planetmath.org

Math for the people, by the people.

insertion operation on languages

Canonical name	InsertionOperationOnLanguages
Date of creation	2013-03-22 18:56:53
Last modified on	2013-03-22 18:56:53
Owner	CWoo (3771)
Last modified by	CWoo (3771)
Numerical id	6
Author	CWoo (3771)
Entry type	Definition
Classification	msc 68Q70
Classification	msc 68Q45
Synonym	insertion-closed
Synonym	insertion-closure
Related topic	DeletionOperationOnLanguages
Related topic	ShuffleOfLanguages
Defines	insertion
Defines	insertion closed
Defines	insertion closure
Defines	k-insertion

Let Σ be an alphabet, and u, v words over Σ. An *insertion* of u into v is a word of the form $v_1 u v_2$, where $v = v_1 v_2$. The concatenation of two words is just a special case of insertion. Also, if w is an insertion of u into v, then v is a deletion of u from w.

The insertion of u into v is the set of all insertions of v into u, and is denoted by $v \rhd u$.

The notion of insertion can be extended to languages. Let L_1, L_2 be two languages over Σ. The insertion of L_1 into L_2, denoted by $L_1 \rhd L_2$, is the set of all insertions of words in L_1 into words in L_2. In other words,

$$L_1 \rhd L_2 = \bigcup \{u \rhd v \mid u \in L_1, v \in L_2\}.$$

So $u \rhd v = \{u\} \rhd \{v\}$.

A language L is said to be *insertion closed* if $L \rhd L \subseteq L$. Clearly Σ^* is insertion closed, and arbitrary intersection of insertion closed languages is insertion closed. Given a language L, the *insertion closure* of L, denoted by $\mathrm{Ins}(L)$, is the smallest insertion closed language containing L. It is clear that $\mathrm{Ins}(L)$ exists and is unique.

More to come...

The concept of insertion can be generalized. Instead of the insertion of u into v taking place in one location in v, the insertion can take place in several locations, provided that u must also be broken up into pieces so that each individual piece goes into each inserting location. More precisely, given a positive integer k, a *k-insertion of u into v* is a word of the form

$$v_1 u_1 \cdots v_k u_k v_{k+1}$$

where $u = u_1 \cdots u_k$ and $v = v_1 \cdots v_{k+1}$. So an insertion is just a 1-insertion. *The k-insertion* of u into v is the set of all k-insertions of u into v, and is denoted by $u \rhd^{[k]} v$. Clearly $\rhd^{[1]} = \rhd$.

Example. Let $\Sigma = \{a, b\}$, and $u = aba$, $v = bab$, and $w = bababa$. Then

- w is an insertion of u into v, as well as an insertion of v into u, for $w = vu\lambda = \lambda vu$.
- w is also a 2-insertion of u into v:
  - the decompositions $u = (ab)(a)$ and $v = (b)(ab)\lambda$
  - or the decompositions $u = \lambda u$ and $v = \lambda v \lambda$

produce $(b)(ab)(ab)(a)\lambda = \lambda\lambda vu\lambda = w$.

- Finally, w is also a 2-insertion of v into u, as the decompositions $u = \lambda u\lambda$ and $v = v\lambda$ produce $\lambda vu\lambda\lambda = w$.

- $u \rhd v = \{ababab, babaab, baabab, bababa\}$.

From this example, we observe that a k-insertion is a $(k+1)$-insertion, and every k-insertion of u into v is a $(k+1)$-insertion of v into u. As a result,

$$u \rhd^{[k]} v \subseteq (u \rhd^{[k+1]} v) \cap (v \rhd^{[k+1]} u).$$

As before, given languages L_1 and L_2, *the k-insertion* of L_1 into L_2, denoted by $L_1 \rhd^{[k]} L_2$, is the union of all $u \rhd^{[k]} v$, where $u \in L_1$ and $v \in L_2$.

Remark. Some closure properties regarding insertions: let $\mathscr{R}$ be the family of regular languages, and $\mathscr{F}$ the family of context-free languages. Then $\mathscr{R}$ is closed under $\rhd^{[k]}$, for each positive integer k. $\mathscr{F}$ is closed $\rhd^{[k]}$ only when $k = 1$. If $L_1 \in \mathscr{R}$ and $L_2 \in \mathscr{F}$, then $L_1 \rhd^{[k]} L_2$ and $L_2 \rhd^{[k]} L_1$ are both in $\mathscr{F}$. The proofs of these closure properties can be found in the reference.

References

[1] M. Ito, *Algebraic Theory of Automata and Languages*, World Scientific, Singapore (2004).