

Working Paper on the Relationship between Subject Indexing Principles and Online Subject Searching Strategies

Prudence W. Dalrymple

University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Jennifer A. Younger

University of Wisconsin-Madison, Madison, WI 53706, USA

The outcome of the subject indexing process has a direct effect on the success of searching strategies. One of the factors determining this outcome, and therefore a central problem in information storage and retrieval is the point at which subject indexing principles and online subject searching strategies intersect. The problem can be defined in straightforward terms as a communication issue. Based on many years of experience and observation, it appears that users do not understand how subject headings, e.g., LCSH, MeSH, are constructed and applied to materials and that subject catalogers do not take into account how users will go about finding the topic.

Subject indexing principles are referred to by several terms in the literature, including content analysis, subject analysis, indexing, and cataloging, and are defined here as the principles that determine how the indexing terminology is established and applied. Online subject searching strategies are the conscious approaches taken by searchers when approaching a subject file. In focusing on end-user searching, we are not including the pre-search interview and are concentrating on the methods, heuristics, and strategies that are used in searching online catalogs or indexes. Subject indexing principles create subject headings or descriptors that are subsequently found through the use of various subject searching strategies, the point at which they meet being the intersection. We propose to examine the nature of the relationship between subject indexing principles and online subject searching strategies.

Much is known independently about indexing and about searching strategies. Previous studies have focused on related but discrete aspects of this study, including the identification of subject searching processes in library catalogs (Markey, 1984), the interaction between indexing and user-assigned terminology (Bates, 1977), why engineers who are problem solvers and prefer to search by function don't use the databases created by indexers using a subject approach (Breton, 1981), and user-responsive subject control (Tague, 1981). Others have written extensively on why it is not acceptable to turn card catalogs and printed indexes into online services without redesigning the indexing procedures (Cochrane, 1981). Theories of indexing (Borko, 1977) and document aboutness (Maron, 1977) as well as practical manuals on the principles of subject headings (Chan, 1986) discuss the indexing processes.

We have identified the following two indexing principles as being of major importance to convey to users:

- specificity in relationship to the hierarchical structure of a particular indexing language, to the document being indexed, and to the policies relating to depth of indexing, and

- the means of coordination used to cover all aspects of a subject discussed in a document.

We have divided search strategy discussion into two major areas: database selection, which is not included in this paper, and search heuristics. At the risk of oversimplifying, we set out the following principles as the two basic principles underlying search heuristics. In practice, the pure form of these techniques is rarely encountered, and searchers draw upon both techniques at will. These are:

- the specific to general approach of "pearl growing" in which a keyword or known entity is used to enter the file, and
- the general to specific approach that uses a broad term or subject heading as the means of entering the file.

An elementary example illustrates how indexing and searching practices collide and what difficulties are encountered in communication between indexer and searcher. Regarding specificity, e.g., indexers follow guidelines that relate specificity to the piece in hand and to the subject cataloging rules and make choices accordingly. For example, an indexer can describe a book about 4 countries in South America under the heading "South America." This practice follows a Library of Congress (LC) rule that says when the number of specific objects exceeds 3, apply the next level of heading. Searchers, on the other hand, relate specificity to their information need and intuitive thinking as to the entry term most likely to yield results. Searchers may well enter the name of a specific country for which they want information and be disappointed with a low retrieval rate because the index contains primarily works under the subject heading of "South America." If they knew the indexing practice, they would search under "South America" (and in fact might anyway). But the question remains, assuming bibliographic instruction stresses the importance of specificity can we expect users to remember that specificity applies only to three or fewer items?

Secondly, the application of a broader heading such as South America (as indicated by the rule of three) does allow efficient searching in that users can search once under South America. However, the use of search techniques that enable searchers to "OR" any number of terms is also intended to provide efficient searching and furthermore is intended to facilitate searching according to the principle that "indexers index to the most specific." It can be seen in this example how the use of Boolean operators changes the interaction between cataloging and searching.

A difficulty in bringing together indexing and searching is that each side is making a different decision about specificity. The problem is further amplified when several indexes and catalogs are accessed from the same workstation.

This example leads to a second kind of problem. Lack of consistent feedback is frustrating to users, but more damaging perhaps is the fact that it undermines the user's confidence in understanding and using the system. Continuing with our South American example, suppose that a user is looking for a known book on archeological excavations in Peru, searches under Peru "AND" excavations, but does not find the book in question because the particular includes information on excavations in four South American countries and subsequently was entered under South America. The most typical interpretation is that the library does not have the item

but in reality, the user does not know how the indexing was applied and therefore does not understand how to search for it "correctly."

Another area is those cases where users do not know how subject headings and subdivisions are applied. Scope notes inform indexers as to current practices, but searchers by and large do not have ready access to these scope notes. A case in point applies to LC cataloging rules for assigning subject headings to books of Christian poetry. General books of Christian poetry are given the combination topic/literary form "Christian poetry." If, however, the poems are solely on a specific topic such as the Trinity, then a subject heading is assigned to bring out this specific topic, e.g., "Trinity - Poetry." A second heading is assigned for the literary form as appropriate (e.g., "American poetry") except that "Christian poetry" would not be used. Thus, if searchers do not know these rules, they will not find all the books of Christian poetry they might otherwise discover.

We are in the process of designing a research project to examine the relationship between standard and accepted subject indexing principles and online subject searching strategies. Specific objectives will include:

- identification of standard indexing principles as exemplified in Library of Congress Subject Heading Manual, 3rd ed., and *ERIC Indexing Handbook: Clearinghouse Indexing Practices*, May 1989,
- identification of standard subject search strategies as exemplified in major textbooks and guides, e.g., *Online Information Retrieval* by Steven P. Harter, *Manual of Online Search Strategies* by C.J. Armstrong and J.A. Large, and *Searching DIALOG: the Complete Guide* from DIALOG Information Services.
- identification of similarities and differences between the subject indexing principles and subject searching strategies, and
- preliminary determinations as to the extent of congruence between subject indexing and subject strategies, and
- preliminary exploration of the implications for revising and enhancing indexing and retrieval systems.

Technological advances in computer-supported indexing as well as in computer-enhanced retrieval techniques bring with them the opportunities for change. In doing so, it is important that we establish common ground between indexing and searching practices and techniques. Our study will attempt to do this.

SELECTED REFERENCES

- Bates, Marcia J. 1977. "Factors Affecting Subject Catalog Search Success." *Journal of the American Society for Information Science* 2: 161-169.
- Borko, H. 1977. "Toward a Theory of Indexing." *Information Processing & Management* 13:355-65.

Breton, E.J. 1981. "Why Engineers don't use Databases." *Bulletin of the American Society for Information Science* 7:20-23.

Chan, Lois Mai. 1986. *Library of Congress Subject Headings: Principles and Applications*. 2d ed. Littleton, CO: Libraries Unlimited.

Cochrane, P. 1981, July. "Improving the Quality of Information Retrieval Online to a Library Catalog or Other Access Services, or, Where do we go from here?" *Online* 5:30-42.

Markey, Karen. 1984. *Subject Searching in Library Catalogs: Before and After the Introduction of Online Catalogs*. Dublin, OH: Online Computer Library Center OCLC.

Maron, M.E. 1977. "On Indexing, Retrieval and the Meaning of About." *Journal of American Society for Information Science* 28:38-43.

Pao, Miranda Lee. 1989. *Concepts of Information Retrieval*. Englewood, CO: Libraries Unlimited.

Tague, Jean M. 1981. "User-Responsive Subject Control in Bibliographic Retrieval Systems," *Information Processing & Management* 17:149-159.