equation for average path length for random insertions y_R is

$$y_R = 1.0474 \ln(N) + 0.106$$

Equation for average path length for sorted insertions y_S is

$$y_S = 0.25N + 0.54$$

ANSWER 3

NOTE: For this question, I am generating the data randomly (in the range 1-tree size) for all the trials, for a given tree size.

On average percentage of red nodes in a given RBT for tree size 10000 is: 25.3585

On average percentage of red nodes in a given RBT for tree size 100000 is: 25.414

On average percentage of red nodes in a given RBT for tree size 1000000 is: 25.3872

Therefore, taking the average of the above 3 values, we can say that percentage of red nodes for a RBT is approximately 25.4%

ANSWER 4

NOTE: The program generates two output files path_length_avg.txt and path_length_std.txt containing the desired values.

Since, the program was taking a long time for running, I collected the data in two separate steps:-

1. Data for N=1-1991 in steps of 10.
2. Data for N=2001-9901 in steps of 100.

The plot for the data collected is shown in Figure:3 (average path length Vs tree size) and Figure:4 (standard deviation Vs tree size). While calculating the internal path length I have taken both red and black links into consideration for computing values for internal path length.

Formula for internal path length(ipl)

$$ipl(t) = ipl(t_L) + ipl(t_R) + s$$

Where t_L, t_R are the left and right subtrees of tree t, s is the size of t.

The average path length for a particular tree size is given by:-

$$avg\ path\ length = ipl\ of\ the\ tree / tree\ size$$

The final value for avg path length was calculated by averaging the avg path length over the number of iterations (1000).

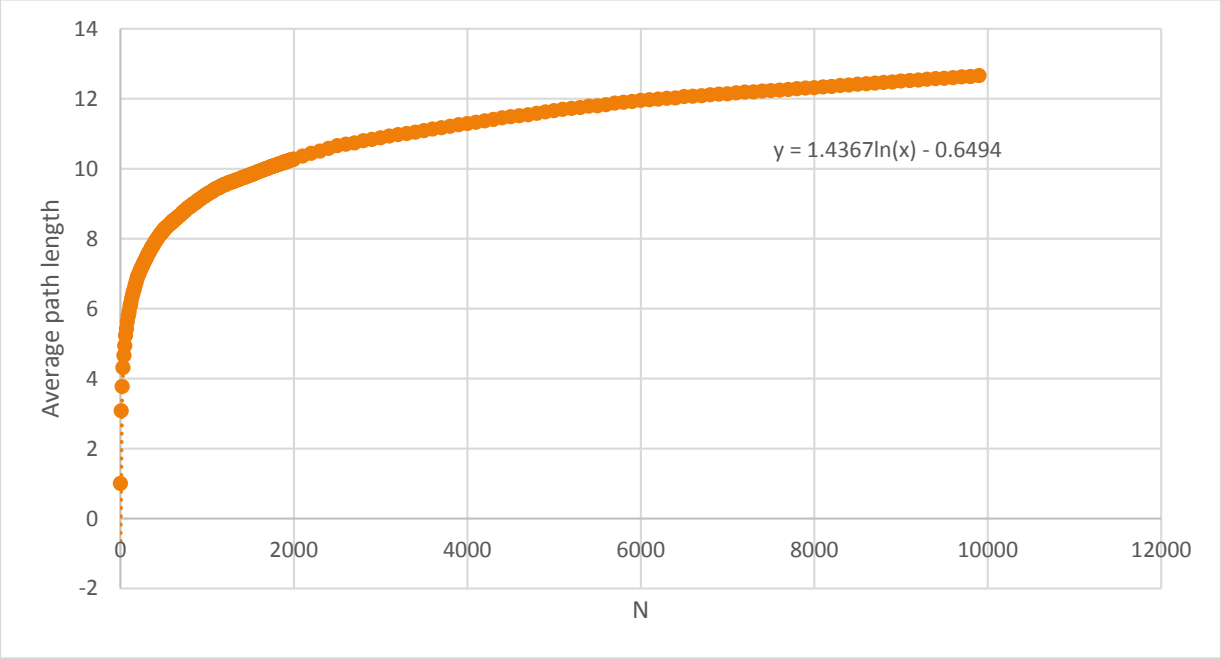

Figure:3

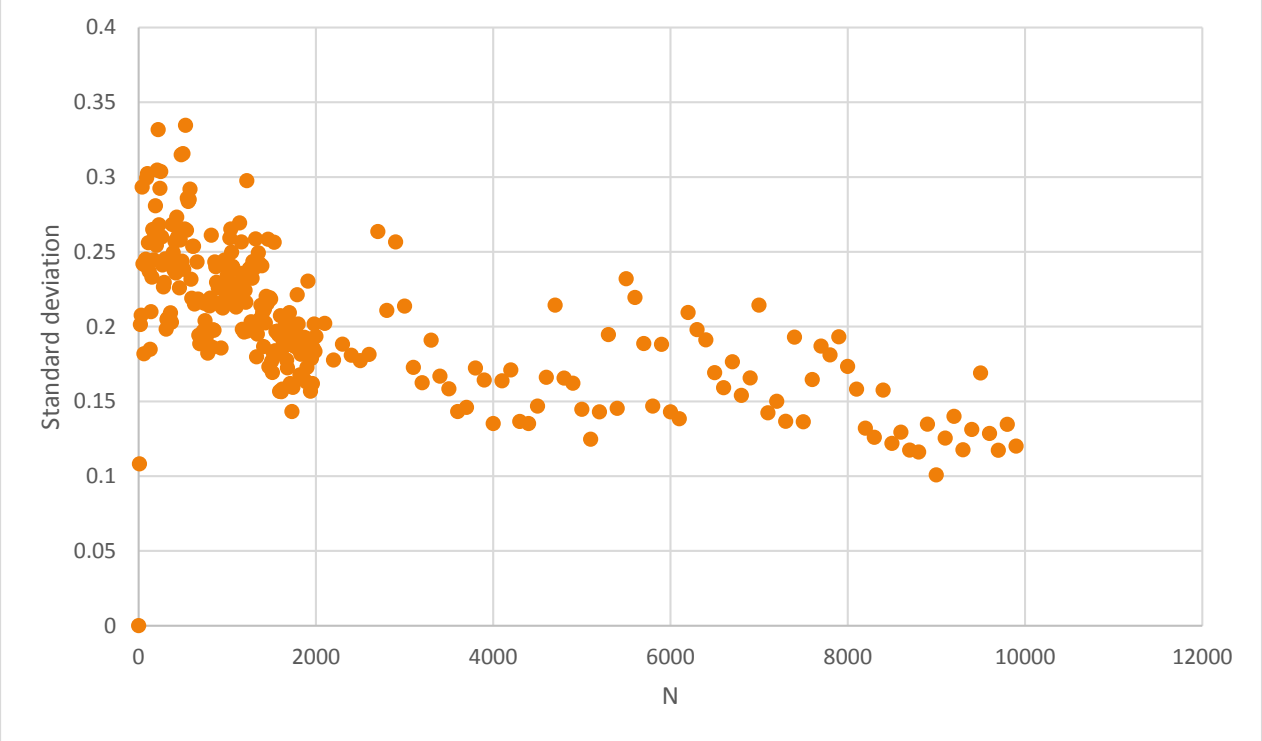

Figure:4

Equation for average path length y is

$$y = 1.4367 \ln(N) - 0.6494$$

Average value for standard deviation is 0.2

ANSWER 5

The value for select(7) is: 8

The value for rank(7) is: 6