RHEINISCH-
WESTFÄLISCHE
TECHNISCHE
HOCHSCHULE
AACHEN

LEHRSTUHL FÜR INFORMATIK VII
PROF. DR. W. THOMAS

Lehrstuhl für Informatik VII • RWTH Aachen • D-52056 Aachen

Dienstgebäude:
Ahornstraße 55, 52074 Aachen

Telefon: 0241 / 80 - 21 700(direkt)
0241 / 80 - 21 701(Sekr.)
Telefax: 0241 / 80 - 22 215
E-mail: thomas@informatik.rwth-aachen.de

Report on the Dissertation

Saaed Salehi: Varieties of Tree Languages

The subject of this thesis is located in the intersection of universal algebra and formal tree language theory, with tree automata as an implicit link. The field has two roots: the use of tree automata, or equivalently finite algebras, to define languages of trees (proposed by Büchi, Elgot, Mezei, Wright in the sixties), and the variety theory of monoids and classes of regular word languages which culminates in the celebrated variety theorem of Eilenberg and Schützenberger. This theorem establishes a correspondence between classes of regular languages that satisfy some simple closure properties and pseudovarieties of monoids, respectively semigroups. (For simplicity, the term "variety" rather than "pseudovariety" is used in the thesis and in this report.) The correspondence result constitutes the frame for a beautiful and powerful classification of the regular word languages, which has also led to some interesting decidability results (allowing to check algorithmically whether a given regular language has a property like "star-free", locally testable", etc.).

Attempts to lift this classification theory to tree languages (and on the algebraic side to proceed from monoids to more complex structures) started in the eighties, and from the beginning it became clear that this is a much more difficult enterprise. Already simple examples of tree language properties (like reverse definite) turned out non-trivial for a characeterization by syntactic algebras, and to capture syntactically the most fundamental proper subclass of the regular tree languages, the first-order definable sets, continues to be a prominent open problem. Furthermore, there seems to be no "canonical" theory of tree language varieties, as seen from the different proposals made by several authors, among them Steinby, Almeida, Esik, Wilke, and also the present referee.

The thesis of Saaed Salehi offers a number of valuable and technically nontrivial results which considerably refine the algebraic instrumentarium for classifying regular tree languages and which sharpen our understanding of the known approaches. The main contributions are the following:

1. An extension of Steinby's theorem on tree language varieties to the many-sorted case,
2. the transfer of Pin's positive variety theorem (capturing language classes which are not necessarily closed under complementation) to the domain of trees,
3. an intriguing analysis of the power of monoid-oriented variety theory for charac-terizing tree language properties, also clarifying the difference to Steinby's theory,
4. further results which combine the techniques of 3. and 4.,
5. a study of Wilke's tree algebras, invoking techniques from rewriting theory and providing an interesting specialization.

It is rather difficult to treat all these contributions on a couple of pages. I can only give some short remarks.

The first two chapters (not counting the introduction) offer generalizations for two known results. First, Steinby's tree language variety theorem is lifted to the case of many-sorted tree languages and algebras. The feature of many-sortedness is needed, for example, in Courcelle's approach to introduce recognizable sets of finite graphs (an aspect which is not pursued in the thesis). While the many-sorted framework does not offer surprises, it requires a very diligent technical treatment. The author shows high competence in setting up this framework, and he carefully links the complementary tracks of syntactic algebras, syntactic congruences, and language varieties. In the final section of the chapter, also the case of generalized varieties (involving different label alphabets of trees) is covered. - Similar remarks are in order for the subsequent chapter on the positive varieties, starting from Pin's work of 1995. Here quasi orders on ordered algebras and the induced ideals take the role of congruences of ordinary algebras and their varieties. As the typical examples, the classes of finite, respectively co-finite tree languages (and their description in terms of nilpotent, respectively co-nilpotent ordered algebras and suitably defined homomorphisms) are considered.

A very enlightening study follows in Chapters 4 and 5, where definability via syntactic monoids is captured as a natural special case of definability via syntactic algebras. The weakness of the monoid based approach is well-known (as observed by Steinby in 1992), but it was open how to set up a variety theorem on recognizability via translation monoids. The author solves this problem (raised by Steinby as well as Esik) by a nice and natural condition on syntactic algebras (Propositions 4.1.7 and 4.1.9) and converting this to extra conditions for a variety theorem (Proposition 4.2.14, which - as many others - rather should be called "theorem"). He shows that the definability by syntactic monoids and by syntactic theories in the sense of Esik are mutually incomparable specializaions of the general case (syntactic algebras), and he clarifies the role of the extra conditions by the examples of 1-definite, finite/ co-finite, and aperiodic tree languages. These results and comparisons give very valuable methodological insight. As an aside, some claims of Nivat and Podelski (1989) are refuted. The final section of the chapter addresses the definability in terms of syntactic semigroups rather than monoids.

In Chapter 5 the author takes the laborious path to develop a variant of Theorem 4.1.14 mentioned above for the case of ordered monoids in the spirit of Chapter 3. This yields the variety theorem 5.2.11. A very interesting case study follows: Here the tree languages definable by semilattice monoids are characterized (Theorem 5.3.14).

The final chapter is devoted to the theory of tree algebras as proposed by Wilke (1996) for a characterization of frontier testable tree languages. The idea is to introduce six constructors which gener-

ate objects of three sorts: labels, contexts (also called special trees in the literature), and trees in the proper sense. Recognizability is defined modulo some axioms which connect the constructors. For simplicity (and I wonder whether there are deeper reasons) only the case of binary trees is considered. Three interesting results are proved: First, Wilke's axioms are shown complete via a conversion into a rewriting system and the representation of trees as unique irreducible terms. Second, in a quite dense treatment a variety theorem is shown for a subclass of the "label"-generated tree algebras, called "reduced" (Definition 6.2.10). Thus a nice domain is singled out where Wilke's algebras fit in a general sense. The third result is another intriguing fact on the power of Wilke's approach: The six constructors preserve the syntactic congruences attached to the three mentioned sorts over binary trees if the leaf alphabet contains at least seven symbols. The tedious and demanding proof is only outlined; in full detail the construction appears in the two papers [43], [44] of the bibliography.

So the thesis offers a wealth of results, refining, deepening, and connecting the numerous existing theories of recognizability of tree languages. While this analysis sharpens our understanding of the known approaches, it leaves open (so far) fundamental questions on decidability. For example, the decidability of the conditions of Theorem 4.2.14 (characterizing recognizability via monoids) remains unsettled. That this is an active field is documented also by a recent thesis (M. Bojanczyk, Warsaw) where some progress on decidability questions regarding the expressiveness of weak logics over trees is obtained.

The thesis is written in a clear and mathematical style. The concise and often quite terse writing is lucid but addressed to experts; and more motivation, examples, and intuitive discussions would be necessary to make this work easily accessible to non-specialists. At several places the author cites from key papers when addressing the background theory – in this respect, when connecting himself with the existing frameworks, I found his approach a bit too modest in view of the clear merits of his own (similarly to his insistence on showing just "propositions" rather than "theorems").

In summary, I consider this work as a highly substantial, technically demanding contribution to algebraically oriented tree language theory. The results clearly constitute methodological progress in this complex area and are already now well visible on the international level.

In conclusion, I recommend to the Faculty of Mathematics and Natural Sciences of the University of Turku without any reservation to accept this valuable thesis as doctoral dissertation.

Aachen, 22.6.2005

(W. Thomas)