Preface

The popularity of the Web has led to a tremendous growth in the volume of available online information. The result is that people have now come to depend on the Web to meet their information needs via search engines, portals, digital libraries, and other information retrieval systems. However, the amount of information and its growth is a double-edged sword due to the problem of information overload, exacerbated by the fact that not all content on the Web is relevant or of acceptable quality to information seekers. Information overload has led to a situation where users are swamped with too much information and have to sift through the materials in search of relevant content.

A variety of techniques have been adopted on the Web to address these problems inherent in information search, drawing from the fields of information retrieval, information filtering, human computer interaction, and the study of information seeking behavior. Work in these areas has yielded many novel and useful algorithmic and user interface techniques. Research in information seeking behavior suggests yet an alternative promising approach in helping users meet their information needs. Many studies have found that interaction and collaboration with other people is an important part in the process of information seeking and use. It is not uncommon that in searching for information, we tap on our social networks—friends, colleagues, librarians—to help locate what we need.

Social information retrieval refers to a family of techniques that assist users in meeting their information needs by harnessing the collective intelligence of other users—their expert knowledge or search experience. Elements of social information retrieval may be found on the Web through the hyperlinks that connect different Web sites (e.g., bookmark lists), subject directories (e.g., Yahoo, Open Directory Project), Google’s PageRank algorithm, and user annotations of resources (e.g., Amazon.com’s book reviews and ratings). More contemporary techniques include social tagging, collaborative querying, social network analysis, subjective relevance judgments, and collaborative filtering.

Social information retrieval is an emerging area and a promising avenue for the design and implementation of a new generation of information retrieval systems. It has drawn interest in academia as well as in industry. This book introduces readers to this area as well as discusses the state-of-the-art techniques in social information retrieval. It serves as a resource for those dealing with information retrieval systems, services, and research, and is written for academics, researchers, information retrieval product managers and software developers, librarians, and students.
ORGANIZATION

The book’s chapters are organized into six sections with the following themes:

- Collaborative Querying
- Collaborative Classification and Organization
- Using Social Networks for Information Retrieval
- Social Issues
- Social Information Seeking Models
- Applications and Case Studies in Social Information Retrieval

Section I deals with collaborative querying, discussing various techniques that support searching by harnessing other users’ search experiences. This section consists of two chapters.

Chapter I, “Collaborating to Search Effectively in Different Searcher Modes Through Cues and Specialty Search” by Naresh Kumar Agarwal and Danny C.C. Poo, argues that searchers generally have difficulty searching in knowledge repositories because of the quantity of data involved and because searcher mechanisms are not tailored to their differing needs. Collaboration among searchers is one possible solution. They review concepts related to social information retrieval and other collaborative mechanisms, and discuss two collaborative searching mechanisms—cues and specialty search. A case study of an online educational portal is also presented to show how collaboration could enhance learning and search efficacy.

Chapter II, “Collaborative Querying Using a Hybrid Content and Results-Based Approach” by Ray Chandrani Sinha, Dion Hoe-Lian Goh, Schubert Foo, Nyein Chan Soe Win, and Khasfariyati Razikin, describes the concept of collaborative querying, a technique that makes use of past users’ search experiences in order to help the current user formulate an appropriate query. Here, related queries are extracted from query logs, clustered, and used as candidates for recommendations. Query similarity is determined using a combination of query terms as well as search result documents. For the latter, features such as titles, URLs, and snippets from the result documents are used. Experimental results reveal that the best clusters are obtained by using a combination of these sources rather than using only query terms or only result URLs alone.

Collaborative classification and organization is covered in Section II. Chapters in this section examine how classification schemes and metadata can be constructed collaboratively. The section consists of three chapters.

Chapter III, “Collaborative Classification for Group-Oriented Organization of Search Results” by Keiichi Nakata and Amrish Singh, begins this section by examining the use of collaborative classification to support social information retrieval by organizing search results. Two approaches, collaborative indexing and search result classification based on shared classification schemes, are described and compared.

Chapter IV, “A Case Study of Use-Centered Descriptions: Archival Descriptions of What Can Be Done with a Collection” by Richard Butterworth, argues that there is a mismatch between current metadata standards for the description of archival holdings and what many users actually want to know about a collection. Use-centered descriptions are proposed as a way of systematically describing what can be done with a collection, and are, in effect, an encoding of library staff’s knowledge about what can be done with a collection. An example of its use by the University of London to encourage wider access to their archival holdings is presented.
Chapter V, “Metadata for Social Recommendations: Storing, Sharing, and Reusing Evaluations of Learning Resources” by Riina Vuorikari, Nikos Manouselis, and Erik Duval, discusses how social information retrieval systems can benefit greatly from sharable and reusable evaluations of online resources in the form of metadata. To achieve interoperability among various systems, a common framework to describe the evaluation of such resources is required. Through a review of various approaches, they present an evaluation framework and apply it to learning resources.

Section III focuses on using social networks for information retrieval. Although the idea of using social networks to find information is not new, it has gained popularity since the introduction of the Google search engine and therefore warrants an in-depth examination of the techniques involved. There are three chapters in this section:

Chapter VI, “Social Network Models for Enhancing Reference-Based Search Engine Rankings” by Nikolaos Korfiatis, Miguel-Ángel Sicilia, Claudia Hess, Klaus Stein, and Christoph Schlieder, begins this section by discussing the integration of information retrieval information from two sources—a social network and a document reference network—for enhancing reference-based search engine rankings. The authors elaborate on the basic intuitions that highlight the contribution of the social context, which can be mined from social networks, into the improvement of the rankings provided in reference-based search engines. A case study on the Web-based encyclopedia Wikipedia is presented as an illustration of the ideas introduced in this chapter.

Chapter VII, “From PageRank to Social Rank: Authority-Based Retrieval in Social Information Spaces” by Sebastian Marius Kirsch, Melanie Gnasa, Markus Won, and Armin B. Cremers, presents methods for utilizing social networks for information retrieval by applying graph authority measures to the social network. The authors present techniques for integrating authority measures in an information retrieval algorithm. To demonstrate the applicability of their algorithm, the authors examine the structure and statistical properties of social networks, and present examples of social networks as well as evaluation results.

Chapter VIII, “Adaptive Peer-to-Peer Social Networks for Distributed Content-Based Web Search” by Le-Shin Wu, Ruj Akavipat, Ana Gabriela Maguitman, and Filippo Menczer, employs social networks in information retrieval from the perspective of collaborative peer-to-peer networks. Their system, called 6Search (6S), aims to address the scalability limitations of centralized search engines. Each peer crawls the Web in a focused way, guided by its user’s information context. Each peer acts as a search “servent” by submitting and responding to queries to/from its neighbors. Prototypes of the 6S system are evaluated via simulations that model users based on actual Web crawls. The quality of the results obtained is also compared against centralized search engines such as Google.

Section IV shifts its attention to examine social issues pertaining to social information retrieval systems. This section consists of two chapters.

Chapter IX, “The Ethics of Social Information Retrieval” by Brendan Luyt and Chu Keong Lee, attempts to examine social networking and social information retrieval in the context of Habermas’s concepts of public sphere and communication actions against the problem of homophily (where a contact between similar people occurs at a higher rate than among dissimilar people), and posits that such activity is likely to increase the fragmentation of society and a reduction in social diversity as groups become more homogenous and isolated from rest of society who also have important roles to play towards learning, among others. The authors conclude with a call for more responsible and aware technological designs with an emphasis on values so that the effects of homophily are addressed.
Chapter X, “The Social Context of Knowledge” by Daniel Memmi, demonstrates that collaborative information management is now the dominant type of occupation with human information processing predominately taking place in a large social context. Interactions are supported by a range of collaborative information tools and systems that are designed to support the seeking, diffusion, and management of explicit and tacit knowledge. While some of these systems are virtually grouping similar users together (thereby promoting the problem of homophily, as discussed in Chapter IX), there is also a need to find new solutions by better understanding and modeling human cognitive processes in information processing from diverse heterogeneous sources to enable the creation of new design ideas for future systems.

Section V on social information seeking models presents a set of different information seeking models in different contexts and highlights the wide ranging applicability of this new emerging field of social information retrieval. This section consists of three chapters.

Chapter XI, “Social Information Seeking in Digital Libraries” by George Buchanan and Annika Hinze, first showed a number of contrasting uses of the social aspects of information seeking and abstracts a number of different approaches to present underlying principles, architectures, and models. Using digital library as a technological platform, the authors propose the provision of social context by the addition and integration of recommendation, alerting, and communication services into the architecture. They suggest that effective social information seeking pivots on closing the gap between human communication and the digital library.

Chapter XII, “Relevant Intra-Actions in Networked Environments” by Theresa Dirndorfer Anderson, provides a conceptual framework for relevance as a socially situated phenomenon, and goes on to describe an ethnographic study of academics engaged in research projects making relevant judgments of information when working with networked information systems. Relevance assessment when theorized as intra-action shows such judgments as emergent constructions that arose as a result of interplay between social and personal, technical and human elements in such a networked environment. The understanding of such a perspective can enable better collaborative systems to be designed, such as the facilitation for creating collaborative metadata schemes to enable alternative representations of content and to cater for different information seeking behaviors in different contexts.

Chapter XIII, “Publication and Citation Analysis as a Tool for Information Retrieval” by Ronald Rousseau, defines citation analysis in the context of applied social network theory, and highlights the relationship between citation linking as a source for information retrieval based on social interaction where authors cite and co-cite each other’s publications. Relations between citation analysis and Web links in the emerging field of Webometrics are also distinguished. The author takes a peek into the future where he envisages: the integration of local and regional citation indexes into a virtual world citation atlas, the spot translation of existing scientific non-English literature on the Web to increase the knowledge base and visibility and citation levels of such authors, and the establishment of global repositories for research and others.

Section VI concludes this book by presenting applications and case studies in social information retrieval. The focus of this section is to examine where and how social information retrieval systems have been applied. There are three chapters in this section.

Chapter XIV, “Personalized Information Retrieval in a Semantic-Based Learning Environment” by Antonella Carbonaro and Rodolfo Ferrini, applies social information retrieval techniques to the education domain. The authors introduce a personalized learning environment providing intelligent support to achieve the expectations of active learning. The system exploits collaborative and semantic approaches to extract concepts from documents, and maintain user and resources profiles based on domain ontolo-
gies. With this approach, the information retrieval process is able to produce personalized views of the learning environment.

Chapter XV, “Multi-Agent Tourism System (MATS)” by Soe Yu Maw and Myo-Myo Naing, addresses the tourism domain. The authors argue that tourism information on the Web is dynamic, and it is not easy to obtain relevant and updated information to meet an individual’s needs. To address this issue, they developed the multi-agent tourism system (MATS) with the goal of providing relevant and updated information tailored to the user’s interests. Key to the system is the rule-based personalization with collaborative filtering technique for personalization in MATS. The technique is able to address the limitations of pure collaborative filtering, including scalability, sparsity, and cold-start problems.

Chapter XVI, “Hybrid Recommendation Systems: A Case Study on the Movies Domain” by Konstantinos Markellos, Penelope Markellou, Aristotelis Mertis, Ioanna Mousourouli, Angeliki Panayiotaki, and Athanasios Tsakalidis, examines how social information retrieval can applied to e-commerce sites, focusing in particular on recommendation systems. The authors investigate hybrid recommendation systems and the way they can support movie e-shops to suggest movies to customers. Specifically, the authors introduce a recommendation approach where knowledge about customers and movies is extracted from usage mining and ontological data in conjunction with customer-movie ratings and matching techniques between customers.