

Introduction to the Theory of Computation

Homework 2

February 4, 2003; Due February 20, beginning of class

“A problems” are for practice only, and should not be turned in.

Problem A1. Recall that two regular expressions R and S are equivalent, denoted by $R \cong S$, iff they denote the same regular language $\mathcal{L}[R] = \mathcal{L}[S]$. Show that the following identities hold for regular expressions:

$$\begin{aligned}R^{**} &\cong R^* \\(R + S)^* &\cong (R^* + S^*)^* \\(R + S)^* &\cong (R^*S^*)^* \\(R + S)^* &\cong (R^*S)^*R^*\end{aligned}$$

Problem A2. Recall that a homomorphism $h: \Sigma^* \rightarrow \Delta^*$ is a function such that $h(uv) = h(u)h(v)$ for all $u, v \in \Sigma^*$. Given any language $L \subseteq \Sigma^*$, we define $h(L)$ as

$$h(L) = \{h(w) \mid w \in L\}.$$

Given any language $L' \subseteq \Delta^*$, we define $h^{-1}(L')$ as

$$h^{-1}(L') = \{w \in \Sigma^* \mid h(w) \in L'\}.$$

Prove that if $L \subseteq \Sigma^*$ is regular, then so is $h(L)$.

Problem A3. Construct an NFA accepting the language $L = \{aa, aaa\}^*$. Apply the subset construction to get a DFA accepting L .

“B problems” must be turned in.

Problem B1 (40 pts). (a) Prove again that the intersection, $L_1 \cap L_2$, of two regular languages, L_1 and L_2 , is regular, **using the Myhill-Nerode characterization** of regular languages.

(b) Let $h: \Sigma^* \rightarrow \Delta^*$ be a homomorphism, as in A2. For any regular language, $L' \subseteq \Delta^*$, prove that $h^{-1}(L')$ is regular, **using the Myhill-Nerode characterization** of regular languages. Prove that the number of states of any minimal DFA for $h^{-1}(L')$ is at most the number of states of any minimal DFA for L' . Can it be strictly smaller?

Problem B2 (120 pts). The purpose of this problem is to investigate the notion of mapping between NFA's. It is assumed that all DFA's and NFA's considered in this problem are defined over some fixed alphabet Σ . For simplicity, we also assume that we are considering NFA's *without* ϵ -transitions.

Given two NFA's $N_1 = (Q_1, \Sigma, \delta_1, q_{01}, F_1)$ and $N_2 = (Q_2, \Sigma, \delta_2, q_{02}, F_2)$, we say that a relation $\varphi \subseteq Q_1 \times Q_2$ is a *simulation of N_1 by N_2* , denoted by $\varphi: N_1 \rightarrow N_2$, if the following properties hold:

- (1) $(q_{01}, q_{02}) \in \varphi$.
- (2) 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 $a \in \Sigma$.
- (3) Whenever $(p, q) \in \varphi$, if $p \in F_1$ then $q \in F_2$.

(i) If N_1 and N_2 are actually DFA's, show that a map $\varphi: N_1 \rightarrow N_2$ of DFA's is a simulation of N_1 by N_2 (viewing the function φ as a relation, in the obvious way).

(ii) Let $\varphi: N_1 \rightarrow N_2$ be a simulation of N_1 by N_2 . Prove that for every $w \in \Sigma^*$, for every $q_1 \in \delta_1^*(q_{01}, w)$, there is some $q_2 \in \delta_2^*(q_{02}, w)$, so that

$$(q_1, q_2) \in \varphi.$$

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

(iii) If N_1 is an NFA and D_2 is a DFA, prove that if $L(N_1) \subseteq L(D_2)$, then there is some simulation $\varphi: N_1 \rightarrow D_2$ of N_1 by D_2 .

Hint. Consider the relation $\varphi = \{(q_1, q_2) \mid q_1 \in \delta_1^*(q_{01}, w), q_2 = \delta_2^*(q_{02}, w), w \in \Sigma^*\}$.

Remark: If D_1 and D_2 are DFA's and $L(D_1) \subseteq L(D_2)$, then there may not exist any DFA map from D_1 to D_2 , but the above shows that there is always a simulation of D_1 by D_2 .

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

In order to salvage (iii), we modify conditions (2) and (3) of the definition of a simulation $\varphi: N_1 \rightarrow N_2$. Let N_1, N_2 be NFA's, and let n_1 be the number of states of N_1 and n_2 the number of states of N_2 . Then, we say that $\varphi: N_1 \rightarrow N_2$ is a *generalized simulation*, for short, a *g-simulation*, if

- (1) $(q_{01}, q_{02}) \in \varphi$.
- (2b) Whenever $(p, q) \in \varphi$, for every $r \in \delta_1(p, a)$, if $\delta_2(q, a) \neq \emptyset$, then there is some $s \in \delta_2(q, a)$ so that $(r, s) \in \varphi$, for all $a \in \Sigma$.
- (3b) For all $w \in \Sigma^*$ with $|w| < n_1 2^{n_2}$, for every $q_1 \in \delta_1^*(q_{01}, w) \cap F_1$, there is some $q_2 \in \delta_2^*(q_{02}, w) \cap F_2$ so that $(q_1, q_2) \in \varphi$.

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

Remark: Condition (3b) is very strong, since by itself, it implies that $L(N_1) \subseteq L(N_2)$. Thus, this “quick fix” is not very satisfactory. A more natural condition (if any), remains to be found!

(v) We say that $\varphi: N_1 \rightarrow N_2$ 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 $\varphi^{-1} = \{(q, p) \in Q_2 \times Q_1 \mid (p, q) \in \varphi\}$).

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

(vi) We say that an NFA N is *trim* if for every state q , there is some $w \in \Sigma^*$ so that $q \in \delta^*(q_0, w)$. Let N be a trim NFA and D a DFA. Give a counter-example to fact that if a simulation $\varphi: N \rightarrow D$ exists, then it is unique.

To fix the above problem we define *reduced simulations*. We say that a simulation $\varphi: N_1 \rightarrow N_2$ is *reduced*, for short, a r -simulation, if for all $(q_1, q_2) \in \varphi$, there is some $w \in \Sigma^*$ with $|w| < n_1 n_2$, so that $q_1 \in \delta_1^*(q_{01}, w)$ and $q_2 \in \delta_2^*(q_{02}, w)$ (n_1 and n_2 are the number of states of N_1 and N_2).

(vii) Let $\varphi: N_1 \rightarrow N_2$ and $\psi: N_2 \rightarrow N_3$ be two simulations. Prove that $\varphi \circ \psi: N_1 \rightarrow N_3$ is also a simulation (where $\varphi \circ \psi$ is the composition of the *relations* φ and ψ). Prove that this is **not true** if φ, ψ are r -simulations.



In the rest of this problem, we will be dealing with r -simulations. Also, \circ denotes composition of *relations*. This means that in $\varphi \circ \psi$, the relation φ is applied *before* the relation ψ . This is the *opposite* of the conventional notation for the composition $\psi \circ \varphi$ of *functions*, where the function φ is applied before the function ψ .

Say that an r -simulation $\varphi: N_1 \rightarrow N_2$ is an *isomorphism between N_1 and N_2* if there is a r -simulation $\psi: N_2 \rightarrow N_1$ such that $\varphi \circ \psi = \text{id}_{N_1}$ and $\psi \circ \varphi = \text{id}_{N_2}$. What can you conclude if there is an isomorphism $\varphi: N_1 \rightarrow N_2$? Does this imply that N_1 and N_2 have the same number of states?

(viii) Given an NFA N (without ϵ -transitions), let $\mathcal{D}(N)$ be the trim DFA obtained by applying to N the subset construction given in class (slides, page 45). Observe that the states of $\mathcal{D}(N)$ are the subsets of the form $\delta^*(q_0, w)$, for all $w \in \Sigma^*$. Prove that there is a r -simulation $\eta_N: N \rightarrow \mathcal{D}(N)$. For every DFA D , for every r -simulation $\varphi: N \rightarrow D$, prove that there is a unique r -simulation $\varphi^\#: \mathcal{D}(N) \rightarrow D$ such that $\varphi = \eta_N \circ \varphi^\#$.

Remarks:

1. Unfortunately, if $\varphi: N_1 \rightarrow N_2$ is an r -simulation,

$$\varphi \circ \eta_{N_2}$$

is **not** necessarily an r -simulation!

2. Simulations and bisimulations play an important role in models of concurrency and some data base models.

Open Problem. Find a reasonable notion of r -simulation between NFA's and DFA's, so that the composition of r -simulations is an r -simulation, and the beginning of (viii) holds. Then, every r -simulation $\varphi: N_1 \rightarrow N_2$ yields an r -simulation $\mathcal{D}(\varphi): \mathcal{D}(N_1) \rightarrow \mathcal{D}(N_2)$ defined by

$$\mathcal{D}(\varphi) = (\varphi \circ \eta_{N_2})^\sharp.$$

If this can be done, let \mathcal{DFA} be the set of trim DFA's (over Σ) and let the maps between DFA's be r -simulations. Similarly, let \mathcal{NFA} be the set of (trim) NFA's (over Σ) and let the maps between NFA's be r -simulations. Then, there are maps $\mathcal{D}: \mathcal{NFA} \rightarrow \mathcal{DFA}$ and $\mathcal{N}: \mathcal{DFA} \rightarrow \mathcal{NFA}$, where $\mathcal{N}(D)$ is the DFA D viewed as an NFA, and $\mathcal{D}(N)$ is the DFA associated with the NFA N . A r -simulation $\varphi: D_1 \rightarrow D_2$ of DFA's is mapped to the same r -simulation $\mathcal{N}(\varphi): \mathcal{N}(D_1) \rightarrow \mathcal{N}(D_2)$ viewed as a r -simulation of NFA's, and a r -simulation $\varphi: N_1 \rightarrow N_2$ of NFA's is mapped to the r -simulation $\mathcal{D}(\varphi): \mathcal{D}(N_1) \rightarrow \mathcal{D}(N_2)$. Then, \mathcal{DFA} and \mathcal{NFA} would be categories and \mathcal{D} and \mathcal{N} would be adjoint functors. Indeed, there would be natural bijections

$$\theta_{N,D}: \text{Hom}_{\mathcal{DFA}}(\mathcal{D}(N), D) \rightarrow \text{Hom}_{\mathcal{NFA}}(N, \mathcal{N}(D)),$$

for all $D \in \mathcal{DFA}$ and all $N \in \mathcal{NFA}$.

Problem B3 (60 pts). let Σ be an alphabet. For any language L and any string $x \in \Sigma^*$, the *left derivative of L w.r.t. x* , denoted by $x \setminus L$, or $D_x L$, or $\frac{dL}{dx}$, is the language

$$D_x L = \{y \in \Sigma^* \mid xy \in L\}.$$

(1) Prove the following identities for all languages L, A, B over Σ :

$$\begin{aligned} D_{xy} L &= D_y(D_x L), \\ D_\epsilon L &= L, \\ D_x(A \cup B) &= D_x A \cup D_x B, \end{aligned}$$

and for every symbol $a \in \Sigma$,

$$\begin{aligned} D_a(AB) &= (D_a A)B \cup (A \cap \{\epsilon\})D_a B, \\ D_a(L^*) &= (D_a L)L^*. \end{aligned}$$

Given a regular expression R and a string $x \in \Sigma^*$, we define the (left) derivative $D_x R$ of R w.r.t. x so that

$$\mathcal{L}[D_x R] = D_x \mathcal{L}[R].$$

We let

$$D_\epsilon R = R \quad \text{and} \quad D_{xa} R = D_a(D_x R)$$

where $a \in \Sigma$ and $x \in \Sigma^*$,

$$D_a \emptyset = \emptyset, \quad D_a \epsilon = \emptyset, \quad D_a a = \epsilon, \quad D_a b = \emptyset,$$

for all $a, b \in \Sigma$, $a \neq b$,

$$\begin{aligned} D_a((R + S)) &= (D_a R + D_a S), \\ D_a(R^*) &= (D_a R R^*), \\ D_a(RS) &= \begin{cases} (D_a R S) & \text{if } \epsilon \notin \mathcal{L}[R], \\ ((D_a R S) + D_a S) & \text{if } \epsilon \in \mathcal{L}[R], \end{cases} \end{aligned}$$

where R, S are any regular expressions.

(2) Give a simple algorithm to decide whether $\epsilon \in \mathcal{L}[R]$, where R is any given regular expression.

Prove that every regular expression has finitely many distinct derivatives (by distinct derivatives, we mean inequivalent derivatives).

Hint. Use an induction on the number of occurrences of the symbols from $\Sigma \cup \{\epsilon, \emptyset, +, \cdot, *, \}$. When $R = (S \cdot T)$, prove that $D_x R$ is equivalent to an expression of the form

$$(D_x S T + D_{v_1} T + \cdots + D_{v_k} T),$$

where for every i , $1 \leq i \leq k$, there is some $u_i \in \mathcal{L}[S]$ such that $x = u_i v_i$. When $R = S^*$, prove that $D_x R$ is equivalent to an expression of the form

$$(D_{v_1} S + \cdots + D_{v_k} S) S^*,$$

where for every i , $1 \leq i \leq k$, there is some $u_i \in \mathcal{L}[S^*]$ such that $x = u_i v_i$.

(3) Assuming that R has n distinct derivatives, prove that every derivative of R belongs to the finite set

$$\{D_x R \mid x \in \Sigma^*, 0 \leq |x| < n\}.$$

Show that the upper bound on the number of derivatives is a product of towers of exponentials (in terms of the length of R).

(4) Prove that if D is a DFA accepting $\mathcal{L}[R]$ and D has n states, then R has at most n distinct derivatives.

If $\nu(R)$ is the number of occurrences in R of the symbols from $\Sigma \cup \{\epsilon, \emptyset, +, \cdot, *, \}$, prove that R has at most

$$2^{2\nu(R)} \leq 4^{|R|}$$

distinct derivatives (where $|R|$ denotes the length of R).

(5) If L is any regular language over Σ^* , prove that the number of states of every minimal DFA for L is equal to the number of distinct derivatives, $D_u(L)$, of L .

(6) Prove that the regular expression

$$R = (a + b)^* a \underbrace{(a + b) \cdots (a + b)}_n$$

has $\nu(R) = 3n + 5$ (if we do not count \cdot , otherwise, $\nu(R) = 4n + 6$) and that R has 2^{n+1} distinct derivatives.

Prove that there is a 2-state DFA accepting the language denoted by $(a + b)^* a$ and there is an $(n + 2)$ -state DFA accepting the language denoted by $\underbrace{(a + b) \cdots (a + b)}_n$.

Yet, prove that any minimal DFA for the language denoted by R above has 2^{n+1} states.

Problem B4 (40 pts). Let $D = (Q, \Sigma, \delta, q_0, F)$ be a deterministic finite automaton. Define the relations \approx and \sim on Σ^* as follows:

$$\begin{aligned} x \approx y & \text{ if and only if, for all } p \in Q, \\ & \delta^*(p, x) \in F \text{ iff } \delta^*(p, y) \in F, \end{aligned}$$

and

$$x \sim y \text{ if and only if, for all } p \in Q, \delta^*(p, x) = \delta^*(p, y).$$

(a) Show that \approx is a left-invariant equivalence relation and that \sim is an equivalence relation that is both left and right invariant. (A relation R on Σ^* is *left invariant* iff uRv implies that $wuRvw$ for all $w \in \Sigma^*$, and R is *right invariant* iff uRv implies that $uwRvw$ for all $w \in \Sigma^*$.)

(b) Let n be the number of states in Q (the set of states of D). Show that \approx has at most 2^n equivalence classes and that \sim has at most n^n equivalence classes.

(c) Given any language $L \subseteq \Sigma^*$, define the relations λ_L and μ_L on Σ^* as follows:

$$u \lambda_L v \text{ iff, for all } z \in \Sigma^*, zu \in L \text{ iff } zv \in L,$$

and

$$u \mu_L v \text{ iff, for all } x, y \in \Sigma^*, xuy \in L \text{ iff } xvy \in L.$$

Prove that λ_L is left-invariant, and that μ_L is left and right-invariant. Prove that if L is regular, then both λ_L and μ_L have a finite number of equivalence classes.

Hint: Show that the number of classes of λ_L is at most the number of classes of \approx , and that the number of classes of μ_L is at most the number of classes of \sim .

Problem B5 (60 pts). Let L be any regular language over some alphabet Σ . Define the languages

$$\begin{aligned} L^\infty &= \bigcup_{k \geq 1} \{w^k \mid w \in L\}, \\ L^{1/\infty} &= \{w \mid w^k \in L, \text{ for all } k \geq 1\}, \text{ and} \\ \sqrt{L} &= \{w \mid w^k \in L, \text{ for some } k \geq 1\}. \end{aligned}$$

Also, for any natural number $k \geq 1$, let

$$L^{(k)} = \{w^k \mid w \in L\},$$

and

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

(a) Prove that $L^{(1/3)}$ is regular. What about $L^{(3)}$?

(b) Let $k \geq 1$ be any natural number. Prove that there are only finitely many languages of the form $L^{(1/k)} = \{w \mid w^k \in L\}$ and that they are all regular. (In fact, if L is accepted by a DFA with n states, there are at most 2^{n^2} languages of the form $L^{(1/k)}$).

(c) Is $L^{1/\infty}$ regular or not? Is \sqrt{L} regular or not? What about L^∞ ?

Problem B6 (40 pts). Which of the following languages are regular? Justify each answer.

(a) $L_1 = \{w c w \mid w \in \{a, b\}^*\}$

(b) $L_2 = \{x y \mid x, y \in \{a, b\}^* \text{ and } |x| = |y|\}$

(c) $L_3 = \{a^n \mid n \text{ is a prime number}\}$

(d) $L_4 = \{a^m b^n \mid \gcd(m, n) = 17\}$.

TOTAL: 360 points.