CISC-471 WINTER 2016

HOMEWORK 7

Please work on these problems and be prepared to share your solutions with classmates in class on Tuesday March 15. Assignments will **not** be collected for grading.

Problems

These questions come from *An Introduction to Bioinformatics Algorithms* by Neil C. Jones and Pavel A. Pevzner.

Problem 7.3: 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. (NOTE: the stated costs are expressed for answering one query. The set up cost, or the pre-processing can be ignored for the purposes of this question.)

Problem 8.6: Find the shortest common superstring for eight 3-mers:

{AGT, AAA, ACT, AAC, CTT, GTA, TTT, TAA}

and solve the following two problems:

- Construct the graph with 8 vertices corresponding to these 3-mers (Hamiltonian path approach) and find a Hamiltonian path (7 edges) which visits each vertex exactly once. Does this path visit every edge of the graph? Write the superstring corresponding to this Hamiltonian path.
- Construct the graph with 8 edges corresponding to these 3-mers (Eulerian path approach) and find an Eulerian path (8 edges) which visits each edge exactly once. Does this path visit every vertex of the graph exactly once? Write the superstring corresponding to this Eulerian path.

Problem 8.8: Find the shortest common superstring for all 8 3-digit binary numbers in 0-1 alphabet.