Tutorial 7

Exercise 1: For each of the following languages, give an example of 5 words belonging to the language, and an example of 5 words that do not belong to the language.

a) $L_1 = \{w \in \{0,1\}^* \mid \text{the length of word } w \text{ is less than } 5\}$

Solution: Examples of words that belong to language L_1: ε, `0`, `1`, `00`, `01`, ...
Examples of words that do not belong to language L_1: `00000`, `00001`, `00010`, `000000`, `1111111`, ...

b) $L_2 = \{w \in \{a,b\}^* \mid \text{the number of occurrences of symbol b in word } w \text{ is even}\}$

Solution: Examples of words that belong to language L_2: ε, `a`, `aa`, `bb`, `aaa`, `abb`, ...
Examples of words that do not belong to language L_2: `b`, `ab`, `ba`, `aab`, `aba`, ...

c) $L_3 = \{w \in \{0,1\}^* \mid \text{in } w \text{ is every 0 (directly) followed by 1}\}$

Solution: Examples of words that belong to language L_3: ε, `1`, `01`, `11`, `101101`, ...
Examples of words that do not belong to language L_3: `0`, `10`, `001`, `010`, `1010`, ...

d) $L_4 = \{w \in \{0,1\}^* \mid w \text{ begins and ends with the same symbol}\}$

Solution: Examples of words that belong to language L_4: `0`, `1`, `00`, `11`, `000`, `010`, ...
Examples of words that do not belong to language L_4: ε, `01`, `10`, `001`, `011`, ...

e) $L_5 = \{w \in \{a,b\}^* \mid w \text{ contains as a subword the sequence abb}\}$

Solution: Examples of words that belong to language L_5: `abb`, `aabb`, `abba`, `abbb`, `babb`, ...
Examples of words that do not belong to language L_5: ε, `a`, `b`, `aa`, `ab`, `ba`, `bb`, `aaa`, ...

Exercise 2: Let us assume $\Sigma = \{a, b\}$ and $n \in \mathbb{N}$.

a) How many words in Σ^* are of length n?

Solution: 2^n

b) How many words in Σ^* are of length at most n?

Solution:

$$2^0 + 2^1 + \cdots + 2^n = \sum_{i=0}^{n} 2^i = \frac{2^{n+1} - 1}{2 - 1} = 2^{n+1} - 1$$

Exercise 3: Consider the following languages:

$L_1 = \{w \in \{0,1\}^* \mid \text{in } w \text{ is every 0 (directly) followed by 1}\}$
$L_2 = \{w \in \{0,1\}^* \mid w = w^R\}$

a) Enumerate the first 5 words of each of languages L_1, L_2 (the smallest words with respect to order $<_L$).

Solution:

L_1: ε, 1, 01, 11, 011
L_2: ε, 0, 1, 00, 11

b) Enumerate the first 5 words of each of languages $\overline{L_1}$, $\overline{L_2}$.

Solution:

$\overline{L_1}$: 0, 00, 10, 000, 001
$\overline{L_2}$: 01, 10, 001, 011, 100

c) Enumerate the first 5 words of language $L_1 \cap L_2$.

Solution:

$L_1 \cap L_2$: ε, 1, 11, 101, 111

d) Enumerate the first 5 words of language $L_1 \cup L_2$.

Solution:

$L_1 \cup L_2$: ε, 0, 1, 00, 01

Exercise 4: Consider languages over $\{a, b\}$. Write down all the words in the concatenation of $L_1 = \{\varepsilon, abb, bba\}$ and $L_2 = \{a, b, abba\}$.

Solution: $L_1 \cdot L_2 = \{a, b, abba, abbb, abbabba, bbaa, bbab, bbaabba\}$

Exercise 5: Consider languages over the alphabet $\{0, 1\}$. Write down all words in the concatenation

$$\{0, 001, 111\} \cdot \{\varepsilon, 01, 0101\}$$

Solution: $\{0, 001, 00101, 0010101, 111, 11101, 1110101\}$

Exercise 6: Consider languages over the alphabet $\{0, 1\}$. Describe the language of all words in the iteration $\{00, 111\}^*$ and write the first 10 words of the language.

Solution: The language contains exactly those words that can be divided into sequences of zeroes of even length and sequences of ones of length divisible by three.

Exercise 7: Consider the following languages:

$L_1 = \{w \in \{0, 1\}^* \mid |w|_1 \leq 1\}$
$L_2 = \{w \in \{0, 1\}^* \mid w = w^R\}$

Describe the words in the language $L_1 \cap L_2$.

Solution: Those words with 0s only, or having just one 1 in the middle, i.e., ε, 0, 00, 000, $\ldots$, 1, 010, 00100, $\ldots$

Exercise 8: Write regular expressions for the following languages:

a) The language $\{ab, ba, abb, bab, abbb, babb\}$

Solution: $ab + ba + abb + bab + abbb + babb$ or $(ab + ba)(\varepsilon + b + bb)$

b) The language over alphabet $\{a, b, c\}$ containing exactly those words that contain subword abb.

Solution: $(a + b + c)^*abb(a + b + c)^*$

c) The language over alphabet $\{a, b, c\}$ containing exactly those words that start with prefix bca or end with suffix $ccab$.

Solution: $bca(a + b + c)^* + (a + b + c)^*ccab$

d) The language $\{w \in \{0, 1\}^* \mid |w|_0 \bmod 2 = 0\}$.

Solution: $1^*(01^*01^*)^*$

e) The language $\{w \in \{0, 1\}^* \mid |w|_0 \bmod 3 = 1\}$.

Solution: $1^*01^*(01^*01^*01^*)^*$

f) The language $\{w \in \{0, 1\}^* \mid w$ contains subwords 010 and $111\}$

Solution: $(0 + 1)^*010(0 + 1)^*111(0 + 1)^* + (0 + 1)^*111(0 + 1)^*010(0 + 1)^*$

g) The language $\{w \in \{a, b\}^* \mid w$ contains subword bab or $|w|_b \leq 3\}$

Solution: $(a + b)^*bab(a + b)^* + a^*(ba^* + \varepsilon)(ba^* + \varepsilon)(ba^* + \varepsilon)$

h) The language $\{w \in \{a, b\}^* \mid w$ contains subword bab and $|w|_b \leq 3\}$

Solution: $a^*ba^*baba^* + a^*baba^*ba^* + a^*baba^*$ or $(\varepsilon + a^*b)a^*baba^* + a^*baba^*ba^*$

i) The language of all words over $\{a, b, c\}$ that contain no two consecutive a's.

Solution: $\big((b + c + a(b + c)\big)^*(\varepsilon + a)$

Exercise 9: Let us have two languages K and L 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 $L_1 \cap L_2$?

Solution: The shortest words is ε and the longest 01100110, since the language L_2 does not contain any word where the same symbol would be repeated more than twice.

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

Solution: Because $1 \in L_1 - L_2$ and $010101 \in L_2 - L_1$.

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

Solution: `10101`, it is unambiguous.

Exercise 10: 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).