Chapter XI
Social Information Seeking in Digital Libraries

George Buchanan
University of Wales, Swansea, UK

Annika Hinze
University of Waikato, New Zealand

ABSTRACT

Information seeking is a complex task, and many models of the basic, individual seeking process have been proposed. Similarly, many tools now exist to support “sit-forward” information seeking by single users, where the solitary seeker interacts intensively with a search engine or classification scheme. However, in many situations, there is a clear interaction between social contexts beyond the immediate interaction between the user and the retrieval system. In this chapter we demonstrate a number of contrasting uses of the social aspects of information seeking, and through those propose, demonstrate, and realize social models of information seeking that complement existing information seeking models and technologies. These include: information sharing among humanities researchers; creation of profiles for continuous, ongoing searching of medical material; and the capture of models of user behaviors in an interactive, mobile tourist information system. From the human perspective we illustrate differing social techniques and issues including: explicit and implicit sharing; seeking facilitated by subject (medical, academic) experts and search experts (librarians); and anonymized and attributed social environments. Whereas many papers focus on particular social retrieval technologies, in this chapter we abstract a number of different approaches to present underlying principles, architectures, and models that can be adopted for a wider range of applications. We focus on digital library (DL) technology, as DLs have well-accepted architectures that support a wide variety of information seeking tools. We also address the key related issue of models of information seeking—models that have strongly influenced the design of DL technologies.
BACKGROUND

Social information seeking studies the many ways in which communication and interaction between people influence their information seeking. To take one simple example, humans often share “pearls of wisdom” through informal, off-line discussions. This form of interchange has been observed in many different domains including engineering, academe, and the clinical world. Though such off-line communications are inaccessible to computer systems, they form a critical foundation for much of the actual information seeking of the users of digital libraries and online encyclopedias. Failing to understand the consequences of this hidden activity on interactive information seeking will result in a shortfall in the quality and effectiveness of online information retrieval tools.

The pervasive nature of social interaction within information seeking already leaves its fingerprints on public Web sites. For example, many news and technical Web sites support simple “e-mail this article” tools that facilitate information sharing between collaborators at a basic level. Similarly, Amazon.com has placed a considerable investment in providing collaborative filtering to support recommendations for its customers. Finally, Google’s PageRank algorithm (Brin & Page, 1998) and the related work of Chakrabarti, Joshi, Punera, and Pennock (2002) on online communities of interest are, in fact, prime examples of the successful harnessing of social information—the human creation of Web links between sites—to improve the effectiveness of automated document retrieval systems. However, persuasive as these “real-world” examples may be, they represent a fragmented veneer beneath which is, at present, a lack of systematic science and, often, a considerable degree of conjecture as to which elements should be evaluated to create a “social” information retrieval infrastructure.

Technical Background

The projects reported and synthesized in this chapter all originate in the domain of digital library research. There are many competing definitions of digital libraries, from those emphasizing the significance of the institution that operates the library, to highly techno-centric definitions that focus upon the combination of particular functions. Our definition, and the one upon which this chapter is built, is that a digital library is an indexed and organized collection of digital documents, where inclusion in the library is determined by the institution or individual who operates the library. Implicit in this definition is an assumption that selection of a document for inclusion in the library is based upon its intellectual quality and its relevance to the purpose of the collection. Similarly, the indexation and organization of the library is assumed to include features analogous to a traditional library of printed documents—for example, subject classification hierarchies, and author and title indexes. These indexes are expected to be supplemented by those that can only realistically be created by a computerized index—for example, of the full text of the documents.

One key concept that we have borrowed from hypertext research is also critical to understanding the role of digital library systems in our research. Halasz (1988), in a seminal work, introduced the distinction between computation within a hypertext and computation over a hypertext. The first concept applies to internal processing that occurs in the creation and operation of the system itself; the second to the ability to access the content of a hypertext whilst it is in operation, and perform some external processing over the content. One practical example may clarify the distinction: a Web server may operate an online store, such as Amazon, while an external computer (e.g., Google) creates a searchable index over the content of many Web servers, including the store, and indexes it. In this context, the store’s Web server