Université de Fribourg
Universität Freiburg

Département de Mathématiques
Mathematik Departement

Mathematical Methods for Computer Science I

Fall 2016

Outline 9

2.4. Regular expression (RE).

Operations on languages which define new languages:

1) **Union:** $L_1 \cup L_2 = \{l \mid l \in L_1 \text{ or } l \in L_2\}$
2) **Concatenation:** $L_1 L_2 = \{l_1 l_2 \mid l_1 \in L_1,\, l_2 \in L_2\}$
3) **Closure** (also called star or Kleene closure) of a language L, denoted by L^* is a language formed by concatenating any number of strings from L

Definition: Regular expressions are strings of symbols and operations which define languages according to certain syntax rules.
Regular expressions can be described recursively as follows:

1) Any symbol $\varepsilon, a \in \Sigma$ is a regular expression with $L(\varepsilon) = \{\varepsilon\}, L(a) = \{a\}$
2) If R_1 and R_2 are REs, then
   (a) $R_1 + R_2$ is a RE: $L(R_1 + R_2) = L(R_1) \cup L(R_2)$
   (b) $R_1 R_2$ is a RE: $L(R_1 R_2) = L(R_1) L(R_2)$
   (c) R_1^* is a RE: $L(R_1^*) = L(R_1)^*$

Theorem. *Let R be a RE defining a language $L = L(R)$. Then there exists an ε-NFA which accepts L.*

Theorem. *If a language L is accepted by a DFA D, then there is a RE R s.t.:*

$$L = L(D) = L(R).$$

Definition: A **generalized NFA** (GNFA) has regular expressions as labels on the edges and

(1) the start state has no incoming edges and is not final;
(2) there is only one final state and it has no outgoing edges;
(3) for every pair of states p, q there is at most one edge from p to q.