Journal of Automata, Languages and Combinatorics **26** (2021) 1–2, 17 – 53
© Institut für Informatik · Justus-Liebig-Universität Giessen

THE PREFIX AUTOMATON

Sabine Broda[A] Eva Maia[B] Nelma Moreira[A] Rogério Reis[A]

[A] *CMUP& DCC, Faculdade de Ciências da Universidade do Porto*
Rua do Campo Alegre, 4169-007 Porto, Portugal
{sabine.broda,nelma.moreira,rogerio.reis}@fc.up.pt

[B] *GECAD – Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, ISEP*
Rua Dr. António Bernardino de Almeida, 431, 4249-015 Porto, Portugal
egm@isep.ipp.pt

ABSTRACT

There are many different constructions when converting regular expressions to finite automata. In this paper we focus on the prefix automaton, $\mathcal{A}_{\mathrm{Pre}}$, introduced by Yamamoto in 2014. We present two different methods for the construction of $\mathcal{A}_{\mathrm{Pre}}$. First, an inductive one, based on a system of expression equations. A second one using an iterative function for computing the states and transitions. We establish relationships between $\mathcal{A}_{\mathrm{Pre}}$ and other constructions, such as the position automaton, partial derivative automaton and their double reversal (dual) counterparts. We study the average size of these constructions, both experimentally and from an analytic combinatorics point of view. Finally, we extend the construction of the prefix automaton to regular expressions with intersection and show that the relationships with the other automaton constructions also hold for these expressions.

Keywords: regular expressions, nondeterministic finite automata, prefix automata, average complexity, regular expressions with intersection

1. Introduction

Conversions from regular expressions to equivalent nondeterministic finite automata can be with or without spontaneuos (ε) transitions. The classic construction with ε transitions is the Thompson construction ($\mathcal{A}_{\varepsilon\text{-T}}$) [20], while the Glushkov/position automaton is a standard ε-free construction ($\mathcal{A}_{\mathrm{POS}}$) [14]. It is well known that if ε transitions are eliminated from the Thompson automaton, the result is the Glushkov automaton [13]. In 2014, Yamamoto [21] presented a new construction of an ε-free automaton starting from the Thompson automaton. For that, each state s of $\mathcal{A}_{\varepsilon\text{-T}}$ was labelled with two regular expressions, one corresponding to the left language of the state, $\mathsf{LP}(s)$, and the other to its right language, $\mathsf{LS}(s)$. Merging states with the

[A] Work partially supported by CMUP (UIDB/00144/2020), which is funded by FCT- Fundação para Ciência e Tecnologia, I.P.