Work Integrated Learning Programmes Division
M.Tech (Data Science and Engineering)

(S1-20_DSECFZG519)

(Data Structures and Algorithms Design)

Academic Year 2020-2021

Assignment 1 – PS23 – [Employee Tracker] - [Weightage 12%]

1. Problem Statement

Given the current pandemic, an organization intends to use the unique employee ID and smart card to identify the whereabouts of their employees within the office premises. Whenever the employee swipes his card in the office the employee ID is recorded. When an employee enters into the office for the first time, the counter is set to 1. From then onwards, each time an employee swipes out of the office premises the counter is incremented, and incremented again when he enters back. If the counter is odd, it means the employee is in the office premises and if the counter is even, it means he is out of the premises.

The organization uses this data to perform the following analytics:

1. How many employees came today?
2. Number of employees that have come today and are currently in office?
3. Check specific employee whereabouts?
4. List of employees that have swiped (in or out) more than x number of times?
5. Which employee ids within a range of IDs attended office, the swipe counter for them, and whether they are inside or outside office?

Requirements:

1. **Implement the above problem statement in Python 3.7 using BINARY TREE ADT.**
2. **Perform an analysis for the questions above and give the running time in terms of input size: n.**

The basic structure of the employee node will be:

```
class EmpNode:
    def __init__(self, EId):
        self.EmpId = EId
        self.attCtr = 1
        self.left = None
        self.right = None
```

Operations:

1. **def _recordSwipeRec(self, eNode, Eid):** This function reads from the **inputPS23.txt** file the ids of employees entering and leaving the organization premises. One employee id should be populated per line (in the input text file) indicating their swipe (entry or exit). The input data is used to populate the tree. If the employee id is already added to the tree, then the swipe counter is incremented for every subsequent occurrence of that employee id in the input file. Use a trigger function to call this recursive function from the root node.

2. **Def _getSwipeRec(self, eNode):** This function counts the total number of employees who came to office today. The output is entered in the **outputPS23.txt** file as shown below.

   Total number of employees recorded today: ***xx***

   Use a trigger function to call this recursive function from the root node.

3. **def _onPremisesRec(self, eNode, eId):** This function searches for all employees that are still on the premises and outputs the total count of such employees. The function is triggered with the following tag from the file **promptsPS23.txt.**

   onPremises:

   If there are employees found who are on the premises the below string is entered into the **outputPS23.txt** file

   xx employees still on premises.

   If there are no employees on premises, the below string into the **outputPS23.txt** file

   No employees present on premises.

   Use a trigger function to call this recursive function from the root node.

4. **def _checkEmpRec(self, eNode, EId):** This function counts the number of times a particular employee swiped today and if the employee is currently in the office or outside.

The function reads the employee id from the file **promptsPS23.txt** where the search id is mentioned with the tag as shown below.

checkEmp:12

checkEmp:22

The search function is called for every checkEmp tag the program finds in the promptsPS23.txt file.

If the employee id is found with an odd swipe count the below string is output into the **outputPS23.txt** file

Employee id ***xx*** swiped **yy** times today and is currently in office

If the employee id is found with an even swipe count the below string is output into the **outputPS23.txt** file

Employee id ***xx*** swiped **yy** times today and is currently outside office

If the employee id is not found it outputs the below string into the **outputPS23.txt** file

Employee id ***xx*** did not swipe today.

Use a trigger function to call this recursive function from the root node.

5. **def _frequentVisitorRec(self, eNode, frequency):** This function generates the list of employees who have swiped more than x number of times. The function reads the x number from the file **promptsPS23.txt** with the tag as shown below.

freqVisit: 5

The function outputs the below string into the **outputPS23.txt** file.

Employees that swiped more than xx number of times today are:

Employee id, count.

Use a trigger function to call this recursive function from the root node.

6. **def printRangePresent(self, StartId, EndId):** This function prints the employee ids in the range **StartId** to **EndId** and how often they have swiped and if they are inside or outside the office into the **outputPS23.txt** file.

The input should be read from the **promptsPS23.txt** file where the range is mentioned with the tag as shown below.

range: 23:125

If Input range is given as 23 to 125 the output file should show:

Range: 23 to 125

Employee swipe:

23, 1, in

41, 3, in

121, 2, out

For this purpose, the tree needs to be **inorder traversed** and the id and frequency of the employees in the range must be printed into the file. If the Id is found in the BT, its frequency cannot be zero as the person had entered the organization at least once.

Sample file formats

The sample files shown below are only indicative of the structure that needs to be followed and the outputs need not be in line with the input and prompt file.

Sample Input file

Each row will have one employee id indicating a single swipe for that employee. The input file name must be called **inputPS23.txt**.

Sample inputPS23.txt

23
22
41
121
41
22
41
121

Sample promptsPS23.txt

onPremises:
checkEmp: 12
checkEmp: 22
freqVisit: 3
range: 23:125

Sample outputPS23.txt

Total number of employees recorded today: 4

2 employees still on premises.

Employee id 12 did not swipe today.

Employee id 22 swiped 2 times today and is currently outside office

Employees that swiped more than 3 number of times today are:

41,3

Range: 23 to 125

Employee swipe:

23, 1, in

41, 3, in

121, 2, out

.

Note that the input/output data shown here is only for understanding and testing, the actual file used for evaluation will be different.

2. Deliverables

1. Word document **designPS23_<group id>.docx** detailing your design and time complexity of the algorithm.
2. **[Group id]_Contribution.xlsx** mentioning the contribution of each student in terms of percentage of work done. Download the Contribution.xlsx template from the link shared in the Assignment Announcement.
3. **inputPS23.txt** file used for testing
4. **promptsPS23.txt** file used for testing
5. **outputPS23.txt** file generated while testing
6. **.py file** containing the python code. Create a single *.py file for code. Do not fragment your code into multiple files

Zip all of the above files including the design document and contribution file in a folder with the name:

[Group id]_A1_PS23_Employee.zip and submit the zipped file.

Group Id should be given as **Gxxx** where xxx is your group number. For example, if your group is 26, then you will enter G026 as your group id.

3. Instructions

1. It is compulsory to make use of the data structure(s) / algorithms mentioned in the problem statement.
2. Ensure that all data structure insert and delete operations throw appropriate messages when their capacity is empty or full. Also ensure basic error handling is implemented.

3. For the purposes of testing, you may implement some functions to print the data structures or other test data. But all such functions must be commented before submission.
4. Make sure that your read, understand, and follow all the instructions
5. Ensure that the input, prompt and output file guidelines are adhered to. Deviations from the mentioned formats will not be entertained.
6. The input, prompt and output samples shown here are only a representation of the syntax to be used. Actual files used to evaluate the submissions will be different. Hence, do not hard code any values into the code.
7. Run time analysis is to be provided in asymptotic notations and not timestamp based runtimes in sec or milliseconds.

Instructions for use of Python:

1. Implement the above problem statement using Python 3.7.
2. Use only native data types like lists and tuples in Python, do not use dictionaries provided in Python. Use of external libraries like graph, numpy, pandas library etc. is not allowed. The purpose of the assignment is for you to learn how these data structures are constructed and how they work internally.
3. Create a single *.py file for code. Do not fragment your code into multiple files.
4. Do not submit a Jupyter Notebook (no *.ipynb). These submissions will not be evaluated.
5. Read the input file and create the output file in the root folder itself along with your .py file. Do not create separate folders for input and output files.

4. Deadline

1. The strict deadline for submission of the assignment is **27th Dec, 2020.**
2. The deadline has been set considering extra days from the regular duration in order to accommodate any challenges you might face. No further extensions will be entertained.
3. Late submissions will not be evaluated.

5. How to submit

1. This is a group assignment.
2. Each group has to make one submission (only one, no resubmission) of solutions.
3. Each group should zip all the deliverables in one zip file and name the zipped file as mentioned above.
4. Assignments should be submitted via Canvas > Assignment section. Assignment submitted via other means like email etc. will not be graded.

6. Evaluation

1. The assignment carries 12 Marks.
2. Grading will depend on
    a. Fully executable code with all functionality working as expected
    b. Well-structured and commented code
    c. Accuracy of the run time analysis and design document.
3. Every bug in the functionality will have negative marking.
4. Marks will be deducted if your program fails to read the input file used for evaluation due to change / deviation from the required syntax.
5. Use of only native data types and avoiding libraries like numpy, graph and pandas will get additional marks.
6. Plagiarism will not be tolerated. If two different groups submit the same code, both teams will get zero marks.
7. Source code files which contain compilation errors will get at most 25% of the value of that question.

7. Readings

Text book: Algorithms Design: Foundations, Analysis and Internet Examples Michael T. Goodrich, Roberto Tamassia, 2006, Wiley (Students Edition). **Chapters:** 2.3