THE HONG KONG POLYTECHNIC UNIVERSITY

DEPARTMENT OF COMPUTING

EXAMINATION

Course : PG Scheme in COMP-61030, MSc Blockchain Technology-61036, MSc AI & BD Computing-61037, MSc Data Science & Analytics-63027

Subject : COMP5112 Data Structures and Database Systems

Group : 1011, 1311, 1111, 1071

Session : 2023 / 2024 Semester I

Date : 7 Dec 2023 Time : 12:30-14:30

Time Allowed: 2 Hours Subject Lecturer: Dr YIU Ken

This question paper has ____5___ pages (cover included).

Instructions to Candidates:

Write ALL your answers in the answer book.
You may use a calculator.

Do not turn this page until you are told to do so!

Question 1. **[25 marks]**

1(a) **[8 marks]**

Draw a binary tree such that all the following conditions are satisfied:

- the data element of each node is a character,
- the preorder traversal prints the sequence: `HOUSEWARMING`, and
- the inorder traversal prints the sequence: `OSUEHARWIMNG`.

Hint: find the root of the tree first by using the properties of tree traversal.

1(b) **[10 marks]**

Insert the following sequence of keys (one-by-one) into an empty AVL tree:

`2, 3, 5, 7, 11, 13, 17, 19.`

Draw the trees as follows:

- **Draw** the tree <u>just before you perform a rotation</u>.
- **Draw** the final tree.

1(c) **[7 marks]**

Write an algorithm to find the *k*-th smallest value in the unsorted array A[0..n–1].

For example, the 4-th smallest value of `5,3,13,7,11,2,17,19` is `7`.

Hint: you may use a Heap (either Max-Heap or Min-Heap) and call its operations (e.g., size( ), isEmpty( ), top( ), push(e), pop( )).

Analyze the time complexity of your algorithm.
Marks will be given based on the correctness and the time complexity of your algorithm.

Question 2. **[20 marks]**

2(a) **[10 marks]**

You are given two doubly linked lists of sorted integers (in the ascending order).
Let L and R be the head pointers of these two doubly linked lists.
Write an algorithm to merge them into a linked list of sorted integers.
Marks will be given based on the correctness and the memory size used by your algorithm (e.g., avoid creating new objects/arrays if possible).

2(b) **[10 marks]**

We are given a string of brackets that contains round brackets only: `'('` and `')'`.
A string `S` of brackets is said to be well matched if it can be expressed as any of the following:

- the sequence `'(', ')'`
- the sequence `'(', U, ')'`, where `U` is well matched
- the sequence `U, V`, where both `U` and `V` are well matched

Here are examples of strings that are well matched.

- `"()"`
- `"()()()"`
- `"((()))"`
- `"(()(())())"`

Write an algorithm to check whether a string S (i.e., an array of characters S[0..n–1]) is well matched.
Hint: you may use a Stack and call its operations (e.g., size( ), isEmpty( ), top( ), push(e), pop( )).

Question 3. **[30 marks]**

A movie recommendation website stores the following four tables in a database.

- Schema of *Person*: (person_id, name, email, phone)
- Schema of *Movie*: (movie_id, title, publisher, year, budget)
- Schema of *Role*: (movie_id, person_id, role)
- Schema of *Review*: (person_id, movie_id, review_date, comments)

All string values (in name, title, publisher, role, comments) are stored in the uppercase format. Examples include: `"KEVIN FEIGE"`, `"INCEPTION"`, `"PRODUCER"`.

3(a) **[10 marks]**

Write a SQL query for each task below:

(i) Find the person_id of each person who has reviewed at least two movies whose `PRODUCER` is `"KEVIN FEIGE"`.

(ii) Find the average budget of the movies such that `"KEVIN FEIGE"` is a `PRODUCER` and `"CHRISTIAN BALE"` is an `ACTOR`.

3(b) **[10 marks]**

Write relational algebra expression(s) for each task below:

(i) Find the name of each person who has reviewed only movies whose `PRODUCER` is `"KEVIN FEIGE"`.

(ii) Find the name of each person who has reviewed all movies whose `PRODUCER` is `"KEVIN FEIGE"`.

Note: If your relational algebra expression is too complicated, you may write it by using multiple assignment statements, e.g.,

$$\begin{array}{|l|}\hline \text{Temp1} \leftarrow \Pi_Y(\sigma_{X=1}(\text{R} \rhd\lhd \text{T})) \\ \text{Temp2} \leftarrow \Pi_Y(\sigma_{X=2}(\text{S} \rhd\lhd \text{T})) \\ \text{Result} \leftarrow \text{Temp1} \cup \text{Temp2} \\ \hline \end{array}$$

3(c) **[10 marks]**

Suppose that the following type of query is frequently used in DBMS

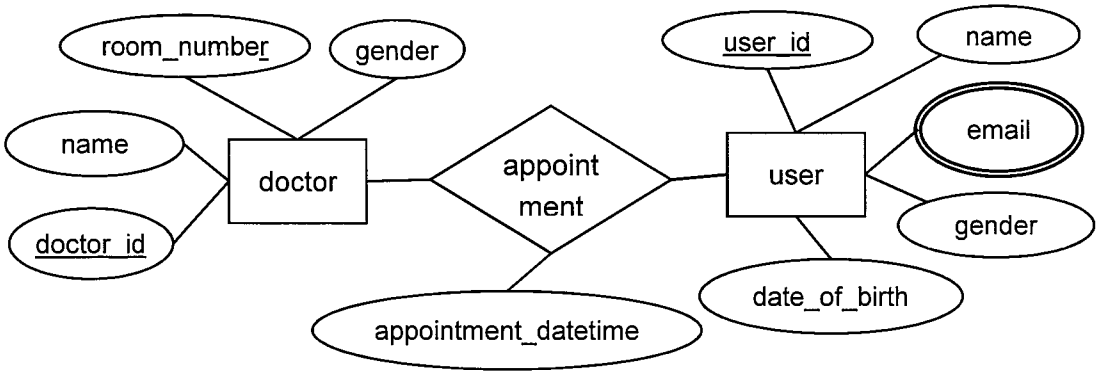
- Find the person_id of each person who has reviewed at least [X] movies in which the person named [Y] takes the role [Z].

Suggest the file organization and indexing that can be used to answer the above query as fast as possible.

Question 4. **[25 marks]**

4(a) **[8 marks]**

Convert the following ER diagram to relational schemas.
Explain your decisions briefly.



4(b) **[10 marks]**

Suppose that a relation R(A,B,C,D,E) satisfies the following functional dependencies:

- A → C
- A → E
- B → D
- CD → A
- CD → B
- CD → E

Find all the candidate keys for R.
Show your steps.

4(c) **[7 marks]**

Check whether the relation in Question 4(b) is in BCNF.
If it is not in BCNF, **decompose** it into BCNF.
Show your steps.

~~~ END OF THE PAPER ~~~
~~~