planetmath.org

Math for the people, by the people.

word

Canonical name	Word
Date of creation	2013-03-22 16:04:44
Last modified on	2013-03-22 16:04:44
Owner	juanman (12619)
Last modified by	juanman (12619)
Numerical id	31
Author	juanman (12619)
Entry type	Definition
Classification	msc 03B10
Classification	msc 03B05
Classification	msc 03B65
Classification	msc 03B99
Classification	msc 03D40
Classification	msc 08A99
Classification	msc 20A05
Classification	msc 20-00
Synonym	well-formed formula
Synonym	wff
Synonym	infinite word
Related topic	Group
Related topic	StraightLineProgram
Related topic	Alphabet
Related topic	Language
Related topic	Concatenation
Related topic	AlternativeTreatmentOfConcatenation
Related topic	FreeSemigroup
Defines	empty word
Defines	well formed formula
Defines	subword
Defines	occur in
Defines	occurrence
Defines	ω-word

Given a set Σ, a *word* (or a string) over Σ is a juxtaposition (variously called concatenation or multiplication) of a finite number of elements in Σ. The juxtaposition is taken as an associative binary operation on Σ. A word with zero number of elements is called an *empty word*, typically denoted by λ or ϵ. The set of words over Σ is denoted Σ^*.

Examples.

1. If $\Sigma = \{a, b, c, \ldots, x, y, z\}$, the English alphabet written in the lower case, then "good", "mathematics", "fasluiwh" are all words (without the double quotes) over Σ, where as "PlanetMath" is not, because it contains upper case letters, which are not in Σ.

2. Let $\Sigma = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, +, =\}$. Then "12", "0345", "9 + 3", "87 = 123", "++231++", "6 + 7 = 13", "7 =" are also words over Σ.

3. The notion of words is used extensively in group theory. The juxtaposition here is the group multiplication, as the multiplication is associative. In other words, if $g_1, g_2, ..., g_m$ are elements in G then we can form the word $w = g_1 g_2 \cdots g_m \in G$. For example, in the free group $\langle a, b | \ \rangle$ a word could be the commutator $[a, b] = aba^{-1}b^{-1}$.

Remarks

- Σ^* is a monoid with juxtaposition as the monoid multiplication, and λ, the empty word, as the multiplicative identity.

- For any element a in Σ, define $a^0 = \lambda$, and $a^{n+1} = a^n a$ for non-negative integers n. Then $a^{n+m} = a^n a^m$ since juxtaposition is associative.

- Words, by definition, are finite in length. This notion can be generalized: an *infinite word*, or more precisely, a ω-*word*, over an alphabet Σ is just a function from $\mathbb{N}$ to Σ. The set of all words over Σ, finite or infinite, is $\Sigma^* \cup \Sigma^{\mathbb{N}}$, and is denoted by Σ^∞ or Σ^ω.

Subwords

A word u is called a *subword* of v if $v = xuy$, for some words x and y (may be empty words). If u is a subword of v, we also say that u *occurs in* v, or that v *contains* u. For example, "math" is a subword of "mathematics".

Given the equation $v = xuy$, we call the triple (x, u, y) an *occurrence* of u in v. The collection of occurrences of u in v is denoted $O(u, v)$. The *number*

of occurrences of u in v defined as the cardinality of $O(u,v)$, and written $|u|_v$. The *position of occurrence* (x, u, y) of u in v is the length of x plus 1.

For example, the number of occurrences of subword a^3 in a^3ba^5c is 4, since

$$O(a^3, a^3ba^5c) = \{(\lambda, a^3, ba^5c), (a^3b, a^3, a^2c), (a^3ba, a^3, ac), (a^3ba^2, a^3, c)\}.$$

The positions of these occurrences are $1, 5, 6$, and 7, respectively.

Generating Words using Rules

Some of the words in the second example above, such as "$+ + 231 + +$" and "$7 =$", do not make any mathematical sense. The way to define words that make sense is through a process called definition by recursion. First, we declare that certain words over Σ are sensible. Then, we have a set of rules or a grammar that dictates how new sensible words can be formed from the old ones. Any word that can be formed from the old words by these rules in a finite number of steps is called sensible.

In the last example, we could declare that all symbols $0, 1, \ldots, 9$ are sensible words. To form new sensible words, we have the rules:

1. if a, b do not contain either $+$ or $=$, then ab is a sensible word;

2. if a two sensible words a, b do not contain the symbol $=$, then $a + b$ and $a = b$ are sensible words;

3. the only sensible words are the initially declared sensible words and those that can be formed by the previous two rules.

It is not hard to see based on the initially declared sensible words and the rules has one of the forms

- a

- $a_1 + a_2 + \cdots + a_n$

- $a_1 + a_2 + \cdots + a_n = b_1 + b_2 + \cdots + b_m$.

where a, a_i, b_j are words without any occurrence of $+$ and $=$, over Σ. As a result, we see that all words in the previous example are sensible (whether they are right or wrong), except "$++231++$" and "$7=$", since they are not in any one of the forms specified above. Note that the third rule above ensures

2

that "++231++" and "7=" are not sensible. Without it, we would be unable to say for sure if these words are sensible or not.

Generally, any collection of words is called a *language*. The collection of all sensible words described above is called *the language generated by* $0, \ldots, 9$ under the rules above. In logic, one calls these sensible words *well-formed formulas*, or *formulas* or *wff* for short.