Where Do the Query Terms Come from? An Analysis of Query Reformulation in Collaborative Web Search

Zhen Yue
University of Pittsburgh
135 North Bellefield Ave.
Pittsburgh, PA, 15260
Zhy18@pitt.edu

Jiepu Jiang
University of Pittsburgh
135 North Bellefield Ave.
Pittsburgh, PA, 15260
jij29@pitt.edu

Shuguang Han, Daqing He
University of Pittsburgh
135 North Bellefield Ave.
Pittsburgh, PA, 15260
{shh69,dah44}@pitt.edu

ABSTRACT
This paper presents a user study aiming to investigate the query reformulation in collaborative Web search. 7 pairs of participants were recruited and each pair worked as a team on two collaborative exploratory Web search tasks. Through the log analysis, we compared possible sources for participants to draw query terms from. The results show that both search and collaborative actions are possible resources for new query terms. Traditional resources for query expansion such as previous search histories and relevant documents are still important resources for new query terms. The content in chat and workspace generated by participants themselves seems more likely to be the resource for new query terms than that of their partners. Task types also affect the influences on query reformulations. For the academic task, previously saved relevance documents are the most important resources for new query terms while chat histories are the most important resources for the leisure task.

Categories and Subject Descriptors
H.3 INFORMATION STORAGE AND RETRIVAL H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces – Collaborative computing, Computer-supported cooperative work

General Terms
Experimentation, Human Factors

Keywords
Collaborative web search; Query reformulation

1. INTRODUCTION
When the information needs are complex and explorative, it may be in the searchers' best interests to collaboratively explore the information space and participate in shared learning [9]. For instance, healthcare providers may collaboratively search for information to diagnose a patient's illness, or family members may collaboratively search the Web to buy a car. Therefore, although information seeking has been traditionally studied as an individual search activity, collaborative information seeking behavior has attracted much attention [2].

Complex and explorative Web searches often involve iterative interaction with retrieval systems, so that query formulation and reformulation have been important topics in individual explorative searches, which cover issues like patterns of query reformulation [4] and the reliable sources for query expansion [5]. However, as a much more complex form of explorative search, collaborative web searches have seldom been the focus for query reformulation. A few studies on query reformulation in collaborative information retrieval have been based on implicit collaborative search where information from a community of like-minded searchers is shared without explicitly modeling the collaboration [1]. Previous learned queries and relevant documents are reused in new and similar search sessions to improve the overall retrieval quality [3]. Often simulated experiments rather than user studies involving human subjects are employed in these studies. In this paper, we focus on examining query reformulation behaviors of a group of users involved in explicit collaborative exploratory web searches. Previous research pointed out that successful assistance to query-reformulation must be designed based on the understanding of users' query behavior [4]. Therefore, it is important to study what are the possible resources for users in an explicit collaborative search to generate query terms for query reformulation. Particularly, we are interested in the following research questions:

RQ1: Users in collaborative web searches perform unique collaborative actions; do these actions influence users' query reformulations?

RQ2: Users in collaborative searches still perform several search actions observed in individual searches; do these actions still influence users' query reformulations? If they do, how do their influences differ from collaborative search actions?

RQ3: How do task differences affect the above influences on users' query reformulation?

2. EXPERIMENT DESIGN
Our study was designed as a set of control experiments with human participants using CollabSearch, a collaborative search system developed by the authors.

2.1 CollabSearch: a Collaborative Search System
CollabSearch [1] is a Web search system for group users. The system has both search and collaboration features. CollabSearch's interface contains three frames: topic statement, search and team workspace. The topic statement frame shows the task description on which the user is currently working. Team members can also

CIKM'12, October 29– November 2, 2012, City, State, Country.

[1] http:// crystal.sis.pitt.edu:8080/CollaborativeSearch/

post their comments below the task description. The search frame connects the user's query to Google and displays the Google search results. Users can also see their search histories as well as those of their teammates. Users examine search results for relevant information, and can save a whole Web page or a snippet of the page. All the saved web pages and snippets, collected by the user and the teammate, are stored in the team workspace frame. A notice is displayed at the top when new items are saved to the team workspace. Users can click to view more details of an item in the workspace or comment on any item.

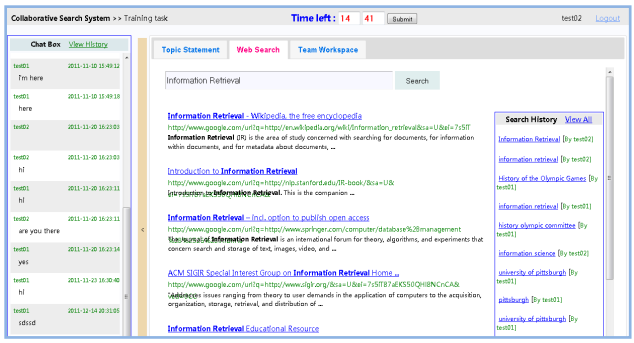

Figure 1: The web search frame of CollabSearch

2.2 Participants

14 participants (7 pairs) were recruited from the University of Pittsburgh for this study. 10 of these participants are male and 4 are female. All of them are experienced searchers. All the participants signed up as pairs and the members of each pair knew each other before the study so that it was easy for them to form a team. Participants in the same team worked on the same task simultaneously. As we were trying to simulate remotely-located collaboration, the participants in the same team could communicate with each other by sending instant text messages or reading each other's search histories and the collected results shared in team workspace but no face-to-face communication was allowed.

2.3 Search Tasks

Two exploratory web search tasks were used in this study. Both of them had been used in other collaborative web search studies [7, 8], so their validity for collaborative search has been examined before. One task (T1) is related to academic work, which asks participants to collect information for a report on the effect of social networking service and software [8]. The other task (T2), which is about leisure activities, asks participants to collect information for planning a trip to Helsinki [7]. Morris' [6] identified that travel planning and academic literature search are two common collaborative search tasks. Therefore, both tasks here are representative in studying collaborative web search. The task description carefully states the kind of information that the participants need to collect and the goal is to collect as many relevant snippets as possible.

2.4 Experiment Procedure

The experiment procedure was: each team worked on both tasks. The order of the two tasks was rotated to avoid the learning and fatigue effect. During the experiment, after being introduced to the study and the system, and filling out an entry questionnaire to establish their search background, these participants worked on a training task to get familiar with the system for 10 minutes. Then they worked on task 1 or task 2, depending on the task order assigned for each team. They had 30 minutes for each task. At the end of each task, each of them also worked on a post-search questionnaire collecting information about their satisfaction with the search results. Before the end of the experiment, participants were asked several open-ended questions for their experience with both tasks.

3. DATA ANALYSIS METHODS

3.1 Categorizing user search actions

In order to identify the possible sources for participants to draw query terms from, we first need to analyze what actions the participants have taken during the whole exploratory Web search. Typical actions recognized in this study and listed in Table 1 include Query, View, Collect, Workspace, Topic and Chat. All these actions were categorized and mapped from the transaction logs recorded in CollabSearch system.

Table 1: User actions

Actions	Descriptions
Query (Q)	A user issues a query or clicks a query from search history.
View (V)	A user clicks a result in the returned result list
Save (S)	A user saves a snippet or bookmarks a webpage
Workspace (W)	A user clicks or edits or comments an item saved in the workspace
Topic (T)	A user clicks the topic statement or leaves comments
Chat (C)	A user sends an message or views the chat history

3.2 Log analysis

Given a previous search query q_{n-1} and its follow-up reformulation q_n, we are interested in: 1) what are the new terms in q_n compared with q_{n-1}; 2) what are the possible sources that influence the uses of the new terms? We count new query terms as those appearing in q_n but not in q_{n-1}. For example, for "Helsinki Finland" and its follow-up reformulation "Helsinki Finland art", we count "art" as a new term. Then, we set up a windows size k and evaluate the influence of a type of action (A) on a new query term in q_n (X) via a Bayesian approach: assuming X in q_n comes from the contexts of an action before q_n; given the observed X, what is the probability that X comes from the contexts of a specific action A, i.e. P(A|X)? Because $P(A|X) \propto P(X, A)$, we can further calculate P(A|X) as its equivalent form P(X, A), i.e. the probability that we can find the type of action A that includes the new query term X within the range of k actions prior to q_n. For example: considering five actions (k=5) prior to q_n, of which 3 actions are viewing search results (V) and 2 of the viewed results contains X, P(X, V) equals to 2/5. Admittedly, observing X in the viewed results (V) does not necessarily mean a causal relationship. However, we can at least adopt the measure as an indicator for the chances that such causal relations exist.

Specifically, we can explain P(X, A) for each type of action A as follows:

P(X, T): the probability that the participant has viewed the topic statement within the range of k actions and we can also find the new query terms in the topic statement.

P(X, V) and P(X, S): the probability that the participant has viewed/saved some results within the range of k actions and we can find the new query term from any of the results (no matter in the titles or snippets).

P(X, W) and **P(X, WP)**: the probability that the participant has viewed the workspace within the range of k actions and we can find the new query terms in articles from the workspaces of the participant (W) and his/her partner (WP).

P(X, Q) and **P(X, QP)**: the probability of finding the new query terms in queries of the participant and his/her partner within the range of k actions. The new terms in q_n obviously do not exist in q_{n-1}. However, within the range of k actions if more than one query had been issued, say q_{n-1} and q_{n-2}, there is a chance we can find the newly added term in q_{n-2}. For this reason, we consider P(X, Q). We here calculate P(X, QP) because one participant's queries will be displayed to the partner, which may also help participants reformulate queries.

P(X, C) and **P(X, CP)**: the probability of finding the new query terms in chat histories between two team members within the range of k actions. We use C and CP for the part of chat histories said by the participant and the partner respectively.

When calculating whether we can find a query term from a chunk of text (e.g. it is the contexts associated with an action), we remove stop words (the 419 stop words in Indri system) in both queries and the texts, and stem texts using Krovertz stemmer.

4. RESULT ANALYSIS AND DISCUSSION

4.1 Actions on Search

In this section, we report results on those actions that aim for search only and are not closely related to the collaboration between two team members. They are T, Q, V, S and W, which can also be found in individual searches. Figure 2 shows the probabilities of finding new query terms in each type of actions under two different task settings.

The left part of Figure 2 shows results of T1 (academic search). We observed the highest probability of finding new query terms from saved documents (S) and viewed documents (V). Both are highly indicative of the participant's judgments of relevant documents. Thus, it is not surprising that the relevant contexts associated with the two types of actions (S and V) can suggest useful terms in the search topic for the participants. Comparatively, we found lower probability of finding new terms in W than S. This is also reasonable considering the fact that W includes information of all historically found documents, but the focus of the participants may evolve during search. Thus, it is not surprising that the recently saved articles (S) may better suggest new query terms to the user. Participants may also get indications of new query terms by checking the topic statement (T), but the likelihood is lower than that from S and V. Such results are also consistent with the nature of task 1 (academic search), in which relevant documents are mostly related to several topics and thus the topic-related relevance feedback is likely to work.

In comparison, we observed differences in results of T2, as shown in the right part of Figure 2. Comparatively, S and V are not the most possible sources for the new query terms. This may indicate that the nature and difficulty of these two tasks are very different. Documents saved in workspace may still be indicative of participants' relevant judgments. However, in T2, topical relatedness may not serve as the main criteria for a document's relevance. For example, when planning a trip, the homepage of Hilton hotels may provide topically relevant results, but participants may not select it for some other reasons. Thus, it is not surprising that the probability of finding new terms from S is comparatively lower than that of from many of the other actions, e.g. C, T, and QP (we will further discuss the indication of these actions in the following sections). V has very low probability of observing the new query terms, which may again indicate the differences in nature and difficulty between T1 and T2. For example, it may suggest the T2 is more difficult in nature so that the majority of webpages checked by participants are irrelevant. Besides, it may also indicate that in such exploratory tasks, it is fairly difficult for participants to just view results' snippets to correctly evaluate the usefulness of a webpage. Instead, the participants may need to frequently check the whole page.

In both tasks, we found very clear trends that, although within a short range (k<5) the user is unlikely to reuse terms in previous queries, it is still fairly common that the participants may get back to adopt previously used query terms within proper range of actions (k>=5). Such results are also consistent with previous studies, e.g. participants' previous search histories are found to be reliable resources for relevance feedback and can consistently improve search results [5].

In conclusion, there is no surprising difference between our results of collaborative search and previous studies of individual search in terms of users' search actions. Task difference is still a dominating factor in users' search behaviors. Existing techniques for individual searches (e.g. relevance feedback by query history and click through) could still help users' collaborative searches.

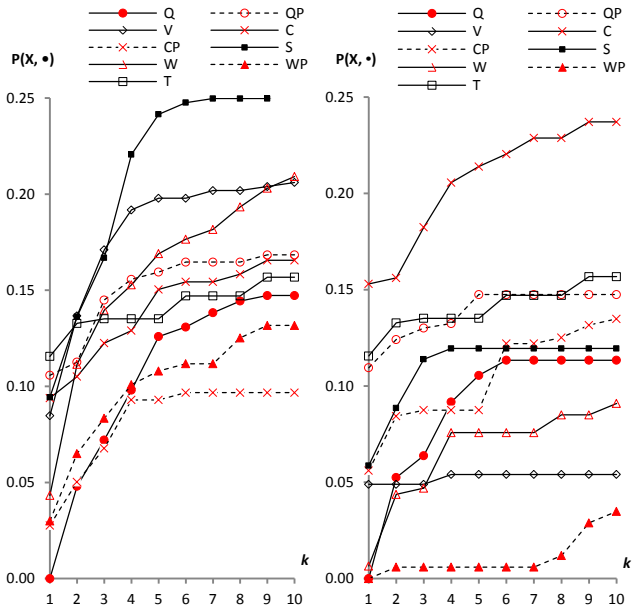

Figure 2. Probability that new query terms exist in each type of action resource (left: T1 and right: T2)

4.2 Actions on Collaboration

In collaborative web searches, participants not only need to perform search related actions, but also rely on teamwork to collaborate. The collaboration-related actions can also help to generate new query terms. There are two types of communication between teammates: the explicit communication – chat, and the implicit communication, from which teammates can check each other's search history or the results saved in the shared team workspace.

In Figure 2, it is obvious C is the most important resource for new query terms in task T2. This indicates that team members often discuss what to search through chatting. The probabilities that C and CP are the resources for new query terms are not that high in T1. This might occur because team members explicitly communicate with each other more often in T2 than in T1. In terms of the comparison between C and CP as the resource for new query terms, we selected the data for k=5 and performed

significance tests. The reason that we selected k= 5 is that the trend become stable after k>=5. As shown in Table 2, the probability that C is the source for new query terms is significantly higher than that of CP. This suggests that the participants tend to tell their team member what they want to search rather than suggesting their team members what to search.

As the search history of both team members are shown in the search frame, it is a possible resource for new query terms. Q as a resource for new query terms indicates that the participants reused the query terms they had used before. QP being a resource for new query terms suggests that the participants got ideas of new query terms from his/her partner's search history. The results in Table 2 show that the probability of QP being the resource for new query terms is slightly higher than that of P in both tasks. However, there is no significant difference between Q and QP. This may suggest that the query terms in both the participant's and their partner's search histories are equally important for future use. To a certain extent, this evidenced that the participants adopted their partner' search vocabularies for their own searches.

Table 2: Probability that new query terms exist in C, CP, Q, QP, W and WP

		Mean (SD)		Statistical Test
		Self	Partner	
Chat (k=5)	**T1**	0.150 (0.160)	0.093 (0.139)	P=.169
	T2	0.214 (0.179)	0.087 (0.096)	**P=.025**
Query (k=5)	**T1**	0.126 (0.115)	0.159 (0.110)	P=.424
	T2	0.106 (0.139)	0.147 (0.143)	P=.333
Workspace (k=5)	**T1**	0.169 (0.136)	0.108 (0.108)	**P=.038**
	T2	0.076 (0.080)	0.006 (0.022)	**P=.018**

Shared team workspace is another important way for implicit communication between team members. Figure 2 shows that W and WP are important resources for new query terms in task T1. This is understandable as the workspace contains the previously saved results. For T1, the topical relevance is the major criterion for finding relevant documents, which means that previous results are important resources for new query terms. W and WP are less likely to be sources for new query terms according to our results. This indicates that T2 is more difficult since participants were unable to generate many new query terms from the previously found documents. By comparing W and WP, we can see that the probability of W being the source for new query terms is significantly higher than that of WP. In other words, the participants are more likely to generate new query terms from their own saved results. This may indicate that team members divide the search topic into sub-topics and each of them is in charge of part of the sub-topics.

4.3 Comparison of Actions on Search and Actions on Collaboration

The focus of this section is on further comparison between the influences of search actions and that of collaborative actions on the participants' query reformulation. In individual searches, the previously clicked documents can be seen as relevant documents, which are important resources for query reformulation. Therefore, we selected V, S, and W actions, which represent the types of resources that users can encounter in individual searches. CP, WP and QP represent the type of resources that users can meet only in collaborative searches from their partners. We examined the joint influence of search (V+S+W) vs. collaborative actions (CP+WP+QP) by the probability of finding new query term X from any of the search or collaborative actions within the range of *k* actions. Note that the joint influence of multiple actions is not the sum of P(X, A) for those actions.

As shown in Table 2, the combination of V, S and W is less likely to be the source for new query terms compared with the combination of CP, WP and QP in T1. For academic search tasks, although each of the three actions (i.e. V, S, W) are very important sources for query terms separately, a combination does not increase the probability of finding new query terms remarkably. This may indicate that V, S and W seem to be homogenous resources for users to obtain new query terms. By homogenous we mean these resources are likely to be the sources for the same query terms. On the other hand, WP, CP and QP seem to be heterogeneous resources for new query terms. Therefore, combining them together significantly increases the probability of finding new query terms. This reveals a benefit of collaborative search, i.e. helping participants generate diverse ideas for query terms that may not exist in previously seen documents.

Table 3: Comparison of search and collaborative actions

	Mean (SD)		Statistical Test
	Search	Collaborative	
T1	0.113 (0.139)	0.293 (0.172)	**P=.013**
T2	0.250 (0.164)	0.242 (0.159)	P=.929

5. CONCLUSION

Both explicit (Chat) and implicit (Workspace) communication in collaborative search can influence users' query reformulations. However, the content generated by users themselves seems more likely to be the resource for new query terms than that of their partners. Traditional resources for query expansion such as previous search histories and relevant documents are still important resources for new query terms. For the academic task, the previously seen documents either viewed or saved in workspace are likely to be homogenous resources for new query terms, while information obtained from the partner tending to provide heterogeneous resources. In the academic task, the previously saved documents are the most important resource while chat is the most important resources for new query term in the leisure task. Further studies are needed to confirm these findings, especially whether the resources provide aspect/facet for query terms or exact keywords for search.

6. REFERENCES

[1] Balfe, E. & Smyth, B. (2005). A comparative Analysis of Query Similarity Metrics for Community-Based Web Search. *ICCBR* 2005, 63-67.

[2] Hansen, P. & Jarvelin, K. (2005). Collaborative Information Retrieval in an information-intensive domain. *Information Processing & Management*, 41(5), 1101-1119.

[3] Hust, A., Klink, S. Junker, M., & Dengel, A. (2002). Query Reformulation in Collaborative Information Retrieval. Proceedings of the International conference on Information and Knowledge Sharing.

[4] Janse, B. J., Booth, D. L. & Spink, A.(2009). Patterns of Query Reformulation During Web Searching. *Journal of American Society of Information Science and Technology*, 60(7), 1358-1371.

[5] Jiang, J., Han, S., Wu, J., and He, D. (2011). PITT at TREC 2011 session track. *In Proc. of TREC 2011*.

[6] Morris, M. R. (2008). A survey of collaborative web search practices. CHI 2008.

[7] Paul, S. (2010). Understanding together: Sense making in collaborative information seeking. Doctoral Dissertation, Pennsylvania State University.

[8] Shah, C, and Marchionini, Gary (2010). Awareness in Collaborative Information Seeking. *Journal of American Society of Information Science and Technology*, 61(10), 1970-1986.

[9] White, R. W. & Roth, R. A. (2009). Exploratory search: Beyond the Query-Response Paradigm. Synthesis Lectures on Information Concepts, Retrieval, and Services, 1(1), 1-9.