Tutorial 2

Exercise 1: Write regular expressions for the following languages:

a) The language $\{ab, ba, abb, bab, abbb, babb\}$
b) The language over alphabet $\{a, b, c\}$ containing exactly those words that contain subword abb.
c) The language over alphabet $\{a, b, c\}$ containing exactly those words that start with prefix bca or end with suffix $ccab$.
d) The language $\{w \in \{0, 1\}^* \mid |w|_0 \bmod 2 = 0\}$.
e) The language $\{w \in \{0, 1\}^* \mid |w|_0 \bmod 3 = 1\}$.
f) The language $\{w \in \{0, 1\}^* \mid w$ contains subwords 010 and $111\}$
g) The language $\{w \in \{a, b\}^* \mid w$ contains subword bab or $|w|_b \leq 3\}$
h) The language $\{w \in \{a, b\}^* \mid w$ contains subword bab and $|w|_b \leq 3\}$
i) The language of all words over $\{a, b, c\}$ that contain no two consecutive a's.

Exercise 2: Let us have two languages L_1 and L_2 described by the regular expressions

$$L_1 = \mathcal{L}(\, 0^*1^*0^*1^*0^* \,), \qquad L_2 = \mathcal{L}(\, (01 + 10)^* \,).$$

a) What are the shortest and the longest words in the intersection $L_1 \cap L_2$?

b) Why none of the languages L_1 and L_2 is a subset of the other?

c) What is the shortest word that does not belong to the union $L_1 \cup L_2$? Is it unambiguous?

Exercise 3: Let us say that we would like to devise a syntax for representation of simple arithmetic expressions by words over alphabet

$$\Sigma = \{\mathtt{A}, \mathtt{B}, \ldots, \mathtt{Z}, \mathtt{a}, \mathtt{b}, \ldots, \mathtt{z}, \mathtt{0}, \mathtt{1}, \ldots, \mathtt{9}, \mathtt{.}, \mathtt{+}, \mathtt{-}, \mathtt{*}, \mathtt{/}, \mathtt{(}, \mathtt{)}\}.$$

a) Propose how identifiers will look like, and deribe them using a regular expression.

b) Propose how number constants will look like, and describe them using a regular expression.

   Remark: Allow the number constants that would represent integers, e.g., `129` or `0`, and also floating-point number constants, e.g., `3.14`, `-1e10`, or `4.2E-23`. Consider also the possibility of representing number constants in other number systems except the decimal number system (e.g., hexadecimal, octal, binary).

Exercise 4: For each of the following languages, construct a DFA accepting the given language. Represent the constructed automata by graphs and tables.

a) $L_1 = \{w \in \{a, b\}^* \mid w = a\}$

b) $L_2 = \{b, ab\}$

c) $L_3 = \{w \in \{a, b\}^* \mid \exists n \in \mathbb{N} : w = a^n\}$

d) $L_4 = \{w \in \{a, b, c\}^* \mid |w|_a \geq 1\}$

e) $L_5 = \{w \in \{0, 1\}^* \mid w \text{ contains subword } 011\}$

f) $L_6 = \{w \in \{a, b, c\}^* \mid |w| > 0 \ \wedge \ |w|_a = 0\}$

g) $L_7 = \{w \in \{a, b\}^* \mid |w| \geq 2 \text{ and the last two symbols of } w \text{ are not the same}\}$

h) $L_8 = \{w \in \{a, b\}^* \mid |w|_a \bmod 3 = 1\}$

Exercise 5: Construct DFA accepting words beginning with $abaab$, ending with $abaab$, and containing $abaab$, i.e., construct deterministic finite automata accepting the following three languages:

a) $L_1 = \{abaabw \mid w \in \{a, b\}^*\}$

b) $L_2 = \{wabaab \mid w \in \{a, b\}^*\}$

c) $L_3 = \{w_1abaabw_2 \mid w_1, w_2 \in \{a, b\}^*\}$

Exercise 6: Describe how to find out for a given DFA $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ if:

a) $\mathcal{L}(\mathcal{A}) = \emptyset$

b) $\mathcal{L}(\mathcal{A}) = \Sigma^*$

Exercise 7: Construct DFA $\mathcal{A}_1, \mathcal{A}_2$ such that:

$$\mathcal{L}(\mathcal{A}_1) = \{w \in \{a, b\}^* \mid |w|_a \bmod 2 = 0\}$$
$$\mathcal{L}(\mathcal{A}_2) = \{w \in \{a, b\}^* \mid \text{every occurence of symbol } b \text{ in } w \text{ is followed with symbol } a\}$$

Using automata $\mathcal{A}_1, \mathcal{A}_2$, construct DFA accepting the following languages:

a) $L_1 = \{w \in \{a, b\}^* \mid |w|_a \bmod 2 = 0$ and every occurence of symbol b in w is followed with symbol $a\}$

b) $L_2 = \{w \in \{a, b\}^* \mid |w|_a \bmod 2 = 0$ or every occurrence of symbol b in w is followed with symbol $a\}$

c) $L_3 = \{w \in \{a, b\}^* \mid$ some occurrence of symbol b in w is not followed with symbol $a\}$

d) $L_4 = \{w \in \{a, b\}^* \mid |w|_a \bmod 2 = 0$ and some occurrence of symbol b in w is not followed with symbol $a\}$

e) $L_5 = \{w \in \{a, b\}^* \mid \text{if } |w|_a \bmod 2 = 0 \text{ then every occurrence of symbol } b \text{ in } w \text{ is followed with symbol } a\}$

f) $L_6 = \{w \in \{a, b\}^* \mid |w|_a \bmod 2 = 0 \text{ iff every occurrence of symbol } b \text{ in } w \text{ is followed with symbol } a\}$

Exercise 8: For each of the following languages, construct a DFA accepting the given language. Represent the constructed automata by graphs and tables.

a) $L_1 = \{w \in \{a, b\}^* \mid |w| \geq 4 \text{ and the second, third, and fourth symbol of } w \text{ are the same}\}$

b) $L_2 = \{w \in \{a, b\}^* \mid |w| \geq 4 \text{ and the third symbol and the last symbol of } w \text{ are the same}\}$

c) $L_3 = \{\, w \in \{a, b, c, d\}^* \mid w \text{ does not start with } a\text{, the second symbol is not } b\text{, the third symbol is not } c\text{, and the fourth symbol is not } d \,\}$

Remark: This language includes also those words w where $|w| < 4$.

d) $L_4 = \{\, w \in \{a, b, c, d\}^* \mid w \text{ does not start with } a \text{ or the second symbol is not } b \text{ or the third symbol is not } c \text{ or the fourth symbol is not } d \,\}$

Exercise 9: Desribe how to find out for given DFA $\mathcal{A}_1 = (Q_1, \Sigma, \delta_1, q_1, F_1)$ and $\mathcal{A}_2 = (Q_2, \Sigma, \delta_2, q_2, F_2)$ if $\mathcal{L}(\mathcal{A}_1) = \mathcal{L}(\mathcal{A}_2)$.