regular language

Canonical name	RegularLanguage
Date of creation	2013-03-22 12:26:31
Last modified on	2013-03-22 12:26:31
Owner	mps (409)
Last modified by	mps (409)
Numerical id	21
Author	mps (409)
Entry type	Definition
Classification	msc 68Q45
Classification	msc 68Q42
Synonym	type-3 language
Synonym	type-3 grammar
Synonym	regular set
Synonym	regular event
Related topic	Language
Related topic	DeterministicFiniteAutomaton
Related topic	NonDeterministicFiniteAutomaton
Related topic	RegularExpression
Related topic	KleeneAlgebra
Related topic	ContextFreeLanguage
Related topic	KleenesTheorem
Defines	regular grammar

A *regular grammar* is a context-free grammar where a production has one of the following three forms:

$$A \to \lambda, \qquad A \to u, \qquad A \to vB$$

where A, B are non-terminal symbols, u, v are terminal words, and λ the empty word. In BNF, they are:

$$\begin{aligned}
<\!\textit{non-terminal}\!> &::= \textit{terminal} \\
<\!\textit{non-terminal}\!> &::= \textit{terminal}\, <\!\textit{non-terminal}\!> \\
<\!\textit{non-terminal}\!> &::= \lambda
\end{aligned}$$

A *regular language* (also known as a *regular set* or a *regular event*) is the set of strings generated by a regular grammar. Regular grammars are also known as Type-3 grammars in the Chomsky hierarchy.

A regular grammar can be represented by a deterministic or non-deterministic finite automaton. Such automata can serve to either generate or accept sentences in a particular regular language. Note that since the set of regular languages is a subset of context-free languages, any deterministic or non-deterministic finite automaton can be simulated by a pushdown automaton.

There is also a close relationship between regular languages and regular expressions. With every regular expression we can associate a regular language. Conversely, every regular language can be obtained from a regular expression. For example, over the alphabet $\{a, b, c\}$, the regular language associated with the regular expression $a(b \cup c)^*a$ is the set

$$\{a\} \circ \{b, c\}^* \circ \{a\} = \{awa \mid w \text{ is a word in two letters } b \text{ and } c\},$$

where $\circ$ is the concatenation operation, and $*$ is the Kleene star operation. Note that w could be the empty word λ.

Yet another way of describing a regular language is as follows: take any alphabet Σ. Let $\mathcal{R}(\Sigma)$ be the smallest subset of $P(\Sigma^*)$ (the power set of the set of words over Σ, in other words, the set of languages over Σ), among all subsets of $P(\Sigma^*)$ with the following properties:

- $\mathcal{R}(\Sigma)$ contains all sets of cardinality no more than 1 (i.e., $\varnothing$ and singletons);
- $\mathcal{R}(\Sigma)$ is closed under set-theoretic union, concatenation, and Kleene star operations.

Then L is a regular language over Σ iff $L \in \mathcal{R}(\Sigma)$.

Normal form. Every regular language can be generated by a grammar whose productions are either of the form $A \to aB$ or of the form $A \to \lambda$, where A, B are non-terminal symbols, and a is a terminal symbol. Furthermore, for every pair (A, a), there is exactly one production of the form $A \to aB$.

Remark. Closure properties on the family of regular languages are: union, intersection, complementation, set difference, concatenation, Kleene star, homomorphism, inverse homomorphism, and reversal.

References

[1] A. Salomaa *Computation and Automata, Encyclopedia of Mathematics and Its Applications, Vol. 25.* Cambridge (1985).