Examining Collaborative Query Reformulation: A Case of Travel Information Searching

Abu Shamim Mohammad Arif, Jia Tina Du, Ivan Lee
School of Information Technology and Mathematical Sciences
University of South Australia, South Australia, Australia
shamimarif@yahoo.com, tina.du@unisa.edu.au, ivan.lee@unisa.edu.au

ABSTRACT
Users often reformulate or modify their queries when they engage in searching information particularly when the search task is complex and exploratory. This paper investigates query reformulation behavior in collaborative tourism information searching on the Web. A user study was conducted with 17 pairs of participants and each pair worked as a team collaboratively on an exploratory travel search task in two scenarios. We analyzed users' collaborative query (CQ) reformulation behavior in two dimensions: firstly, CQ reformulation strategies; and secondly, the effect of individual queries and chat logs on CQ reformulation. The findings show that individual queries and chat logs were two major sources of query terms in CQ reformulation. The statistical results demonstrate the significant effect of individual queries on CQ reformulation. We also found that five operations were performed to reformulate the CQs, namely: addition, modification, reordering, addition and modification, and addition and reordering. These findings have implications for the design of query suggestions that could be offered to users during searches using collaborative search tools.

Categories and Subject Descriptors
H.3.3 [Information Search and Retrieval]: Search process, Query formulation

General Terms
Experimentation, Measurement

Keywords
Collaborative search, Interactive IR, Query reformulation

1. INTRODUCTION
Web searching is a process of querying and reformulating queries to satisfy certain information needs. Web searchers frequently modify their queries to obtain better results and this process is referred as query reformulation [3]. Typical searchers have little tolerance for viewing low-ranked search results and they prefer to reformulate the query rather than see the complete result lists [3]. Previous studies have shown that 52% of users reformulate their queries [2]. When information needs become complex and exploratory, it might be in searchers' best interests to collaboratively explore the information space and participate in shared learning [10]. For instance, healthcare providers tend to collaboratively search for information to diagnose a patient's illness [9]. Therefore, although information seeking has been traditionally studied as an individual search activity, collaborative information seeking and query reformulation have attracted much attention in recent years [1, 9, 11].

Complex and exploratory Web searches often involve iterative interactions with retrieval systems. Query formulation and reformulation are important topics not only in individual exploratory searches but in collaborative searches which may cover issues such as patterns of query reformulation [3] and the reliable sources for query expansion. Understanding query reformulation behavior and being able to accurately identify reformulation queries have several benefits. One of these benefits is learning from user behavior to better suggest automatic query refinements or query alterations. Another benefit is that, if we are able to accurately identify query reformulations, we will be in a better position to evaluate the satisfaction of users with query results. Identifying query terms for query reformulation can be very useful in cases where the retrieved results are irrelevant to users' information needs. Successful assistance to query reformulation must be designed based on the understanding of users' query behavior [3]. However, as a much more complex form of exploratory search, collaborative web search has seldom been the focus of query reformulation research.

Previously learned queries and relevant documents have been reused in new and similar search sessions to improve the overall retrieval quality in collaborative information retrieval (IR) [1]. Reddy et al. [5] investigated the role of communication in collaborative information searching, from the perspective that the communication is necessary to establish common grounds among members of the team. They found that users communicated their plans, thoughts, and search queries through the chat functionality of the retrieval system while seeking information. Furthermore the communication (i.e. chat) function played a critical role in query reformulation and in enhancing search results by enabling the team members to share common views. A collaborative querying system has been sketched [4] to show how communication and prior queries help collaborators with query formulation and reformulation. In [11], the authors identified that actions, such as querying, saving, viewing, and chatting, seem more likely to be the sources for new query terms in collaborative searches. They observed that despite the dominance of traditional resources such as previous search histories and relevant documents in query reformulation, chat log is an important resource for new query terms in the leisure (travel) task. Yet little is known about the impact of individual queries and chat logs generated by team

SIGIR'14, July 6-11, 2014, Gold Coast, QLD, Australia.

http://dx.doi.org/10.1145/2600428.2609463

members on collaborative query (CQ) reformulation in the collaborative search process. In this study, we are particularly interested in addressing the following two research questions:

RQ1: What are the strategies for CQ reformulation?
RQ2: What is the interplay between individual querying, chatting, and CQ reformulation in collaborative Web searching?

2. EXPERIMENTAL DESIGN

Based on the results of previous studies [6, 7] which examined the characteristics and strategies of tourists' collaborative search behavior, we designed a collaborative information search interface - Collaborative Tourism Information Search (ColTIS) - which supports the searching for travel-related information in both standalone (a single user) mode and collaborative mode (multiple users).

ColTIS allows various numbers of users to perform collaborative searching. It provides not only the features of stand-alone query formulation (query composed by a single user), but also the CQ formulation (query composed by more than one user). Of the multiple windows in the interface, three relevant windows are illustrated (Figure 1) and described in the present paper. Window 1 is a standalone search window and Window 2 is used for collaborative searching. The query terms are forwarded to the Google search engine and the search results are then displayed. A Chat window (Window 3) facilitates real-time communication between multiple users while searching for information. ColTIS supports the function of text-based messaging among multiple users.

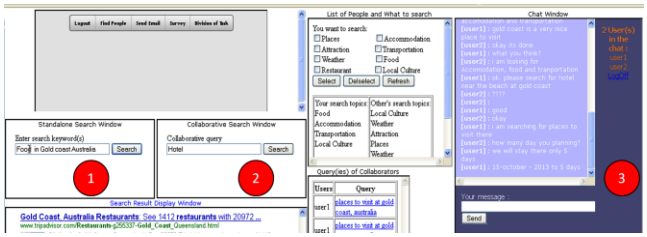

Figure 1. Collaborative tourism information searching (ColTIS) interface.

2.1 Participants

A total of 34 staff and postgraduate students (17 males and 17 females) were recruited from a large university in Australia to participate in the study. All of them had been searching the Web for an average of 11.5 years. To facilitate collaborative search, all the participants signed up as pairs (17 pairs in total) and the members of each pair knew each other and had co-working experience: they either did collaborative work or traveled together before the study. We simulated remotely-located collaboration and the participants in the team could communicate with each other by sending instant text messages.

2.2 Search Task

A tourism-related exploratory Web search task with two different scenarios was used in this study: in the first scenario, pairs of participants were asked to search with ColTIS; and in the second scenario, they were asked to conduct the same search using Tripadvisor.com with Google Talk as the means of communication. Searching on travel planning is believed to be a common collaborative search task and similar types of exploratory search tasks had been used in other collaborative Web search studies [8]. Each team was asked to look for travel information for preparing a travel plan including destination, accommodation, attraction, transport, food and so on. The participants were asked to collect as much relevant information as possible for planning the trip.

2.3 Procedure

The study was conducted in a computer laboratory in the university. As we aimed to simulate remotely-located (different places) synchronous (real-time) collaboration, the participants were placed in different places separated by a partition in the laboratory so that they could not talk to each other directly or see what the other person was doing. The team members could only communicate with each other by sending instant text messages or reading each other's search histories. Each pair of participants was firstly introduced to the study and both systems (i.e. ColTIS and Tripadvisor.com). The participants in each pair then filled out an entry questionnaire about their educational backgrounds and experiences on Web searching, traveling, and collaborative work. After that, they were asked to work together on the exploratory search task which took around 30 minutes on average for each search scenario. At the end of the search, each participant completed a post-search questionnaire collecting information about their satisfaction with the collaborative search systems. Both the search logs and chat logs were stored in the database for further analysis.

2.4 Data Analysis

The objective of collaborative IR is to obtain relevant information to the solution of a common goal. In such a process, forming queries is a very important activity to meet every party's interest. This paper focuses only on the analysis of CQ reformulation behavior involved in the collaborative search process. The data for analysis included the search logs including individual queries and CQs and users' chat logs. Individual queries refer to those queries composed by individual users using the standalone query window, while CQs refer to the queries composed by both participants in a pair using the collaborative query window. The data were analyzed in two dimensions: i) finding out CQ reformulation patterns and ii) identifying the contribution of individual queries and chat logs to the CQ reformulation. We collected a total of 210 tourism-related individual queries and 140 CQs. Among 507 chat log entries, 486 were considered to be valid and the rest were either blank or meaningless. The summary of the data used in this paper are shown in Table 1.

Table 1. Summary of data

Type of Data	No.
Total valid individual queries	210
Total valid CQs	140
Total valid chat entries	486

3. RESULTS AND DISCUSSION

3.1 Collaborative Querying Architecture

The CQ formulation architecture (Figure 2) was drawn to illustrate how a CQ is initiated, reformulated and executed. The figure shows that a CQ is initiated by any of the collaborating team members using Window 2 (in Figure 1) which is visible to all members. The initiator or any other team member can add any term(s) to the CQ, edit any term(s) of the CQ, or delete any term(s) from this CQ. After the end of the modification, the CQ is

executed to retrieve the results which are similar in function to traditional IR systems. If the results are not satisfactory, then the users refine the CQ. This modification of the CQ is repeated until users are satisfied with it or the retrieved results. This satisfaction is defined as the agreement on CQ or on results retrieved by the collaborators.

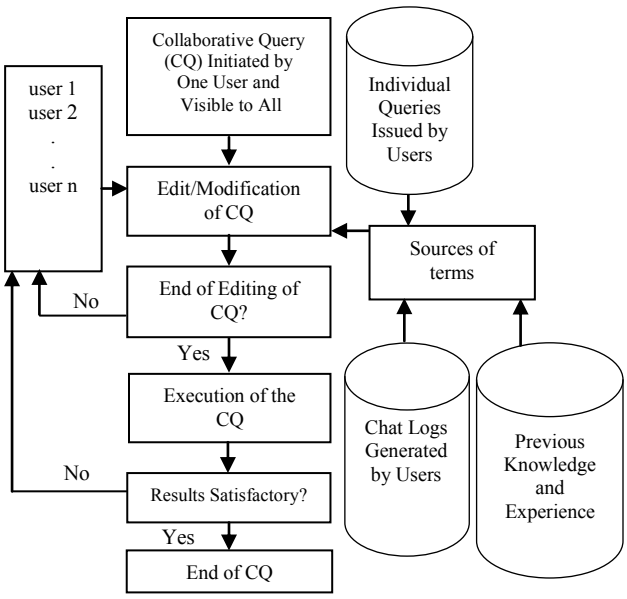

Figure 2. Flow of collaborative query reformulation.

3.2 Collaborative Query Reformulation Strategy

In the process of formulating CQ, an obvious question may emerge: which terms should be selected in the query formulation and how does the CQ reformulation evolve? Our analysis shows that, three main sources were employed by the team for the selection of terms earmarked for 140 CQ formulation, including individual queries (51%), chat logs (26%), and previous knowledge and experiences (23%) (Table 2).

Table 2. Source of terms for collaborative query reformulation

Source term(s) for collaborative query	Instances	%
Individual queries	72	51
Chat logs	36	26
Previous knowledge and experiences	32	23
Total	140	100

3.2.1 Sources of Terms for Collaborative Queries

• **Individual queries:** It is evident that during CQ formulation, the collaborating team members employed the terms from either their own queries or their partner's queries. These terms were used to add to or replace any existing terms in CQ. Individual queries were identified as the mostly used source for CQ reformulation. Here is an example:

Individual query

Crown hotel Melbourne (issued by User 4, Pair 2)

Collaborative query

hotel in Melbourne (initiated by User 3, Pair 2)

Crown Towers hotel in Melbourne (modified by User 3, Pair 2)

Here, the query term "*Crown*" came from the individual query.

• **Chat logs:** Chat logs also contributed over a quarter to CQ reformulation. The participants borrowed the terms from their own chat logs or their teammate's chat logs to form a CQ. An example is given below:

Chat log

I want to find the local culture in Pukhet (issued by User 13, Pair 7)

Collaborative query

Restaurant in Thailand (initiated by User 14, Pair 7)

Restaurant in Thailand, Pukhet (the term "*Pukhet*" is added by User 13, Pair 7)

Here, the query term "*Pukhet*" came from the chat logs.

• **Previous knowledge and experiences:** The sources of these terms originated in searchers' experiences, retrieved relevant results, preferences, or previous knowledge, which accounted for 23% of the total CQ reformulations. An example comes as follows:

Collaborative queries

Food in Seattle (initiated by User 34, Pair 17)

Australian food in Seattle (the term "*Australian*" is added by User 33, Pair 17)

The term "*Australian*" was from neither individual queries nor chat logs. It was derived from users' preferences.

3.2.2 Operations Performed for Collaborative Query Reformulation

It is interesting to note that the pairs of participants performed several operations to reformulate CQs. These operations are defined based on two consecutive CQs: CQ_i and CQ_{i+1}, where CQ_{i+1} is the query immediately following the query CQ_i in the same session. The detailed descriptions of these operations are provided below:

• **Addition:** CQ_i and CQ_{i+1} contain at least one term in common and CQ_{i+1} contains more terms than CQ_i **.** Example: Food in Gold Coast Australia (CQ_i) → Indian food in Gold Coast Australia (CQ_{i+1})

• **Modification:** CQ_i and CQ_{i+1} contain at least one term in common and at least one term of CQ_i has been modified /replaced. Example: Parkview hotel in Melbourne (CQ_i) → Crown tower hotel in Melbourne (CQ_{i+1})

• **Reordering:** CQ_i and CQ_{i+1} contain exactly the same terms but the order of these terms may be different. Example: Machu pichu hotel (CQ_i) → Hotel machu pichu (CQ_{i+1})

• **Addition and modification:**
CQ_i and CQ_{i+1} contain at least one term in common; CQ_{i+1} contains more terms than CQ_i and at least one term of CQ_i has been modified/replaced. Example: Restaurant in Sydney (CQ_i) → Restaurant near hotel CBD Sydney (CQ_{i+1})

• **Addition and reordering:** CQ_i and CQ_{i+1} contain at least one term in common; CQ_{i+1} contains more terms than CQ_i and the order of the terms in CQ_{i+1} is different from CQ_i . Example: Gold Coast flight from Adelaide (CQ_i) → Flight from Adelaide to Gold Coast (CQ_{i+1})

Out of 140 CQs, 56 were initial CQs and the remaining 84 were the outcomes of the CQ reformulation operation (Table 3).

Table 3. Operations performed during collaborative query reformulation

Operation	Number	%
Addition	51	61
Modification	17	20
Addition and reorder	12	14
Reorder	3	4
Addition and modification	1	1
Total	84	100

It was observed that in most cases, the participants added query terms to the previous CQ (61%) and modified the previous CQ (20%) for reformulating the query.

3.2 Impact of Individual Queries and Chat Logs on Collaborative Query Reformulation

In this section we report the results regarding the usage of individual queries and chat logs during collaborative searching and their influences on the formulation of CQs. We performed significance tests to show the differences between individual queries and chatting logs as sources of CQ reformulation. The results show that during the collaborative search process, users collaborated more through chat logs (486 instances, mean= 28.59, SD = 12.31) than through queries (210 instances, mean= 12.35, SD = 5.49). The usage of chatting is statistically more significant than the usage of queries (p=.00005) collaboration (Table 4).

Table 4. Comparison between query and chat for collaboration and CQ reformulation in collaborative information searching

	Mean (SD)		Statistical Test
	Individual Query	Chat	
Collaboration	12.35(5.49)	28.59(12.31)	p=0.00005
Collaborative query reformulation	8.00(5.29)	5.14(3.00)	p=0.206

However, when comparing the usage of individual queries and chat logs as the source of query terms, the statistical test (Table 4 Row 4) indicates that individual query is not significantly different from chat logs in terms of formulating CQ (p=0.206).

The individual queries which contributed to the CQ reformulation were further divided into two types: self-generated queries and partner's queries, and the chat logs used for CQ reformulation were also classified into two categories: self-generated chat logs and partners' chat logs. As we further analyzed self-generated and partners query transactions and chat logs (Table 5), we found that self-generated query terms differ significantly from self-generated chat logs (p = 0.0054). This suggests that as a resource of CQ, the participants reused more query terms they had used before than their own chat logs. However, there is no statistical difference between partners' queries and partners' chat logs (p = 0.0909) with respect to CQ formulation. This may suggest that the query terms in partners' query histories and partners' chat histories are equally important for the CQ (re)formulation.

Table 5. Comparison of self-query versus self-chat and partners-query versus partners-chat as the source for collaborative query

	Mean (SD)		Statistical Test
	Query	Chat	
Self	2.50(1.41)	1.30(0.48)	p=0.0054
Partner	2.28(1.13)	1.64(0.74)	p=0.0909

In summary, collaborators pose individual queries and chats with teammate as they conduct collaborative searching on the Web. These individual queries and chat logs influence users' collaborative query reformulation, adoption of search strategies, and results evaluation. The evaluation of search results might further lead to the reformulation of CQ during the collaborative searching.

4. CONCLUSION AND FUTURE WORK

The goal of this study was to explore CQ reformulation behavior in travel information search. Our study differs from previous work [1, 4, 11] in that we investigated sources of terms for CQ reformulation, operations used to reformulate CQs, and the impact of individual queries and chat logs in reformulating CQs in the collaborative Web search. Our results provide some evidence that both individual queries and chat logs influence users' CQ reformulation. Self-generated individual queries seem more likely to be the source for new query terms than those that are partner-generated. The sources for CQ formulation such as chat logs and previous experience and previous knowledge are also important for new query terms. The findings have implications to support users when they reformulating CQs in the collaborative searching. For example, to minimize the efforts needed for CQ formulation, we may be able to identify the similarity among the query terms obtained from different sources. Currently we are working on automatic extraction of query terms from individual queries and chat logs for CQ reformulation.

5. REFERENCES

[1] Hust, A., Klink, S., Junker, M., and Dengel, A. 2002. Query reformulation in collaborative information retrieval. In *Proceedings of the International conference on Information and Knowledge Sharing*.

[2] Jansen, B. J., Spink, A., and Pedersen, J. 2005. A temporal comparison of altavista web searching. *Journal of the American Society for Information Science and Technology*, 56(6), 559–570.

[3] Jansen, B. J., Booth, D. L., and Spink, A. 2009. Patterns of query reformulation during Web searching. *Journal of American Society of Information Science and Technology, 60*(7), 1358-1371.

[4] Fu, L., Goh, D. H. L., Foo, S. S. B., and Supangat, Y. 2004. Collaborative querying for enhanced information retrieval. *Springer Berlin,* 378-388.

[5] Reddy, M. C., Jansen, B. J., and Krishnappa, R. 2008. The role of communication in collaborative information searching. In *Proceedings of the American Society for Information Science and Technology*, *45*(1), 1-10.

[6] Mohammad Arif, A. S., Du, J. T., and Lee, I. (in press). Understanding tourists' collaborative information retrieval behavior to inform design. *Journal of the Association for Information Science and Technology.*

[7] Mohammad Arif, A. S., Du, J. T., and Lee, I. 2013. Exploring tourists' collaborative Web search: Implications for system design. In *Proceedings of the Australasian Web Conference (AWC2013*) pages 49-58.

[8] Morris, M. R. 2008. A survey of collaborative web search practices. In *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems,* pages 1657–1660.

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

[10] White, R. W. and Roth, R. A. 2009. Exploratory search: Beyond the query-response paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 1(1), 1-9.

[11] Yue, Z., Jiang, J., Han, S., and He, D. 2012. Where do the query terms come from?: An analysis of query reformulation in collaborative web search. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM),* pages 2595-2598.