COMP 7270/76 Spring 2013 Final Exam 150 Minutes Time Limit Strictly Enforced 5 Questions 100 Points

Name:________________________________

Write and draw neatly and clearly. If we can't read it, you will NOT get credit even if it is correct.
Each question carries 20 points.

1. Consider the efficient algorithm to compute the Maximum Subsequence Sum:

```
MSS-Algorithm-3 (A:array[p..q] of integer)
sum = max = 0
for i = p to q
    sum = sum + A[i]
    if sum > max then
        max = sum
    else if sum < 0 then
        sum = 0
return max
```

Modify this algorithm to return the starting and ending indexes of the maximum subsequence instead of its sum. A partially modified algorithm is given below. Your task is to understand this algorithm and correctly complete it by filling in the blanks. Completely rewriting it to identify and return the indexes some other way is not an acceptable answer.

```
MSS-ALG-5 (A:array[p..q])

sum = max = 0

start = start-index = p

end = end-index = 0

for i = p to q do

    sum = sum + A[i]

    end = _____

    if sum > max then

        max = sum

        start-index = ________

        end-index = _________

    else if sum < 0 then

        sum = 0

        start = ________

return (start-index, end-index)       {this return statement returns both values}
```

2. Genetic information is encoded as a sequence of nucleotides (guanine, adenine, thymine, and cytosine), represented using the letters G, A, T, and C. Design a finite automaton to detect all occurrences of the nucleotide pattern GAGA in a strand of DNA represented as a string T with characters from this 4-letter alphabet. Provide its transition table. You do not need to draw the automaton.

3.. Given the matrix chain consisting of 4 matrices <A1, A2, A3, A4>, where the size of each matrix is provided in the table below:

Matrix	Size
A1	5 X 10
A2	10 X 5
A3	5 X 20
A4	20 X 10

Compute the optimal costs of the matrix-chain multiplication and show them in the matrices **m and s**. Provide the **fully parenthesized** expression for the multiplication.

4. Given the following set of symbols and their frequency of appearance in a document, design the optimal variable length binary code for these symbols so that the file size of the document can be minimized. **State the codes and the tree** constructed by the Huffman algorithm.

Λ	Γ	Π	Δ	Θ	Σ	Ω
2	5	15	3	7	1	8

5. A simple cycle in a graph is a loop that starts from one node and returns to that starting node without visiting any node more than once. The length of a simple cycle is the number of its edges.

The Hamiltonian Circuit Problem (HCP): Given an unweighted graph of n nodes determine whether it has a simple cycle of length n that visits all n nodes. This is a known NP-Complete problem.

The Longest Simple Cycle Problem (LSCP) is the problem of finding a simple cycle of maximum length in a graph, if there is at least one simple cycle of any length in the graph.

(5a) LSCP is in the class of problems NP because (a) no polynomial time algorithm to solve it is known but (b) a potential solution to it can be checked in polynomial time. Explain the strategy of a polynomial time algorithm to check a potential solution to LSCP.

(5b) We can show that LSCP is an NP-Complete problem by (1) showing that LSCP is in NP and (2) that HCP, a known NP-Complete problem, can be reduced or transformed to LSCP in polynomial time. Explain briefly but clearly why these two steps are sufficient to prove that LSCP is an NP-Complete problem.

(5c) The reduction or transformation from HCP to LSCP is trivial in that no changes to the graph are needed. That is, given an instance of the HCP problem on a graph G, the corresponding instance of the LSCP problem uses the same graph G. The last part of proving the NP-Completeness of LSCP is to show two properties:
(1) if G has a Hamiltonian Circuit, then from it we can derive a simple cycle of maximum length in G
Explain how:

(2) if G has a simple cycle of maximum length, then from it we can determine whether or not there is a Hamiltonian Circuit in G. Explain how: