CS 341-008, Spring 2021, Hybrid Section
Solutions for Midterm 1

1. (a) False. The language a^* is regular but infinite.
   (b) True. Homework 3, problem 2b.
   (c) False. Let $A = \{ a^n b^n \mid n \geq 0 \}$ and $B = (a \cup b)^*$. Then $A \subseteq B$, A is nonregular, and B is regular.
   (d) False. Homework 6, problem 2a.
   (e) False. If A is recognized by an NFA, then A must be regular by Corollary 1.40.
   (f) False. $A = \{ a^n b^n \mid n \geq 0 \}$ is context-free but not regular.
   (g) True. If A is recognized by an NFA, then A is regular by Corollary 1.40. Then by Corollary 2.32, A is context-free, so Theorem 2.9 ensures that A has a CFG in Chomsky normal form.
   (h) False. Let $A = \{ a^n b^n c^n \mid n \geq 0 \}$ and B have regular expression $(a \cup b \cup c)^*$. Then $A \subseteq B$, A is not context-free (see slide 2-96), and B is context-free because it is regular (Corollary 2.32).
   (i) False. Let $A = \emptyset$ and $B = \{ a^n b^n c^n \mid n \geq 0 \}$. Then A is regular because it is finite (slide 1-95), so A is also context-free (Corollary 2.32). Language B is not context-free (see slide 2-96). But $A \circ B = \emptyset$ (e.g., see slide 0-30), which is regular so also context-free.
   (j) False. 1^*0^* generates the string $1000 \notin A$, so the regular expression is not correct. In fact, A is nonregular, so it can't have a regular expression.

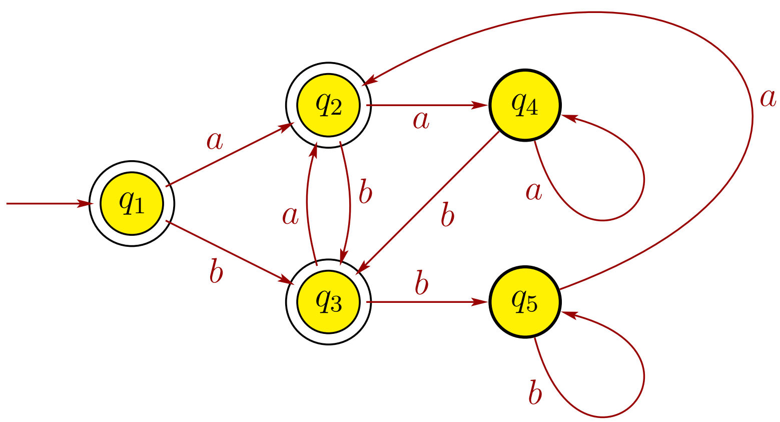
2. (a) $(bb \cup a)b^*aa^*$ Another regular expression is $bbb^*aa^* \cup ab^*aa^*$. There are infinitely many correct regular expressions for the language.
   (b) $G_2 = (V_2, \Sigma, R_2, S_2)$ with $S_2 \notin V_1$, where
      - $V_2 = V_1 \cup \{S_2\}$,
      - S_2 is the (new) starting variable,
      - Σ is the same alphabet of terminals as in G_1, and
      - $R_2 = R_1 \cup \{S_2 \to S_1 S_2 \mid \varepsilon\}$.
   (c) $M_3 = (Q_3, \Sigma, \delta_3, q_3, F_3)$, where
      - $Q_3 = Q_1 \times Q_2$;
      - Σ is the same alphabet as M_1 and M_2 have;
      - the transition function δ_3 satisfies $\delta_3((q,r),\ell) = (\delta_1(q,\ell), \delta_2(r,\ell))$ for $(q,r) \in Q_3$ and $\ell \in \Sigma$;
      - the starting state $q_3 = (q_1, q_2)$; and
      - $F_3 = (Q_1 \times F_2) \cup (F_1 \times Q_2)$

(d) After the one step of removing $A \to \varepsilon$, the CFG is then

$$\begin{aligned} S_0 &\to S \\ S &\to 10A1SA \mid 101SA \mid 10A1S \mid 101S \mid A101 \mid 101 \mid \varepsilon \\ A &\to 110A0 \mid 1100 \end{aligned}$$

3. (a) For the language $C = \{\, w \in \Sigma^* \mid w \text{ does not end in a double letter } \}$ with $\Sigma = \{a, b\}$, consider its complement $\overline{C} = \{\, w \in \Sigma^* \mid w \text{ ends in a double letter } \}$, which has DFA on slide 1-17 of the notes. The complement of $\overline{C}$ is $\overline{\overline{C}} = C$, so we can obtain a DFA M for C by swapping the accepting and non-accepting states of a DFA for $\overline{C}$.

A DFA for C is below:



A 5-tuple description of the DFA above is $M = (Q, \Sigma, \delta, q_1, F)$, where

- $Q = \{q_1, q_2, q_3, q_4, q_5\}$
- $\Sigma = \{a, b\}$
- The transition function $\delta : Q \times \Sigma \to Q$ is defined as

	a	b
q_1	q_2	q_3
q_2	q_4	q_3
q_3	q_2	q_5
q_4	q_4	q_3
q_5	q_2	q_5

- q_1 is the start state
- $F = \{q_1, q_2, q_3\}$

There are infinitely many other correct DFAs for C.

(b) A regular expression for C is $\varepsilon \cup a \cup b \cup (a \cup b)^*(ab \cup ba)$. There are infinitely many other correct regular expressions for C.

4. A CFG for $D = \{ b^i a^j b^k \mid i, j, k \geq 0, \text{ and } k = i + j \}$ is $G = (V, \Sigma, R, S)$ with set of variables $V = \{S, X\}$, where S is the start variable; set of terminals $\Sigma = \{a, b\}$; and rules

$$\begin{aligned} S &\rightarrow bSb \mid X \\ X &\rightarrow aXb \mid \varepsilon \end{aligned}$$

There are infinitely many other correct CFGs for D.

5. Language $D = \{ b^i a^j b^k \mid i, j, k \geq 0, \text{ and } k = i + j \}$ is nonregular. We prove this by contradiction. Suppose that D is a regular language. Let p be the "pumping length" of the Pumping Lemma. Consider the string

$$s = b^p a^p b^{2p}.$$

Note that $s \in D$ because the number of b's at the end $(2p)$ equals the sum of the number of b's at the beginning (p) and the number of a's in the middle (p). Also, the length of s is $|s| = 4p > p$, so the Pumping Lemma will hold. Thus, there exists strings x, y, and z such that $s = xyz$ and

(i) $xy^i z \in D$ for each $i \geq 0$,

(ii) $|y| > 0$,

(iii) $|xy| \leq p$.

Since the first p symbols of s are all b's, the third property implies that x and y consist only of b's. So z will be the rest of the b's at the beginning, followed by $a^p b^{2p}$. The second property states that $|y| > 0$, so y has at least one b. More precisely, we can then say that

$$\begin{aligned} x &= b^j \text{ for some } j \geq 0, \\ y &= b^k \text{ for some } k \geq 1, \\ z &= b^m a^p b^{2p} \text{ for some } m \geq 0. \end{aligned}$$

Since $b^p a^p b^{2p} = s = xyz = b^j b^k b^m a^p b^{2p} = b^{j+k+m} a^p b^{2p}$, we must have that

$$j + k + m = p, \text{ where } k \geq 1.$$

The first property implies that $xy^2 z \in D$, but

$$\begin{aligned} xy^2 z &= b^j b^k b^k b^m a^p b^{2p} \\ &= b^{p+k} a^p b^{2p} \notin D \end{aligned}$$

because the number of b's at the end $(2p)$ does not equal the number of b's at the beginning $(p + k)$ plus the number of a's in the middle (p) since $p + k > p$ by $k \geq 1$. Because the pumped string $xy^2 z \notin D$, we have a contradiction. Therefore, D is a nonregular language.