Journal of Automata, Languages and Combinatorics 9 (2004) 2/3, 333–345

CONTEXT DERIVATION SETS AND CONTEXT-FREE NORMAL FORMS [1]

CHARLOTTE MILLER
Department of Computer Science, The University of Western Ontario
London, Ontario, Canada
e-mail: cmiller@csd.uwo.ca

ABSTRACT

A set of transformations is presented that will convert an arbitrary context-free grammar to six of the normal forms for which the right hand side of any production has at most two occurrences of nonterminal symbols. These transformations form the basis of a meta-normal form algorithm for context-free grammars. The algorithm takes as input an arbitrary context-free grammar and a target normal form, expressed as an extended two-symbol grammar form, and converts the grammar to that normal form. The number of nonterminals and productions in the output grammars of each of the base transformations is minimal.

Keywords: Normal form, grammar form, context-free grammar, Greibach normal form, Nerode equivalence, descriptional complexity

In formal language theory, a normal form is a restriction of the finite description (grammar, automaton, expression) of a language in such a way that the language itself is not altered. The use of normal forms permits some formal language properties to be more easily proved. On a more practical note, a context-free language defined in terms of a context-free grammar in normal form simplifies language recognition and parsing, both in natural language processing and computer program compilation. In this work, the basis of a meta-normal form algorithm for context-free grammars is presented, that is, of an algorithm to transform an arbitrary context-free grammar into any normal form, expressed as a two-symbol grammar form. In each step of the algorithm, a single transformation to one of six base normal forms is applied. This transformation is essentially the same for each base form, changing only in certain details according to the normal form being targeted. The number of nonterminals in the output grammar of this transformation is minimal.

[1] Full version of a submission presented at the 4th Workshop on *Descriptional Complexity of Automata, Grammars and Related Structures* (London, Ontario, Canada, August 21–24, 2002).