Recognisability for infinite trees

Achim Blumensath

TU Darmstadt

Abstract. We develop an algebraic language theory for languages of infinite trees. This theory is based on so-called *ω-hyperclones.* We show that a language of infinite trees is regular if, and only if, it is recognised by a finitary path-continuous ω-hyperclone. Aiming at applications to decision procedures we also introduce the corresponding notion of a Wilke algebra and we prove that the omega-power of a Wilke algebra uniquely determines the infinite product. As an application we give a purely algebraic proof of Rabin's Tree Theorem.

Instead of using finite automata to develop the theory of regular languages, one can also employ semigroup theory. By now this approach has a long tradition and there exists an extended structure theory connecting varieties of languages with finite semigroups. This theory is particularly effective if one is interested in characterising subclasses of the class of all regular languages. For instance, the only known decidable characterisation of the class of first-order definable languages is based on semigroup theory.

Naturally, there have been attempts to generalise this theory to other notions of regularity. For languages of ω-words, such a generalisation has largely been achieved. A detailed account can be found in the book of Pin and Perrin [11]. There also have been several contributions to an algebraic theory for languages of finite trees [15, 18, 16, 5, 12, 7, 6, 13, 3, 1, 8]. But the resulting theory is still fragmentary with several competing approaches and formalisations. Our own work has been influenced in particular by the following two articles: Ésik and Weil [7] have developed an approach using *preclones,* while Bojańczyk and Walukiewicz [3] use *forest algebras.* As far as the algebraic setting is concerned, the formalisation in the present article most closely resembles the work on *clones* by Ésik [5].

So far, an algebraic theory for languages of *infinite trees* is still missing. The main obstacle is the lack of appropriate combinatorial tools, like Ramseyan factorisation theorems for infinite trees. In particular, a purely combinatorial proof that every nonempty regular language of infinite trees contains a regular tree is still missing. There is recent work of Bojańczyk and Idziaszek [2] on characterisation results for classes of infinite trees that manages to circumvent these problems by a technical trick: since every regular language of infinite trees is determined by the regular trees it contains, it is sufficient to consider only regular trees.

In this talk I provide steps in the development of an algebraic theory for recognisability of classes of infinite trees. Inspired by the work of Ésik and Weil on preclones, we define suitable algebras of infinite trees called *ω-hyperclones.* We can show that every regular language is recognised by some homomorphism into such a (finitary, path-continuous) ω-hyperclone.

The proof is performed in two steps. First, we define a special class of ω-hyperclones called *path-hyperclones* that directly correspond to tree-automata. The problem with path-hyperclones is that their definition is not axiomatic, but syntactic. That is, given an arbitrary ω-hyperclone we cannot tell from the definition whether or not this ω-hyperclone is isomorphic to some path-hyperclone.

In the second step, we therefore give an algebraic characterisation of the main properties of such path-hyperclones (they are *path-continuous*). Using this result we can transfer our characterisation from path-hyperclones to path-continuous ω-hyperclones.

Finally, we prove that the class of path-continuous ω-hyperclones is closed under products and a certain power-set operation. From these results we can deduce a second (equivalent) version of our main theorem: recognisability by finitary path-continuous ω-hyperclones is the same as definability in monadic second-order logic.

Historically, one of the main advantages of algebraic characterisations of regularity has been their suitability for deriving decision procedures for subclasses of regular languages. We hope that the theory developed in this talk will also be useful in this respect. One of the prerequisites for obtaining decision procedures is that the algebras one is dealing with are finitely representable. For the algebras introduced above this is, in fact, the case.

As an application of this result, we present a new proof of Rabin's Tree Theorem, which states that the monadic second-order theory of the infinite complete binary tree is decidable. The standard proof of this result is based on the translation of monadic second-order formulae into tree automata. The required automata-theoretic machinery in turn rests on two main results: (i) the positional determinacy of parity games and (ii) either a determinisation construction for Büchi automata, or an analogous result for tree automata (see [9, 10]).

For Büchi's Theorem – the corresponding result for the natural numbers with successor relation – there exists, besides the usual automata-theoretic proof, an alternative proof due to Shelah [14, 17]. It is purely combinatorial in nature and it is based on Feferman-Vaught like composition arguments for monadic second-order theories.

For a long time it has been an open problem to extend these results to the theory of the binary tree. What was missing in order to transfer Shelah's proof was a suitable variant of Ramsey's Theorem for trees. Such a theorem has recently been provided by Colcombet [4]. In this article we use Colcombet's result to prove that our algebras have finite representations. This fact is then used to give an alternative proof of Rabin's theorem without references to automata or games.

References

1. P. Baltazar. M-Solid Varieties of Languages. *Acta Cynbernetica*, 18:719–731, 2008.
2. M. Bojańczyk and T. Idziaszek. Algebra for Infinite Forests with an Application to the Temporal Logic EF. In *Proc. 20th International Conference on Concurrency Theory, CONCUR, LNCS 5710*, pages 131–145, 2009.
3. M. Bojańczyk and I. Walukiewicz. Forest Algebras. In J. Flum, E. Grädel, and T. Wilke, editors, *Logic and Automata: History and Perspectives*, pages 107–132. Amsterdam University Press, 2007.
4. T. Colcombet. On Factorization Forests. Technical Report hal-00125047, Irisa Rennes, 2007.
5. Z. Ésik. A variety theorem for trees and theories. *Publ. Math. Debrecen*, 54:711–762, 1999.
6. Z. Ésik. Characterizing CTL-like logics on finite trees. *Theoretical Computer Science*, 356:136–152, 2006.
7. Z. Ésik and P. Weil. Algebraic recognizability of regular tree languages. *Theoretical Computer Science*, 340:291–321, 2005.
8. Z. Ésik and P. Weil. Algebraic Characterization of Logically Defined Tree Languages. *International Journal of Algebra and Computation*, 20:195–239, 2010.
9. E. Grädel, W. Thomas, and T. Wilke. *Automata, Logic, and Infinite Games*. LNCS 2500. Springer-Verlag, 2002.
10. D. E. Muller and P. E. Schupp. Simulating alternating tree automata by non-deterministic automata: New results and new proofs of the theorems of Rabin, McNaughton and Safra. *Theoretical Computer Science*, 141:69–107, 1995.
11. D. Perrin and J.-E. Pin. *Infinite Words – Automata, Semigroups, Logic and Games*. Elsevier, 2004.
12. T. Petkovic and S. Salehi. Positive varieties of tree languages. *Theoretical Computer Science*, 347:1–35, 2005.
13. S. Salehi and M. Steinby. Tree algebras and varieties of tree languages. *Theoretical Computer Science*, 377:1–24, 2007.
14. S. Shelah. The Monadic Second Order Theory of Order. *Annals of Mathematics*, 102:379–419, 1975.
15. M. Steinby. A theory of tree language varieties. In *Tree Automata and Languages*, pages 57–82. North-Holland, 1992.
16. M. Steinby. General Varieties of Tree Languages. *Theoretical Computer Science*, 205:1–43, 1998.
17. W. Thomas. Ehrenfeucht Games, the Composition Method, and the Monadic Theory of Ordinal Words. *LNCS*, 1261:118–143, 1997.
18. T. Wilke. Algebras for Classifying Regular Tree Languages and an Application to Frontier Testability. In *Proc. 20th Int. Colloquium on Automata, Languages and Programming, ICALP 1993, LNCS 700*, pages 347–358, 1993.