ASSIGNMENT 2

LVRMIG001

Miguel Lavarinhas

OOP Design:

- In this assignment 5 classes were created:
    1. Statement class
    2. GenericsKbAVLApp class
    3. BinaryTreeNode class
    4. AVLTree class
    5. TestSubset class

- The first class created was the statement class in which each statement object represents an entry in the knowledge base and consists of a term, sentence, and confidence score. It also implements the Comparable interface. There is one compareTo() method which is used later in the AVLTree class.
- The GenericsKbAVLApp class is a program which uses the user input to enter a file filled with statements which are used to create Statement objects and fill an AVL Tree. The second part is then to read a file with just terms which then get compared to the statements in the AVL tree and are outputted if found. The last part is how it calls the TestSubset class to create a subset of the first file entered and compare it to the second file entered and output as usual while counting the operations used in the AVLTree class to insert and search. As creativity a statistics option was also added.
- The next class created was the BinaryTreeNode class in which each Node object represents an entry in the AVLTree knowledge base and stores a Statement object as data. This is a class of generic type which extends the Comparable interface.
- The last class AVLTree is used to reference a knowledge base and has many methods such as insert(), find(), balance() and rotateright() which are used in the GenericsKbAVLApp and TestSubset classes. This is a generic class which extends the Comparable interface.

Goal:

- The goal of this assignment is to use an AVL Tree as a knowledge base and use reading lines from a textfile and inserting them and then reading lines from another textfile to search for them in the AVL Tree to inspect the actual results of how man operations get performed versus the theoretical value.

Experimental Testing:

- Description of testing protocol: For the first part of the assignment the testing is done using insert and search counters to count the operations done. In the second part of this assignment the testing was automated by creating a separate class called TestSubset.

This class reads the statements in GenericsKB.txt and creates a subset of certain length which the user defines. There is another method in TestSubset which runs through the GenericsKB-queries.txt file and compares the terms in that file to the terms in the subset and print whether the term has been found in the main program and prints the number of search and insert operations used in the AVLTree class. This process is repeated 10 times with a different amount of entries in the subset and the operation counters are entered onto a graph to compare the theoretical values of the search and insert operations to the ones found in this program.

- Part 1 Testing:

```
5. End program
Enter your choice 1

Enter file name:
GenericsKB.txt
Knowledge base loaded successfully.
Choose an action from the menu:
1. Load a knowledge base from a file
2. Search for statements by entering new file
3. Test with random subset
4. Get Statistical summary of main dataset
5. End program
Enter your choice 2

Enter file name:
TenQueries.txt
Statement found for amphidiploid: Amphidiploids are tetraploids with a complete diploid genome complement from each species parent.(0.8168812990188599)
Statement found for sponger: A sponger is a follower(1.0)
Statement found for moon jelly: Moon jellies catch small plankton with tentacles, covered with stinging cells, called nematocysts.(0.8074753880500793)
Statement found for follicular cast: Follicular casts are tightly adherent scales around the hair shaft.(0.7716030478477478)
Statement found for riptide: Riptides are current.(1.0)
Statement for term Tickle not found.
Statement for term Pickle not found.
Statement for term Sickle not found.
Statement for term Bar not found.
Statement for term Tar not found.
Knowledge base loaded successfully.
The number of insert operations:
Best case: 0
Average case: 14.43768
Worst case: 19
The number of search operations:
Best case: 11
Average case: 15.1
Worst case: 16
```
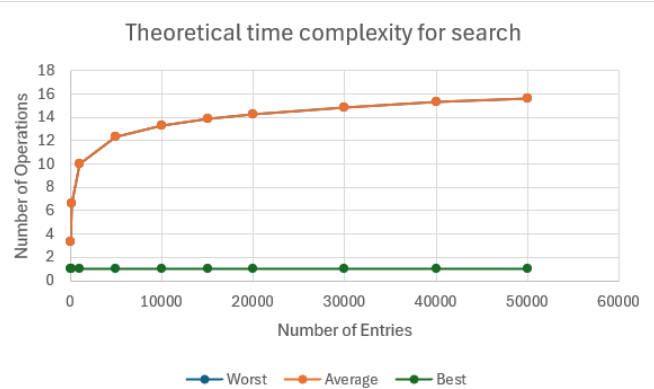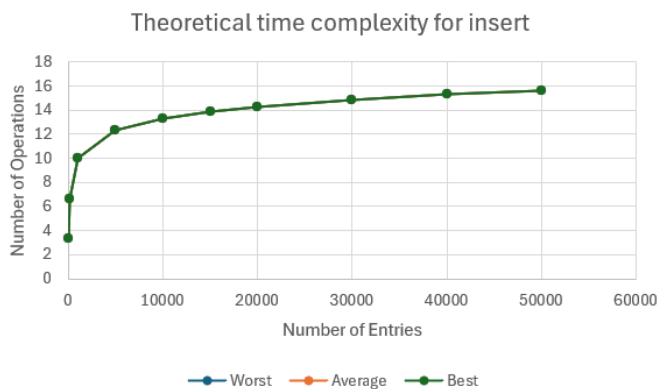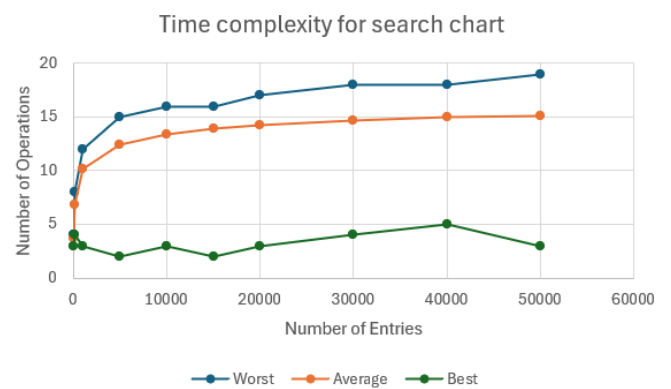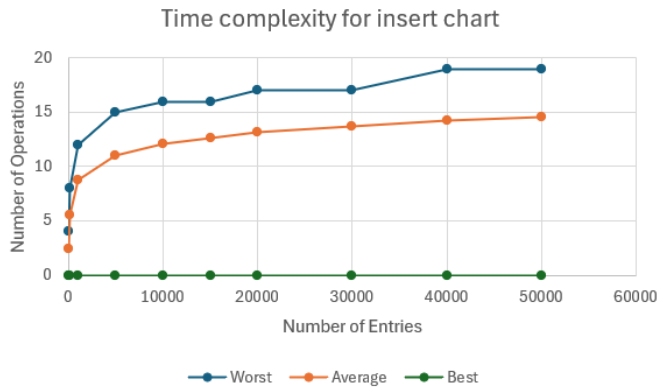
Code to show the output of the Code with a manually made query textfile of 5 terms that are in the GenericsKB.txt file and 5 that are not.

As seen above the program functions and the count operations were recorded as:

1. Insert operations:
   Best case: 0
   Average case: 14.438
   Worst case: 19
2. Search operations:
   Best case: 11
   Average case: 15.1
   Worst case: 16

- Part 2 Testing:

Time complexity for insert chart

Time complexity for search chart

Theoretical time complexity for insert

Theoretical time complexity for search

Conclusion: As seen from above the theoretical graphs and actual graphs from the program are similar as they both have the same logarithmic pattern and follow similar values, however, there are a few differences. The first difference is that the Best case scenario for the insert complexity is O(0) for the program as the first if statement where it checks if the node is equal to null does not increment the counter as the assignment states it is not comparing keys. The second difference is the search Best Case for the search is theoretically O(1), however, in reality, the program has a slightly higher varying value as the subset is always randomised and the first node in the set may not exist in the query file. Lastly, the Average Cases are the same as the Worst Cases in the theoretical graphs yet differ in the actual graphs which could be due to the differences in each subset as they are randomised. In conclusion, the results are similar to the theoretical values and the differences might be due to the way the program was implemented.

Creativity:

- For creativity the program has a statistics option for the user when the user selects option 4 which shows statistics regarding the confidence score of the statements in the

file entered and displays the time taken in milliseconds to run the insert and search operations regarding the first two files. The time is also displayed for the operations of each subset.

1. Example for the statistics option:

```
Statement found for xenophobia: Xenophobias are social phobia.(1.0)
Statement found for religious activity: Religious activities are cultural activities.(1.0)
Statement found for rightist: A rightist is a conservative(1.0)
Statement found for different gas: Different gases comprise particles with different masses.(0.7551524043083191)
Statement found for ethnic study: Ethnic studies are fields of study.(1.0)
Statement found for splayed leg: Splayed legs are legs that extend outward from the hips.(0.8099573254585266)
Statement found for amphidiploid: Amphidiploids are tetraploids with a complete diploid genome complement from each species parent.(0.8168812990188599)
Statement found for sponger: A sponger is a follower(1.0)
Statement found for moon jelly: Moon jellies catch small plankton with tentacles, covered with stinging cells, called nematocysts.(0.8074753880500793)
Statement found for follicular cast: Follicular casts are tightly adherent scales around the hair shaft.(0.7716030478477478)
Statement found for riptide: Riptides are current.(1.0)
Knowledge base loaded successfully.
The number of insert operations:
Best case: 0
Average case: 14.43768
Worst case: 19
The number of search operations:
Best case: 5
Average case: 15.0774
Worst case: 19
Choose an action from the menu:
1. Load a knowledge base from a file
2. Search for statements by entering new file
3. Test with random subset
4. Get Statistical summary of main dataset
5. End program
Enter your choice 4

Statistics summary:
Highest confidence score: 1.0
Lowest confidence score: 0.0
Average confidence score: 0.9362020874249936
Time taken to run: 764 ms.
```

As seen above the statistics option displays the statistics of the confidence score and displays the time of 764ms it took to run the code.

2. Example of the subset time:

```
Statement for term plus not found.
Statement for term barium sulfide not found.
Statement for term accordion door not found.
Statement for term osteoblastic lesion not found.
Statement for term bayonet not found.
Statement for term aqueous fluid not found.
Statement for term nosegay not found.
Statement for term xenophobia not found.
Statement for term religious activity not found.
Statement for term rightist not found.
Statement found for different gas: Different gases comprise particles with different masses.(0.7551524043083191)
Statement for term ethnic study not found.
Statement for term splayed leg not found.
Statement for term amphidiploid not found.
Statement for term sponger not found.
Statement for term moon jelly not found.
Statement for term follicular cast not found.
Statement for term riptide not found.
The number of insert operations:
Best case: 0
Average case: 5.29
Worst case: 7
The number of search operations:
Best case: 4
Average case: 6.789
Worst case: 8
Time Taken to run subset: 474 ms.
```

As seen above the time taken to run the subset of 100 entries took 474 ms.

Git Log:

0: commit 1030287d2980a7ae6d18c3cad335fd53fa9a9c80

1: Author: Miguel Lavarinhas <lvrmig001@nightmare.cs.uct.ac.za>

2: Date: Wed Mar 20 12:36:45 2024 +0000

3:

4: Assignment 2: Added javadoc to all classes.

5:

6: commit 2d5fd1d4b7b12368f85840605a8f8f90a7e19ca1

7: Author: Miguel Lavarinhas <lvrmig001@nightmare.cs.uct.ac.za>

8: Date: Wed Mar 20 09:17:53 2024 +0000

9:

...

13: Author: Miguel Lavarinhas <lvrmig001@nightmare.cs.uct.ac.za>

14: Date: Mon Mar 18 18:31:18 2024 +0000

15:

16: Added creativity to GenericsKbAVLApp

17:

18: commit b7b2dd8a072b7bfe2f1fb8399bfbde1eef7c7739

19: Author: Miguel Lavarinhas <lvrmig001@nightmare.cs.uct.ac.za>

20: Date: Mon Mar 18 16:27:06 2024 +0000

21:

22: Assignment 2: My first commit.