COSC 2P03 Advanced Data Structures: Assignment 2

Instructor: Yifeng Li*[1]

[1]Department of Computer Science, Brock University

September 30, 2022

1 DrugBank and Drug Design

AI for drug design is one of several major research projects being conducted by our Brock Biomedical Data Science Laboratory. In these projects, we develop machine learning models that capture key knowledge from existing data and then apply them for prediction or molecular generation. The DrugBank dataset along with other data are used in our projects. In this assignment, we focus on playing with the DrugBank data. You are expected to finish the following tasks.

2 Your Tasks (Total: 12 marks)

You should define classes named `BinaryNode`, `Drug`, and `DrugBank` with the following requirements (feel free to define extra variables, classes, and methods if needed). Please note that, `DrugDesign` is actually a class for binary search tree.

1. Define the class `Drug` with data attributes `drugBankID`, `genericName`, `SMILES`, `url`, `drugGroups`, and `score`. This class has at least one method, named `displayDrug` to print out the information on the screen. **(0.5 mark)**

2. Define the `BinaryNode` class with an instance of the `Drug` class as its data attribute. This class has a method called `displayNode`. **(0.5 mark)**

3. Define a method named `readData` under the `DrugBank` class to load all the provided information from the given text file to an array variable (named `data` which is an attribute of the class). **(1 mark)**

4. Define an (either iterative or recursive) method named `create` to insert all drugs in the array into a binary search tree using `drugBankID` as the key. The root of the binary search tree is a data attribute of the class, and is named `root` of BinaryNode type. Hint: you may define a separate `insert` method. **(2 marks)**

5. Define a recursive method named `inOrderTraverse` under the `DrugBank` class to output the information of each drug (using the same format as the given file) into an output text file (named `dockedApprovedSorted.tab`) in the order of the traversal. **(2 marks)**

6. Define an (either iterative or recursive) method named `search` under the `DrugBank` class to search for a drug entry based on a given `drugBankID`. **(1 mark)**

7. Define a method named `delete` under the `DrugBank` class to remove a node with a drug entry of the given `drugBankID` from the binary search tree. **(2 marks)**

8. Define an (iterative or recursive) method named `depth1` under the `DrugBank` class to find the the depth of a node with a given key (i.e., `drugBankID`). **(1 mark)**

*E-mail address: `yli2@brocku.ca`

9. Define a recursive method named `depth2` under the `DrugBank` class to find the depth of the deepest node in the tree. **(1 mark)**

10. In the **main** function, instance (named `db`) of the `DrugBank` should be created. The following methods of this instance should be called sequentially:

    (a) `db.create(···)`

    (b) `db.inOrderTraverse()`

    (c) `db.depth1(‘‘DB01050’’)`

    (d) `db.depth2()`

    (e) `db.search(‘‘DB01050’’)`

    (f) `db.search(‘‘DB00316’’)`

    (g) `db.delete(‘‘DB01065’’)`.

    , **(0.5 mark)**

11. Your code should be well commented. **(0.5 mark)**

3 Submission

- Your source code.
- A PDF printout of your source code.
- The text file with result required in Task 10; and the another output file `dockedApprovedSorted.tab`.
- Compress the above files in a zipped folder named `COSC2P03_A2_Firstname_Lastname_StudentNumber.zip` and submit it through Brightspace before indicated due time.
- If any of the above require files are not submitted, 0 mark will be given to the whole assignment.
- Late submissions will not be accepted.

4 Academic Integrity

This assignment should be tackled individually. Outsourcing or teamwork is not allowed. Violation of this requirements will be seriously processed in accordance with university policies.