CSE 484 Midterm II Exam

Olcay Taner YILDIZ

I. Question (12 points)

a) Two players play the following game with two sequences of length n and m nucleotides. At every turn a player can either delete an arbitrary number of nucleotides from one sequence or an equal (but still arbitrary) number of nucleotides from both sequences. The player who deletes the last nucleotide wins. Who will win? Describe the winning strategy for each n and m.

b) Two players play the following game with a nucleotide sequence of length n. At every turn a player may delete either one or two nucleotides from the sequence. The player who deletes the last letter wins. Who will win? Describe the winning strategy for each n.

II. Question (18 points)

What are the optimal local alignments for MOAT and BOAST? Show all optimal alignments and the corresponding paths under the scoring matrix below and indel penalty -1.

	A	B	M	O	S	T
A	1	-1	-1	-2	-2	-3
B		1	-1	-1	-2	-2
M			2	-1	-1	-2
O				1	-1	-1
S					1	-1
T						2

III. Question (18 points)

Suppose that you are given an array A of n words sorted in lexicographic order and want to search this list for some arbitrary word, perhaps w (we write the number of characters in w as $|w|$). Design three algorithms to determine if w is in the list: one should have O($n|w|$) running time; another should have O($|w| \log n$) running time but use no space (except for A and w); and the third should have O($|w|$) running time but can use as much additional space as needed.

IV. Question (20 points)

Use a) the Hamiltonian path and b) Eulerian path approach to solve the SBH problem for the following spectrum:

$$S = \{ATG, GGG, GGT, GTA, GTG, TAT, TGG\}$$

Label edges and vertices of the graph, and give all possible sequences s such that Spectrum(s, 3) = S.

V. Question (12 points)

Find the shortest common superstring for all 9 2-digit ternary numbers in 0-2 alphabet (00, 01, 02, 10, 11, 12, 20, 21, 22).

VI. Question (20 points)

Design two algorithms for finding the longest string shared by two given texts of length n.

a) The first algorithm will have $O(n^3)$ time complexity and compares substrings of two texts one by one.

b) The second algorithm will have $O(n^2)$ time complexity and uses Suffix trees.