Model Completeness of an Algebra of Languages *

David Haussler

CU-CS-178-80

University of Colorado at Boulder
DEPARTMENT OF COMPUTER SCIENCE

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

MODEL COMPLETENESS OF AN ALGEBRA OF LANGUAGES

by

David Haussler
Department of Computer Science
University of Colorado at Boulder
Boulder, Colorado

#CU-CS-178-80 March, 1980

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

Abstract

An algebra $<\mathcal{L}, f, g>$ of languages over a finite alphabet $\Sigma = \{a_1, \ldots, a_n\}$ is defined with operations $f(L_1, \ldots, L_n) = a_1 L_1 \cup \ldots \cup a_n L_n \cup \{\lambda\}$ and $g(L_1, \ldots, L_n) = a_1 L_1 \cup \ldots \cup a_n L_n$ 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 $<\mathcal{L}, f, g>$ is given and it is shown that the subalgebra $<\mathcal{R}, f, g>$ where $\mathcal{R}$ is the set of regular languages is a prime model for the theory of $<\mathcal{L}, f, g>$.

Let $\Sigma = \{a_1,\ldots,a_n\}$ be a finite alphabet and Σ^* the free semigroup with empty word λ generated by Σ. Let 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 L:

$$f(L_1,\ldots,L_n) = a_1 L_1 \cup \ldots \cup a_n L_n \cup \{\lambda\}$$

$$g(L_1,\ldots,L_n) = a_1 L_1 \cup \ldots \cup a_n L_n$$

where $a_i L_i$ 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 $< L, f, g >$ is model complete.

Proof: Let us define a simple bijection between 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 $< L, f, g >$ 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], $< L, f, g >$ is model complete. □

We now consider sets of equations for $< L, f, g >$, 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 $< L, f, g >$ and every assignment to the variables of E which satisfies E assigns L to x_p.

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 $\{x_i = t_i : 1 \le i \le m\}$, 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

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

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 $D'' = \{x_{i_k} = t_k : 1 \le k \le r\}$ 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 $\langle L, f, g \rangle$ 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 eg. [2]) in terms of sets of equations in $<\mathcal{L}, f, g>$.

Theorem 2: The following are equivalent:

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

(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 $\{x_i = \phi_i (x_{i_1}, \ldots, x_{i_n}) : 1 \le i \le m\}$ 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 $E_0 = \{x_i = t_i : 1 \le i \le m\}$ 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 $k : 1 \le k \le n, u_k$ 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. $E = \{a_1, \ldots a_n\}$ 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 R be the class of regular languages over E. Since R is closed under the operations f and g, $\langle R, f, g \rangle$ is a subalgebra of $\langle L, f, g \rangle$. From Theorem 2 we may easily deduce the following corollary:

<u>Corollary 1</u>: Every finite set of equations in f and g which has a solution in $\langle L, f, g \rangle$ has a solution in $\langle R, f, g \rangle$.

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

Theorem 3: $<\mathcal{R}, f, g>$ is an elementary subalgebra of $<\mathcal{L}, f, g>$ and is a prime model for its theory.

Proof: By the isomorphism of the proof of Theorem 1, $<\mathcal{R}, f, g>$ 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]. $\square$

In closing, let us mention a few open problems.

1. Is the theory of $<\mathcal{L}, f, g>$ decidable?

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

$$\frac{\partial}{\partial a_1}(L) = \{w : a_i w \in L\}$$

which are existentially first order definable from f and g. Does the theory of the algebra $<\mathcal{L}, f, g, \frac{\partial}{\partial a_n}, \ldots, \frac{\partial}{\partial a_n}>$ admit elimination of quantifiers?

References

[1] J. H. Conway, Regular Algebra and Finite Machines, Chapman and Hall, London, 1971.

[2] J. E. Hopcroft and J. D. Ullman, Introduction to Automata Theory, Languages and Computation, Addison-Wesley, Reading, Mass., 1979.

[3] Mycielski and Perlmutter, "Model Completeness of Some Metric Completions of Absolutely Free Algebras," Algebra Universalis, 10 (1980), to appear.

Acknowledgment

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