A Reduction of Scattered Context Grammars

Ing. Radovan Klembara (iklembara@fit.vut.cz)

Supervisor: prof. RNDr. Alexander Meduna CSc.

Extended Abstract

This presentation explains how to reduce the size of scattered context grammars concerning the number of both non-context-free productions and nonterminals. It proves that every recursively enumerable language is generated by a six-nonterminal scattered context grammar with a single non-context-free production.

It is known that the scattered context grammars with a single non-context-free production characterize the family of recursively enumerable languages [2]. This presentation shows that this can be improved by not only reducing a number of non-contex-free productions but also by reducing the number of nonterminals. The resulting scattered grammar $G = (N, T, P, S)$ contains exactly one non-context-free production of the form $(1, 2, 0, 3, 0, 2, 1) \rightarrow (2, \epsilon, \epsilon, \epsilon, \epsilon, \epsilon, 2)$ and six nonterminals. This demonstration is based on the special case of queue grammars called left-extended queue grammars (see [1]). Presentation proves reduction via an algorithm that can create a reduced scattered context grammar from left-extended queue grammar.

To understand the algorithm this presentation must first establish the first normal form of left-extended queue grammar. This form requires that grammar doesn't rewrite its terminals. It is shown in a presentation that it is possible to transform any left-extended queue grammar to this first normal form. The idea behind this transformation is that new grammar will use sepecial states and two-component symbols to generate only nonterminals and after the start of the resulting string is reached it will be only rewriting nonterminals. Next, the second normal form is introduced. This form requests from any left-extended queue grammar in a first normal form that $(a, q, x, p) \in R \wedge (a', q, x', p') \in R \Rightarrow a = a'$. The transformation from the first normal form to the second normal form makes use of creating new states for each state in Q and for each possible nonterminal that can be changed in that state. New states are pairs $< q, a >$, where q is the original state and a is a nonterminal to be rewritten.

The main conversion algorithm to be presented uses two codings generated by context-free productions of G that handle any context dependency. These codings use special nonterminals and states of left-extended queue grammar simulated by scattered grammar. Nonterminals are $\{0, 1, 2, 3\}$, and by their succesful removal from resulting string it is possible to check whether generated strings are correct. The codings derive string uxv from its start symbol where u and v are the codes.

Recall that the conversion algorithm's input is left-extended queue grammar, this means it records the prefix of all symbols rewritten and all states through which it passes through on the left side. On the right side before symbols it generates the states from which it exits while at the end there are states it enters. To verify symulation algorithm takes advantage of these records which are coded by two introduced codings. The simulation consists of two phases. The first phase generates the resulting string and records each production. The second phase mainly uses only one non-context-free production to check whether a resulting string is correct. This can be verified if and only if coded recorded history on the left side of string is equal to coded reversed right side of the string. Between these two phases it is essential to add central nonterminals from which the second phase can start, for this purpose it is essential to add two new nonterminals $\{4, 5\}$, which ensure this step.

The presentation concludes with construction of simulating scattered context grammar for a simple left-extended queue grammar.

References

[1] Kleijn, H. C. M. a Rozenberg, G. On the generative power of regular pattern grammars. *Acta Informatica*. 1983, roč. 20, č. 4. S. 391–411. Dostupné na: https://doi.org/10.1007/BF00264281.

[2] Křivka Zbyněk, M. A. Scattered context grammars with one non-context-free produc- tion are computationally complete. *Acta Informatica*.