star height

Canonical name	StarHeight
Date of creation	2013-03-22 18:57:55
Last modified on	2013-03-22 18:57:55
Owner	CWoo (3771)
Last modified by	CWoo (3771)
Numerical id	9
Author	CWoo (3771)
Entry type	Definition
Classification	msc 20M35
Classification	msc 68Q70
Related topic	StarFree

The *star height* of a regular expression p over an alphabet Σ, denoted by $\mathrm{ht}(p)$, is defined recursively as follows:

1. $\mathrm{ht}(\varnothing) = \mathrm{ht}(a) = 0$ for any $a \in \Sigma$;
2. $\mathrm{ht}(p \cup q) = \mathrm{ht}(pq) = \max(\mathrm{ht}(p), \mathrm{ht}(q))$;
3. $\mathrm{ht}(p^*) = \mathrm{ht}(p) + 1$.

Let $\Sigma = \{a, b, c\}$. The following expressions have star height $0, 1, 2, 3$:

$$(a \cup b)c \qquad a^*b^* \qquad (a^*b)^*c \qquad ((ab^*a \cup c)^*b)^*$$

Since any regular expression p describes a regular language $L(p)$, we may extend the definition of star height to regular languages. However, since a regular language may be described by more than one regular expressions, we need to be a little careful:

The *star height* of a regular language L is the least integer i such that L may be described by a regular expression of star height i. In other words:

$$\mathrm{ht}(L) = \min\{h(p) \mid L = L(p),\, p \in R(\Sigma)\},$$

where $R(\Sigma)$ is the set of all regular expressions over Σ.

Remarks.

- Any regular language over a singleton alphabet has star height at most 1, which can be proved by the pumping lemma for regular languages.
- If the alphabet Σ consists of at least two letters, then for every positive integer n, there is a regular language whose star height is n. In fact, it can be shown that if $|\Sigma| = 2$, then for every n, there is a regular language L over Σ such that $\mathrm{ht}(L) = n$.
- It was an open question whether an algorithm exists for determining $\mathrm{ht}(L)$ for an arbitrary regular language L. In 2005, the question was resolved by D. Kirsten in the positive, and the algorithm is that of a non-deterministic finite automaton.
- We may also define star height on generalized regular expressions. For a regular language L, define $\mathrm{ht}_g(L) = \min\{h(p) \mid L = L(p),\, p \in R_g(\Sigma)\}$, where $R_g(\Sigma)$ is the set of all generalized regular expressions. It is clear

that $\mathrm{ht}_g(L) \leq \mathrm{ht}(L)$. However, it is still an open question whether, for every integer n, there is a regular language L with $\mathrm{ht}_g(L) = n$.

A star-free language has star height 0 with respect to representations by generalized regular expressions, but may be positive with respect to representations by regular expressions, for example $L = \{ab\}^*$.

References

[1] A. Salomaa, *Formal Languages*, Academic Press, New York (1973).

[2] A. Salomaa, *Jewels of Formal Language Theory*, Computer Science Press (1981).