



planetmath.org

Math for the people, by the people.

context-free language

Canonical name	ContextfreeLanguage
Date of creation	2013-03-22 12:26:28
Last modified on	2013-03-22 12:26:28
Owner	CWoo (3771)
Last modified by	CWoo (3771)
Numerical id	33
Author	CWoo (3771)
Entry type	Definition
Classification	msc 68Q45
Classification	msc 68Q42
Synonym	type-2 language
Synonym	type-2 grammar
Related topic	Language
Related topic	RegularLanguage
Related topic	DeterministicFiniteAutomaton
Related topic	NonDeterministicFiniteAutomaton
Related topic	PushdownAutomaton
Related topic	ContextSensitiveLanguage
Related topic	DeterministicPushdownAutomaton
Defines	context-free
Defines	context-free grammar

A *context-free language* is a language over some alphabet that can be generated from a particular kind of a known as a *context-free grammar*.

Formally, a *context-free grammar* is a formal grammar $G = (\Sigma, N, P, \sigma)$ whose productions in P have the form

$$X \rightarrow w,$$

where $X \in N$ and $w \in \Sigma^*$. A *context-free language* is the formal language $L(G)$ generated by a context-free grammar G . In other words,

$$L(G) := \{v \in T^* \mid \sigma \xrightarrow{*} v\},$$

where $T := \Sigma - N$ is the set of terminals, and $\sigma \xrightarrow{*} v$ simply means that the productions in P are applied a finite number of times to the starting symbol σ so that v is reached in the end.

This formal definition is not initially very enlightening. As an example, consider typical infix arithmetic notation. An expression typically consists of one or more terms, joined together by operators with varying precedence. Parentheses can be used to override operator precedence. This syntax is a context-free language. Let us describe such expressions with the operators $+$ and $-$ with lowest precedence, and $*$ and $/$ with highest precedence, where all operators are left-associative and the only other allowed symbols are integer literals and parentheses.

$$\begin{aligned} T &:= \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, (,), +, -, *, /\} \\ N &:= \{S, A, B, C, D\} \\ P &:= \{(S, A), (S, S+A), (S, S-A), \\ &\quad (A, B), (A, A*B), (A, A/B), \\ &\quad (B, C), (B, (S)), (C, D), (C, CD), \\ &\quad (D, 0), (D, 1), (D, 2), (D, 3), (D, 4), (D, 5), (D, 6), (D, 7), (D, 8), (D, 9)\} \\ \sigma &:= S \end{aligned}$$

It can be shown that the languages of the classical propositional logic and predicate logic are both context-free.

Context-free grammars are also known as Type-2 grammars in the Chomsky hierarchy. The Chomsky hierarchy specifies Type-2 grammars as consisting only of production rules of the form $A \rightarrow \gamma$, where A is a non-terminal and γ is a string of terminals and non-terminals.

Remarks.

1. A context-free grammar can be represented by a pushdown automaton. The automaton serves both as an *acceptor* for the language (that is, it can decide whether or not any arbitrary sentence is in the language) and as a *generator* for the language (that is, it can generate any finite sentence in the language in finite time).
2. The syntaxes of most programming languages are context-free grammars (or very close to it). A very useful subset of context-free languages are regular languages.
3. A context-free grammar G is context-sensitive if none of its productions has the form $X \rightarrow \lambda$, where λ is the empty word.
4. The family of context-free languages has the following closure properties: union, concatenation, Kleene star, homomorphism, inverse homomorphism, reversal, and intersection with a regular language. However, it is not closed under intersection, set difference, and complementation.

References

- [1] S. Ginsburg *The Mathematical Theory of Context-Free Languages*. McGraw-Hill, New York (1966).