Classification Systems as a Visualization and Navigational Tool in Online Searching: A Technique for Supporting Multi-Cultural Information Seeking

Jiqun Liu and Nina Wacholder
School of Communication and Information (SC&I)
Rutgers, the State University of New Jersey
4 Huntington Street, New Brunswick, NJ, 08901
jl2033@scarletmail.rutgers.edu; ninwac@rutgers.edu

Abstract

We propose that a promising approach to support information seekers from different cultures is to integrate the visualization of different classification systems into advanced information retrieval (IR) systems. This will allow information seekers to retrieve documents that reflect semantic associations made by a classification system that relies on familiar cultural norms and also documents based on a classification system that relies on less familiar cultural norms. On the assumption that there exist underlying connections between a Web searcher's mental model and a classification system from the searcher's own cultural environment, this paper explores the possibility of integrating classification systems into IR systems. We present a prototype integrated search interface and discuss how classification systems could help users unlock better search performance and improve user experience. This paper concludes with a discussion of promising directions for future research, especially for IR systems that support multi-cultural approaches.

Introduction

Classification and IR systems are similar in the sense that both are intended to help information seekers find documents that satisfy an information need. We use classification in this paper to refer both to issues of classification and of categorization as distinguished in Jacob (2004). By integrating classification systems into sophisticated IR systems which track user activities and user intentions, it will be possible to conduct studies on whether this integration can help users' mental models of Web information resources in real-time searching, thus improving the results of the information search process and the user experience. The two primary research questions are:

- How can we integrate classification systems into IR systems for a Web search process?
- Does this integration improve search results and/or user comfort with the IR system?

Because of the preliminary nature of this proposal, we start with a simple but realistic information need: an information seeker is looking for popular books that would help them learn more about the ideas expressed in *Das Kapital* and about the book's social and cultural influence.

Background

According to Zerubavel (1991), classification is a mental process of grouping "similar" things together in distinct clusters and dividing "different" clusters from one another. In this sense, classification posits "islands of meanings" which do not exist "out there" as objectively observable entities. Instead, it is generated by both our minds and the process of cognitive socialization (Zerubavel, 1991). Thus, a classification system can be understood as a mental model of interrelated concepts, constructs, and "islands of meanings" which can be differentiated from each other and also linked together at various levels. Use of a classification system from an unfamiliar culture (e.g., someone from China using LCSH or someone from the U.S. using CLCS) can require mental effort well beyond the normal effort expended by a user of a classification scheme.

With the development of advanced IR techniques, most human information seeking activities are conducted online using web-based search engines. Slone (2002) studied the effects of mental models on Web search behavior and found that the differences in mental models can lead to significant behavioral-level divergence, especially in query formulation strategy. Hence classification systems as a reflection of mental models can contribute semantically related categories and subcategories to Web information organization and retrieval. The impact of using different classification systems may affect the ways people interact with Web information and IR systems.

In information organization practice, classification systems are adopted by librarians and other information professionals to organize library collections. Interestingly, people with different cultural backgrounds and mental models adopt different classification systems in information organization. For example, the Library of Congress Subject Headings (LCSH) is widely employed in U.S. public and university libraries, whereas public libraries and bookstores in China usually use the Chinese Library Classification System (CLCS) when organizing their digital and physical collections. To the best of our knowledge, classification systems have been used in IR systems primarily as assistant tools that assist with information filtering and faceted browsing, for example in the context of health information search (Zhang, 2014) and patent search (Larkey, 1999).

RQ1: The Integration of Classification System and IR System

To better illustrate the idea of integration, we take book search as an example. Suppose a user's search task is finding books that are relevant to Karl Marx's *Das Kapital*. After the user issues a query (in this very simple example, the book title), the integrated IR system will retrieve the result of exact match: the book *Das Kapital*. The system will also return books on related topics. In this context, the relevance of a topic or book can be partly determined by the classification system closest to the information seeker's mental model. If the user is more familiar with CLCS than LCSH, then they may interact with the results that are recommended by CLCS (e.g., *the biography of Mao Zedong*) since these results are more

likely to seem relevant or useful. By contrast in LCSH, *Das Kapital* is classified under economic theory. If the user mainly employs LCSH in organizing books, he or she may identify the books from economic areas and a variety of economic models as relevant.

In a state of the art IR system, users' implicit preferences and mental model of classification can be inferred from their interaction with the system (e.g., clicks on the search result; snippets from different classification systems; dwell time on different search results). Once a user's preference is clarified, the IR system will provide the option of setting up the preferred system as default assistive classification system, thereby better support user's book search (e.g., recommending more books from related categories in CLCS). The interactive process is illustrated in Figure 1.
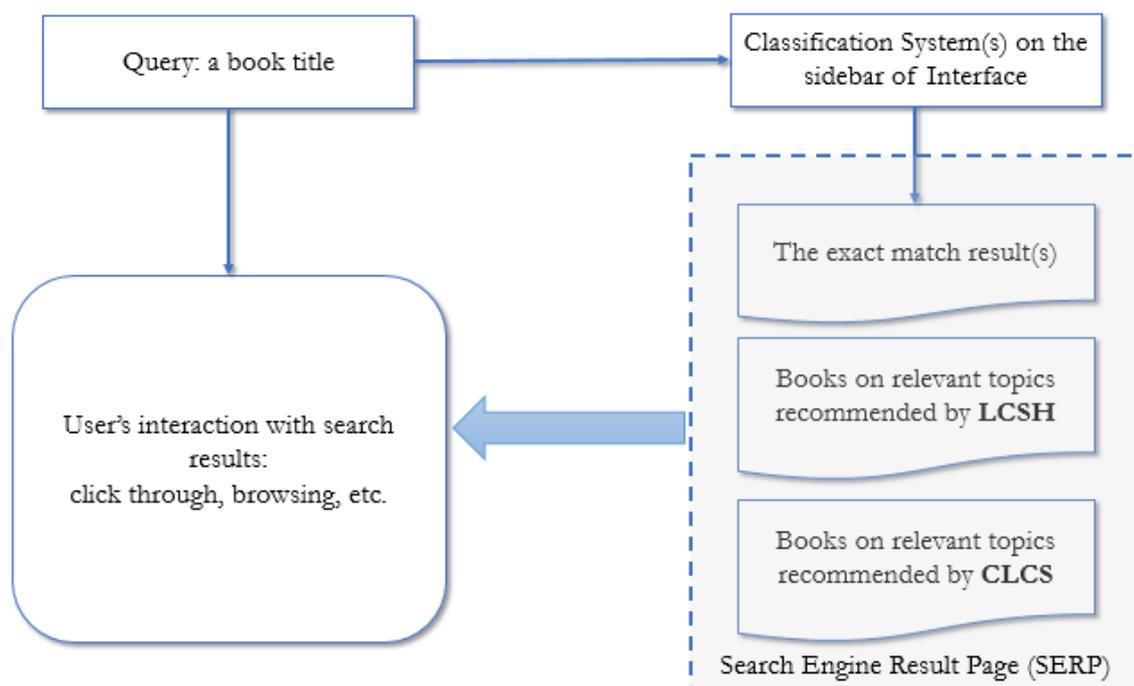

Figure 1. The interaction between user and the integrated system

RQ2: Improve Search Performance with Classification-Based Mental Model

By using multiple classification systems, a user will be able to search potentially useful information in different ways. In information search practices, users should be allowed to choose between a series of classification systems so they may search and browse in the manner that best matches their information needs, search tasks, and mental models. For example, if a user wishes to find more relevant books about *Das Kapital*, they can examine the search result snippets recommended by different classification systems and see which books are good matches with their needs. Based on the user's preferences and past search behaviors, the relevant documents may cover economic domains. However, if the user is more familiar with the CLCS system and wants to better explore the political and ideological aspects of Marxist theory, searching and browsing with the help of CLCS may be a better

choice. Figure 2 illustrates the integrated search interface. When the user issues the query *Das Kapital*, the different sets of relevant books would be identified and recommended by the default classification system (e.g., LCSH or CLCS). In LCSH, *Principles of Microeconomics* may be a relevant result as it also belongs to the subcategory HB (Economic theory, Demography). By contrast, based on CLCS, the system may recommend the book *Mao Zedong: A Biography* because both this book and *Das Kaiptal* are under category A (Marxism, Leninism, Maoism, and Deng Xiaoping theory). Space limitations preclude a longer discussion of the reasons for these differences. Since users can to choose the classification system they are more familiar with, they may locate relevant books in a more efficient manner and with less cognitive load. Search results obtained using LCSH *and* CLCS would reflect, at least partially, the cultural models reflected in the different classification schemes.
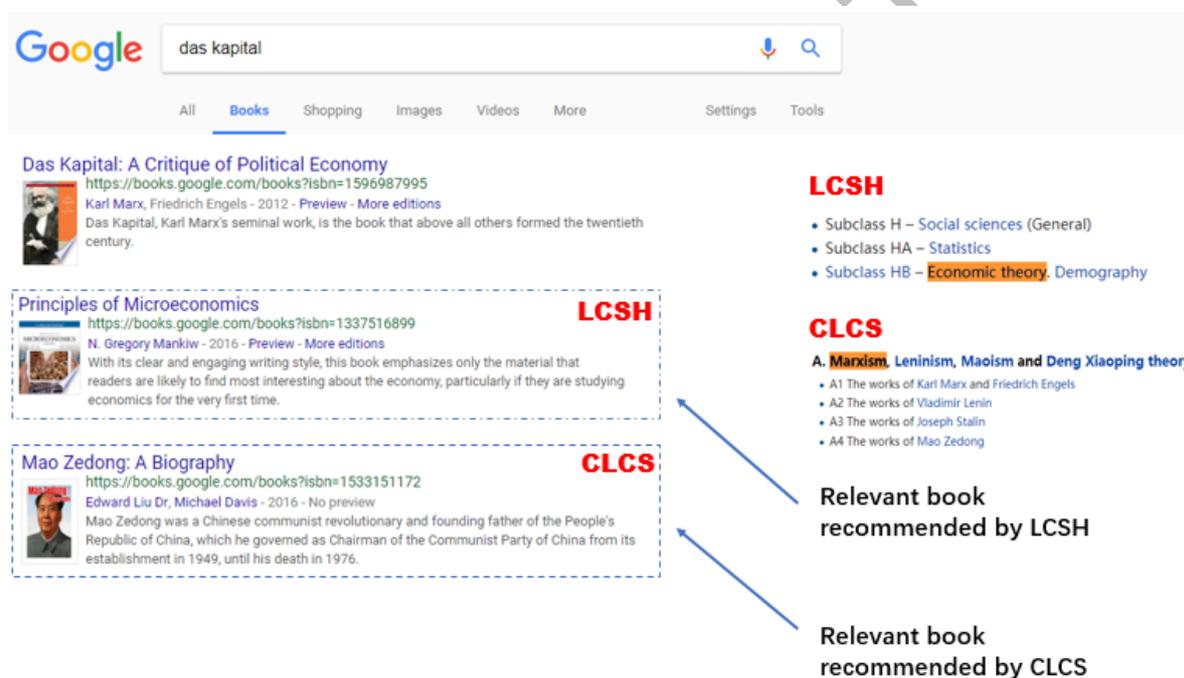

Figure 2. Prototype of the integrated search interface

Conclusion

The idea of integrating classification systems into IR system essentially speaks to the cultural aspects of information search. Specifically, it takes the differences between culture-specific mental models into consideration in IR system design and thus can serve as a supplement to mainstream technology-oriented IR research.

Future research can go beyond book search and investigate more deeply the underlying connections between cultural background, classification systems, and mental models in general Web searching. This work will also support evaluation of the usefulness of classification systems in IR by comparing them with other assistant tools and

recommendation techniques. To further reduce the cognitive load of users in Web searching, future research can explore the possibilities of developing customized classification systems based on the explicit and implicit feedback provided by human-system interaction.

References

Jacob, E. K. (2004). Classification and categorization: A difference that makes a difference. *Library Trends*, 52(3), 515-540.

Larkey, L. S. (1999). A patent search and classification system. In *Proceedings of the Fourth ACM Conference on Digital Libraries* (pp. 179-187). ACM: New York, NY.

Slone, D. J. (2002). The influence of mental models and goals on search patterns during web interaction. *Journal of the Association for Information Science and Technology*, *53*(13), 1152-1169.

Zerubavel, E. (1991). *The fine line: Making distinctions in everyday life*. New York, NY: Free Press.

Zhang, Y. (2014). Searching for specific health-related information in MedlinePlus: Behavioral patterns and user experience. *Journal of the American Society for Information Science and Technology*, *65*(1), 53–68.