Coding Collected Research Papers to Create a Simple Information System

James J. Sosnoski, Department of Communication, University of Illinois at Chicago, Chicago, IL, USA
Gordon S. Carlson, Department of Communication Studies, Fort Hays State University, Hays, KS, USA

ABSTRACT

A major issue with research archives is the difficulty of finding contextually relevant information when developing a theory or project. Utilizing a sociotechnological approach, this paper proposes a descriptive markup language to function as a notational system to be used to convert collections of research papers in various disciplines into an information system. Markup codes attached to research papers can significantly improve data retrieval. A theoretical framework for the notational system is proposed, drawn from Charles Fillmore’s frame semantics and Thomas Kuhn’s conception of a lexicon specific to numerous research areas. Examples of how the notational system can be used are offered and the steps needed for the implementation of the notational system are outlined.

Keywords: Descriptive Markup Language, Information System, Markup Codes, Notational System, Research Papers

INTRODUCTION

Digital libraries of research publications are usually spread across various online sites. And these sites are not structured in the same way. This makes the retrieval of data that goes beyond keyword searches from them difficult—e.g., relations among concepts forming a model of the situation, conceptual changes in concepts drawn from previous research.

To make these issues more concrete, consider the digital libraries that contain publications about communication research. Because it does not have a uniform naming pattern, communication research is enigmatic. It is, for instance, difficult to relate the assumptions underlying research projects to each other. Consulting bibliographies is often futile. The titles of research publications in communication studies are not related semantically.
example, agenda setting theory is not semantically related to framing theory which is often considered a species of agenda setting theory. Looking only at the terms, the conceptions would not seem connected. On the other hand, speech codes theory would seem to be related to speech act theory because they are semantically related. However, Philipsen’s (1997) conception of “codes” is quite different than Searle’s (1969) conception of “acts.” Similarly, Kenneth Burke’s concept of symbolic action (1966) is semantically related to and considered an instance of symbolic interactionism. However, Burke views communication from the point of view of its author in one-way transactions and Mead/Blumer (1969) emphasize the dialogical character of communication. There is no semantic connection between standpoint theory and action assembly theory, yet conceptually they are more closely related than speech acts and speech codes. There are many other instances of the absence of cohesiveness in the discourses of communication researchers that can illustrate the difficulty of finding patterns of interrelations in their conceptions. Semantic coherence, however, is not the problem. The problem is the lack of a patterned, descriptive system capable of marking the similarities and differences among the theories in the research archives more precisely.

Why is this a problem and how serious is it? Whenever publications reach a certain mass, they go beyond the capability of humans to remember where they are so that they can be found when needed. Take a simple example: a person begins to purchase books. At first, she has no difficulty in locating a book which she needs. With even a small bookshelf, at some point she will have to put her books in order to form a library. To do so, she must categorize them. Let’s say she organizes her books in the categories: very interesting, mildly interesting, not very interesting, and not at all interesting. This idiosyncratic system might work well for her in as much as she would be likely to refer more often to the very interesting books than the totally uninteresting books.

The problem with an idiosyncratic filing system is that it is suited to an individual. The categories are meaningful to that individual but are not shared by specific groups. Categorization is often at the heart of data retrieval. Since it would not be feasible to change the names of communication theories, we believe that there is a need for a notational system for describing communication research that could be used as a markup language. Such a notational system would not be in any way prescriptive. It would be designed to show the inter-relatedness of communication research.

We can’t predict what the mark up system will provide; we can only create the framework and then through using it see what others will find. Put another way, we seek to avoid a framework of notation which is deterministic with predictable outcomes because this will lead to a narrower set of possible outcomes. With less potential outcomes it is possible for “false positives” where two documents which are not the same are marked up in such a similar way that a user would be confused or misled. While this problem of naming conceptions exists in all research fields, we use the field of Communication Studies (referred to as communication) as a case study since it draws from so many other fields and is a nexus for many competing and contradictory concepts. The problem of the idiosyncratic naming of communication research and our proposed remedy are socio-technological and directly impact the creation and management of knowledge. Solutions to this problem will lie at the intersection between people and technology as a once necessary print culture fades behind new technological information management regimes that allow people to interact with information in radically divergent ways.

Data retrieval is a common problem in large organizations. In many instances, the data and its organizing technologies are structured in ways that make it difficult to locate the information you need; for example, papers about contextually similar situations. From our point of view, the problem is in the way the data in a digital
Guiding Design for Waiting
www.igi-global.com/chapter/guiding-design-waiting/30369?camid=4v1a