

사회적 네비게이션 기반 사회적 검색

Social Search in the Context of Social Navigation

Jae-wook Ahn*, Rosta Farzan†, Peter Brusilovsky*

초 록

웹기반 교육 자료들이 폭발적으로 증가함에 따라 적합한 자료들에 보다 효과적으로 접근할 수 있는 방법이 요구되고 있다. 이러한 새로운 방법들 중의 하나로 사회적 네비게이션(social navigation) 기반의 사회적 검색(social searching)이 정보 검색 분야에서 제시되었는데, 이는 동료 이용자들로부터 제공된 정보를 바탕으로 검색 결과의 향상을 추구하는 기법이다. 본 연구에서는 개인화와 사회적 네비게이션에 근거한 웹 기반 사회적 검색 시스템을 구축하였으며 이용자 연구를 통해 이용자에게 적합하고 필수적인 정보를 제공할 수 있는 방법이라는 것을 검증하려 하였다.

ABSTRACT

The explosive growth of Web-based educational resources requires a new approach for accessing relevant information effectively. Social searching in the context of social navigation is one of several answers to this problem, in the domain of information retrieval. It provides users with not merely a traditional ranked list, but also with visual hints which can guide users to information provided by their colleagues. A personalized and context-dependent social searching system has been implemented on a platform called KnowledgeSea II, an open-corpus Web-based educational support system with multiple access methods. Validity tests were run on a variety of aspects and results have shown that this is an effective way to help users access relevant, essential information.

키워드: 사회적 검색, 사회적 네비게이션, 개인화, 적응 시스템

social search, social navigation, personalization, adaptive systems

* School of Information Sciences, University of Pittsburgh ([jaa38](mailto:jaa38@pitt.edu), [peterb](mailto:peterb@pitt.edu))@pitt.edu

† Intelligent Systems Program, University of Pittsburgh ([rostaf](mailto:rostaf@pitt.edu))@pitt.edu

1. Introduction

Since the appearance of the World Wide Web in the 1990's, its contents have been growing explosively. This phenomenon has increased the need to develop effective tools to help people access the information they seek efficiently. Various tools are available to meet this need based on strategies like browsing, filtering, and searching information. Searching, especially, has become one of the most popular methodologies for helping users to find the information they need. Numerous commercial and experimental Web search systems have been developed, tested, and placed into use. It has been established as one of the major methods for accessing useful information on the Web.

However, most of the Web search services which are being widely used lack one important aspect, *personalization*. They are based on the traditional information retrieval approach, which assumes that the query space on the user side and the document space on the system side are identical (Smyth et al. 2004). The queries which users formulate and the document indices generated by the systems are comprised of terms or keywords. The searching process is described as matching those query terms with document index terms. Yet there are abundant situations

where the user's terms and the author's terms do not coincide. This mismatch is more evident in the Web environment where the materials are being produced exponentially and are authored by millions of users with different backgrounds. Moreover, due to the nature of the Web environment, automatic indexing and collecting prevail, making it even harder to reduce the gap between these two spaces. Also, the average number of query terms formulated by users ranges between two and three words on average (Lawrence & Giles 1998) which is quite different in information retrieval experiment environment where enough query terms are often provided in the form of sentences and make the mismatch more problematic.

Further, many of the current Web search services are based on a traditional information retrieval model which is not personalized and is context-free. They adopt a "one size fits all" approach which gives the same set of search results to different users whenever they use the same queries. Different users have varying backgrounds and levels of knowledge within each domain of interest and thus user needs should be differentiated in terms of these variances. We can expect that a search service which supports context-dependent search results will help various users more effectively.

Social search is an attempt to overcome

these query-document space mismatching, non-personalization, and context-independence problems. It makes use of new features to promote the effectiveness of search results. A group of different users who share the same interests can use similar query terms for the same task and their search experiences can benefit other users. Based on these observations, the social search approach takes advantage of past search histories. When a user enters a query, the system looks up the search history of the group which the user is a member of and thus can provide better search results by re-ranking the results according to the clues extracted from the search history or by providing the user with more guidance, beyond the results given by the baseline term-matching retrieval systems.

Motivated by this goal, we have designed and implemented a social search system with social navigation support. It was constructed within the adaptive social navigation support system KnowledgeSea II (Brusilovsky, Chavan, & Farzan 2004) as a platform for sharing user information. In order to test the validity of our approach, we conducted user surveys and transaction log analysis. This paper describes our social search approach in the context of social annotation in section 2 and discusses the evaluation results in section 3. The last section concludes our discussion and

suggests future plans for research.

2. Related Studies

2.1 Social navigation

Social navigation (Dieberger et al. 2000) research explores methods for organizing users' explicit and implicit feedback in order to support meaningful information navigation. This approach includes two features. The first feature supports a known social phenomenon, that people tend to follow the "footprints" of other people. The second feature is self-organization, which allows social navigation systems to function with little or no manual interference from humans. Two well-known examples are Web auctions or blogs.

Jon Dron and others (2001) introduced CoFind (Collaborative Filter in N Dimensions), which structures and selects learning resources for teachers. This system was inspired by the concept of "stigmergy." Stigmergy was coined by Grasse and refers to systems explored by termites when building mounds (Heylighen 1999). When termites build mounds or ants form trails, they produce these by following their colleagues' traces, collaboratively. The outputs become stronger as time passes and more group members participate. They may also dissipate if a specific purpose disappears or

the members' participation decrease. For example, when food runs out, the trail to the location of the food dissipates.

2.2 Social search

Social search or collaborative search is an approach used to promote the effectiveness of Web searches by relying on past search histories (Freyne & Smyth 2004). Smyth and others implemented and tested a social search engine, I-SPY, which is based on the observation that for specialized topic searches, the number of repetitions of query terms is higher than that found within general topic searches. Therefore, they stored query-document frequency matrices from past search histories of the community of users and re-ranked search results by looking up query-document frequencies. They reported an improvement in search results with this approach.

2.3 KnowledgeSea II

KnowledgeSea II (KSII) is a social adaptive navigation support system for open and closed corpus educational resources. Currently, it supports various domains like C-programming language resources, information retrieval, and human computer interaction (HCI) textbooks. In general, open corpus documents are Web resources and are indexed and maintained automatically. Closed corpus resources are stored in a local machine and the indexing is

done manually by domain experts. The evaluations discussed in this paper were made with the C-programming language resources that incorporate over 25,000 documents, including open corpus online resources (online tutorials and textbooks) and closed corpus resources (lecture notes).

These resources are automatically indexed and maintained by the system and users are allowed to access any document, using various methodologies. The default access method is map-based horizontal navigation (Brusilovsky & Rizzo 2002). The resources are presented within an information map which has been generated with a Self Organizing or SOM technology (Kohonen 1997), where similar documents are grouped together and visually represented in an 8×8 matrix. Each cell of this SOM contains keywords which represents the contents of the documents it incorporates. The positioning of each cell also represents the relevancy of the documents contained in them, thanks to the nature of the SOM. Adjacent cells represent similar contents while cells positioned further away on the map represent dissimilar contents, thus generating related clusters.

Even though the SOM map is the core technology for arranging and providing relevant information to users, it also supports other various methods, including social search. The next section describes

other methods which users can use within KSII to fulfill their information needs.

2.3.1 Information access methods in KnowledgeSea II

KnowledgeSea supports not merely one method for accessing information; it supports two-level visualization, hypertext browsing, social search, and recommendation. The following are detailed descriptions of each method.

(1) Two-level visualization

As mentioned above, KSII supports map-based visualization with SOM technology. SOM provides an efficient means of organizing information access by grouping and visualizing similar resources into cells on the map. However, as the number of documents to visualize increases, each cell becomes too crowded for users to make meaningful choices. Multi-level SOM, introduced by Roussinov and Chen (1998), solves this problem by providing users with homogeneous multi-level maps where a group of cells in the original map can be expanded into a second level map and those maps expand again recursively so that they can create a hierarchy. It embeds another problem, however, in that the map now becomes static, and allowing only passive exploration to users unlike other visualization methods, such as MovieFinder (Ahlberg & Shneiderman 1994) or VIBE

(Visual Information Browsing Environment, Olsen et al. 1993; Olsen & Korfhage 1994).

To avoid this problem, a two-level heterogeneous visualization is supported by the current KSII. The first level is visualized with SOM while the second level is created through the dynamic, relevance-based visualization method used by VIBE. The original VIBE framework was developed by researchers at Mode College and at the School of Information Sciences, University of Pittsburgh and was later re-implemented and extended for Web-based environments. The documents clustered in the SOM cells are positioned and visualized on a 2-dimensional space by their relationships to keywords which represent the contents of the cells. For more detail, refer to Ahn, Farzan, and Brusilovsky 2006.

(2) Hypertext browsing

Typically, C-programming Web tutorials or textbooks from KSII collections are organized as a tree of topics connected by links. KSII preserves these links and encourages learners to use them. Often, students discover relevant pages by using maps or searching and then continue to explore the same tutorial along its own authored links. Alternatively, students can start browsing from the root page of each tutorial (root pages of all tutorials are listed on the front page of the system). While pages of some tutorials are not immediately

accessible through the map, every page in the system is accessible through browsing from the given links.

(3) Social search

Users can search the corpus of KSII for the resources they need by entering query terms in the new social search interface. The search results are provided with conventional ordered lists sorted by similarity values between documents and queries. The search system, however, also incorporates a social search approach unlike other traditional search engines. It displays not only the relevance or similarity information but provides additional social navigation information, such as the traffic and annotation information accumulated from each individual user as well as from the group of users that the individual is a member of. Detailed descriptions will be provided in section 3.

(4) Recommendations

Several tutorial pages are explicitly recommended by the instructor. Links to these pages are added to the list of resources for corresponding lectures in the course portal. It is the easiest way for students to access resources, since the portal is frequently used for a range of educational activities.

2.4 Social navigation in KnowledgeSea II

2.4.1 Traffic-based social navigation support

The concept behind traffic-based social navigation support is based on the visualization of the navigational history of each group of learners. The intensity of the background color of each cell of the map represents the intensity of group traffic (Brusilovsky et al 2004). The map starts with a very light shade of blue for each cell and, as the students progress, the more frequently visited cells become a darker and darker blue, so that students can easily follow the footprints of others by looking for cells with darker backgrounds.

operator, loop, expression L11	operator, loop, expression	operator, expression, value L14
loop, operator, statement	operator, expression, loop	language, operator, type
loop, statement, operator L12 L15	statement, loop, operator L16	language, statement, problem
statement, compiler, loop	language, statement, compiler	language, problem, run

Figure 1 A fragment of the KSII map

Students can also view the history of their own interaction with the system. The color of the “human” icon shown in each cell reflects the number of cell visits by the individual student. The more visits, the darker the color of the icon. In this way, students can compare their interaction with the system with those of their total community members by comparing the

color of the human icon to the background color (Figure 1). This information helps the students to decide which cell to visit next.

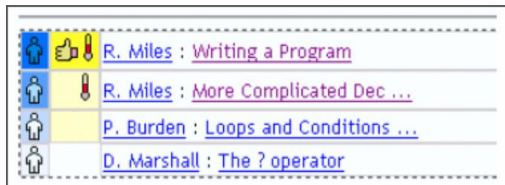


Figure 2 A fragment of a map cell, including the list of resources

When a student clicks on one of the cells of this map, the cell “opens” and shows the list of available resources inside the cell. Typically, behind each of these links is a section from a web-based C-programming tutorial or book. Figure 2 presents a fragment of a map cell. A finer-grained, traffic-based social navigation support is offered for each resource inside each cell. The left hand column of the list of resources shows visual navigation cues that reflect individual and group traffic information. The same visualization approach is repeated here: the background color represents group traffic and the color of the human icon, individual traffic. The same traffic-based, adaptive visual cues are consistently provided next to each resource link in KSII, in any context where this link may appear. The browsing support component of the system dynamically processes all tutorial pages and adds corresponding traffic-based cues to the right of each link previously

visited by the user of the group. All links on the tutorial page are initially created by the author but are then dynamically enhanced with adaptive visual cues at runtime. Finally, to keep the interface consistent, the KSII search engine generates the same traffic-based visual cues for each link when presenting the results of each search (Figure 3).

2.4.2 Annotation-based social navigation support

Annotation-based social navigation support was pioneered in the last version of KSII. Based on student feedback collected during the evaluation of the second version of the system, we attempted to provide a more reliable form of social navigation support that would supplement simple traffic information. Annotation-based support harnesses the power of page highlighting and annotations made by students. An important feature of the KnowledgeSea system is the ability to annotate tutorial pages. While reading each tutorial page, students are able to write note and highlight different sections of that page. In addition to the personal benefits of annotating while reading, the presence of annotation creates an obvious trail for other learners to follow. AnnotatED, the annotation component of KnowledgeSea, provides a rich set of features in support of this annotation activity. Students can

specify the usefulness of a marked page by associating a positive or general type to their notes. Also they can specify whether or not they want to share the notes with others and also if they want to write the note anonymously or signed. All this information assists the author of the annotation and other learners when they later consider accessing the annotated material.

As can be seen in Figure 2 and Figure 3, some of the links are enhanced with more than traffic information. The additional icons are designed to support learners with more emphatic information, based on annotations provided by the individual or the group of students. The “thumbs up” icon shows that a positive comment about this page was made by the current student. If the comment is neutral note, the icon will be a sticky note. In this way, the student can easily locate previously annotated pages. In particular, it helps the student to navigate back to pages that were previously found useful. In addition, some of the links have been annotated with a thermometer icon. This icon represents the overall “annotation temperature.” If the thermometer shows a positive temperature, the page has a larger balance of positive annotations. Therefore, the thermometer icon assists students in deciding which pages to visit by recommending pages with positive temperatures. Similar to traffic-based

visual cues, annotation-based cues also apply the color-intensity scheme to express information about the density of group/user annotations. The intensity of the yellow background color represents the quantity of group annotations and the intensity of the foreground icon color (thumbs up or sticky note) represents the quantity of individual annotations.

3. System Design and Implementation

Current social search functionality is closely coupled with the social navigational support of KSII. Since KSII has been adopted as a platform for this social navigational support, the traffic and annotation information is also shareable with the social search and other basic information access methods of KSII, such as SOM map and hypertext browsing. When a user enters a query, the social search system returns documents which match the query with high similarity values and provides an ordered list of documents just as in other ad-hoc retrieval systems. In addition to the traditional form of search results, the social search model adds in social navigation information based on user traffic and annotation.

3.1 Ad-hoc retrieval model

3.1.1 Document processing

The documents in the target corpus are indexed by the system and made searchable. The basic search part follows the popular vector space model (Salton 1989). In this model, the whole collection of documents is represented as a matrix where each document is understood as a row while the columns represents each of the terms that appear in the corpus. The component values of each vector (document) are weights or importance of corresponding terms. The TF-IDF weighting scheme (Equation 1) was used for implementing the search service for KSII.

$$tfidf(d, t) = tf(d, t) \cdot \log\left(\frac{|D|}{df(t)}\right) \quad (1)$$

where

$tf(d, t)$: number of occurrences of term t in document d

$df(t)$: number of documents where term t appears

$|D|$: total number of documents in the corpus

This formula uses TF (Term Frequency) times IDF (Inverse Document Frequency), where TF means the number of occurrences of a term in one document, and IDF is an inverse of document frequency—the number of documents which include the term in question. IDF can decrease the

effect of very general terms, which appear in a lot of documents and thus have greater DF values but less importance.

The length of each document vector is identical to the number of terms appearing in the corpus. If a term occurs in a document, the corresponding TF-IDF weight is filled into the vector component and if not, 0 is applied to that column. Therefore, the resulting matrix becomes very sparse with most of the components are filled in with 0's.

The procedure to implement this document-term matrix is shown below.

- (1) *Collect document contents* - the documents in the open or closed corpus are collected and their contents are extracted for indexing.
- (2) *Extract meaningful tokens from documents* - from the documents collected in the previous step, meaningful tokens are extracted and incorporated to the index. In this step, a stopword list is also applied, so that very common terms or terms which carry little information, such as articles and prepositions, are excluded.
- (3) *Extract term stems* - because the same terms can be expressed in different forms (e.g. the plural form of a noun), only their stems are extracted, to be used for indexing. In this study, Porter's stemming algorithm (Porter 1980) was used for stem extraction.

(4) *Count TF and IDF* - TF and IDF values for the index terms are calculated, as in Equation 1, and stored in the database.

3.1.2 Matching

In the vector space retrieval model, user queries are represented as term vectors, similar to document vectors. The terms in query vectors are weighted as 1 and then the similarity between this query vector and every other document vector is calculated. The documents are then ordered by this similarity value.

The cosine similarity coefficient (Equation 2) was used for the calculation of the similarity value between the query and document vectors, which represents the cosine angle between the two vectors.

$$SIM(x, y) = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2} \sqrt{\sum y_i^2}} \quad (2)$$

where

x_i, y_i : i th term weight in vector x and y .

These values range from 0 to 1, where 0 means two vectors are totally dissimilar (or there is no matching term between them) and 1 means that they are identical. Therefore, if a document completely matches a user query, the equation produces a similarity value of 1 and if a document contains no matching term, it produces a similarity value of 0. In reality, most of the

similarity values are in between these extremes and a specific cutoff point is chosen in order to retrieve documents with higher similarity values. In this study, the cut off point was set to 0.01; thus, only the documents which have a similarity value higher than 0.01 are retrieved and presented to the user (at the rate of 20 documents per page).

3.2 Social search model

The social search approach is achieved by synchronizing social navigation to support KSII. This synchronization is achieved by sharing social navigation information. The social search system retrieves information from shared storage and provides it with ad-hoc information retrieval results. When a user performs operations such as viewing documents and leaving annotations about them, relevant information is stored and the information for the corresponding user is updated so that it can be used for other components of KSII which support social navigation.

The system supports two types of social navigation as described in section 2: traffic-based social navigation and annotation-based social navigation. Traffic includes user and group traffic. User traffic means the number of times the current user has viewed a page and group traffic means counting the viewing frequency of other users who are included in a specific group.

Annotation type (Praise, General, and Question) as well as user attitude toward a resource (Positive or Negative) are logged.

These various types of information about a resource are represented using different colors and icons. Background color intensity represents group traffic, which gets darker as group traffic increases. Meanwhile, user traffic is represented by the foreground color intensity of the human icons. The contrast between the two gives more information: A comparatively darker foreground color means that the user has higher traffic than the group's average while a lighter foreground color means that the user has lower traffic than the average.

Annotation is represented with color and icon too. The intensity of the yellow background color represents the number of annotations made by group users, which gets darker as the number increases. In order to show different annotation types, sticky-note, thumbs-up, and question-mark icons are used for the three different annotation types discussed above. User's attitudes are represented with a thermometer icon. For positive and negative annotations, red and blue thermometer icons are used respectively.

3.3 Search result presentation

The retrieval results of the social search approach in this study are presented in a similar way to the conventional search

results at first glance—as an ordered list. Figure 3 shows the social search results for the sample query “pointer memory function.” The query terms are stemmed and passed to the search engine. 504 documents with higher than the cutoff value are retrieved and 20 of them are displayed in the first page. Document sources, titles, and similarity scores are displayed for users who can then click to view the content of any document. In addition to these ad-hoc retrieval results, social navigation information appears in the state column. This represents traffic- and annotation-based social navigation information, as described in the previous section.

In Figure 4, two different records with equal similarity values are presented. Even though their similarity to a given query is identical, the traffic information is different. Users can easily understand that group users have visited the second document more than the first one by its darker background color. They can also see that the current user visited these two documents as frequently as other group users because the foreground and background colors of each are identical.

Figure 5 shows an example of annotation-based social navigation information for a document retrieved by the social search system. Along with the title and similarity score, traffic- and annotation-based social navigation information is also provided in the form of thumbs-up and thermometer

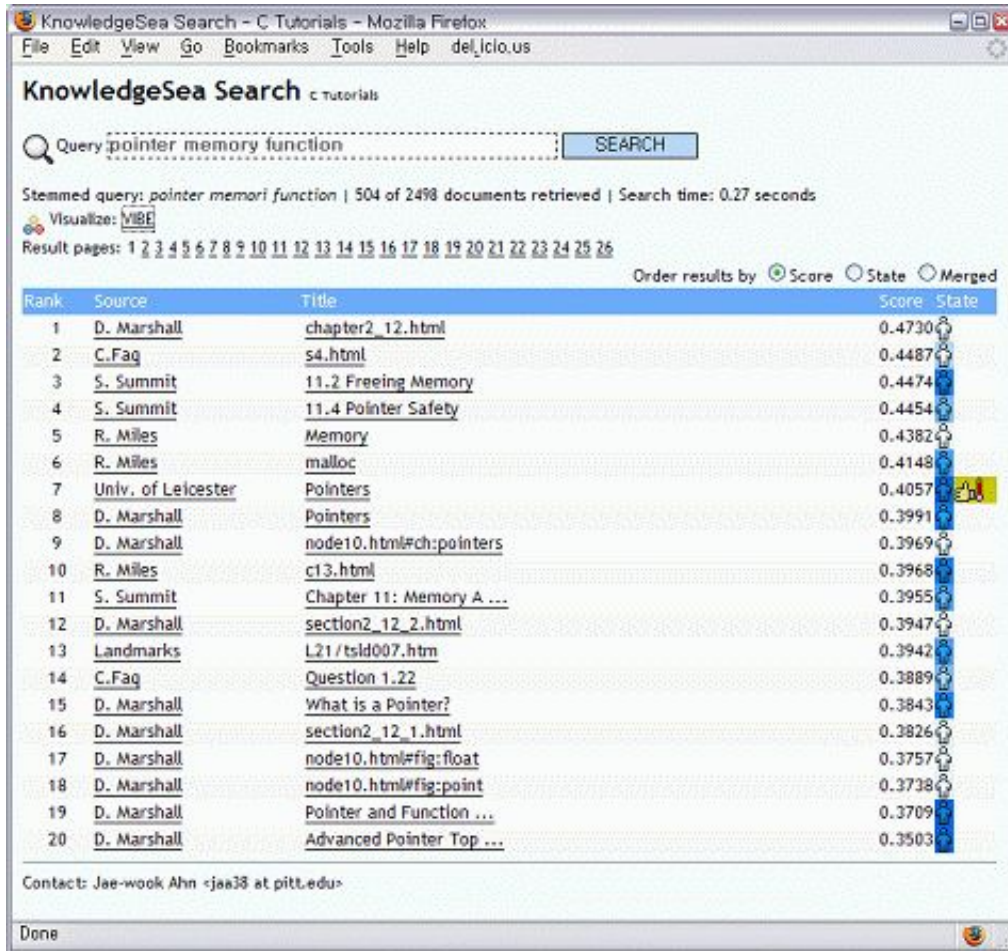


Figure 3 Social search results



Figure 4 Traffic-based social navigation



Figure 5 Annotation-based social navigation

icons in pertinent colors. With this information, users can see that the document has been annotated many times (there is a darker background color). They can also see that it embeds Praise-type annotations (thumbs-up icon) and the users had a

positive regard for it (red thermometer icon).

Along with the ordered list presenting the search results, the current system also supports a VIBE visualization. The VIBE visualization framework is used for a two-level visualization for KSII but it is also

simultaneously applied to the search result visualization. Figure 6 is a visualization example for the same query and the results of the ordered list in Figure 3. It displays three query terms and 20 retrieved documents (one page of ordered lists). The position of each document is determined by its relative similarity to the POIs (Point Of Interest) or query terms by the following algorithm (Olsen et al., 1993).

- (1) If a document is influenced by one POI only, its icon is positioned on top of this POI.

- (2) If a document is influenced by two POIs, the position of this document will be on the line between the POIs, closer to the one that has the higher similarity score.

- (3) If a document is influenced by more than two POIs, a temporary position is determined by the first two POIs as in step (2), and then a new position between the temporary position and the next POI is determined in the same manner used in step (2).

This positioning function is based on

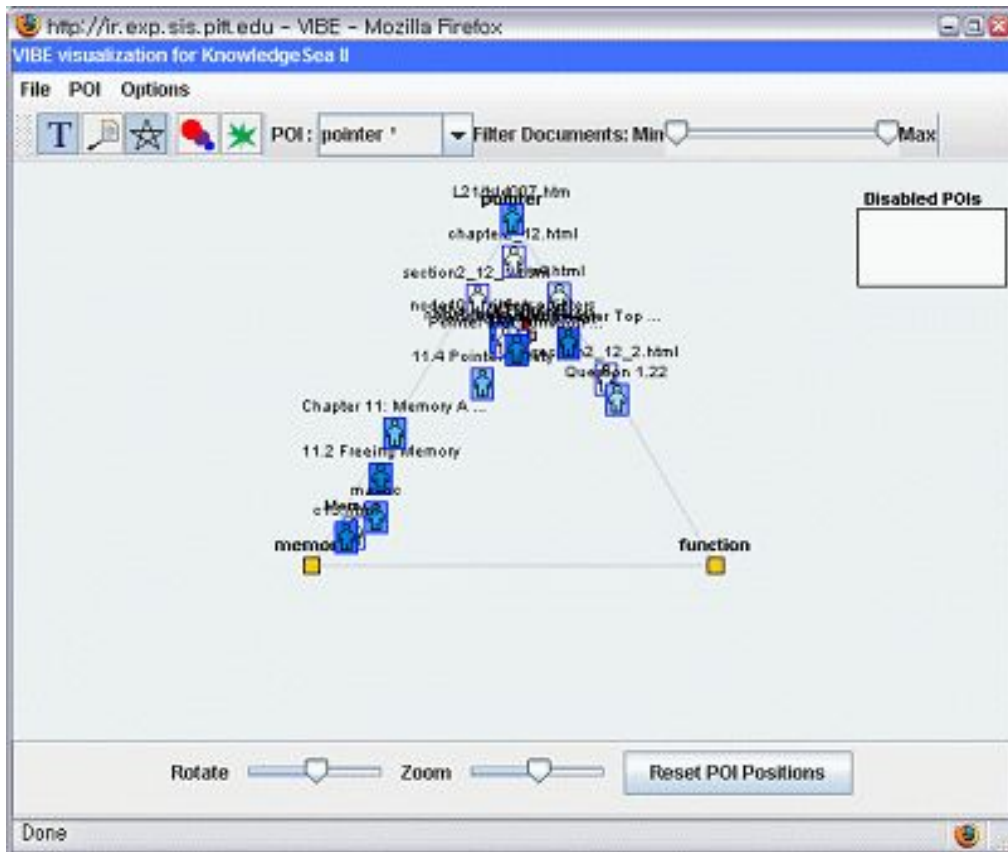


Figure 6 VIBE visualization of the search results

the similarity ratios between documents and POIs with the distance from a POI to a document representing its influence on the document in terms of similarity. If a document is closer to a POI, then users can assume it is more similar to that POI than other POIs. By dragging the POIs, users can explore and understand the relationships between query terms and documents because the document positions are interactively updated by POI position changes. For detailed information about this visualization framework, see Ahn, Farzan, Brusilovsky 2006; Olsen & Korfhage 1994; Olsen et al. 1993.

4. Evaluations

In order to test the validity of our social search approach in regards to the addressed problems, we split the evaluation procedure into two stages for answering the following questions.

- 1) User preference for the social search approach in the context of social navigation
- 2) Usefulness of the social search approach in the context of social navigation

We analyzed the transaction log and a questionnaire for the first evaluation stage and conducted a log analysis for the second stage. KSII logs all transaction data that occurs when users access the system. The

data had been collected in a classroom during the Fall 2004 semester when the system was made available to students in an introductory C language programming course at the University of Pittsburgh and they were asked to fill in an anonymous questionnaire. The use of the system and the questionnaire was not mandatory. 15 students were enrolled in the class; 13 used the KSII system, and 10 responded to the questionnaire. The sections below describe the results of the analyses.

4.1 Analysis of user preference for social search

KSII supports four different types of information access methods, including the social search system. Before we test whether the social search approach is valid for helping users access relevant resources, we need to know whether users like the present system and how they use it. For this purpose we conducted log analyses and asked about user preference, using a questionnaire. Table 1 shows the log analysis result for the top 20 accessed pages and the frequency of each access method used to reach them. These pages were accessed 214 times with the map-based visualization being the most popular method to access them. Searching accounted for 10% of the total accesses. This result was different from our expectation that the search service would increase the overall

access rate of users.

Map-based visualization	Hypertext browsing	Recommendation	Social search
43%	35%	12%	10%

Table 1 Accessing method for the top 20 accessed pages

However, according to the answers given in the questionnaire about their attitude towards the need for the search interface and social navigation, most students (over 80%) responded with positive answers:

“Strongly agree” or “Agree” (Figure 7). Their preference for the search interface was slightly higher than their preference for social navigation. Even though the usage rate of social search is far below that for the map-based visualization or hypertext browsing, the users showed very high preference for this new access method.

4.2 Analysis of usefulness of the social search approach

4.2.1 Random versus effective access rate

We compared the effective and random relative access rates for links with higher rankings (placed on the top of retrieval lists) and links with traffic-based cues. The

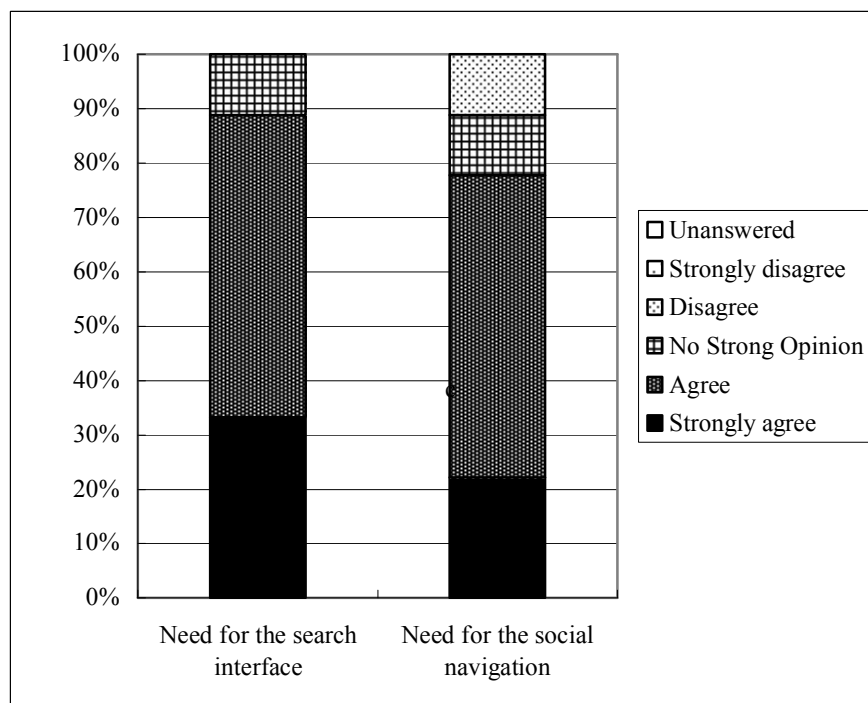


Figure 7 Students' attitude towards social search system

random relative access rate reveals which fraction of clicks will be made on links with a specific property if the user selects links in the search results list randomly. It shows how often the links with this property appear in the search results. The effective relative access rate shows the real fraction of links with the target quality, among all accessed links. If the effective relative access rate is higher than random, it means that the links with this quality encourage users to access them. In the reverse case, links with rates lower than random discourage users from accessing them.

We tried to answer two questions. The first one was: “Do students prefer links that are ranked higher?” We considered items from rank 1 to 3 as high rank and the random relative access rate for these top three would be $3/20 = 0.15$. Students accessed 53 documents from different search result lists and 16 of them were among the top 3 documents. Therefore, the effective relative access rate is $16/53 = 0.3$, which is twice the random. This shows that students preferred links on the top of the list, which is similar to the results from the previous section.

The second question was: “Do students prefer links with traffic-based social navigation cues?” We separated links with any visible past traffic (number of past clicks > 1) and links with higher traffic (number of past clicks > 2) in order to

answer this question. Even though the links with one past click shows a different color cue than the one without traffic, the difference is not very big and users might ignore the difference. However, the color cue when there are at least 3 past clicks shows quite an evident difference, meaningful enough to separate these two cases visually. The comparison between random and effective access rate is provided in Table 2. Out of 53 cases, students chose 17 documents from the visible traffic (at least one traffic) category and the effective relative access rate for links with visible traffic is $17/53 = 0.32$, which is four times higher than random access rate (0.08). For items with higher traffic (group traffic > 2) the random versus effective rate was 0.05 to 0.19, where the latter case also showed a higher preference.

	Random Chance	Access Rate
Group Traffic > 1	0.08	0.32
Group Traffic > 2	0.05	0.19

Table 2 Effect of traffic-based social navigation on visiting pages within search results

4.2.2 Time spent reading pages

Users cannot see the contents of the retrieved items until they open and examine

them. We can assume that the time spent reading (TSR) those items by users reflects the “true relevance” of the documents retrieved. The larger the TSR of pages are, the higher the implication of a “true relevance.” We first compared high and low ranked items in the retrieved sets in terms of the number of accesses to them. Table 3 shows that the users’ preferred high rank items, about half of the total accesses. We then compared the TSR of documents (Table 4), which shows that the users spent a lot of time reading items with high rankings. However, comparing the group traffic of high rank items, students spent more times reading items with high group traffic (56.5) than those with low group traffic (8). These results show that the users are influenced by the presence of visual social navigation cues as well as a high ranking in the list of retrieval items.

Number of documents			
	Low Rank	High Rank	Total
Low group traffic	19	12	31
High group traffic	9	8	17
Total	28	20	48

Table 3 Document distribution by rank and group traffic

Median TSR			
	Low Rank	High Rank	Total
Low group traffic	50	8	25
High group traffic	21	56.5	31
Total	41.5	13	26.5

Table 4 Effect of social navigation on accessing from search results, by TSR

Conclusions

This paper discussed a social searching methodology in the context of social navigation in order to solve the problems embedded in traditional information retrieval strategies. Traditional strategies lack personalization and cannot support the search task’s context. A social search system was implemented on an existing platform, the open-corpus Web-based educational support system called KnowledgeSea II. The new version supports various access methods to help users find relevant information within a pool of Web-based educational resources. The idea which binds this variety of methods together is called social navigation, which incorporates social information from groups of users and exploits it to help guide individual users. Two types of social

navigation information have been used in this framework: traffic- and annotation-based social navigation information. The social search methodology is synchronized with other services by sharing this information and can provide users with not only the conventional ad-hoc retrieval results, but also some additional social information, representing the opinion and popularity of the items as rated by colleagues.

The validity of this approach was tested in various aspects. First, we tried to measure the preference for the search service as compared to other methods provided by KSII. Even though social searching accounted for smaller amounts of usage than other methods, according to log analysis results, users themselves agreed to the importance and the need of it by answering very positively in the questionnaires. We then tried to assess the usefulness of social searching by comparing random versus effective access rates and time spent reading pages (TSR). The results show that the proposed social searching methodology was effective in serving user goals to access relevant information by incorporating and providing social navigation to users.

Our future plan includes more extensive evaluations. The current system supports additional domains, such as information retrieval, HCI, and digital libraries'

educational resources, beyond the C language resources introduced in this study. Not enough data has been collected because of the new domains were only recently introduced, but they will be evaluated and compared in the near future. Also, new technologies are being added, including the VIBE visualization for search result presentation. The evaluation of them will include the evaluation of system-to-user interaction and the use of social search with social navigation characteristics.

References

- Ahlberg, C., and Shneiderman, B. 1994. "Visual information seeking: Tight coupling of dynamic query filters with starfield displays." In: *Human Factors in Computing Systems*. ACM Press, New York, 313-317.
- Ahn, J., Farzan, R., and Brusilovsky, P. 2006. "A Two-Level Adaptive Visualization for Information Access to Open-Corpus Educational Resources." Accepted to *Adaptive Hypermedia 2006 Workshop on the Social Navigation and Community-based Adaptive Techniques*.
- Brusilovsky, P., Chavan, G., and Farzan, R. 2004. "Social adaptive navigation support for open corpus electronic textbooks." In: P. De Bra (ed.) *Proceedings of Third International Conference on Adaptive Hypermedia*

- and Adaptive Web-Based Systems (AH'2004)*, Eindhoven, the Netherlands.
- Brusilovsky, P., and Rizzo, R. 2002. "Using maps and landmarks for navigation between closed and open corpus hyperspace in Web-based education." *The New Review of Hypermedia and Multimedia* 9:59-82.
- Dieberger, A., Dourish, P., Hook, K., Resnick, P., and Wexelblat, A. 2000. "Social Navigation: Techniques for Building More Usable Systems." *Interactions* 7(6):36-45.
- Dron, J., Boyne, C., and Mitchell, R. 2001. "Footpaths in the Stuff Swamp. In: Fowler, W. and Hasebrook, J. (eds.): *Proc. of WebNet'2001, World Conference of the WWW and Internet*, Orlando, FL, AACE, 323-328.
- Freyne J., Smyth B. 2004. "An Experiment in Social Search." In: P. De Bra (ed.) *Proceedings of Third International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'2004)*, Eindhoven, the Netherlands.
- Heylighen F. 1999. "Collective Intelligence and its Implementation on the Web: algorithm to develop a collective mental map." *Computational and Mathematical Theory of Organizations* 5(3):253-280.
- Kohonen, T. 1997. *Self-Organizing Maps*. Springer-Verlag.
- Lowrence, S., and Giles, C.L. 1998. "Context and Page Analysis for Improved Web Search." *IEEE Internet Computing*, July-August, pp38-46.
- Olsen, K.A., and Korfhage, R.R.. 1994. "Desktop visualization." In *Proceedings 1994 IEEE symposium on visual languages*, St. Louis, MO: IEEE. pp 2349-244.
- Olsen, K.A., Korfhage, R.R., Sochats, K.M., Spring, M.B., and Williams, J.G. 1993. "Visualization of a document collection: The vibe system." *Information Processing and Management* 29(1):69-81.
- Porter, M.F. 1980. "An algorithm for suffix stripping." *Program* 14(3):130-137.
- Roussinov, D.G., and Chen, H. 1998. "A Scalable Self-organizing Map Algorithm for Textual Classification: A Neural Network Approach to Thesaurus Generation." *CC-AI, The Journal for the Integrated Study of Artificial Intelligence, Cognitive Science and Applied Epistemology* 15(1):81-111.
- Salton, G. 1989. *Automatic Text Processing*. Addison-Wesley Publishing Co., Reading, MA.
- 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-Adapted Interaction* 14(5): 383-423.