Formal Language | Homework 1 Finite State Automata and Regular Expression | 2024 Spring

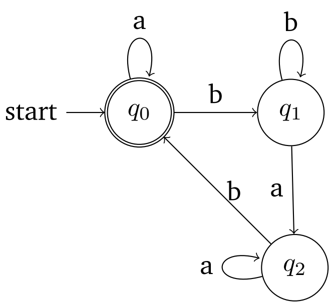
Homework 1
Finite State Automata and Regular Expression

Homework rules

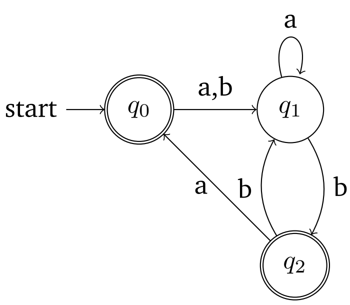
- Group discussion is *highly recommended*. But you have to write down your solution by *yourself*. Aim for clarify and conciseness in your solution, emphasizing the main ideas over low-level details.
- You should *cite* all the references that help you to solve the homework, including the people you have discussed with and the online material you looked at.
- You are required to type your solution using latex and submit a *pdf* and a *tex* file by email to csie.tamc@gmail.com. You may type your solution in English (preferable) or Chinese.
- You are not expected to solve *all* problems. Just try your best! If you don't solve a problem, try your best to write down your thinking process.
- There are total 10 problems, each weight 10 pts and 1 bonus problem which weights additional 20 pts. Total: $100 + 20$(bonus) points.

1. (Fun with `DFA`) For each of the following languages, construct a `DFA` that accepts the language. In all cases, the alphabet is $\{0, 1\}$.
   (a) $\{w \mid |w| \equiv 0 \mod 3\}$.
   (b) $\{w \mid 110$ is not a substring of $w\}$.
   (c) $\{w \mid$ every odd position in w is $1\}$.
   (d) $\{w \mid w$ contains at least two 1s and at most two 0s$\}$.
   (e) $\{\epsilon, 0\}$.

2. Convert the `DFAs` to a `regex`:



(a)



(b)

3. (Fun with NFA) For each of the following languages, construct an NFA, with the specified number of states, that accepts the language. In all cases, the alphabet is $\{0, 1\}$.

   (a) $\{w \mid w \text{ ends with } 10\}$ with 3 states.

   (b) $\{w \mid w \text{ contains the substring } 1011\}$ with 5 states.

   (c) $\{w \mid w \text{ contains an odd number of 1s or exactly two 0s}\}$.

   (d) $\{w \mid w \text{ begins with 1 or ends with } 0\}$ with whatever number of states you like.

   (e) $\{11, 111\}^*$ with whatever number of states you like.

4. (Fun with regex) Give regular expressions describing the following languages. In all cases, the alphabet is $\{0, 1\}$.

   (a) $\{w \mid w \text{ contains with at least three 1s}\}$.

   (b) $\{w \mid w \text{ contains exactly two 0s and at least two 1s}\}$.

   (c) $\{w \mid \text{ every odd position in } w \text{ is } 1\}$.

5. If $n \in \mathbb{N}$ and $w = a_1 \cdots a_n$ is a string, for each $i \in [n-1] \cup \{0\}$, the string $a_1 \cdots a_i$ is called a *proper prefix* of w. For any language L, we define

$$\texttt{MIN}(L) := \{w \in L \mid \text{ no proper prefix of } w \text{ belongs to } L\}.$$

   Prove that if L is regular, then $\texttt{MIN}(L)$ is regular as well.

6. (a) Given an alphabet Σ, for any language $L \subseteq \Sigma^*$, prove that $L^{**} = L^*$ and $L^*L^* = L^*$.

   (b) Prove that every finite language is regular.

   (c) Given an example of a non-regular language A and a regular language B such that $A \subseteq B$.

   (d) Given an example of a non-regular language A and a regular language B such that $B \subseteq A$.

   Let L ba nay regular language over some Σ. Define the languages:

$$L^\infty := \bigcup_{k \geq 1} \{w^k \mid w \in L\},$$
$$L^{1/\infty} := \{w \mid w^k \in L \text{ for all } k \geq 1\},$$
$$\sqrt{L} := \{w^k \in L \text{ for some } k \geq 1\}.$$

   And also for any $k \in \mathbb{N}$, let

$$L^{(k)} := \{w^k \mid w \in L\},$$
$$L^{(1/k)} := \{w \mid w^k \in L\}.$$

   (e) Prove that $L^{(1/3)}$ is regular.

   (f) What about $L^{(3)}$?

   (g) Let $k \in \mathbb{N}$. Prove that there are only finitely many languages of the form $L^{(1/k)}$ and that they are all regular.

   (h) Is $L^{1/\infty}$ regular or not?

   (i) Is $\sqrt{L}$ regular of not?

   (j) What about L^∞?

7. Which of the following languages are regular? Justify each answer.

   (a) $L = \{wcw \mid w \in \{a,b\}^*\}$.

   (b) $L = \{xy \mid x, y \in \{a,b\}^* \text{ and } |x| = |y|\}$.

   (c) $L = \{a^n \mid n \text{ is a prime number}\}$.

   (d) $L = \{a^m b^n \mid \gcd(m,n) = 17\}$.

8. Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA. A state $q \in Q$ is *reachable* iff there is some string $w \in \Sigma^*$ such that $\hat{\delta}(q_0, w) = q$. Consider the following method for computing the set $Q_r \subseteq Q$ of reachable states: define the sequence of sets $Q_r^i \subseteq Q$ where

$$Q_r^0 := \{q_0\}$$
$$Q_r^{i+1} := \{q \in Q \mid \exists p \in Q_r^i, \exists a \in \Sigma,\ q = \delta(p,a)\}.$$

   (a) Prove by induction on i that Q_r^i is the set of all reachable states from q_0 using paths of length i.

   (b) Give an example of a DFA such that $Q_r^{i+1} \neq Q_r^i$ for all $i \geq 0$.

   (c) Change the inductive definition of Q_r^i as follows:

$$Q_r^{i+1} := Q_r^i \cup \{q \in Q \mid \exists p \in Q_r^i, \exists a \in \Sigma,\ q = \delta(p,a)\}.$$

   Prove that there exists an i_0 such that $Q_r^{i_0+1} = Q_r^{i_0} = Q_r$.

   Define the DFA M_r as follows: $M_r = (Q_r, \Sigma, \delta_r, q_0, F \cap Q_r)$, where $\delta_r : Q_r \times \Sigma \to Q_r$ is the restriction of δ to Q_r.

   (d) Explain why M_r is indeed a DFA.

   (e) Prove that $L(M_r) = L(M)$. A DFA is called *reachable* or *trim* if $M = M_r$.

9. Let $\Sigma = \{a_1, \ldots, a_n\}$ be an alphabet of n symbols.

   (a) Construct an NFA with $2n+1$ states accepting the set L_n of strings over Σ such that, every string in L_n has an odd number of a_i, for some $a_i \in \Sigma$. Equivalently, if L_n^i is the set of strings over Σ with an odd number of a_i, then $L_n = L_n^1 \cup \cdots \cup L_n^n$.

   (b) Prove that there is a DFA with 2^n states accepting the language L_n.

   (c) Prove that every DFA accepting L_n has at least 2^n states.

10. Given two DFAs $M_i = (Q_i, \Sigma, \delta_i, q_{0i}, F_i)$, $i = 1, 2$, a *morphism* $h : M_1 \to M_2$ of DFAs is a function $h : Q_1 \to Q_2$ satisfying the following:

    - $h(\delta_1(p,a)) = \delta_2(h(p), a)$, for all $p \in Q_1$ and all $a \in \Sigma$.
    - $h(q_{01}) = q_{02}$.

    An F-map $h : M_1 \to M_2$ is a morphism h satisfying $h(F_1) \subseteq F_2$. A B-map $h : M_1 \to M_2$ is a morphism h satisfying $h^{-1}(F_2) \subseteq F_1$. A *proper homomorphism* of DFAs is an F-map and also a B-map, i.e. $h^{-1}(F_2) = F_1$.

    (a) Prove that if $f : M_1 \to M_2$ and $g : M_2 \to M_3$ are morphisms (resp. F-maps, resp. B-maps) of DFAs, then $g \circ f : M_1 \to M_3$ is also a morphism (resp. F-map, resp. B-map).

(b) If $h : M_1 \to M_2$ is a morphism, prove that

$$h(\hat{\delta}_1(p, w) = \hat{\delta}_2(h(p), w)$$

for all $p \in Q_1$ and all $w \in \Sigma^*$.

(c) Prove that if $h : M_1 \to M_2$ is a proper homomorphism, then $L(M_1) = L(M_2)$.

11. (Bonus) (Morphisms between NFAs) In this problem we assume that we are considering NFAs without ϵ-transitions.

Given two NFAs $N_i = (Q_i, \Sigma, \delta_i, q_{0i}, F_i)$, $i = 1, 2$, we say that a relation $\varphi \subseteq Q_1 \times Q_2$ is a *simulation* of N_1 by N_2, denoted by $\varphi : N_1 \to N_2$, if the following properties hold:

- $(q_{01}, q_{02}) \in \varphi$.
- Whenever $(p, q) \in \varphi$, for every $r \in \delta_1(p, a)$, there is some $s \in \delta_2(q, a)$ so that $(r, s) \in \varphi$, for all $a \in \Sigma$.
- Whenever $(p, q) \in \varphi$, if $p \in F_1$ then $q \in F_2$.

(a) If N_1 and N_2 are actually DFAs, show that an F-map $\varphi : N_1 \to N_2$ of DFAs is a simulation of N_1 by N_2.

(b) Let $\varphi : N_1 \to N_2$ be a simulation of N_1 by N_2. Prove that for every $w \in \Sigma^*$, for every $q_1 \in \hat{\delta}_q(q_{01}, w)$, there is some $q_2 \in \hat{\delta}_2(q_{02}, w)$ so that $(q_1, q_2) \in \varphi$.

(c) Conclude that $L(N_1) \subseteq L(N_2)$.

(d) If N_1 is an NFA and N_2 is a DFA, prove that if $L(N_1) \subseteq L(N_2)$, then there is some simulation $\varphi : N_1 \to N_2$ of N_1 by N_2. *Hint*. Consider the relation $\varphi = \{(q_1, q_2) \mid q_1 \in \hat{\delta}_1(q_{01}, w),\ q_2 = \hat{\delta}_2(q_{02}, w),\ w \in \Sigma^*\}$.

Remark. If N_1 and N_2 are DFAs and $L(N_1) \subseteq L(N_2)$, then there may not exist any DFA map from N_1 to N_2, but above shows that there is always a simulation of N_1 by N_2.

(e) Give a counter-example showing that (c) is generally *false* for NFAs, i.e., if N_1 and N_2 are both NFAs and $L(N_1) \subseteq L(N_2)$, there may not be any simulation $\varphi : N_1 \to N_2$.

In order to salvage (c), we modify the conditions of the definition of a simulation: we say that $\varphi : N_1 \to N_2$ is a *generalized simulation* (or g-simulation) if

- $(q_{01}, q_{02}) \in \varphi$
- Whenever $(p, q) \in \varphi$, for all $a \in \Sigma$, if $\delta_1(p, a) \neq \emptyset$ and $\delta_2(q, a) \neq \emptyset$, then for every $r \in \delta_1(p, a)$, there is some $s \in \delta_2(q, a)$ so that $(r, s) \in \varphi$.
- For all $w \in \Sigma^*$ with $|w| < n_1 2^{n_2}$, for every $q_1 \in \hat{\delta}_1(q_{01}, we) \cap F_1$, there is some $q_2 \in \hat{\delta}_2(q_{02}, w) \cap F_2$ so that $(q_1, q_2) \in \varphi$.

(f) Prove that $L(N_1) \subseteq L(N_2)$ iff there is some g-simulation $\varphi : N_1 \to N_2$.

(g) We say that $\varphi : N_1 \to N_2$ is a *g-bisimulation between N_1 and N_2* if φ is a g-simulation between N_1 and N_2 and φ^{-1} is a g-simulation between N_2 and N_1 (recall that $\varphi^{-1} = \{(q, p) \in Q_2 \times Q_1 \mid (p, q) \in \varphi\}$).

(h) Prove that $L(N_1) = L(N_2)$ iff there is some g-bisimulation between N_1 and N_2.