An Investigation of Search Processes in Collaborative Exploratory Web Search

Zhen Yue  
School of Information Sciences  
University of Pittsburgh  
135 North Bellefield Ave.,  
Pittsburgh, PA  
zyh18@pitt.edu

Shuguang Han  
School of Information Sciences  
University of Pittsburgh  
135 North Bellefield Ave.,  
Pittsburgh, PA  
shh69@pitt.edu

Daqing He  
School of Information Sciences  
University of Pittsburgh  
135 North Bellefield Ave.,  
Pittsburgh, PA  
dah44@pitt.edu

ABSTRACT
This paper presents a user study aiming to investigate search processes in collaborative exploratory Web search. Our analysis of search processes focuses on the distribution and transition of user search actions captured in a collaborative web search system called CollabSearch. The results show that a large proportion of users’ actions in collaborative searches were related to explicit communication, which is one of the sources for users to obtain query ideas. This paper concludes with some insights on the range of behaviors and activities that a collaborative search system should support.

Keywords
Collaborative web search, search process, exploratory search, collaborative information behavior

INTRODUCTION
When the search task is exploratory, it may be in the searchers’ best interest to collaboratively explore the information space and participate in shared learning (White & Roth, 2009). A collaborative search system need not only support the interaction between a user and the system, but also support the interaction among users. A successful collaborative search system relies on good understanding of the group activities involved in the search process. Studies which seek to describe collaborative search process can help developers understand the range of behaviors and activities that systems need to accommodate. Therefore, understanding the various manifestations of collaborative information behavior involved in a search process is crucial for designing and evaluating systems supporting collaborative information seeking.

Collaborative information seeking has been studied in various environments including both organizational and Web setting (Hansen & Järvelin, 2005; Evans & Chi, 2008). Our study focuses on collaborative exploratory search in the Web search environment.

In individual information seeking, researchers explored many methods to investigate a single user’s search process. Models of the individual search process are well established such as Kuhlthau’s (1991) model. Shah and Gonzalez-Ibáñez (2010) attempted to map Kuhlthau’s ISP model to collaborative information seeking. Through a laboratory study with 42 pairs of participants, they investigated similarities and disparities between individual and collaborative information seeking processes. In the conclusion, they declared that social elements are missing when applying the ISP model in a collaborative setting. However, current investigations of collaborative search processes are limited to exploring the application of an individual search process model in a collaborative setting.

From the discussion above, it is clear that investigating the collaborative search process is crucial for designing and evaluating systems that support collaborative information seeking. Modeling the search process remains a challenge in exploratory search. It is even more challenging to model group activities involved in the collaborative exploratory search process.

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

CollabSearch: a Collaborative Search System
CollabSearch 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 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

1 http://crystal.exp.sis.pitt.edu:8080/CollaborativeSearch/
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

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 know each other before the study so that it was natural 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.

Search Tasks
Two exploratory web search tasks were used in this study. Both of them had been used in other collaborative web search studies (Shah, 2010; Paul, 2010) so their validity for collaborative search has been examined before. One task is related to academic work, which asks participants to collect information for a report on the effect of social networking service and software (Shah, 2010). The other task, which is about leisure activities, asks participants to collect information for planning a trip to Helsinki (Paul, 2010). Morris (2008) identified that travel planning and academic literature search are two common collaborative search tasks. Therefore, both tasks here are representative for studying collaborative web search. The task description carefully states the kind of information that the participants need to collect and the goal to collect as many relevant snippets as possible.

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.

DATA ANALYSIS METHODS
Categorizing user search actions
In terms of search process analysis, we are interested in what kind of actions the participants have taken during the whole process of exploratory Web search. Typical actions recognized in this study include Query, View, Collect, Workspace, Topic and Chat, of which the details are listed in Table 1. All these actions were categorized and mapped from the transaction logs recorded in CollabSearch system.

<table>
<thead>
<tr>
<th>Actions</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query (Q)</td>
<td>A user issues a query or clicks a query from search history.</td>
</tr>
<tr>
<td>View (V)</td>
<td>A user click a result in the returned result list</td>
</tr>
<tr>
<td>Collect (C)</td>
<td>A user collects a snippet or bookmarks a webpage</td>
</tr>
<tr>
<td>Workspace (W)</td>
<td>A user clicks, edits or comments an item saved in the workspace</td>
</tr>
<tr>
<td>Topic (T)</td>
<td>A user clicks the topic statement for view or leaves comments</td>
</tr>
<tr>
<td>Chat (CH)</td>
<td>A user sends an message to the other user or views the chat history</td>
</tr>
</tbody>
</table>

Temporal analysis of search actions
The study of the temporal distribution was applied throughout the whole search process in order to understand when participants perform certain search actions. We first divided the entire search session for each task into four segments with equal time, and then calculated the frequency distribution within each time segment. The first time segment can be viewed as the beginning phase of the search process, and the last time segment represents the ending phase. The middle phase of the search process is further divided into two segments, so that the lengths of the time segment are roughly the same. Given that participants had 30 minutes to complete a search task, each time segment is 7.5 minutes long.

Transition analysis of search actions
The above two methods focus on describing search action alone. We also examined the relationship between different search actions, which we called the transitions of search actions.

In the transition analysis of search actions, we consider the sequential dependence order of user actions. Each search action has one predecessor action and one successor action.
Since all search actions are categorized into six different types, there are total 36 possible action transition pairs, such as from Query to View (Q→V), from View to Collect (V→C), and so on. Our study here focused on two aspects. First, we analyzed the percentage of each of the 36 action transition pairs in order to find out the most frequent action transition pairs. Second, we conducted pre-action analysis for Chat and Query. Pre-action analysis is defined as, for a given type of search action, analyzing the percentage distribution of its predecessor actions. The reason we conducted pre-action analysis for Chat is because we want to see what triggered participants to explicitly communicate with each other, and a similar analysis for Query is intended to find out possible sources for generating queries.

RESULTS

Temporal distribution of search actions
The temporal distribution of search actions provides us with an overall sense of when participants conduct what actions. Figure 4 using area chart present the average percentages of each search action per participant within the four time segments of the entire search session for each task.

It can be seen that for both tasks, team members constantly chat with each other throughout the whole search session. In the academic task, the percentage of Chat action fluctuates through the four time segments. It might because team members first communicate to come up with a search plan or strategy and then focus on their own sub-task. But later, after they have some results, they feel the need to communicate with each other again. Another difference between the two tasks is the temporal distribution of Workspace action. The percentage of Workspace actions is relative constant in the leisure task, whereas its percentage in the academic task is higher in the beginning than in the ending phase. This might indicate that team members in the academic task tended to utilize the “Team Workspace” to discuss final results whereas members in the leisure task preferred to chat with each other to make final agreements.

Another important message is that participants tended to issue relatively more queries and collect more results in the first half of the search session than in the second half. In the literature, such as Kuhlthau’s model, formulation and collection are in the second half of the six stages in information seeking process. This might indicate that in collaborative search, participants start collecting documents sooner.

Transition of search actions
The results of transition analysis, including frequency distribution of action pairs, pre-Chat analysis and pre-query analysis are presented in this section.

Distribution of search action pairs
Since we have 6 types of search actions, in total there are 36 possible action transition pairs. In Table 2, we listed the top 5 frequent action (larger than 10%) transition pairs. Since there is not much difference between the two tasks on action transition pair distribution, here we didn’t distinguish the tasks. It can be seen that top two frequent actions pairs are CH-CH and W-W. CH-CH indicates constant chatting activities and W-W represents continuous actions in “Team workspace”, both of which are not directly related to search but actions that support the collaboration. The next three most frequent action transition pairs include V-C, Q-V and V-V. They represent typical search behavior patterns. For example, after issuing a query, viewing the results is very likely to happen. And participants may continue viewing several results on the result page. After viewing a result that is relevant, the participant would collect that result.

Table 2: Most frequent action pairs for both academic task and leisure task

<table>
<thead>
<tr>
<th>Rank</th>
<th>Action (Frequency %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CH-CH (42.29%)</td>
</tr>
<tr>
<td>2</td>
<td>W-W (15.57%)</td>
</tr>
<tr>
<td>3</td>
<td>V-C (14.57%)</td>
</tr>
<tr>
<td>4</td>
<td>Q-V (14.36%)</td>
</tr>
<tr>
<td>5</td>
<td>V-V (14.21%)</td>
</tr>
</tbody>
</table>

Pre-Chat analysis
Figure 2 illustrates pre-Chat action distribution. We labeled the transition probability from each type of action to Chat on each link. It is clear that the most common action before Chat is Chat itself, which suggests the cyclical nature of Chat actions. The second common action before Chat is Topic. The reason for this relationship might be after viewing the topic statement, the participants tend to discuss with each other on the task requirements and allocate sub-tasks. The next possible action before Chat is Workspace. This suggest that after viewing, editing or commenting the items saved in the “Team workspace”, the participants also need to explicitly communicate with each other to inform or discuss the updates and changes in the shared “Team Workspace”.

Figure 2: Pre-Chat analyses for both academic task and leisure task

Pre-query analysis
Figure 3 visualizes the proportion of predecessors of Query action. The visualized result shows the most common action before Query is Chat. This indicates that the explicit communication in collaborative search helps the
participants to generate query ideas. The second common action before query is Topic. It is easy to understand that the participants need to check the task requirements before issuing a query. Another interesting finding is that the probability of transition from Query to Query is very low. This may suggest that in collaborative search, participants issued good queries that did not need reformulation.

![Diagram](image)

**Figure 3: Pre-query analysis for both academic task and leisure task**

**DISCUSSION AND CONCLUSION**

In this paper, we report a study examining the search processes in collaborative exploratory Web search. The participants worked in pairs on two exploratory Web search tasks using our CollabSearch system. We found that there are some differences on the temporal distributions of user actions in two tasks. We also found that participants tended to issue more queries and collect more results in the first half of session than in the second half, which is different from the findings in individual search scenarios. Through the analysis of action transition pairs, we found that actions related to collaboration are more frequent than actions related to search. The pre-Chat analysis revealed that the reasons that trigger Chat might include needs for discussing task requirement and item collected. Through the pre-query analysis, we found that Chat might be the source for participants to generate query ideas. This study provides some insights for designers on the range of behaviors and activities that a collaborative search system should support. Further studies are needed to fully understand them.

**ACKNOWLEDGMENTS**

This work was partially supported by the National Science Foundation under Grant No. 0704628 and IIS-1052773.

**REFERENCES**


![Graph](image)

**Figure 4: Temporal distribution of actions in SYN for academic task (left) and leisure task (right)**