



planetmath.org

Math for the people, by the people.

equivalent regular expressions

Canonical name	EquivalentRegularExpressions
Date of creation	2013-03-22 18:57:58
Last modified on	2013-03-22 18:57:58
Owner	CWoo (3771)
Last modified by	CWoo (3771)
Numerical id	10
Author	CWoo (3771)
Entry type	Definition
Classification	msc 20M35
Classification	msc 68Q70

Let  $\Sigma$  be an alphabet, and  $E(\Sigma)$  be the set of all regular expressions over  $\Sigma$ . Two expressions  $p, q$  are said to be *equivalent*, written  $p \equiv q$ , if they describe the same language:  $L(p) = L(q)$ .

This relation is clearly an equivalence relation on  $E(\Sigma)$ , and therefore partitions  $E(\Sigma)$  into equivalence classes. Furthermore, if  $\cup, \cdot$ , and  $*$  are interpreted as operations on  $E(\Sigma)$ , then it is clear that  $\equiv$  respects each of these operations, and so is a congruence relation on  $E(\Sigma)$ .

Let  $E = E(\Sigma)/\equiv$ , the set of equivalence classes. Members of  $E$  are denoted  $[p]$ . For simplicity, we drop the square brackets around  $p$  from now on.

The following identities (in  $E$ ) are easily established: for any  $p, q, r \in E$ :

- |                                            |                               |                                    |
|--------------------------------------------|-------------------------------|------------------------------------|
| 1. $p \cup q = q \cup p$                   | 6. $p(q \cup r) = pq \cup pr$ | 11. $p\emptyset^* = p$             |
| 2. $p \cup p = p$                          | 7. $(p \cup q)r = pr \cup qr$ | 12. $pp^* = p^*p$                  |
| 3. $p \cup \emptyset = p$                  | 8. $\emptyset p = \emptyset$  |                                    |
| 4. $p \cup (q \cup r) = (p \cup q) \cup r$ | 9. $p\emptyset = \emptyset$   | 13. $p^*p \cup \emptyset^* = p^*$  |
| 5. $p(qr) = (pq)r$                         | 10. $\emptyset^*p = p$        | 14. $(p \cup \emptyset^*)^* = p^*$ |

Identities 1,3,4 establish that  $E$  is a commutative monoid with  $\cup$  as the “addition”, and  $\emptyset$  as the identity. Likewise, identities 5,10,11 establish that  $E$  is a monoid with  $\cdot$  as the “multiplication”, and  $\emptyset^*$  as the identity element. By identities 6 through 9,  $E$  with the two operations form a semiring ( $\cup$  being the addition and  $\cdot$  the multiplication). Lastly, identity 2 says that  $E$  is an idempotent semiring.

Now, as a idempotent semiring, the binary relation  $\leq$  such that  $p \leq q$  iff  $p \cup q = q$  (or  $L(p) \subseteq L(q)$ ). It is not hard to see the following implication:

$$pq \cup r \leq q \quad \text{implies} \quad p^*r \leq q. \quad (1)$$

Assume the left hand side of the implication. In other words,  $L(pq \cup r) \subseteq L(q)$ . Then  $L(p)L(q) \cup L(r) \subseteq L(q)$ , which implies that  $L(r) \subseteq L(q)$ , and  $L(p)L(q) \subseteq L(q)$ , which, by induction, implies that  $L(p)^n L(q) \subseteq L(q)$ , and hence  $L(p)^+ L(q) \subseteq L(q)$ . Now,  $L(p^*r) = L(p)^* L(r) = L(r) \cup L(p)^+ L(r) \subseteq L(q) \cup L(p)^+ L(q) \subseteq L(q)$ . Hence we arrive at the right hand side of the implication.

This implication, together with identities 12 and 13, show that  $E$ , with binary operations  $\cup, \cdot$  and the unary operation  $*$ , is a Kleene algebra.

**Remarks.**

1. If we impose the condition  $\emptyset^* \not\leq p$ , the above implication can be written as

$$pq \cup r = q \quad \text{implies} \quad p^*r = q. \quad (2)$$

Suppose  $x \in L(q) = L(pq \cup r) = L(p)L(q) \cup L(r)$ . We use induction on the length of  $x$ . If  $|x| = 0$ , then  $x \in L(r) \subseteq L(p)^*L(r)$ , since  $L(p)$  does not contain the empty word  $\lambda$ . Suppose now that  $|x| > 0$ . Then either  $x = yz$  where  $y \in L(p)$  and  $z \in L(q)$ , or  $x \in L(r)$ . In the former case, since  $y$  is not the empty word by the imposed condition,  $z$  is a strictly shorter word than  $x$ , which, by induction, is in  $L(p^*r) = L(p)^*L(r)$ . As a result,  $x = yz \in L(p)L(p)^*L(r) \subseteq L(p)^*L(r)$ . In the latter case, we have  $x \in L(p)^*L(r)$ . In either case,  $x \in L(p^*r)$ , and the implication is proved.

2. Regular expressions can be thought of as well-formed formulas in a formal system. A sentence is of the form  $p = q$  where  $p, q$  are wffs. An interpretation of the sentence  $p = q$  may be defined as the equation  $L(p) = L(q)$ . A sentence is valid if its interpretation is true. The list of identities above are all valid sentences, and can in fact be thought of as axioms of the system. There are two rules of inferences:

- (a) formal variable substitution, and
- (b) *from*  $pq \cup r = q$  *infer*  $p^*r = q$ , given that  $p \cup \emptyset^* = p$  is not valid (implication 2 above).

The system is *complete* if all valid sentences may be derived from the axioms by rules of inferences. We have the following results:

- If the set of axioms is finite, and (a) as the sole rule of inference, then the system is not complete.
- However, the system is complete if the (finite) set of axioms above, and both rules (a) and (b) are used.
- In fact, with (a) and (b), all the axioms we need are 1, 4, 5, 6, 7, 8, 10, 13, 14, and none can be removed to keep the system complete.

## References

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