CS 181 Spring 2020 Homework Week 2

Assigned Tue 4/7; Due via GradeScope Mon 4/13 6:00pm

1. Recall from last week: Let alphabet $\Sigma = \{c, d, e\}$. Let

$$L_5 = \{w \in \Sigma^+ \mid w = yx, \quad \text{where } y \in \Sigma^+,\ x \in \Sigma \text{ and } w \text{ contains substring } xxxx\}.$$

   Show an NFA which recognizes this language. Specify the NFA as a fully specified state diagram. Be sure to clearly indicate your initial state and accepting state(s). Briefly describe how your design works.

2. Let $\Sigma = \{a, b\}$. Recall from lecture we defined the notation $\#(a, w)$ to be the number of occurrences of symbol a in word w. Let

$$L_1 = \{w \in \Sigma^+ \mid \#(a, w) = \#(b, w)\} \quad \text{and} \quad L_{11} = L_1 \circ L_1.$$

   Also let

$$L_2 = \{w \in \Sigma^* \mid \#(a, w) = \#(b, w)\} \quad \text{and} \quad L_{22} = L_2 \circ L_2.$$

   (a) Prove or disprove: $L_1 = L_{11}$,

   (b) Prove or disprove: $L_2 = L_{22}$.

3. Recall from lecture that a *directed rooted ordered tree* is a connected, directed graph which is acyclic, even when viewed as an undirected graph, where one node is designated as the root and all edges in the tree are directed away from the root towards the leaves. It is called an "ordered" tree because the order of the edges coming out of a node matters, but that is not important for this problem.

   The height of a directed rooted tree is defined as the length of a longest path from the root to a leaf. The height can be expressed in terms of the number of edges on a longest path or in terms of the number of nodes on a longest path. I.e., "n edges" and "$n + 1$ nodes" are two different ways of expressing the same tree height. We define a k-ary directed rooted ordered tree to be a directed rooted ordered tree in which each node has at most k children.

   Let T be a k-ary directed rooted ordered tree of height n edges ($n + 1$ nodes), where $n \geq 0$.

   (a) What is the fewest number of leaf nodes T can have in terms of n?

   (b) What is the largest number of leaf nodes T can have in terms of n? Prove your answer by induction on n.

4. Let $\Sigma = \{a, b, c\}$. Give a DFA or NFA for the following language:

$$L_4 = \{w \in \Sigma^+ \mid w \text{ contains two non-overlapping substrings consisting of the same symbol repeated 3 times consecutively}\}.$$

   E.g., *abbbcbbbba*, *cccbbbcccbbb*, *cccccc*, and *aaabaaacaaa* are in L_4, but *aaacbbb* and *abbbbc* are not. Briefly explain how your finite automaton works. Part of your score will be based on demonstrating effective use of the finite automaton model, including determinism vs. nondeterminism.