Weighted Automata and Weighted Logics over Tree-like Structures

DISSERTATION

Christian Mathissen

Fakultät für Mathematik und Informatik
Universität Leipzig
mathissen@informatik.uni-leipzig.de

In theoretical computer science the connection between automata and logic is of fundamental importance. This connection was first considered by Büchi and Elgot in the 1960s, when they showed that the languages accepted by finite automata are precisely those languages that can be defined in monadic second-order logic (MSO). In this thesis we consider extensions of Büchi's and Elgot's theorem into two directions. First, we consider classes of objects which are more general than words and carry a tree-like structure. Second, we consider quantitative aspects and investigate weighted automata operating on these structures. The study of weighted automata goes back to the work of Schützenberger. He equipped the transitions of an automaton additionally with a weight and studied the behavior of such a device which is now a formal power series, i.e. a mapping assigning to a word an element of a semiring. A semiring is the algebraic structure that carries the weights. For example, the natural numbers form a semiring, but also the probabilistic semiring given by the interval $[0, 1]$ together with the max-operator and the usual multiplication is a semiring.

This thesis investigates different weighted automaton models over tree-like structures such as texts, nested words and hedges, which have already been considered in the literature. We characterize all automaton models logically. This is achieved by considering suitable adaptions of weighted logics. The formalism of weighted logics was introduced by Droste and Gastin [1] in 2005 and provides an extension of classical MSO which is now enriched with values from a semiring in order to add quantitative expressiveness. Since already for words the full weighted MSO is expressively stronger than weighted automata, we restrict the consideration to a syntactically defined fragment called sRMSO which was proposed in [2]. Now, rather than proving for each class of structures a characterization on its own, for instance by an induction over the structure of formulae, we use a translation technique and reduce the results to ones that have already been shown. This admits the advantage that it gives insight into the similarities of the different structures. More importantly, by using the translation technique we get decidability results for the emptiness and equivalence problems from corresponding results for trees.

For the case of hedges and nested words the automaton models are straightforward generalizations of the unweighted models which had already been presented in the literature. For texts, however, no automaton model had been considered so far and the model of weighted

branching and parenthesizing automata we present is new, even for the unweighted case. It provides a joint generalization of a model of Ésik & Németh [3], the so-called parenthesizing automata, and a model of Lodaya & Weil [5], the so-called branching automata. Thomas [6] established a general approach for defining automata on different classes of finite graphs (of bounded degree). He calls the automata resulting from this approach "graph acceptors". We show in this thesis that graph acceptors are strictly weaker than branching and parenthesizing automata. Note that this is rather surprising, as for words graph acceptors and finite automata are equivalent.

Apart from logical ones we also give algebraic characterizations. For this we introduce the notion of a recognizable series over a general algebra as well as the notion of a syntactic algebra of a series. The notion of a recognizable series always depends on a set of operations defined on the class under consideration. Recognizable series over an appropriately chosen set of operations turn out to coincide with the behaviors of automata in all the cases we consider.

In the last chapter of this thesis we consider algebraic formal power series. Algebraic formal power series form an important generalization of context-free languages. They had already been considered by Chomsky and Schützenberger and have since been intensively studied. We apply the logical characterization of regular nested word series and obtain characterizations of algebraic formal power series in terms of weighted logics generalizing a result of Lautemann, Schwentick and Thérien [4] for context-free languages. Let us point out, that even though the characterizations of algebraic formal power series are generalizations of the results of [4], in contrast to the latter paper, we give a different proof, also for their result, using our connection between texts and nested words as well as using weighted nested word automata and weighted branching and parenthesizing automata.

References

[1] M. Droste and P. Gastin. Weighted automata and weighted logics. In *Proc. of the 32nd ICALP, Lisbon*, volume 3580 of *Lecture Notes in Computer Science*, pages 513–525, 2005.

[2] M. Droste and P. Gastin. Weighted automata and weighted logics. Chapter 5 in *Handbook of Weighted Automata*. (M. Droste, W. Kuich, and H. Vogler, eds.) EATCS Monographs on Theoretical Computer Science. Springer, 2009. to appear.

[3] Z. Ésik and Z.L. Németh. Higher dimensional automata. *Journal of Automata, Languages and Combinatorics*, 9(1):3–29, 2004.

[4] C. Lautemann, T. Schwentick, and D. Thérien. Logics for context-free languages. In *Proc. of the 8th CSL, Kazimierz*, volume 933 of *Lecture Notes in Computer Science*, pages 205–216, 1994.

[5] K. Lodaya and P. Weil. Rationality in algebras with a series operation. *Information and Computation*, 171(2):269–293, 2001.

[6] W. Thomas. On logics, tilings, and automata. In *Proc. of the 18th ICALP, Madrid*, volume 510 of *Lecture Notes in Computer Science*, pages 441–454, 1991.