Collecting Community Wisdom: Integrating Social Search & Social Navigation

Jill Freyne
University College Dublin
School of Computer Science & Informatics
Belfield, Dublin 4, Ireland
jill.freyne@ucd.ie

Rosta Farzan
University of Pittsburgh
School on Information Science
Pittsburgh, PA, 15260,USA
rosta@cs.pitt.edu

Peter Brusilovsky
University of Pittsburgh
School on Information Science
Pittsburgh, PA, 15260, USA
peterb@pitt.edu

Barry Smyth
University College Dublin
School of Computer Science & Informatics
Belfield, Dublin 4, Ireland
barry.smyth@ucd.ie

Maurice Coyle
University College Dublin
School of Computer Science & Informatics
Belfield, Dublin 4, Ireland
maurice.coyle@ucd.ie

ABSTRACT

The goal of this paper is to detail the integration of two "social Web" technologies – social search and social navigation – and to highlight the benefits of such integration on two levels. Firstly, both technologies harvest and harness "community wisdom" and in an integrated system each of the search and navigation components can benefit from the additional community wisdom gathered by the other when assisting users to locate relevant information. Secondly, by integrating search and browsing we facilitate the development of a unique interface that effectively blends search and browsing functionality as part of a seamless social information access service. This service allows users to effectively combine their search and browsing behaviors. In this paper we will argue that this integration provides significantly more than the simple sum of the parts.

ACM Classification: H.3.1 [Content Analysis and Indexing]: Indexing method; H.3.7 [Digital Libraries]: User issues; H.5.2 [User Interfaces]: Graphical user interfaces (GUI); H.5.4 [Hypertext/Hypermedia]: Navigation.

General terms: Design, Human Factors.

Keywords: Social Navigation, Social Search, Community-Based Adaptation, Hypermedia, User Interfaces.

INTRODUCTION

With information growing at an exponential pace the information access tools that have served us well in the past are now creaking under the weight of the Web. Navigating through the ever-changing information multiverse is becoming increasingly difficult and even the latest search engine technologies are struggling to cope with our limited ability to declare our information needs. Recent research efforts have highlighted the interactive nature of information access behavior and promoted the potential value of harnessing user activity patterns to drive the next generation of social information access tools. For example, researchers have demonstrated how social navigation techniques can leverage the browsing behavior of past users (pages accessed, time spent reading, annotations left on pages) to guide others to interesting and relevant information; see [12, 23]. At the same time complementary research on social search (e.g. [21]) has highlighted how the search patterns (queries and selections) of users and communities of users can be used when responding to future searches in order to adapt a result-list to the needs and preferences of a particular community. These research efforts are significant because they demonstrate how the actions and behaviors of users can enrich content collections and thereby aid information access. However, we believe that these various approaches suffer from a lack of integration: each approach encapsulates within its own information access paradigms; each accumulates its own collection of community wisdom which is never shared with other approaches. This is clearly flawed. Why shouldn't the browsing patterns and annotations of a community of users enrich their search results, and vice versa?

For example, consider a student working on a term-paper on social navigation, as part of a class on advanced information access. The class forms a community of interest, and as the students browse and search for information relevant to their term-papers their activities, queries, comments, tags and annotations all combine to express the collective wisdom of the class, a wisdom that can be captured and exposed to other users. So when a student searches for a paper on "social navigation" they might see, among other results, a link to

IUI'07, January 28–31, 2007, Honolulu, Hawaii, USA.

Wexelblat and Maes' original paper on Footprints. While the student may not know that this paper is important, the community wisdom can guide the student to this result with some indication that it has been viewed before and perhaps annotated by a significant percentage of users within this community. Perhaps some of these annotations will bridge the gap between the paper and class context by remarking on the critical role that this early work has had on social navigation systems and history-based browsing. Similarly, when another user is browsing their class Web and notices a link to this paper, they might appreciate a visual cue indicating that other classmates have searched for this work with queries such as "social navigation". Replaying one of these queries will lead the user to another set of relevant papers, some of which will also have been selected and annotated by other classmates. The point is that all of this interaction information, whether explicit annotations or implicit navigation trails and search histories, can be combined to add a rich layer of social knowledge to facilitate all kinds of future information access opportunities.

In this paper we explain how we have combined social search and social navigation to take full advantage of the synergies that these technologies have to offer. We will describe how the resulting integration has been used to provide a community-based access to the ACM Digital Library, allowing community members to benefit from a wide range of fully integrated social search and navigation features. Importantly, we will stress how this integration represents an important conceptual step forward by bridging the domains of navigation and search and allowing users to move more fluidly and naturally between information sources via a richer form of linkages.

COMMUNITY-BASED INFORMATION ACCESS

Over the last decade a range of social information access technologies has attracted attention of both researchers and practitioners as they strive to help users find their way in the rapidly expanding information space. Social information access technologies capitalize on the natural tendency of people to follow direct and indirect cues of activities of others, e.g. going to a restaurant that seems to attract many customers, or asking others what movies to watch. The pioneering work on social information access in the early 90's attempted to formalize this social phenomenon in two main forms: collaborative filtering and history-enriched environments. Collaborative filtering attempted to propagate information items between users with similar interests. This technology enabled a social form of information filtering and recommendation. For example, the pioneering collaborative filtering system GroupLens [20] allowed cross-recommendation of netnews articles. History-enriched environments attempted to make the aggregated or individual actions of other community members visible. This technology enabled social navigation through an information space. For example, the 'Read Wear and Edit Wear' system [15] 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. More recently, the set of social information access technologies was extended with social search and social bookmarking systems. Social search systems such as I-SPY [21] attempted to help new searchers by capitalizing on past successful searches of similar users. Social bookmarking systems such as WebTagger [16] applied tagging to help new users locate useful information already discovered and classified by others. Information access using social tagging systems was recently popularized by such systems as del.icio.us (http://del.icio.us) and Flickr (www.flickr.com).

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. The goal of our work is to integrate different social information access technologies and allow them to share the accumulated community wisdom of different information access paradigms. Our current work focuses on social search and browsing. To provide more information about these technologies, and about the starting point of our work, the remaining part of this section reviews these two social information access technologies and our past work in this regard.

Social Browsing

Social browsing can be considered as an application of the more general idea of social navigation, applied to the specific context of hypertext link browsing. Social navigation in an information space was originally introduced by Dourish and Chalmers as "moving towards cluster of people or selecting objects because others have been examining them"[11]. The classic example of social navigation in a hypertext environment was provided later by Wexelblat and Maes in the Footprints system, which visualizes usage paths in a Web site [23]. The system allows users to leave activity traces in the virtual environment and visualizes these traces to guide future users. The idea of footprint-based navigation was later used in other systems such as CoWeb [9]. In a similar vein, the SearchGuide [6] is a Web search support system that annotates content and links within result pages reflecting the interactions of previous users of the system as they browse away from a result page.

An elaborated example of social browsing is provided by Knowledge Sea II [4]. Knowledge Sea II was developed to help students in the same class discover most useful pages in multiple open corpus textbooks. It supports information access through search, visualization, and browsing and guides users with extended footprint-based and annotation-based social navigation support. Footprint-based navigation support is based on keeping track of visiting behavior of the students. Knowledge Sea II takes into account the time spent reading each page, to capture a more precise insight into the intention of the users and to eliminate unreliable (fleeting) footprints [12]. Social navigation support in Knowledge Sea II is also based on the annotations provided by the students. Like reading, annotating is a natural activity for users, but at the same time provides a stronger source of evidence for the importance of the viewed page, compared with a simple visit. Footprints in Knowledge Sea II are presented by different color densities; darker color represents higher traffic. For example, Figure 1 shows the knowledge map of the Knowledge Sea II system. Resources are organized into cells in a table and the background color of each cell represents the density of visits to resources inside the cell. Resources with students' annotation are augmented with visual cues. A small sticky-note represents the presence of annotations, left by students, and the thermometer icon represents the "temperature" of the annotations. The temperature grows warmer as more students associate positive annotations with the page.

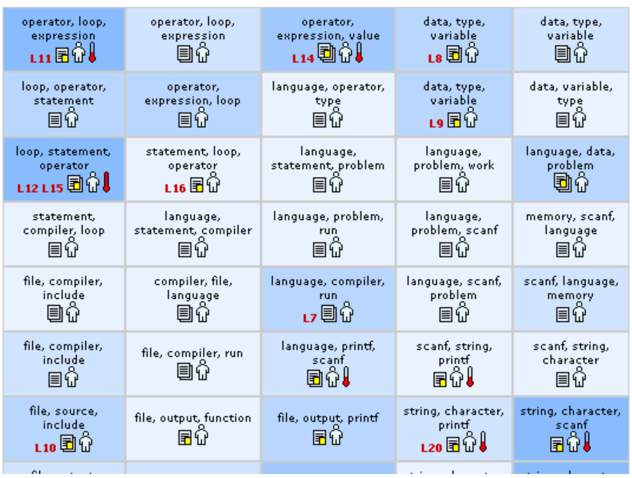

Figure 1: Footprints in KnowledgeSea II.

Social Search

Modern search engines continue to struggle when it comes to delivering the right results to the right users at the right time. The term-based matching engines at the core of modern Web search systems have their origins in early information retrieval research, which focused primarily in supporting the specialized retrieval tasks of information retrieval experts. In contrast the diversity of Web content is a far cry from a specialized document collection and today's Web searchers are not information retrieval experts; see for example, [2, 8, 13, 19]. As a result Web search engines have had to evolve to cope with this new reality. For example, the works of [3] and [17] complement term-based matching with document connectivity information, while others seek to exploit context as a means of supplementing vague queries [18]. Alternatively, other researchers have looked at clustering techniques as a way to impose order on a collection of search results, with a view to identifying different conceptual groupings of results [9, 14].

One core problem that developers have little control over concerns the average searcher's inability to express their search needs effectively in the form of a query. For example, typical search queries are vague and under specified when it comes to revealing a searcher's true information needs. The average search query contains only 2-3 terms [18] and users often use terms in their query that are different from those used in the documents they seek [13]. Just as Google and others have leveraged connectivity information as a new source of knowledge to guide search, others have looked to users themselves for guidance. For instance, in the late 90's the ill-fated DirectHit (www.directhit.com) search engine sought to introduce searcher behaviour into its result-ranking machinery, by allowing the popularity of result selections to influence the ranking of a result for a given query. Unfortunately the approach did not succeed, for a number of reasons: first query repetition in general search tends to be low and so popularity information was sparse; secondly, popularity-based rankings are readily abused by malicious searchers when it comes to promoting their favorite sites. Nevertheless the concept of harnessing search history information (the queries and selections of users) is a powerful one, especially if this usage information can be used in a more targeted fashion. Below we will describe a successful demonstration of just such an approach – called collaborative Web search (CWS) – which relies on the availability of communities of like-minded searchers, recording and reusing their focused search activities to produce result-lists that are better adapted to the needs of the community. Studies have shown, for example, that there is a lot more query repetition within the searches of communities of likeminded users [21]. CWS is an example of social search, which is particularly well-adapted to one of the most powerful drivers of the social Web, namely social networking services. These services (e.g., MySpace, Bebo, Frienster, FaceBook, etc.) provide services for like-minded groups of online users to interact and communicate and provide an ideal platform for social search.

Briefly, CWS is designed to manipulate the results of some underlying search engine, post-processing them so that they better conform to the inferred preferences of a community of users. To do this CWS records the queries and result selections (search interaction histories) of each community of searchers and reuses these patterns to identify results for promotion when faced with a new target query. Specifically, when presented with a new query, in addition to retrieving the appropriate results from the underlying search engine, CWS retrieves any search sessions associated with similar queries and combines and scores the results selected during these sessions. These results are candidates for promotion and the top-ranking candidates are typically promoted ahead of the results returned by the underlying search engine. We will discuss CWS in more detail in a later section.

I-SPY [21] is an implementation of CWS accessible at http://ispy.ucd.ie. Very briefly, I-SPY facilitates the creation of online search communities and relies on a variety of underlying search engines for its core results. Figure 2 shows an example result-list for the query 'sun' for a particular community of computer science students. The top promotions are for sites that relate to community interests (Java Sun Microsystems sites), both of which have been previous selected for a range of similar past queries; these related queries are also shown alongside each promoted result and can themselves be used to initiate new searches. In contrast, the top matching result from the underlying search engines is for the UK Sun newspaper. In this way search becomes a more social activity with the shared search experiences of the community helping to deliver more relevant results and offer new opportunities for further search.

Figure 2: I-SPY result page

INTEGRATED SOCIAL INFORMATION ACCESS

The goal of the work presented in this paper is to achieve a true integration of social search and social browsing. Starting with a social search component based on I-SPY [21], and the social navigation component, AnnotatEd, used in Knowledge Sea II [4], we designed a community-based information access system where these two technologies could reinforce each other. Our "true integration" attempt was based on two critical ideas. First, the "community wisdom" collected by the search component should also be used to support usage of the navigation component, and vice versa. Second, social search and social browsing should be seamlessly integrated at the interface level: the user should be able to move from search to browsing without losing the context of work and community-based support, and such movement should be actively enhanced and

To evaluate our ideas, we developed a system that incorporates the framework for community-based search result personalization introduced in [6] and which facilitates community-based access to the Communications of ACM (CACM) magazine. CACM is well known for its broad scope and provides a good example of a collection of documents accessed by multiple communities. Our system is designed to aid users in their quest to locate relevant research articles in CACM, by harnessing the collective wisdom of the community. This section explains how community-based access to CACM works, from a user perspective, stressing the integrative aspects of the system. To do this we refer back to the example in the introduction, where students in an advanced information access class are busy completing a term-paper on social navigation.

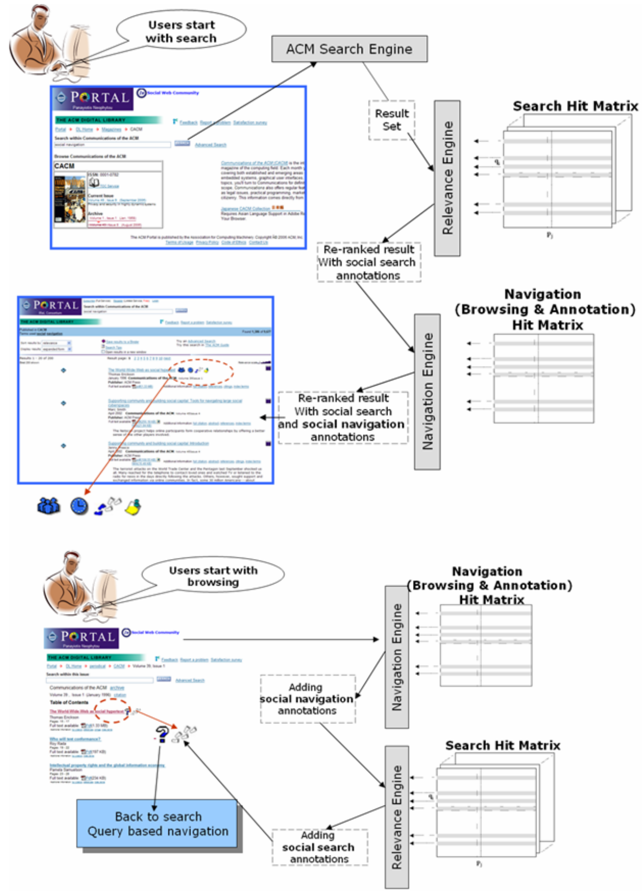

Figure 3: Integrated community-based information access in the search and browsing contexts.

From a user perspective, the system looks like an augmented version of the ACM Digital Library (which was used as the back-end of our system). As standard, the ACM Digital Library Web site (http://portal.acm.org) provides users with two information access strategies to uncover information in its corpus – searching and browsing. The example below demonstrates how the augmented system helps students through each of these processes. The overall view of the system and the two information access scenarios are shown in Figure 3.

Let's start with a search scenario (top of Figure 3). Upon entering the query "social navigation" to the familiar ACM Digital Library search box the users are presented with a result page as seen in Figure 4. Our social information access system alters the results returned by the standard ACM search facility in two ways. Firstly, I-SPY's search technology re-ranks the result-list according to the preferences of the community, i.e. promoting CACM articles previously selected by classmates. Secondly, we utilize and extend a technique described in [6] and [7], whereby supplementary search and navigational access information is appended to any results in the results-list which the system has encountered previously. This information is added to the results in the form of easily interpreted visual indicators. In total, five icons are used, each denoting a different type of interaction information. Each icon appears with varying levels of filling to denote the level of interaction history recorded.

Figure 4: Search Result Page

For example, the result that appears in position one of the result-list shown in Figure 4 is augmented with all 5 types of icon. The community icon (first icon from left) shows the relevance of the result for the current query relative to the community in which it was entered. Its relevance value denotes the percentage of times this result has been selected by community users using the query (see Equation 3). In Figure 4, a Wexelblat and Maes article "Footprints: history-rich tools for information foraging" has a relevance of 100% to the query "social navigation", thus the result has been selected every time this query has been entered. The related icon (second icon from left) informs the user of any related queries for this result; that is, other queries that have led to the selection of this result by the current community. A simple mouse-over reveals a list of queries; this provides the student with suggestions of other queries that may be relevant to their search task and also informs the user of other topics that the article may also be related to. For example, in Figure 4 the related queries for the article are "social navigation" and "social interaction". An examination of the title and snippet of the result reveals that the word "social" is not mentioned at this stage, so the result might otherwise have been missed. However, from the queries used to find the article it is clear that the Footprints system is a social system. The color of the icon represents the average relevance of the article in the results list across all related queries.

The recency icon (third from left in Figure 4) provides users with information relating to the last time the result was encountered by users through searching or browsing. This allows users to form a view of the freshness of the interaction trail. Results that have not been accessed in a number of months may not be as useful as results that have been accessed more recently. The footprints icon (fourth from left in Figure 4) explicitly connects the social search system to social navigation system by indicating the browsing popularity of the article behind the link. The fill-level of this icon indicates the relative popularity of the article in question. For example, the most visited article by a community is 100% filled.

The annotation icon (fifth from left in Figure 4) informs the user of the presence of annotations by community members. Annotated articles contain valuable additional user-provided information, which the authors of the annotations deemed relevant to members of the community. Once again a mouse-over of this icon shows annotation popularity.

Figure 5: Social navigation support for navigating

Figure 6: - Article augmented with related community activities

The browsing task (bottom part of Figure 3) is also assisted by social visual cues (icons), as can be seen in Figure 5. Each link to a CACM article encountered during browsing is augmented with social information. It includes table-of-contents pages, as shown in Figure 5, as well as article information pages with cross-citation links to other articles. Again the annotation icon (third icon from left) denotes the presence of annotations. The browsing popularity of the link is represented by the footprints icon. In addition, we complement the browsing interface with search information in the form of a search related icon; another explicit connection between the search and browsing components. For example, the query icon (the question-mark icon, first on left in Figure 5) indicates that the article has been selected as relevant to at least one query issued by the community users. A mouse-over the icon reveals a list of such queries.

Choosing an article, from either a search result-list or through a link on one of the browsed pages, brings the user to the article's information page (see Figure 6). In addition to the standard article information provided by the CACM (authors, abstract, source etc.) the users are presented with additional social information pertaining to article's search history. Integrated into the information page is a panel displaying the list of queries that have been used to locate the article. Each query is augmented with icons that represent the recency of the last search and each query's relevance to the article. This unique feature had several motivations. Firstly, listing past queries provides an insight to the article content, as mentioned in a previous paragraph. The second and more important motivation is to provide another connection between browsing and search knowledge and functions. By selecting a query in the list, for example, a search can be instantly initiated. Since this query was originally used to locate the viewed article, the results of the new search are also likely to be related to the viewed article. In effect selecting the query is like asking the system to find "similar" articles, or articles that are "more like this" without the user having to browse further, using the related query as an additional contextual cue. Returning to our example, selecting the query "social interaction" returns articles on social browsing, social gaming, social navigation and interaction designs, all of which have similarities to the footprints social interaction system, which the social interaction query listed among its results.

The last context where our system augments the ACM Digital Library interface is the full content view of the article in HTML format. This view can be reached directly from the search results and table-of-contents pages as well as from the article information page. While reading the article, users can contribute to yet another information access attribute, annotation. By performing searches and navigating, the users contribute implicit feedback to the "community wisdom". Annotations, however, require explicit feedback from users in the form of textual comments. In our system, the usual article view page is augmented by a side panel (see Figure 7), which contains a full list of previous annotations and a form allowing users to add new annotations. Annotations can be in the form of page highlighting or note writing. The notes (textual annotations) can be categorized into positive or general annotations. The user can choose whether to make their annotations visible to just their community members or to the general public. Finally, the users also have the option to sign the note with their name, or keep it anonymous. Thus, users can add any comments they like relating to this article and its relevance to their needs and tasks, and the community. The shared users' annotation is used to provide navigation support for the community as presented above.

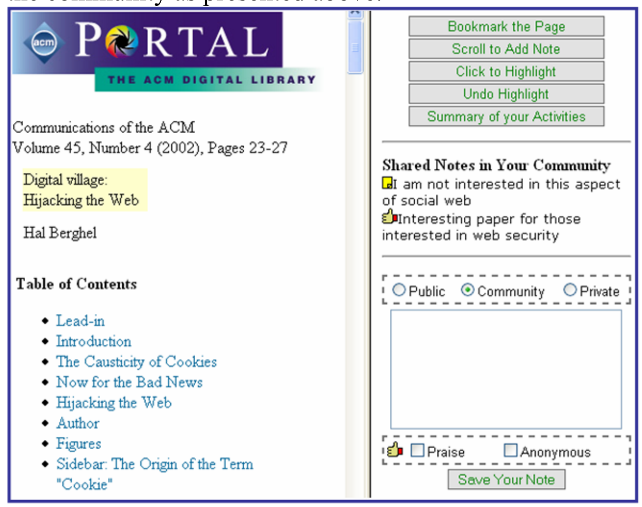

Figure 7: Annotation interface

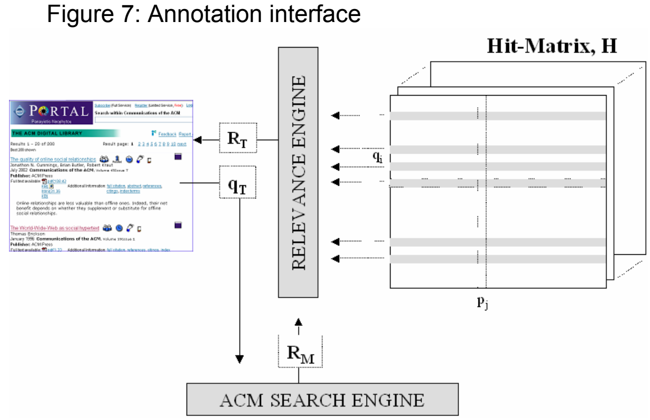

Figure 8: I-SPY system architecture

THE MECHANISMS OF COMMUNITY-BASED INFORMATION ACCESS

In this section we describe in details how the abovementioned system works at a technical level.

The Search Component

Very briefly, the proxy-based architecture of the search component of our system intercepts ACM Library search requests, sends a parallel request to the CWS module and integrates the CWS and ACM result lists as described below (see [7] for more details on this aspect of the system). The key novelty of the search component of our system stems from how the result-list (R_M) returned from the regular ACM search engine is processed to produce a new, personalized result-list that reflects the learned preferences of a community of like-minded searchers, according to the CWS approach reviewed above.

Critical in this is a data structure known as the hit-matrix, H, which is the repository for community search experience; see Figure 8. Briefly, each community is associated with its own hit-matrix and each hit-matrix relates article selections to past queries for that community. Basically, each element of the hit matrix, H_{ij}, stores a value v (to indicate that page p_j has been selected for query, q_i in v search sessions). In other words, each time a user selects a page, p_j, for a query term, q_i, their community hit matrix is updated.

Collaborative search exploits the hit matrix as a direct source of relevancy information. Thus at retrieval time, for each target query q_T we retrieve a set of similar search queries to serve as a source of relevant results. Query similarity can be measured using a simple term-overlap metric (Equation 1); alternative similarity metrics have been evaluated [1] but for now we will use this simple measure. During the retrieval stage, this allows collaborative search to rank-order queries according to their similarity to the target query so that all, or a subset of, these similar queries might be reused during result ranking.

$$Sim(q_T, q_i) = \frac{|q_T \cap q_i|}{|q_T \cup q_i|} \quad (1)$$

The results associated with these similar queries are then scored to evaluate their relevance to q_T. Basically, the relevance of a page p_j to some query q_i is estimated as the relative number of selections that this page has attracted for qi (see Equation 2). Then, to compute the relevance of p_j for a target query q_T, across some set of similar queries $q_1, \ldots, q_n$, we calculate an overall relevance score as the weighted sum of the relevance of p_j for each of the $q_1, \ldots, q_n$, as shown in Equation 3. Each Relevance(p_j ,q_i) is weighted by Sim(q_i, q_T) to discount the relevance of results from less similar queries; Exists(p_j, q_i)=1 if H_{ij}>0 and 0 otherwise.

$$Relevance = (p_j, q_i) = \frac{H_{ij}}{\sum_j H_{ij}} \quad (2)$$

$$WRelevance(p_j, q_T, q_1, ,,, q_n) = \frac{\sum_{i=1...n} Relevance(p_j, q_i) \bullet Sim(q_T, q_i)}{\sum_{i=1...n} Exists(p_j, q_i) \bullet Sim(q_T, q_i)} \quad (3)$$

Finally, the k top ranking results according to Equation 3 are promoted ahead of the other results returned by the ACM search engine, producing the final result-list R_T that is returned to the user.

The Browsing Component

The browsing component – called AnnotatEd – is the main component behind the social navigation engine. AnnotatEd takes care of tracking all of the browsing activities of the system's users and generates visual social navigation cues as illustrated in Figure 9. As shown in Figure 9, when a page is accessed, its navigation information is updated by AnnotatEd. Since we are not providing annotation facility for articles in PDF format and on the other hand clicking on PDF format of article represent stronger interest in the article, we give twice more weight to clicks on PDF version of the articles. AnnotatEd then extracts all links contained in the page. Two categories of link exist, those within the CACM domain and those outside it. Outside links remain unchanged, links within the domain are augmented with community information and redirected through AnnotatEd. Each qualifying link will potentially be augmented with three types of icons representing browsing popularity, annotation popularity, and search popularity. Browsing popularity is represented by footprints icon using six levels of filling to present different level of popularity. The exact popularity number is shown on mouse over the icon.
Annotation popularity is defined based on the absolute number of annotations an article has received. We see articles that have any annotations as being potentially more important than those that do not have annotations, independent of their percentage of total annotations. Three levels of annotation popularity are considered as shown in Table 1. Annotation popularity is represented by a sticky note icon. Three levels of filling are used for the icon to present three different levels of popularity as shown in Table 1. The exact popularity figure is shown on mouse over the icon.

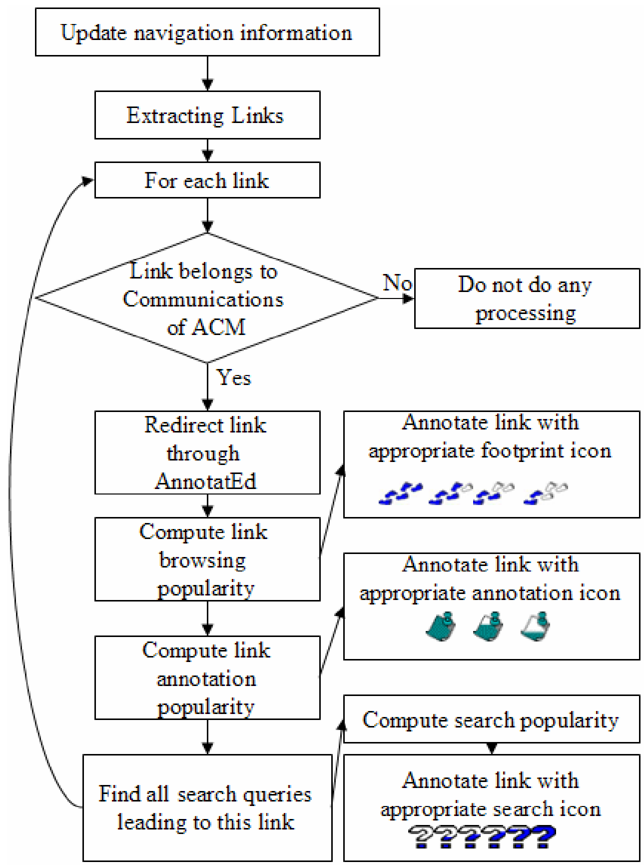

Figure 9: User tracking and link annotation performed by the AnnotatED browsing component.

Table 1: Annotation popularity

Number of annotations	Popularity Level	Icon
1	1	
1< <5	2	
>=5	3	

Similarly, the fill level of the question mark icon represents search popularity which is the percentage of the searches inside the community that has been directed to this link.
An important part of navigation support offered by the browsing component is query-based navigation. When a user chooses to view an article in HTML, the paper information page is shown. One small but vital change to the standard page is made by AnnotatEd. As shown in Figure 9, for each article, the search queries that have lead to that article being selected in the past are displayed and the user can navigate back to the search space by clicking on the link. The list of queries is augmented with search icons as described earlier and shown in Figure 6.

PRELIMINARY EVALUATION

We conducted a preliminary evaluation of the system to assess the general idea of social support for information access and integration of social search and social navigation. Ten PhD students at the School of Computer Science and Informatics in University College Dublin served as subjects in our study. The subjects were randomly and evenly divided into control and experimental groups. They were asked to find CACM articles on the topic of "the social Web". Each subject was asked to use the system for 20 minutes to find as many papers as possible, and then submit the titles of the papers for review.

For the purpose of the study we prepared a control version of our system in which all social visual cues were turned off but all user interactions were tracked. The subjects in the control group were asked to perform the task using the control version of the system. The information collected from the control group was then used to populate the "community wisdom" database for the test system. In addition two of the authors used the system for the same task, i.e. to find articles relevant to "the social Web". This was done to alleviate the cold-start problem. The subjects of the experimental group then performed the task using the full community-based support interface showing the interactions of the control group and the authors. The subjects of the experimental group were asked to answer a short questionnaire providing their subjective feedback on the system. The questionnaire focused on the social cues provided by the system. A subset of the questions and possible answers that are most critical in the context of this paper are provided in Figure 10.

All subjects indicated that they noticed the social cues. Answering the question whether they have found themselves being drawn to articles which were augmented by these cues, one subject mentioned that he was "always" drawn to the articles augmented with social cues and the four other subjects answered they were "often" drawn to them. A set of questions was designed to determine which cues were the most useful when searching and browsing for the relevant papers. The possible answers ranged from always to never. To compare the usefulness of different cues we assigned numerical importance rating for each answer ranging from 4 for "always" to 0 for "never" and averaged the answers for each question. The results are shown in Figures 11 & 12

Figure 10 - Sample questions from questionnaire

The footprint icon, showing the browsing popularity of an article in the community, was found to be most useful during search sessions. Interestingly this icon was generated using the "community wisdom" accumulated by the browsing component. In turn, the most useful community-based navigation support during browsing was provided by the question mark icon, indicating the presence of relevant community queries and constructed from the "community wisdom" assembled by the search component. Thus, in both contexts we find that the social information collected from the complementary context is considered to be more useful than the 'local' social information: search wisdom is most useful during browsing and browsing wisdom is most useful during search. This key result speaks to the added value provided by our integration of search and navigation: the sum of the parts is greater than the whole.

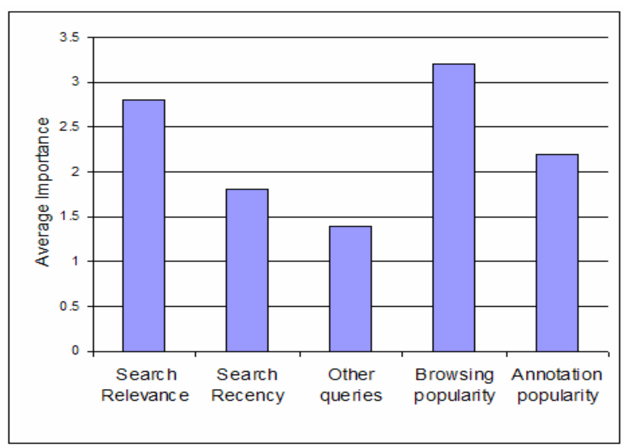

Figure 11: Importance of social cues while searching

At the end of the questionnaire, the subjects were asked to provide their overall opinion about the usefulness of social annotation in information retrieval tasks in general and in the social ACM system. All the subjects responded that the social annotations are useful in general and that the presence of social icons in this task made it easier to locate information in ACM digital library.

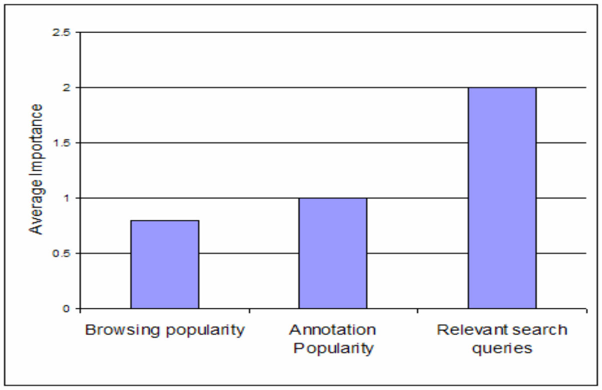

Figure 12: Importance of social cues when browsing

DISCUSSION

The work presented in this paper can be considered as proof-of-concept evidence to support the notion that different community-based information access technologies can be used in an integrated system to reinforce each other and provide unique added-value to the user. More specifically, we demonstrated that social search and social browsing could be integrated on both the interface level and the level of internal mechanisms.

On the level of the internal mechanisms, the integration was achieved by taking into account the "community wisdom" collected by both browsing and search component when supporting the user work with either of the components. We have detailed how the information collected by the search component can be used to help user in the browsing process and, vice versa, the information collected by the browsing component can be used to provide an additional support during search. Moreover, as indicated by the user study data, the information collected by one component was considered to be even more useful in the context of the other.

On the interface level the integration was achieved by a seamless connection from search to browsing and from browsing to search, which allowed the user to use both ways of information access, in parallel, without loosing the context of work and social navigation support. That is, after selecting a specific article by following a link in the list of search results, the users could continue browsing for similar articles by following socially annotated forward and backward citation links provided by the ACM Digital Library. Among other information, the social cues informed the users about the relevance of all articles discovered on the way to similar past searches. In turn, the visible list of queries that was used in the past, to search for a specific page, which was encountered by users during browsing was used to highlight useful queries that could be immediately executed by simply clicking on a query link.

This two-way connection between search and browsing opens a whole new way of navigating the collection of articles provided by the system, which we call query-based navigation. Instead of relying solely on the links between articles provided by the original collection, a user can click on a specific query that was used to find this article in the past and that stresses a specific aspect of the article that the user is mostly interested in. This will re-run the search and presents the user with a whole list of articles that are similar to the original article in this specific aspect. From this list (possibly with further help of community social cues) the user can navigate to a useful article that may have a lot in common with the original article yet is not connected to it in the hyperspace. The query-based navigation between these two articles is possible due to the "community wisdom" collected by the system in the form of past queries and navigation history: the linkages that are implicitly formed are social linkages, which complement the explicit navigation linkages provided by the content or site author. The fact that the question icon (indicating the presence of relevant queries) was considered most useful during navigation, is a powerful endorsement of the importance of query-based navigation to the users.

CONCLUSION & FUTURE WORK

The Social Web has highlighted the willingness of everyday users to play the role of content producers as they take advantage of features that allow them to add their mark on Web content. Social bookmarking, reviews, and content annotations are all examples of explicit types of user generated social content. In addition there is a wealth of implicit social content waiting to be mined, including the search and browsing history of users. While others, in the past have sought to take advantage of different forms of explicit and implicit social information, the focus has tended to be limited and narrow, often relating to one particular information access paradigm, such as browsing or search.

In this paper we have taken a significant step forward, by describing a tightly-coupled integration of social navigation and search. The result is an information access system that captures a broad range of user activities and uses this community wisdom to annotate content in a way that facilitates new styles of interaction. For example, users can seamlessly initiate new searchers, based on social cues, as they browse and navigate through information. Similarly, focused browsing can be initiated directly from the search interface. In this way we have effectively bridged the browsing-search gap in a way that facilitates improved information access. The social information helps users to better understand the relevance of annotated information from the perspective of their own community's interests.

We have presented the results of a preliminary evaluation, which speak to the value of the proposed integration. These results indicate that users found social cues appealing and relevant and that their use of such cues helped them in their quest for relevant content. Future work will seek to further evaluate the system as well as exploit new opportunities for additional community-based personalization and adaptation of content.

ACKNOWLEDGMENTS

The authors thank Science Foundation Ireland for making this collaborative work possible through Grant No. 03/IN.3/I361, and through the E.T.S. Walton Award to the third author. This research is partially supported by the National Science Foundation under Grant No. 0447083, Vodafone Ireland, and the National Science Foundation graduate fellowship to the second author.

REFERENCES

1. Evelyn Balfe and Barry Smyth (2005) Case-Based Collaborative Web Search. In Proceedings of 16th European Conference on Case-Based Reasoning, pages 489–503. Springer-Verlag, 2004.
2. Bollmann-Sdorra P. and Raghavan, V. (1993) On the Delusiveness of Adopting a Common Space for Modeling IR Objects: are Queries Documents?, J. Am. Soc. Inf. Sci., 44 No.10, pp 579-587.
3. Brin S. and Page L. (1998) The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer Networks and ISDN Systems, 30(1–7):107–117, 1998.
4. Brusilovsky, P., Chavan, G., and Farzan, R. (2004) Social adaptive navigation support for open corpus electronic textbooks. In: P. De Bra and W. Nejdl (eds.) Proceedings of Third International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'2004), Eindhoven, the Netherlands, August 23-26, 2004, pp. 24-33
5. comScore Networks (2006) Social Networking Sites Continue to Attract Record Numbers, June 15, 2006, http://www.comscore.com/press/release.asp?press=906
6. Coyle, M. and Smyth, B. (2007) On the Community-based Explanation of Search Results. In Proceedings of the 10th Annual Conference on Intelligent User Interfaces (IUI'07), Hawaii, U.S.A., (in press).
7. Coyle, M. and Smyth, B. (2007) Supporting Intelligent Web Search. In ACM Transactions on Internet Technology Special Issue on Intelligent Techniques for Web Personalization, (in press).
8. Cui, H., et al, (2002) Probabilistic Query Expansion Using Query Logs, Proceedings of the Eleventh International Conference on World Wide Web, Honolulu, Hawaii, USA, pp 325-332.
9. Dell Zhang, Y.D. (2004) Semantic, Hierarchical, Online Clustering of Web Search Results. In Proceedings of the 6th Asia Pacific Web Conference (APWEB)., Hangzhou, China, 2004.
10. Dieberger, A. and Guzdial, M. (2003) CoWeb - experiences with collaborative Web spaces. In: C. Lueg and D. Fisher (eds.): From Usenet to CoWebs: Interacting with Social Information Spaces. New York: Springer-Verlag, pp. 155-166.
11. Dourish, P. and Chalmers, M. (1994). Running out of space: models of information navigation. Proceedings of HCI'94, Glasgow, August 1994.
12. Farzan, R. and Brusilovsky, P. (2005) Social navigation support in E-Learning: What are real footprints. In: S. S. Anand and B. Mobasher (eds.) Proceedings of IJCAI'05 Workshop on Intelligent Techniques for Web Personalization, Edinburgh, U.K., August 1, 2005, pp. 49-56.
13. Furnas, G. W. et al, (1987), The Vocabulary Problem in Human System Communication, Communications of the ACM, Vol. 30 No.11, 964-971.
14. Hamilton, N., (2003) The mechanics of a deep net metasearch engine. In Proceedings of the Twelfth International World Wide Web Conference, Budapest, Hungary, 2003.
15. Hill, W. C., Hollan, J. D., Wroblewski, D., and McCandless, T., (1992) Edit Wear and Read Wear. In: Proceedings of ACM Conference on Human Factors in Computing Systems, CHI'92. ACM Press, New York City, New York. 3-9
16. Keller, R. M., Wolfe, S. R., Chen, J. R., Rabinowitz, J. L., and Mathe, N. (1997) A bookmarking service for organizing and sharing URLs. In: Proceedings of Sixth International World Wide Web Conference, Santa Clara, CA, April, 1997, pp. 1103 - 1114.
17. Kleinberg, Jon M., (1999) Authoritative sources in a hyperlinked environment. Journal of the ACM, 46(5): 604–632, 1999.
18. Lawrence, S. and C. Lee Giles, (1999), Accessibility of Information on the Web, Nature, Vol. 400, No.6740, pp 107-109.
19. Raghavan, V. and Sever, H., (1995), On the Reuse of Past Optimal Queries, Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, United Stated, pp 344-350.
20. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994) GroupLens: An open architecture for collaborative filtering of netnews. In: Proceedings of ACM conference on Computer supported cooperative work, Chapel Hill, North Carolina, ACM Press, pp. 175-186.
21. Smyth, B., Balfe, E., Freyne, J., Briggs, P., Coyle, M., and Boydell, O. (2004). Exploiting Query Repetition and Regularity in an Adaptive Community-Based Web Search Engine. User Modeling and User-Adaptated Interaction, 14(5). 383-423, 2004.
22. Svensson, M., Höök, K. and Cöster, R., (2005). Designing and Evaluating Kalas: a Social Navigation System for Food Recipes. ACM Transactions on Computer-Human Interaction, 12 (3). 374-400.
23. Wexelblat, A. and Maes, P. (1999) Footprints: History-rich tools for information foraging. In ACM Conference on Human-Computer Interaction (CHI'99), Pittsburgh, PA. 270-277 .