User Generated Multi-Dimensional Classification in an Adaptive Network Library Interface[1]

Paul B. Kantor
Department of Library and Information Studies
The Alexandria Project Lab
School of Communication, Information, and Library Studies
Rutgers. The State University of New Jersey
New Brunswick, NJ 08903

Ronald Rice
Noshir Contractor
Department of Communication
School of Communication, Information, and Library Studies
Rutgers. The State University of New Jersey
New Brunswick, NJ 08903

Classification can be thought of as defining subject matter classes, and assigning information bearing items (IBEs) to those classes as a way to support organization and retrieval of those IBEs. This corresponds to a Platonic view in which subjects reside in a world of abstractions, and real world IBEs are mapped to them (many-to-many) as accurately as possible.

An alternative, very open and empirical view of classification stresses the relatedness among IBEs as the starting point. In the abstract formulation, IBEs will be related if they share "enough" subject descriptors or terms, or other abstract representations of content. The extreme of relatedness expresses itself in the library by shelf-list adjacency. This forms the support mechanism for subject matter browsing.

Under the alternative view, relatedness can be taken as a fundamental concept. The nature of relatedness can not be captured in a linear graph, such as shelf list. There are more relations than just "predecessor" and "successor". In fact, one can conceive of an empirically constructed directed graph in which a link runs from A to B if some knowledgeable user of the collection has said "if you find A interesting, you will want to look at B also." The relation might also be symmetric in nature: "A and B are related to each other".

Defined in this empirical way, the graphical structure can be a supplementary finding tool, for library users in a well-defined common interest group. With this possibility in mind, we are constructing a front end interface, called the Adaptive Network Library Interface (ANLI), which stands between a community of users and the online public catalog.

ANLI interfaces have been developed for the Online Public Catalogs at both Rutgers and the University of Illinois. It is to be installed in the Fall of 1992 for data collection. The present paper discusses the information being gathered, and the conceptual framework for its analysis.

Communication Network.

The members of eight user communities (the Illinois University Beckman Institute, the Math and Statistics Departments at Illinois, and, at Rutgers, the Departments of Philosophy, Mathematics, Statistics, Computer Science and Operations Research) has some present patterns of communication related to research. These patterns will be mapped, by interviews and network analysis, prior to installation of the ANLI. After one year, the patterns will be analyzed to see if they have changed, and whether any observed changes seem attributable to the ANLI.

When using the ANLI, users will be prompted to add pointers whenever they reach a full bibliographic display. They will first be able to browse the list of existing pointers, and endorse some or all of them. Then they may access the catalog through a second window, to add further items to the list of pointers, provided those further items are in the catalog. No free form pointers are accepted.

The ANLI keeps track of the participants in each communication link. If several different persons provide or affirmed the same link, that represents a relation among those people. These relations can be used to form a weighted graph, in which the weight of the arc from Smith to Jones is the number of links on which they concurred. It can also be used to create hypergraph structures, whose nodes are sets of related individuals. These graphs, whose nodes represent individuals, can be compared with the communication structure obtained by direct interview.

Implications for Classification

In addition, the proximity of disparate IBEs, as revealed by the number of pointers connecting them, and by the length of the path needed to connect them can be compared with proximity as defined by shelf-list separation, and by vector measures in the space of subject headings. These analysis will reveal the degree to which the "structure" of information space, as revealed by actual human behavior, is reflected (or not reflected) in the traditional linear library format.

Finally, the links themselves can be "colored" to represent the nature of the relation. Examples that come to mind are: B is an example of A; B elaborates on A; B corrects A; B uses the same conceptual approach as A, and so forth. Analyses of the reasons that people give for proposing links may suggest a classification of link types, which can then be used as part of the input session, to enrich the network of pointers.

Details of the analytical plan will be presented. The screens and modes of the system will be demonstrated with transparencies. If suitable hardware, and an internet connection are available, a demonstration of either the Rutgers or Illinois system can be made available after the presentation.