MIDDLE EAST TECHNICAL UNIVERSITY, NORTHERN CYPRUS CAMPUS
CNG213 Data Structures – Programming Assignment 3

**Date handed out: 31 May 2021, Monday**
**Date submission due: 14 June 2021, Monday 23:55 (Cyprus Time)**

**Data Analysis @Instagram**

This assignment aims to help you practice the AVL Tree ADT. You will write a program that can be used to analyse the data collected on the METU NCC page on Instagram. You will use this application to index the posts and find out the most popular post in a given period of time.

**Requirements:**
In this assignment, you are given a text file called "insta.txt" that includes the data about the posts on the METU NCC page on Instagram. This file includes the following information:

- **Post ID**: This is the unique id of the post made.
- **Permalink**: This is the unique URL for the post made.
- **Total reach**: The total number of people the post has been served to.
- **Engaged Users**: The number of unique people who engaged in certain ways with the post. For example, by commenting on, liking, sharing or saving the post.
- **Type**: It shows the type of the post which can be "Photo" or "Video".
- **Posted**: It shows the date and time of the post.

And sample data is as follows where each part is separated by a "," character:

```
12493332_1592322190;https://www.instagram.com/p/CPQJ3TyrnNI;157;478;Photo;24/
05/2021 23:39
12493000_1592322896;https://www.instagram.com/p/CPKzBxUKVCM;145;256;Photo;20/
05/2021 09:30
11193254_8962322522;https://www.instagram.com/p/COsNXRnL9v1;100;870;Photo;10/
05/2021 11:40
```

Your task here is to process this file, and generate an AVL tree based on the **total reach** data. If there is more than one post with the same total reach value, they should then be stored together in the same node. Based on this data representation, you need to write a program that provides the following functionalities to the user:

1. **Display the full posts:** This will display the full AVL tree constructed. For traversal, you need to display the posts in descending order based on total reach data.
2. **The post with the maximum total reach:** This will display the details (Post ID, Permalink, Type, Posted, Total reach and Engaged Users) of the post with the maximum total reach data.
3. **The post with the maximum engaged users:** This will display the details (Post ID, Permalink, Type, Posted, Total reach and Engaged Users) of the post with the maximum engaged users.

**Programming Requirements:**

You will start by taking the file name as a command line argument and then you will need to implement at least the following functions:

- **`readData()`:** This function will mainly process the external file. As an input, it will take the file name and it will return an AVL tree.
- **`insertPost()`:** This function will take an AVL tree, and the details of a post, and then it will insert the post to the AVL tree. The post will be interested into the tree based on the total reach value. You cannot make assumptions about the number of posts that has been taken at the same number of total reach value. Therefore, if the node with the given total reach value exists then you will add your post details to that node.
- **`displayPosts()`:** This function will mainly take an AVL tree and display the data in the tree in descending order according to the total reach value.
- **`mostReach()`:** This function will mainly take an AVL tree and display the details of the posts with the maximum total reach value. In the comment part of this function, discuss the complexity of this function based on your current representation of data. You also need to discuss if there is a way you could improve this.
- **`mostEngaged()`:** This function will mainly take an AVL tree and will find the post with the maximum engaged users. In the comment part of this function, discuss the complexity of this function based on your current representation of data. You also need to discuss if there is a way you could improve this.

Please note that in this assignment, you can make use of the functions in the **string.h** library and similar external libraries. You cannot assume about the number of posts in this external file.

**Submission Requirements:**

In this assignment, you need to have a header file (avltree.h) which includes the major functionality of the AVL Tree ADT. If you will use other ADTs, you need to create a separate header file for each of them. You also need to have a C source file (instaAnalysis.c) that includes the main function and other functions. You need put all these files into the "cng213a3" folder and then submit the compressed version of the folder to ODTU-CLASS. If you do not follow this structure, you will lose %10 from the overall grade.

**Programming Style Tips!**

Please follow the modular programming approach. In C programming, we use functions referred to modules to perform specific tasks that are determined/guided by the solution. Remember the following tips!

• Modules can be written and tested separately!
• Modules can be reused!
• Large projects can be developed in parallel by using modules!
• Modules can reduce the length of the program!
• Modules can also make your code more readable!

**Important Notes:**

- Remember to have good programming style (Appropriate comments, variable names, formulation of selection statements and loops, reusability, extensibility etc.). Each of the items above will include10% for good programming style.
- Read rules regarding to assignments from the Syllabus carefully.
- If your code does not compile due to syntax errors, you will automatically get zero.
- If your code includes a variable declaration inside a for loop such as for(int i=0; i<5;i++), you will automatically get zero.
- If your code includes global varibles, you will automatically get zero.

**Grading:**
Your program will be graded as follows:

| **Grading Point Mark** | **(out of 100)** |
|---|---|
| AVL Tree Data Structure | 5 |
| Processing data file (**`readData()`**) | 15 |
| Inserting/Updating a node in the tree (**`insertPost()`**) | 25 |
| Displaying the posts (**`displayPost()`**) | 10 |
| Displaying the details of the post which has maximum total reach value (**`(mostReach)`**) (complexity discussion – 2pts) | 15 |
| Displaying the details of the post which has maximum engaged users (**`mostEngaged()`**) (complexity discussion – 2pts) | 15 |
| The main function | 15 |

**NOTE**: Remember to have good programming style (Appropriate comments, variable names, formulation of selection statements and loops, reusability, extensibility etc.). Each of the items above will include10% for good programming style. Good programming style also includes modularity approach explained above.

**Sample run:**

```
>>instaAnalysis insta.txt
Welcome to Data Analysis @ Instagram
====================================

Data Processed:
===============
Post ID: 12493332_1592322190
Total Reach: 157 Engaged users: 478
Post ID: 12493000_1592322896
Total Reach: 145 Engaged users: 256
Post ID: 11193254_8962322522
Total Reach: 100 Engaged users: 870

Maximum Total Reach:
====================
Post Id: 12493332_1592322190
Permalink: https://www.instagram.com/p/CPQJ3TyrnNI
Type: Photo
Posted: 24/05/2021 23:39
Total Reach: 157
Engaged users: 478

Maximum Engaged Users:
====================
Post Id: 11193254_8962322522
Permalink: https://www.instagram.com/p/COsNXRnL9v1
Type: Photo
Posted: 10/05/2021 11:40
Total Reach: 100
Engaged users: 870
```