Summer 1, 2006 CIS 511

Introduction to the Theory of Computation
Jean Gallier

Homework 2

June 30, 2006; Due June 8, 2006

"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 $h\colon \Sigma^* \to \Delta^*$ 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.

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 $h\colon \Sigma^* \to \Delta^*$ be a homomorhism, 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 (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 $x\backslash L$, 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 B3 (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:

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

and

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

(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 $wuRwv$ 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:

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

and

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

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 B4 (60 pts). Let L be any regular language over some alphabet Σ. Define the languages

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

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^n} 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 B5 (40 pts). Which of the following languages are regular? Justify each answer.

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

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

(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: 240 points.