Journal of Automata, Languages and Combinatorics **17** (2012) 2–4, 123–143

DESCRIPTIONAL COMPLEXITY OF UNION AND STAR ON CONTEXT-FREE LANGUAGES

Jürgen Dassow and Ronny Harbich
Fakultät für Informatik, Otto-von-Guericke-Universität Magdeburg
PSF 4120, 39016 Magdeburg, Germany
e-mail: `dassow@iws.cs.uni-magdeburg.de, ronny@raubvogel.net`

ABSTRACT

We investigate context-free languages with respect to the measures Prod and Symb of descriptional complexity, which give the minimal number of productions and the minimal total number of symbols in productions, respectively, necessary to generate the language by context-free grammars. In particular, we consider the behaviour of these measures with respect to operations. For a measure $K \in \{\mathrm{Prod}, \mathrm{Symb}\}$, given natural numbers $c_1, c_2, \ldots, c_n$ and an n-ary operation τ on languages, we discuss the set $g_{\tau,K}(c_1, c_2, \ldots, c_n)$ which is the range of $K(\tau(L_1, L_2, \ldots, L_n))$ where, for $1 \leq i \leq n$, L_i is a context-free language with $K(L_i) = c_i$. The operations under discussion are union and Kleene closure.

Keywords: production complexity, symbol complexity, context-free languages, union, Kleene star

1. Introduction

One interesting question which has been investigated very intensively over the last 20 years concerns the behaviour of syntactic complexity under operations. More precisely, for a language family $\mathcal{L}$, a measure K of descriptional complexity, an n-ary operation τ on languages under which $\mathcal{L}$ is closed, and natural numbers $m_1, m_2, \ldots, m_n$, one tries to determine the set $g_{\tau,K}(m_1, m_2, \ldots, m_n)$ of all values $K(\tau(L_1, L_2, \ldots, L_n))$, where L_i is a language of $\mathcal{L}$ with $K(L_i) = m_i$ for $1 \leq i \leq n$, and the number $f_{\tau,K}(m_1, m_2, \ldots, m_n)$ which is given by the maximal number in $g_{\tau,K}(m_1, m_2, \ldots, m_n)$.

Most of the papers study the function $f_{\tau,sc}$ where sc is the state complexity of regular languages (i. e., $sc(L)$ is given by the number of states of a minimal deterministic automaton accepting L) and τ is union, intersection, concatenation, complementation, Kleene star, reversal, or a combination of these operations. For example, for union and concatenation, one has $f_{\cup,sc}(m, n) = mn$ and $f_{\cdot,sc}(m, n) = (2m - 1)2^{n-1}$. For further results, we refer to [1, 17, 13, 12] and the summarizing articles [15, 16].

For some operations, the number $f_{\tau,nsc}$ where nsc is the state complexity with respect to nondeterministic automata is studied in [5] and [11], and research on $f_{\tau,tr}$ where tr is the complexity measure given by the number of transitions is done in [4].