Assignment 2 solutions

1. Let $\Sigma = \{0, 1, 2\}$, and consider the following three languages over Σ:

$$A_1 = \{x2y : x, y \in \{0,1\}^*, |x| = |y|\},$$
$$A_2 = \{x2y : x, y \in \{0,1\}^*, x \neq y\},$$
$$A_3 = \{xy : x, y \in \{0,1\}^*, x \neq y\},$$
$$A_4 = \{xy : x, y \in \{0,1\}^*, |x| = |y| \text{ and } x \neq y\}.$$

Determine whether each of the languages A_1, A_2, A_3 and A_4 is regular or not, and prove that your answers are correct.

Solution. To prove non-regularity of a language we can use the pumping lemma or the Myhill-Nerode theorem. For all languages proved to be non-regular below either approach works, but we choose to use the Myhill-Nerode theorem. However, for language A_2 we also present a detailed proof which uses the pumping lemma.

The language A_1 is **non-regular**. Consider the infinite set of strings $B = L(0^*)$. For every two stings $x, y \in B$ such that $x \neq y$ we have $x2x \in A_1$, but $y2x \notin A_1$, and therefore $x \not\equiv_{A_1} y$, where $\equiv_{A_1}$ is the equivalence relation used in the Myhill-Nerode theorem. Because $B \subset \Sigma^*$ is infinite, $\equiv_{A_1}$ divides Σ^* in infinitely many equivalence classes. Hence A_1 is non-regular by the Myhill-Nerode theorem.

The language A_2 is **non-regular**. Again, let $B = L(0^*)$. Now, for every two stings $x, y \in B$ such that $x \neq y$ we have $x2x \notin A_2$, but $y2x \in A_2$, which gives us $x \not\equiv_{A_2} y$. The equivalence relation $\equiv_{A_2}$ divides Σ^* in infinitely many equivalence classes, therefore A_2 is non-regular by the Myhill-Nerode theorem.

We can also prove that A_2 is non-regular using the pumping lemma for regular languages. Suppose the contrary: A_2 is regular. Let $p \in \mathbb{N}$ be the pumping length for A_2. Consider a string $s = 0^p 2 0^{p+p!}$, which is in A_2 because $0^p \neq 0^{p+p!}$. We have $|s| \geq p$, therefore according to the pumping lemma there exist strings $x, y, z \in \Sigma^*$ such that $s = xyz$, $|xy| \leq p$, $|y| \geq 1$, and $xy^n z \in A_2$ for each $n \in \mathbb{N}$. Hence, there exist $l, m \in \mathbb{N}$ such that $l + m \leq p$, $m \geq 1$, $x = 0^l$, $y = 0^m$ and $z = 0^{p-l-m} 2 0^{p+p!}$. Now let us choose $n = 1 + p!/m$, which is natural number due to the fact that $m \leq p$. According to the pumping lemma we have $xy^n z = 0^l 0^{m(1+p!/m)} 0^{p-l-m} 2 0^{p+p!} = 0^{p+p!} 2 0^{p+p!} \in A_2$, which is obviously not true. Hence, we have obtained a contradiction and A_2 is non-regular.

The language A_3 is **regular**. Note that $\varepsilon \notin A_3$. However, every non-empty string $w \in \{0,1\}^*$ can be written as $w = xy$, where $x = \varepsilon$ and $y = w$. Clearly $x \neq y$, thus $w \in A_3$. This means that $A_3 = \{0,1\}^* \setminus \{\varepsilon\} = L((0 \cup 1)(0 \cup 1)^*)$ is regular.

The language A_4 is **non-regular**. Consider the infinite set of strings $C = L((00)^* 1)$. For every two stings $x, y \in C$ such that $x \neq y$ we have $xx \notin A_4$, while $yx \in A_4$, where the latter follows from the fact that yx has even length and, when it is divided into two parts of equal length, the last symbol of the first part is 0, but the last symbol of the second part is 1. Hence $x \not\equiv_{A_4} y$, and similarly as above A_4 is non-regular by the Myhill-Nerode theorem.

Without going into details, here I just mention that we could also consider strings $0^{p_1} 2 0^{p_1} \in A_1$ and $0^{p_4} 1 0^{p_4 + 2p_4!} 1 \in A_4$ to prove non-regularity of A_1 and A_4 using the pumping lemma, where p_1 and p_4 are respective pumping lengths.

2. Give context-free grammars for each of the languages A_1, A_2, A_3 and A_4 defined in question 1.

Solution. A context-free grammar for A_1:

$$\begin{aligned} S &\to ESE \mid 2 \\ E &\to 0 \mid 1 \end{aligned}$$

A context-free grammar for A_2:

$$\begin{aligned} S &\to EWC \mid CEW \mid A_0 1W \mid A_1 0W \\ C &\to ECE \mid 2 \\ W &\to WE \mid \varepsilon \\ E &\to 0 \mid 1 \\ A_0 &\to EA_0E \mid 0W2 \\ A_1 &\to EA_1E \mid 1W2 \end{aligned}$$

From EW we obtain all non-empty strings in $\{0,1\}^*$, and from C all the strings of odd length having a unique 2 in the middle. Therefore $EWC \mid CEW$ gives all the strings in $\{x2y : x, y \in \{0,1\}^*, |x| \neq |y|\}$. From A_0 we get all the strings in $\{u_1 0 v 2 u_2 : u_1, u_2, v \in \{0,1\}^*, |u_1| = |u_2|\}$, which means that $A_0 1W$ gives all the strings w for which there exist $x, y \in \{0,1\}^*$ and $n \in \mathbb{N}$ such that $w = x2y$, n-th symbol of x is 0 and n-th symbol of y is 1. The same way we treat $A_1 0W$. Two strings are different if either they have different lengths or for some $n \in \mathbb{N}$ their n-th symbol differs (or both); hence the grammar above gives us A_2.

A context-free grammar for A_3:

$$S \to SS \mid 0 \mid 1$$

A context-free grammar for A_4:

$$\begin{aligned} S &\to C_0C_1 \mid C_1C_0 \\ C_0 &\to EC_0E \mid 0 \\ C_1 &\to EC_1E \mid 1 \\ E &\to 0 \mid 1 \end{aligned}$$

Suppose we use the rule $S \to C_0C_1$ first, then $n \geq 0$ times use the rule $C_0 \to EC_0E$ followed by $C_0 \to 0$ and $m \geq 0$ times the rule $C_1 \to EC_1E$ followed by $C_1 \to 1$, and finally any rules for the instances of E. This way we can obtain all the strings of form $u0v1w$, where $u, v, w \in \{0,1\}^*$, $|u| = n$, $|v| = m + n$, and $|w| = m$, that is, all the strings of form xy such that $xy \in \{0,1\}^*$, $|x| = |y| = n + 1 + m$, $n+1$-st bit of x is 0 and $n+1$-st bit of y is 1. The case when we first use the rule $S \to C_1C_0$ is analogous. Therefore it is easy to see that the grammar above generates A_4.

3. Let $\Sigma = \{0,1\}$, and define a language $A \subseteq \Sigma^*$ as follows:

$$A = \{x \in \Sigma^* : x \text{ contains the substrings 01 and 10 an equal number of times}\}.$$

For example, 10001001 is in A, because

(i) it contains the substring 01 two times: $100\underline{01}0\underline{01}$

(ii) and it contains the substring 10 two times: $\underline{10}00\underline{10}01$

Determine which of the following statements is true:

(a) A is regular.
(b) A is not regular, but it is context-free.
(a) A is not context-free.

Give a detailed proof of whichever statement you select.

Solution. The language A is **regular**. We show that A can be given by the regular expression $\varepsilon \cup 0 \cup 1 \cup 0(0 \cup 1)^*0 \cup 1(0 \cup 1)^*1$, in other words, A consists of the empty string and all the strings which start and end with the same symbol. Note that the substring 01 appears in a string in places where a block of consecutive 0s (at least one 0) is followed by a block of consecutive 1s (at least one 1), and 10 appears in places where a block of consecutive 1s is followed by a block of consecutive 0s. Therefore, the number of these appearances can differ by at most one, and in order for the substrings 01 and 10 to appear an equal number of times, the string must start and end with the same symbol (unless it is empty).

4. Suppose α is a real number satisfying $\alpha > 1$. Define a language

$$E_\alpha = \left\{0^{\lfloor \alpha^n \rfloor} : n \in \mathbb{N}\right\}.$$

Prove that E_α is not context-free.

Solution. Suppose the contrary, that is, suppose that E_α is context-free. Then the pumping lemma for context-free languages states that there exists a number $p \in \mathbb{N}$ such that, if $s \in E_\alpha$ has the length at least p, then s may be divided in five pieces $s = uvxyz$ so that $0 < |vy| \le |vxy| \le p$ and for each $q \in \mathbb{N}$ we have $uv^qxy^qz \in E_\alpha$. Choose a natural number n such that $n > \log_\alpha(\frac{p+1}{\alpha-1})$, and therefore $\lfloor \alpha^{n+1} \rfloor - \lfloor \alpha^n \rfloor \ge \alpha^{n+1} - \alpha^n - 1 = \alpha^n(\alpha - 1) - 1 > p$. Let $s = 0^{\lfloor \alpha^{n+1} \rfloor} \in E_\alpha$. Because $|s| > p$, there exist $u, v, x, y, z \in \Sigma^*$ such that $s = uvxyz$, $1 \le |vy| \le p$ and $s' = uxz \in E_\alpha$. But this cannot be the case, because $\lfloor \alpha^{n+1} \rfloor > |s'| \ge \lfloor \alpha^{n+1} \rfloor - p > \lfloor \alpha^n \rfloor$ and there is no $m \in \mathbb{N}$ such that $s' = 0^{\lfloor \alpha^m \rfloor}$.

5. Let $\Sigma = \{0, 1\}$ and let $A \subseteq \Sigma^*$ be a given regular language. Define a new language $B \subseteq \Sigma^*$ as

$$B = \{x \in \Sigma^* : xx \in A\}.$$

Is B necessarily a regular language? If your answer is "yes," prove that this is so, and if your answer is "no," give a counter-example.

Solution. Given that A is regular, there must exist a DFA $K = (Q, \Sigma, \delta, q_0, F)$ with $L(K) = A$. For every $q \in Q$ define two DFA $M_q = (Q, \Sigma, \delta, q_0, \{q\})$ and $N_q = (Q, \Sigma, \delta, q, F)$. We claim that

$$B = \bigcup_{q \in Q} L(M_q) \cap L(N_q).$$

Because regular languages are closed under intersection and union, B is **regular**. It remains to prove that the claim above is true.

Suppose $x \in B$. Then $xx \in A$. Let $p = \delta^*(q_0, x)$, and, because $\delta^*(q_0, xx) \in F$, we have $\delta^*(p, x) \in F$. Therefore $x \in L(M_p)$ and $x \in L(N_p)$, which means $x \in \bigcup_{q \in Q} L(M_q) \cap L(N_q)$.

Now suppose $y \in \bigcup_{q \in Q} L(M_q) \cap L(N_q)$, and thus there exists $p \in Q$ such that $y \in L(M_p)$ and $y \in L(N_p)$. Then $\delta^*(q_0, y) \in \{p\}$ and $\delta^*(p, y) \in F$, which implies $\delta^*(q_0, yy) \in F$. Hence $yy \in A$ and $y \in B$. This concludes the proof.

Another way to prove that B is regular is by considering the following NFA. Define an NFA $H = (\{s\} \cup Q^3, \Sigma, \mu, s, F')$, with its transition function μ defined as

1. $\mu(s, \varepsilon) = \{(q_0, q, q) : q \in Q\}$, and
2. $\mu((q, p, r), \sigma) = \{(\delta(q, \sigma), \delta(p, \sigma), r)\}$ for each $(q, p, r) \in Q^3$ and $\sigma \in \Sigma$,

and the set of accepting states being $F' = \{(r, q_f, r) : r \in Q,\ q_f \in F\}$. Along the similar lines as for the proof above we can prove that $B = L(H)$.