DLDTHA021 REPORT:

I have created a class/program GenericskbAVLApp which

uses an AVL tree (AVL.java) to perform operations on data

input. These operations include insertion and searching.

The AVL tree itself uses the Binary Trees and its nodes and node classes

GenericskbAVLApp reads in input from a text file called “GenericsKB.txt” and

stores it in an AVL tree, the input is data if the form of an

item consisting of a term which can be used as a key,

sentence and confidence score.

We have another text file “`GenericsKB-queries.txt`” used for queries from which we

read item names and then use that to search for an item

with the same term in the AVL tree.

Experimental Setup and the goal of the experiment:

The goal of the experiment was to analyze and compare the time taken by the avl tree which is a self-balancing tree as the input size varied and was randomized.

I created a class called Experiment, to interact with the main class and print out the results for each case of the input size n (n=5, n=50, n-=500, etc.). Each input size was randomized and then each had 3 cases from which I was going to choose the best, average and worst case for each input size n.

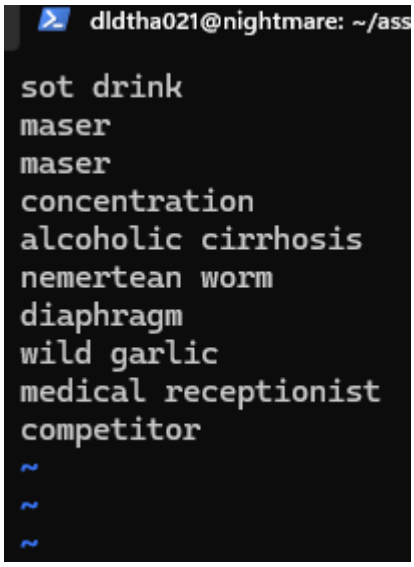
This is an example of the experiment that was run in conjunction with the main program:

```
Do you want to do the experiment? (y/n)
y
Generated file: bestcase.txt
Generated file: averagecase.txt
Generated file: worstcase.txt

For 5 lines,
Case 1 is:

Total number of comparisons made: 30743
Total number of insert operations: 5
Total number of search operations: 25726

Case 2 is:

Total number of comparisons made: 31787
Total number of insert operations: 5
Total number of search operations: 26770
```

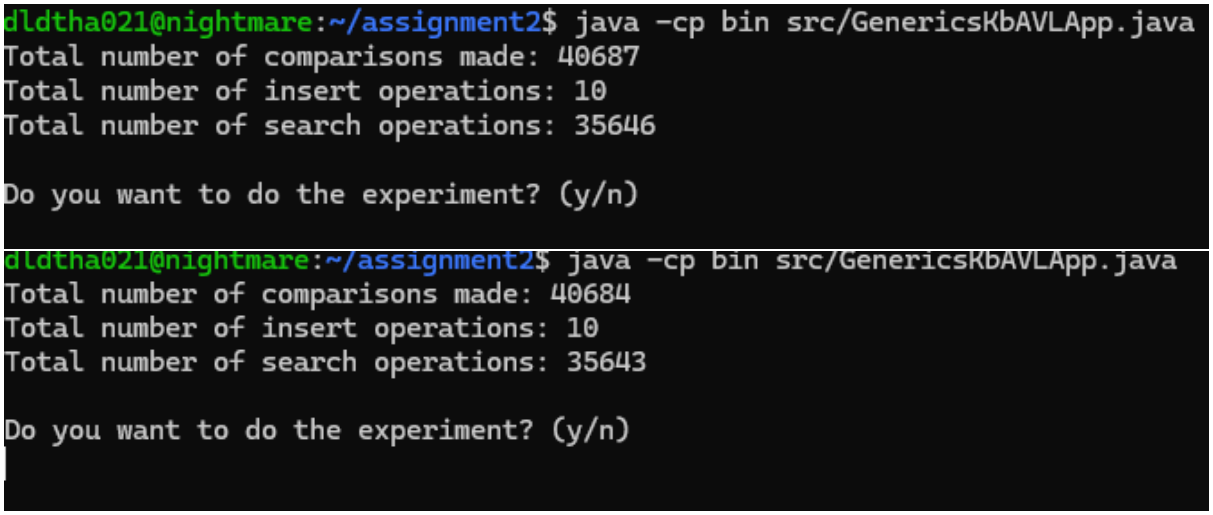
To help with the experiment I created counting variables in the avl tree itself to make sure that regardless of whether I'm doing other experiments the counters are always present and actively counting, my counting variables are for searching , inserting and then the variable that counts the overall variables.

For the testing from part 1:

I manually created a text files to check how my program behaves for when the queries have duplicate terms , and this is the files I created :



And these were the results:



These were the two case for when it had duplicates as show earlier and from when they were removed and I also made an error on purpose for the 1st term "sot drink" to add to the testing case and see how my program would behave.

From my instrumentation:

I got these results from the process I showed earlier:

Input Size	Best Case (Insert)	Best Case (Search)	Average Case (Insert)	Average Case (Search)	Worst Case (Insert)	Worst Case (Search)
5 lines	5	25726	5	25726	5	25726
50 lines	50	57820	50	57430	50	58150
500 lines	500	91622	500	91570	500	89794
5000 lines	5000	125544	5000	125386	5000	123548
50000 lines	50000	159548	50000	159218	50000	156894

```
Do you want to do the experiment? (y/n)
y
Generated file: bestcase.txt
Generated file: averagecase.txt
Generated file: worstcase.txt

For 5 lines,
Case 1 is:

Total number of comparisons made: 30743
Total number of insert operations: 5
Total number of search operations: 25726

Case 2 is:

Total number of comparisons made: 31787
Total number of insert operations: 5
Total number of search operations: 26770
```

For creativity:

I added the class Experiment but instead of just doing it all in the background I added to my program, making it print out the overall experimenting process for a person who has the required files so that they can do the analysis and interpret the results for different purposes as programmers or users.

Summary statistics from git:

```
dldtha021@nightmare:~/assignment2$ git log | (i=0; while
0: commit db26909775d6a068995fb455808aea826a338e52
1: Author: Thabo Dladla <DLDTHA021@myuct.ac.za>
2: Date: Fri Mar 28 21:42:18 2025 +0000
3:
4: I fixed my makefile ,it was not documenting correclty
5:
6: commit ccf6986648844afe0c1a9794c599756cf6357a7d
7: Author: Thabo Dladla <DLDTHA021@myuct.ac.za>
8: Date: Fri Mar 28 20:27:41 2025 +0000
9:
...
55: Author: Thabo Dladla <DLDTHA021@myuct.ac.za>
56: Date: Wed Mar 26 23:54:43 2025 +0000
57:
58: I added my incomplete makefile
59:
60: commit c0a54dcfe88fc5ac3076c22cfa83c04b2beaf955
61: Author: Thabo Dladla <DLDTHA021@myuct.ac.za>
62: Date: Wed Mar 26 23:53:55 2025 +0000
63:
64: I moved my work to my unix and to my git repo
dldtha021@nightmare:~/assignment2$
```