Using arrays and binary search trees to store, search for and insert data from a file.

CSC2001F Assignment 1

Zuhayr Halday (HLDZUH001)

3/8/24

Overall OOP Design

Array application

While programming the version of the app which uses the array data structure, adding multiple classes for small classes for unnecessary reasons would have obfuscated the code, and as such, I have chosen to only use one main class for that program. This main class contains all functions which were used in the program. These included:

- LineCount – counts all lines in a given file in order to determine the size of the array which will hold the data.
- ReadIntoArray – creates an array using LineCount and stores each line from the file as an entry in the array.
- TermSearch – searches the array for an entry matching a user-input term.
- TermAndStatementSearch - searches the array for an entry matching a user-input term and statement.
- The main function.

Binary Search Tree application

In the binary search tree version of the app, two additional classes were used alongside the main class of the program. These included:

Node:

- An object representing the node of a binary search tree. This object includes a data, left and right property.

BST:

- An object representing the entire binary search tree data structure, which includes a Node property.
- This object also housed all BST-related functions, such as the Search, Insert and preOrder traversal functions. The algorithms in these functions were almost identical to the ones learnt in lectures this semester.
- All BST-related functions made use of recursive functions.

Relationships between classes

In the case of the BST app, Nodes were used as a property of the BST class, and this BST class was used in the main function of the program. User input from the menu of the program was used to call various functions of the BST object. Some functions also produce an output to the screen, and thus, to the user. The classes and user all interact with one another to form a fully functioning program.

Testing

GenericsKbArrayApp:

```
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 3

Missing data, knowledge base not loaded yet.
```

```
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

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. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 3
Enter the term to search: cheese

Statement found: Mozzarella cheese is cheese. (Confidence score: 1.0)


Statement found: Cheese spread is paste (Confidence score: 1.0)


Statement found: Vegetarian cheese is produced using microbial or fungal enzymes. (Confidence score: 0.7017971277236938)


Statement found: A cheese sauce is a sauce (Confidence score: 1.0)


Statement found: Domestic cheese uses milk. (Confidence score: 1.0)


Statement found: Process cheese is cheese (Confidence score: 1.0)


Statement found: Laugh cow light cheese has calories. (Confidence score: 1.0)


Statement found: Farm cheese is cheese (Confidence score: 1.0)


Statement found: Brick cheese is cheese (Confidence score: 1.0)
```

```
Statement found: Hard cheese is misfortune. (Confidence score: 1.0)

Statement found: A cheese dip is a dip (Confidence score: 1.0)

Statement found: Cheese whey is produced during the separation of casein and whey as the curd for cheese is produced. (Confidence score: 0.7176533
937454224)

Statement found: Cow light cheese has calories. (Confidence score: 1.0)

Statement found: Big cheese is an adult (Confidence score: 1.0)

Statement found: Cottage cheese is cheese. (Confidence score: 1.0)

Statement found: American cheese has moisture. (Confidence score: 1.0)

Statement found: Firm cheeses are cheeses that have been cooked and pressed. (Confidence score: 0.7462358474731445)

Statement found: Grate cheese adds flavor. (Confidence score: 1.0)

Statement found: Cheese production relies on applications. (Confidence score: 1.0)

Statement found: Swiss cheese is located in shops. (Confidence score: 1.0)

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice:
```

```
Statement found: Swiss cheese is located in shops. (Confidence score: 1.0)

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 4
Enter the term: cheese sauce
Enter the statement to search for: A cheese sauce is a sauce

The statement was found and has a confidence score of  1.0.

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice:
```

```
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 2
Enter the term: shoe
Enter the statement: A shoe is worn on a foot.
Enter the confidence score: 1.0

Statement for term shoe has been updated.

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 3
Enter the term to search: shoe

Statement found: Shoe lace is lacing (Confidence score: 1.0)


Statement found: Comfortable shoes are shoes. (Confidence score: 1.0)


Statement found: A wooden shoe is a shoe (Confidence score: 1.0)


Statement found: Brake shoes are restraint. (Confidence score: 1.0)


Statement found: Formal shoes are shoes. (Confidence score: 1.0)


Statement found: Soft shoes are tap dancing. (Confidence score: 1.0)
```

```
Statement found: A wooden shoe is a shoe (Confidence score: 1.0)

Statement found: Brake shoes are restraint. (Confidence score: 1.0)

Statement found: Formal shoes are shoes. (Confidence score: 1.0)

Statement found: Soft shoes are tap dancing. (Confidence score: 1.0)

Statement found: Old shoes are located in trash. (Confidence score: 1.0)

Statement found: Walking shoes are shoes. (Confidence score: 1.0)

Statement found: A running shoe is a shoe (Confidence score: 1.0)

Statement found: Tennis shoes are court shoes. (Confidence score: 1.0)

Statement found: Golf shoes have rubber soles. (Confidence score: 1.0)

Statement found: Basketball shoes are court shoes. (Confidence score: 1.0)

Statement found: Tight shoes cause blisters. (Confidence score: 1.0)

Statement found: A shoe is worn on a foot. (Confidence score: 1.0)

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice:
```

```
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 4
Enter the term: shoe
Enter the statement to search for: A shoe is worn on a foot.

The statement was found and has a confidence score of  1.0.
```

```
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 5
PS C:\Users\mrkew\OneDrive - University of Cape Town\UCT\3rd Year\1st Semester\CSC2001F\
```

```
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 1
Enter file name: file.txt
Ran into an error while trying to read file: file.txt (The system cannot find the file specified)
```

```
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 3
Enter the term to search: tiktok
'tiktok' was not found in the current knowledge base.
```

```
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 4
Enter the term: tiktok
Enter the statement to search for: Tiktoks are fun to watch.
Term and statement: 'tiktok' and 'Tiktoks are fun to watch.' was not found in the knowledge base.
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice:
```

GenericsKbBSTApp

```
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 3

Knowledge base has not been loaded yet.

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 1
Enter file name: Gen.txt
Ran into an error while trying to read file: Gen.txt (The system cannot find the file specified)
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

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. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 3
Enter the term to search: cheese


Statement found: Cheese spread is paste (Confidence score: 1.0)


Statement found: A cheese sauce is a sauce (Confidence score: 1.0)


Statement found: A cheese dip is a dip (Confidence score: 1.0)


Statement found: Cheese production relies on applications. (Confidence score: 1.0)


Statement found: Cheese whey is produced during the separation of casein and whey as the curd for cheese is produced. (Confidence
score: 0.7176533937454224)


Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 4
Enter the term: cheese sauce
Enter the statement to search for: A cheese sauce is a sauce

Statement found: A cheese sauce is a sauce (Confidence score: 1.0)
```

```
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 2
Enter the term: tiktok
Enter the statement: Tiktoks are fun to watch.
Enter the confidence score: 1.0

Statement for term tiktok has been updated.

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 3
Enter the term to search: tiktok


Statement found: Tiktoks are fun to watch. (Confidence score: 1.0)
```

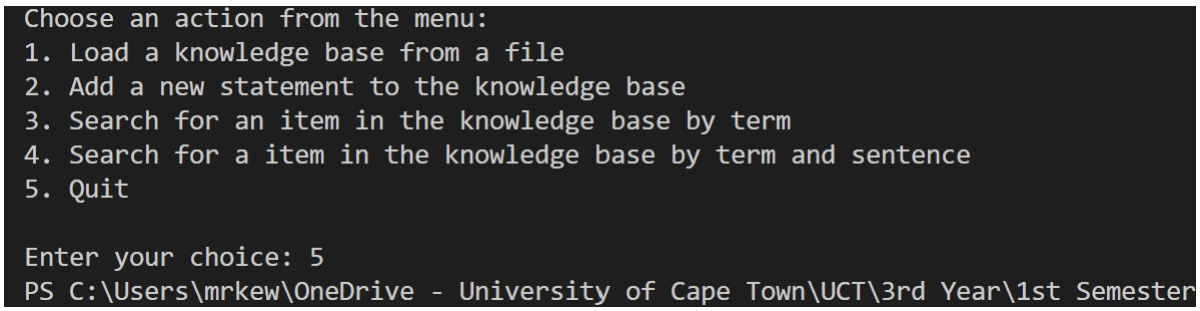
```
Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 3
Enter the term to search: spongebob

'spongebob' was not found in the current knowledge base.

Choose an action from the menu:
1. Load a knowledge base from a file
2. Add a new statement to the knowledge base
3. Search for an item in the knowledge base by term
4. Search for a item in the knowledge base by term and sentence
5. Quit

Enter your choice: 4
Enter the term: spongebob
Enter the statement to search for: Spongebob is a sponge.
Term and statement: 'spongebob' and 'Spongebob is a sponge.' was not found in the knowledge base.
```



Description of testing protocol

In the above tests, almost all possibilities were covered. This was done by considering each possibility after each number was input and where that could lead to. This is akin to White-Box testing, as I, as the programmer, have detailed knowledge of how the program can and should operate. This testing method is also what allowed me to detect and resolve the many bugs encountered while creating these programs, such as error checking when files are being loaded or read, or providing user friendly messages for searches that yield no results, etc. As can be seen in the screenshots, all of these potential bugs were tested and accounted for. The basic requirements of the program, such as the populating, updating and searching of the knowledge bases were also tested.

Creativity

Extensive error checking

While it was not easy trying to be creative while also staying within the requirements of the program, extensive error checking was performed at multiple points throughout both programs. Try/catch blocks were used multiple times, specifically when reading from an external file and importing the data into either an array or a BST. Other than that, global Boolean variables were created in order to ensure that no operations can be performed by the user until a knowledge base is loaded from a file.

Enhanced search functionality

When using the search by term functionality of each program, users need not input exactly what is written for each term of a generic. Simply searching for one word or any other partial match of a search term will still result in all partial matches being returned as search results.

Using a preOrder traversal algorithm

When using a BST to store data from a file, there are a handful of traversal algorithms which have been covered in lectures. When using the search functionality of the BST program, search results will appear in a specific order, different to that of the array program's order. This order is in fact determined by a preOrder traversal algorithm when searching through the BST.

Git Log

First 10 commits:

1. 81dc96ea5aff8c85f294c10b2b1a8fdf725788d2

   ZuhayrHalday <halday.zuhayr@gmail.com> 1709912276 +0200

   commit (initial): First Commit

2. 81dc96ea5aff8c85f294c10b2b1a8fdf725788d2
   2cb86d79aeb25e82df8bbf34591c808a6aca598b

   ZuhayrHalday <halday.zuhayr@gmail.com> 1709913081 +0200

   commit: Added all necessary class imports

3. 2cb86d79aeb25e82df8bbf34591c808a6aca598b
   1226b06934be54feeae55d68e12dd9ddaf9b3c18

   ZuhayrHalday <halday.zuhayr@gmail.com> 1709913363 +0200

   commit: Added an array

4. 1226b06934be54feeae55d68e12dd9ddaf9b3c18
   5fc0b5ca694d7319513557eb8cb19f49175c5f04

   ZuhayrHalday <halday.zuhayr@gmail.com> 1709913559 +0200

   commit: Added text files to project folder

5. 5fc0b5ca694d7319513557eb8cb19f49175c5f04
   f4320f301de0f5353eddd22b27f70a55e41c493c

   ZuhayrHalday <halday.zuhayr@gmail.com> 1709914643 +0200

   commit: Updated LineCount function to read each line of an imported file

6. f4320f301de0f5353eddd22b27f70a55e41c493c
   bdb5bd372a0922a5db4779e62c05c47fb9b3da39

   ZuhayrHalday <halday.zuhayr@gmail.com> 1709914694 +0200

   commit: Populated GenericsKbArrayApp class with first function, LineCount

7. bdb5bd372a0922a5db4779e62c05c47fb9b3da39
   628afb8b7c85c8aa8b20c13696d4d5734416a796

   ZuhayrHalday <halday.zuhayr@gmail.com> 1709914867 +0200

   commit: Enclosed file reading in a try/catch block to detect errors while trying to read a file

8. 628afb8b7c85c8aa8b20c13696d4d5734416a796
   58c38a6c891e46edf64cd0546775f4e5fbc75756

   ZuhayrHalday <halday.zuhayr@gmail.com> 1709916139 +0200

   commit: Added similar try/catch block to the ReadIntoArray function, while reading lines from file into an array

9. 58c38a6c891e46edf64cd0546775f4e5fbc75756
   6f9a82d6b1eff0eff6bb719651beb0d57105c415

   ZuhayrHalday <halday.zuhayr@gmail.com> 1709917000 +0200

   commit: Added TermSearch function. Will be used when a user wants to search by terms

10. 6f9a82d6b1eff0eff6bb719651beb0d57105c415
    c023bff736663a616c16713a4dbe3a9aea19810b

    ZuhayrHalday <halday.zuhayr@gmail.com> 1709917309 +0200

    commit: Added functionality to check if a file has been read into the array before attempting to search the array

Last 10 commits:

1. 7af701316c7257f7f8077ea50b3361fe9bc6bc74
   c9a408fde87c55f8fcb70cb33ce60855734bf15a

   ZuhayrHalday <halday.zuhayr@gmail.com> 1709927077 +0200

   commit: Added function to search the BST using term and statement using the same BST search algorithm as the previous search function

2. c9a408fde87c55f8fcb70cb33ce60855734bf15a
   0fc845691650f3b2d45442b4587a1a2a6dccb85f

   ZuhayrHalday <halday.zuhayr@gmail.com> 1709927807 +0200

   commit: Used the preOrder traversal algorithm to determine the order in which search results will be found

3. 0fc845691650f3b2d45442b4587a1a2a6dccb85f
   07f547dca7fd1affbd21f642c891862873ea3b4b

   ZuhayrHalday <halday.zuhayr@gmail.com> 1709928262 +0200

   commit: Added loadFile function to main class, used similar try/catch implementation as in the array version of the program

4. 07f547dca7fd1affbd21f642c891862873ea3b4b
   009fa73e1c64cb11296b57b228a965a4c89574c2

   ZuhayrHalday <halday.zuhayr@gmail.com> 1709928813 +0200

   commit: Copied frame of main function from array app, made changes to adapt the function to a BST based data structure

5. 009fa73e1c64cb11296b57b228a965a4c89574c2
   b356fedf631f56a81e55b28526f44fc5f73a7af3

   ZuhayrHalday <halday.zuhayr@gmail.com> 1709928895 +0200

   commit: Minor formatting changes

6. b356fedf631f56a81e55b28526f44fc5f73a7af3
   cc47df56556e745ffa68869fcbec75252f758081

   ZuhayrHalday <halday.zuhayr@gmail.com> 1709929308 +0200

   commit: Fixed bug with fileError variable

7. cc47df56556e745ffa68869fcbec75252f758081
   48b1b1d9a266f1538984bbbd50d251162c7f9631

   ZuhayrHalday <halday.zuhayr@gmail.com> 1709929782 +0200

   commit: Fixed bug with preOrderTraversal function, not all search results were displayed

8. 48b1b1d9a266f1538984bbbd50d251162c7f9631
   f033aff8c1630d408d020019641da29bb412c6d1

   ZuhayrHalday <halday.zuhayr@gmail.com> 1709930017 +0200

   commit: Changed the way search results are displayed to the user

9. f033aff8c1630d408d020019641da29bb412c6d1
   73b25232c9030071cdbe99599c6316fa9aa583d8

   ZuhayrHalday <halday.zuhayr@gmail.com> 1709930609 +0200

   commit: Added javadoc comments

10. 73b25232c9030071cdbe99599c6316fa9aa583d8
    1e98a8690720adf544bca18af3616a4a3ea533ed

    ZuhayrHalday <halday.zuhayr@gmail.com> 1709931299 +0200

    commit: Added javadoc comments