

MULTI-BAR AND MULTI-TILDE REGULAR OPERATORS

PASCAL CARON

*LITIS, Université de Rouen,
76801 Saint-Étienne du Rouvray Cedex, France*
e-mail: Pascal.Caron@univ-rouen.fr

JEAN-MARC CHAMPARNAUD

*LITIS, Université de Rouen,
76801 Saint-Étienne du Rouvray Cedex, France*
e-mail: Jean-Marc.Champarnaud@univ-rouen.fr

and

LUDOVIC MIGNOT

*LITIS, Université de Rouen,
76801 Saint-Étienne du Rouvray Cedex, France*
e-mail: Ludovic.Mignot@univ-rouen.fr

ABSTRACT

The aim of this paper is to introduce two families of regular operators that fit with the Glushkov automaton construction. The multi-bar (resp. multi-tilde) operators are based on a simple mechanism: eliminating (resp. adding) the empty word to the language of several possibly overlapping factors of a concatenation product. We consider regular expressions containing multi-bar or multi-tilde operators, and we show that any such extended expression with n symbol occurrences can be converted into an automaton with $n + 1$ states.

Keywords: Extended regular expression, multi-bar operator, multi-tilde operator, Glushkov automaton

1. Introduction

The framework of this paper is the translation from a regular expression into a finite automaton as well as the inverse translation. Numerous studies have been devoted to this topic, leading to algorithms for both sides of the conversion. The first ones are due to McNaughton and Yamada [23] for both constructions, to Glushkov [14] for constructing an automaton, and to Brzozowski and McCluskey [3] for constructing an expression.

Many research works have focused on producing a small automaton as efficiently as possible. For example there exist quadratic algorithms for converting an expression