Tree Algebras

Saeed Salehi
Turku Center for Computer Sciences
Lemminkaisenkatu 14A
FIN-20520 Turku
saeed@cs.utu.fi

Many important families of regular languages have effective characterizations in terms of *syntactic monoids* or *syntactic semigroups* (see e.g. [1]). Definition of the syntactic monoid (resp. syntactic semigroup) of a language $L \subseteq X^*$ (resp. a language $L \subseteq X^+$) over an alphabet X requires regarding L as a subset of the free monoid X^* (resp. free semigroup X^+). On the other hand tree languages have traditionally been regarded as subsets of term algebras. So it appears natural to base the classification of regular tree languages over a ranked alphabet Σ on their *syntactic algebras* as was done by Steinby [4]. Also the *syntactic monoids/semigroups* of tree languages introduced by Thomas [5] have been used for characterizing families of tree languages. Nivat and Podelski's approach was, in a sense, a step further in this direction by treating (binary) trees as elements of an infinitely generated free monoid (see [2] and [3].)

Here we consider another rather new formalism introduced by Wilke [6] in which trees are not directly viewed as elements of any algebraic structure but are represented by terms over a signature Γ which consists of six operation symbols involving three sorts Alphabet, Tree and Context. The trees thus represented are binary trees over a given label alphabet. A *tree algebra* is a Γ-algebra satisfying certain identities which identify (some) pairs of Γ-terms which represent the same tree. The *syntactic tree algebra* of tree language L is defined in a natural way. Its component of sort Tree is the syntactic algebra of L in the sense of [4], while its CONTEXT-component is a semigroup almost identical to the syntactic monoid of [5].

We shall study various aspects of tree algebras as well as the relationships between the various approaches mentioned above.

Firstly, we consider the canonical morphism from the Γ-term algebra generated by a given label alphabet A. Its TREE-component yields the A-trees from their representations as Γ-terms, and by showing that its kernel is the fully invariant congruence relation generated by Wilke's axioms of tree algebras, one gets immediately the soundness and completeness of the axiom system (i.e. any two Γ-term representing the same tree are proved to be equal in the axiom system). Along the proof of the above mentioned statement, a complete term rewriting system is constructed for Wilke's axiom system. Hence, the word problem for tree algebras is seen to be effectively solvable.

It also turns out that some other formalizations using a different signature or another axiomatization for defining tree algebras are possible.

Furthermore, we characterize the tree algebras isomorphic to syntactic tree algebras of a regular tree language. The conditions are extensions of those given in [4] for traditional syntactic algebras.

Finally, Wilke's axiomatization for characterizing k-frontier testable tree languages is simplified so that the axioms and the proof can be written by taking out the sort CONTEXT.

References

[1] Eilenberg S., *Automata, Languages and Machines*, Vol. B, Academic Press, New York 1976.

[2] Nivat M. & Podelski A., "Tree monoids and recognizability of sets of finite trees", *Resolution of Equation in Algebraic Structures*, Vol. 1, Academic Press, Boston, MA, 1989, pp. 351-367.

[3] Podelski, A., "A monoid approach to tree languages", in: Nivat M. & Podelski A. (ed.) *Tree Automata and Languages*, Elsevier-Amsterdam (1992) pp.

[4] Steinby M., "A theory of tree language varieties", in: Nivat M. & Podelski A. (ed.) *Tree Automata and Languages*, Elsvier-Amsterdam (1992) pp. 57-81.

[5] Thomas, W., "Logical aspects in the study of tree languages", *Ninth Colloquium on Trees in Algebra and in Programming* (Proc. CAAP'84), Cambridge University Press, Cambridge 1984, 31-51.

[6] Wilke T., "An algebraic characterization of frontier testable tree languages", *Theoretical Computer Science*, Vol. 154, N. 1 (1996) pp. 85-106.