

Social Information Access: The Other Side of the Social Web

Peter Brusilovsky

School of Information Sciences, University of Pittsburgh
135 N. Bellefield Ave. Pittsburgh, PA 15260, USA
peterb@pitt.edu

Abstract. Modern Web, which is frequently called Social Web or Web 2.0, celebrates the power of the user community. Most frequently it is associated with the power of users as contributors or 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 work on social information access techniques, which help users get to the right information using “community wisdom” distilled from tracked actions of those who worked with this information earlier. The paper provides an overview of this research stream focusing on social search, social navigation, and social visualization techniques.

1 Introduction

Modern Web, which is frequently called Social Web or Web 2.0, celebrates the power of the user community. Most frequently it is associated with the power of user community as content contributors through Wikis, blogs, and resource sharing sites. However, the power of community impacts not only the production of Web content, but also the access to all kinds of Web content. A number of research groups worldwide work on *social information access* technologies, which help users get to the right information using “community wisdom” distilled from the 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. This stream has to be considered as emerging. 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.

The social information access is frequently considered as an alternative to the traditional (content-based) information access. 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 [10]. 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.

As a type of information access, which can offer multiple benefits, while being relatively easy to organize and maintain, social information access attracts more and more researchers and practitioners. The paper provides an overview of this emerging research stream while focusing primary on social search, social navigation, and social visualization techniques.

2 The Emergence of Social Information Access

The pioneer work on social information access appeared in the early 1990s, but the emergence of social information access as a research area happened between 1994 and 1996. This period brought many innovations in all areas related to information access, as research teams investigated new approaches to help users in the rapidly expanding information space. In the area of social information access, these years produced two well-defined research streams: collaborative filtering and social navigation.

Collaborative filtering [36; 45] attempted to propagate information items between users with similar interests. This technology enabled social forms of *information filtering and recommendation*. The emergence of collaborative filtering is typically traced back to the *Information Tapestry* project and a seminal paper about it, which coined the term [23]. Information Tapestry employed an approach, which was later called *pull-active* collaborative filtering [36]: to receive social guidance, its users had to actively query the community feedback left by earlier users. Another early example of “active” filtering (in this case *push-active*) was provided by Lotus Notes recommender [39], which encouraged users to send interesting documents directly to their colleagues. While both of these examples were highly influential, the mainstream work in the area of collaborating filtering focused on *automatic* approaches based on matching users with similar interests and cross-recommending positively-rated items. This direction was lead by such pioneer systems as GroupLens [43], Ringo [47], and Video Recommender [26].

Social navigation [13] in its early forms attempted to visualize the aggregated or individual actions of community users. The motivation behind this work was that that these “footprints” can help community users to navigate through information space.

By its nature, social navigation supported user browsing activity. The ideas of social navigation are frequently traced back to the pioneer *Read Wear and Edit Wear* system [27]. This system visualized the history of authors' and readers' interactions with a document enabling new users to quickly locate the most viewed or edited parts of the document. Social navigation in information space as well as the term *social navigation* was introduced two years later by Dourish and Chalmers as "moving towards cluster of people" or "selecting subjects because others have examined them" [16].

The pioneer systems Juggler [12] and Footprints [50] used the ideas of social navigation to help users navigating in two kinds of information spaces – a Web site and a text-based virtual environment (MOO). Both systems attempted to visualize "wear" traces left by the system users in order to guide future users. In addition to this *indirect* social navigation, Juggler also implemented several types of direct social navigation (for example, allowing users to guide each other directly through chat). This allowed Dieberger [12] to start the process of generalizing the ideas of social navigation.

Further generalization of the field of social navigation was propelled by several workshops, which gathered like-minded researchers, and publications, which steamed from these workshops [14; 28; 41]. As a result of this active idea exchange, the scope of social navigation was broadened to cover all approaches, which use social feedback as a source of knowledge in assisting the users. The notion of social feedback was also broadened to include a variety of options – from navigation traces to rich explicit feedback and resource annotations. The newly-defined field of social navigation included two main groups of systems - collaborative filtering systems and Footprints-type systems, which were referred to as history-enriched environments.

More recently, the set of social information access technologies was extended with *social search* and *social bookmarking* systems. The research on collaborative information retrieval [33] produced some truly social search systems that attempted to help new searchers by capitalizing on past successful searches of similar users. The pioneer AntWorld system [32] was soon followed by several similar systems such as I-SPY [48] and SERF [30]. Another group of pioneer systems such as Siteseer [44], PowerBookmarks [38], and WebTagger [34] started the collaborative bookmarking research stream. Researchers working on collaborative bookmarking explored different ways of bookmark sharing to help new users locate useful information already discovered and classified by others and invented social tagging mechanism. In less than 10 years social bookmarking and tagging system popularized by such systems as *del.icio.us* or *Flickr.com* grew into a new major Internet technology [24].

It's important to stress again that each social information access technology achieved success by collecting community wisdom in a specific form, and enabling users working with a specific information access paradigm to benefit from this. Being quite different by nature, the social information access technologies have a lot in common. Unfortunately, no recent attempts were made to provide a comprehensive overview of these technologies. The important "unifying" books and papers on social navigation [14; 28; 41] appeared too early to cover and integrate more recent technologies. This paper attempts to fill this gap to some extent and provide a unified view on all technologies listed above. It looks at the problem from the prospective of information access, which is a slightly different and in some sense more narrow angle

than the one taken in the books cited above. The next section suggests a framework, which can be used to classify existing social information access technologies. The remaining sections focus on social search, browsing, and visualization, which are arguably the least known social information access technologies.

3 A Taxonomy of Social Information Access Technologies

To understand differences between modern social information access systems, it is useful to understand which type of information access they are attempting to support. In an earlier paper we distinguished four major *information access* paradigms to classify adaptive information access systems: ad-hoc information retrieval, information filtering, hypertext browsing, and information visualization [9]. 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 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 analysis of modern social information access technologies shows that different technologies were, in fact, developed in conjunction with different information access paradigms. The type of information access supported by a specific technology to a large extent determines the nature of this technology and its difference from other paradigms. 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" [13] by augmenting links with digital "wear" indicators. The natural approach to collect this wisdom is to track user page visits [8] or link traversals [50].

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 [30; 48] or stress, which of the returned documents are not only relevant, but also appreciated by the community [2; 32]. A reliable approach to collecting this wisdom is to track connections between queries and items selected or rated by the community members in the context of these queries [30; 32; 48].

The presence of the context (such as current query or location in the hyperspace) helps both IR and browsing systems to identify similar users without more sophisticated tracking. These systems can accumulate community wisdom by query, by link, or by page – but not necessarily by user. As a result, there is no need for a user to login, although both kinds of systems can provide better service if the user can pick up one of the sub-communities tracked by the system [19; 48]. As a result, social search and browsing systems are very easy to integrate into any kind of existing Web systems.

In contrast, collaborative filtering technologies were designed to work in the most challenging situation – long-term information filtering, i.e., in a situation with no current context. As a result, collaborative filtering technologies have to track all user activity on the individual level, construct detailed user profiles and apply sophisticated approaches to match similar users.

Let's skip social visualization, since this area is not sufficiently developed, and move to social bookmarking. Social bookmarking technology presents an interesting case in social information access. Unlike other listed technologies, modern social bookmarking does not really augment any of the traditional information access paradigms. Instead, it provides an alternative mechanism to access information using community-contributed tags. In this case the social wisdom, which other social systems accumulate in some hidden form “behind the stage”, becomes visible as a tightly interlinking *tag space*. The information can be accessed through this tag space using traditional access paradigms such as tag searching, tag navigation, or tag visualization in the form of *tag cloud*. So, social tagging augments several traditional information access paradigms by providing additional community-created space where these paradigms can be applied.

As could be noticed from the discussion above, there are at least three different levels on which social information technologies accumulate user information. The most relaxing is the *all-users* level where a system does not distinguish its users and accumulates all past usage activity “in a single pile”. This approach was used in several early social search and browsing systems [12; 21; 32; 50] and is still appropriate in the situations where the body of users is reasonably sized and their information needs are similar [30; 37].

In the situation where the body of users is large and diverse (which is the case for the majority of “open Web” systems), the presence of context such as the current query or Web page become insufficient to reliably identify similar users. The information needs of the users passing a specific link or issuing a specific query may still be too diverse. This situation caused several recent social information access projects such as I-SPY [48], Knowledge Sea II [17], Conference Navigator [19], and ASSIST [20] to start accumulating social wisdom for different subsets of the whole body of users independently. The subsets could be of different nature and size – from a *community* of like-minded users [19; 48], which can include hundreds, to a small *group* of users (such as a college class) joined by the same information goal [17]. This approach can be called *community-level* or *group-level* information access. In this paper we will use the latter term, since it stresses the clear difference from “all-users” approach. Group level access provides an attractive compromise between all-users and individual level. It can provide very reliable social guidance without a requirement to authenticate, which can be a stumbling point for several reasons [48].

Comparable with all-user level, group-level approaches sort user feedback in multiple “group bins”, which may result in community wisdom becoming too sparse. The challenge for the developers of group-level social systems is to engineer the groups of the proper size to make sure that the volume of social feedback is sufficient for useful guidance. If the volume of social information provided by an average user is large and the volume of resources used by a group is not large, some good social guidance can be provided even for small groups of users. In addition, the problem of sparsity in group-level systems can be addressed by propagation of social feedback between groups [22].

The most fine-grained level of wisdom collection is *user-level*, where each piece of feedback is associated with an individual user and accumulated in the profile of this user. User-level tracking is a standard for *personalized information access* techniques such as adaptive navigation support [7] or personalized search [40]. Among social information access systems, user-level tracking is critical only for collaborative filtering systems. However, modern approaches to increasing user contributions in social systems such as incentives [11] or “do it for yourself” [18; 19] may require user-level tracking in any kind of social information access systems. User-level tracking adds another burden to the developers of a social information system: how to attribute all kinds of feedback left by an individual user to the profile of that user. To some limited extent, it can be done by tracking a user within a single session, but in general case it requires a long-term user profiling and reliable user authentication (such as password-protected login).

Due to the lack of space, this paper offers no further attempts to classify social information access systems. Instead, we refer to several collaborative filtering and social navigation papers [13; 36; 45; 49], which offer several useful dimensions to classify social systems. While each of these dimensions was suggested to classify a special group of systems, most of them can be successfully applied to classify all kinds of social information access approaches.

The remaining sections attempt to provide examples of social navigation technologies discussed above. Since both collaborative filtering and social tagging are rather well publicized in research literature, the presentation is focused on the least known social navigation, social search, and social visualization technologies.

4 Social Browsing

Social browsing systems use “community wisdom” to assist their users in the process of browsing a hyperspace or another virtual environment (such as MUD). As mentioned before, social browsing support systems collect the community wisdom by tracking two kinds of information – link traversals (link-centric approach) and page visits (page-centric). While these kinds of information look similar, they are quite different and each has its own advantage. Link traversal is the most browsing-specific kind of information. It allows not just counting how many users visited a specific page, but also distinguishing where they came from, i.e. the context in which this visit took place. The context in hypertext is quite important. Visiting the same page may be very relevant for the community members in one context (i.e., from one page linked

to it) and much less relevant in the other context (i.e., from another linked page). While link-centric approaches take this context into account, they still do not track user activities within a page and this can't distinguish links to really useful pages from "tar pits" – low-value pages hidden behind attractive links. In contrast, page-centric approaches do not distinguish how the community users get to a specific page. It decreases the precision of system advice and makes it necessary to group users in reasonably homogeneous communities and track navigation within each community. At the same time, it is easier for page centric approaches to take into account user behavior within each page (time spent, browsing, annotation) and thus distinguish good pages from tar pits.

Most of the early social browsing systems were link centered, although they still differ in respect to how the accumulated link traversal information was used. One kind of systems followed the traditions of *intelligent hypertext* and used link traversal information to periodically modify and expand the hyperspace link structure. For example, a system presented in [5] can add a link from page A to page C if existing links from A to B and from B to C were frequently used in succession. This kind of re-structuring typically requires global log analysis and is better performed *off-line*. The other kind of systems was inspired by the social navigation ideas and used link traversal information to dynamically generate history-enriched environments, where the behavior of past users is made visible.

A classic example of a link-centric history-enriched system is Juggler [12]. Juggler is a MOO system, a text-based virtual environment, which is conceptually similar to hypertext. A MOO system consists of rooms connected by passages. Every room exit in Juggler tracked how frequently it was used and showed this information (textually) as wear on a door mat. Wear decayed over time in order to reflect the behavior of natural environments, such as a path in the forest, which may fade and disappear if not used frequently. This feature was introduced to guide people towards popular locations and make encounters in the environment more likely. The same idea was implemented in the Web navigation context in Footprints system, which visualizes usage paths in a web site [50]. The Footprints system allows users to leave activity traces in the virtual environment and visualizes these traces to guide future users. With the Footprints system, new users can see the popularity of each link on the current page and make navigation decisions.

An example of a simple page-centric social browsing system is CoWeb [13; 15]. CoWeb is a history-enriched Wiki system. To increase user awareness of what is going on in the Wiki space and to guide the users to most recently updated or visited pages all links inside the CoWeb were annotated with activity markers (Fig. 1). An *access marker* showed access information using a metaphor of *footprints*. Small footprint symbols in three different colors (gray, orange, red) were placed right next to links to indicate the amount of traffic the page behind that link received in the past 24 hours. A *novelty marker* indicated another kind of community activity, which is specific to Wiki: page updates. Using three different novelty levels, it indicated how long ago that page was last modified.

A more sophisticated example of page-centric social browsing system is the Knowledge Sea II [8]. Knowledge Sea II uses ideas of social navigation to support both browsing and visualization access to information. The visualization-based access is provided through an 8 by 8 cell-based map of the information space. This map is

assembled using Kohonen's Self-Organized Map (SOM) technology [35] from about 25,000 Web pages devoted to C programming language. Every cell on a resulting map provides access to a subset of these pages. By clicking on a cell, the user can open it and get access to the set of pages located in this cell (Fig. 2). An interesting property of SOM technology is that it places similar pages into the same or adjacent cells on the map, so the result presents a reasonably good semantic map of the information space. The cells of the map are marked by keywords, which are most frequently found in its pages and by landmark resources located in the cell.

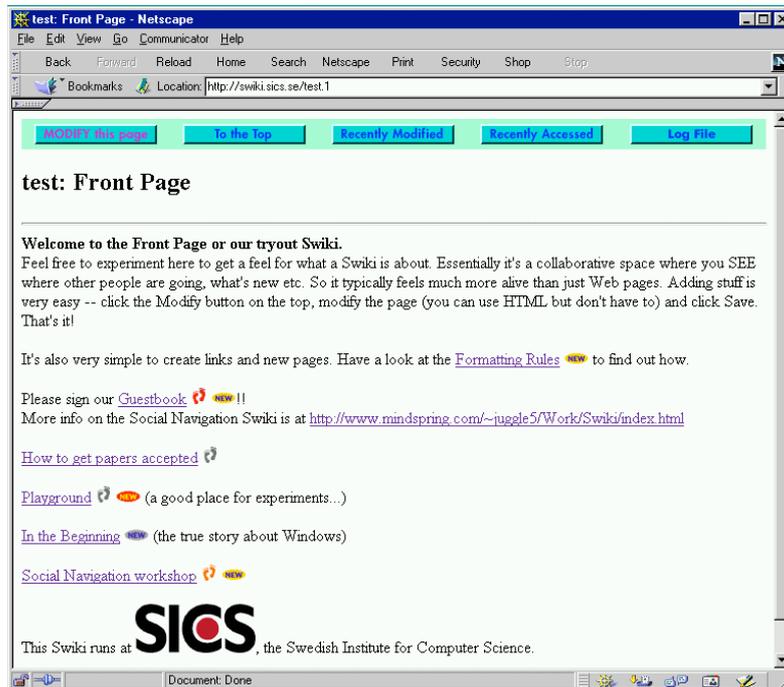


Fig. 1. Page-centric social browsing support in CoWeb. Two kinds of activity markers indicate when the page behind the link was last modified and also whether it was recently accessed. Used from [13] with the permission from the author.

The browsing-based access is provided through the hierarchical structure of the C programming tutorials assembled by the system. Each tutorial site is organized as a tree with table of contents, sections, and subsections. The home page of Knowledge Sea II provides access to the root pages of all these tutorials. Starting from that, users can navigate down to the sections or subsections of interest.

The community wisdom in Knowledge Sea II is collected by tracking two kinds of page-centric user information: timed page visits (traffic) and page annotations. This information is used to generate a history-enriched environment with two types of visual cues, which change the appearance of links on the pages and map cells presented to the user (Fig. 2). These cues are based on the two kinds of tracked information and are known respectively as traffic- and annotation-based social

navigation support. The system generates appropriate cues individually for each user by analyzing past individual activities of the user and other users belonging to the same group.

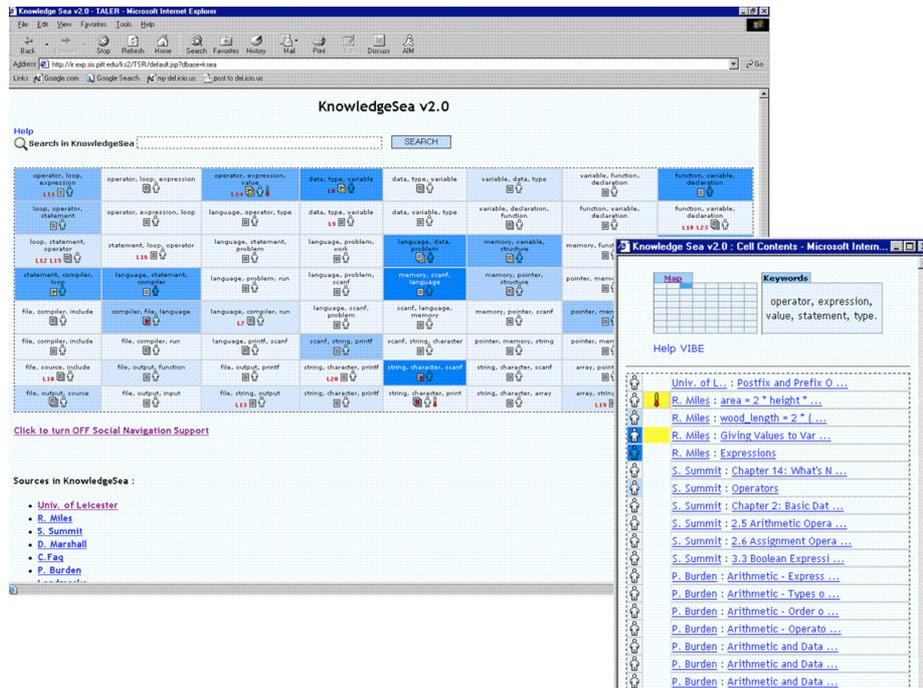


Fig. 2. Social navigation support in the Knowledge Sea II system. The knowledge map is shown on the top left and an opened cell on the right. The list of links to the tutorial roots is shown on the bottom left. A darker blue background indicates documents and map cells that have received more attention from users within the same group. Human icons with darker colors indicate documents and cells that have received more attention from the user herself. Similarly, a yellow background indicates density of annotations.

Traffic-based navigation support attempts to express how much attention the user herself and other users from the same group paid to each of 25,000 pages that the system monitors. The level of attention for a page is computed taking into account both number of visits and time spent on the page and is displayed to the user through an icon that shows a human figure on a blue background. The color saturation of the figure expresses the level of the user's own attention while the background color expresses the average level of group attention. The higher the level of attention is, the darker the color appears to the user. The contrast between colors allows the user to compare her navigation history with the navigation of the entire group. For example, a light figure on a dark background indicates a page that is popular among group members but remains under-explored by the user. The color of the map cell and the human figure shown in the cell is computed by integrating attention parameters of all pages belonging to that cell.

Annotation-based navigation support uses a similar approach to represent the number of page annotations made by the users from the same group. Users can annotate each page in the system. Users can also indicate that a note is praise (i.e., the page is good in some aspect). While users make annotations mainly for themselves, Knowledge Sea II allows all users of the same group to benefit from collective annotation behavior. The yellow annotation icon shown next to the blue traffic icon shows the density and the “praise temperature” of annotation for each page. The more annotations a page has, the darker the yellow background color appears to the user. The temperature shown on a thermometer icon indicates the percentage of praise annotations.

5 Social Search

Let’s start with narrowing down the notion of social search. Over the last 10 years, researchers and practitioners suggested a number of creative approaches to use social information for the improvement of Web search. The most well-known example is Google PageRank [6], which improves ranking by using Web link structure - a communal product of Web page authors. However, in the light of definition of social information access suggested in the introduction, only a subset of these approaches qualify as social search – those based on taking into account the past behavior of information system *users*. Similar to the case of social browsing, the approaches based on past search behavior can be further classified into *off-line* and *on-line* approaches. Off-line approaches seek to use various kinds of social information to improve indexing and ranking delivered by search engines. For example, weights of query terms may be increased in the index of documents, which were selected in response to this query [46]. On-line approaches seek to assist users in their search dynamically by accessing the accumulated social information during the very search process. By *social search*, modern sources traditionally mean on-line approaches. We will follow this tradition below.

Social search approaches do not attempt to modify the behavior of search engines, but instead apply the community wisdom before the search engine is invoked or after the results are returned. Pre-search approaches use community wisdom for social query expansion. This idea was suggested by Fitzpatrick and Dent [21] (who also coined the term *social search*) and expanded in several other projects [3; 29]. Due to the lack of space, we will not provide more details about these approaches.

Post-search approaches use the community wisdom to manipulate results returned by a search engine. It can be done in several ways: re-ranking results returned by the engine according to their social value, inserting additional results that were not in the list originally, and adding social visual cues to the listed results. A pioneer attempt to use social information for ranking search results was made at the second part of the 1990's by DirectHit (www.directhit.com) search engine. DirectHit used query logs to measure the popularity of result selections for each given query. This data was used for “social ranking” for future occurrences of this query. Unfortunately this approach turned out to be too simplistic for a large-scale search engine: it never became a success story. The study [4] showed that DirectHit falls below the satisfaction of an

average user. The most cited reason of DirectHit failure was low query repetition, which made the social data collected by it too sparse to use frequently and reliably. User diversity was another likely contribution: users with different goals and interests may prefer different results returned by the same query. Finally, the proposed approach to link ranking was too easy to abuse by malicious users who wanted to promote their favorite pages.

A more sophisticated approach was pioneered at the same time [31] in AntWorld system [32] and later re-used in SERF [30]. AntWorld introduced the concept of a *quest*, which is an information goal pursued by a user over a sequence of queries (Fig. 3). The system successfully encouraged its users to describe their quests in natural language and used this description to determine inter-quest similarity. During their search, the users were able to rank search results by their relevance to the original quest (not a query used to obtain this result!). These innovations allowed the system to address to some extent the sparsity and reliability problems. To determine documents, which are socially relevant for a particular quest, the system looked for positively ranked documents in past similar quests. The system assisted the user by adding socially relevant documents to the list of search results and also adding a small *ant* icon to socially relevant links returned during each search within the quest.

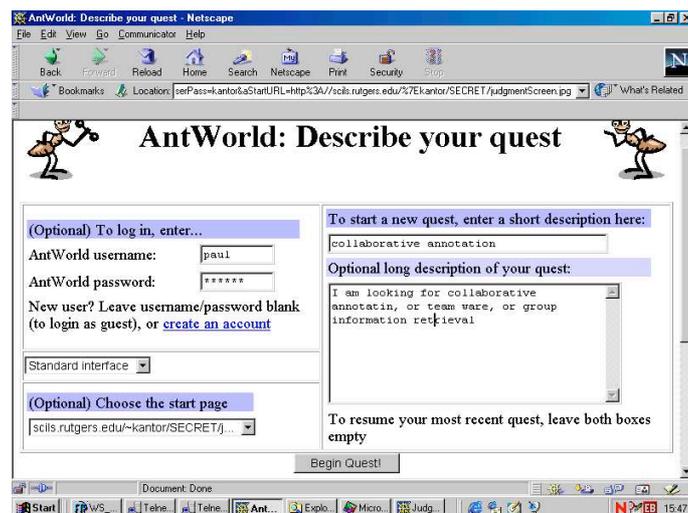


Fig. 3. AntWorld encouraged its user to describe their quests in natural language.

A different approach to improve the reliability of social recommendation was suggested in I-SPY, which worked as a post-filter for regular search engines [48]. I-SPY combined and extended DirectHit and AntWord approaches. Like DirectHit, it used indirect feedback (the fact of document selection from the list of results) and accumulated social information on the level of a single query, not a multi-query quest. Thus it targeted typical search engine users who are less likely to specify wordy quests or rate search results. At the same time, I-SPY used query similarity to fight sparsity in the same way as AntWorld used quest similarity. When presented with a

new query, in addition to retrieving the appropriate results from the underlying search engine, I-SPY retrieved any search sessions associated with *similar* queries and combined the results selected during these sessions. Results with high social scores were promoted ahead of the results returned by the search engine (Fig. 4).

The key innovation introduced by I-SPY was community-based search. I-SPY allowed users to join one of many communities and perform the search from the community prospect. All social feedback was collected and used independently for each community, i.e., on a group-level, which increased the reliability of search results. A more recent example of social search system with group-level collection of social feedback is the search component of Knowledge Sea II [2].

The screenshot shows the I-SPY search engine interface. At the top, there is a search bar with the text 'shakey' and a 'Search' button. To the right of the search bar is a 'PRIVATE SEARCH' checkbox, which is checked. Below the search bar, there are two tabs: 'Communities' and 'About I-Spy'. The main content area is divided into several sections:

- Related Information:** A list of recent queries and web pages.
- I-Spy Recommends:** A section showing 2 of 4 promoted results. It includes a slider bar and a tip: 'Tip: Move the slider to see more recommended results. Recommended results are marked with [eyes icon]'. The first result is 'SRI Technology: Shakey the Robot' with a pair of eyes icon. Below it is 'Moravec Robot book figure' also with a pair of eyes icon.
- Other Matching Results:** A list of other results, including 'shakey.com - Welcome'.

Annotations in red text and arrows point to specific features: 'Private Search check box' points to the checkbox; 'Slider Bar' points to the slider; 'Promotional Results' points to the 'SRI Technology: Shakey the Robot' result; 'Related Queries' points to the 'Related Queries' link; and 'Meta-search Results' points to the 'shakey.com - Welcome' result.

Fig. 4. Promotion of community-relevant results in I-SPY search engine. Promoted links are placed on the top of the results list and annotated with a “pair of eyes” icon.

6 Social Visualization

Social visualization is the least investigated area of social information access, however, it is one of the most promising contexts of the application of social access ideas, due to the highly expressive power of information visualization (IV). Information visualization allows users to see a set of information resources as a whole, while still being able to recognize individual resources, their properties, and their relationships to each other using relative positioning and visual cues such as shape, color, and size. In contrast to the 1-dimensional guidance provided by the ordered list of objects in IR and IF, a typical IV can use two spatial dimensions in addition to the item appearance to express any information important for the users.

This could be critical for social information access systems, which have to present community wisdom to their users *in addition* to regular information presented in traditional information systems. Giving an individual the ability to manipulate the relative positioning of documents and visual cues allows a social information access system to present more aspects that are valuable to their users. In addition, the higher level of interactivity supports more reliable user tracking techniques.

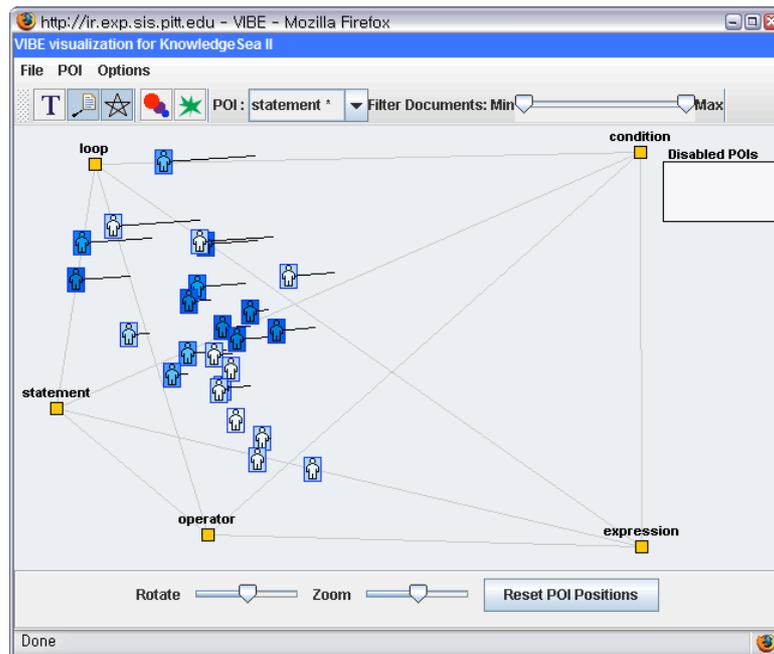


Fig. 5. A visualization of a Knowledge Sea II cell in Social VIBE system. Information resources are represented by “human on a background” icons. The color density of the background indicates timed page traffic for a group of users.

To some extent, social visualization resembles social browsing and social information filtering. Like in social hypertext browsing, information visualization supports user-driven exploration of information items. However, this exploration is done not by moving from item to item using links, but by exploring and manipulating the visual representation of these items. As a result the context (such as current page or query) either does not exist or is hard to define. It makes the visualization close to information filtering. These similarities hint that approaches from social browsing and collaborative filtering could be appropriate for building social visualization. Indeed, both user-profile based approaches from the area of collaborative filtering and page-centric approaches from the area of collaborative browsing are appropriate for collecting community wisdom in the context of social visualization. A benefit of page-centric approaches is that they may not require individual login, it may be sufficient for users to indicate their community as in I-SPY [48], ASSIST [20] or Conference Navigator [19] systems. If the number of visualized objects is not very

large (up to a few thousands) and the groups are reasonably homogeneous, community-level tracking may work quite well. For larger information spaces or to achieve better precision individual profiling and profile matching may be necessary. On the presentation side, the traditional ranked recommendation list used by collaborative filtering system is hardly appropriate for expressing the community wisdom in social visualization. The ideas of history-enriched environments used by social browsing looks much more relevant. I.e., the representation of information items on the visualization can be altered to express past interaction of similar users (users from the same community or with matching profiles).

Knowledge Sea II presented in section 4 provides an example of social visualization, which uses page-centric group-level activity tracking and expresses it by creating a history-enriched environments with color visual cues. However, it is not the most useful example since the visualization itself is not very typical. It is area-based, not item-based (as in the majority of visualization approaches) and the user can't explore this visualization by manipulating its parameters.

A more straightforward example can be provided by Social VIBE [1]. This system was originally developed in conjunction with the Knowledge Sea II and used to visualize documents within a SOM cell. Social VIBE is based on VIBE [42] interactive spatial visualization approach, which uses document content analysis to present a collection of documents in two dimensions, relative to the *points of interest* (POI). By manipulating the location of the POI, a user can explore the collection and locate relevant documents. In Social VIBE the top cell keywords (the cell focus) are used as POI so the visualization helps users can to discover relationships between the focal keywords and the documents located inside the cell. On Fig. 5 the POIs are shown as small orange squares and the documents are displayed using the same social icons as used by Knowledge Sea II social browsing context (the color of the human figure indicates user personal traffic and its background indicates group traffic). The document positions are determined by their similarities to POIs: the closer a document to a POI, the more similar its contents to the POI. For example, we can see that the documents that are displayed on Fig. 5 are more similar to terms like "loop," "statement," and "operator" than the terms "condition" and "expression." When a user drags a POI around the screen, related documents follow the move, according to their similarity to that POI. Therefore, the user can easily understand the related document by observing these movements. If one document moves a greater distance than the other documents do, when a POI is moved, it means that the document is more similar to that POI. Trails of the movements may optionally be displayed, as in Fig. 5. Social VIBE also provides several other ways to manipulate the visualization (see [1]), but they are not essential for the focus of this paper.

8 Conclusions

This paper reviewed social information access, a new stream of research on the crossroads of information access and social computing. It reviewed the origins of this stream, classified social information access technologies according to the supported information access paradigm, and provided examples of three less explored types of

social information access systems. Due to the space limits, some interesting technologies were just briefly mentioned and some were completely left out. The results of empirical evaluation or reviewed technologies were not presented either. In addition, there was no space to discuss two important integrative topics, which go beyond the technology-by-technology structure of the review. The first of these topics is the emergence of social information access systems, which integrate creatively several traditional technologies such as visualization+browsing in Knowledge Sea II [17] or search+browsing in ASSIST [20]. This direction of research is important since integration expands the volume of community information available for the social guidance algorithms. Another important topic is the modern stream of research on encouraging user contributions in social systems. Ensuring a reliable stream of user feedback is critical for social systems and a number of modern approaches to increase the flow of feedback seriously influence the design of social systems. More information about it can be found in [11; 20; 25]. With all these shortcomings, the author hopes, that this review will be useful to those interested in exploring and implementing social information access technologies.

Acknowledgments. This material is partially based upon work supported by the National Science Foundation under Grants No. 0310576 and 0426021. A number of systems presented in this paper were developed in collaboration with Rosta Farzan, who was supported by National Science Foundation Graduate Fellowship.

References

1. Ahn, J.-w., Farzan, R., Brusilovsky, P.: A two-level adaptive visualization for information access to open-corpus educational resources. In: Brusilovsky, P., Dron, J., 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. (2006) 497-505
2. Ahn, J.-w., Farzan, R., Brusilovsky, P.: Social search in the context of social navigation. Journal of the Korean Society for Information Management 23, 2 (2006) 147-165
3. Amitay, E., Darlow, A., Konopnicki, D., Weiss, U.: Queries as anchors: selection by association. In: Proc. of Proceedings of the 16th ACM Conference on Hypertext and Hypermedia. (2005) 193-201
4. Beg, M.M.S., Ravikumar, C.P.: Measuring the quality of web search results. In: Proc. of 6th Int. Conf. on Computer Science and Informatics, A Track at the 6th Joint Conference on Information Sciences (JCIS 2002). (2002) 324-328
5. Bollen, J., Heylighen, F.: A system to restructure hypertext networks into valid user models. The New Review of Multimedia and Hypermedia 4 (1998) 189-213
6. Brin, S., Page, L.: The anatomy of a large-scale hypertextual (Web) search engine. In: Ashman, H., Thistewaite, P. (eds.) Proc. of Seventh International World Wide Web Conference. Vol. 30. Elsevier Science B. V. (1998) 107-117
7. Brusilovsky, P.: Adaptive navigation support. In: Brusilovsky, P., Kobsa, A., 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) 263-290

8. Brusilovsky, P., Chavan, G., Farzan, R.: Social adaptive navigation support for open corpus electronic textbooks. In: De Bra, P., Nejd, W. (eds.) Proc. of Third International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'2004). Lecture Notes in Computer Science, Vol. 3137. Springer-Verlag (2004) 24-33
9. Brusilovsky, P., 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
10. Burke, R.: Hybrid Web recommender systems. In: Brusilovsky, P., Kobsa, A., 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
11. Cheng, R., Vassileva, J.: Design and evaluation of an adaptive incentive mechanism for sustained educational online communities. *User Modelling and User-Adapted Interaction* 16, 2/3 (2006) 321-348
12. Dieberger, A.: Supporting social navigation on the World Wide Web. *International Journal of Human-Computer Interaction* 46 (1997) 805-825
13. Dieberger, A.: Where did all the people go? A collaborative Web space with social navigation information (2000) available online at <http://homepage.mac.com/juggle5/WORK/publications/SwikiWriteup.html>
14. Dieberger, A., Dourish, P., Höök, K., Resnick, P., Wexelblat, A.: Social navigation: Techniques for building more usable systems. *interactions* 7, 6 (2000) 36-45
15. Dieberger, A., Guzdial, M.: CoWeb - experiences with collaborative Web spaces. In: Lueg, C., Fisher, D. (eds.): *From Usenet to CoWebs: Interacting with Social Information Spaces*. Springer-Verlag, New York (2003) 155-166
16. Dourish, P., Chalmers, M.: Running out of space: Models of information navigation. In: Cockton, G., Draper, S.W., Weir, G.R.S. (eds.) *Proc. of HCI'94*. Cambridge University Press (1994)
17. Farzan, R., Brusilovsky, P.: Social navigation support through annotation-based group modeling. In: Ardissono, L., Brna, P., Mitrovic, A. (eds.) *Proc. of 10th International User Modeling Conference*. Lecture Notes in Artificial Intelligence, Vol. 3538. Springer Verlag (2005) 463-472
18. Farzan, R., Brusilovsky, P.: Social navigation support in a course recommendation system. In: Wade, V., Ashman, H., Smyth, B. (eds.) *Proc. of 4th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'2006)*. Lecture Notes in Computer Science, Vol. 4018. Springer Verlag (2006) 91-100
19. Farzan, R., Brusilovsky, P.: Community-based Conference Navigator. In: Dimitrova, V., Tzagarakis, M., Vassileva, J. (eds.) *Proc. of 1st Workshop on Adaptation and Personalisation in Social Systems: Groups, Teams, Communities at the 11th International Conference on User Modeling, UM 2007*. (2007) 30-39
20. Farzan, R., Coyle, M., Freyne, J., Brusilovsky, P., Smyth, B.: ASSIST: adaptive social support for information space traversal. In: *Proc. of 18th conference on Hypertext and hypermedia, HT '07*. ACM Press (2007) 199-208
21. Fitzpatrick, L., Dent, M.: Automatic feedback using past queries: social searching? In: *Proc. of ACM SIGIR '97*. (1997) 306-313
22. Freyne, J., Smyth, B.: Cooperating search communities. In: Wade, V., Ashman, H., Smyth, B. (eds.) *Proc. of 4th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'2006)*. Lecture Notes in Computer Science, Vol. 4018. Springer Verlag (2006) 101-111
23. Goldberg, D., Nichols, Oki, B., Terry, D.: Using collaborative filtering to weave an information tapestry. *Communications of the ACM* 35, 2 (1992) 61-70
24. Hammond, T., Hannay, T., Lund, B., Scott, J.: Social Bookmarking Tools (I): A General Review. *D-Lib Magazine* 11, 4 (2005)

25. Harper, F.M., Li, X., Chen, Y., Konstan, J.: An economic model of user rating in an online recommender system. In: Ardissono, L., Brna, P., Mitrovic, A. (eds.) Proc. of 10th International User Modeling Conference. Lecture Notes in Artificial Intelligence, Vol. 3538. Springer Verlag (2005) 307-316
26. Hill, W., Stead, L., Rosenstein, M., Furnas, G.: Recommending and evaluating choices in a virtual community of use. In: Proc. of SIGCHI conference on Human factors in computing systems, CHI '95. ACM (1995) 194-201
27. Hill, W.C., Hollan, J.D., Wroblewski, D., McCandless, T.: Edit wear and read wear. In: Proc. of SIGCHI Conference on Human Factors in Computing Systems, CHI'92. ACM Press (1992) 3-9
28. Hook, K., Benyon, D., Munro, A.J. (eds.): Designing Information Spaces: The Social Navigation Approach. Springer Berlin (2003)
29. Huang, C.K., Chien, L.F., Oyang, Y.J.: Relevant term suggestion in interactive web search based on contextual information in query session logs. Journal of the American Society for Information Science and Technology 54, 7 (2003) 638-649
30. Jung, S., Harris, K., Webster, J., Herlocker, J.L.: SERF: Integrating human recommendations with search. In: Proc. of ACM 13th Conference on Information and Knowledge Management, CIKM 2004. (2004) 571-580
31. Kantor, P.B.: Historical Note on the archaeology and the history of recommender systems (2007) available online at http://www.scils.rutgers.edu/~kantor/AntWorld/OCLC_ANTWORLD_HistoricalNote.doc
32. Kantor, P.B., Boros, E., Melamed, B., Meňkov, V., Shapira, B., Neu, D.J.: Capturing human intelligence in the net. Communications of the ACM 43, 8 (2000) 112-116
33. Karamuftuoglu, M.: Collaborative information retrieval: toward a social informatics view of IR interaction. Journal of the American Society for Information Science 49, 12 (1998) 1070-1080
34. Keller, R.M., Wolfe, S.R., Chen, J.R., Rabinowitz, J.L., Mathe, N.: A bookmarking service for organizing and sharing URLs. In: Proc. of Sixth International World Wide Web Conference. Computer Networks and ISDN Systems, Vol. 29, 8-13 (1997) 1103 - 1114
35. Kohonen, T.: Self-organizing maps. Springer Verlag, Berlin (1995)
36. Konstan, J., Riedl, J.: Collaborative filtering: Supporting social navigation in large, crowded infospace. In: Höök, K., Benyon, D., Munro, A.J. (eds.): Designing Information Spaces: The Social Navigation Approach. Springer, Berlin (2003) 43-82
37. Lehtikoinen, J., Salminen, I., Aaltonen, A., Huuskonen, P., Kaario, J.: Meta-searches in peer-to-peer networks. Personal and Ubiquitous Computing 10, 6 (2006) 357-367
38. Li, W.-S., Vu, Q., Agrawal, D., Hara, Y., Takano, H.: PowerBookmarks: a system for personalizable Web information organization, sharing, and management. In: Proc. of 8th International World Wide Web Conference. Elsevier (1999) 297-311
39. Maltz, D., Ehrlich, K.: Pointing the way: active collaborative filtering. In: Proc. of CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM (1995) 202-209
40. Micarelli, A., Gasparetti, F., Sciarrone, F., Gauch, S.: Personalized search on the World Wide Web. In: Brusilovsky, P., Kobsa, A., 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) 195-230
41. Munro, A.J., Hook, K., Benyon, D. (eds.): Social Navigation of Information Space. Springer Berlin (1999)
42. Olsen, K.A., Korfhage, R.R., Sochats, K.M., Spring, M.B., Williams, J.G.: Visualisation of a document collection: The VIBE system. Information Processing and Management 29, 1 (1993)

43. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.: GroupLens: An open architecture for collaborative filtering of netnews. In: Proc. of ACM 1994 Conference on Computer Supported Cooperative Work. ACM Press (1994) 175-186
44. Rucker, J., Polano, M.J.: Sitemeer: Personalized navigation for the Web. Communications of the ACM 40, 3 (1997) 73-75
45. Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S.: Collaborative filtering recommender systems. In: Brusilovsky, P., Kobsa, A., 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) 291-324
46. Scholer, F., Williams, H.E.: Query association for effective retrieval. In: Proc. of ACM 11th Conference on Information and Knowledge Management, CIKM 2002. (2002) 324-331
47. Shardanand, U., Maes, P.: Social information filtering: Algorithms for automating "word of mouth". In: Katz, I., Mack, R., Marks, L. (eds.) Proc. of CHI95. ACM (1995) 210-217
48. Smyth, B., Balfe, E., Freyne, J., Briggs, P., Coyle, M., 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
49. Svensson, M., Höök, K., Cöster, R.: Designing and evaluating kalas: A social navigation system for food recipes. ACM Transactions on Computer-Human Interaction 12, 3 (2005) 374-400
50. Wexelblat, A., Mayes, P.: Footprints: History-rich tools for information foraging. In: Proc. of ACM Conference on Human-Computer Interaction (CHI'99). (1999) 270-277