THE AUSTRALIAN NATIONAL UNIVERSITY

Second Semester 2006

COMP1140
(Data Structures and Algorithms)

Writing Period: 3 hours duration

Study Period: 15 minutes duration

Permitted Materials: None

Answer ALL Questions

All your answers must be written in the boxes provided in this booklet. You may be provided with scrap paper for working, but it must **not** *be used to write final answers.*

There is additional space at the end of the booklet in case the boxes provided are insufficient. Label such overflow boxes with the question number.

Do not remove this booklet from the examination room.

Name (family name first):

Student Number:

Official use only:

Q1 (10)	Q2 (15)	Q3 (25)	Q4 (25)	Q5 (25)	Total (100)

QUESTION 1 [10 marks]

(a) Given an array of n elements, containing three distinct keys, `true, false` and `maybe`. Devise an $O(n)$ algorithm to rearrange the list so that all `false` elements precede the `maybe` elements, which in turn precede `true` elements. You are allowed to use only constant extra space.

QUESTION 1(a) **[4 marks]**

(b) Given an integer sequence $16, 4, 28, 42, 15, 38, 7, 20, 9, 31$, which is stored in an array $A[1..10]$.

(i) Construct a binary max-heap for the numbers in the sequence.

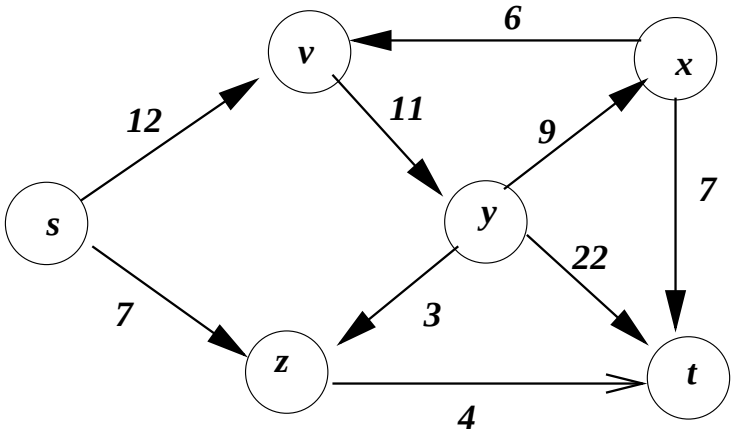
QUESTION 1(b) (i) **[4 marks]**

(ii) Show the total time for building a binary max-heap is bounded by $O(n)$ if the heap contains n elements.

QUESTION 1(b) (ii) **[2 marks]**

QUESTION 2 [15 marks]

(a) Given the following directed graph $G(V, E)$ with source s, the problem is to find the shortest path from s to every other vertex $v \in V - \{s\}$.



Apply Dijkstra's algorithm to the problem, and show the main intermediate steps.

QUESTION 2(a) **[8 marks]**

QUESTION 2(a) (continue) **[8 marks]**

(b) How to modify Floyd-Warshall's algorithm for all pairs of shortest paths problem to find the transitive closure of a directed graph $G(V, E)$? A directed graph $G(V, E')$ is the transitive closure of $G(V, E)$, if there is a directed path in G from node u to node v, then there is a directed edge $\langle u, v \rangle \in E'$. Describe the major steps of the proposed algorithm.

QUESTION 2(b) **[7 marks]**

QUESTION 3 [25 marks]

In each of the following questions, justify your answers with the appropriate formulas, principles, and/or explanations.

(a) In the game of "Hollywood squares", X's and 0's may be placed in any of the nine squares of a tic-tac-toe board (a 3x3 matrix) in any combination (i.e., unlike ordinary tic-tac-toe, it is not necessary that X's and 0's be placed alternately, so, for example, all the squares could wind up with X's). Squares may also be blank, i.e., not containing either an X or and 0. How many different boards are there?

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

(b) How many distinct ways are there to seat 6 people around a circular table? Note that because a round table has no obvious starting point, seatings are not "distinct" if they are rotations of one another.

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

(c) Suppose we have infinite supplies of red balls, blue balls, and green balls. How many ways are there of choosing m balls? Order of the balls is **un**important. *Hint: Solve the problem for two colors first, and then by "induction" for three colors. A simple formula (and there is such a formula!) that works for any m gives full credit, an algorithms that computes the number gives partial credit.*

QUESTION 3(c) **[6 marks]**

(d) Calculate the probabilities of the following events. The probability space is the deal of two cards in order from an ordinary 52-card deck.

- At least one card is an Ace
- The cards are of the same rank
- The cards are of the same suit
- The cards are of the same rank and suit
- The cards are the same either in rank or in suit
- The first card is of a higher rank than the second card.

QUESTION 3(d) **[6 marks]**

QUESTION 3(d) (continue) **[6 marks]**

(e) There are three prisoners, A, B, and C. They are told that one and only one is randomly chosen to get sentenced, the other two get free, and that the guard knows who. A asks the guard to tell him the name of one of the other prisoners who will get free. The guard answers that B gets free.

A reasons that either he or C gets sentenced, so the probability that A gets free is 1/2. On the other hand, reasons A, no matter who gets sentenced, the guard knows somebody besides A who gets free, so he always has an answer to A's question. Therefore, the asking and answering of the question provides no information about whether or not A gets sentenced, so the probability that A gets sentenced is still 1/3, as it was before the question was asked.

What is the true probability that A gets sentenced after the sequence of events described above? *Hint: You need to construct an appropriate probability space, one that represents not only the experiment in which a prisoner is chosen to get sentenced but also the possibility that the guard has a choice of whether to answer "B" or "C", and the experiment in which he chooses one if so.*

QUESTION 3(e) **[7 marks]**

QUESTION 3(e) (continue) **[7 marks]**

QUESTION 4 [25 marks]

Consider the following function for computing CNF of a formula:

```
class Formula {
  int var;            //variable index
  int opr;            //logical operator
  boolean isVar;      //true for variable nodes
  Formula sub1, sub2; // subformulas
}
// Logical operators
int FALSE_OP = 0, TRUE_OP = 1, NOT_OP = 2, OR_OP = 3, AND_OP = 4;

Formula CNF(Formula f)
{
   if(f == null || f.isVar || f.opr == NOT_OP) return f;
   if(f.opr == AND_OP) {
      f.sub1 = CNF(f.sub1); f.sub2 = CNF(f.sub2);
   }
   else if(f.opr == OR_OP)
   {
      if(f.sub1.opr == AND_OP)
      {
         Formula f1 = new Formula(OR_OP, copy(f.sub1.sub1), copy(f.sub2));
         Formula f2 = new Formula(OR_OP, copy(f.sub1.sub2), copy(f.sub2));

         f1 = CNF(f1); f2 = CNF(f2);
         f.opr = AND_OP;
         f.sub1 = f1; f.sub2 = f2;
      }
      else if(f.sub2.opr == AND_OP)
      {
         Formula f1 = new Formula(OR_OP, copy(f.sub1), copy(f.sub2.sub1));
         Formula f2 = new Formula(OR_OP, copy(f.sub1), copy(f.sub2.sub2));
         f1 = CNF(f1);  f2 = CNF(f2);
         f.opr = AND_OP;
         f.sub1 = f1; f.sub2 = f2;
      }
      else {
         Formula f1 = CNF(f.sub1);
         Formula f2 = CNF(f.sub2);
         f.sub1 = f1; f.sub2 = f2;
      }
   }
   return f;
}
```

The function `CNF` assumes the input formula is already in NNF. The function is incomplete. That is, there are some inputs on which the function will give wrong answers.

(a) Give an input formula (which is in NNF but not CNF) such that the output of the function applied to this formula is not in CNF. What should be added to the function to fix the problem?

QUESTION 4(a) **[10 marks]**

(b) Prove that the formula

$$((p \rightarrow q) \vee r \vee t) \rightarrow (p \rightarrow (q \vee r \vee t))$$

is a tautology using the resolution method.

QUESTION 4(b) **[5 marks]**

(c) Consider the following statement:

The number of non-terminal nodes of the reduced BDD of a given formula F is bigger than or equal to the number of distinct variables in F.

Assume that the variables in F are indexed by natural numbers and they are ordered according to their indices. Is the statement true or false? If it is true, prove it, otherwise give a counterexample, i.e., find a formula F such that its reduced BDD has fewer number of non-terminal nodes than the number of distinct variables in F.

QUESTION 4(c) **[5 marks]**

(d) Give the reduced BDD representation of the formula

$$x_1 \wedge (x_2 \vee x_3) \vee \neg(x_4 \vee \neg x_5) \vee x_4 \wedge x_6.$$

The variables are ordered according to their indices. (Hint: finding the reduced form of a BDD can often be done faster if the formula is in disjunctive normal form. Convert first the formula to DNF and then compute the reduced BDD.)

QUESTION 4(d) **[5 marks]**

Answer box continued over page.

QUESTION 5 [25 marks]

Given the following database scheme for representing information about sellers, their products and prices etc.:

	Relation	Explanation
1.	seller_loc(seller, address, postcode, city)	location of sellers
2.	seller_offer(seller, product, price)	products offered by sellers and current price
3.	seller_rating(seller, average rating, #ratings)	average ratings of sellers,
4.	product_list(product, product category)	category of products
5.	buyer_loc(buyer, address, postcode, city)	location of buyers
6.	transactions(buyer, product, price, seller, rating, date)	all previous transactions: "buyer" bought "product" from "seller" at "price" on "date" and "rating" of the transaction by the "buyer" (-5 to +5)
7.	distances(postcode, postcode, distance)	distance between suburbs

1. Formulate the following queries in relation algebra. You can use the additional operators: $min_X(Y)$ and $max_X(Y)$ which select the tuple of relation Y with the minimum/maximum X value. $sum_X(Y)$ gives the sum of all numerical values of attributes X in table Y.

   (a) Buyer "John123" wants to buy product "Fridge15". From which seller can he get the cheapest price?

   QUESTION 5. 1.(a) **[1 mark]**

   (b) Same as (a), but only consider sellers located within 50km of where "John123" lives.

   QUESTION 5. 1.(b) **[2 marks]**

(c) Same as (b), but only sellers with a positive average rating.

QUESTION 5. 1.(c) **[2 marks]**

(d) List all buyers who bought the same product categories as “John123”.

QUESTION 5. 1.(d) **[2 marks]**

(e) How much could “John123” have saved if he had bought all products at the cheapest price within two weeks of when he bought them?

QUESTION 5. 1.(e) **[3 marks]**

2. Make a change to the database scheme so that at least one query is faster while the run-time for the other queries does not increase. Please justify briefly. (Do not just remove redundant information!)

QUESTION 5.2 **[3 marks]**

3. What attribute(s) could be added/changed so that the information contained in the above database scheme would be practically more useful? Please justify briefly.

QUESTION 5.3 **[3 marks]**

4. What are the keys for the relations mentioned above?

QUESTION 5.4 **[2 marks]**

5. What is best suited as primary index for the relations mentioned above taking into account the given queries? Please justify briefly.

QUESTION 5.5 **[3 marks]**

6. Prove or disprove the following equality:

$$((R \bowtie_{A=B} S) \bowtie_{C=D} T) \equiv (R \bowtie_{A=B} (S \bowtie_{C=D} T))$$

where A is an attribute of R, B, C are different attributes of S, and D is an attribute of T.

QUESTION 5.6 **[4 marks]**

Additional answers. Clearly indicate the corresponding question and part.

Additional answers. Clearly indicate the corresponding question and part.

Additional answers. Clearly indicate the corresponding question and part.

Additional answers. Clearly indicate the corresponding question and part.