Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Information and Computation 197 (2005) 55–89
Information and Computation
www.elsevier.com/locate/ic

The equational theory of regular words

Stephen L. Bloom [a,*], Zoltán Ésik [b,1]

[a] Department of Computer Science, Stevens Institute of Technology, Hoboken, NJ 07030, USA
[b] Institute for Informatics, University of Szeged, Szeged, Hungary

Received 15 April 2004; revised 18 January 2005

Abstract

Courcelle introduced the study of regular words, i.e., words isomorphic to frontiers of regular trees. Heilbrunner showed that a nonempty word is regular iff it can be generated from the singletons by the operations of concatenation, omega power, omega-op power, and the infinite family of shuffle operations. We prove that the algebra of nonempty regular words on the set A, equipped with these operations, is freely generated by A in a variety which is axiomatizable by an infinite collection of some natural equations. We also show that this variety has no finite equational basis and that its equational theory is decidable in polynomial time.

Keywords: Words; Arrangement; Regular; Linear order; Equational theory

1. Introduction

By "word" we understand a labeled linear order, extending the familiar notion of a labeling of $\{1, 2, \ldots, n\}$, for some $n \geqslant 0$. Courcelle [10] introduced the study of such words ("arrangements,"

* Corresponding author.
E-mail address: bloom@cs.stevens.edu (S.L. Bloom).
[1] Partially supported by the National Foundation of Hungary for Scientific Research, Grant T46686. Currently visiting Rovira i Virgili University, Tarragona, Spain.

0890-5401/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ic.2005.01.004

in his terminology). He showed that every finite or countable word is isomorphic to the frontier of a complete binary tree, where the linear order on the leaves of the tree is the lexicographic (left to right) order. He introduced several operations on such words, including concatenation (or product), omega and omega-op powers. He proved that initial solutions of finite systems of fixed point equations

$$x_i = u_i, \quad i = 1, 2, \ldots, k \tag{1}$$

where $x_1, \ldots, x_k$ are variables and the u_i are finite words on the letters in a set A and the variables, are isomorphic to frontiers of *regular* trees. Further, he showed that the solutions of certain kinds of systems can be expressed by "quasi-rational" expressions, formed from single letters in A using the operations of concatenation, omega and omega-op power. Courcelle asked for a complete set of axioms for these operations. In [1], a complete set of axioms for just the concatenation and omega power operation on words was given, and in [3] Courcelle's question was answered.

We call a word which is isomorphic to the frontier of a regular binary tree a **regular word**. Several results on regular words have been obtained by Heilbrunner [13], Thomas [15], and the authors [2]. Heilbrunner showed that all nonempty regular words on the set A can be generated from single letters by means of the above mentioned operations, namely concatenation, omega and omega-op power, together with (infinitely many) "shuffle" operations. Terms formed from letters in A and these operations are called "terms on A." Heilbrunner gave an algorithm which, given a finite system of fixed point equations of the form (1) such that the first component (i.e., x_1) of the initial solution is nonempty, produces a term denoting it. Thomas gave an algorithm to determine when two terms denote isomorphic words. His algorithm is based on Rabin's theorem on automata for infinite trees.

Heilbrunner discussed several identities involving the terms with both Courcelle's operations, as well as the shuffle operations, but did not obtain a completeness result. Our paper gives a set Ax of axioms and

- in Theorem 76, shows them to be complete. This result implies that
- for any alphabet A, the algebra of regular words on an alphabet A is freely generated by A in the variety defined by these equations.
- We show also that the equational theory of this variety is decidable in polynomial time, (see Theorem 79), and is not finitely based, (see Theorem 82).

The completeness theorem, and the corresponding complexity result, provide a solution to a problem that has been open for over 20 years.

We describe our method, which may be of independent interest.

We find an appropriate "condensation" [14] of the linear order of a regular word u, and replace certain subwords by appropriately labeled points. Given a word u on the alphabet A, we show that its underlying linear order L_u is partitioned into **blocks** of an equivalence relation: two points $p < q$ are in the same block iff they both belong to some "uniform" subword (defined below) or neither does and the interval $\{x \in L_u : p \leqslant x \leqslant q\}$ is finite. The blocks of L_u are also linearly ordered in the obvious way, and we denote this linearly ordered set by $\widehat{L}_u$. The blocks of regular words are denoted by what we call the "primitive terms" below. If two primitive terms s, t denote isomorphic words, then our axioms Ax are strong enough so that $\mathsf{Ax} \vdash s = t$.