Objective of this assignment:

* To verify empirically Theorem 12.4 (Chapter 12, Textbook) that states:

“*The expected height h(n) of a randomly built binary search tree on n distinct keys is O(lg n).*”

What you need to do:

1. (10 points) Implement the *Tree-Insert(T,z)* operation on a binary search tree.
2. **Repeatedly** insert randomly picked numbers in a binary search tree and **collect** the height of the obtained tree. is the height of the binary search tree containing n elements. For simplicity, no need to enforce that the keys are distinct. In other words, you do need to check whether a number is already in the tree.
3. (60 points = (20 points per function/plot)) Plot on three separate graphs the plots of the functions , , and versus ***n***.
4. (30 points: 10 points per plot) Discuss the three plots and conclude regarding the asymptotic growth of h(n). Decide whether your measurements confirm Theorem 12.4.

**Objective**:

The objective of this programming assignment is to verify empirically Theorem 12.4 that states that “*the expected height of a randomly built binary search tree on n distinct keys is O(lg n)*.”

In order to conduct this experiment, you must implement in your preferred language (available on Tux Machines) the Tree-Insert(T,z) operation. Below, I explain in pseudocode the program you must write to collect data and verify Theorem 12.4. You do not need to enforce that the keys are distinct.

**Program to implement**

collectData()

for n = 250 to 20,000 by 250 // 250, 500, 750, 1,000, …. 20,000

sum\_heightn = 0

for j = 1 to 5 do //Take 5 measurements mj for j=1 to 5

for i = 1 to n

pick randomly a number p in the range [0-40,000]

create a node z

set z.key = p

Tree-Insert(T,z)

Measure the height hj of the tree

Discard Tree (free memory)

sum\_heightn += hj

collect Height(n)= sum\_heightn/5 // Average height for n

Write in a file F the value n and Height(n)

**Data Analysis**

Use any plotting software (e.g., Excel) to plot the functions , , and where is the value Height(n) you collected in File F. File F is the file produced by the program you implemented. Discuss your results based on the plots. Decide whether Theorem 12.4 (Chapter 12, Textbook) is correct. This theorem states “*The expected height h(n) of a randomly built binary search tree on n distinct keys is O(lg n).*”

Start inserting your answers on the next page >>>>>>>>>>>>>>>>>>>>

Start Answering On This Page Where Indicated In Blue:

1. (10 points) Implement the *Tree-Insert(T,z)* operation on a binary search tree.

Insert here a snapshot of your implemented *Tree-Insert(T,z)* method in your preferred language...

Text

Description automatically generated

1. **Repeatedly** insert randomly picked numbers in a binary search tree and **collect** the height . is the height of the binary search tree containing n elements. For simplicity, no need to enforce that the keys are distinct. In other words, you do need to check whether a number is already in the tree.

Submit/Attach your program *CollectData* (to collect data) **separately** on Canvas for this assignment. The program is your implementation of the pseudocode provided above to collect data. You can use any language as long as it is already available on Engineering Unix Tux machines. Include here the directives to compile and execute your program CollectData.

**Navigate to the directory (I was on bmm0066@tux251) The file is just in that home directory.**

**programming\_9.java should be there**

**type in “javac programming\_9.java”**

**After compilation finishes, type in “java programming\_9”**

**The program should run and it will output to the system how many iterations are done.**

**Once done, there should be a file called “Output.csv” with all the output data.**

1. (60 points (20 points per function/plot)) Plot on three separate graphs the same plot of the functions , , and versus ***n***.

Insert here the plot for ........

Chart, scatter chart

Description automatically generated

Insert here the plot for ........

Chart

Description automatically generated

Insert here the plot for ........

Chart

Description automatically generated

1. (30 points: 10 points per plot) Discuss the three plots and conclude regarding the asymptotic growth of h(n). Decide whether your measurements confirm Theorem 12.4.

Discuss here the plot ........

**If we are assuming that h(n) is a function O(logn), then the graph for this function should be a horizontal line. If we examine our plot for from above, we can see that the graph shows nearly a straight horizontal line. Below is a graph I found online of log(x)/log(x). It also shows a straight horizontal line, so our experimental data matches our theoretical data. This supports our theory that h(n) is O(logn).**

Graphical user interface, application

Description automatically generated

Discuss here the plot ........

**If we are assuming that h(n) is a function O(logn), then the graph for this function should approach zero as its domain grows. If we examine our plot for from above, we can see that the graph does exactly this. Below is a graph I found online of log(x)/xsqrt(x). It also shows a graph that approaches zero as its domain grows, the only difference being that this graph starts in the negative range. Our experimental data is different from this because the heigh of a tree cannot be negative, so this function cannot reach negative values. Other than that though, our experimental graph matches our theoretical one, helping to confirm our theory that h(n) is O(logn).**

A screenshot of a computer

Description automatically generated with medium confidence

Discuss here the plot ........

**If we are assuming that h(n) is a function O(logn), then the graph for this function should approach zero as its domain grows, and should appear in the same shape as a graph of 1/x. If we examine our plot for from above, we can see that the graph does exactly this. Below is a graph I found online of log(x)/xlog(x). It also shows a graph that approaches zero as its domain grows, and demonstrates the 1/x graph. Since our experimental graph matches our theoretical one, this helps to confirm our theory that h(n) is O(logn).**

A screenshot of a computer

Description automatically generated with medium confidence

Conclude the validity of Theorem 12.4

**From theorem 12.4 we assume that the height of a BST grows as the log(n) since we are generating the BST through random insertions. To prove this assumption, we can use the function h(n) = log(n), graph it, and graph it transformed by other functions, and see if the results are what we expect them to be. If we look back at the previous three steps of this exercise, we can see that every function that involved our h(n) function came out in the same form that we expected. Due to this, we can confirm that h(n) is O(logn), and confirm the validity of Theorem 12.4.**

**Report**

* **Insert** your answers where indicated. This file with your inserted answers will **contain**, explain, and discuss the plots.
* In addition, the above report (this file) must contain the following information:
  + whether your data collection program works or not (this must be just ONE sentence)
  + the directions to compile and execute your program
* Good writing is expected.
* Recall that answers must be well written, documented, justified, and presented to get full credit.

**What you need to turn in:**

* Electronic copy of your source data collection program (separately attached to this assignment)
* Electronic copy of this file (including your inserted answers) (separately attached). Submit the file as a Microsoft Word or PDF file.

**Grading**

* See points distribution with questions/tasks.