Formal Languages and Automata (CS322)

Homework 1

Lecturer: Eunjung KIM

Use proper English in your answers, and make sure to write them clearly. Any answer not written in English will receive a score of zero. Your scores may also be deducted for unclear or disorganized writing. You must also fully justify your answer. You can earn partial points for providing rough ideas, but not the full credit. If the notations in your answers differ from those introduced in the lecture or homework, your score may be deducted unless you have defined them in your solutions. Late submissions will be accepted up to 24 hours after the due date, but half of the score received will be deducted.

Remark: We denote the set of nonnegative integers by $\mathbb{N}$ and the set of positive integers by $\mathbb{Z}^+$.

Q1. Fix an alphabet Σ (i.e., a nonempty finite set of symbols). For a string $s \in \Sigma^*$, we denote its length by $|s|$. For two strings $s, t \in \Sigma^*$, we write $s \sqsubseteq t$ to mean that s is a substring of t (i.e., there exist strings $x, y \in \Sigma^*$ such that $t = xsy$). The *substring problem* has the following input and output:

- Input: two strings $s, t \in \Sigma^*$.
- Output: YES if $s \sqsubseteq t$, NO otherwise.

(Remark: In LaTeX, $\sqsubseteq$ is `\sqsubseteq`.)

(a) Let $s \in \Sigma^*$ be an arbitrary string. Define a language $L_s = \{t \in \Sigma^* : s \sqsubseteq t\}$. Construct a deterministic finite automaton (DFA) with at most $|s| + 1$ states that recognizes L_s. Justify your answer. (Remark: You cannot express the answer using a transition diagram since the answer depends on an arbitrary string s. You must describe a transition function (as well as other entries of a DFA) in a well-defined way.)

(b) Present an algorithm for the substring problem in time $O(|s| + |t|)$. Justify your answer (i.e., provide a correctness proof and running-time analysis).

Q2. Let $k \in \mathbb{Z}^+$ be fixed and let $\Sigma = \{\mathtt{a}, \mathtt{b}\}$. Define $L_k \subseteq \Sigma^*$ to be the language of all strings of length at least k whose k-th symbol from the end is $\mathtt{b}$. Equivalently, $t \in L_k$ if and only if $t = x\mathtt{b}y$ for some $x, y \in \Sigma^*$ with $|y| = k - 1$.

(a) Propose a DFA M recognizing L_k. Argue that $L(M) = L_k$.

(b) Let $P = \{\mathtt{a}, \mathtt{b}\}^k$. Let $M = (Q, \Sigma, \delta, q_0, F)$ be any DFA recognizing L_k, and let $\hat{\delta}$ be the extended transition function of M. Define the function $f \colon P \to Q$ by $f(w) = \hat{\delta}(q_0, w)$. Prove that f is injective. Conclude by arguing further that there is no DFA with fewer than 2^k states that recognizes L_k.

Q3. A set $A \subseteq \mathbb{N}$ is *linear* if $A = \{p + qn : n \in \mathbb{N}\}$ for some $p, q \in \mathbb{N}$. A subset of $\mathbb{N}$ is *quasi-linear* if it is the union of finitely many linear sets.

(a) Let $A \subseteq \mathbb{N}$ be an arbitrary set. Prove that the set

$$A^{\oplus} := \{a_1 + \cdots + a_k : k \in \mathbb{N},\ a_1, \ldots, a_k \in A\}$$

is quasi-linear.

(b) Prove that a language $L \subseteq \{\mathtt{a}\}^*$ is regular if and only if the set $\{n : \mathtt{a}^n \in L\}$ is quasi-linear.

(c) Prove that for any $X \subseteq \{\mathtt{a}\}^*$, the language X^* is regular.

Q4. Consider One-Dimensional Minesweeper (ODM for short) played on a board of shape $1 \times n$ where $n \in \mathbb{N}$. Let $\Sigma = \{\mathtt{0}, \mathtt{1}, \mathtt{2}, \mathtt{\#}\}$. For $w \in \Sigma^*$, interpret $w = a_1 a_2 \ldots a_n$ as the configuration of the board, where the symbol a_i represents the i-th cell from the left on the board. The i-th cell and j-th cell are considered adjacent if and only if $|i - j| = 1$. A string w encodes a valid configuration if, for all i with $1 \leq i \leq n$, the following conditions hold:

- If $a_i = \mathtt{\#}$, then the ith cell contains a mine.
- If $a_i \in \{\mathtt{0}, \mathtt{1}, \mathtt{2}\}$, then the ith cell does not contain a mine, and the value of a_i equals the number of mines in its adjacent cells.

Then, consider Cyclic One-Dimensional Minesweeper (CODM for short). The rules are the same as ODM, except that the first cell which represented by a_1 and the last cell which represented by a_n are also considered as adjacent. Notice that '1#' is a valid configuration and '2#' is invalid configuration of CODM when $n = 2$.
Give a DFA M recognizing the language $\{w \in \Sigma^* : w \text{ encodes a valid configuration of CODM}\}$.

Hint: Start with find a DFA recognizing the language $\{w \in \Sigma^* : w \text{ encodes a valid configuration of ODM}\}$.

Q5. Construct a DFA or an NFA recognizing the following languages.

(a) Let $\Sigma_1 = \{0, 1\}^3$ be an alphabet, where each symbol is written as columns. Construct an NFA M_1 recognizing the following language L_1:

$$L_1 = \left\{ \begin{pmatrix} a_1 \\ b_1 \\ c_1 \end{pmatrix} \begin{pmatrix} a_2 \\ b_2 \\ c_2 \end{pmatrix} \cdots \begin{pmatrix} a_n \\ b_n \\ c_n \end{pmatrix} : [a_1 a_2 \ldots a_n]_2 + [b_1 b_2 \ldots b_n]_2 = [c_1 c_2 \ldots c_n]_2 \right\}.$$

Here, $[w]_2$ denotes the numerical value of the binary sequence w, e.g., $[101]_2 = 5$. Along with an NFA, provide its transition diagram recognizing L_1.

(b) Let $\Sigma_2 = \{0, 1, \ldots b-1\}$ for $b \geq 2$. Construct a DFA M_2 recognizing the base b representations of numbers divisible by $k \in \mathbb{Z}^+$.

(c) Let $\Sigma_3 = \{0, 1, \ldots b-1\}$ for $b \geq 2$. Construct a DFA M_3 recognizing the reverse of the base b representations of numbers divisible by $k \in \mathbb{Z}^+$.

For (b) and (c), you may assume that the empty string ϵ is 0, and the leading 0's of a base-b-representation are allowed and does not affect its value, i.e., 0001 is equal to 1. Furthermore, you may provide transition functions for (b) and (c) if they are succinct and easy to understand.

Q6. A *monoid* is a pair $(M, \circ)$ where M is a set and $\circ : M \times M \to M$ is a binary operation on M that satisfies the following axioms:

(M1) The operation is associative, that is, $a \circ (b \circ c) = (a \circ b) \circ c$ for every $a, b, c \in M$; and

(M2) there is an element $1_M \in M$ such that $1_M \circ a = a \circ 1_M = a$ for every $a \in M$.

Given two monoids $(M, \circ)$ and $(N, \star)$, a function $h : (M, \circ) \to (N, \star)$ is a *monoid homomorphism* if $h(a \circ b) = h(a) \star h(b)$ for every $a, b \in M$ and $h(1_M) = 1_N$. Observe that for any set of alphabets Σ, the language Σ^* together with the concatenation operation forms a monoid. In this vein, we always assume that Σ^* is equipped with the concatenation operation.

Let Σ be an alphabet. Prove that $L \subseteq \Sigma^*$ is regular if and only if there is a finite monoid $(M, \circ)$ (i.e., a monoid where M is a finite set) and a monoid homomorphism $h : \Sigma^* \to (M, \circ)$ satisfying the following.

There exists $F \subseteq M$ such that for every $w \in \Sigma^*$, $w \in L$ if and only if $h(w) \in F$.

♣ Submit your solution via KLMS-gradescope by 15 September at 23:59:59.