COSE215: Theory of Computation

Lecture 6 — Regular Expressions

Hakjoo Oh
2017 Spring

Regular expression

A regular expression denotes a language.
E.g., $(a + (b \cdot c))^*$ stands for:

$$\{\epsilon, a, bc, aa, abc, bca, bcbc, aaa, aabc, \ldots\}$$

Syntax

Definition (Syntax of regular expressions)

Regular expressions over alphabet Σ are constructed recursively:

1. (Basis) $\emptyset$, ϵ, and $a \in \Sigma$ are regular expressions.
2. (Induction)
   - If R_1 and R_2 are regular expressions, so are $R_1 + R_2$ and $R_1 \cdot R_2$.
   - If R is a regular expression, so are R^* and (R).

$$
\begin{array}{rcl}
R & \rightarrow & \emptyset \\
 & | & \epsilon \\
 & | & a \in \Sigma \\
 & | & R_1 + R_2 \\
 & | & R_1 \cdot R_2 \\
 & | & R^* \\
 & | & (R)
\end{array}
$$

Semantics

Definition (Semantics of regular expressions)

A regular expression R means a set of strings, denoted $L(R)$, which is defined inductively:

$$
\begin{aligned}
L(\emptyset) &= \emptyset \\
L(\epsilon) &= \{\epsilon\} \\
L(a) &= \{a\} \\
L(R_1 + R_2) &= L(R_1) \cup L(R_2) \\
L(R_1 \cdot R_2) &= L(R_1)L(R_2) \\
L(R^*) &= (L(R))^* \\
L((R)) &= L(R)
\end{aligned}
$$

Example

$$L(a^* \cdot (a + b)) =$$

Exercises

Find the languages of the regular expressions and equivalent finite automata.

- $(a + b)^*$
- $(a + b)^*(a + b)$
- $(a \cdot a)^*(b \cdot b)^*b$

Exercises

Find regular expressions for the languages:

- $L = \{w \in \{0,1\}^* \mid$ 0 and 1 alternate in $w\}$
- $L = \{w \in \{0,1\}^* \mid w$ has at least one pair of consecutive zeros$\}$
- $L = \{a^n b^m \mid n \geq 3, m$ is even$\}$
- $L = \{a^n b^m \mid (n+m)$ is even$\}$
- $L = \{a^n b^m \mid n \geq 4, m \leq 3\}$