Homework 2: due 17:00, 17 October 2016

For questions 1–3, we assume that $\Sigma = \{a, b\}$.

(1) Construct the NFA for each of the following languages.

(a) The language L that consists of all the words in which a appears exactly once.
(b) The language L that consists of all the words that starts with a and ends with a.
(c) The language L that consists of all the words that contains aba.
For example: aba and $aaaabaaaaa$ are in L, since they contain aba. On the contrary, $aaaaa$ and $bbbbabbaaaaa$ are not in L, since they do not contain aba.
(d) The language L that consists of all the words that do *not* contain aa.
(e) The language L that consists of all the words w such that if w contains aa, then w ends with bb.

(2) Construct the regular expression for each of the languages above.

(3) Determine and verify which of the following languages are regular. That is, supply its regular expression or NFA, if you claim that a language is regular, and conversely, supply the proof, if you claim that a language is not regular.

(a) L consists of all the words in which a appears exactly 3 times.
(b) L consists of all the words in which a appears even number of times.
For example: $bbbbb$ is in L, since a appears zero times, and zero is even. ab is not in L, since a appears once, and one is odd.
(c) L consists of all the words of even length.
For example: ϵ and $ababaa$ are in L, since their lengths are 0 and 6, respectively, but a and abb are not, since their lengths are 1 and 3, respectively.
(d) L consists of all the words of the form $a^m b a^n$, where m and n are positive integers such that $m \leqslant n$.
For example: aba is in L, and so is $aaabaaaaa$. However, $aaaaba$ is not in L since a appears 4 times before b, but appears only once after b, and neither do $aaabaa$ and $aaba$. Likewise, $abbaaa$ and $abaabaaa$ are not in L, since b appears more than once.
(e) L consists of all the words in which the number of occurrences of a is a prime number.

(4) For a language $L \subseteq \Sigma^*$ (not necessarily regular), we define the equivalence relation $\sim_L$ on Σ^* as follows. $u \sim_L v$, if the following holds: For every $w \in \Sigma^*$, $uw \in L$ if and only if $vw \in L$.

(a) Prove that $\sim_L$ is an equivalence relation.
(b) Prove that if $u \sim_L v$, then either both $u, v \in L$ or both $u, v \notin L$.
(c) Suppose L is regular, and $\mathcal{A}$ is its DFA, i.e., $L(\mathcal{A}) = L$. For a word w, we denote by $\mathcal{A}(w)$ the state of $\mathcal{A}$ after reading w. Or, more formally, if $w = a_1 \cdots a_n$ and $q_0 a_1 q_1 \cdots a_n q_n$ is the run of $\mathcal{A}$ on w, then $\mathcal{A}(w) = q_n$.
Prove that if u and v are words such that $\mathcal{A}(u) = \mathcal{A}(v)$, then $u \sim_L v$.

In the following, let $\#(\sim_L)$ (read: the index of $\sim_L$) denote the number of equivalence classes of $\sim_L$. Recall that since $\sim_L$ is an equivalence relation on Σ^*, by Theorem 1.2, $\sim_L$ partitions Σ^*. Also, note that it is possible that $\sim_L$ has *infinitely* many equivalence classes.

(d) Following (c), prove that if L is regular with $\mathcal{A}$ being its DFA, then $\sim_L$ has finitely many equivalence classes and $\#(\sim_L) \leqslant |Q|$, where Q is the set of states of $\mathcal{A}$.

(5) **(bonus point)** Let L be a language over Σ, where $\sim_L$ has finitely many equivalence classes $C_1, \ldots, C_m$. Using the notation in Lecture 1, we can represent each C_i as $[w]$, for every $w \in C_i$.

In the following, we assume that $L \neq \emptyset$.

(a) Prove that there is $i_1, \ldots, i_k \subseteq \{1, \ldots, m\}$ such that $L = C_{i_1} \cup \cdots \cup C_{i_k}$. (You can use (4.b) here.)

(b) Consider the following DFA $\mathcal{A} = \langle \Sigma, Q, q_0, F, \delta \rangle$.

- $Q = \{p_1, \ldots, p_m\}$, i.e., the number of states is precisely the number of equivalence classes in $\sim_L$.
- q_0 is p_j, where j is such that $\epsilon \in C_j$.
- $F = \{p_{i_1}, \ldots, p_{i_k}\}$, where $i_1, \ldots, i_k$ are the indices in (5.a).
- $\delta : Q \times \Sigma \to Q$ is defined as follows. For every $p_i \in Q$, for every $a \in \Sigma$, we pick an arbitrary $w \in C_i$, and define $\delta(p_i, a) = p_j$, where $[wa] = C_j$.

Prove that δ is a well-defined function, i.e., for every $w_1, w_2 \in C_i$, $[w_1 a] = [w_2 a]$. In other words, the end result p_j remains the same for whichever w we pick, as long as w is from C_i.

(c) Let $\mathcal{A}$ be as in (b). Prove that $L(\mathcal{A}) = L$.

(d) **[Myhill-Nerode theorem]** Prove that a language L is regular if and only if $\sim_L$ has finitely many equivalence classes.