Assignment 2

1. Suppose $M = (Q, \Sigma, \delta, q_0, A)$ with extended transition function $\delta^\star : Q \times \Sigma^\star \to Q$. Prove that for any $q \in Q$ and $x, y \in \Sigma^\star$ that $\delta^\star(q, xy) = \delta^\star(\delta^\star(q, x), y)$. Hint: prove by induction on $|y|$.

   Solution: Proof by induction on $|y|$. Base case $y = \epsilon$. Assume that $x \in \Sigma^\star$ is an arbitrary string. Then $\delta^\star(\delta^\star(q, x), y) = \delta^\star(\delta^\star(q, x), \epsilon) = \delta^\star(q, x) = \delta^\star(q, x\epsilon) = \delta^\star(q, xy)$, as needed.

   Now assume $\delta^\star(q, xy) = \delta^\star(\delta^\star(q, x), y)$ for $|y| = n$. Prove for $|y| = n + 1$. Write $y = za$ where $a \in \Sigma, |z| = n$. Then $\delta^\star(q, xy) = \delta^\star(q, xza) = \delta(\delta^\star(q, xz), a) = \delta(\delta^\star(\delta^\star(q, x), z), a) = \delta^\star(\delta^\star(q, x), za) = \delta^\star(\delta^\star(q, x), y)$, as needed.

2. Prove by induction on n that if L is a language and R is a regular expression such that $L = L(R)$ then there exists a regular expression R_n such that $L(R_n) = L^n$. Be sure to use the fact that if R_1 and R_2 are regular expressions then $L(R_1R_2) = L(R_1) \cdot L(R_2)$.

   Solution: Proof by induction on n. Base case $n =$ then $L^0 = \{\epsilon\}$ and so $R_0 = \epsilon$.

   Inductive hypothesis: assume there exists a value k such that when $n = k$ there exists R_n such that $L(R_n) = L^n$.

   Inductive step: prove this for $n = k+1$. $L^{k+1} = L \cdot L^k$. We know that R is a regular expression for L and that R_k is a regular expression for L^k. By definition of concatenation of regular expressions, $L(R_1R_2) = L(R_1)L(R_2)$ and so $L(RR_k) = L(R) \cdot L(R_k) = L(R) \cdot L(R)^k = L(R)^{k+1}$, as needed.

3. Every finite-sized language L has a regular expression that accepts it. Informally, the expression for $\{w, x, y\}$ is $w + x + y$. Prove this is true using induction on the number of words in L. You may need to use these definitions in your proof: (i) if w is a word then w is also a regular expression with $L(w) = \{w\}$, (ii) if R_1 and R_2 are regular expressions then $R_1 + R_2$ is a regular expression, and (iii) $L(R_1 + R_2) = L(R_1) \cup L(R_2)$.

   Solution: Proof by induction on the number of words. Base case $|L| = 0 \Rightarrow L = \{\}$. Then $R = \emptyset$ is a regular expression for L.

   Inductive hypothesis: assume there exists a k such that if $|L| = k$ then there exists R_k such that $L(R_k) = L$.

   Inductive step: prove that the statement holds for $|L| = k + 1$. Let $w \in L$ be an arbitrary word. We know such a word exists because $|L| = k + 1 \geq 1$. Consider $L' = L \setminus \{w\}$. Since $w \in L$ we have $|L'| = k$. By the inductive hypothesis, we know there exists a regular expression R_k such that $L(R_k) = L'$. Moreover, by definition, the regular expression w has $L(w) = \{w\}$. Thus we can create the regular expression $w + R_k$ such that $L(w + R_k) = L(w) \cup L(R_k) = \{w\} \cup L' = \{w\} \cup (L \setminus \{w\}) = L$, as needed.

4. Give a generic construction for a finite state machine that accepts a non-empty finite-size language $L = \{w_1, \dots w_n\}$. Define all parts of the machine. (You do not need to prove it accepts the langauge). Hint: do this for a simple language like $\{\epsilon, a, aa, ba, ab\}$ and then generalize what you did so it can work for any finite language. Hint: make $Q \subseteq \Sigma^\star$.

   Solution: We construct a machine $M = (Q, \Sigma, \delta, q_0, A)$ to accept it as follows. Let n be the length of the longest word in L. Since L is finite then there is a unique longest word and n is a finite value. Then construct the following machine: $Q = \bigcup_{i=0}^{n} \Sigma^i \cup \{E\}$, that is, there is a unique state for every possible word of length up to and including n, as well a an error state E. We define $\delta(q, a) = qa$ if $|q| < n$ and $\delta(q, a) = E$ otherwise, $q_0 = \epsilon$ and $A = L$.

   Note how we use the idea that "states" here are actually the words, not an arbitrarily named state that we associate with the word, and the transitions effectively concatenate a letter to the word, so the start state is empty word. This technique makes it also easier to prove that it works, which we provide in addition: we prove that $\delta^\star(q_0, w) = w$ for all $|w| \leq n$. Base case, $w = \epsilon$ then $\delta^\star(q_0, \epsilon) = q_0 = \epsilon$, as needed. Now assume for length until k that our statement holds. Prove for $|w| = k+1 \leq n$. Write $w = va$. We know that $\delta^\star(q_0, v) = v$, so $\delta^\star(q_0, w) = \delta(\delta^\star(q_0, v), a) = \delta(v, a) = va$ by the definition of δ and that fact that $|v| < n$.

   To prove $L = L(M)$, consider $L(M) = \{w \in \Sigma^\star : \delta^\star(q_0, w) \in A\}$ so $L(M) = \{w \in \Sigma^\star : w \in A\}$ and $L(M) = \{w \in \Sigma^\star : w \in L\} = L$, as needed.

5. Suppose L is a regular language, and $M = (Q, \Sigma, \delta, q_0, A)$ is a deterministic finite state machine such that $L(M) = L$. Prove that if $|Q| = 2$ then at least one of the following hold: (i) $L = \emptyset$ (ii) $\epsilon \in L$, or (iii) $\exists a \in \Sigma$ such that $a \in L$. Note that different M with $|Q| = 2$ may have a different property that holds. Hint: do a case analysis for configurations of A.

   Solution: Let $Q = \{q_0, q_1\}$. There are four possible values for A, which are $\emptyset$, $\{q_0\}$, $\{q_1\}$, and $\{q_0, q_1\}$. We divide this into three cases: (1) $A = \emptyset$, (2) $q_0 \in A$, and (3) $q_0 \notin A \wedge q_1 \in A$.

   Case 1. $A = \emptyset$. Then there cannot exist $w \in \Sigma^\star$ such that $\delta^\star(q_0, w) \in A$ and therefore $L = \emptyset$ by definition of $L(M)$.

   Case 2. $q_0 \in A$. Then $\delta^\star(q_0, \epsilon) = q_0 \in A$ so $\epsilon \in L$.

   Case 3. $A = \{q_1\}$. There are two subcases: (i) there exists $a \in \Sigma$ such that $\delta(q, a) = q_1$ or (ii) there does not exist any such a. If $\delta(q, a) = q_1$ then $\delta^\star(q, a) \in A$ and so $a \in L$. Otherwise, $\delta^\star(q_0, w) = q_0$ for all $w \in \Sigma^\star$, which we now prove by contradiction. Assume there exists a string w such that $\delta^\star(q_0, w) = q_1$ and assume that w is shortest. We know $w \neq \epsilon$ since $\delta^\star(q_0, \epsilon) = q_0$. So $w = w_1 \dots w_n$, and $\delta^\star(q_0, w_1 \dots w_{n-1}) = q_0$ because there are only two states and it is not q_1 since w is shortest. Then $\delta(q_0, w_n) = q_1$, a contradiction. Therefore, $\delta^\star(q_0, w) = q_0 \notin A$ and so $L = \emptyset$.