Summary: Most modern libraries for regular expression matching allow back-references (i.e., repetition operators) that substantially increase expressive power, but also lead to intractability. In order to find a better balance between expressiveness and tractability, we combine these with the notion of determinism for regular expressions used in XML DTDs and XML Schema. This includes the definition of a suitable automaton model, and a generalization of the Glushkov construction. We demonstrate that, compared to their non-deterministic superclass, these deterministic regular expressions with back-references have desirable algorithmic properties (i.e., efficiently solvable membership problem and some decidable problems in static analysis), while, at the same time, their expressive power exceeds that of deterministic regular expressions without back-references.

MSC: 68Q45 Formal languages and automata

Keywords: deterministic regular expression; regex; Glushkov automaton

Software: XPath; XQuery; PERL

Full Text: DOI Link

References:
[16] Cox, R., Regular expression matching can be simple and fast (but is slow in Java, Perl, PHP, Python, Ruby, ...), (2007)
[18] Diekert, V., Makanin’s Algorithm, (2002), in: [43], chapter 12
[34] Hopcroft, J. E.; Ullman, J. D., Introduction to Automata Theory, Languages, and Computation, (1979), Addison-Wesley · Zbl 0426.68001
[37] Kay, M., XPath and XQuery Functions and Operators 3.0 W3C Recommendation, (2014), W3C
[40] Latte, M.; Niewerth, M., Definability by weakly deterministic regular expressions with counters is decidable, (Proc. MFCS 2015, (2015)) · Zbl 1402.68146
[53] Sperberg-McQueen, C. M.; Thompson, H., XML schema, (2005)

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.