SCHOOL OF COMPUTING (SOC)

Diploma in Applied AI and Analytics

ST1507 DATA STRUCTURES AND ALGORITHMS (AI)

2022/2023 SEMESTER 2
ASSIGNMENT ONE (CA1)
~ Thesaurus Based Text Processing Application~

Objective of Assignment

To practice what you have learnt in the module on data structures, algorithms, and object-oriented programming by developing a Thesaurus Based Text Processing Application.

Instructions and Guidelines:

1. This is an individual assignment, and it accounts to **30%** of your final grade.

2. The submission date is **Friday 25 November 5:00 pm**.

3. The development will be carried out in Python using Anaconda.

4. The interviews will be conducted during the DSAA lessons in week 7. You are expected to explain on your code and program logic. Take note that the interview is compulsory. In case you are absent from the interview without valid reasons, you will not get any marks for the assignment.

5. No marks will be given, if the work is copied, or you have allowed others to copy your work.

6. **50% of marks** will be deducted for submission of assignment within **ONE** calendar day after the deadline. **No marks shall** be awarded for assignments submitted **more than one day** after the deadline.

Warning: Plagiarism means passing off as one's own the ideas, works, writings, etc., which belong to another person. In accordance with this definition, you are committing plagiarism if you copy the work of another person and turning it in as your own, even if you would have the permission of that person. Plagiarism is a serious offence and disciplinary action will be taken against you. If you are guilty of plagiarism, you may fail all modules in the semester, or even be liable for expulsion.

Overview of the system

Your job is to implement a *Thesaurus Based Text Processing Application* that serves the following three purposes:

1. It will allow a user to set up a thesaurus (*) from scratch.
   (*) A thesaurus is a collection of keywords with associated synonyms.
2. It will allow a user to sort the synonyms in a thesaurus according to different criteria.
3. It will allow the user to make use of the thesaurus to process an input text by either simplifying it (by replacing synonyms with keywords) or to make the text more elegant (by replacing keywords with synonyms)

An example of a thesaurus:

Next is an example of a thesaurus with 3 keywords: 'brave', 'cat' and 'ferocious'. For each of the keywords is then spelled out a list of synonyms.

cat:
kitty, feline, kitten, pussycat

brave:
courageous, valiant, heroic, bold, daring, fearless, plucky

ferocious:
fierce, vicious, violent, cruel, brutal, aggressive, wild, unruly, savage, merciless, barbarous

Example of sorting synonyms:

The keywords and synonyms in a thesaurus can be sorted, for instance, as shown below, in alphabetic order.

brave:
bold, courageous, daring, fearless, heroic, plucky, valiant

cat:
feline, kitten, kitty, pussycat

ferocious:
aggressive, barbarous, brutal, cruel, fierce, merciless, savage, unruly, vicious, violent, wild

Example of using a thesaurus to simplify a text:

Input Text

A stray **cat** was recently spotted in the neighborhood protecting her little **kitten** as the two were being attacked by a **ferocious** dog. The **brave feline** eventually managed to chase the **aggressive** dog away. This goes to show that we should never underestimate a **pussycat**'s maternal instincts when looking after her little **kitty**.

Output Text

A stray **cat** was recently spotted in the neighborhood protecting her little **cat** as the two were being attacked by a **ferocious** dog. The **brave cat** eventually managed to chase the **ferocious** dog away. This goes to show that we should never underestimate a **cat**'s maternal instincts when looking after her little **cat**.

Example of using a thesaurus to make a text more elegant:

Input Text

A stray **cat** was recently spotted in the neighborhood protecting her little **cat** as the two were being attacked by a **ferocious** dog. The **brave cat** eventually managed to chase the **ferocious** dog away. This goes to show that we should never underestimate a **cat**'s maternal instincts when looking after her little **cat**.

Output Text

A stray **feline** was recently spotted in the neighborhood protecting her little **kitten** as the two were being attacked by a **aggressive** dog. The **bold kitty** eventually managed to chase the **barbarous** dog away. This goes to show that we should never underestimate a **pussycat**'s maternal instincts when looking after her little **feline**.

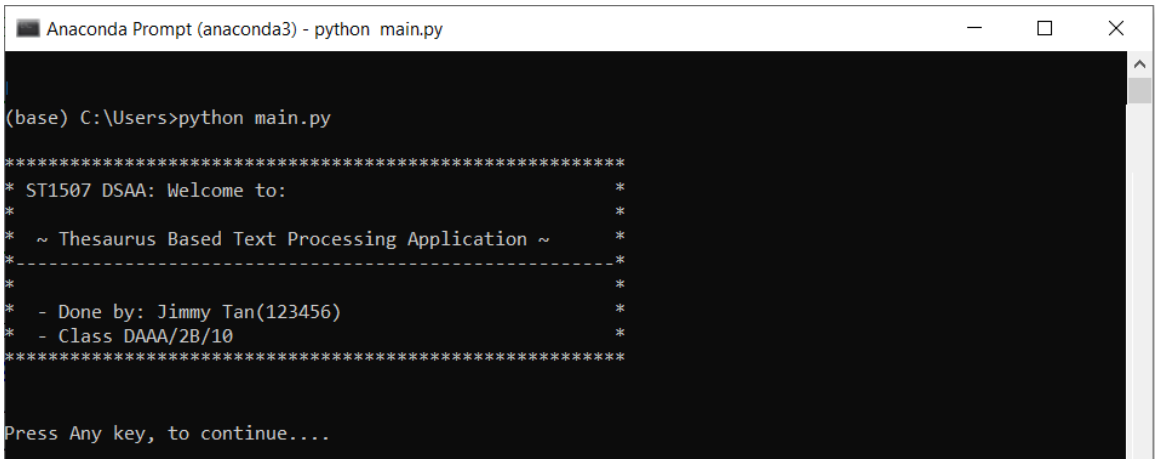
(*) Take note we are using the alphabetically sorted thesaurus from the previous page.

Starting the application

Your application must be able to start from the Anaconda Command prompt by typing:

`python main.py`.

It will then display a title bar with essential information, next it waits for the user to press any key before the application continues.



- You are required to follow the above format.
- Please, ensure that you display your name, student ID and class in the format as is shown in the example

Selection menu

When the application continues, the user will then be presented with a menu as shown below. The menu allows the user to choose from 10 options.

```
Please select your choice: (1,2,3,4,5,6,7,8,9,10)
        1. New
        2. Open
        3. Sort
        4. Process Text
        5. Extra Option One
        6. Extra Option Two
        7. Print
        8. Save
        9. Save As
        10. Exit
Enter choice:
```

- You are required to follow the above format for the menu display.
- Option 5 and 6 are reserved as extra options that you will need to design and implement yourself.
- The user will be able to repeatedly select the various options from the menu.
- The application will terminate once the user selects option 10 Exit (displaying an Exit message as shown below).

```
Press Any key, to continue....

Please select your choice: (1,2,3,4,5,6,7,8,9,10)
        1. New
        2. Open
        3. Sort
        4. Process Text
        5. Extra Option One
        6. Extra Option Two
        7. Print
        8. Save
        9. Save As
        10. Exit
Enter choice: 10

Bye, thanks for using ST1507 DSAA: Thesaurus Based Text Processor
```

Setting up a new Thesaurus

The user selects option '1' to start a new thesaurus. To set up a new thesaurus, the user will be prompted to enter the first keyword (for instance enters the word 'cat').

```
Please select your choice: (1,2,3,4,5,6,7,8,9,10)
        1. New
        2. Open
        3. Sort
        4. Process Text
        5. Extra Option One
        6. Extra Option Two
        7. Print
        8. Save
        9. Save As
        10. Exit
Enter choice: 1

We will be starting a new Thesaurus.
You may now enter a series of keywords and their synonyms.

Enter keyword: cat
```

For each keyword, the user may enter one or multiple synonyms. For the keyword 'cat', for instance, the user may enter as first synonym the word 'kitty'.

```
We will be starting a new Thesaurus.
You may now enter a series of keywords and their synonyms.

Enter keyword: cat

You may enter one or more synonyms for "cat"
(please press "Enter" once done).
Enter synonym for "cat": kitty
```

After the first synonym has been entered, the user may then add additional synonyms. Once done the user presses enter. The user will then be asked whether he/she would like to add more keywords or not.

```
Enter keyword: cat

You may enter one or more synonyms for "cat"
(please press "Enter" once done).
Enter synonym for "cat": kitty
Enter synonym for "cat": feline
Enter synonym for "cat": kitten
Enter synonym for "cat": pussycat
Enter synonym for "cat":

Do you want to add more keywords? y/n:
```

If the user selects ‘y’, he/she will be prompted to enter the next keyword and synonyms to be added to the thesaurus (for instance adding keyword ‘brave’ and a collection of synonyms for ‘brave’).

```
Enter synonym for "cat": pussycat
Enter synonym for "cat":

Do you want to add more keywords? y/n: y

Enter keyword: brave

You may enter one or more synonyms for "brave"
(please press "Enter" once done).
Enter synonym for "brave": courageous
Enter synonym for "brave": valiant
Enter synonym for "brave": heroic
Enter synonym for "brave": bold
Enter synonym for "brave": daring
Enter synonym for "brave": fearless
Enter synonym for "brave": plucky
Enter synonym for "brave":

Do you want to add more keywords? y/n:
```

(*) Take note, the application should not allow the same synonym for a particular keyword to be entered twice. Neither should it allow the same keyword to be entered twice.

After the user has entered all keywords and synonyms, the application will then display the new thesaurus

```
Enter keyword: ferocious

You may enter one or more synonyms for "ferocious"
(please press "Enter" once done).
Enter synonym for "ferocious": fierce
Enter synonym for "ferocious": vicious
Enter synonym for "ferocious": violent
Enter synonym for "ferocious": cruel
Enter synonym for "ferocious": brutal
Enter synonym for "ferocious": aggressive
Enter synonym for "ferocious": wild
Enter synonym for "ferocious": unruly
Enter synonym for "ferocious": savage
Enter synonym for "ferocious": merciless
Enter synonym for "ferocious": barbarous
Enter synonym for "ferocious":

Do you want to add more keywords? y/n: n

Your new Thesaurus is ready and printed here....
brave: bold, courageous, daring, fearless, heroic, plucky, valiant
cat: feline, kitten, kitty, pussycat
ferocious: aggressive, barbarous, brutal, cruel, fierce, merciless, savage, unruly, vicious, violent, wild


Press Any key, to continue....
```

Take note the new thesaurus (per default) is sorted in alphabetically order for both the keywords as well as for the synonyms.

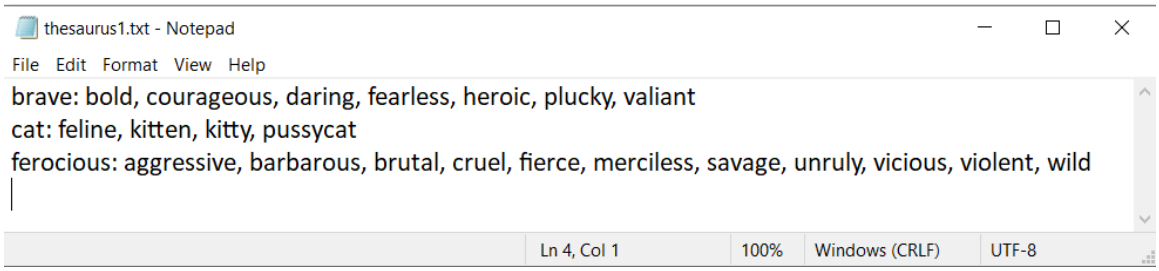
Saving a thesaurus

Once a user has finished setting up a new thesaurus, he/she may save it to a text file. For instance, the thesaurus could be saved under the filename 'thesaurus1.txt'

```
Please select your choice: (1,2,3,4,5,6,7,8,9,10)
        1. New
        2. Open
        3. Sort
        4. Process Text
        5. Extra Option One
        6. Extra Option Two
        7. Print
        8. Save
        9. Save As
        10. Exit
Enter choice: 9
Save As
Please enter new filename: thesaurus1.txt
Your file "thesaurus1.txt" has been saved

Press Any key, to continue....
```

You must ensure that your application stores the thesaurus with the same text format as is shown in below example:



- Keyword and synonyms are separated with ':' character.
- Synonyms are separated with ',' character

An existing thesaurus can be saved at any time under the same name by selecting option 9 (Save As).

Opening a thesaurus from a file

A user can read a thesaurus from a file by selecting option 2. The application will then print the thesaurus on the screen and wait for the user to press any key before the application continues.

```
Please select your choice: (1,2,3,4,5,6,7,8,9,10)
        1. New
        2. Open
        3. Sort
        4. Process Text
        5. Extra Option One
        6. Extra Option Two
        7. Print
        8. Save
        9. Save As
        10. Exit
Enter choice: 2

We will be opening an exisiting Thesaurus.
Please enter input file: thesaurus1.txt

Thesaurus "thesaurus1.txt" has been loaded and is printed here....
brave: bold, courageous, daring, fearless, heroic, plucky, valiant
cat: feline, kitten, kitty, pussycat
ferocious: aggressive, barbarous, brutal, cruel, fierce, merciless, savage, unruly, vicious, violent, wild


Press Any key, to continue....
```

Sorting the synonymns within a thesaurus

After a user sets up a new thesaurus, or opens a thesaurus from file, he/she may sort the synonyms within a thesaurus by selecting option 3. A second menu appears allowing the user to (repeatedly) choose a certain sorting method, until he/she presses option 5 as to return back to the main menu.

```
Press Any key, to continue....

Please select your choice: (1,2,3,4,5,6,7,8,9,10)
        1. New
        2. Open
        3. Sort
        4. Process Text
        5. Extra Option One
        6. Extra Option Two
        7. Print
        8. Save
        9. Save As
        10. Exit
Enter choice: 3

Please select your choice: (1,2,3,4,5)
        1. Alphabetically (Default)
        2. Length/Alphabetically
        3. Length/Random Alphabetically
        4. Randomly
        5. Back to Main Menu
Enter choice: 2
```

The application supports the following 4 sorting methods.

1. Alphabetically (Default)

```
Please select your choice: (1,2,3,4,5)
       1. Alphabetically (Default)
       2. Length/Alphabetically
       3. Length/Random Alphabetically
       4. Randomly
       5. Back to Main Menu
Enter choice: 1

Sorting Synonyms: Alphabetically (Default)
brave: bold, courageous, daring, fearless, heroic, plucky, valiant
cat: feline, kitten, kitty, pussycat
ferocious: aggressive, barbarous, brutal, cruel, fierce, merciless, savage, unruly, vicious, violent, wild


Press Any key, to continue....
```

2. Length/Alphabetically

```
Please select your choice: (1,2,3,4,5)
       1. Alphabetically (Default)
       2. Length/Alphabetically
       3. Length/Random Alphabetically
       4. Randomly
       5. Back to Main Menu
Enter choice: 2

Sorting Synonyms: Length/Alphabetically
brave: bold, daring, heroic, plucky, valiant, fearless, courageous
cat: kitty, feline, kitten, pussycat
ferocious: wild, cruel, brutal, fierce, savage, unruly, vicious, violent, barbarous, merciless, aggressive


Press Any key, to continue....
```

3. Length/Random Alphabetically

```
Please select your choice: (1,2,3,4,5)
       1. Alphabetically (Default)
       2. Length/Alphabetically
       3. Length/Random Alphabetically
       4. Randomly
       5. Back to Main Menu
Enter choice: 3

Sorting Synonyms: Length/Random Alphabetically
brave: bold, heroic, daring, plucky, valiant, fearless, courageous
cat: kitty, kitten, feline, pussycat
ferocious: wild, cruel, unruly, brutal, fierce, savage, violent, vicious, merciless, barbarous, aggressive


Press Any key, to continue....
```

4. Randomly

```
Please select your choice: (1,2,3,4,5)
       1. Alphabetically (Default)
       2. Length/Alphabetically
       3. Length/Random Alphabetically
       4. Randomly
       5. Back to Main Menu
Enter choice: 4

Sorting Synonyms: Randomly
brave: courageous, heroic, plucky, bold, fearless, daring, valiant
cat: pussycat, kitty, feline, kitten
ferocious: wild, cruel, merciless, barbarous, savage, fierce, violent, aggressive, brutal, vicious, unruly


Press Any key, to continue....
```

Processing text with a thesaurus

After a user sets up a new thesaurus, or opens a thesaurus from file, he/she may proceed with processing text by selecting option 4. The user will be requested to enter a filename for the text that needs to be processed. The text will then be displayed, and application waits until user presses any key.

```
Please select your choice: (1,2,3,4,5,6,7,8,9,10)
        1. New
        2. Open
        3. Sort
        4. Process Text
        5. Extra Option One
        6. Extra Option Two
        7. Print
        8. Save
        9. Save As
        10. Exit
Enter choice: 4

Select the file you want to process
Please enter input file: localnews1.txt

The text before processing:
A stray cat was recently spotted in the neighborhood protecting her little kitten as the two were being attacked
by a ferocious dog. The brave feline eventually managed to chase the aggressive dog away. This goes to show that
we should never underestimate a pussycat's maternal instincts when looking after her little kitty.

Press Any key, to continue....
```

After pressing any key, a second menu appears allowing the user to (repeatedly) choose a certain processing method, until he/she presses option 3 as to return to the main menu.

```
Next choose a Text Processing option.

Please select your choice: (1,2,3)
        1. Simplified Writing
        2. Elegant Writing
        3. Back to Main Menu
Enter choice: 1
```

The application supports the following 2 text processing methods, Simplified Writing and Elegant Writing.

Simplified Writing

Below screenshot shows the text input before processing and text output after Simplified Writing processing.

```
Select the file you want to process
Please enter input file: localnews1.txt

The text before processing:
A stray cat was recently spotted in the neighborhood protecting her little kitten as the two were being attacked
by a ferocious dog. The brave feline eventually managed to chase the aggressive dog away. This goes to show that
we should never underestimate a pussycat's maternal instincts when looking after her little kitty.

Press Any key, to continue....

Next choose a Text Processing option.

Please select your choice: (1,2,3)
        1. Simplified Writing
        2. Elegant Writing
        3. Back to Main Menu
Enter choice: 1

Processing text for: Simplified Writing

The text after processing:
 A stray cat was recently spotted in the neighborhood protecting her little cat as the two were being attacked by
 a ferocious dog. The brave cat eventually managed to chase the ferocious dog away. This goes to show that we sho
uld never underestimate a cat's maternal instincts when looking after her little cat.

Press Any key, to continue....

Do you want to save the text to a file? y/n: y
Please enter new filename: simplified.txt
The text has been saved in  "simplified.txt"

Press Any key, to continue....
```

Elegant Writing

Below screenshot shows the text input before processing and text output after Elegant Writing processing.

```
Select the file you want to process
Please enter input file: localnews2.txt

The text before processing:
A stray cat was recently spotted in the neighborhood protecting her little cat as the two were being attacked
 by a ferocious dog. The brave cat eventually managed to chase the ferocious dog away. This goes to show that
 we should never underestimate a cat's maternal instincts when looking after her little cat.

Press Any key, to continue....

Next choose a Text Processing option.

Please select your choice: (1,2,3)
        1. Simplified Writing
        2. Elegant Writing
        3. Back to Main Menu
Enter choice: 2

Processing text for: Elegant Writing

The text after processing:
 A stray feline was recently spotted in the neighborhood protecting her little kitten as the two were being a
ttacked by a aggressive dog. The bold kitty eventually managed to chase the barbarous dog away. This goes to
show that we should never underestimate a pussycat's maternal instincts when looking after her little feline.


Press Any key, to continue....

Do you want to save the text to a file? y/n: y
Please enter new filename: elegant.txt
The text has been saved in  "elegant.txt"

Press Any key, to continue....
```

Printing a Thesaurus

A thesaurus can be printed at any time by selecting option 7 from the menu.

```
Please select your choice: (1,2,3,4,5,6,7,8,9,10)
        1. New
        2. Open
        3. Sort
        4. Process Text
        5. Extra Option One
        6. Extra Option Two
        7. Print
        8. Save
        9. Save As
        10. Exit
Enter choice: 7

The Thesaurus "thesaurus1.txt" is  printed here....
brave: bold, daring, heroic, plucky, valiant, fearless, courageous
cat: kitty, feline, kitten, pussycat
ferocious: wild, cruel, brutal, fierce, savage, unruly, vicious, violent, barbarous, merciless, aggressive


Press Any key, to continue....
```

Exiting the application

The user can repeatedly select Options 1,2,….,9. Option 10 is to exit the program. You should follow the output as is shown below.

```
Press Any key, to continue....

Please select your choice: (1,2,3,4,5,6,7,8,9,10)
        1. New
        2. Open
        3. Sort
        4. Process Text
        5. Extra Option One
        6. Extra Option Two
        7. Print
        8. Save
        9. Save As
        10. Exit
Enter choice: 10

Bye, thanks for using ST1507 DSAA: Thesaurus Based Text Processor
```

Basic requirements:

- You are required to design and write the Python application using an object-oriented programing approach (OOP). You should leverage on your knowledge of encapsulation, function/operator overloading, polymorphism, inheritance etc.

- You may make use of Python's already build in data structures, such as list, tuple, dictionary etc., however you should refrain from using the classes from the collection library. Instead, you are required to write you own classes to support the various data structures that you may need (for instance for sorting you may need a class *SortedList*). Of course, you may refer to the lecture slides and lab tasks and expand further on those classes that we had previously developed and discussed in tutorials and lab sessions.

- The classes that you develop must be placed in separate python files.

- Pay attention to user input validation. Your application should not crash if a user types in the wrong input. Instead, when a user enters wrong input, you should notify the user, and allow him/her to enter again.

Deliverables

Your deliverables must include all the python files (.py files).

- Make sure that the code you submit is complete, and that it can run from the Anaconda prompt as:

```
python main.py
```

- Make sure your code is properly commented, and that for each python file you write on top in comments, your name, student ID and class.

- Include some examples of thesaurus, and text files for processing that you may have used for testing your application (take note assessors may test your application with their own thesaurus and text files).

Submission deadline

Final Submission is on **Friday 25 November 5:00 pm**.

-Submit all the deliverables (Source Code, test files) in the designated BrightSpace drop box before the deadline.

- You must submit it as one Zipped folder (RAR will not be accepted, only zip). Label your submission as:

CA1_Name_StudentID_Class.zip

For example: *CA1_JimmyTan_12345_2B10.zip*

Assessment Criteria

The assignment will be assessed based on the following criteria:

Assessment criteria	Marks awarded
User Interface and File IO - An interactive user interface that follows the required format, as shown in screenshots. - Proper user input validation (handles wrong user inputs correctly). - Reading and writing to files is functional.	Max 20
Functionality of the application: - Able to setup a thesaurus from scratch. - Able to sort and print thesaurus - Able to simplify and make text more elegant.	Max 30
Programming techniques, efficiency, robustness, and readability of code: - Usage of classes and OOP technology. - Usage of data structures (python build in data structures as well as those data structures you made yourself). - Functions and programming constructs. - Code is properly commented. - Application if free of crashes.	Max 30
Two Extra Features: - Features will be judged according to technical sophistication and useability of feature.	Max 10
Demo: - Ability to demonstrate and explain the application and the code clearly. - Q&A **(*) Multiplier on total score may be applied if student demonstrate poor understanding of code.**	Max 10
Grand Total	**100**

-- End --