

**PhD Qualifier Examination**  
**Department of Computer Science and Engineering**

Date: 29-Mar-2017

Part A, SIT Syllabus

Maximum Marks: 100

---

Answer any five questions from Part A.

---

A.1 Supply short answers to the following parts.

(2 × 5)

(a) Write the asymptotic tight bound of  $T(n)$  defined as

$$T(n) = \begin{cases} T(3n/4) + 1 & \text{if } n \geq 4, \\ n & \text{otherwise.} \end{cases}$$

(b) A binary search tree contains an odd number of integers. Assume that each node in the tree stores only a key and two child pointers. What is the best-case time complexity to find the median of these integers from the tree?

(c) An unordered one-dimensional array  $X$  contains  $4n$  distinct elements. We need to partition  $X$  into  $X_1, X_2, X_3, X_4$  with  $n$  elements in each part, such that they are ordered, that is,  $a_1 < a_2 < a_3 < a_4$  for all  $a_1 \in X_1, a_2 \in X_2, a_3 \in X_3, a_4 \in X_4$ . The elements in the individual parts need not be ordered after the partition. Write the time complexity of the best algorithm for this.

(d)  $L$  and  $M$  are two linked lists, each containing  $n$  distinct elements sorted in increasing order. There may, however, be some elements common to  $L$  and  $M$ . What is the worst-case time complexity for the construction of a new linked list  $N$  that would contain all elements of  $L$  and  $M$  but without any duplicate?

(e) A star graph of order  $n$  is a tree having a central node of degree  $n$  and  $n$  other nodes of degree one each. Given the adjacency-list representation of a (general) graph  $G$ , what would be the worst-case time complexity to decide whether  $G$  is a star graph?

A.2 Let  $G = (V, E)$  be an undirected graph with  $n$  vertices numbered  $0, 1, 2, \dots, n-1$ . You maintain a max-priority queue  $Q$  of the vertices with the vertex degrees used for heap ordering. Suppose that the graph is dynamic in the sense that new edges are added and existing edges are deleted frequently. Since the addition or deletion of an edge changes vertex degrees, the queue  $Q$  should be appropriately and efficiently modified after each such operation. In the following parts, you do not need to consider the storage of  $G$  or its modifications necessitated by edge additions and deletions.

(a) What data structures should you maintain in addition to the storage of  $G$ ? (3)

(b) Mention how  $Q$  can be efficiently modified after each addition of a new edge. (5)

(c) What is the running time of your algorithm of Part (b)? (2)

A.3 Prove that any comparison-based sorting algorithm requires  $\Omega(n \log n)$  comparisons in the worst case. (10)

A.4 Let  $A$  be an  $m \times n$  matrix of distinct integers, that is, no two entries of the matrix are the same. For each  $i \in \{1, 2, \dots, m\}$ , let  $f(i)$  denote the index of the column containing the minimum element of row  $i$ . It is given that  $f(1) \leq f(2) \leq \dots \leq f(m)$ .

Here is a description of a divide-and-conquer algorithm that computes  $f(i)$  for each row  $i \in \{1, 2, \dots, m\}$ .

“Construct a submatrix  $A'$  of  $A$  consisting of the even numbered rows of  $A$ . Recursively determine the index of the minimum element of each row of  $A'$ . Then compute the index of the minimum element in the odd numbered rows in  $A$ .”

(a) Suppose we know  $f(i)$  for each even-numbered row  $i$ . Explain how to use this information to compute  $f(j)$  for all odd-numbered row  $j$  in  $O(m+n)$  time. (5)

(b) Write the recurrence describing the running time of the divide-and-conquer algorithm described above. Show that the running time of your algorithm is  $O(m+n \log m)$ . (5)

A.5 Let  $G = (V, E)$  be an undirected graph on  $n$  vertices  $v_1, v_2, \dots, v_n$ . Let  $N(v_i) := \{v_j \in V \mid (v_i, v_j) \in E\}$  denote the *neighborhood* of  $v_i$ , and  $N^h(v_i) = \{v_j \in N(v_i) \mid j > i\}$  the *higher neighborhood* of  $v_i$  (that is, the set of neighbors with higher indices). Your task is to find out whether for every  $i \in \{1, 2, \dots, n\}$ ,  $N^h(v_i)$  is a clique. (A subset of vertices of a graph is called a *clique* if the subgraph induced on this subset is complete. That is, every two distinct vertices in such a subset are neighbors of each other.) Assume that you are given an adjacency-matrix representation of the graph  $G$ .

(a) Design a  $\Theta(n^3)$ -time algorithm to solve the problem. (3)

(b) Design a  $\Theta(n^2)$ -time algorithm to solve the problem. (Hint: Let  $N^h(v_i) = S = \{v_j, v_k, v_l, v_m, \dots\}$  with  $i < j < k < l < m < \dots$ . If  $S$  is a clique, then  $S \setminus \{v_j\} \subseteq N^h(v_j)$ ,  $S \setminus \{v_j, v_k\} \subseteq N^h(v_k)$ , and so on.) (7)

A.6 Let  $A[0 \dots n-1]$  be an array of  $n$  distinct integers. We call  $A$  to be of zigzag type if for some indices  $i, j$  satisfying  $0 < i < j < n-1$ ,  $A[0 \dots i]$  is sorted in increasing order,  $A[i \dots j]$  is sorted in decreasing order, and  $A[j \dots n-1]$  is sorted in increasing order. For example, the array  $[2, 3, 9, 7, 6, 4, 1, 5]$  is of zigzag type with  $i = 2$  and  $j = 6$ . Write a C/C++/Java function that takes  $A$  and its size  $n$  as input. The function first checks whether  $A$  is of zigzag type. If not, it reports *failure*, and returns. Otherwise, it prints the three subarrays  $A[0 \dots i]$ ,  $A[i \dots j]$ , and  $A[j \dots n-1]$  in three separate lines, and returns. (10)

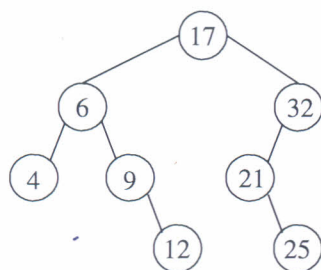
A.7 A sequence  $a_0, a_1, a_2, \dots$  is defined inductively as follows.

$$\begin{aligned} a_0 &= 1, \\ a_{2k+1} &= a_{2k} + a_k + 1 \text{ for all } k \geq 0, \\ a_{2k+2} &= a_{2k+1} + a_{k+1} \text{ for all } k \geq 0. \end{aligned}$$

(a) Write a recursive C/C++/Java function to compute and return  $a_n$  (given  $n$  as input). (5)

(b) Write an efficient iterative C/C++/Java function to compute and return  $a_n$  (given  $n$  as input). (5)

A.8 An array-based representation of a binary search tree  $T$  is demonstrated in the following figure.



Size	8
Root	3
Key	9 12 21 17 32 4 25 6 ...
Left	-1 -1 -1 7 2 -1 -1 5 ...
Right	1 -1 6 4 -1 -1 -1 0 ...
	0 1 2 3 4 5 6 7

We store the size (number of nodes) of  $T$ ; call it  $n$ . We store information about the  $n$  nodes in the first  $n$  locations of three arrays Key, Left, and Right. Each index  $i \in \{0, 1, 2, \dots, n-1\}$  stores the particulars of one node. Key[i] stores the key stored at this node, Left[i] stores the index of the left child ( $-1$  if this child does not exist), and Right[i] stores the index of the right child (or  $-1$  if no right child exists). The index of the root node is maintained separately. The dotted box in the above figure is such a representation of the tree on the left. The three arrays are not to be kept sorted, that is, any node can appear at any index.

(a) Propose a user-defined data type to store a binary search tree in this format (the dotted box in the above figure). (3)

(b) Write an efficient C/C++/Java function to insert a key  $x$  into a binary search tree  $T$  stored in this representation. Recall that if  $x$  already resides in  $T$ , no change is made. Otherwise, a new node is inserted at an appropriate position such that the binary-search-tree ordering is maintained. (7)



## Comprehensive Exam Spring 2016-17

### DBMS (Total Marks 25)

1. Consider a hospital having several departments (like cardiology, orthopedics, neurology, etc.). Each doctor can belong to only one department. A patient may be admitted under multiple departments and be assigned to one or more doctors. There is a date of admission and a date of release for each patient.

A patient is identified by a patient\_id. For each patient, we also need to maintain information like name, address and one or more phone numbers. For each doctor, we need to maintain doctor\_id, name, date of birth, department and highest degree. A department is identified by its department\_id. Other attributes of a department are department name, name of the HoD and number of wards it has.

- (a) Draw an Entity-relationship (E-R) diagram for capturing the above-mentioned information, clearly identifying all the entities, relationships, cardinalities and attributes (including the type of attribute like key attribute, multi-valued attribute, derived attribute, etc.).
- (b) Derive the relational schema from the above E-R model. The schema should not have any redundant relation and it should be in 3NF. For each relation, clearly identify all the columns, primary and foreign keys. **[5+5=10]**

2. Consider a relational schema with the following relations P(ABC), Q(ABD) and R(AE):

P(ABC)			Q(ABD)			R(AE)	
A1	B1	C1	A1	B1	2	A1	E1
A1	B1	C3	A1	B2	5	A2	E3
A2	B2	C3	A2	B3	1	A3	E2
A2	B1	C2	A1	B1	3	A1	E3
A3	B3	C2	A2	B2	6	A2	E4
			A3	B3	1	A3	E4

Show the intermediate results and the final result of evaluating the following relational algebra query on the above schema:

$$\Pi_{ACE}[(\sigma_{(B=B2 \vee C=C3)}[P \bowtie R]) \bowtie (\sigma_{(D>1)}[Q \bowtie R])] \quad [5]$$

3. Assume that the maximum number of key values that can fit in any leaf as well as non-leaf node of a B+ tree is 3.
- (a) Construct the tree with the following set of key values entered in the order shown: ACB, ACE, BCD, BDE, CGH, ACD, CAB, ABD, ABE, ABF. Show the intermediate tree at the end of each insert.
- (b) If the key value ACD is now deleted, show the tree that will result. **[4+2=6]**
4. Consider a schedule of 3 transactions T1, T2 and T3 as shown below where r1(x) denotes "transaction T1 reads data item x", w3(m) denotes "transaction T3 writes data item m", etc.

S: r1(x), r1(y), w3(x), w2(y), w1(z), r3(z), r2(z), r1(m), w3(m)

Is the schedule S given above view serializable? If so, it is view equivalent to which serial schedule? If not, why not? **[2+2=4]**

**Comprehensive Exam Spring 2016-17**

**Software Engineering (Total Marks =25)**

**Q.1** What is the difference between the functional and the non-functional requirements of a system? Identify at least two functional requirements of a bank ATM (Automated teller machine) system. Also identify one non-functional requirement for an ATM system. **[2+2+2 Marks]**

**Q.2** Give meaningful examples of each of the following types of relationships among classes in an object-oriented program. Write only a class diagram in UML syntax in each case. **[4 X 1 Marks]**

- i) Inheritance
- ii) Association
- iii) Aggregation
- iv) Dependency

**Q.4** Design the black-box test suite for a function named quadratic-solver. Quadratic-solver accepts three floating point numbers (a,b,c) representing a quadratic equation of the form  $ax^2 + bx + c = 0$ . It computes and displays the solution. **[5]**

**Q.5** Answer all the following questions by selecting the most appropriate option. **[20X 0.5 Marks]**

- i) The operation phase in the waterfall model is a synonym for which one of the following phases?
  - a) Coding and unit testing phase
  - b) Integration and system testing phase
  - c) Maintenance phase
  - d) Design phase
- ii) Which one of the following phases accounts for the maximum effort during development of a typical commercial software?
  - a) Coding
  - b) Testing
  - c) Designing
  - d) Specification
- iii) Which one of the following life cycle models lacks the characteristics of iterative software development?
  - a) Spiral model
  - b) Prototyping model
  - c) Classical waterfall model
  - d) Evolutionary model
- iv) The radial dimension of the spiral model represents which one of the following?
  - a) Cost incurred so far
  - b) Number of features implemented so far
  - c) Progress in the implementation of the current feature
  - d) Number of risks that have been resolved so far
- v) Effort is measured using which one of the following units?
  - a) Persons
  - b) Person-months
  - c) Months
  - d) Rupees
- vi) COCOMO estimation model can be used to estimate which one of the following?
  - a) LOC

- b) Effort
  - c) Function points
  - d) Defect density
- vii) If a project is already delayed, then adding manpower to complete it at the earliest would be:
- a) Always counter productive
  - b) Can help to a very limited extent
  - c) Most effective way to tackle the situation
  - d) Can cause project completion in the shortest time
- viii) Who among the following is a stakeholder in a software development project?
- a) A shareholder of the organization developing the software
  - b) Anyone who is interested in the software
  - c) Anyone who is a source of requirements for the software
  - d) Anyone who might be affected by the software
- ix) Which of the following is not a goal of requirements analysis?
- a) Weed out ambiguities in the requirements
  - b) Weed out inconsistencies in the requirements
  - c) Weed out non-functional requirements
  - d) Weed out incompleteness in the requirements
- x) The modules in a good software design should have which of the following characteristics:
- a) High cohesion, low coupling
  - b) Low cohesion, high coupling
  - c) Low cohesion, low coupling
  - d) High cohesion, high coupling
- xi) Which of the following is a black box testing approach:
- a) Path testing
  - b) Boundary value testing
  - c) Mutation testing
  - d) Branch testing
- xii) Which of the following can be considered as a program validation technique:
- a) Unit testing
  - b) Integration testing
  - c) Code review
  - d) Acceptance testing
- xiii) Alpha and Beta testing are considered to be which one of the following types of testing?
- a) Regression testing
  - b) Unit testing
  - c) Integration testing
  - d) System testing
- xiv) For a large program which one of the following integration testing strategy is rarely used:
- a) Big-bang
  - b) Top-down
  - c) Bottom-up
  - d) Mixed
- xv) Which one of the following is the basic focus of modern quality paradigms?



- a) Process assurance
- b) Product assurance
- c) Thorough testing
- d) Thorough testing and rejection of bad products

xvi) Which of the following statements about Cyclomatic complexity metric of a program is FALSE?

- a) It is a measure of the testing difficulty of the program.
- b) It is a measure of understanding difficulty of the program.
- c) It is a measure of the linearly independent paths in the program
- d) It is a measure of the size of the program

xvii) A legacy software product refers to a software that is:

- a) Developed at least 50 years ago.
- b) Obsolete software product.
- c) Software product that has poor design structure and code.
- d) Software product that could not be tested properly before product delivery.

xviii) Which one of the following best characterizes *inheritance*?

- a) It is same as encapsulation.
- b) Aggregation of information.
- c) Generalization and specialization.
- d) Polymorphism

xix) Which of the following indicates "Is a kind of" relationship?

- a) Aggregation
- b) Association
- c) Dependency
- d) Inheritance

xx) Which of the following is a characteristic of a good object-oriented design:

- a) Deep class hierarchy
- b) Large number of methods per class
- c) Large number of message exchanges per use case
- d) Moderate number of methods per class