# star-free

| | |
|---|---|
| Canonical name | Starfree |
| Date of creation | 2013-03-22 18:59:05 |
| Last modified on | 2013-03-22 18:59:05 |
| Owner | CWoo (3771) |
| Last modified by | CWoo (3771) |
| Numerical id | 16 |
| Author | CWoo (3771) |
| Entry type | Definition |
| Classification | msc 68Q42 |
| Classification | msc 68Q45 |
| Synonym | star free |
| Synonym | non-counting |
| Synonym | noncounting |
| Related topic | StarHeight |

Let $\mathscr{R}$ and $\mathscr{GR}$ be the families of languages represented by regular expressions and generalized regular expressions respectively. It is well-known that

$$\mathscr{R} = \mathscr{GR}.$$

This means that the additional symbols $\cap$ and $\neg$ in a generalized regular expressions are extraneous: removing them will not affect the representational power of the expressions with respect to regular languages.

What if we remove $^*$, the Kleene star symbol, instead? With regard to regular expressions, removing $^*$ severely limits the power of the expressions. By induction, the represented languages are all finite. In this entry, we briefly discuss what happens when $^*$ is removed from the generalized regular expressions.

**Definition**. Let $\Sigma$ be an alphabet. A language $L$ over $\Sigma$ is said to be *star-free* if it can be expressed by a generalized regular expression without $^*$. In other words, a star-free language is a language that can be obtained by applying the operations of union, concatenation, and complementation to $\varnothing$ and atomic languages (singleton subsets of $\Sigma$) a finite number of times.

If we denote $\mathscr{SF}$ the family of star-free languages (over some alphabet $\Sigma$), then $\mathscr{SF}$ is the smallest set of languages over $\Sigma$ such that

- $\varnothing \in \mathscr{SF}$,
- $\{a\} \in \mathscr{SF}$ for any $a \in \Sigma$,
- if $L, M \in \mathscr{SF}$, then $L \cup M, LM, L^c \in \mathscr{SF}$.

A shorter characterization of a star-free language is a language with star height 0 with respect to representations by generalized regular expressions.

In relations to finite and regular languages, we have the following:

$$\mathscr{F} \subseteq \mathscr{SF} \subseteq \mathscr{R}, \tag{1}$$

where $\mathscr{F}$ denotes the family of finite languages over $\Sigma$.

Furthermore, it is easy to see that $\mathscr{SF}$ is closed under Boolean operations, so that $\mathscr{SF}$ contains infinite languages, for example $\neg\varnothing$ represents $\Sigma^*$. As a result, the first inclusion must be strict. This example also shows that languages representable by expressions including the Kleene star may still be star-free. Here's another example: $\{ab\}^*$ over the alphabet $\{a, b\}$. This language can be represented as

$$\lambda \cup (ab\Sigma^* \cap \Sigma^* ab \cap \neg(\Sigma^* a^2 \Sigma^*) \cap \neg(\Sigma^* b^2 \Sigma^*))$$

The expression above, of course, is not star-free, and includes the symbol $\lambda$ representing the empty word. However, $\Sigma^*$ is just $\neg\varnothing$, and $\lambda$ is just $\Sigma^* \cap \neg(a\Sigma^*) \cap \neg(b\Sigma^*)$. Some substitutions show that $\{ab\}^*$ is indeed star-free.

Is the second inclusion strict? Are there regular languages such that representations by expressions including the Kleene star is inevitable? The following proposition answers the question:

**Proposition 1.** *A language $L$ is star-free iff there exists a non-negative integer $n$ such that, for any words $u, v, w$ over $\Sigma$, $uv^n w \in L$ iff $uv^{n+1}w \in L$.*

A language satisfying the second statement in the proposition is known as *noncounting*, so the proposition can be restated as: a language is star-free iff it is noncounting.

As a result of this fact, we see that languages such as $\{(ab)^2\}^*$ is not star-free, although it is regular. Indeed, if we pick $u = v = w = ab$ as in the statement of the proposition above, we see that $uv^n w$ is in the language iff $uv^{n+1}w$ is not in the language, for any $n \geq 0$. Therefore, the second inclusion in chain (1) above is also strict.

The above proposition also strengthens chain (1): denote by $\mathscr{T}(\infty)$ the family of locally testable languages, then

$$\mathscr{T}(\infty) \subset \mathscr{S}\mathscr{F} \subset \mathscr{R}. \tag{2}$$

The first inclusion is due to the fact that, for any $k$-testable language $L$ (over some $\Sigma$), we have

$$\mathrm{sw}_k(uv^k w) = \mathrm{sw}_k(uv^{k+1}w)$$

(the definition of $\mathrm{sw}_k(u)$ is found in the entry on locally testable languages). Note the first inclusion is also strict. For example, the language represented by $(ab)^* \cup (ba)^*$ is star-free but not locally testable.

# References


[1] A. Ginzburg, *Algebraic Theory of Automata*, Academic Press (1968).

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