A Generalization of the Nerode Characterization of Regular Sets Using Well Quasi Orderings *

David Haussler

CU-CS-218-81

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.

A GENERALIZATION OF THE NERODE CHARACTERIZATION OF REGULAR SETS USING WELL QUASI ORDERINGS

by David Haussler

Department of Computer Science
University of Colorado
Boulder, Colorado 80302

CU-CS-218-81 Dec. 1981

This work was supported by the University of Colorado Doctoral Fellowship Program.

ABSTRACT

A quasi order generalizes the notion of an equivalence relation (by not demanding that the relation be transitive) and a well quasi order generalizes the notion of an equivalence relation of finite index. Given a quasi order $\leq$ on a finitely generated free monoid Σ^* and a set $S \subseteq \Sigma^*$, the closure of S under $\leq$ is defined by $cl_{\leq}(S) = \{w \varepsilon \Sigma^* : s \leq w$ for some $s \varepsilon S\}$. A quasi order $\leq$ on Σ^* is monotone if and only if $x \leq x'$ and $y \leq y'$ implies that $xy \leq x'y'$ for all $x, x', y, y' \varepsilon \Sigma^*$. The Nerode characterization of the regular sets can be stated as follows: For any $S \subseteq \Sigma^*$, S is regular if and only if there exists a monotone equivalence relation $\equiv$ of finite index on Σ^* and a (finite) set S' such that $S = cl_{\equiv}(S')$. We generalize this characterization by showing that S is regular if and only if there exists a monotone well quasi order $\leq$ on Σ^* and a (finite) set S' such that $S = cl_{\leq}(S')$.

The importance of certain equivalence relations of finite index in Kleene's theory of regular events ([Sha, Mc 56]) was first observed by J. Myhill ([My 57]) and A. Nerode ([Nr 58]).

Definition. A binary relation R on a finitely generated free monoid Σ^* is *monotone* if and only if xRx' and yRy' implies that $xyRx'y'$. A monotone equivalence relation is called a *congruence*.

It is easily verified that if $R \subseteq \Sigma^* \times \Sigma^*$ is reflexive, then R is monotone if and only if for all $x, y, z \varepsilon \Sigma^*$, xRy implies that $xzRyz$ and $zxRzy$.

Definition. An equivalence relation $\equiv$ on Σ^* is a *right* (resp *left) congruence* if and only if for all $x, y, z \varepsilon \Sigma^*$, if $x \equiv y$ then $xz \equiv yz$ (resp. $zx \equiv zy$).

The following characterization of the regular sets over Σ^* in terms of congruences on Σ^* of finite index is usually attributed to Nerode (see e.g., [Lal 79]).

Proposition 1. For any set $S \subseteq \Sigma^*$, S is regular if and only if S is a union of equivalence classes under some congruence on Σ^* of finite index.

As it turns out, it suffices to consider only left or right congruences.

Proposition 2. For any set $S \subseteq \Sigma^*$, S is regular if and only if S is a union of equivalence classes under some right (left) congruence on Σ^* of finite index.

Let us now consider a type of ordering more general than the equivalence relations.

Definition. A *quasi order* is a reflexive and transitive relation.

Obviously, an equivalence relation is a special type of quasi order which is in addition a symetric relation. Furthermore, an equivalence relation of finite index is a special kind of quasi order contained in the class of *well quasi orders* (see [Hig 52], [Erd, Rad 52] and [Kru 72]). Higman ([Hig 52]) gives the following definitions of a well quasi order and proves them equivalent.

Definition. Given a set T and a quasi order $\leq$ on T, then $\leq$ is a well quasi order on T if and only if any of the following hold

i) $\leq$ is well founded on T and each set of pairwise incomparable elements in T is finite,

ii) for each infinite sequence $\{x_i\}$ of elements in T, there exist $i<j$ such that $x_i \leq x_j$,

iii) each infinite sequence of elements in T contains an infinite ascending subsequence,

iv) T has the "finite basis property," i.e., for each set $S \subseteq T$ there exists a finite $B_S \subseteq S$ such that for every $s \in S$ there exists a $b \in B_S$ such that $b \leq s$, and

v) Every sequence of $\leq$-closed subsets of T which is stricly ascending under inclusion is finite.

From these definitions, it is obvious that a symetric quasi order on a set T is a well quasi order on T if and only if it is an equivalence relation of finite index. It follows that the class of congruences of finite index is exactly the class of *symetric* monotone well quasi orders. We now demonstrate how the Nerode Theorem (Proposition 1) can be generalized to arbitrary monotone well quasi orders.

Since an arbitrary monotone well quasi order on a set T does not induce a partition of T into disjoint subsets, we make the following definition.

Definition. Given a quasi order $\leq$ on a set T, for any $S \subseteq T$, $cl_{\leq}(S) = \{t \in T : s \leq t \text{ for some } s \in S\}$. A set $S \subseteq T$ is *$\leq$-closed* if and only if $cl_{\leq}(S) = S$.

It is obvious that for any equivalence relation $\equiv$ on Σ^*, a set $s \subseteq \Sigma^*$ is $\equiv$-closed if and only if it is a union of equivalence classes under $\equiv$. Hence the following is a straightforward generalization of the Nerode theorem.

Theorem 1. For any $S \subseteq \Sigma^*$, S is regular if and only if S is $\leq$-closed under some monotone well quasi order $\leq$ on Σ^*.

Proof. Since the "only if" part follows from Proposition 1, it suffices to show that for any monotone well quasi order $\leq$ on Σ^* each $\leq$-closed set S in Σ^* is regular. Let us assume to the contrary that we are given a monotone well quasi order $\leq$ on Σ^* and a $\leq$-closed set $S \subseteq \Sigma^*$ which is not regular. For each $w \varepsilon \Sigma^*$ let $f(w)=\{x \varepsilon \Sigma^* : wx \varepsilon S\}$. Let $\equiv$ be the binary relation on Σ^* defined by $u \equiv v$ if and only if $f(u)=f(v)$. It is readily verified that $\equiv$ is a right congruence on Σ^*. Thus since S is not regular, $\equiv$ is not of finite index by Proposition 2. Hence we can find an infinite sequence $\{w_i\}$ of words in Σ^* such that $w_i \not\equiv w_j$ for $i \neq j$. Since $\leq$ is a well quasi order, there exists an infinite subsequence of $\{w_i\}$ which is ascending with respect to $\leq$, using definition iii. Hence, we may assume that $\{w_i\}$ itself is chosen as an ascending sequence.

Since $\leq$ is monotone and $\{w_i\}$ is ascending, for any $x \varepsilon \Sigma^*$ and $i<j$, $w_i x \leq w_j x$. Hence since S is $\leq$-closed, $w_i x \varepsilon S$ implies that $w_j x \varepsilon S$. Thus the sequence $\{f(w_i)\}$ is ascending with respect to inclusion. Further, since $w_i \not\equiv w_j$ for $i \neq j$, $\{f(w_i)\}$ is strictly ascending. Now by the same reasoning as above, for any i and any $x \leq y$, if $w_i x \varepsilon S$ then $w_i y \varepsilon S$. Hence each $f(w_i)$ is $\leq$-closed. Thus $\{f(w_i)\}$ forms an infinite strictly ascending sequence of $\leq$-closed sets, contradicting the fact that $\leq$ is a well quasi order, using definition v. We conclude that every $\leq$-closed set S is regular.

Using the "finite basis property" (Definition iv) of a well quasi order, we have the following corollary of Theorem 1.

Corollary 1. A set $S \subseteq \Sigma^*$ is regular if and only if there exists some monotone well quasi order $\leq$ on Σ^* and a finite set $F \subseteq S$ such that $S=cl_{\leq}(F)$.

Definition. Given a quasi order $\leq$ on a set T, $\mathbf{A}_{\leq}$ denotes the smallest Boolean algebra of subsets of T containing the $\leq$-closed sets.

Since the regular languages are closed under the Boolean operations of union, intersection and complement, we also have the following result.

Corollary 2. For any $S \subseteq \Sigma^*$, S is regular if and only if $S \varepsilon \mathbf{A}_{\leq}$ for some monotone well quasi order $\leq$ on Σ^*.

We close with the following historical note.

In 1972, Leonard Hains published a regularity result concerning the subsequence relationship on words in Σ^* ([Hai 72]). Define $x \leq y$ if and only if there exist $a_1, \cdots, a_k \varepsilon \Sigma$ and $u_1, \cdots, u_{k+1} \varepsilon \Sigma^*$ such that $x = a_1 \cdots a_k$ and $y = u_1 a_1 \cdots u_k a_k u_{k+1}$. Hains showed that for any set $S \subseteq \Sigma^*$, the sets $\overline{S} = \{w \varepsilon \Sigma^* : s \leq w \text{ for some } s \varepsilon S\}$ and $\underline{S} = \{w \varepsilon \Sigma^* : w \leq s \text{ for some } s \varepsilon S\}$ are both regular. In proving this result, Hains "rediscovered" the theory of well quasi orders, as have several other researchers (see [Kru 72]). Shortly after this result appeared, J. H. Conway published a simpler proof of this result, using Higmans work on well quasi orders ([Con 71]).

It is apparent that $\leq$, as defined above, is a monotone quasi order, and Higman showed that it is a well quasi order ([Hig 52]). Furthermore $\overline{S} = cl_{\leq}(S)$ and $\underline{S} = \Sigma^* - cl_{\leq}(\Sigma^* - \underline{S})$, hence $\overline{S}$ and $\underline{S}$ are contained in $\mathbf{A}_{\leq}$ for any $S \varepsilon \Sigma^*$. The results of this paper further generalize the work of Hains and Conway by incorporating it into a natural extension of the Nerode theorem to well quasi orders. Further applications of Theorem 1 appear in [Hau 81], where the well quasi orders of [Ehr, Hau 81] are used to obtain a regularity characterization result for a certain class of generalized semi-Dyck languages.

ACKNOWLEDGEMENTS

I would like to acknowledge Grzegorz Rozenberg and Andrzej Ehrenfeucht for numerous suggestions concerning the presentation of this material. I am especially indebted to Dr. Rozenberg for first suggesting that these results be viewed as a generalization of the Nerode theorem.

REFERENCES

[Con 71] J. H. Conway, *Regular Algebra and Finite Machines*, Chapman and Hall, London, 1971, 63-64.

[Ehr,Hau 81] A. Ehrenfeucht and D. Haussler, "A generalization of a theorem of Graham Higman," submitted for publication in *J. Combinatorial Theory*.

[Erd,Rad 52] P. Erdos and R. Rado, "Sets having the divisor property, solution to problem 4358," *American Math. Monthly*, 59 (1952), 255-257.

[Hai 72] L. H. Hains, "On free monoids partially ordered by embedding," *J. Combinatorial Theory*, 6 (1969), 94-98.

[Hau 81] D. Haussler, "Insertion languages: a class of generalized semi-Dyck languages," in preparation.

[Hig 52] G. Higman, "Ordering by divisibility in abstract algebras", *Proc. London Math. Soc.* 3 (2) (1952), 326-336.

[Kru 72] J. B. Kruskal, "The theory of well-quasi ordering: a frequently discovered concept", *J. Combinatorial Theory* (A) 13 (1972), 297-305.

[Lal 79] G. Lallement, *Semigroups and Combinatorial Applications*, John Wiley and Sons, New York, 1979.

[My 57] J. Myhill, "Finite automata and the representation of events," Wright Air Development Command Technical Report 57-624 (1957) 112-137.

[Nr 58] A. Nerode, "Linear automaton transformations," *Proc. American Math. Soc.* 9(1958), 541-544.

[Sha, Mc 56] C. F. Shannon and J. McCarthy (Eds.), *Automata Studies*, Princeton University Press, Princeton, 1956.