FINAL EXAMNATION

TIME ALLOCATION: 3 HOURS

INSTRUCTION TO STUDENTS: ATTEMPT ALL THE QUESTIONS

Question 1: Based on natural selection, name and explain the three operators that ensure survival of a species. (6 Marks)

Question 2: What do you understand by the following terms (10 Marks)

a. Dynamic programming
b. Machine learning
c. Combinatorial pattern matching
d. Algorithm
e. Big O notation

Question 3: Discuss FIVE ways in which biologists have learned to analyze DNA (14 Marks)

Question 4: (10 Marks)

The following algorithm is a cousin of one of the motif finding algorithms we have considered in this chapter. Identify which algorithm is a cousin of ANOTHERMOTIFSEARCH and find the similarities and differences between these two algorithms.

```
ANOTHERMOTIFSEARCH(DNA, t, n, l)
s ← (1, 1, ..., 1)
bestMotif ← FINDINSEQ(s, 1, t, n, l)
return bestMotif
```

Please Turn Over

```
FINDINSEQ(s, currentSeq, t, n, l)
bestScore ← 0
for j ← 1 to n − l + 1
    s_currentSeq ← j
    if currentSeq ≠ t
        s ← FINDINSEQ(s, currentSeq + 1, t, n, l)
    if Score(s) > bestScore
        bestScore ← Score(s)
        bestMotif ← s
return bestMotif
```

Question 5: (10 Marks)

Implement an algorithm that counts the number of occurrences of each l-mer in a string of length n. Run it over a bacterial genome and construct the distribution of l-mer frequencies. Compare this distribution to that of a random string of the same length as the bacterial genome.

Question 6: (10 Marks)

Given a long *text* string T, one shorter *pattern* string s, and an integer k, find the first occurrence in T of a string (if any) s' such that $d_H(s, s') \leq k$. What is the complexity of your algorithm?

Question 7: (10 Marks)

The search trees in the text are *complete k-ary trees*: each vertex that is not a leaf has exactly k children. It is also *balanced*: the number of edges in the path from the root to any leaf is the same (this is sometimes referred to as the *height* of the tree). Find a closed-form expression for the total number of vertices in a complete and balanced k-ary tree of height L.

Question 8: (10 Marks)

Given a multiset of integers $A = \{a_i\}$, we call the polynomial $A(x) = \sum_i x^{a_i}$ the *generating function* for A. Verify that the generating function for ΔA is $\Delta A(x) = A(x)A(x^{-1})$. Given generating functions for U and V, find generating functions for $A = U \oplus V$ and $B = U \ominus V$. Compare generating functions for ΔA and ΔB. Are they the same?

Question 9: (10 Marks)

Write an algorithm that, given an n-element multiset, generates all m-element subsets of this set. For example, the set $\{1, 2, 2, 3\}$ has four two-element subsets $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$, and $\{2, 2\}$. How long will your algorithm take to run? Can it be done faster?

Question 10: (10 Marks)

You are given an 8×8 table of natural numbers. In any one step, you can either double each of the numbers in any one row, or subtract 1 from each of the numbers in any one column. Devise an algorithm that transforms the original table into a table of all zeros. What is the running time of your algorithm?