Testing Attribute Grammars for Circularity

Kari-Jouko Rääihä \(^1\) and Mikko Saarinen \(^2\)

\(^1\) Department of Computer Science, University of Helsinki, Tukholmankatu 2, SF-00250 Helsinki 25, Finland
\(^2\) Technical Research Center of Finland, Computing Services, Vuorimiehentie 5, SF-02150 Espoo 15, Finland

Summary. The problem of deciding whether a given attribute grammar is noncircular is known to require exponential time for infinitely many grammars. Here the time requirement of a simple circularity test is analyzed. It is shown that the reason for the exponential time requirement is the number of graphs in a collection formed for every nonterminal. By practical experiments it is argued that for real grammars the number is very small. Therefore it is feasible to actually perform the circularity test in practice. Different techniques to improve the implementation of the circularity test are discussed, too.

1. Introduction

It has been shown that the time complexity of the circularity test for attribute grammars \([8]\) is inherently exponential \([4, 5]\). The lower bound obtained is \(2^{cn^{1/2}}\), where \(n\) is the size of the description of the grammar. Hence it can be concluded that for a broad class of attribute grammars the circularity test is intractable.

The "size" of an attribute grammar is a rather abstract measure and does not show the real reason for the exponentiality. In this paper we analyze the simple circularity test from \([9]\) to find the factor which makes the test exponential.

It is shown that the only exponential factor in the complexity of the algorithm is the number of graphs in a collection constructed for each nonterminal. We say that the grammar belongs to class \(G(k)\) if the circularity test can be carried out with no more than \(k\) graphs in any collection at one time.

The \(G(k)\) grammars form a natural extension of the so-called absolutely noncircular attribute grammars of Kennedy and Warren \([7]\). The \(G(1)\) grammars are (almost) the same as the absolutely noncircular grammars. The "erroneous" circularity test in \([8]\) works exactly for the class of \(G(1)\) grammars.
Because of the exponentiality result [4], we know that there are infinitely many grammars which do not belong to any $G(k)$ class where $k$ is polynomial in the size of the grammar. The main purpose of this paper is to show that such grammars do not appear in practice. In particular, we have tested the wide variety of grammars developed for the compiler writing system HLP [13]. All of these grammars are $G(4)$. Thus we conjecture that for practical grammars the circularity test is feasible, after all!

We also discuss different ways to perform the circularity test, and analyze the implications of implementation decisions with practical grammars.

2. Notation

An attribute grammar describes the meaning of strings that may be produced by the underlying context-free grammar. This is accomplished by associating a fixed set of attributes with each nonterminal and a fixed set of semantic rules with each production. The set of attributes associated with nonterminal $X$ is denoted by $A(X)$. It is partitioned into two disjoint subsets: the inherited and the synthesized attributes of $X$. These sets are denoted by $I(X)$ and $S(X)$, respectively.

Production $p$ is denoted by

$$X_{p_0} \rightarrow X_{p_1} X_{p_2} \ldots X_{p_{n_p}}.$$ 

The semantic rules associated with production $p$ are functions defining the values of attributes $s \in S(X_{p_0})$ and $i \in \bigcup_{j=1}^{n_p} I(X_{p_j})$ on the values of attributes $a \in \bigcup_{j=0}^{n_p} A(X_{p_j})$. An attribute $a$ associated with $X_{p_j}$ is denoted by $a_j$ and called an occurrence of $a$ within $p$.

An attribute grammar defines the semantics of a specific string with the help of its parse tree in the following way. If production $p$ is applied at some node $n$, the values for the attributes $s \in S(X_{p_0})$ and $i \in \bigcup_{j=1}^{n_p} I(X_{p_j})$ are defined according to the corresponding semantic rules that are associated with production $p$. The values for $a \in A(X_{p_0})$ are attached to node $n$ and the values for $a \in A(X_{p_j})$, $1 \leq j \leq n_p$, to the $j$:th son (from the left) of node $n$. The process of evaluation is repeated until no more attribute values can be defined. With respect to the order of evaluation of semantic rules the only guidance is that the evaluation of a given rule can take place only after the values of its argument attributes have already been evaluated.

We say that an attribute grammar is noncircular if and only if at each node in every parse tree conceivable by the underlying context-free grammar all the attribute values attached to that node can be defined. Otherwise the grammar is said to be circular.