The Other Side of the Social Web: A Taxonomy for Social Information Access

Peter Brusilovsky
School of Information Sciences
University of Pittsburgh
Pittsburgh, PA 15217, USA
+1 412 6249404
peterb@pitt.edu

ABSTRACT
The power of the modern Web, which is frequently called the Social Web or Web 2.0, is frequently traced to the power of users as contributors of various kinds of contents through Wikis, blogs, and resource sharing sites. However, the community power impacts not only the production of Web content, but also the access to all kinds of Web content. A number of research groups worldwide explore what we call *social information access* techniques that help users get to the right information using "collective wisdom" distilled from actions of those who worked with this information earlier. This invited talk offers a brief introduction into this important research stream and reviews recent works on social information access performed at the University of Pittsburgh's PAWS Lab lead by the author.

Categories and Subject Descriptors
H.3.3 [**Information Systems**]: Information Search and Retrieval – *search process, information filtering, relevance feedback;* H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces – *collaborative computing;* H.5.4 [**Information Interfaces and Presentation**]: Hypertext and Hypermedia – *navigation;*

General Terms
Algorithms, Human Factors

Keywords
Social information access, social navigation, social search, social tagging, social visualization.

1. INTRODUCTION
The power of the modern Web, which is frequently called the Social Web or Web 2.0, is frequently traced to the power of users as contributors of various kinds of contents through Wikis, blogs, and resource sharing sites. However, the community power impacts not only the production of Web content, but also the access to all kinds of Web content. A number of research groups worldwide explore what we call *social information access* techniques that help users get to the right information using "collective wisdom" distilled from actions of those who worked with this information earlier. Social information access technologies capitalize on the natural tendency of people to follow direct and indirect cues of others' activities, e.g. going to a restaurant that seems to attract many customers, or asking others what movies to watch.

Social information access can be formally defined as a stream of research that explores methods for organizing users' past interaction with an information system (known as explicit and implicit feedback), in order to provide better access to information to the future users of the system. It covers a range of rather different systems and technologies operating on a different scale - from a small closed corpus site to the whole Web. While the technologies located on the different sides of this stream may not even recognize each other as being a part of the same whole, the whole stream is driven by the same goals: to use the power of a user community for improving information access.

An important feature of all social information access systems is self-organization. Social information access systems are able to work with little or no involvement of human indexers, organizers, or other kinds of experts. They are truly powered by a community of users. Due to this feature, social information access technologies are frequently considered as an alternative to the traditional (content-oriented) information access technologies. In most of the cases, social information access can run in parallel with the traditional one, helping users to find resources that would be hard to find in a traditional way. In other cases where traditional information access is hard to organize (for example, in a collection of non-indexed images), social mechanisms (such as tagging) can provide a handy replacement. However, it has been more and more frequently demonstrated that most benefits could be obtained by integrating social and traditional technologies, for example, building hybrid recommender systems, which integrate collaborative and content-based recommender mechanisms [1].

The goal of this invited talk is to provide a brief overview of the emerging social information access research stream and to promote integrative and hybrid social information access technologies. To achieve this goal we offer two approaches. First, we introduce a brief taxonomy that defines the space of social information access and shows both, place of traditional technologies within this space and the opportunities for integrative work. Second we provide a review of recent works on social information access performed at the University of Pittsburgh's PAWS Lab lead by the author. These works provide various examples of bringing together different social information access technologies.

Conference'10, Month 1–2, 2010, City, State, Country.

2. A TAXONOMY FOR SOCIAL INFORMATION ACCESS

Two aspects have to be considered to understand differences between modern social information access systems and technologies. Most important is the type of information access they are attempting to support. In earlier papers on the topic we distinguished four major information access paradigms to classify adaptive information access systems: ad-hoc information retrieval, information filtering (recommendation), hypertext browsing, and information visualization [2; 3]. In ad-hoc information retrieval (IR), users get access to relevant information by issuing a query to an IR system or search engine and analyzing/accessing a ranked list of documents (for example, book records), which are returned as a result. In information filtering (IF) an information system attempts to recommend documents, which match the user's long-term interests. Traditional IF systems match a user-provided profile against a flow of incoming documents (for example, news articles) to select the most relevant items for the user. Modern recommender systems (often considered as an extension of IF) construct dynamic user profiles by observing user interactions, and as a result can produce new recommendations even in stable document collections. In hypertext browsing, a user attempts to find relevant documents by browsing links that connect documents in a collection. In information visualization, a set of documents is presented to the user using some visualization metaphor in 2 or 3 dimensions; the user observes or, in the case of interactive visualization, interacts with the visualized set to find the most relevant documents.

The type of supported information access is, however, only a part of the picture. From the "social" prospect it is also important to consider what kinds of actions of past users a specific system or technology uses to enhance traditional information access. In the field of personalized and social systems these actions are frequently called as *user feedback*. The reason for this name is that in the oldest adaptive filtering and retrieval system user feedback (typically provided as a yes/no or a multi-point scale rating) was the only kind of user actions that these systems were able to trace. User ratings are an example of so-called explicit feedback - actions that the users explicitly provide to express some opinion about an information item. While users ratings are still very popular source of information in some social information access systems (for example, collaborative recommendation systems), it has long been recognized that user ratings form a very small fraction of user interaction with information. As a result, more recent work focused on various *implicit feedback* indicators [4] that assess user attitude to an information items provided *implicitly* i.e., through various users actions that were not specifically performed to express this attitude. Most popular source of implicit feedback is user clicks and dwell time (known together as *clickstream*). Both clicks and time provide a good evidence of user interest to an item or item usefulness for a given user. While this evidence is less reliable, it is more readily available in many contexts. A more reliable evidence could be extracted by tracing various kinds of user manipulations with a document – from mousing over document to saving, bookmarking, or printing a document [5].

Explicit ratings and clickstream are the extreme points of implicit-explicit continuum of feedback. Nowadays, social information access system employ a whole range of user actions that bridge the gap between these two extremes such as user queries, comments, and tags. These new kinds of actions are harder to classify into explicit/implicit bins. For example, tagging is certainly an explicit action and to some extent it expresses user attitude to an item. Yet, tagging is typically done not to express feedback or attitude, but to organize personal or social information space. As for comments, while in some systems (like e-commerce) comments are done to express attitude (i.e., clearly explicit feedback), in other systems (like online reading) comments are mostly notes to oneself.

The kind of supported information access and the kind of considered user actions form two dimension of the taxonomy that defines the space of social information access technologies (Table 1). Presenting this space in two dimensions shows that this space is not really well explored. The table shows some well-explored pockets on the crossroads of information access types and trackable actions surrounded by larger *terra incognita*. The reason for this is that classic social information access technologies were developed in tight conjunction with corresponding information access paradigms and with little interest to what *other* paradigms can offer. As a result, the type of information access supported by a specific technology to a large extent determined the kind of social enhancement provided by the technology and the kind of user actions that the "social" version of this paradigm harvested to provide social access.

For example, classic *social navigation* technologies (history-enriched environments) were developed to support browsing-based access. This context requires navigation support systems, which can help the users to decide, which of many links on the current page to follow. The natural approach to using the community wisdom is to show "where did the people go" [6] by augmenting links with digital "wear" indicators. The natural approach to collect this wisdom is to track user page visits [7] or link traversals [8].

Collaborative filtering technologies were designed as an extension of traditional information filtering that has been long based on user explicit feedback [9]. As a result, the majority of collaborative filtering systems still use ratings as the only kind of feedback. Ratings are tightly integrated into the very algorithms of traditional collaborative filtering.

Social search technologies were developed to support traditional IR information access. In this context, users expect to see a ranked list of relevant resources. The natural approach to using the community wisdom is to insert community-relevant links into the list or results [10; 11] or stress, which of the returned documents are not only relevant, but also appreciated by the community [12; 13]. A reliable approach to collecting this wisdom is to track connections between queries and items selected [11] or rated [10; 13] by the community members in the context of these queries.

While the social search technology shows an example of horizontal integration (using several kind of feedback to support one kind of information access), *social bookmarking* technology presents a different integration case in social information access. In this case, a single kind of user feedback is used (tagging), yet this feedback is engineered to enhance at least two different kinds of information access. Not only social tags are used to improve search by bridging the well-known author-user mismatch, they also provide an alternative mechanism to browse information through tag indexes and tag clouds.

The simple integration examples provided by social search and social tagging shows that the kind of supported information access and the kind of tracked user actions should not be necessary connected into tight couples. I.e., within one system, one kind of

information access can be supported by social wisdom extracted from several different kinds of feedback. Excellent examples of horizontal integration are provided by modern recommender systems that explored how almost every possible kind of feedback can be used for recommendations. Yet, even in this area a single system uses one, rarely two kinds of feedback. The cases of vertical integration where a specific kind of tracked actions could be used to support more than one kind of access are even more rare.

One of the goals of our recent work at PAWS Lab was exploring the space of integrated social information access technologies. Next sections attempt to provide a few examples of social systems that integrate several technologies and spans several "cells" on the social information access map (Table 2).

Table 1. The space of social information access

Past User Behavior	**Explicit Actions**			**Implicit Actions**		
Supported Access Types	**Tags**	**Comments**	**Ratings**	**Queries**	**Clicks/Time**	**Manipulations**
Search	Social tagging systems		Social search			
Browsing					Social Navigation	
Recommendation			Collaborative Filtering			
Visualization						

Table 2. Examples of integrative social information access systems

Past User Behavior	**Explicit Actions**			**Implicit Actions**		
Supported Access Types	**Tags**	**Comments**	**Ratings**	**Queries**	**Clicks/Time**	**Manipulations**
Search		KS, ASSIST		ASSIST	KS, ASSIST	
Browsing	CoMeT	KS, ASSIST	CourseAgent	ASSIST	CoMeT, KS, ASSIST	CoMeT
Recommendation			CourseAgent			
Visualization		KS			KS	

3. TOWARDS INTEGRATED SOCIAL INFORMATION ACCESS SYSTEMS

3.1 CourseAgent

One of the simplest kinds of integration is using a specific kind of user feedback to support two types of social information access: one that is traditionally based on this feedback and one that usually exploits another kind of feedback. An example of this integration is CourseAgent [14]. CourseAgent is a system that helps graduate students at our School to select courses that are most relevant to their career goals. The kind of feedback exploited by CourseAgent is user ratings. A student is expected to rate the relevance of each taken course to one or more career goals using explicit 5-point scale. This kind of feedback typically drives social recommendation mechanism. CourseAgent does offer course recommendations. However, more visible form of social information access offered by this system is social navigation support. Relevant courses are annotated with different adaptive visual cues in every context where the courses are shown to students like course schedules or course catalogs. This allows the system to provide in-context guidance beyond the usual rating list. Title and Authors

3.2 CoMeT

Another simple kind of integrated information access is using a range of observable user actions to offer an extended version of a specific kind of social information access. A good example of this is CoMeT, a system that guides its users to most interesting research talks offered by several universities in Pittsburgh [15]. CoMeT recognizes that users are rarely interested to rate research talks so it attempts to use several implicit evidences of interest: accessing a talk, adding a talk to one's schedule or calendar, sending information about a talk to friends, tagging a talk. This feedback is used to offer integrated social navigation: i.e., it guides users to interesting talks right in the context where these talk are shown – day, week, and month calendars.

3.3 Knowledge Sea (KS)

An extensive exploration of social information access was performed in the Knowledge See (KS) project. Knowledge See is an educational system that was designed to guide students in a class to most useful and relevant reading fragments. KS started as a traditional social navigation system [7]: it tracked class document access and offered visual cues that indicated cumulative student traffic. Thus the system guided the users to the pages that were most visited by a class as a whole. Traffic, however, appeared to be a relatively noisy indicator of interest. Subsequent versions explored more reliable indicators – time spent reading [16] and annotations created by the users while reading [17]. The system was also extended with a social search [12] and a social visualization [18] components that used both the clickstream and the annotations to support search and visualization-based access to class readings.

3.4 ASSIST

Our most advanced attempt to create an integrated system for social information access was performed collaboratively with a research team from Dublin lead by Barry Smith. When we started this collaboration, our team accumulated a lot of experience in various kinds of social navigation while Barry's team with its work on several versions of I-Spy [11] had a great experience in social search. Yet, at that time both teams followed traditional approaches. Our team delivered browsing support using clickstream left by the users when browsing and Barry's team focused on queries and page access combinations left by the users in the search process. Both teams recognized, however, that a typical Web information access session includes a combination of search and browsing where users most typically start from search, but then follow a sequence of link to the destination. What we attempted to do is to design an integrated system that can support both kinds of information access using both kinds of implicit feedback. The result was the ASSIST architecture. It was implemented and explored in several contexts such as social access to ACM Digital Library and social exploration of YouTube videos [19; 20].

4. ACKNOWLEDGMENTS

A number of students working in PAWS Lab provided critical contributions to the work presented in this talk. The author wants to specifically acknowledge Rosta Farzan, Jaewook Ahn, Chirayu Wongchokprasitti, Sherry Sahebi, and Denis Parra. This work was partially supported by NSF grants IIS 0447083 and DUE 0840597 to the author.

5. REFERENCES

[1] Burke, R. Hybrid Web recommender systems. In: Brusilovsky, P., Kobsa, A. and Neidl, W. (eds.): The Adaptive Web: Methods and Strategies of Web Personalization. Lecture Notes in Computer Science, Vol. 4321. Springer-Verlag, Berlin Heidelberg New York, 2007, 377-408.

[2] Brusilovsky, P. and Tasso, C. Preface to special issue on user modeling for Web information retrieval. User Modeling and User Adapted Interaction, 14, 2-3 (2004), 147-157.

[3] Brusilovsky, P. Social Information Access: The Other Side of the Social Web. In: Geffert, V., et al. (eds.) Proc. of SOFSEM 2008, 34th International Conference on Current Trends in Theory and Practice of Computer Science, Springer Verlag, 5-22.

[4] Claypool, M., Le, P., Wased, M., and Brown, D. Implicit interest indicators. In: Proc. of 6th International Conference on Intelligent User Interfaces, ACM Press, 33-40.

[5] Goecks, J. and Shavlik, J. Learning user's interests by unobtrusively observing their normal behavior. In: Lieberman, H. (ed.) Proc. of 2000 International Conference on Intelligent User Interfaces, ACM Press, 129-132.

[6] Dieberger, A., Dourish, P., Höök, K., Resnick, P., and Wexelblat, A. Social navigation: Techniques for building more usable systems. interactions, 7, 6 (2000), 36-45.

[7] Brusilovsky, P., Chavan, G., and Farzan, R. Social adaptive navigation support for open corpus electronic textbooks. In: De Bra, P. and Nejdl, W. (eds.) Proc. of Third International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'2004), Springer-Verlag, 24-33.

[8] Wexelblat, A. and Mayes, P. Footprints: History-rich tools for information foraging. In: Proc. of ACM Conference on Human-Computer Interaction (CHI'99), 270-277.

[9] Konstan, A., Bradley, N.M., Maltz, D., Herlocker, J.L., Gordon, L.R., and Riedl, J. GroupLens: Applying collaborative filtering to usenet news. Communications of the ACM, 40, 3 (1997), 77-87.

[10] Jung, S., Harris, K., Webster, J., and Herlocker, J.L. SERF: Iintegrating human recommendations with search. In: Proc. of ACM 13th Conference on Information and Knowledge Management, CIKM 2004, 571-580.

[11] Smyth, B., Balfe, E., Freyne, J., Briggs, P., Coyle, M., and Boydell, O. Exploiting Query Repetition and Regularity in an Adaptive Community-Based Web Search Engine. User Modeling and User-Adapted Interaction, 14, 5 (2004), 383-423.

[12] Ahn, J.-w., Farzan, R., and Brusilovsky, P. Social search in the context of social navigation. Journal of the Korean Society for Information Management, 23, 2 (2006), 147-165.

[13] Kantor, P.B., Boros, E., Melamed, B., Meñkov, V., Shapira, B., and Neu, D.J. Capturing human intelligence in the net. Communications of the ACM, 43, 8 (2000), 112-116.

[14] Farzan, R. and Brusilovsky, P. Encouraging User Participation in a Course Recommender System: An Impact on User Behavior. Computers in Human Behavior, 27, 1 (2011), 276-284.

[15] Sahebi, S., Wongchokprasitti, C., and Brusilovsky, P. Recommending research colloquia: a study of several sources for user profiling. In: Proc. of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec 2010) at the 2010 ACM conference on Recommender systems, RecSys '10, ACM, 32-38

[16] Farzan, R. and Brusilovsky, P. Social navigation support in E-Learning: What are real footprints. In: Anand, S.S. and Mobasher, B. (eds.) Proc. of IJCAI'05 Workshop on Intelligent Techniques for Web Personalization, 49-56.

[17] Farzan, R. and Brusilovsky, P. AnnotatEd: A social navigation and annotation service for web-based educational resources. New Review in Hypermedia and Multimedia, 14, 1 (2008), 3-32.

[18] Ahn, J.-w., Farzan, R., and Brusilovsky, P. A two-level adaptive visualization for information access to open-corpus educational resources. In: Brusilovsky, P., Dron, J. and Kurhila, J. (eds.) Proc. of Workshop on the Social Navigation and Community-Based Adaptation Technologies at the 4th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, 497-505.

[19] Farzan, R., Coyle, M., Freyne, J., Brusilovsky, P., and Smyth, B. ASSIST: adaptive social support for information space traversal. In: Proc. of 18th conference on Hypertext and hypermedia, HT '07, ACM Press, 199-208.

[20] Coyle, M., Freyne, J., Brusilovsky, P., and Smyth, B. Social Information Access for the Rest of Us: An Exploration of Social YouTube. In: Nejdl, W., Kay, J., Pu, P. and Herder, E. (eds.) Proc. of 5th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'2008), Springer Verlag, 93-102.