Theory of Computation

Fall 2016, Midterm Exam.

Due: Nov. 14, 2016

1. (20 pts) True or False? (Score=max$\{0, Right - \frac{1}{2}Wrong\}$.)

   (1) If A, B and C are languages, then $A \cdot (B \cap C) = A \cdot B \cap A \cdot C$.
   Solution: $\times$ E.g. $\{1, 11\}(\{1\} \cap \{11\}) = \emptyset$ but $\{1, 11\}\{1\} \cap \{1, 11\}\{11\} = \{111\}$.

   (2) The set $\{a^m b^n c^n d^m | m, n \geq 0\}$ is not context-free.
   Solution: $\times$

   (3) If A is not context-free, then the complement of A is also not context-free.
   Solution: $\times$

   (4) For any language A, $(A^*)^* = A^*$
   Solution: ◯

   (5) If A is a language, then $A^2 \subseteq A$ implies $A = A^+$.
   Solution: ◯

   (6) $(a \cup b)^* = (a^* b^*)^*$.
   Solution: ◯

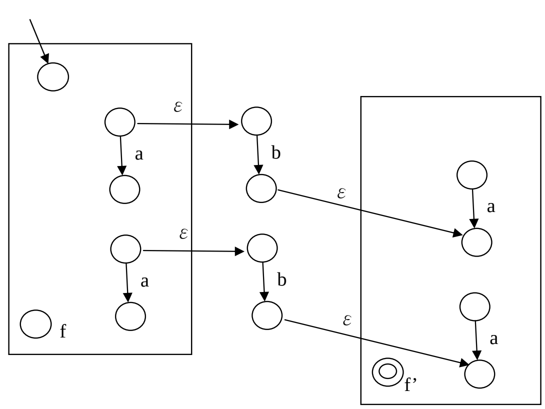
   (7) Not all finite languages are regular.
   Solution: $\times$

   (8) If L is nonregular and both of L' and $L \cap L'$ are regular, then $L \cup L'$ is nonregular.
   Solution: ◯ because $((L \cup L') - L') \cup (L \cap L') = L$.

   (9) Given a PDA A and an NFA B, there is an algorithm to decide whether $L(A) \cap L(B) = \emptyset$.
   Solution: ◯

   (10) There is a non-context-free language L such that L^* is regular.
   Solution: ◯ E.g. $L = \{0^p \mid p \text{ is prime }\}$.

2. (10 pts) Let Subst-One-Char$(w, a, b) = \{x \mid w$ can be written as $w = uav$ and $x = ubv\}$, where $u, v \in \Sigma^*$ and $a, b \in \Sigma$. This is the set of strings obtained by replacing an arbitrary one of the a's in w by a b. E.g. Subst-One-Char$(cacac, a, b) = \{cbcac, cacbc\}$. Let Subst-One-Char$(L, a, b)$ $= \{x \mid x \in$ Subst-One-Char(w, a, b) for some $w \in L\}$. Prove that if L is regular, then so is Subst-One-Char(L, a, b).
   Solution: Consider the following construction



3. (10 pts) Prove that at most $k^{2k+1}2^k$ languages over the binary alphabet $\{0,1\}$ can be recognized by a DFA with k states.
Solution: Simply count the number of distinct DFAs with k states. Name the states $1, 2, 3, \cdots, k$. Then a DFA is a tuple $(\{1,2,...,k\}, \{0,1\}, \delta, q_0, F)$ where $q_0 \in \{1,2,...,k\}$, $F \subseteq \{1,2,...,k\}$, $\delta : \{1,2,...,k\} \times \{0,1\} \to \{1,2,...,k\}$. Thus, the number of distinct ways to choose (q_0, F, δ) is $k \times 2^k \times k^{2k}$.

4. (10 pts) Let L be a nonempty language in which the shortest string has length k. Prove that L cannot be recognized by a DFA with fewer than $k+1$ states.
Solution:
(Proof 1): Let $q_0 \overset{w_1}{\to} q_1 \overset{w_2}{\to} q_2 \cdots \overset{w_k}{\to} q_{k+1}$ be an accepting computation. Argue that all the $q_1, q_2, \cdots q_{k+1}$ are distinct, for if $q_i = q_j, i \neq j$, then $w_1 w_2 ... w_k$ is not the shortest string in the language.
(Proof 2): Take any string $w = w_1 w_2 \cdots w_k$ in L. We claim that the $k+1$ strings ϵ, w_1, $w_1 w_2$, $w_1 w_2 w_3$, $\cdots$, $w_1 w_2 w_3 \cdots w_k$ are each in a different equivalence class of $\equiv_L$. Indeed, for any $i < j$,
$(w_1 w_2 \cdots w_i)(w_{j+1} \cdots w_k) \notin L$;
$(w_1 w_2 \cdots w_j)(w_{j+1} \cdots w_k) \in L$,
where the first line holds because that string is shorter than k. Since $\equiv_L$ has at least $k+1$ equivalence classes, the Myhill-Nerode theorem implies that any DFA for L must have at least $k+1$ states.

5. (10 pts) Let $L_1 \subseteq (a \cup b)^*$ be a set of strings. In each string in L_1, delete every b immediately following an a to get the set L_2. For instance, if $L_1 = \{aabba, aa\}$, then $L_2 = \{aaba, aa\}$. You are asked to define two homomorphisms $h_1, h_2 : \{a, b, \hat{b}\}^* \to \{a, b\}^*)$, and write an expression for L_2 in terms of $h_1, h_2, h_1^{-1}, h_2^{-1}, R, L_1$, for some regular expression R. Explain why your answer is correct.
Solution: Let $h_1(a) = a, h_1(b) = b, h_1(\hat{b}) = b$, $h_2(a) = a, h_2(b) = b, h_2(\hat{b}) = \epsilon$. Then

$$L_2 = h_2(h_1^{-1}(L_1) \cap b^*(a \cup a\hat{b}b^*)^*)$$

.

6. (10 pts) Consider the language L defined by the regular expression $(a^* \cup ba)^*$. Describe the equivalence classes of $\{a,b\}^*$ w.r.t. the Myhill-Nerode relation $\equiv_L$ defined by:
$x_1 \equiv_L x_2 \Leftrightarrow \forall y \in \Sigma^*, (x_1 \cdot y \in L \Leftrightarrow x_2 \cdot y \in L)$
Present these equivalence classes through regular expressions. Use $\equiv_L$ to construct a minimal automaton $M_{\equiv_L}$ for the language L, and draw the graph of the automaton. (Hint: choose equivalence classes from the following a^*, b^*, $(a \cup b)^*$, $(ab)^*$, $(a \cup ba)^*$, $(ab \cup b)^*$, $(ab \cup ba)^*$, $(a \cup ba)^*a$, $(a \cup ba)^*b$, $(a \cup ba)^*bb(a \cup b)^*$, $(a \cup ba)^*aa(a \cup b)^*$, $(a \cup ba)^*bab(a \cup b)^*$.)
Solution: The automaton is

- Q=$\{\{(a \cup ba)^*, (a \cup ba)^*b, (a \cup ba)^*bb(a \cup b)^*\}\}$
- $\delta((a \cup ba)^*,\ a) = (a \cup ba)^*$
  $\delta((a \cup ba)^*,\ b) = (a \cup ba)^*b$
  $\delta((a \cup ba)^*b,\ a) = (a \cup ba)^*$
  $\delta((a \cup ba)^*b,\ b) = (a \cup ba)^*bb(a \cup b)^*$
  $\delta((a \cup ba)^*bb(a \cup b)^*,\ a) = (a \cup ba)^*bb(a \cup b)^*$
  $\delta((a \cup ba)^*bb(a \cup b)^*,\ b) = (a \cup ba)^*bb(a \cup b)^*$
- $q_0 = (a \cup ba)^*$

Without using a rigorous method to construct the set of states, you may try the sequence of strings $\epsilon, a, b, aa, ab, ba, bb, ...$ to find out the equivalence classes. Also note that among a^*, b^*, $(a \cup b)^*$, $(ab)^*$, $(a \cup ba)^*$, $(ab \cup b)^*$, $(ab \cup ba)^*$, only one may appear in the final result as all of them contain ϵ. Also $(a \cup ba)^*a$ and $(a \cup ba)^*aa(a \cup b)^*$ both contain aa, and $(a \cup ba)^*b$ and $(a \cup ba)^*bab(a \cup b)^*$ both contain bab.

7. (10 pts) Give a right-linear grammar for the following regular language: $(00 \cup 1)^*$. Show your work in sufficient detail.
Solution: Let $G = (\{S, X\}, \{0, 1\}, \{S \to 0X \mid 1S \mid \epsilon; \quad X \to 0S\}, S)$

8. (10 pts) Is the language $\{a^{2^n} \mid n \geq 0\}$ context-free? Prove your answer.
Solution: Let p be the pumping constant. According to the pumping lemma, $a^{2p} = uvxyz$ with $|vy| \leq p$. Then $|uv^2xy^2z| > 2^p$ and $|uv^2xy^2z| \leq 2^p + p < 2^p + 2^p = 2^{p+1}$. Hence, uv^2xy^2z has length strictly between 2^p and 2^{p+1}, and therefore, the string cannot be in L.

9. (10 pts) A context-free grammar is called a *linear context-free grammar* if its production is of the following form $X \to aY$, $X \to Zb$, $X \to c \mid \epsilon$ (where X, Y, Z are nonterminals and a, b, c are terminals), i.e., the right-hand side of a production contain at most one nonterminal. Give a pumping lemma for linear context-free languages. Show why your statement is correct.
Solution: Pumping lemma for linear context-free languages.
Let L be a linear context-free language. Then there exists some positive integer m such that any $w \in L$ with $|w| \geq m$ can be decomposed as $w = uvxyz$, such that

- $|uvyz| \leq m$,
- $|vy| \geq 1$,
- $uv^ixy^iz \in L$, *for all* $i \geq 0$

Proof: see the following figure.