BIT 28 (1988), 755–784

ENCROACHING LISTS AS A MEASURE OF PRESORTEDNESS

STEVEN S. SKIENA[1]

Department of Computer Science, University of Illinois, Urbana, Illinois 61801, USA

Abstract.

Encroaching lists are a generalization of monotone sequences in permutations. Since ordered permutations contain fewer encroaching lists than random ones, the number of such lists m provides a measure of presortedness with advantages over others in the literature. Experimental and analytic results are presented to cast light on the properties of encroaching lists. Also, we describe a new sorting algorithm, *melsort*, with complexity $O(n \log m)$. Thus it is linear for well ordered sets and reduces to mergesort and $O(n \log n)$ in the worst case.

CR category: F.2.2. *AMS category:* 68R05.

Keywords: sorting, presortedness, encroaching lists, melsort, permutations.

1. Introduction.

A measure of *presortedness* is an integer function on a permutation σ of a totally ordered set reflecting how much σ differs from the total order. Examples of presortedness measures include the number of runs or inversions in σ. This paper describes a new measure of presortedness based on the size of a combinatorial structure and uses it to define an efficient sorting algorithm, *melsort*.

Let us consider an *ordered list* to be an ordered set of elements. The head of an ordered list l, $head(l)$, is the smallest element in l, and similarly we can define the $tail(l)$ to be the largest element in l. An *encroaching list set* is an ordered set of ordered lists $l_1, \ldots, l_m$ such that $head(l_i) \leq head(l_{i+1})$ and $tail(l_i) \geq tail(l_{i+1})$ for $1 \leq i < m$. Thus the lists nest or *encroach* upon one another. For convenience, we will call $l_1, \ldots, l_m$ *sublists* of the encroaching list set.

In this paper, we give a heuristic for extracting encroaching list sets from a permutation. The size of the extracted encroaching list set, m, represents an

[1] Current address: Department of Computer Science, State University of New York, Stony Brook, NY 11794. This work was partially supported by National Science Foundation Grant CCSR-8714565.

Received January 1988. Revised June 1988.

interesting measure of presortedness since permutations which appear to be well ordered have smaller list set sizes than more random ones. We then define a sorting algorithm with complexity $O(n \log m)$. For completely or nearly completely ordered data it is $O(n)$, while in worst case and average cases it is $O(n \log n)$ as it reduces to mergesort. It may prove a practical sorting algorithm, particularly if there has been some previous attempt at ordering the data. Finally, we consider interesting combinatorial questions on the size and distribution of encroaching lists, and provide evidence that encroaching list size is a better measure of presortedness than others in the literature.

2. Encroaching lists.

Encroaching lists can be described by an algorithm which constructs them. Consider a permutation σ on an ordered set. We can accumulate ordered lists with the following strategy: if the item σ_i fits on either end of an existing sublist, put it there; otherwise form a new list. Place each item on the oldest sublist which it fits on. Initially, we start with just one sublist consisting of item σ_1.

An example will make this process clear. Consider the permutation of elements $\sigma = (4, 6, 5, 2, 9, 1, 3, 8, 0, 7)$. Initially, l_1 consists of 4, and the second elements fits at the tail of it. 5 is between 4 and 6, so l_2 is created to hold it. The next three elements all fit on the oldest sublist and are placed there: 3 fits at the end of l_2 and 0 on l_1, but the last element requires a new sublist. The final sublists:

$$l_1 : 012469$$

$$l_2 : 358$$

$$l_3 ; 7$$

Notice that the oldest fitting list heuristic maintains the relative ordering of the list heads and tails as sorted without additional work.

This algorithm is similar in spirit to the bumping procedure associated with Young tableaux, and in fact Young tableaux will prove helpful in the obvious question of how many encroaching lists result. Knuth [1] defines a *Young tableau of shape* $(n_1, n_2, \ldots, n_m)$, where $n_1 \geq n_2 \cdots \geq n_m \geq 0$, as an arrangement of $n_1 + n_2 + \cdots + n_m$ distinct integers in an array of left-justified rows, with n_i elements in row i, such that the entries of each row are in increasing order from left to right, and the entries of each column are increasing from top to bottom. The similarities between encroaching lists and Young tableaux result from both being an ordered set of ordered lists, although the number of elements in an encroaching list does not decrease monotonically and the ordering of the tail elements is different with Young tableaux. We will exploit these similarities in many of our proofs.