# Midterm #1

Mark all correct answers in each of the following questions.

$\Sigma$ denotes an arbitrary alphabet unless otherwise specified. $L$, $L_1$, $L_2$ are languages over $\Sigma$.

1. A language $L$ is *closed under concatenation* if for every $w_1, w_2 \in L$ we have $w_1 w_2 \in L$.

   (a) If $L$ is closed under concatenation, then $L^* = L$.

   (b) If $L$ is closed under concatenation, then the language $L^* - L$ is finite.

   (c) If $L$ is closed under concatenation, then it is infinite.

   (d) If $L_1 L_2 = L_2 L_1$, then $w_1 w_2 = w_2 w_1$ for every $w_1 \in L_1,\ w_2 \in L_2$.

   (e) If $L^*$ is finite, then $a^2 b^5 a^2 \notin L^*$ (where $a, b \in \Sigma$).

   (f) If $L_1 L_2$ is infinite, then both $L_1$ and $L_2$ are such.

   (g) None of the above.

2. (a) If $L_1$ and $L_2$ are finitely describable, then so are both $L_1 \cup L_2$ and $L_1 L_2$.

   (b) If $L$ is regular, then it is finitely describable.

   (c) If $L$ is finitely describable, then there exists an NFA accepting it, and therefore a DFA accepting it as well.

   (d) A finitely describable language may be infinite, and even uncountable.

   (e) If $L_1$ is finitely describable and the symmetric difference $(L_2 - L_1) \cup (L_1 - L_2)$ is finite, then $L_2$ is finitely describable.

(f) If there exist DFA's accepting the languages $L_1$, $L_2$, $L_3$, then there exists an NFA accepting $L_1 \cup L_2 L_3^*$.

(g) None of the above.

3. Let $\alpha$, $\beta$ be regular expressions.

(a) If $|\alpha| \neq |\beta|$, then $L(\alpha) \neq L(\beta)$.

(b) If $L(\alpha) \supseteq L(\beta)$, then $|\alpha| \geq |\beta|$.

(c) There exist infinitely many mutually distinct regular expressions $\alpha_1, \alpha_2, \ldots$, such that $L(\alpha_n) = L(\alpha)$ for each $n$.

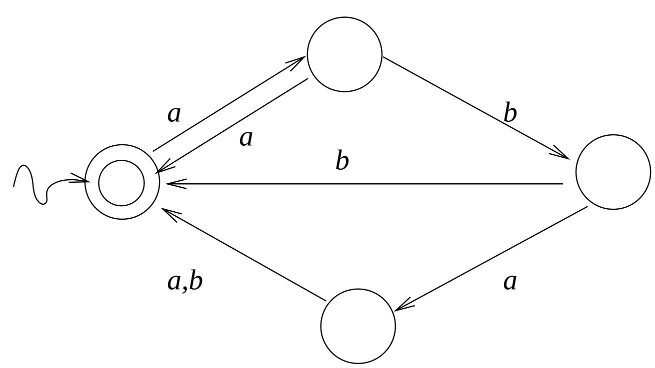
(d) If $L(\alpha) \neq \emptyset$ and $L(\beta) \neq \emptyset$, then $L((\alpha\beta)) \supseteq L(\alpha)$ and $L((\alpha\beta)) \supseteq L(\beta)$.

(e) If $\alpha \neq \phi$, then there exists a regular expression $\gamma$ (obtained by solving the equation $\alpha x = \beta$) for which $L((\alpha\gamma)) = L(\beta)$.

(f) If $\alpha = (010 \cup 101)^*$, then $L(\alpha)$ is strictly contained in the set of all words over $\{0, 1\}$ whose length is divisible by 3 and which contain neither of the words $0^3$ and $1^3$ as a subword.

(g) None of the above.

4. Let $M$ be the following automaton:




(a) $L(M) = L((a^2 \cup ab^2 \cup aba(a \cup b))^*)$.

(b) $L(M) \supseteq \{a^{2n}b^{2n} : n \geq 0\}$.

(c) $L(M) \supsetneqq L\left(((a^2)^*ab(b \cup a^2 \cup ab))^*\right)$.

(d) $L(M) = L\left((a^2)^* \cup (ab^2)^* \cup (aba(a \cup b))^*\right)$.

(e) $L(M)$ is finite.

(f) There exists a finite automaton $M_1$ such that $L(M_1)$ is finite and $L(M) = L(M_1)^*$.

(g) None of the above.

5. Let $M = (Q, \Sigma, \Delta, s, A)$ be an NFA such that for every $(q, w, p) \in \Delta$ we have $|w| \leq 1$. Let $M' = (Q', \Sigma, \delta', s', A')$ be the equivalent DFA constructed according to the algorithm presented in class.

   (a) $|Q'| = 2^{|Q|}$ (where $|Q|$ and $|Q'|$ are the sizes of $Q$ and $Q'$, respectively).

   (b) If $A = Q$, then $A' = Q'$.

   (c) We have $s' \in A'$ if and only if $s \in A$.

   (d) If the set $A$ is augmented by turning one of the non-accepting states into an accepting state, then the set $A'$ increases as well. In particular, the language $L(M')$ contains words not accepted before the change.

   (e) If for every $(q, w, p) \in \Delta$ we have $|w| = 1$, then in $M'$ it is impossible to get from the state $s'$ to any state $K \in Q'$ with $|K| \geq 2$.

   (f) If for some $q \in Q$ and $\sigma \in \Sigma$ we have $(q, \sigma, q) \in \Delta$, then there exists some $K \in Q' - \{\emptyset\}$ for which $\delta'(K, \sigma) = K$.

   (g) None of the above.

## Solutions

1. If $L$ is closed under concatenation, then (by induction) any concatenation of a positive number of words in $L$ again belongs to $L$, so that $L \supseteq L^+$. However, we do not necessarily have $\varepsilon \in L$. For example, the

language $L = \Sigma^* - \{\varepsilon\}$ is closed under concatenation. Thus, $L^* - L$ is either empty or consists of the word $\varepsilon$ only.

The languages $\emptyset$ and $\{\varepsilon\}$ are closed under concatenation, yet they are finite.

Taking $L_1 = L_2 = \Sigma^*$ we have $L_1 L_2 = L_2 L_1 = \Sigma^*$, yet words in $L_1$ and $L_2$ do not necessarily commute; for example, $ab \neq ba$.

If $L$ contains a non-empty word, then $L^*$ contains all the powers of this word, which are mutually distinct, and in particular $L^*$ is infinite. Hence $L^*$ is finite if and only if $L = \emptyset$ or $L = \{\varepsilon\}$. In both of the latter cases we have $L^* = \{\varepsilon\}$.

The product $\{a\}^*\{a\}$ is infinite although the second factor consists of a single word.

Thus, only (b) and (e) are true.

2. Any language which is finitely describable in terms of finitely many finitely describable languages is itself finitely describable. In particular, if $L_1$ and $L_2$ are finitely describable, then so are $L_1 \cup L_2$ and $L_1 L_2$.

   A regular expression is a particular form of a finite description of a language, so that a regular language is finitely describable.

   The language $\{a^n b^n : n \geq 0\}$ is finitely describable, yet is not accepted by any DFA.

   The language $\{a, b\}^*$ is infinite and finitely describable. However, as $\Sigma^*$ is countable, every language is finite or countable.

   A word belongs to $L_2$ if and only if it either belongs to $L_1$ and not to the symmetric difference $(L_2 - L_1) \cup (L_1 - L_2)$ or it does not belong to $L_1$ but does belong to the symmetric difference. If the symmetric difference is finite, listing all its words in the above statement provides a finite description of $L_2$.

   If the languages $L_1$, $L_2$, $L_3$ are accepted by finite automata, then, by the closure properties of the family of languages accepted by finite automata, there exists a finite automaton accepting $L_1 \cup L_2 L_3^*$. Since for every NFA there exists an equivalent DFA, in the above it is immaterial whether we consider DFA's or NFA's.

Thus, only (a), (b), (e) and (f) are true.

3. Taking, say, $\alpha_1 = (\alpha \cup \phi)$, $\alpha_2 = (\alpha_1 \cup \phi)$, and in general $\alpha_n = (\alpha_{n-1} \cup \phi)$ for all $n$, we obtain infinitely many pairwise distinct regular expressions representing the same language.

   If $\alpha = \beta = a$, then $L(\alpha) = L(\beta) = \{a\}$ is not contained in $L((\alpha\beta)) = \{a^2\}$.

   For $\alpha = a$ and $\beta = b$ there exists no regular expression $\gamma$ for which $L((\alpha\gamma)) = L(\beta)$.

   The language $L((010 \cup 101)^*)$ consists of all words of the form $w_1 w_2 \dots w_n$ with $n \geq 0$ and each $w_i$ being either 010 or 101. Clearly, the length of any such word is divisible by 3, and the word cannot contain more than 2 consecutive 0's or 1's. The word 001, for example, satisfies the latter two conditions, yet does not belong to $L((010 \cup 101)^*)$.

   Thus, only (c) and (f) are true.

4. Words accepted by $M$ are those leading from $s$ back to $s$. A word has this property if it is a concatenation of any number of words leading from $s$ back to $s$ without being at $s$ in the meantime. The only words having the latter property are $a^2$, $ab^2$, $aba^2$ and $abab$. Hence $L(M) = L((a^2 \cup ab^2 \cup aba^2 \cup abab)^*)$.

   The word $a^2b^2$ is clearly not accepted by $M$.

   Every non-empty word in $L\left(((a^2)^* ab(b \cup a^2 \cup ab))^*\right)$ contains at least one occurrence of $b$, and in particular this language does not contain the word $a^2$, which is accepted by $M$. On the other hand, every word of one the three forms $a^{2k+1}b^2$, $a^{2k+1}ba^2$ and $a^{2k+1}bab$ moves us in $M$ from $s$ to itself. Hence any concatenation of such words has the same property, and therefore $L(M) \supseteq L\left(((a^2)^* ab(b \cup a^2 \cup ab))^*\right)$.

   The word $a^3b^2$ is clearly accepted by $M$, yet it does not belong to $L\left((a^2)^* \cup (ab^2)^* \cup (aba(a \cup b))^*\right)$.

   Since the language $\{a^2,\ ab^2,\ aba^2,\ abab\}$ is finite, there exists an automaton $M_1$ accepting it. Then $L(M) = L(M_1)^*$.

Thus, only (a), (c) and (f) are true.

5. In our construction of an equivalent DFA we have $Q' = 2^Q$, and therefore $|Q'| = 2^{|Q|}$.

   Since $A'$ consists of those sets of elements intersecting $A$, the set $\emptyset$ never belongs to $A'$ in our constructuion.

   We have $s' \in A'$ if and only if at least one of the states of $Q$ to which we can get from $s$ by means of $\varepsilon$-transitions belongs to $A$. This condition is clearly strictly weaker than the condition $s \in A$.

   If the state $q$ is adjoined to $A$, then the state $\{q\}$ is adjoined to $A'$. However, this change does not necessarily imply that the language accepted by the automaton increases since the state $q$ may be unreachable from $s$. In fact, in this case the change does not add words to the language accpeted by the automaton.

   Even if there are no $\varepsilon$-transitions in $M$, the subset $\delta'(s', \sigma)$ of $Q$ may consist of more than one element. For example, if both $(s, a, q_1) \in \Delta$ and $(s, a, q_2) \in \Delta$, then both $q_1$ and $q_2$ belong to $\delta'(s', a)$.

   If $(q, \sigma, q) \in \Delta$, then $\delta'(\{q\}, \sigma) \supseteq \{q\}$. Let $K_0 = \{q\}$, $K_1 = \delta'(K_0, \sigma)$, $K_2 = \delta'(K_1, \sigma)$, and so forth. A simple induction shows that $K_{i+1} \supseteq K_i$ for each $i$, and since all $K_i$'s are contained in $Q$, we must have $K_{i+1} = K_i$ from some place on. Taking $K = K_i$ for such an $i$, we have $\delta'(K, \sigma) = K$.

   Thus, only (a) and (f) are true.