PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 83, Number 2, October 1981

MODEL COMPLETENESS OF AN ALGEBRA OF LANGUAGES[1]

DAVID HAUSSLER

ABSTRACT. An algebra $\langle \mathfrak{L}, f, g \rangle$ of languages over a finite alphabet $\Sigma = \{a_1, \ldots, a_n\}$ is defined with operations $f(L_1, \ldots, L_n) = a_1L_1 \cup \cdots \cup a_nL_n \cup \{\lambda\}$ and $g(L_1, \ldots, L_n) = a_1L_1 \cup \cdots \cup a_nL_n$ and its first order theory is shown to be model complete. A characterization of the regular languages as unique solutions of sets of equations in $\langle \mathfrak{L}, f, g \rangle$ is given and it is shown that the subalgebra $\langle \mathfrak{R}, f, g \rangle$ where $\mathfrak{R}$ is the set of regular languages is a prime model for the theory of $\langle \mathfrak{L}, f, g \rangle$. We show also that the theory of $\langle \mathfrak{L}, f, g \rangle$ is decidable.

Let $\Sigma = \{a_1, \ldots, a_n\}$ be a finite alphabet and Σ^* the free semigroup with empty word λ generated by Σ. Let $\mathfrak{L}$ be the class of all languages over Σ, i.e., all subsets of Σ^*. We introduce two n-ary operations on the languages of $\mathfrak{L}$:

$$f(L_1, \ldots, L_n) = a_1L_1 \cup \cdots \cup a_nL_n \cup \{\lambda\},$$
$$g(L_1, \ldots, L_n) = a_1L_1 \cup \cdots \cup a_nL_n,$$

where a_iL_i denotes the language obtained by prefixing all the words of L_i with the letter a_i.

Our first result is the following theorem which follows from Theorem 3 of Mycielski and Perlmutter [3].

THEOREM 1. *The first order theory of the algebra $\langle \mathfrak{L}, f, g \rangle$ is model complete.*

PROOF. Let us define a simple bijection between $\mathfrak{L}$ and the set of infinite, oriented trees with nodes labeled from $\{f, g\}$, each node having n successors. Given such a tree, label the edges emanating from each node with the letters a_1 through a_n from left to right. Associate with the tree the language consisting of all words of Σ^* corresponding to the consecutive labels of the edges of any path leading from the root to a node labeled f. It follows that the algebra $\langle \mathfrak{L}, f, g \rangle$ is isomorphic to the algebra R_σ of [3], where σ specifies the two n-ary function symbols f and g. Thus by Theorem 3 of [3], $\langle \mathfrak{L}, f, g \rangle$ is model complete. □

We now consider sets of equations for $\langle \mathfrak{L}, f, g \rangle$, i.e., sets of equations written solely in terms of the function symbols f and g and variables x_i.

Let us say that a language L is uniquely determined by a set of equations E and a variable x_p iff E is satisfiable in $\langle \mathfrak{L}, f, g \rangle$ and every assignment to the variables of E which satisfies E assigns L to x_p.

Received by the editors November 3, 1980.

1980 *Mathematics Subject Classification*. Primary 03C60, 68D30.

[1]This research was conducted while the author was supported by a University of Colorado doctoral fellowship.

0002-9939/81/0000-0482/$02.00

Lemma 1. *If L is uniquely determined by some set of equations and a variable, then L is uniquely determined by a set of equations E and x_1, where E has the unknowns $x_1, \ldots, x_m$ and is of the form $\{x_i = t_i\colon 1 < i < m\}$, the t_i's being terms which are not variables.*

Proof. Assume that L is uniquely determined by the set of equations D and the variable x_p. We define an equivalence relation, $\equiv$, on the variables appearing in D by

$$x_i \equiv x_j \quad \text{iff } D \Rightarrow x_i = x_j \text{ is true in } \langle \mathcal{L}, f, g \rangle.$$

From each equivalence class, we choose a representative, insuring that x_p is chosen as a representative of its class. We then replace all the variables in D by their representatives, obtaining a set of equations D' which is equivalent to D with respect to the remaining variables.

Using the isomorphism from Theorem 1 and Lemma B, Case I from [3], we convert D' to an equivalent set of equations $D'' = \{x_{i_k} = t_k\colon 1 < k < r\}$ where the x_{i_k}'s are distinct variables and the t_k's are terms which are not variables. Now notice that the system D'' is satisfiable in $\langle \mathcal{L}, f, g \rangle$ for every assignment of the variables which do not occur on the left-hand side of any equation of D'' (see [3, Formula (2)]). Hence x_p appears on the left-hand side of some equation in D''. To finish the proof, we substitute every variable of D'' which does not appear on the left-hand side of any equation by the variable x_p. Finally, we rename the variables to obtain a set of equations E of the desired form. $\square$

Our second theorem provides a characterization of the class of regular languages (see e.g. [2]) in terms of sets of equations in $\langle \mathcal{L}, f, g \rangle$.

Theorem 2. *The following are equivalent.*

(i) *L is uniquely determined by some set of equations and a variable in $\langle \mathcal{L}, f, g \rangle$.*

(ii) *L is uniquely determined by a set of equations E in unknowns $x_1, \ldots, x_m$ and the variable x_1, where E is of the form $\{x_i = \phi_i(x_{i_1}, \ldots, x_{i_n})\colon 1 < i < m\}$ and $\phi_i \in \{f, g\}$ for each i.*

(iii) *L is regular.*

Proof. We first show (i) $\Rightarrow$ (ii). By Lemma 1, we may assume that L is uniquely determined by $E_0 = \{x_i = t_i\colon 1 \leqslant i \leqslant m\}$ and the variable x_1 where E_0 has the properties stated in the lemma. From E_0 we will produce a set of equations E of the form specified in (ii) in the following way. Initially let $E = E_0$. Then, given any equation of E of the form $x_j = \phi(u_1, \ldots, u_n)$ where the u_i's are terms and for some $k\colon 1 \leqslant k \leqslant n$, u_k is not a variable, replace this equation with the two equations $x_j = \phi(u_1, \ldots, u_{k-1}, x_i, u_{k+1}, \ldots, u_n)$ and $x_i = u_k$ where i is the least integer such that x_i does not appear in any equations of E up to this point. We continue this operation as long as feasible. Since terms are of finite depth, this process terminates and it is apparent that it produces a set of equations E of the required form which is equivalent to E with respect to the original variables.

To show (ii) $\Rightarrow$ (iii), we transform E into a finite automaton $M = \langle Q, \Sigma, \delta, x_1, F \rangle$ accepting precisely the language L. Q, the set of states of M, is defined to be the set of variables of E. $\Sigma = \{a_1, \ldots, a_n\}$ is the alphabet of M. δ, the transition function, is defined by $\delta(x_k, a_j) = x_{i_j}$ iff E has an equation of the form $x_k = \phi(x_{i_1}, \ldots, x_{i_n})$. x_1 is the start state and F, the set of accepting states, is the set of those variables x_k for which an equation of the form $x_k = f(x_{i_1}, \ldots, x_{i_n})$ is in E. In view of the definition of the operations f and g, it is obvious that M must accept L, hence L is regular.

To see that (iii) $\Rightarrow$ (i) it suffices to observe that given any deterministic finite automaton $M = \langle Q, \Sigma, \delta, x_1, F\rangle$ with $Q \subset \{x_1, x_2, \ldots\}$ and $\Sigma = \{a_1, \ldots, a_n\}$ we can easily reverse the above construction, obtaining a set of equations E such that E and x_1 uniquely determine the language L accepted by M. □

Let $\mathfrak{R}$ be the class of regular languages over E. Since $\mathfrak{R}$ is closed under the operations f and g, $\langle \mathfrak{R}, f, g\rangle$ is a subalgebra of $\langle \mathfrak{L}, f, g\rangle$. From Theorem 2 we may easily deduce the following corollary.

COROLLARY 1. *Every finite set of equations in f and g which has a solution in $\langle \mathfrak{L}, f, g\rangle$ has a solution in $\langle \mathfrak{R}, f, g\rangle$.*

However, using [3] again, we obtain the following stronger result.

THEOREM 3. *$\langle \mathfrak{R}, f, g\rangle$ is an elementary subalgebra of $\langle \mathfrak{L}, f, g\rangle$ and is a prime model for its theory.*

PROOF. By Theorem 2 and the isomorphism of the proof of Theorem 1, $\langle \mathfrak{R}, f, g\rangle$ is isomorphic to the algebra A_σ of [3] where σ is as before. Our result follows from part (ii) of Theorem 3 of [3]. □

THEOREM 4. *The theory of $\langle \mathfrak{L}, f, g\rangle$ is decidable.*

PROOF. We can define the operations f and g in the monadic second-order theory $\langle SnS\rangle$ of n successor functions $S_1, \ldots, S_n$ as follows.

$$g(L_1, \ldots, L_n) = L \underset{\text{def}}{\Leftrightarrow} \forall x\Big[x \in L \Leftrightarrow \bigvee_{1<i<n} \exists y \in L_i[x = S_i(y)]\Big],$$

$$f(L_1, \ldots, L_n) = L \underset{\text{def}}{\Leftrightarrow} \forall x\Big[x \in L \Leftrightarrow \Big(\bigvee_{1<i<n} \exists y \in L_i[x = S_i(y)]\Big) \vee x = \lambda\Big],$$

where

$$x = \lambda \underset{\text{def}}{\Leftrightarrow} \forall y \bigwedge_{1<i<n} x \neq S_i(y).$$

Since Rabin [4] has shown that $\langle SnS\rangle$ is decidable, our theorem follows. □

REMARK. By a slight modification of the proof, Theorem 4 generalizes to the algebra R_σ of [3].

Problem. In [1], J. H. Conway defines and studies the operations:

$$\partial(L)/\partial a_i = \{w: a_i w \in L\}$$

which are existentially first order definable from f and g. Does the theory of the algebra $\langle \mathfrak{L}, f, g, \partial/\partial a_1, \ldots, \partial/\partial a_n\rangle$ admit elimination of quantifiers?

ACKNOWLEDGEMENT. I am indebted to Jan Mycielski for numerous refinements of my original observations and proofs.

References

1. J. H. Conway, *Regular algebra and finite machines*, Chapman & Hall, London, 1971.

2. J. E. Hopcroft and J. D. Ullman, *Introduction to automata theory, languages and computation*, Addison-Wesley, Reading, Mass., 1979.

3. J. Mycielski and P. Perlmutter, *Model completeness of some metric completions of absolutely free algebras*, Algebra Universalis **12** (1981) (to appear).

4. M. O. Rabin, *Decidability of second-order theories and automata on infinite trees*, Trans. Amer. Math. Soc. **141** (1969), 1–35.

Department of Computer Science, University of Colorado at Boulder, Boulder, Colorado 80309