Exploiting Knowledge-in-the-head and Knowledge-in-the-social-web: Effects of Domain Expertise on Exploratory Search in Individual and Social Search Environments

Ruogu Kang, Wai-Tat Fu, & Thomas George Kannampallil
Human Factors Division and Beckman Institute
University of Illinois at Urbana-Champaign
{kang57, wfu, tgk2}@illinois.edu

ABSTRACT
Our study compared how experts and novices performed exploratory search using a traditional search engine and a social tagging system. As expected, results showed that social tagging systems could facilitate exploratory search for both experts and novices. We, however, also found that experts were better at *interpreting* the social tags and *generating* search keywords, which made them better at finding information in both interfaces. Specifically, experts found more general information than novices by better interpretation of social tags in the tagging system; and experts also found more domain-specific information by generating more of their own keywords. We found a dynamic interaction between knowledge-in-the-head and knowledge-in-the-social-web that although information seekers are more and more reliant on information from the social Web, domain expertise is still important in guiding them to find and evaluate the information. Implications on the design of social search systems that facilitate exploratory search are also discussed.

Author Keywords
Search behavior, domain expertise, exploratory search.

ACM Classification Keywords
H.3.3 Information Search and Retrieval: Search Process. H.5.3 Group and Organization Interfaces: Collaborative computing, Web-based interaction. H5.4. Information interfaces and presentation (e.g., HCI): User Issues.

General Terms
Human Factors, Performance.

INTRODUCTION
While information seeking is evolving into a social and collaborative activity, the nature of search is no longer limited to simple query-based fact retrieval. The Web has become a diverse platform for exploring, learning, organizing and sharing information. Indeed, many researchers have argued that the page-ranked search results returned by traditional search engine cannot satisfy the growing variety of users' information needs (e.g., [13]).

In this paper, we focus on how users with different domain expertise perform *exploratory search* - exploring for information in order to learn about a topic - when using traditional search engines such as Google (www.google.com) and social tagging systems such as Delicious (http://delicious.com/). There have been discussions comparing the advantages and disadvantages of social search systems and traditional search engines. Previous studies suggest that users could benefit from a social search environment by reading information cues (e.g., social tags) left by other users, that act as "signposts", guiding them to the right information (e.g., [2, 9, 11, 14]). By using these tags as trails for information search leads users through an exploratory process of learning and knowledge acquisition [17]. While direct fact retrievals have well-defined information goals, exploratory search often involves an iterative process of exploration and goal refinement as knowledge is incrementally acquired during the process [5]. It is therefore reasonable to assume that, for exploratory search, social information systems are more desirable than traditional search engines, as social tags can act as navigational cues that guide the iterative search and understanding process. In fact, many researchers have argued that the traditional search engines are becoming less useful for people to *make sense* of the search results, which is much more important than fact retrievals for multiple purposes [5,6]. Although this idea may seem intuitive, to our knowledge, there is no direct comparison that shows specifically in what ways a social information system may assist exploratory search more effectively than traditional search engines, and it is also still unclear how domain expertise may facilitate exploratory search in different search interfaces.

Previous research has shown that domain expertise influences how people search on traditional search engines (e.g., [4, 16]). Search performance is found to be better when users are searching in their own domain of expertise. As social tagging systems enable users to see other users'

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.

interpretation of the same information contents, we hypothesize that users' domain knowledge would also impact what queries they use to search for information, how they evaluate the relevance of search results, how they interpret the content tagged by others and what tags they assign to different information content. In other words, domain knowledge may play an even more important role in exploratory search than simple fact-retrievals, as it may influence the *interpretation* and *learning* process in the iterative exploratory search process as users are interacting with a social tagging system.

To understand the role of domain expertise in exploratory information search, we conducted a laboratory experiment to investigate experts and novices' search behavior in two search interfaces. The main goal of the study is to investigate: (1) how experts and novices search and interpret information in individual and social search environments, (2) how and why a social tagging system may facilitate exploratory search, and (3) given the knowledge about *how* users with different levels of expertise perform exploratory search, what are the implications for the design of next generation of social information systems.

RELATED WORK

In addition to serving the function of categorization and description of information contents, many researchers have argued for social tags' roles as navigational cues (e.g., [9, 14, 15]). More recently, there is an added focus on comparing social tagging systems with traditional search engines in terms of exploratory search efficiency. Morrison [15] argued that social tags, interpreted as folksonomies, would have as much precision as search engine. He also claimed that folksonomies have an advantage for finding new information compared to search engines. Heymann et al. [9] analyzed large datasets from Delicious and suggested several interesting relations about tags and URLs. Similar to the results obtained by Morrison [15], they found that users in Delicious were more interested in newly added URLs. This recency effect, similar to Cattuto's work [3], may be attributed to the possibility that those URLs created recently are tagged more frequently. They also found a relatively high overlap between popular query terms and popular tags. They therefore argued that most tags in social bookmarking system are relevant and effective for information search. They also proposed that tags associated with bookmarks in a bookmarking system are more useful than typical link text (e.g., page titles) returned from a search engine. Krause et al. [12] compared user activity and behavior from Delicious, MSN, AOL and Google by analyzing tags and queries. By comparing the total number of query terms in MSN and tags in Delicious, they showed that MSN has a significantly larger number of terms but the average frequency of each item was quite similar in both systems, indicating that Delicious users focused on fewer topics, *but* each topic was reached by users equally often.

Despite all these similarities on accuracy, recency and richness of search results, the social influence in tagging system is not replicable in traditional search engines. Golder and Huberman [8] found that tag choices were influenced by tags created by others even if they had different information needs when tagging. Sen et al. [16] showed that available tags in a tag community could directly impact a user's tendency in choosing tag vocabulary. Therefore, we hypothesize that the social environment in tagging systems, as suggested in these prior research reports, would have the potential to support knowledge exchange during the information search process through bookmarks and categories.

Prior research has also found that differences in domain expertise may affect search behavior. White et al. [18] investigated the queries, search sessions, website selection, and rates of search successes for domain experts and non-experts using a log-based analysis. They found that domain experts were more successful in search and used more domain specific vocabularies. Zhang et al. [19] found that experts generated more queries in search and more words in each queries, but the search efficiency did not show any difference between experts and non-experts. Duggan and Payne [4], however, found that greater topic knowledge resulted in shorter search time per webpage, faster decision to leave a webpage, and shorter search queries. They found that greater knowledge could increase the ability of users to select more related information. As search query is a primary measurement of search behavior, the impact of domain knowledge on query generation or search vocabularies has received significant research attention in recent years. Allen [1] examined the impact of topic knowledge on information catalog searching. They found that higher-knowledge participants used more search expressions in catalog search than lower-knowledge participants. They also suggested a possible trend that participants with higher level of topic knowledge and who expressed difficulty in search are more likely to introduce new vocabulary into the search. Hsieh-Yee [10] found that when users have certain levels of search experience, subject-matter knowledge would play an important role affecting experienced searcher's reliance on their own language, and the use of the external term source. The result indicated that when users were searching out of their domain, they made more effort in preparing for the search, monitored the search more closely, and tried out more term combinations, but when experts searched in their domain, they used more of their own terms.

Although there are many comparisons between social tagging systems and traditional search engines, as well as studies about the effect of domain expertise in information search, there are fewer studies on how domain knowledge impacts users' search behavior in different types of search interfaces. Kammerer et al. [11] suggested that a tag-based search browser could potentially compensate for differences in prior domain knowledge. They found that exploratory tag

search system was beneficial for a novice and there was a learning effect during exploratory information search.

In summary, previous research has provided sufficient support in distinguishing the influence of expertise difference on information search. Some researchers have pointed out the usefulness of exploratory search interfaces for both experts and novices. Others have proposed that social search systems can potentially compensate for the traditional keyword-based search engines to improve information seekers' knowledge gain. However, what is still missing is how the expertise profiles of users affect their exploratory search behavior in a social search system.

METHOD

We conducted a laboratory experiment to collect data on expert and novice users' searching and tagging behavior in Delicious and Google. The main goal of the experiment is to understand the effects of domain expertise on search behavior and information interpretation in individual and social search environments as users are engaged in an exploratory topic search task.

Experimental Design and Participants

We used a 2 × 2 between-subject design to investigate the difference in users' search behavior when they were using a traditional search engine (Google) and a social tagging system (Delicious), and how users with different levels of domain expertise may interpret information differently using these two different search interfaces. While Google provides a traditional search environment for keyword-based queries, Delicious provides tagged social bookmarks created by other users that allow participants to use either tag-based or keyword-based queries. Under the assumption that domain expertise might influence how well participants could generate keywords to search, interpret search results, and select social tags, we also expected that experts and novices could adopt different search strategies when they performed exploratory search using the two interfaces.

	Google	**Delicious**
Expert	12	12
Novice	12	12

Table 1. 2 × 2 experimental design

A total of 48 participants were recruited for the study (22 female, 26 male, mean age = 24.4). All participants were skilled computer users with more than 10 years of computer usage experience (mean = 13.8). All participants reported Google as their most familiar search engine and that they performed Internet searches with an average frequency of 3.95 on a 5-point scale (4 means "use search engine several times a day"). 24 of the participants claimed to have expertise knowledge in finance or related area (such as holding an advanced degree or had current or prior employment experiences in the finance industry). The other 24 did not have any training or knowledge in finance or related fields. In addition to the self-claims on their knowledge backgrounds, we used three additional methods to further verify participants' expertise level (discussed later). Expert and novice participants were randomly assigned to one of the two interfaces in the 2 × 2 experimental design (see Table 1). Participants were paid $25 for their participation in the experiment.

The Exploratory Search Tasks

We used "financial crisis" as the topic for our exploratory search task. This topic was chosen for its current relevancy and differences in the depth of knowledge about the topic between subject matter experts and the general public (topic novices). Participants were asked to imagine that they were to collect information from the Web to give a talk on the current financial crisis. They were encouraged to explore information using their assigned search interfaces (Google or Delicious) to enrich or supplement their own knowledge. During their search activities, participants were asked to *save and tag* useful websites as bookmarks. In Delicious, they could save websites as bookmarks to their assigned web account, while in Google they were instructed to save bookmarks in a given folder in browser and create tags for the resource. They were instructed to search, read, and select information, but not spend too excessive time on a single web page. We collected the following data from each of the participants in both conditions.

Self-report

Participants were asked to complete a short survey with 5 questions about their knowledge of finance and economics as well as their familiarity of the current financial crisis on a 5-point scale. Sample questions include: "I know the causes and backgrounds of the current financial crisis", "I can give my own opinion about what should be done to deal with this crisis", etc. We found a high reliability for the self-report questionnaire (Cronbach's alpha = 0.921).

Knowledge Questionnaire

We used a knowledge questionnaire to test the participants' specific knowledge about the financial crisis. There were 20 questions in total, of which 10 were general questions such as, "Which event precipitated the current financial crisis?" The rest 10 questions are specific questions that required professional training in finance or economics (e.g., "Which of the following is the investors' strategy against the unsystematic risk?"). Questions in the knowledge questionnaire were collected from online quizzes about the financial crisis and from textbooks. The questionnaire was reviewed by two graduate students majoring in finance and one professional with more than 15 years experience in a financial holding company.

Topic description

The participants were also asked to perform a topic description task before and after they did the information

search. In this task, the participants were given the topic "financial crisis" and were asked to write down words or phrases to describe the topic. This task tested the fluency of the concepts that the user generated to associate with the topic. Our purpose was to measure their understanding about the topic based on their retrieval of terms and concepts from the memory [7]. We expected to see a conceptual knowledge change by analyzing the topic description results before and after information search.

Categorizing

Bookmarks and tags created by the participants were presented to them after the information search task. The participants were required to categorize their bookmarks and provide a label to each category. The categorization of bookmarks is considered as a direct measure of whether the selection of bookmarks is influenced by their domain knowledge. We designed this categorization task in order to understand their knowledge gain from the information search process on a higher level and see how experts and novices interpret the search results differently based on their own knowledge structures.

Procedure

Participants were first given general information about the experiment and the goal of this study, and were then asked to read and sign the consent form for participating in the experiment. Participants then filled out a general survey about their demographics and a self-report survey on their knowledge background. After that, they were asked to do the pre-test topic description task. For the topic description task, they were asked to write down terms/phrases about the topic on a sheet of paper and stop at anytime when they were done. On average, the topic description task took about 5 minutes. Then participants were randomly assigned to the Google or Delicious condition. The researcher briefly explained the task and demonstrated how to use the search engine or the social tagging system and how to create tags and save bookmarks. Participants were provided enough time to familiarize themselves with their tasks and the interfaces before they started the experiment, during which the experimenter would answer any questions that they had. Participants performed their tasks individually and were given a maximum of 1.5 hours for their task.

The Camtasia recorder was used to record all on-screen actions of the participants including information searching, bookmark selection, tag creation and URL clicks. After finishing the search task, participants performed a post-test topic description task. Then they would complete the knowledge questionnaire. The knowledge questionnaire was given after the search task to avoid potential priming effect on their search behavior by the knowledge questions. After the participants filled out the knowledge questionnaire, the researcher provided a printed copy of the bookmarks and tags that they had generated during their search task. The participant was then asked to categorize the bookmarks into groups and give a label to each group. A short open-ended interview was conducted in the end regarding the participants' opinions about the search interface, tagging process and categorization. The whole experiment took about 2.5 hours.

RESULTS

Identifying Domain Knowledge

From the self-reported expertise ratings, we found significant difference between the groups (mean = 3.8 and 2.87 on a 5-point scale for experts and novices respectively, $p<0.001$). Consistent with the self-reported ratings, we found a significant difference on the general knowledge test score between experts and novices ($p<0.05$), as well as on the 10 professional questions in the questionnaire ($p<0.01$). Experts also generated more terms to describe the topic of "financial crisis" than novices, both before and after the task. Paired-sample t-test showed that only in the expert-Delicious group, there was a significant difference in the knowledge gain after the search tasks ($p<0.05$). This result suggested that the knowledge of experts improved after searching using Delicious, but not for the other groups.

Search Behavior

To analyze the differences in search behavior, we compared the number of bookmarks and tags created, the number of URL visits, and the number of URL visits per bookmark saved for each participant across the groups. While the number of bookmarks and tags created reflected the effectiveness of their search behavior, the number of URL visits per bookmark saved indicated how efficiently participants could find relevant information using the interfaces.

Analysis of variance (ANOVA) showed that the main effect of interface was significant for the total number of tags created ($F(1, 44) = 4.105$, $p <0.05$), but the main effect of interface for other variables did not reach significance. The main effect of expertise was marginally significant for number of bookmarks created ($F(1, 44)=560.3$, $p=0.08$), but was not significant for other measures. The interaction effect of expertise and interface was significant for the number of tags created ($F(1,44) = 6.146$, $p <0.05$). We also found that the significant interaction between expertise and interface for the number of URL visits per bookmark saved ($F(1,40) = 10.148$, $p<0.01$) (see Figure 1).

As Figure 1.a shows, experts saved more bookmarks than novices, but the interaction effect between expertise and interface was not significant ($p = 0.2$). Post-hoc analysis showed that experts collected more bookmarks in Delicious ($p<0.001$) but the difference between experts and novices in Google was not significant. We also found that experts using Delicious generated more tags than novices ($p < 0.05$), but the difference between experts and novices was not significant in the Google group (Figure 1.b).

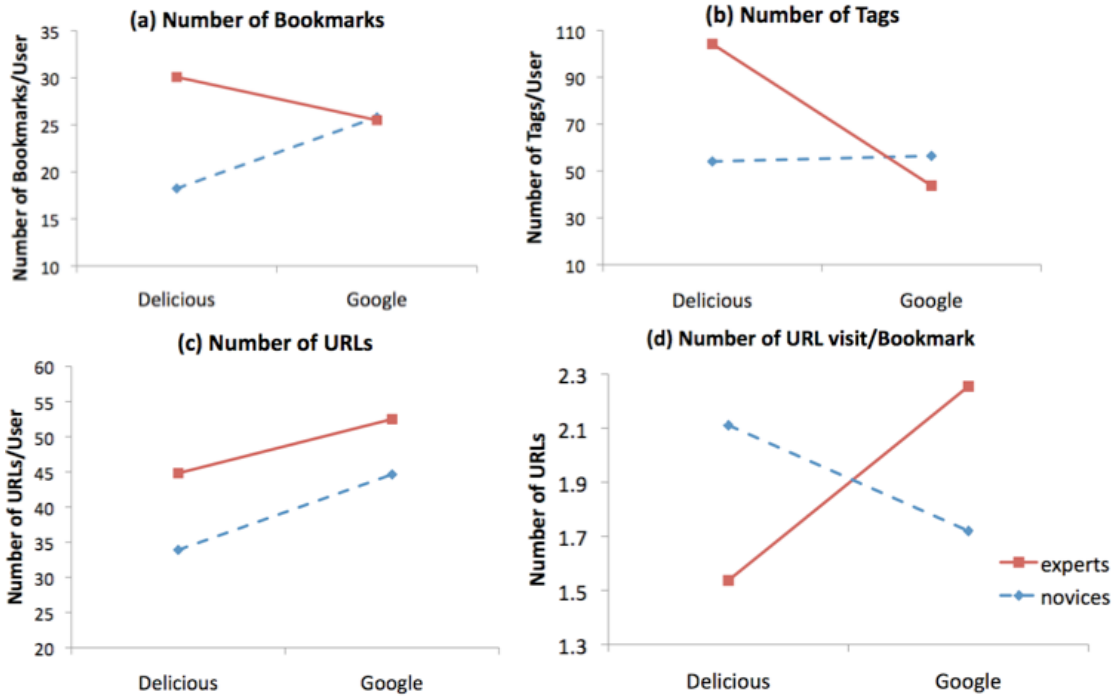

Figure 1. Comparison of search behavior between experts and novices in two interfaces

Results showed that, as expected, domain expertise facilitated information search when they used either the search engine or the social tagging system, as reflected by the higher number of bookmarks saved by experts. Experts also created significantly more tags when using Delicious than novices. This can probably be attributed to the fact that experts were better at interpreting tags created by others, as well as generating terms to describe the bookmarks (we will present further analysis on this). However, domain expertise did not induce the same difference in the Google group. At least in terms of the total number of bookmarks saved and the number of tags, novices and experts were about the same in their performance when using Google.

Figure 1.c shows that experts visited more URLs than novices in both interfaces, although the difference was only marginally significant ($p=0.14$). The main effect of interface and the interaction between interface and expertise was not significant for number of URL visits. However, when we analyzed the number of URL visits per bookmark saved (see Figure 1.d), we found a significant interaction between interface and expertise ($p<0.01$). This measure of the number of URL visits per bookmark saved reflects the efficiency of search, as a smaller number would indicate that the number of relevant bookmarks saved was higher per unit browsing action. Post-hoc analysis confirmed that the search efficiency for experts in the Delicious group was significantly better than novices ($p<0.01$). Interestingly, novices visited more URLs to find a relevant bookmark when they used Delicious compared to Google ($p<0.05$), which could indicate that novices had lower efficiency when searching in Delicious than Google.

The results suggested that experts and novices search differently in two search environments, in terms of the number of bookmarks saved, tag creation, and the search efficiency in finding relevant information. We found that experts collected more information (as reflected by the number of bookmarks), created more tags, and have higher search efficiency in Delicious compared to novices in the same condition. Domain expertise was found to be a major factor influencing the information collection and interpretation. One reason for the higher search efficiency in Delicious could be attributed to the social tags that may facilitate the evaluation of the relevance of links before users click on them.

Types of Queries

We also looked at the query generation of the four groups of participants. Figure 2.a shows the number of keyword-based queries (entering keywords in keyword search box) performed by experts and novices in each interface. ANOVA showed that the main effect of interface was significant ($F(1, 41) = 7.341$, $p<0.01$), as well as for the interaction between expertise and interface ($F(1,41) = 3.109$, $p<0.1$). The main effect of expertise was not significant. As Figure 2 shows, participants in Delicious in general used less keyword-based search than Google, and the interaction was mostly caused by the difference between experts and novices in the Delicious group. Indeed, post-hoc analysis showed that experts performed significantly more keyword-based queries than novices in Delicious ($p<0.01$), but the difference was not significant for the Google group.

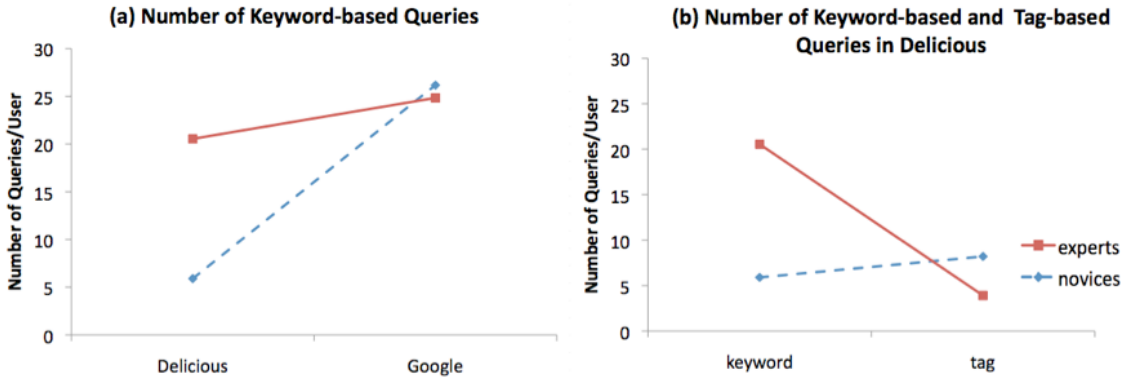

Figure 2. Search strategies employed by four groups of participants

To further understand the difference in search strategy brought by expertise difference in the Delicious group, we performed an separate ANOVA on the use of tag-based queries (selecting tags from the popular tags or other users' tags attached to each website title) and keyword-based queries (entering keywords in keyword search box) for experts and novices in Delicious (as in Google there was no tag-based query). As Figure 2.b shows, the interaction between query type (tag-based or keyword-based) and expertise was significant ($p<0.05$). Post-hoc analysis showed that experts used more keyword-based search than novices ($p<0.05$), while novices used more tag-based queries than experts.

The interaction effect between expertise and query types in Delicious provided additional evidence that experts were more able to come up with their own terms to search. Experts conducted more keyword-search but less tag-search in Delicious. This result implied that experts used their own terms more [10] by using more keyword-based search, but novices relied more on directly using others' tags to search.

Information Sharing by Bookmarks

In addition to search behavior, we also examined the bookmarks saved by different groups. 48 participants selected 1170 bookmarks. Among those 1170 bookmarks, 359 bookmarks were saved by more than 2 participants, 811 bookmarks were saved by only one participant. In total 937 distinct bookmarks were saved by all users.

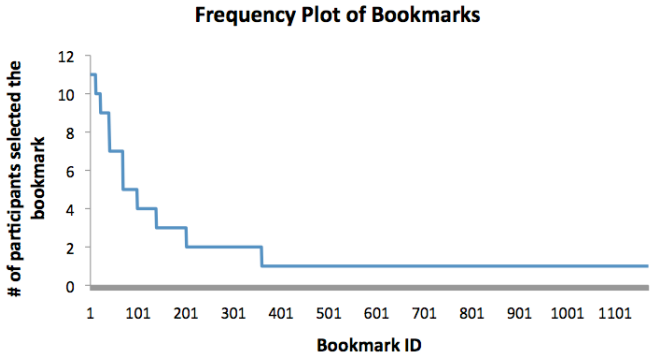

Figure 3. The frequency-rank plot of bookmarks saved by all participants

Figure 3 shows the frequency-rank plot of the bookmarks saved by all users. Consistent with previous studies, the plot of the number of participants sharing each bookmark shows a typical power-law distribution [3], in which there were rapid drops in the frequencies of bookmarks as rank increased, and there was a long tail indicating that many unique bookmarks saved by individual participants. The most common bookmark saved by 11 participants was the wikipedia page on subprime mortgage crisis (http://en.wikipedia.org/wiki/Subprime_mortgage_crisis).

Popular vs. unique bookmarks

To study the extent to which different interfaces and expertise may lead to the saving of more unique or more shared bookmarks, we divided the bookmarks saved by all participants into two groups: the bookmarks shared by 3 or more participants in our experiment were called *popular* bookmarks; the bookmarks shared by 2 or less participants in our experiment were called *unique* bookmarks. Each of the popular and unique bookmarks were shared by an average of 307 and 21 users in Delicious respectively, which at least partially validated the general "popularity" of these bookmarks as reflected by the massive number of users in Delicious.

We were interested in finding out how participants with different level of domain expertise and different interfaces would save bookmarks in the popular or unique groups. To this end, we performed a 2 (shared frequency) × 2 (interface) × 2 (expertise) ANOVA using the number of bookmarks saved by each user as dependent variables.

Results showed that the main effects of expertise and shared frequency (popular/unique) were significant ($p < 0.05$), but the main effect of interface was not significant ($p = 0.91$). The interaction effect of interface × shared frequency and expertise × shared frequency was significant ($p < 0.10$). The interaction effect of expertise and interface was marginally significant ($p = 0.10$). The three-way interaction of interface × expertise × shared frequency was not significant.

As shown in Figure 4, both experts and novices selected more popular bookmarks when using Delicious than when using Google. Specifically, experts saved more popular

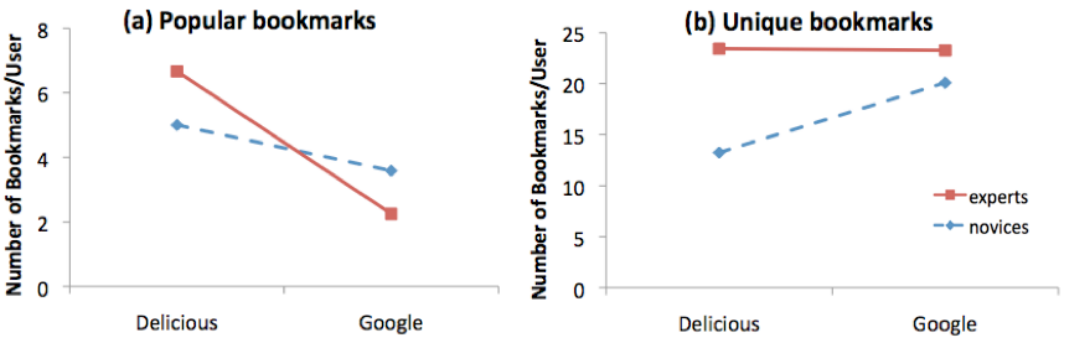

Figure 4. The impact of expertise and interface on bookmark sharing

bookmarks than novices in Delicious but novices saved more popular ones in Google.

For unique bookmarks, we can see that experts selected almost the same number of bookmarks in Delicious and Google. However, novices selected more unique bookmarks in Google than in Delicious. This result gives strong evidence to the fact that experts relied more on their own background knowledge and were influenced less by the social environment in Delicious; but novices selected less unique bookmarks in Delicious because of the stronger social influence in that condition.

Referring back to the very beginning of this paper, the number of bookmarks collected by experts in Delicious was significantly more than novices (see Figure 1.a). Now we can have a clearer picture about the difference between experts and novices in Delicious. Apparently, the higher number of bookmarks created by experts than novices in Delicious was caused by the higher number of unique bookmarks selected by experts, as evidenced by the fact that the difference between experts and novices was significant in unique bookmarks ($p<0.05$), but not in popular bookmarks ($p = 0.262$).

When we examined the content of the bookmarks, we found that most of the unique bookmarks were either specific web sites describing a particular event, or professional websites developed for finance professionals [18]. Therefore, it seemed that in human-generated indexing systems such as Delicious, novices were less likely to benefit from tags that experts gave on those unique websites, as novices might not have the background knowledge to judge whether or not those tags and websites were relevant information.

Categories of Bookmarks

To further understand how domain expertise contributes to the differences in the selection of bookmarks, we analyzed how users categorized their own bookmarks based on the content (e.g., if bookmark A was about the *cause* of the financial crisis, and bookmark B was about the *history* of the financial crisis). 242 categories were generated by 47 participants (one participant's data was lost due to technical problems). On average, each participant generated 5.15 categories, and the average number of bookmarks in each category was 5.18 (s.d. = 2.45). These categories would allow us to understand better about how experts and novices differed in their interpretation of the contents of the Web sites they found, and why they believed the information was relevant to the topic.

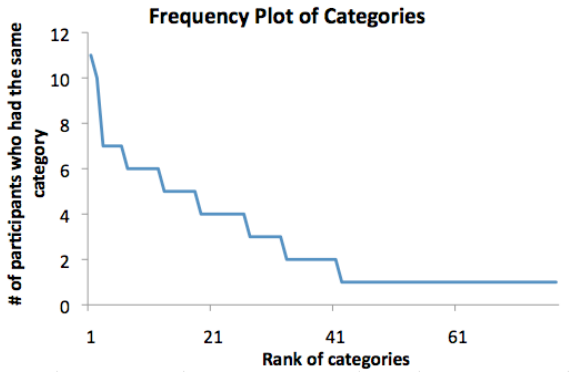

Figure 5. The frequency-rank plot of bookmark categories

Although users were free to select any words to categorize the bookmarks, we noticed that there were overlapping and semantically similar categories. Two raters merged the categories into 77 distinct categories by combining identical or semantically similar categories. The agreement between the two raters was 91%. We also found a power law distribution by plotting the number of shared categories vs. the number of users (see Figure 5). Specifically, one can also see the long tail of the distinct categories that were selected by only a single participant.

General vs unique categories

Similar to the analysis of bookmarks, we also classified the categories into two groups. In the *popular* group, more than three participants used the same category, and the rest were put in the unique group. The most shared category is "Cause of the financial crisis", which was used by 11 users. Under this category, there were 154 distinct bookmarks. The popular categories were all general or common

categories like “causes”, “history” or “explanation of the financial crisis”. In contrast to the highly shared categories, there were more categories that were unique to individual participants or shared by only two participants. Most of these categories were related to a specific company, person, event, or professional terms (e.g. “AIG”, “CDO”, etc). Table 2 shows a list of sample categories.

	Categories	# of users	# of bookmarks
Popular	Cause	11	154
	Effects	10	61
	History	7	35
	How to deal/end/to do/reaction	7	19
Unique	AIG	2	15
	Wall street	2	12
	CMBS/CDO	1	6
	Big three/ US auto	1	2

Table 2. Examples of the categories of bookmarks saved by the participants

In order to further understand how users interpreted the information they collected, we were interested in finding out whether participants with different level of domain expertise using different interfaces might be more likely save bookmarks in the popular or unique category groups. Similar to the analysis of bookmark selection, we performed a 2 (shared frequency) × 2 (interface) × 2 (expertise) ANOVA using the number of bookmarks in each category as dependent variables.

Results from ANOVA showed that the main effects of shared frequency and interface were significant ($F(1, 367)=2491.5$, $p< 0.001$ and $F(1, 367)=100.4$, $p< 0.05$ respectively), but the main effect of expertise was not significant ($p=0.26$). The expertise × interface × shared frequency three way interaction was significant ($F(1,367)=73.7$, $p<0.05$). This three-way interaction was apparently due to the significant two-way interactions between interface and shared frequency ($F(1, 367)=242.1$, $p<0.001$), as no other 2-way interaction was significant.

As the main effect of expertise was not significant, we performed separate ANOVAs on each level of expertise. We found that for the expert group, the main effects of interface and shared frequency were significant ($F(1, 217)=1162.8$, $p < 0.001$ and $F(1, 217)=78.0$, $p < 0.05$ respectively). The interaction between interface and shared frequency was also significant ($F(1, 217)=313.1$, $p < 0.001$). For novices, only the main effect of shared frequency was significant ($p < 0.001$), but all other effects were not significant. As shown in Figure 6 a and b, experts saved more bookmarks in popular categories when using Delicious than when using Google ($p < 0.001$) but the reverse was true when they are selecting bookmarks that belong to the unique categories ($p < 0.001$). Novices also selected more popular bookmarks in Delicious than in Google ($p < 0.001$), but the difference between Delicious and Google was not significant when novices were selecting bookmarks in unique categories. In addition, we compared the difference between experts and novices in four cases, and found that only the difference between experts’ and novices’ bookmark number in unique categories differed significantly in Google condition ($p < 0.001$) as shown in Figure 6.b. All other comparisons were not significant.

This pattern of results suggests that for both experts and novices, when they used Delicious to search for information, they saved more general bookmarks than using Google. Interestingly, when experts used Google, they found more unique bookmarks than when they used Delicious, but novices found similar number of unique bookmarks in Delicious and Google. Compared with the bookmark results, we can see that although experts selected almost the same number of unique bookmarks in two interfaces, they still found more bookmarks in the unique categories when using Google.

The results were consistent with the notion that social information systems such as Delicious is designed to facilitate the finding of general information, as tags created

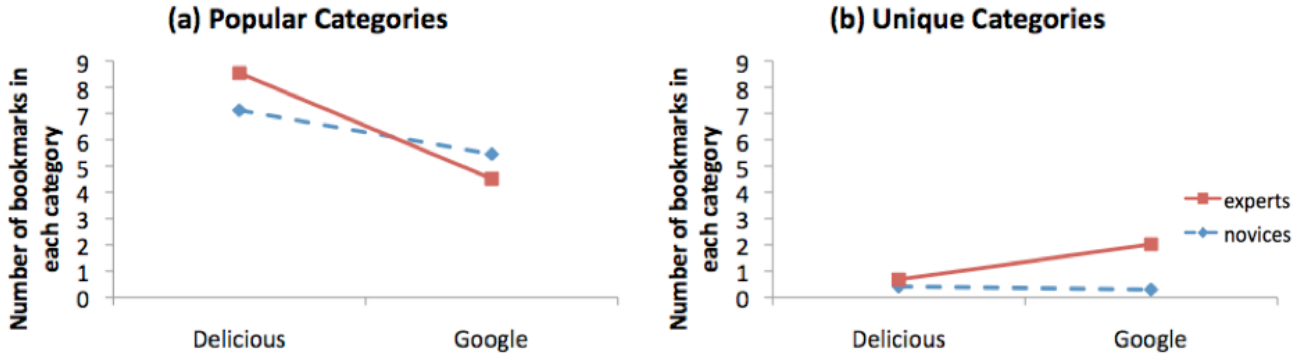

Figure 6. The impact of expertise and interface on bookmark categories

by other users increased the likelihood of finding these popular bookmarks. Therefore, experts got significantly more bookmarks in popular categories when using Delicious than using Google. In contrast, when using Google, experts utilized their domain expertise to generate "expert keywords" to search for information. As a result, they found more bookmarks that belong to the unique categories. It is also possible that because Delicious is a general social system, it may not have indexed as much domain-specific information as Google, which apparently had a much wider coverage of web sites that contain domain-specific information. However, it seemed that only experts were able to locate these unique bookmarks by coming up with keywords based on their domain expertise. For novices, the lack of domain expertise did not allow them to come up with as many keywords as experts, thus they were not able to find as many bookmarks in the unique categories as experts. Supported by the high number of bookmarks in popular categories selected by novices, Delicious did seem to help novices to find general information better than Google. This again demonstrated that social tags in Delicious did facilitate sharing of general information, even for novices who lack the domain expertise. Moreover, this interaction effect showing that experts and novices interpreted information differently in two interfaces is consistent with the bookmark results, which also illustrated the semantic imitation effect [6]. However, finding unique information still needs domain expertise to facilitate the information search, as evidenced by the much higher number of bookmarks in the unique categories saved by experts than novices.

CONCLUSION

Based on results from a laboratory experiment, we found strong evidence supporting that domain expertise (domain experts vs. novices) and search interfaces (traditional search engines vs. social tagging systems) have a dynamic interactive impact on users' information search behavior. Experts in general saved more information (bookmarks and tags) than novices. When we defined search efficiency as the number of URL visits per bookmark saved, experts had higher efficiency in Delicious and novices had higher efficiency in Google. Moreover, each search interface seems to facilitate information search in different ways. Specifically, we found that experts found more general information using Delicious and more unique information using Google, which was supported by the analysis of bookmarks and bookmark categories saved by both groups of users. At the same time, novices were able to find similar number of general information with experts, but much less unique information than experts in both interfaces.

The results on search strategies showed that experts used more keyword-based queries than novices in Delicious, while novices used more tag-based queries. This suggested that experts seemed to be capitalizing on their *knowledge-in-the-head* when performing exploratory search, but novices had to rely more on *knowledge-in-the-social-web* for their search. The results provided further support to the claim that social information systems can facilitate the sharing of useful information among novice users, and social tags do seem to have strong potential to augment exploratory search of general information, even for users who have little knowledge on the topic. However, our analysis of unique information showed a discrepancy with previous studies. The results suggested that domain experts still performed better in finding unique information in both interfaces. The assistance provided by the present social search systems could hardly help domain novices to find specific information related to a topic.

Lastly, all our results consistently showed that experts conducted better information search in the social environment, as evidenced by the higher number of bookmarks saved, the higher number of the topic description terms that they generated, and the higher gain in knowledge test scores after searching in Delicious. Although we did not find any significant improvement on novices' knowledge after their search, as previously found by others [11], this result might be caused by the limited search time in our experiment. Further research is needed to inform the learning effect in information search.

DISCUSSION AND IMPLICATION

Our study focused on how domain expertise interacts with different interfaces to facilitate exploratory search. Because of the difference in their knowledge structures, experts were driven more by a "top-down" process in information search by relying more on their knowledge-in-the-head. While this "top-down" process could bring more use of their own terms in keyword generation, it also limited their use of social tags in a social tagging environment. We also found that novices tended to use the shared knowledge of social systems (e.g., tags) more than experts, but there was no significant improvement reflected in the knowledge tests after information search. However, experts in general gained more information than novices in Delicious, as reflected by the more bookmarks, more tags and better search efficiency by the experts. In short, social search environment allows the exchange of users' knowledge in their head through the media of social web. We found that domain experts performed better as a "knowledge-receiver" in our study, but it seemed that the interface has not yet reached a state in which experts could perform just as well as a "knowledge-giver" to help novices. Future research should focus on how to provide a better interface for domain expertise to naturally "flow" into the social information systems as experts interact with them.

Inspired by the different user behaviors in the two independent search interfaces, we believe that the combined use of traditional and social search environments will be an interesting future topic to investigate. Social search systems, as well as traditional search engine, were both found to bring information gain in different conditions. In our

findings, experts have the potential to provide information cues in a social search environment, but it is not easy for novices to pick up those cues because of their different knowledge backgrounds. In addition, experts could find more general information in the social search environment, but traditional search engine was better at providing unique information.

Our results also indicated some limitations and future directions on the design of social search systems. A possible limitation of social tagging systems from the perspective of supporting exploratory search is the long tail of specific or unique bookmarks. By analyzing the bookmarks and bookmark categories in general and unique groups, we expected to see Delicious encouraging information sharing as well as knowledge exchange. However, the results showed the deficiency of social tagging systems in supporting the discoveries of domain-specific information. As expert users are more likely to generate their own search terms, especially for the unique concepts (lower frequency bookmarks or categories), it is likely that they will provide more unique and specialized tags. Therefore search results returned would have a smaller possibility to be shown on the popular page of Delicious. As a result, some useful results may be harder to be discovered because of its lower popularity.

Another possible direction for future research on social tagging systems could be on how to compress the long tail generated by different users so that they can be more "discoverable". This long tail is not caused by the "vocabulary problems" such as error tags or synonyms, but caused by the specific and unique information that is meaningful only to a relatively smaller group of users (e.g., experts). Although their popularity may be low, their quality may be high. Therefore, a better balance between the two may lead to better presentation of useful information in a social information system. As the trend in the design of information search systems is gradually transforming from the machine-user interaction to the user-user interaction, more attention should be paid to the difference between information seekers in order to design better search systems.

REFERENCES

1. Allen, B. Topic Knowledge and Online Catalog Search Formulation. *The Library Quarterly 61*, 2 (1991), 188-213.
2. Ames, M., Naaman, M. Why we tag: motivations for annotation in mobile and online media. In *Proc. CHI 2007*, ACM Press (2007), 971-980.
3. Cattuto, C., Loreto, V., Pietronero, L. Semiotic dynamics and collaborative tagging. In *Proc. National Academy of Sciences* (2007), 1461-1464.
4. Duggan, G. B., Payne, S. J. Knowledge in the head and on the web: using topic expertise to aid search. In *Proc. CHI2008*, ACM Press (2008), 39-48.
5. Fu, W.-T. The microstructures of social tagging: a rational model. In *Proc. CSCW' 08*, ACM Press (2008), 229-238.
6. Fu, W.-T., Kannampallil, T. G., Kang, R. Semantic Imitation in Social Tagging. *ACM Transactions on Computer-Human Interaction*, accepted pending minor revisions.
7. Griffiths, L. T., Steyvers, M., Firl, A. Google and the Mind: Predicting Fluency with PageRank. *Psychological Science 18*, 12 (2007), 1069-76.
8. Golder, S.A., Huberman, B.A. Usage patterns of collaborative tagging systems. *Journal of Information Scence 32*, 2 (2006), 198-208.
9. Heymann, P., Koutrika, G., Garcia-Molina, H. Can social bookmarking improve web search? In *Proc. WSDM 2008*, ACM Press (2008), 195-206.
10. Hsieh-Yee, I. Effects of Search Experience and Subject Knowledge on the Search Tactics of Novice and Experienced Searchers. *Journal of the American Society for Information Science 44*, 3 (1993), 161-174.
11. Kammerer, Y., Nairn, R., Pirolli, P. L., Chi, E. H. Signpost from the masses: learning effects in an exploratory social tag search browser. In *Proc. CHI 2009*, ACM (2009), 625-634.
12. Krause, B., Hotho, A., Stumme, G. A Comparison of Social Bookmarking with Traditional Search. *Advances in Information Retrieval* (2008), 101-113.
13. Marchionini, G. Exploratory search: from finding to understanding. *Communications of the ACM 49*, 4 (2006), 41-46.
14. Millen, D. R., Yang, M., Whittaker, S., Feinberg, J. Social bookmarking and exploratory search. In *Proc. ECSCW 2007*, ACM Press (2007), 21-40.
15. Morrison, P. J. Tagging and searching: Search retrieval effectiveness folksonomies on the World Wide Web. *Information Processing & Management 44*, 4 (2008), 1562-1579.
16. Sen, S., Lam, S. K., Rashid, A. M., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F. M., Riedl, J. Tagging, communities, vocabulary, evolution, In *Proc. CSCW 2006*, ACM Press(2006), 181-190.
17. White, R.W., Drucker, S. M., Marchionini, G., Hearst, M., Schraefel, M. C. Exploratory search and HCI: designing and evaluating interfaces to support exploratory search interaction. *Ext. Abstracts CHI 2007*, ACM Press (2007), 2877-2880.
18. White, R.W., Dumais, S. T., Teevan, J. Characterizing the influence of domain expertise on web search behavior. In *Proc. WSDM 2009*, ACM Press (2009), 132-141.
19. Zhang, X. M., Anghelescu, H. G. B., Yuan, X.J. Domain knowledge, search behaviour, and search effectiveness of engineering and science students: an exploratory study. *Information Research 10*, 2 (2005).