Institut
für Informatik
Christian-Albrechts-Universität zu Kiel

Theoretische Informatik
Christian-Albrechts-Platz 4
D-24118 Kiel

Prof. Dr. Thomas Wilke
Tel.: +49 431 880 7511
Fax: +49 431 880 7614
wilke@ti.informatik.uni-kiel.de

Christian-Albrechts-Universität · Institut für Informatik · 24098 Kiel

Faculty of Mathematics and Natural Sciences of the
University of Turku

18.08.2005

Assessment of the Doctorol Dissertation "Varieties of Tree Languages" by Saeed Salehi

As the title—Varieties of Tree Languages—suggests, the subject of the thesis sits right on the intersection of universal algebra and formal language theory, the link being provided by the facts that formal languages can be viewed as subsets of free algebras and that acceptance by a *finite*-state device corresponds to being a union of equivalence classes of a *finite*-index congruence (being an inverse image of a homomorphism into a *finite* algebra). Characterizations of classes of finite algebras thus carry over to classes of formal languages and vice versa, similar to what is known from Galois theory. The fundamental results from the sixties and seventies are Schützenberger's characterization of star-free languages by aperiodic monoids (1965) and Eilenberg's theorem (1976), which makes explicit a correspondence between pseudovarieties of finite semigroups and varieties of regular languages of words. The aim of all the efforts has always been to provide a powerful framework for classifying formal languages, ideally resulting in efficient algorithms for checking whether a language has this or that distinctive feature, for example, being definable by a star-free expression.

About a decade after the first fundamental results on word languages had been obtained and one had realized how powerful the concepts developed were, efforts were made to extend the theory in various directions, in particular, to finite trees, a straightforward generalization of words. Different approaches were taken: At one extreme, trees were viewed as terms over arbitrary signatures and the rôle finite semigroups had played in the word case was taken over by finite algebras without any restriction (Steinby, 1979, Almeida, 1990, Ésik, 1999, etc.). At the other extreme, trees were decomposed along paths and thus retained a word structure; in this approach finite semigroups were still the structures of choice for classifying regular tree languages (Nivat, Podelski, Péladeau, Thomas etc.). And, of course, there were other approaches that were trying to combine the two extremes, for instance,one approach that uses so-called tree algebras, three-sorted algebras of a particular shape. Over time, a large zoo of competing approaches had emerged, each having its particular advantages and disadvantages. The relations between all the approaches were, however, neither obvious nor well-studied. It was this task—to compare all the approaches in order to pave the way for further progress—that Saeed Salehi chose as the topic of his PhD project.

In doing this, Saeed Salehi had chosen an important and difficult topic: he had to deal with all

the different approaches, which sometimes were built on deep mathematical theories, and he had to overcome a considerable number of problems, some of which could not have been foreseen when he started and others of which had been open for quite a while, for instance, it was not known whether variety theorems for monoid-definable or tree-algebra-definable languages exist.

The actual results Saeed Salehi presents in his thesis are remarkable. The most important ones are:

1. variety theorems for tree languages: for many-sorted algebras (Chapter 2), for ordered algebras (Chapter 3 and Chapter 5), and for tree algebras (Chapter 6),
2. characterizations of semigroup-definable and monoid-definable varieties of tree languages and their ordered versions (Chapters 4 and 5),
3. a disclosure of a crucial misconception of earlier work by Nivat and Podelski (Chapter 4), and
4. a fairly deep study of tree algebras (Chapter 6).

These results bring about exactly what is needed: a clear picture of the interconnections of all the algebraic approaches for classifying regular tree languages that had been developed.

First, Saeed Salehi's thesis offers new insights. We not only know now that monoid-definable and semigroup-definable classes of languages can indeed be described by closure properties. We also know that some basic classes of languages cannot be described by monoids or semigroups, contrary to what the community had assumed before (but had not been able to prove, of course). In addition, we now have a fairly good understanding of the classes of tree languages which can be characterized by tree algebras.

Second, Saeed Salehi's thesis contains new indeas and introduces new concepts. Clearly, we find straightforward generalizations of well-known notions, for instance, the notion of a "many-sorted variety of finite algebras" or the notion of a "variety of ordered finites algebras" in the thesis. Substantial progress is achieved by definitions of new closure properties (in particular, with respect to semigroup and monoid-definable tree language classes) and new notions of varieties with respect to tree algebras ("b-variety" and "r-variety").

Third, the results obtained in the thesis show that Saeed Salehi masters his field of research. The proofs of some of the theorems are quite involved and require a perfect understanding of what had been proved in the field prior to the start of the PhD project. Proofs of some of the results (see Chapter 6 and the references therein) are very involved and rather long. (For very good reasons, they are not included in the thesis. Instead, the reader is referred to the corresponding publications.)

Fourth, there is no doubt that with the thesis Saeed Salehi has demonstrated that he is familiar with the relevant literature in his field of research and the various directions of research pursued nowadays. In fact, he has a very good overview of all what has been done.

That the results of Saeed Salehi's thesis are important to the field is also reflected by the fact that they have been published in appropriate journals of considerable quality or even selected for presentation at international conferences.— Some of these results were obtained as part of joint work with other researchers, but it is obvious from the statement by Professor Magnus Steinby and also from what one could observe as a member of the research community that Saeed Salehi was at the center of all this research and the driving force for it.

There is only little criticism that needs to be spelled out. The organization of the material is not optimal; it follows too closely the publications that the thesis is based on. In his thesis, Saeed Salehi is very reserved as far as his own opinion is concerned: The thesis is merely a compilation of technical results, with only a few comments, often quoted from the literature. However, the disputation clearly showed that Saeed Salehi is a mature member of the community who has his own opinions and can state them precisely; it seems he was simply a little too modest when he made the write-up for his thesis. Finally, as the ultimate aim of the algebraic framework for classifying formal languages is always the characterization of concrete classes of languages, one would have wished to see more concrete examples in the thesis. There are some, but there could have been more.— All this weighs only little; it does not in the least diminish the overall value of the thesis.

To sum up: Saeed Salehi's thesis offers profound insights, develops new and important concepts, and reveals, for the first time, crucial and clearly non-trivial interconnections between the various approaches that had been taken before. It clearly achieves its goal, namely to clear up the complex scene that had emerged over time and to thus pave the way for new developments. In conclusion, I can only highly recommend to

accept the thesis.

It was a pleasure for me to read Saeed Salehi's thesis and to be his opponent in his disputation.

Prof. Dr. Thomas Wilke