Hirtle, S. C. (1990). Ordered trees: A structure for the mental representation of information. Proceedings of the 1st ASIS SIG/CR Classification Research Workshop, 79-86. doi: 10.7152/acro.v1i1.12466

Ordered Trees: A Structure for the Mental Representation of Information

Stephen C. Hirtle

Interdisciplinary Department of Information Science
University of Pittsburgh, Pittsburgh, PA 15260, USA

ABSTRACT

This paper reviews the ordered tree technique for deriving a structured mental representation, and discusses potential applications for information science. The ordered tree clustering algorithm that we are developing provides an alternative to more traditional approaches of clustering based on similarity matrices. This algorithm is based on patterns of linear strings, typically free-recall orders, but potentially any material organized in a linear fashion. The technique has proven useful in representing ordered classifications, such as an alphabet or a numerical sequence, imbedded within a hierarchical semantic structure. Furthermore, the ordered tree technique, its corresponding data structures, and the data analysis procedures, can be generalized to include cross-classification used by human subjects and the classification of subjects themselves into categories such as experts and novices.

1. MENTAL REPRESENTATIONS AND INFORMATION SCIENCE

The need for sophisticated scaling techniques in the field of information science is becoming increasingly clear as information systems move from simple data storage devices to complex information handlers. In particular, there is little doubt that an information system becomes more efficient or easier to use when the imbedded classification structures used by the system resemble those that users of the system have for the material. It is posited that this is true not only at the global level, but that it is true even at lower levels of system use, such as in menu hierarchies and command structures. For example, recent work by Fisher, Yungkruth, and Moss (1989) suggests that retrieval times will be greatly reduced if a menu is constructed to be consistent with the logical semantics of the domain.

This paper discusses a technique, called the *ordered tree clustering algorithm*, which was developed by Reitman and Rueter (1980). The ordered tree clustering algorithm is useful for developing classifications for material where order is an important component of the classification scheme. This technique was first developed for use with free-recall data within the field of cognitive science, but it has strong potential for applications within the field of information science as well. This technique is described, in detail, in this paper. In particular, the next section describes the raw data of free-recall used to construct the trees. The third section describes an alternative scaling method and its inherent drawbacks. The fourth section describes the ordered tree technique in detail. Finally, the last section proposes some extensions of the technique for the field of information science.

2. FREE-RECALL ORDERS

The first step in deriving a classification is to collect appropriate data to measure the similarity between terms. Within the field of psychology, these measures have included pairwise

ISSN: 2324-9773

Hirtle, S. C. (1990). Ordered trees: A structure for the mental representation of information. Proceedings of the 1st ASIS SIG/CR Classification Research Workshop, 79-86. doi: 10.7152/acro.v1i1.12466

similarity measures, such as direct judgments or confusion indices, and order information such as the order in which items are recalled in a free-recall task (Reitman & Rueter, 1980). Having collected such raw data, further processing is required before use by an information system. Thus, the second step is to represent the information in the data by a simplified structure, such as a hierarchical tree, a Euclidean space, or some other representation.

Many sophisticated techniques, such as multidimensional scaling and hierarchical cluster analysis, have been developed for use with pairwise similarity measures (see Baird & Noma, 1978, for further discussion of these techniques). However, relatively little research has been done with the problem of accounting for the rich information available in free-recall orders. Yet, free-recall orders are easy to collect, and can represent not only clusters of information, but the order of information, as well.

For example, consider the problem of recalling the provinces of Canada, repeatedly. A person would most likely be neither random nor completely consistent in the recall. Rather, permutations from one recall to another would reflect mental clusters. In addition, the ordering within the recalls would reflect any ordered mental information. Such data might be collected in an experimental setting simply by asking a subject to recall from memory a fixed set of items several times. Using the Canadian province example, a hypothetical subject might recall:

1. NFLD NS PEI NB QUE ONT MAN SASK ALTA BC YK
2. NFLD NS PEI NB QUE ONT MAN SASK ALTA BC YK
3. PEI NS NB NFLD QUE ONT MAN SASK ALTA YK BC
4. YK BC QUE ONT MAN SASK ALTA NB PEI NS NFLD
5. BC YK QUE ONT MAN SASK ALTA NS NB PEI NFLD

Inspection of these recalls indicates an eastern cluster consisting of Newfoundland and the maritime provinces, a western cluster of British Columbia and the Yukon, and a strongly ordered cluster across the mainland of Canada from Quebec to Alberta.

In the experiments reported further on, multi-trial free-recall is the primary dependent measure. Subjects were asked to recall a set of items ranging in size from about 10 items to 50 items, and they were asked to recall the set in the range of 14 times to 30 times, depending on the experiment. In addition, subjects on certain trials were cued with certain items and asked to begin with that item. Cuing was used to encourage permutation of the orders and these trials were analyzed in a slightly different fashion, which will not be discussed here. Finally, if the material was new to the subjects, they were required to memorize the material first, before beginning the free-recalls which were to be analyzed.

The remaining sections of this paper will discuss techniques for analyzing free-recall orders, first discussing the use of symmetric distance matrices, and its inherent drawbacks, followed by our ordered tree technique. All these techniques depend on the assumption that items close in a recall order are conceptually close in the subject's mental representation.

ISSN: 2324-9773

Hirtle, S. C. (1990). Ordered trees: A structure for the mental representation of information. Proceedings of the 1st ASIS SIG/CR Classification Research Workshop, 79-86. doi: 10.7152/acro.v1i1.12466

3. SYMMETRIC DISTANCE MATRIX TECHNIQUES

Friendly (1977) discussed several methods of creating a pairwise dissimilarity matrix from free-recall orders. For example, one could take the average number of intervening items in a set of recalls as a measure of distance. Items that were contiguous on *n* trials would be a distance of *n* apart, whereas items more disparate would be assigned distance values of greater than *n*. Using the earlier example free-recall orders and this measure, the resulting matrix of dissimilarities would be:

NS	5.0									
PEI	16.0	15.0								
NB	17.0	16.0	5.0							
QUE	18.0	17.0	10.0	5.0						
ONT	19.0	18.0	15.0	10.0	5.0					
MAN	20.0	19.0	20.0	15.0	10.0	5.0				
SASK	35.0	34.0	19.0	18.0	17.0	16.0	15.0			
ALTA	41.0	40.0	25.0	24.0	23.0	22.0	21.0	6.0		
BC	41.0	40.0	25.0	24.0	23.0	22.0	21.0	8.0	6.0	
YK	45.0	44.0	29.0	28.0	27.0	26.0	25.0	12.0	10.0	8.0
	YK	BC	ALTA	SASK	MAN	ONT	QUE	NB	PEI	NS

The complete link clustering solution using this matrix is shown in Figure 1. Note that order information is lost and that some items, such as Ontario and Quebec, appear in non-appropriate clusters.

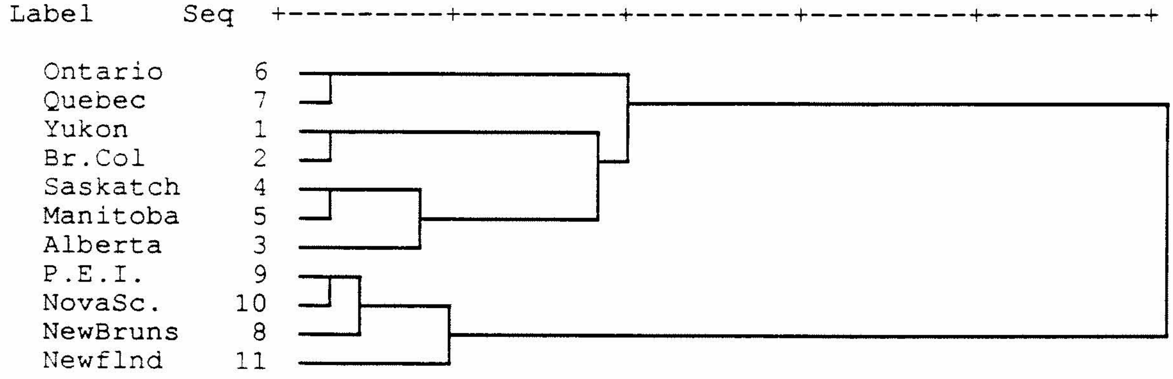


Figure 1. Dendrogram using complete linkage clustering algorithm.

4. ORDERED TREE TECHNIQUE

The ordered tree clustering algorithm developed by Reitman and Rueter (1980) is an alternative to symmetric distance models. This algorithm for producing ordered trees based on multi-trial free recall orders has been implemented as a stand-alone program called TIGER, written in RATFOR, and available from the author in an executable form for the IBM PC.

ISSN: 2324-9773

Hirtle, S. C. (1990). Ordered trees: A structure for the mental representation of information. Proceedings of the 1st ASIS SIG/CR Classification Research Workshop, 79-86. doi: 10.7152/acro.v1i1.12466

The algorithm proceeds top-down, examining the set of recall orders for consistent chunks, that is, items that are consistently recalled together without intervening items. Again, referring to the example we see that {NFLD, NS, PEI, NB} occurs together in some order in each recall. In trials 1 and 2, it appears in the order: NFLD, NS, PEI, NB. In trial 3, it appears in the order: PEI NS NB NFLD. In trial 4, it appears in the order: NB PEI NS NFLD. In trial 5, it appears in the order: NS NB PEI NFLD. Likewise, {NS, PEI, NB} occurs together in each recall. In contrast, the set of items {NFLD, NS, PEI} is not a chunk, as it does not appear without intervening items in either the third, fourth, or fifth recall orders.

The set of all chunks derived from the recall orders can be written into the form of a poset, or partially ordered set. Using the example recalls, the set of all chunks is shown in Figure 2.

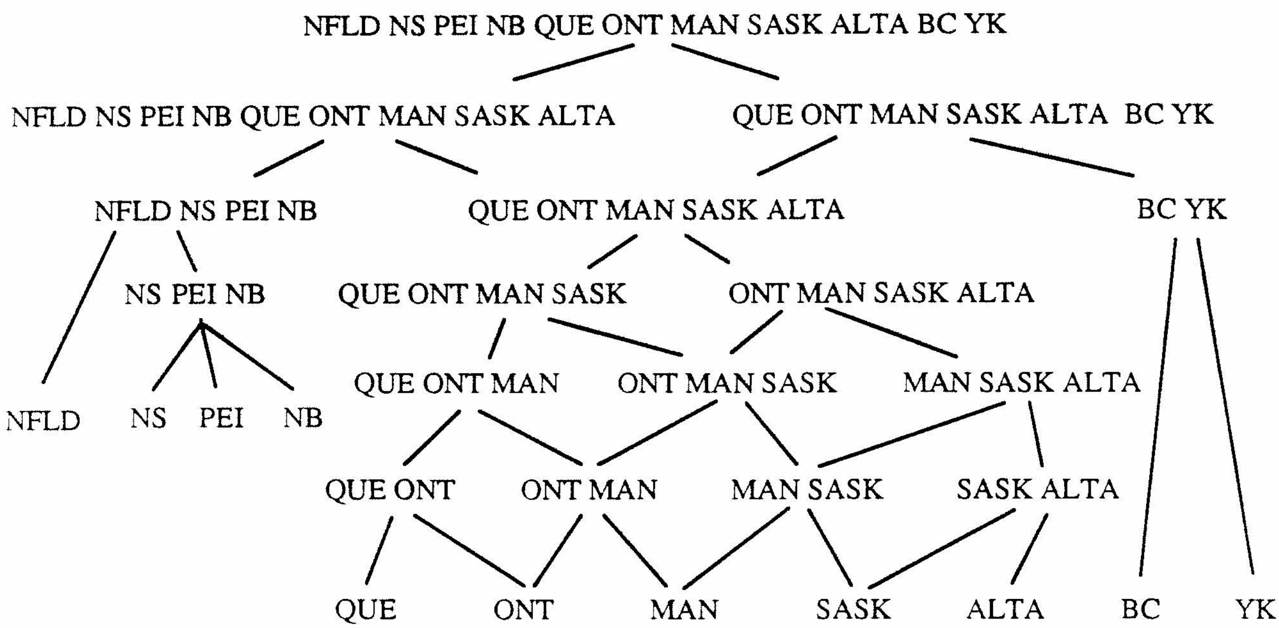


Figure 2. Set of all chunks derived from the example set of recall orders.

The final step is to rewrite the poset as a tree. This final representation is called an ordered tree, as the nodes at any level of the tree can be ordered as either a unidirectional or bidirectional node, or unordered, as a nondirectional node. The ordered tree for the current example is shown in Figure 3. The unidirectional node is indicated by a single-headed arrow, the bidirectional node is indicated by a double-headed arrow, and the nondirectional nodes are indicated by a absence of an arrow. Thus, you will note that the tree represents both the hierarchical clustering and the order information.

ISSN: 2324-9773

Hirtle, S. C. (1990). Ordered trees: A structure for the mental representation of information. Proceedings of the 1st ASIS SIG/CR Classification Research Workshop, 79-86. doi: 10.7152/acro.v1i1.12466

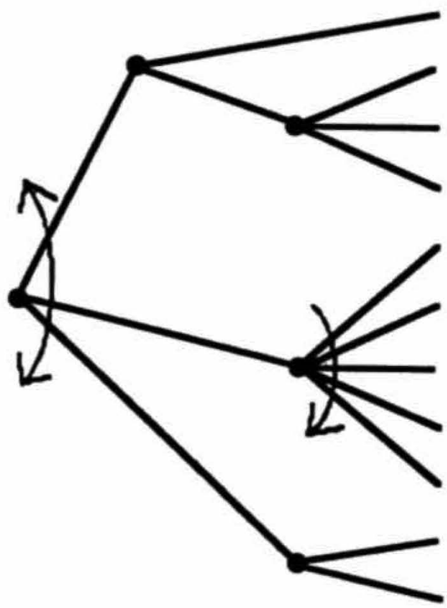


Figure 3. Ordered tree for example set of recall orders.

4.1 APPLICATIONS OF ORDERED TREES

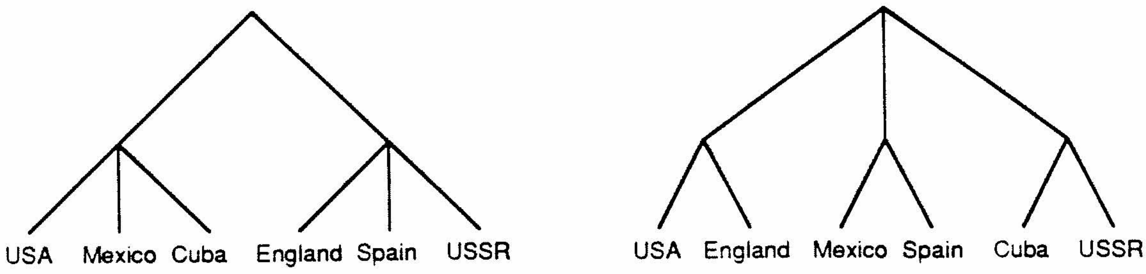
The ordered tree algorithm has been applied to a number of domains within the field of cognitive science, including representing the expert knowledge of computer programmers (McKeithen, Reitman, Rueter, & Hirtle, 1981), representing spatial knowledge (Hirtle & Jonides, 1985; McNamara, Hardy, & Hirtle, 1989) and representing knowledge of classroom material (Naveh-Benjamin, McKeachie, Lin, & Tucker, 1986). Potential applications for information science is discussed in the last section.

4.2 CROSS-CLASSIFICATIONS

In addition to modelling a single hierarchical tree, it is possible to represent multiple trees, or cross-classifications, with ordered tree structures. For example, it is possible that no overall structure exists in the set of recall orders as a whole, but rather the subject accesses one of several memory representations or dimensions of representation. Such strategies may be likely in complex domains where cross-classification is common (e.g., Broadbent, Cooper, & Broadbent, 1978). As a small, but concrete, example, consider the two trees S and T shown in Figure 4. One tree (S) represents a geographical organization of 6 countries, whereas the other tree (T) represents a socio-political organization. Given the two organizations, we assume that any one free-recall order is drawn from a single tree. If the entire set of recalls was subjected to the ordered tree analysis, no structure would emerge as there are no consistent chunks across both trees.

ISSN: 2324-9773

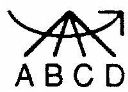
Hirtle, S. C. (1990). Ordered trees: A structure for the mental representation of information. Proceedings of the 1st ASIS SIG/CR Classification Research Workshop, 79-86. doi: 10.7152/acro.v1i1.12466



(S) (T)

Figure 4. Example of a cross-classification of country names.

A technique has been developed recently to parse a set of recall orders into distinct sets, each representing a separate tree structure (Hirtle, 1988; Hirtle & Crawley, 1989). This parsing is accomplished by considering each recall order to be a strict unidimensional tree containing a single recall order. For example, the single recall order "ABCD" would correspond to the unidirectional tree:



Every pair of recall orders can then be compared by constructing a joint tree for the pair. For example, let square brackets {[]} represent unidirectional nodes, angle brackets {<>} represent bidirectional nodes, and parentheses {()} represent nondirectional nodes. Then, the similarity between orders "ABCD" and "ABDC" can be quantified by noting that both orders are contained in a highly ordered tree, namely, [AB<CD>], which is quite similar to the two unidimensional trees [ABCD] and [ABDC]. The joint tree differs from each of individual trees by the addition of a single node. In contrast, the orders "ABCD" and "CADB" are extremely dissimilar. This is seen by the fact that the only tree containing both orders is the trivial tree, (ABCD). In practice, this notion of similarity can be quantified using a branch of mathematics known as lattice theory.

By calculating a similarity (or dissimilarity) measure between all pairs of recall orders, it was possible to parse the recall orders into distinct sets and then calculate distinct trees for each set (Hirtle, 1988; Hirtle & Crawley, 1989). Thus, using this method of parsing of recall orders, it has proven possible to identify circular patterns, multiple outliers, and, in some cases, multiple strategies.

5. ORDERED TREE AND INFORMATION SCIENCE

As mentioned earlier, information systems would be improved if their classification schemes reflected mental representations of users. Therefore, it is the author's belief that ordered tree technique would prove useful in information science applications in a number of tasks. Most certainly, the ordered tree framework should prove useful for analyzing knowledge bases,

ISSN: 2324-9773

Hirtle, S. C. (1990). Ordered trees: A structure for the mental representation of information. Proceedings of the 1st ASIS SIG/CR Classification Research Workshop, 79-86. doi: 10.7152/acro.v1i1.12466

constructing a representation of semantic knowledge, and building hierarchical information trees. Furthermore, applications where it is important to classify users as to the nature of their semantic knowledge would benefit from the use of ordered trees. In particular, ordered trees should prove useful in domains where not only the hierarchical structure is important, but where there is an implicit ordering on the material as well, either because of the structure of the domain or because of the preferences that the user imposes upon the material.

It also is reasonable that such a framework can be modified for data beyond recall orders to any set of repeated items, for example, accessing of repeated menu items, the order of scanning catalog items, or the order of searching keyword terms. These linear strings would then replace the recall orders as the raw data in developing trees. Future research is needed to explore such possibilities.

REFERENCES

Baird, J. C., & Noma, E. (1978). *Fundamentals of scaling and psychophysics.* New York: Wiley.

Broadbent, D. E., Cooper, P. J., & Broadbent, M. H. P. (1978), A comparison of hierarchical and matrix retrieval schemes in recall. *Journal of Experimental Psychology: Human Learning & Memory, 4*, 486-497.

Fisher, D. L., Yungkruth, E., & Moss, S. M. (1989). "Optimal menu hierarchy design: Syntax and semantics." *Manuscript in preparation*, University of Massachusetts.

Friendly, M. L. (1977). In search of the M-gram: The structure of organization in free recall. *Cognitive Psychology, 9*, 188-249.

Hirtle, S. C. (1987). On the classification of recall strings using lattice-theoretic measures. *Journal of Classification, 4*, 227-242.

Hirtle, S. C., & Crawley, E. (1989). "The use of ordered trees for uncovering strategies in free-recall." In N. X. Luong (Ed.), *CUMFID n. 16: Analyse arborée des données textuelles* (pp. 125-138). Nice, France: CNRS.

Hirtle, S. C., & Jonides, J. (1985). "Evidence of hierarchies in cognitive maps." *Memory & Cognition, 3*, 208-217.

McKeithen, K. B., Reitman, J. S., Rueter, H. R., & Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology, 13*, 307-325.

McNamara, T. P., Hardy, J. K., & Hirtle, S. C. (1989). Subjective hierarchies in spatial memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15*, 211-227.

Naveh-Benjamin, M., McKeachie, W. J., Lin, Y., & Tucker, D. G. (1986). "Inferring students' cognitive structures and their development using the "ordered tree technique." *Journal of Educational Psychology, 78*, 130-140.

Reitman, J. S., & Rueter, H. R. (1980). "Organization revealed by recall orders and confirmed by pauses." *Cognitive Psychology, 12*, 554-581.

ISSN: 2324-9773

Hirtle, S. C. (1990). Ordered trees: A structure for the mental representation of information. Proceedings of the 1st ASIS SIG/CR Classification Research Workshop, 79-86. doi: 10.7152/acro.v1i1.12466

ISSN: 2324-9773