

Social Aspects of Information

Felix Chu

*Coordinator, Systems and Operations Unit
Malpass Library
Western Illinois University
Macomb, IL 61455-1390*

Introduction

In recent decades, institutions have been moving away from the positivism evident in many disciplines in the earlier years of the 20th century. Instead of seeking for a single truth or standard, institutional assessments have become local in nature within a framework of best practices. Similarly, librarianship must move in that direction in responding to local needs. With the advent of a new century in an increasingly heterogeneous environment, one needs to understand that a piece of information does not have a single meaning but may be interpreted in various ways by different groups of people.

This paper looks at access points to library collections. The access points to books, journal articles, and other documents are those assigned by catalogers and indexers in the guise of subject headings, descriptors, and call numbers. These are primarily subject-oriented approaches. While the first groups allow browsing on the computer screen, the call numbers allow browsing of the physical shelves. Another source that has become available since the switch from card catalogs and paper-based tools to OPACs and other online tools is searching by keywords with relevance rankings. However, users are still unable to filter out unwanted materials or are not able to access relevant documents, because of factors such as discipline-specific meaning and usage of common terms.

Over the years, many authors have dealt with the difficulty of access. The user and the cataloger agree roughly a third of the time or less about the topics of abstracts, texts or other materials.¹ Connell also comments that differences in the use of terminology, semantics, and individual point of view contribute to how a book is described by an indexer and a user.² While searching by keywords may be presented as an option, as Fugmann says, “here we encounter the failure to realize that the word in its use in natural language is not intended for use in isolation. It is only in its context that the word assumes its meaning and importance.”³ Since a text has embedded meanings, the indexer or cataloger has to interpret and translate concepts into appropriate subject headings. The problem of interpretation was brought out earlier by Ingwersen and Wormell. They state that while retrieval using library tools tends to be based on an “exact match” paradigm dependent on commonality of understandings, expectations, goals and concept interpretation, the dynamic nature of society does not provide such a homogeneous approach among individuals such as the indexer and

the library user.⁴ One's interpretation is based on prior knowledge and experiences that are not shared. Kuhlthau went on to say that while the research paradigm may assume order, the personal interpretation may be a process of sense-making where one selectively attends to certain details.⁵ In trying to understand the text, "the more complex or ambiguous the stimulus, the more the perception is determined by what is already 'in' the individual and less by what is 'in' the stimulus."⁶ One applies search strategies and understands the results based on professional, ethnic or other characteristics that one has acquired. But at the same time, these characteristics will rule out other competing paradigms. Thus, lawyers may infer from the text and try to resolve conflicts among the sources, while engineers may make certain assumptions that sit behind the text, evaluate the sources, and seek out the best answer.⁷

The Social Aspects of Information

The concept of the social nature of information has engendered published research in many fields. While it is evident from the literature cited above that text is subject to interpretation, the literature on the social aspect of information in librarianship is rather sparse. For example, Alfino and Pierce recently presented an analysis on the moral value of information that may impinge on our professional values which are guided by a belief in the neutrality of information.⁸ The notion of "neutrality" has developed over time in interaction between the library profession and the culture. For example, the presence of fiction in the library collection was initially a hotly-debated topic because it marked a shift of the role of the public library as a core educational institution to an outlet for entertainment, and in the role of the librarian from educator to reader's advisor. In the case of fiction, constraints on money and space constrained the growth of those collections, and librarians were able to adopt a neutral stance toward that genre of literature. But the "Internet problem" is not just an updated "fiction problem" because much of what becomes available are from external sources not acquired by the library. Information itself may be neutral, but within a given context, it gains moral value that we librarians must deal with.

Looking more broadly in the information field, one may also refer to Bonnie Nardi and Vicki O'Day, who spoke on the need for creating a holistic context to make information and technology meaningful.⁹ "Technological tools and other artifacts carry social meaning. Social understanding, values, and practices become *integral aspects* of the tool itself."¹⁰ In using the telephone as an example, they spoke on the tacit understanding of the etiquette of placing a call, answering a call, greeting or taking turns during the conversation. The written text is not stable in meaning in that the reader constructs the meaning in interpreting the text. It is an active task. In creating meaning, the reader's cultural setting is different from the author's. Thus information ecology is the interaction between people, practices, values, and technologies within a local environment.

In a similar vein, Brown and Duguid have stressed the need to look at the social network surrounding a piece of information in order to benefit from its meaning. In their book *The Social Life of Information*, they state that within the design of the digital world, many people have tunnel vision, because they concentrate only on what they consider important, without paying attention to the background that forms the context.¹¹ It is only with the context that bits of information may be judged to be relevant or useful. The context may

be viewed as including the infrastructure that supports what is seen within the tunnel. For example, the turn-around time for patron-initiated ILL transactions may be considered long by many users because they do not consider that the supplying library may not have personnel to receive and process the request immediately, the requested item may be in use but not charged out, or the item is missing. In this sense, the authors are saying that, indeed, “less is more” if the information has been filtered through a social network that can provide the context and only a small portion of the total is presented as relevant and contextualized information. A piece of information acquires meaning only when it is socially understood. It is absolutely necessary to read the “background” to understand the relevance and worth of a piece of information. They speak of this background information as “stolen knowledge” that is essential. “Stolen knowledge” is what one learns by observing something being done, not by formal instruction. It is what puts a topic within a context—not just the facts and when to use them, but actually experiencing the process.¹²

In an earlier paper, I said that knowledge is dynamic in nature. But in the process of cataloging, the interpretations of the meaning of the contents are frozen into call numbers. The affective part of the meaning of a text is stripped because that part cannot be reflected through the access points.¹³ As any cataloger will probably agree, some books are easy to catalog because they clearly treat a single well-established topic with a corresponding subject heading and a class number in the classification schedule. Others, especially those on interdisciplinary topics in emerging areas, require more interpretation by the cataloger to fit them into existing subject headings and class schedules. And in trying to use a one-size-fits-all tool such as the Library of Congress Subject Headings, the fit is better in some cases and barely marginal in others. One’s interpretation of the content, however, is based on the social construct of one’s background.

Devlin and Rosenberg¹⁴ analyze communication in the workplace. They use a mathematical theory called “situation theory” to provide a framework for the study of information. A situation may be a limited part of reality. The authors analyze the use of Problem Report Forms (PRF) at an industrial site and at attempts to automate the process.¹⁵ The form has fixed fields encoded with categories such as “Site No:” or “System:”. There are limited areas for free-format text that describes the problem, its resolution, and parts used to fix the equipment. Since the free-text area and the time for each job are limited, the text used by the technicians usually contains many short phrases and abbreviations. Details that may be common knowledge to the technicians are also omitted. In trying to automate the process, interpretation of these forms by software became a bottleneck. Three kinds of knowledge are required to interpret the forms: 1) familiarity with the structure of the PRF; 2) technical knowledge of the domain; and 3) knowledge of the social structure pertaining to the PRF.

The MARC record is also formatted with fixed fields and free-text fields that must be filled in a highly prescribed manner according to cataloging rules. Only certain phrases and abbreviations are normally used even in the note areas with highly-prescribed meaning. To interpret the catalog record one must be: 1) familiar with the structure of the catalog record; 2) know the technical structure of the library and information world; and 3) know the social constraints of information that appear on a catalog record. All three work together to lend composite meaning to the record. The first part relies on a technical knowledge of the parts

of the cataloging record such as title, imprint, pagination and subject tracings. Particularly important is the call number, which functions as a link to the physical item. The second part refers to knowing things such as the physical layout of the library. This includes knowing the “local” version of the filing rules—what it means to be in call number order or alphabetical order. Similarly, one must know what is included in a journal citation and how to go from the citation, a symbol for the real thing, to the correct issue of a journal which may be cataloged or uncataloged, and therefore shelved differently.

The “social meaning” comes with the third part of decoding a journal citation or a catalog record. This is where the knowledge base or the context may not be shared. For example, when I started cataloging in 1975, there was the unwritten practice of assigning no more than three subject tracings per title. The title may of course be a 15-page pamphlet or a 10-volume set. There are also conventions governing the assignment of subject headings. Rarely does one assign a general heading once a more specific one is used. This means that a history book on Buenos Aires would be assigned “Buenos Aires (Argentina)—History”, but not “Argentina—History” even though Buenos Aires is the capital and its history is a major part of Argentine history. Unless a patron, and for that matter a librarian in public services, is aware of the practice, that pamphlet remains inaccessible if one consults the OPAC and looks only under “Argentina—History”. Similarly, the 10-volume set on the history of Argentina will probably only have the subject entry “Argentina—History”. But there may very well be a 50-page article on the history of Buenos Aires. Thus a technically perfect search for history materials concerning Buenos Aires may miss a great deal of relevant information.

A related issue is that subject headings and descriptors are added, changed, or updated periodically. Since additions are trailing the production of new knowledge, literature on new developments may not have the new headings, perhaps because the indexers have not become familiar with the concepts. For example, my Ph.D. dissertation is on collaboration between teaching faculty and librarians using a loosely coupled system framework for analysis. The seminal article used was written in 1976 by Karl Weick and examined educational institutions as loosely coupled systems.¹⁶ It was indexed by ERIC and the identifier “loosely coupled systems” was used for the first time although the authorized list did not appear until 1980. During the time up to 1992 as research for the dissertation progressed, the identifier appeared on fewer than half of the relevant articles indexed by ERIC. Those missing the identifier included titles such as “Loose Coupling Revisited: a Critical View of Weick’s Contribution to Educational Administration” (ED283255) which appeared in 1983 and “Curriculum Change in Loosely Coupled Systems” (EJ255080), published in 1981. Many other articles used were indexed by ABI-Inform and used descriptors dealing with management or organizational behavior. Thus, if there were no attempts to follow up on known authorities and footnotes, a great deal of relevant literature would have been missed, especially earlier literature. Also during this time, another problem with connotations based on tacit knowledge became apparent. To fellow doctoral students working with materials on higher education, a high “retention” rate is a good thing. But to others who are school principals, a *low* “retention” rate is a good thing. What often remains unstated is that to a university administrator, retention rate refers to how many freshmen come back for their sophomore year or sophomores for their junior year, instead of failing, dropping out or transferring to a different school. But to a K-12 educator, retention means

having students repeat a grade. This is a clear example of discipline-specific meanings attributed to the same words that may not be articulated.

Even more confusing is the concept of the picaresque novel. There is an LC subject heading called “Picaresque literature” which was later expanded with national qualifiers such as “Picaresque literature, Mexican” or “Picaresque literature, English”. During the earlier part the 20th century, the topic was treated as a phenomenon in Spanish literature during the Golden Age from 16th to mid-17th centuries, marked by *Lazarillo de Tormes* which appeared in 1554 followed by *Guzmán de Alfarache* (pt.1 1599, pt.2 1604) and ending with *La vida del buscón* in 1626. As scholars from comparative literature became interested in the genre, the base expanded to include works from other national literatures without the temporal boundaries. Thus picaresque novels came to include *Moll Flanders* (1772) by Daniel Defoe, *The Adventures of Huckleberry Finn* (1884) by Mark Twain, *The Adventures of Augie March* (1953) by Saul Bellow, and *The Tin Drum* (1959) by Günter Grass. Over the years, emphasis in critical works went from commentaries on the social to the psychological and later to the mythical.¹⁷ Even though the social climate for understanding the genre has changed over time, the basic structure of subject headings has not. This instance is marked by a shifting definition of the genre where there are multiple communities of practice in the field, each with a variation in the definition of the genre.

Another problem is the social understanding that is reflected through the organization of the classification schemes that influence access. In “Sameness and Difference,” Hope Olson commented on the cultural biases that are reflected through the use of Dewey Decimal Classification (DDC).¹⁸ In trying to group “like” items together, the “likeness” is a socially constructed value that reflects a point of view. Thus Canadian literature in English, French, and Inuit are classed in different areas of the 800s. Similarly in the Library of Congress classification schedules, for the picaresque literature mentioned above, call numbers may be in many different areas of the Ps. While PN3428 is provided for general works on the genre, commentaries devoted primarily to a single national literature are grouped with that national literature. Thus, Spanish is in PQ6147.P5 and German is PT747.P5. Individual works are further spread out within the national literatures by author.¹⁹ So “sameness” and “difference” are social distinctions where one “likeness” is preferred to another “likeness” for organizing materials.

In a similar vein, a topic that interests me is the role that food plays in history. The potato is a New World discovery. But then why does it show up in ethnic cooking in such places as Russia and India? How did it become a staple in Ireland which eventually led to the great famine? How is the subsequent Irish migration to the United States related to the famine, and possibly led to figures such as Tip O’Neill who influenced the political landscape with his saying that all politics is local? Those are interesting questions but hard to answer because subject headings and call numbers are not structured with the same outlook on information as the questions. Some of the useful books are *Seeds of Change* classed in SB71 with subject headings “Plants and history” and “Crops—History”; *Hunger and History* classed in TX353 with subject headings “Food—History”, “Civilization—History” and “Agriculture—History”; *Food in History* classed in GT2850 with subject headings “Food—History” and “Dinners and dining”; and *The History and Social Influence of the Potato*

classed in SB211 with the subject heading “Potatoes—History”.²⁰ Thus, if one only looked for books on the history of the potato, only one of the four books would be found. But there is also no apparent way for a more systematic approach.

Shively has provided the insightful observation that current library resources such as the catalog record describe the container, not the contained.²¹ The catalog or the index tells the fact that a container has milk but not how the milk in that particular container tastes. The dysfunction is that library users want to deal with ideas in the sources, but the subject headings and other access points reflect only the container and not the nature of the contained. What is missing includes the tacit dimension to which one must become acculturated because each discipline or culture has its own set of assumptions for organizing ideas and special use of vocabularies.²² This is evident in eliciting information from the previously-discussed PRF or from a catalog record. Communities of practitioners such as catalogers have assumptions that may not be shared. These are bits that are not verbalized but learned through accumulated social experiences. These bits are procedural in nature and highly contextualized. One pays selective attention to the bits in a passing stream and selectively combine and compare them to one’s past knowledge.²³ These selections are then frozen into access points associated with each item. Thus the social dimension is composed of both the interpretation of the explicit text and the tacit assumptions surrounding the text. While it is not the librarian’s intention to hide information, the way information is organized and access provided may have placed an obstacle to its use. That obstacle is more social than technical in the singular interpretation of the content provided by existing access tools.

Notes

1. See Mary E. Brown, “By Any Other Name: Accounting for Failure in the Naming of Subject Categories,” *Library & Information Science Research* 17 (Fall 1995): 347-85; and Lourdes Y. Collantes, “Agreement in Naming Objects and Concepts for Information Retrieval” (Ph.D. diss., Rutgers, 1992).
2. Tschera Harkness Connell, “Techniques to Improve Subject Retrieval in Online Catalogs: Flexible Access to Elements in the Bibliographic Record,” *Information Technology & Libraries* 10 (June 1991): 87-98.
3. Robert Fugmann, “Obstacles to Progress in Mechanized Subject Access and the Necessity of a Paradigm Change,” in *Saving the Time of the Library User Through Subject Access Innovation*, ed. William J. Wheeler (Champaign, Ill.: Publications Office, Graduate School of Library and Information Science, 2000), 7-45.
4. Peter Ingwersen and Irene Wormell, “Means to Improved Subject Access and Representation in Modern Information Retrieval,” *Libri* 38 (June 1988): 94-119.
5. Carol C. Kuhlthau, “Inside the Search Process: Information Seeking from the User’s Perspective,” *Journal of the American Society for Information Science* 42 (June 1991): 361-371.
6. Ronald Fernandes and Herbert A. Simon, “A Study of How Individuals Solve Complex and Ill-structured Problems,” *Policy Sciences* 32 (1999): 226.
7. Ronald Fernandes and Herbert A. Simon, “A Study of How Individuals Solve Complex and Ill-structured Problems,” *Policy Sciences* 32 (1999): 238ff.

8. Mark Alfino and Linda Pierce, "The Social Nature of Information," *Library Trends* 49 (Winter 2001): 471-485.
9. Bonnie A. Nardi and Vicki L. O' Day, *Information Ecologies: Using Technology with Heart* (Cambridge, Mass.: MIT Press, 1999)
10. Bonnie A. Nardi and Vicki L. O' Day, *Information Ecologies: Using Technology with Heart* (Cambridge, Mass.: MIT Press, 1999), 21.
11. John Seely Brown and Paul Duguid, *The Social Life of Information* (Boston, Mass.: Harvard Business School Press, 2000)
12. John Seely Brown and Paul Duguid, *The Social Life of Information* (Boston, Mass.: Harvard Business School Press, 2000), 136.
13. Felix T. Chu, "The Freezing of Dynamic Knowledge," *Technicalities* 16 (January 1996): 1, 7-9.
14. Keith Devlin and Duska Rosenberg, *Language at Work: Analyzing Communication Breakdown in the Workplace to Inform Systems Design* (Stanford, Calif.: Center for the Study of Language and Information, 1996)
15. Keith Devlin and Duska Rosenberg, *Language at Work: Analyzing Communication Breakdown in the Workplace to Inform Systems Design* (Stanford, Calif.: Center for the Study of Language and Information, 1996), 151ff.
16. Karl E. Weick, "Educational Organizations as Loosely Coupled Systems," *Administrative Science Quarterly* 21 (March 1976): 1-19.
17. See for example, Frank Wadleigh Chandler, *Romances of Roguery* (New York: Macmillan, 1899); Alonso Zamora Vicente, *Qué es la novela picaresca* (Buenos Aires: Columba, 1962); Richard Bjorson, *The Picaresque Hero in European Fiction* (Madison: University of Wisconsin Press, 1977); and Alexander Blackburn, *The Myth of the Picaro: Continuity and Transformation of the Picaresque Novel, 1554-1954* (Chapel Hill: University of North Carolina Press, 1979); and *The Picaresque: a Symposium on the Rogue's Tale*, ed. by Carmen Benito-Vessels and Michael Zappala (Newark: University of Delaware Press, 1994).
18. Hope A. Olson, "Sameness and Difference: a Cultural Foundation of Classification," *Library Resources & Technical Services* 45 (July 2001): 115-122.
19. The class numbers are taken from the most recent editions of the *Library of Congress Classification* schedules.
20. Taken from OPAC search results using DRA Information Gateway at Western Illinois University in November and December of 2001.
21. Daniel Shively, "Idea Retrieval," *Technical Services Quarterly* 6, 1 (1988): 23-30.
22. Don Spanner, "Border Crossing: Understanding the Cultural and Informational Dilemmas of Interdisciplinary Scholars," *Journal of Academic Librarianship* 27 (September 2001): 352-360.
23. Robert J. Sternberg, "Enhancing Education for Immigrants: The Role of Tacit Knowledge," *Educational Policy* 12 (November 1998): 705-718.