

LEXICALIZED SYNTACTIC ANALYSIS BY TWO-WAY RESTARTING AUTOMATA

FRANTIŠEK MRÁZ^(A) FRIEDRICH OTTO^(B) DANA PARDUBSKÁ^(C)
MARTIN PLÁTEK^(D)

^(A) *Charles University, Department of Computer Science
Malostranské nám. 25, 118 00 Prague 1, Czech Republic
frantisek.mraz@mff.cuni.cz*

^(B) *Fachbereich Elektrotechnik/Informatik, Universität Kassel
34109 Kassel, Germany
f.otto@uni-kassel.de*

^(C) *Comenius University in Bratislava, Department of Computer Science
Mlynská Dolina, 84248 Bratislava, Slovakia
pardubska@dcs.fmph.uniba.sk*

^(D) *Institute of Computer Science, Czech Academy of Sciences
P.O. Box 5, 182 07 Prague 8, Czech Republic
martin.platek@mff.cuni.cz*

ABSTRACT

We study *h-lexicalized two-way restarting automata* that can rewrite at most i times per cycle for some $i \geq 1$ (hRLWW(i)-automata). This model is useful for the study of lexical (syntactic) disambiguation, which is a concept from linguistics that is based on certain reduction patterns. We study lexical disambiguation through the formal notion of *h-lexicalized syntactic analysis* (hLSA). The hLSA is a relation between a *basic language* and the corresponding *h-proper language*, which is obtained from the basic language by mapping all basic symbols to input symbols. We stress the sensitivity of hLSA by hRLWW(i)-automata to the size of the window, the number of possible rewrites per cycle, and the degree of (non-)monotonicity of these automata. We introduce the concepts of *contextually transparent languages* and *contextually transparent lexicalized analyses* based on very special reduction patterns, and we present two-dimensional hierarchies of their subclasses based on the size of the window and on the degree of monotonicity. The bottoms of these hierarchies correspond to the regular and the context-free languages. The class of contextually transparent languages forms a proper subclass of the class of context-sensitive languages with syntactically natural properties.

Keywords: analysis by reduction, two-way restarting automaton, h-lexicalization, contextually transparent language, mildly context-sensitive language

This research was partially done with institutional support RVO: 67985807, and it was partially supported by the grant of the Czech Science Foundation No. 19-05704S and by VEGA 2/0165/16. Some of the results have been presented at the 23rd Prague Stringology Conference, August 2019. An extended abstract appeared in the proceedings [10].