A GENERALIZATION OF A THEOREM OF GRAHAM HIGMAN

by

Andrzej Ehrenfeucht and David Haussler

Department of Computer Science
University of Colorado
Boulder, Colorado 80309

CU-CS-219-81 Dec. 1981

This work supported in part by NSF Grant MCS 79-03838 and the University of Colorado Doctoral Fellowship Program.

ABSTRACT

Given a finite alphabet Σ and a set of words $S \subseteq \Sigma^+$, I_S is the relation on Σ^* defined by $x\ I_S\ y$ if and only if there exist $x_1, x_2 \in \Sigma^*$ and $w \in S$ such that $x = x_1x_2$ and $y = x_1wx_2$. The partial order $\leq_S$ on Σ^* is defined as the reflexive transitive closure of I_S. In the special case where $S = \Sigma$, $\leq_S$ is the subsequence relation on Σ^*. Higman has shown that $\leq_\Sigma$ is a well partial order on Σ^* for any finite alphabet Σ. We generalize this result by showing that for arbitrary $S \subseteq \Sigma^+$, $\leq_S$ is a well partial order on Σ^* if and only if there exists a k_0 such that for each word $x \in \Sigma^*$ of length greater than k_0 there exist $x_1, x_2 \in \Sigma^*$ and $w \in S$ such that $x = x_1wx_2$.

INTRODUCTION

Graham Higman's paper [Hig, 52] contains the following corollary. If we let Σ^* be the free monoid generated by the finite alphabet Σ and let $\leq$ be the subsequence relation on Σ^*, i.e., for $x,y \in \Sigma^*$, $x \leq y$ if and only if there exist $a_i, \cdots, a_k \in \Sigma$ and $u_1, \cdots, u_{k+1} \in \Sigma^*$ such that $x = a_1 \cdots a_k$ and $y = u_1 a_1 u_2 a_2 \cdots u_k a_k u_{k+1}$, then $\leq$ is a well partial order on Σ^*, that is, every set of words in Σ^* has a finite and nonempty subset of minimal words with respect to $\leq$. This implies of course that any set of words in Σ^* which is pairwise incomparable with respect to $\leq$ is finite. This result has been rediscovered [HAI 69] and generalized [KRU 60, 72], [LAU 76] repeatedly in the ensuing years as the theory of well partial orders (or more generally, well quasi orders) has developed.

Let us define a more general class of partial orders on Σ^* which includes the subsequence relation. We begin by defining the relation of simple insertion as follows. For any set $S \subseteq \Sigma^*$ and $x,y \in \Sigma^*$, x is related to y by an insertion from S if and only if $x = x_1 x_2$ and $y = x_1 w x_2$ for some $x_1, x_2 \in \Sigma^*$ and $w \in S$. For a given set S, this relation is denoted I_S. By taking the reflexive and transitive closure of I_S, we obtain a partial order on Σ^* which we will denote $\leq_S$. Thus $x \leq_S y$ whenever y can be obtained from x by repeated insertions of words from S.

Taking S to be Σ we obtain the subsequence relation on Σ^*, hence $\leq_\Sigma$ is a well partial order on Σ^*. This is not true for arbitrary $S \subseteq \Sigma^*$. For instance, if we let $\Sigma = \{ (,), [,] \}$ and $S = \{ (), [] \}$ then $x \leq_S y$ if and only if y is obtained from x by inserting strings of well balanced parenthesis of type () and [] (with arbitrary nesting) between the letters of x. (Here λ, the empty string, is considered well balanced). It is easily verified that $\leq_S$ is not a well partial order on Σ^* by observing that all of the strings in the set (* are pairwise incomparable with respect to $\leq_S$. On the other hand, Σ is by no means the only set which defines a well partial order on Σ^* via repeated insertion. If we let $\Sigma = \{a, b\}$ and $S = \{aa, ab, ba, bb\}$ then $x \leq_S y$ if and only if y is obtained from x by inserting strings of even length between the letters of x. In this case it is not hard to show that $\leq_S$ is a well partial order on Σ^*. It is somewhat more difficult to show that $\leq_S$ remains a well partial order on Σ^* if we omit either the string ab or the string ba from S.

In this paper we characterize those sets $S \subseteq \Sigma^*$ such that $\leq_S$ is a well partial order on Σ^*. The property which characterizes these sets is called "subword unavoidability" and is defined as follows. A set $S \subseteq \Sigma^* - \{\lambda\}$ is subword unavoidable in Σ^* if and only if there exists a finite k_0 such that for each word $y \in \Sigma^*$ of length greater than k_0, y has a subword in S, i.e. $x \; I_S \; y$ for some $x \in \Sigma^*$. Since Σ is obviously subword unavoidable in Σ^*, Higman's result follows directly from this characterization theorem.

This result can be used in conjunction with a result from [HAU1 81] to show that certain languages generated by repeated insertions are regular languages. Let us say that a partial order $\leq$ on Σ^* is monotone whenever $x \leq x'$ and $y \leq y'$ implies that $xy \leq x'y'$. Given any set $T \subseteq \Sigma^*$ and partial order $\leq$, we define the closure of T under $\leq$ as $\{w : x \leq w$ for some $x \in T\}$. From the main theorem of [HAU1 81] it follows that for any monotone well partial order $\leq$ on Σ^* and set $T \subseteq \Sigma^*$, the closure of T under $\leq$ is a regular language. Since $\leq_S$ is a monotone partial order for any $S \in \Sigma^*$, the results of this paper imply that the closure of T under $\leq_S$ is a regular language for any $T \subseteq \Sigma^*$ and subword unavoidable $S \subseteq \Sigma^* - \{\lambda\}$. It is not difficult to show that the closure of any finite set of nonempty words under $\leq_S$ will not be a regular language unless S is subword unavoidable (see [HAU2 81]), hence the property of subword unavoidability can be used to characterize those sets which generate regular langauges from finite bases via repeated insertion.

BASICS

Higman gives following definitions of a well quasi order (among others) and proves them equivalent [HIG 52].

Definition: A ***quasi order*** is a reflexive and transitive relation. Given a quasi order $\leq$ on a set S, $\leq$ is a ***well quasi order*** on S (or a ***well partial order*** if $\leq$ is a partial order) if and only if any of the following hold:

i) $\leq$ is well founded on S, i.e., there exist no infinite strictly descending sequences of elements in S, and each set of pairwise incomparable elements is finite.

ii) For each infinite sequence $\{x_i\}$ of elements in S there exist $i<j$ such that $x_i \leq x_j$.

iii) Each infinite sequence of elements in S contains an infinite ascending subsequence.

An obvious property of well quasi orders which will be useful later is the following.

Proposition 1. If $\leq_1$ is a well quasi order on the set S and $\leq_2$ is an extension of $\leq_1$ which is also a quasi order, then $\leq_2$ is a well quasi order on S.

Definition. Given sets S_1 and S_2 and relations R_1 and R_2 on S_1 and S_2 respectively, the relation $R_1 \times R_2$ on $S_1 \times S_2$ is defined by $<a,b> R_1 \times R_2 <c,d>$ if and only if $a \; R_1 \; c$ and $b \; R_2 \; d$.

Another easy consequence of the above definitions of a well quasi order is the following proposition.

Proposition 2. Given sets S_1, S_2 and well quasi orders $\leq_1$ and $\leq_2$ on S_1 and S_2 respectively, the transitive closure of $\leq_1 \bigcup \leq_2$ is a well quasi order on $S_1 \bigcup S_2$ and $\leq_1 \times \leq_2$ is a well quasi order on $S_1 \times S_2$.

One of the earliest results of the theory of well quasi orders is the following, apparently discovered independently by Higman, Neumann and Erdos and Rado around 1950. (See note at the end of [Erd and Rad 52]).

Definition. For any set S, $S^{<\omega}$ is the set of finite sequences of elements of S. Given a set S and a quasi order $\leq$ on S, the ordering $\leq^E$ on $S^{<\omega}$ is defined by $<s_1, \cdots, s_k> \leq^E <t_1, \cdots, t_l>$ if and only if there exists a subsequence $<t_{i_1}, \cdots, t_{i_k}>$ of $<t_1, \cdots, t_l>$ such that $s_j \leq t_{i_j}$ for $1 \leq j \leq k$.

Proposition 3. If $\leq$ is a well quasi order on S then $\leq^E$ is a well quasi order on $S^{<\omega}$.

See [LAV 76] for a very short proof of this result.

Definition. Throughout this paper Σ^* will denote the free monoid with null word λ generated by the finite alphabet Σ. $\Sigma^+ = \Sigma^* - \{\lambda\}$. We will extend the operation of concatenation to subsets of Σ^* in the natural manner, for $S_1, S_2 \subseteq \Sigma^*$, $S_1 S_2 = \{xy : x \in S_1 \text{ and } y \in S_2\}$. In the case that one of the sets is a singleton, say $S_1 = \{x\}$, then $S_1 S_2$ may be denoted xS_2.

Definition. A relation R on Σ^* is *monotone* if and only if for all $x, x', y, y' \in \Sigma^*$, $x \; R \; x'$ and $y \; R \; y'$ implies that $xy \; R \; x'y'$.

Lemma 1. Given $S_1, S_2 \subseteq \Sigma^*$ and a monotone quasi order $\leq$ on Σ^*, if $\leq\times\leq$ is a well quasi order on $S_1 \times S_2$ then $\leq$ is a well quasi order on $S_1 S_2$.

Proof. Let $\{x_i y_i\}$ be an infinite sequence of words in $S_1 S_2$ where for all i, $x_i \in S_1$ and $y_i \in S_2$. Since $\leq\times\leq$ is a well quasi order on $S_1 \times S_2$, we can find i, j such that $i < j$ and $<x_i, y_i> \leq\times\leq <x_j, y_j>$, i.e., $x_i \leq x_j$ and $y_i \leq y_j$. Since $\leq$ is monotone, this implies that $x_i y_i \leq x_j y_j$. Thus $\leq$ is a well quasi order on $S_1 S_2$.

Lemma 2. Given $S \subseteq \Sigma^*$ and a monotone quasi order $\leq$ on Σ^* where $\lambda \leq x$ for all $x \in S$, if $\leq^E$ is a well quasi order on $S^{<\omega}$ then $\leq$ is a well quasi order on S^*.

Proof. Let $\{u_{i,1} \cdots u_{i,k_i}\}$ be an infinite sequence of words in S^* where $u_{i,n} \varepsilon S$ for all i and all n, $1 \le n \le k_i$. Since $\le^E$ is a well quasi order on $S^{<\omega}$, we can find i,j such that $i < j$ and $<u_{i,1}, \cdots, u_{i,k_i}> \le^E <u_{j,1}, \cdots, u_{j,k_j}>$. Hence there exists a subsequence $<u_{j,l_1}, \cdots, u_{j,l_{k_i}}>$ of $<u_{j,1}, \cdots, u_{j,k_j}>$ such that $u_{i,n} \le u_{j,l_n}$ for $1 \le n \le k_i$. Since $\lambda \le x$ for all $x \varepsilon S$, this implies that $u_{i,1} \cdots u_{i,k_i} \le u_{j,1} \cdots u_{j,k_j}$ by monotonicity. Hence $\le$ is a well quasi order on S^*.

Since the subsequence relation $\le$ on Σ^* is monotone and for all $\boldsymbol{a} \varepsilon \Sigma$, $\lambda \le \boldsymbol{a}$, the Higman result cited in the introduction can easily be derived from Proposition 3 using Lemma 2.

MAIN RESULT

In this section we define the partial orders on Σ^* generated by repeated insertion and characterize those insertion sets which generate well partial orders on Σ^*.

Definition. Given $S \subseteq \Sigma^+$ and $x, y \in \Sigma^*$, $x \, I_S \, y$ if and only if there exist $x_1, x_2 \in \Sigma^*$ and $w \in S$ such that $x = x_1 x_2$ and $y = x_1 w x_2$. $\leq_S$ denotes the reflexive transitive closure of I_S. For $u, v \in \Sigma^*$, a *derivation of v from u by $\leq_S$* is a finite sequence of words $<x_1, \cdots, x_k>$ where $k \geq 1$ such that $x_1 = u$, $x_k = v$ and for $1 \leq i < k$, $x_i \, I_S \, x_{i+1}$.

Lemma 3. Given $S \subseteq \Sigma^+$ and $u, v \in \Sigma^*$

i) $\leq_S$ is a partial order and

ii) there exists a derivation of v from u by $\leq_S$ if and only if $u \leq_S v$.

Proof. This is obvious.

Definition. Given a set $S \subseteq \Sigma^+$, S is *subword unavoidable* in Σ^* if and only if there exists a k_0 such that for all words $x \in \Sigma^*$ longer than k_0 there exist $x_1, x_2 \in \Sigma^*$ and $w \in S$ such that $x = x_1 w x_2$. The smallest such k_0 is called the *subword avoidance bound* for S.

Lemma 4. If $S \subseteq \Sigma^+$ is subword unavoidable in Σ^* with subword avoidance bound k_0 then there exists a finite $F \subseteq S$ such that F is subword unavoidable in Σ^* with subword avoidance bound k_0.

Proof. Let $S \subseteq \Sigma^+$ be subword unavoidable in Σ^* and k_0 be the subword avoidance bound for S. Then any word of length k_0+1 has a subword in S, and this subword must have length k_0+1 or less. Thus any word longer than k_0 has a subword in the subset of S of words of length k_0+1 or less. Thus this set is subword unavoidable in Σ^* with subword avoidance bound k_0.

Lemma 5. Given $S \subseteq \Sigma^+$, if S is not subword unavoidable in Σ^* then $\leq_S$ is not a well partial order on Σ^*.

Proof. If S is not subword unavoidable in Σ^* then there exists an infinite set of words $T \subseteq \Sigma^*$ such that for any $x \in T$ there exist no $x_1, x_2 \in \Sigma^*$ and $w \in S$ such that $x = x_1 w x_2$. Hence for no $x, y \in T$ can there be a derivation of y from x by $\leq_S$. Thus the words in T are pairwise incomparable under $\leq_S$ and hence $\leq_S$ is not a well partial order on Σ^*.

Definition. For each $S \subseteq \Sigma^+$ let

$S_0 = S^*$

and $S_{n+1} = \left[\bigcup_{a_1 \cdots a_k \in S \cup \{\lambda\}} S_n a_1 S_n a_2 \cdots S_n a_k S_n \right]^*$.

Lemma 6. For any set $S \subseteq \Sigma^+$ and $n \geq 0$,

i) if $uv \in S_n$ and $w \in S$ then $uwv \in S_{n+1}$,

ii) if $uv \in S_n$, where the number of letters in u is less than or equal to n, and $w \in S$ then $uwv \in S_n$ and

iii) if S is finite then $\leq_S$ is a well partial order on S_n.

Proof.

ad. (i). This is obvious.

ad. (ii). Here we use induction on n. If $n = 0$ then we need only consider the case $u = \lambda$ and the statement follows from the fact that $S_0 = S^*$. Now let us assume that the statement holds for some $n \geq 0$. If $uv \in S_{n+1}$ then $uv = w_1 a_1 w_2 a_2 \cdots w_k a_k w_{k+1}$ where $w_i \in S_n$ for $1 \leq i \leq k+1$ and $a_1 \cdots a_k \in S^*$. Hence for some i, $1 \leq i \leq k+1$, $u = w_1 a_1 \cdots w_{i-1} a_{i-1} w_i'$ and $v = w_i'' a_i \cdots w_k a_k w_{k+1}$ where $w_i', w_i'' \in \Sigma^*$ and $w_i' w_i'' = w_i$. For any $w \in S$, $w_i' w w_i'' \in S_{n+1}$ by part (i). Thus if $i = 1$, then $uwv \in S_{n+1}$ because $a_1 w_2 \cdots a_k w_{k+1} \in S_{n+1}$ and S_{n+1} is closed under concatenation. On the other hand, it is apparent that if $i > 1$ and u has at most $n + 1$ letters, w_i' has at most

n letters. Thus by the inductive hypothesis, for any $w \in S$, $w_i'ww_i'' \in S_n$. But this implies that $uwv \in S_n a_1 \cdots S_n a_k S_n$, thus $uwv \in S_{n+1}$. Thus the statement holds for $n+1$ and the result follows by induction.

ad. (iii). Again we use induction on n. Since S is a finite set, $\leq_S$ is a well partial order on S. Hence by Proposition 3, $\leq_S^F$ is a well partial order on $S^{<\omega}$. Now since $\leq_S$ is a monotone partial order on Σ^* and $\lambda \leq_S w$ for all $w \in S$, $\leq_S$ is a well partial order on S^* by Lemma 2. Thus the statement holds for the case $n = 0$. Let us suppose this statement holds for some $n \geq 0$. Using Proposition 2 part (ii) and Lemma 1 we have that $\leq_S$ is a well partial order on $S_n a_1 \cdots S_n a_l S_n$ for any $a_1 \cdots a_l \in \Sigma^*$. Furthermore, if $a_1 \cdots a_l \in S \cup \{\lambda\}$ then for any $w \in S_n a_1 \cdots S_n a_l S_n$, $\lambda \leq_S w$. Also, since S is finite, $\leq_S$ is a well partial order on $T_n = \bigcup_{a_1 \cdots a_l \in S \cup \{\lambda\}} S_n a_1 \cdots S_n a_l S_n$ using Proposition 2 part (i). Thus using Proposition 3, $\leq_S^F$ is a well quasi order on $T_n^{<\omega}$ and hence by Lemma 2 $\leq_S$ is a well partial order on T_n^*. Furthermore, $\lambda \leq x$ for all $x \in T_n^*$. Since $T_n^* = S_{n+1}$, we have shown that the statement holds for $n+1$. The result follows by induction.

Definition. Given $S \subseteq \Sigma^+$, for each $n \in N$

let $R(S_n) = \bigcup_{a_i, \cdots, a_k \in \Sigma, k \leq n} S_n a_1 S_n a_2 \cdots S_n a_k S_n$.

Lemma 7. For any $S \subseteq \Sigma^+$,

(i) if S is finite then $\leq_S$ is a well partial order on $R(S_n)$ for all n,

(ii) $\{R(S_n)\}$ is an ascending sequence of sets such that $\Sigma^* = \bigcup_{n=1}^{\infty} R(S_n)$ and

(iii) If S is subword unavoidable in Σ^* and k_0 is the subword avoidance bound for S, then $\Sigma^* = R(S_{k_0})$.

Proof.

ad. (i). This follows from Proposition 2 and Lemma 1 using Lemma 6 part (iii).

ad. (ii). This is obvious.

ad. (iii). Assume to the contrary that $\Sigma^* - R(S_{k_0}) \neq \phi$. Let x be among the shortest words in $\Sigma^* - R(S_{k_0})$. Since $R(S_{k_0})$ contains all words of length k_0 or less, x must be longer than k_0 letters. Since k_0 is the subword avoidance bound for S, we can find among the first k_0+1 letters of x a subword in S. Thus $x = uwv$ where $w \varepsilon S$ and u has k_0 or fewer letters. Since x was of minimal length, $uv \varepsilon R(S_{k_0})$. Hence $uv = w_1 a_1 \cdots w_k a_k w_{k+1}$ where $a_i \varepsilon \Sigma$ for $1 \leq i \leq k$ and $w_i \varepsilon S_{k_0}$ for $1 \leq i \leq k+1$. Find i such that $u = w_1 a_1 \cdots w_i'$, $v = w_i'' a_i \cdots w_k a_k w_{k+1}$ and $w_i' w_i'' = w_i$. Now by Lemma 6 part (ii) $w_i' w w_i'' \varepsilon S_{k_0}$, since the number of letters in w_i' is less than or equal to k_0. Hence $x = uwv$ is in $R(S_{k_0})$ contrary to hypothesis.

Theorem 1. Given a set $S \subseteq \Sigma^*$, $\leq_S$ is a well partial order on Σ^* if and only if $S - \{\lambda\}$ is subword unavoidable in Σ^*.

Proof. Obviously $\leq_{S-\{\lambda\}} = \leq_S$. Thus by Lemma 5, if $\leq_S$ is a well partial order on Σ^* then $S - \{\lambda\}$ is subword unavoidable in Σ^*. On the other hand, given any $S \subseteq \Sigma^+$ which is subword unavoidable in Σ^* with subword avoidance bound k_0, by Lemma 4 there exists a finite $F \subseteq S$ such that F is subword unavoidable in Σ^* with subword avoidance bound k_0 . By Lemma 7 parts (i) and (iii), $\leq_F$ is a well partial order on $R(F_{k_0}) = \Sigma^*$. Thus since $\leq_S$ is an extension of $\leq_F$, $\leq_S$ is a well partial order on Σ^* by Proposition 1.

REFERENCES

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

[HAI 69] L. H. Haines, "On free monoids partially ordered by embedding," *J. Combinatorial Theory*, 6 (1969), 94-98.

[HAU1 81] D. Haussler, "A generalization of the Nerode characterization of the regular sets using well quasi orderings," submitted to ???.

[HAU2 81] D. Haussler, "Insertion languages: a class of generalized Dyck languages," submitted to ???.

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

[KRU 60] J. B. Kruskal, "Well-quasi-ordering, the tree theorem, and Vazsonyi's conjecture," *Trans. Amer. Math. Soc.* 95 (May 1960), 210-225.

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

[LAV 76] R. Laver, "Well quasi orderings of sets of finite sequences," *Math Proc. Cambridge Philos. Soc.* 79 (1976), 1-10.

ANY OPINIONS, FINDINGS, AND CONCLUSIONS OR RECOMMENDATIONS EXPRESSED IN THIS PUBLICATION ARE THOSE OF THE AUTHOR AND DO NOT NECESSARILY REFLECT THE VIEWS OF THE NATIONAL SCIENCE FOUNDATION.

THIS MATERIAL IS BASED UPON WORK SUPPORTED BY THE NATIONAL SCIENCE FOUNDATION UNDER GRANT NO. MCS 79-Ø3838.