Automata Theory and Formal Languages
Class 2

Marcin Luckner, PhD
mluckner@mini.pw.edu.pl

Version 1.2
October 16, 2020

Regular expression

A regular expression over an alphabet Σ is a construct defined as follows

- $\emptyset$ is a regular expression,
- ϵ is a regular expression,
- each $a \in \Sigma$ is a regular expression,
- if r are s are regular expressions then
  - $(r + s)$ the sum of regular expressions,
  - (rs) the concatenation of regular expressions,
  - $(r)^*$ the Kleene closure of the regular expression

  are regular expressions.

Language generated by regular expression

A regular expression over an alphabet Σ generate languages:

- $\emptyset$ generates the empty language $\emptyset$
- ϵ generates the language $\{\epsilon\}$
- each letter $a \in \Sigma$ generates the language $\{a\}$
- if s and r are regular expressions then
  - $(r + s)$ generates the language $R \cup S$
  - (rs) generates the language RoS
  - $(r)^*$ generates the language R^*

Regular languages properties

The following equalities hold

1. $\emptyset + r = r + \emptyset = r$
2. $\emptyset r = r\emptyset = r$
3. $\epsilon + r = r + \epsilon$
4. $\epsilon r = r\epsilon = r$
5. $r + s = s + r$
6. $(r + s) + t = r + (s + t) = r + s + t$
7. $(rs)t = r(st) = rst$
8. $r(s + t) = rs + rt$
9. $(r + s)t = rt + st$
10. $(r^*)^* = r^*$
11. $(r^*s^*)^* = (r + s)^*$
12. $(r^* + s^*)^* = (r + s)^*$

Construction of regular expression I

Construct a regular expression that describe the following language
The language L over alphabet $\Sigma = \{a, A, 9\}$ contains words, which are invalid passwords. A valid password has at least 3 letters and at least one position of each letter from the alphabet.
$A9999aaa \in L$, $aaa9 \notin L$

Construction of regular expression II

- A valid password has at least 3 letters

$$(a + A + 9)(a + A + 9)(a + A + 9)(a + A + 9)^*$$

- A valid password has at least one position of each letter from the alphabet

$$\begin{aligned}
&(a + A + 9)^* a (a + A + 9)^* A (a + A + 9)^* 9 (a + A + 9)^* + \\
&(a + A + 9)^* a (a + A + 9)^* 9 (a + A + 9)^* A (a + A + 9)^* + \\
&(a + A + 9)^* A (a + A + 9)^* a (a + A + 9)^* 9 (a + A + 9)^* + \\
&(a + A + 9)^* A (a + A + 9)^* 9 (a + A + 9)^* a (a + A + 9)^* + \\
&(a + A + 9)^* 9 (a + A + 9)^* a (a + A + 9)^* A (a + A + 9)^* + \\
&(a + A + 9)^* 9 (a + A + 9)^* A (a + A + 9)^* a (a + A + 9)^*
\end{aligned}$$

Construction of regular expression III

- An invalid password with less than 3 letters

$$r_2 = (a + A + 9 + \epsilon)(a + A + 9 + \epsilon)$$

- An invalid password without some letters

$$r_1 = (a + A)^* + (a + 9)^* + (9 + A)^*$$

- A composition

$$(a + A)^* + (a + 9)^* + (9 + A)^*$$

because $L(r_1) \subset L(r_2)$

Construction of regular expression IV

- The first condition, a password has to include all kinds of letters can be use as a condition for passwords in practice.
- The second condition should be modified because real passwords should be longer.
- However we can define invalid password for any length n

$$(a+A)^* + (a+9)^* + (9+A)^* + \underbrace{(a+A+9+\epsilon)(a+A+9+\epsilon)\ldots(a+A+9+\epsilon)}_{n-1}$$

Assignments I

1. Construct regular expressions that generate language L and set $\Sigma^* \setminus L$
   - 1.1 The language L over alphabet $\Sigma = \{a, b\}$ without three adjacent identical letters.
     $aabbaa \in L$, $aaab \notin L$
   - 1.2 The language L over alphabet $\Sigma = \{0, 1\}$ contains words, which are a binary representation of odd numbers without useless zeros.
     $10100101 \in L$, $10100100 \notin L$

Assignments II

2. Construct a regular expression that describe the following language
   2.1 The language L over alphabet $\Sigma = \{0, 1\}$ contains words, which have at least two 1's after a pair of adjacent 0's and before the next pair of adjacent 0' or the end of word. A sequence of three 0's is considered as a single pair. $10100101 \in L$, $10100100 \notin L$