planetmath.org

Math for the people, by the people.

concatenation

Canonical name	Concatenation
Date of creation	2013-03-22 17:16:38
Last modified on	2013-03-22 17:16:38
Owner	CWoo (3771)
Last modified by	CWoo (3771)
Numerical id	17
Author	CWoo (3771)
Entry type	Definition
Classification	msc 20M35
Classification	msc 68Q70
Synonym	juxtaposition
Synonym	monoidal
Related topic	Word
Defines	length
Defines	length of a word
Defines	monoidal language

Concatenation on Words

Let a, b be two words. Loosely speaking, the *concatenation*, or *juxtaposition* of a and b is the word of the form ab. In order to define this rigorously, let us first do a little review of what words are.

Let Σ be a set whose elements we call *letters* (we also call Σ an *alphabet*). A (finite) word or a string on Σ is a partial function $w : \mathbb{N} \to \Sigma$, (where $\mathbb{N}$ is the set of natural numbers), such that, if $\operatorname{dom}(w) \neq \varnothing$, then there is an $n \in \mathbb{N}$ such that

$$w \text{ is } \begin{cases} \text{defined for every } m \leq n, \\ \text{undefined otherwise.} \end{cases}$$

This n is necessarily unique, and is called the *length* of the word w. The length of a word w is usually denoted by $|w|$. The word whose domain is $\varnothing$, the empty set, is called the empty word, and is denoted by λ. It is easy to see that $|\lambda| = 0$. Any element in the range of w has the form $w(i)$, but it is more commonly written w_i. If a word w is not the empty word, then we may write it as $w_1 w_2 \cdots w_n$, where $n = |w|$. The collection of all words on Σ is denoted Σ^* (the asterisk * is commonly known as the Kleene star operation of a set). Using the definition above, we see that $\lambda \in \Sigma^*$.

Now we define a binary operation $\circ$ on Σ^*, called the *concatenation* on the alphabet Σ, as follows: let $v, w \in \Sigma^*$ with $m = |v|$ and $n = |w|$. Then $\circ(v, w)$ is the partial function whose domain is the set $\{1, \ldots, m, m+1, \ldots, m+n\}$, such that

$$\circ(v, w)(i) = \begin{cases} v(i) & \text{if } i \leq m \\ w(i-m) & \text{otherwise.} \end{cases}$$

The partial function $\circ(v, w)$ is written $v \circ w$, or simply vw, when it does not cause any confusion. Therefore, if $v = v_1 \cdots v_m$ and $w = w_1 \cdots w_n$, then $vw = v_1 \cdots v_m w_1 \cdots w_n$.

Below are some simple properties of $\circ$ on words:

- $\circ$ is associative: $(uv)w = u(vw)$.
- $\lambda w = w \lambda = w$.
- As a result, Σ^* together with $\circ$ is a monoid.
- $vw = \lambda$ iff $v = w = \lambda$.
- As a result, Σ^* is never a group unless $\Sigma^* = \{\lambda\}$.

- If $a = bc$ where a is a letter, then one of b, c is a, and the other the empty word λ.
- If $ab = cd$ where a, b, c, d are letters, then $a = c$ and $b = d$.

Concatenation on Languages

The concatenation operation $\circ$ over an alphabet Σ can be extended to operations on languages over Σ. Suppose A, B are two languages over Σ, we define

$$A \circ B := \{u \circ v \mid u \in A, v \in B\}.$$

When there is no confusion, we write AB for $A \circ B$.

Below are some simple properties of $\circ$ on languages:

- $(AB)C = A(BC)$; `http://planetmath.org/Iei`.e., concatenation of sets of letters is associative.
- Because of the associativity of $\circ$, we can inductively define A^n for any positive integer n, as $A^1 = A$, and $A^{n+1} = A^n A$.
- It is not hard to see that $\Sigma^* = \{\lambda\} \cup \Sigma \cup \Sigma^2 \cup \cdots \cup \Sigma^n \cup \cdots$.

Remark. A formal language containing the empty word, and is closed under concatenation is said to be *monoidal*, since it has the structure of a monoid.

References

[1] H.R. Lewis, C.H. Papadimitriou *Elements of the Theory of Computation.* Prentice-Hall, Englewood Cliffs, New Jersey (1981).