



planetmath.org

Math for the people, by the people.

generalized regular expression

|                  |                              |
|------------------|------------------------------|
| Canonical name   | GeneralizedRegularExpression |
| Date of creation | 2013-03-22 19:01:34          |
| Last modified on | 2013-03-22 19:01:34          |
| Owner            | CWoo (3771)                  |
| Last modified by | CWoo (3771)                  |
| Numerical id     | 5                            |
| Author           | CWoo (3771)                  |
| Entry type       | Definition                   |
| Classification   | msc 20M35                    |
| Classification   | msc 68Q70                    |

Recall that a regular expression over an alphabet  $\Sigma$  is a finite strings of symbols that are elements of  $\Sigma$ , together with symbols  $\emptyset$ ,  $\cup$ ,  $\cdot$ ,  $*$ , as well as  $($  and  $)$ , which is put together based on a set of expression formation rules. In a generalized regular expression, additional symbols  $\cap$  and  $\neg$  are tossed in also. Formally,

**Definition.** Given an alphabet  $\Sigma$ , let  $S$  be the set  $\{\emptyset, \cup, \cdot, *, \cap, \neg, (, )\}$ , considered to be disjoint from  $\Sigma$ . Let  $X$  be the smallest subset of  $(\Sigma \cup S)^*$  containing the following:

- any regular expression is in  $X$ ,
- if  $u, v \in X$ , then  $(u \cap v), (\neg u) \in X$ .

An element of  $X$  is called a *generalized regular expression* over  $\Sigma$ .

Like regular expressions, every generalized regular expressions are designed to represent languages (it is clear that  $\cap$  and  $\neg$  are intended to mean set-theoretic intersection and complementation). If  $u$  is a generalized regular expression:

- if  $u$  is regular expression, then the language represented by  $u$  as a generalized regular expression is  $L(u)$ , the language represented by  $u$  as a regular expression;
- if  $A$  is represented by  $u$  and  $B$  is represented by  $v$ , then  $A \cap B$  is represented by  $(u \cap v)$
- if  $A$  is represented by  $u$ , then  $\Sigma^* - A$  is represented by  $(\neg u)$ .

By induction, it is easy to see that, given a generalized regular expression  $u$ , there is exactly one language represented by  $u$ . We denote  $L(u)$  the language represented by  $u$ , and  $\mathcal{GR}$  the family of languages represented by generalized regular expressions.

Since regular languages are closed under intersection and complementation, generalized regular expressions in this regard are no powerful than regular expressions. The symbols  $\neg$  and  $\cap$  are therefore extraneous. In other words,

$$\mathcal{R} = \mathcal{GR},$$

where  $\mathcal{R}$  is the family of regular languages.

## References

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