Comparing collaborative and independent search in a recall-oriented task

Hideo Joho, David Hannah, and Joemon M. Jose
Department of Computing Science, University of Glasgow
Sir Alwyn Williams Building, Lilybank Gardens, Glasgow G12 8QQ.
{hideo,hannahd,jj}@dcs.gla.ac.uk

ABSTRACT

Search interfaces are mainly designed to support a single searcher at a time. We therefore have a limited understanding of how an interface can support search where more than one searcher concurrently pursues a shared information need. This paper investigated the performance and user behaviour of concurrent search. Based on a recall-oriented search task, a user study was carried out to compare an independent search condition to collaborative search conditions. The results show that the collaborative conditions helped searchers diversify search vocabulary while reducing redundant documents to be bookmarked within teams. However, these effects were found to be insufficient to improve the retrieval effectiveness. We discussed the implications for concurrent search support based on our findings.

Symposium Themes

Personalised and collaborative information access in context; Interactive information retrieval and interface issues.

1. INTRODUCTION

While the advance of technology has allowed a search engine to process thousands of queries per second, the search interface is still fundamentally designed to support a single searcher at a time. The search process itself can have an intermediate person or agent to formulate queries [12], but we have limited understanding on how an interface can support searches where more than one person pursues a shared information need at the same time. This kind of search is called *concurrent search* in this paper. Concurrent search is a type of collaborative search where the members of a team share a period of time to achieve a goal [2], as opposed to the space[1]. One typical example is a holiday planning where a group of people try to find travel information together.

[1]You can share the space, but it is not the essential property of the concurrent search.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
IIiX '08 London, UK
Copyright 2008 ACM X-XXXXX-XX-X/XX/XX ...$5.00.

One professional group where concurrent search is frequently carried out is the intellectual property (IP) information community [6]. For example, when a firm considers an investment for the development of a new technology or product, they send a team of searchers to survey the IP coverage of existing patents [7]. The outcome of the task is crucial to the firm since the cost of patent infringement can be devastating. An interesting property of patent retrieval is the data management policy concerning search histories, where there is a great reluctance to keep search logs after a task has been completed, for two reasons: one being that it can be held against an organisation as negative evidence in lawsuits; the other being to protect industrial secrets. Such retrieval properties makes patent retrieval challenging since collaborative filtering techniques (e.g., [5]) might not be applicable to support concurrent search.

Motivated by the work task of the IP information community, we decided to investigate concurrent search as a generic research problem. Two existing studies in IR came to our attention in the literature. The first work was the SearchTogether system [8] which offered a range of support for collaborative search on the Web where the users are remotely located. The interface was designed to address the interaction properties such as awareness of other member's activities, division of labour, and persistency of search history. Another work was a collaborative video retrieval system [11] implemeted on a table with a touch-panel screen. Unlike the SearchTogether system, it was designed to support concurrent search via tangible device.

Both studies demonstrated the potential benefit of collaborative search interfaces and challenges for the effective design. However, neither compared their systems to independent search condition in the experiment. This paper presents a comparative study of concurrent search performed in the two conditions. We were interested in measuring the performance of concurrent search using a recall-oriented task since it had common properties with the task performed by the IP community. Based on the findings of the comparative study, we make recommendations of concurrent search environments.

1.1 Research questions

The specific research questions addressed in this paper were as follows.

- RQ1: *What strategy will be employed when people are engaged in a concurrent search?*

  The first question tried to understand the underlying intention of search behavoiur when a collaborative fa-

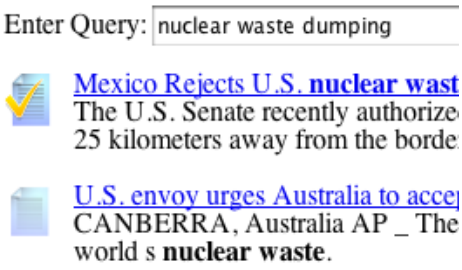

Figure 1: A search result with document status information. The first document was viewed and bookmarked while the second was viewed but not bookmarked.

cility was available in an environment. We also looked at the searching behaviour that was particularly affected by the collaborative search strategies.

- RQ2: *Do collabotive conditions perform more effective search than the independent condition?*

  This was the central question we would like to address in this paper by comparing the performance of collaborative and independent search conditions. We hypothesise that participants performed more effective search in the collaborative conditions compared to the independent condition due to their collaboration. We also looked at the effect of search time (e.g., first 5, 10, 15 minutes) in the behaviour and performance.

- RQ3: *How does the concurrent search affect searchers' perceptions on tasks?*

  The last question addressed the effect of concurrent search on people's perception of tasks. We hypothesised that participants had a more positive perceptions on tasks in the collaborative conditions compared to the independent condition. We also had feedback from participants for potential improvement of tested environment.

The rest of the paper is structured as follows. Section 2 discusses the experimental design of our study. Section 3 presents the results of the experiments. Section 4 discusses the findings of our study and implications for supporting of concurrent search. Finally, Section 5 concludes the paper with future work.

2. EXPERIMENTAL DESIGN

To answer the research questions discussed above, we carried out a user study of within-subject design. This section provides the details of the experiment.

2.1 Experimental conditions

We devised three conditions where participants carried out a search task. The task was to find as many relevant documents as possible for a topic (i.e., recall-oriented search), and participants were asked to bookmark the documents when they found relevant information. The description of the conditions is as follows.

Figure 2: Shared search history with text messaging. The first searcher, `demo_1`, submitted a query (1), viewed two documents (2 and 4), and bookmarked one of them (3). The second searcher, `demo_2`, submitted a similar query (5). Then, the team discussed a search strategy using text messaging (6 and 7).

Condition 1 (C_1): Independent search.

The first condition was devised to measure the performance of independent search. The same topic was given to two members of a team, but each performed the search independently. The search interface used in this condition was designed to be similar to existing search engines.

Since the underlying task in our work was recall-oriented, we added an icon to represent a status of retrieved documents. Different icons were used for the documents that had been viewed or viewed and bookmarked by participants, to facilitate the task. A screenshot of the search result in our interface is found in Figure 1.

Condition 2 (C_2): Shared search history.

To support a team of people performing a concurrent search, we have developed a search interface equipped with an instant messaging (IM) facility. The IM facility allowed the team to monitor the search activity of the other member. More specifically, the IM facility broadcasted the queries submitted, the documents visited, and documents saved to both of the team. There could be other information to be shared, but we chose these actions since they were common in many search tasks. By clicking the hyperlinks on the shared search history, the other member can replicate the actions taken by the partner. However, no communication was allowed in C_2.

The icons shown in the search results were also shared so that participants were able to avoid previously accessed documents if they wished to do so.

Condition 3 (C_3): C_2 + Communication.

In the third condition, participants were allowed to use any form of communication. More specifically, they were encouraged to communicate using either the text messaging facility of the interface or verbal communication, whichever

was more suitable for the situation and comfortable for the team. A screenshot of the sidebar showing shared search history and text messenging is found in Figure 2.

While it would be interesting to compare the effect of text and verbal chats by having a separate condition for each form of communication, we decided to address this issue in a future study. Therefore, in C_3, any form of communication was allowed. We initially considered separating the search history from text messages in the side panel. However, the feedback from the pilot study did not show a clear preference for this. Therefore, we decided to present both items of information in the same panel but highlight the text messages to better capture the searchers' attention. It was considered that this made the interface design simpler, and therefore cause less of a diversion to participants' attention, compared to a separate design.

In all conditions, the document collection was indexed by the Terrier system [9] and was also used for the back-end retrieval function in the experimental systems. We used the out-of-box settings of the Terrier system. Finally, the scenarios for the simulated work task situation [3] were created for each of the conditions to help participants understand the context of each condition. The scenario used in this study is found in the Annex.

2.2 Participants

Twelve pairs of participants (a total of twenty-four people) were recruited for our study. The recruitment was initiated by a call for participation distributed to the mailing lists in our organisation, and the first twelve pairs who agreed to participate were invited for the experiment. Most pairs were friends or colleagues. Of twenty-four people, there were five females and nineteen males. The entry questionnaire established that the average age of participants was 28 ($\sigma = 3.9$), ranging from 22 to 38. All were research students or research assistants in the computer science field. They had on average 8.4 years ($\sigma = 2.0$) of online search experience, and performed several searches every day. The participants' favourite search engines were Google (21), Yahoo (3), and AlltheWeb (1)[2]. Thirteen used advanced features of search engines. In summary, our participants were frequent searchers with an advanced knowledge of computers. We intentionally focused on this profile since the motivation of our work was based on information professionals who were not casual
searchers but who searched as an important part of their jobs. Experiments with different profiles are beyond the scope of this paper and part of our future work.

We asked two further questions in the entry questionnaire. One was about their experience of the concurrent search. The question was "*Have you had an experience of searching for information with someone else using search engines?*" Eight participants reported that they had performed concurrent searches. The searches were mainly transactional purposes; an evaluation of business software, holiday information (e.g., flights, hotels, etc.), and movie/concert information. The other question asked about their experience in using instant messenger tools. This was because our experimental systems used an instant messenger tool to share search history in real-time. It was found that twenty had used an instant messenger tool, and fifteen used it everyday.

[2]One participant selected Google and Yahoo.

2.3 Test collection

We used the Robust Track of TREC 2005 [13] for our experiment since the track's topic and document collection has successfully been used by other interactive experiments [1]. The Aquaint collection used in the Track was roughly 3GB of text and included 1,033,461 documents. There are 50 topics used by the Robust Track 2005. We considered two options for the selection of topics. One was to use three topics selected by the experimenter to better control the experiment, and another was to let participants select three topics based on their interest for a better engagement with the tasks [10]. In this study, we decided to take both of the options by selecting 15 candidate topics from the Track and allowing participants to select three topics based on their interest.

The selection of candidate topics was carried out as follows. 50 topics were first sorted by the number of relevant documents defined in qrels. The median of relevant documents was 110 ranging from 9 (Topic 345) to 376 (Topic 354) in the 50 topics. We took the 15 topics in the middle of the sorted list so that all topics had a sufficient number of relevant documents to find within 15 minutes of concurrent search, but also to ensure the number of relevant documents did not vary too much. Table 1 shows the 15 topics selected by this process along with the number of relevant documents and number of times each topic was selected by our participants (denoted as *Pick*). As can be seen, no obvious concentration on particular topics was found.

Topic	Qrels	Pick	Topic	Qrels	Pick
303	86	4	344	123	3
363	111	3	367	95	5
383	137	1	393	97	1
397	88	3	409	151	0
435	152	0	439	127	5
448	121	3	625	109	2
651	97	2	658	116	3
689	110	1			

Table 1: 15 candidate topics

2.4 Procedure

For each team of participants, the experiment was carried out in the following manner: 1) When the team arrived, they were welcomed and given an information sheet that described the overview of the experiment; 2) When they agreed to participate, they were asked to sign a consent form, followed by an entry questionnaire to establish their search background; 3) They were presented with 15 candidate topics from which they were asked to select three topics that were most interesting to the team; 4) Then they performed a training session based on the first condition they were going to carry out; 5) They performed the first task for 15 minutes, followed by a post-search questionnaire to capture their subjective assessments on the task; 6) A change of topic and scenario occurred, and they performed another training session with the new scenario; 7) Repeat the step 5 and 6 for the second and third task; 8) When they completed three tasks, they were asked to fill in an exit questionnaire to capture feedback about the concurrent search, interfaces, and tasks they performed.

The training sessions consisted of the introduction of the

work-task scenario, description of the database, and tutorial on the search interfaces with a sample topic. The training session tended to take 10 to 15 minutes. The second and third sessions tended to be shorter than the first session since there were duplicated steps in the three scenarios. However, participants were encouraged to verbally discuss their search strategy during the training sessions of C_2 and C_3 to facilitate their collaboration. The entire session took 100 to 120 minutes. They were rewarded with £10 for participation, and the best performing team was awarded with a £50 prize. The order of the three conditions was rotated to reduce the learning effect in the analysis. There were six combinations of three conditions, and participants were systematically assigned to one of the six combinations. The three topics selected by participants were presented in the decreasing order of their interest. In other words, they performed the least interesting topic (out of three) first and most interesting topic last, to compensate for the fatigue effect with their topic interest.

3. RESULTS

This section presents the results of our experiments based on the research questions discussed in Section 1.1. Statistical significance of the results was tested by the Wilcoxon signed rank test with $p \leq .05$. We used C_1 as a control condition, thus the pair-wise comparisons were made between C_1 and C_2 and between C_1 and C_3 unless otherwise stated.

3.1 Search strategies

RQ1 looked at the search strategy employed by participants during the tasks of concurrent search. The objective of this question was to gain an insight into the underlying intentions that can affect the searching behaviour and performance. As discussed in Section 2.4, participants were encouraged to discuss a search strategy in the training session of C_2 and C_3. In addition, it was possible to discuss and revise the initial plan as the tasks progressed in C_3. The strategies employed were self-reported and captured by the exit questionnaire.

By far the most frequent strategy reported was to submit different queries. Almost all teams mentioned this as a strategy. One stated "each one would go through different groups of documents, so the scope of the search could be wider and no time would be lost duplicating the same work". Some teams articulated how to differentiate the queries. For example, one team tried to divide by date (i.e., one searched for between 19xx to 200x, and another searched for a different range). Another team decided to try generic words by one member and specific words by another. The second category of strategy involved the browsing and judgement of retrieved documents. A frequently reported strategy in this category was not to visit the documents that were already accessed or judged by the team. Several teams mentioned this as a strategy. In other words, they trusted the relevance judgement of the other team member. For example, one team tried to submit the same query and browse the documents in different result pages (i.e., one went through the odd pages while another went through the even pages). Furthermore, a couple of teams tried to cross-check the viewed documents before bookmarking, so that they could agree on the relevance judgement. The third category of strategy involved the sharing of findings during the tasks. This was often regarded as a strategy in C_3. For example, participants tried to share the effectiveness of keywords and/or interesting information found in the retrieved documents. Many teams reported that they carried out these actions by just looking at the shared search history shown in the interface.

	C_1		C_2		C_3	
	Ind.	Team	Ind.	Team	Ind.	Team
Query	11.3	21.8	**16.5**	**31.7**	**15.5**	**29.8**
Word	11.0	18.8	**16.4**	**26.8**	**15.8**	26.0
View	31.2	51.3	26.0	51.0	23.8	46.0
Bookmark	23.2	37.6	18.6	36.7	**16.7**	32.3
SCTR	0.74	0.74	0.69	0.71	0.64	0.67

bold: statistically significant ($p \leq .05$).

Table 2: Statistics of searching behaviour

In summary, participants were trying to improve the search efficiency by submitting different queries and by avoiding viewing retrieved documents that have already been accessed by the team. As illustrated in Section 2.1, these strategies were supported by the interface used in C_2 and C_3. Participants were also trying to share useful information where appropriate.

We then looked at the effect of the strategies on the behaviour of the concurrent search. Table 2 shows the average number of queries submitted to complete a task, number of unique words used in the queries, number of documents viewed, number of documents bookmarked, and finally, a successful click-through rate (SCTR) which was defined by the proportion of viewed documents that were judged relevant, and thus, bookmarked by participants. For each condition, the data of individual members (annoted as *Ind.*) and of the team (annoted as *Team*) are shown. Note that the team data of C_1 was generated by treating the results as if they worked as a team to be comparable to the other two conditions. As can be seen, the number of queries and the range of search vocabulary were higher in C_2 and C_3 compared to C_1. The individual statistics and team stastistics were compared across the conditions separately (i.e., C_1 Ind. vs. C_2 Ind., C_1 Team vs. C_2 Team, etc.). The number of documents viewed by the team was relatively similar across the conditions. However, there was a difference in the level of redundancy of viewed documents. Due to the independent nature of C_1, there were more duplicated documents viewed as a team than C_2 and C_3. Finally, participants tended to bookmark fewer documents in C_3 compared to the other two conditions. Consequently, the SCTR decreased as the level of collaboration increased. The bookmarking and SCTR were based on perceived relevance of retrieved documents.

As for the verbal and textual communication in C_3, six teams reported in a post-search questionnaire that they had a text chat while ten teams reported they had a verbal chat. The motivations for initiating communication varied: suggestion of query terms to use or avoid; spelling questions; general questions (e.g., is a boat a ship?); pointing out a wrong relevance judgement; warning of the next move; confirming the relevance of retrieved documents. While some teams found it unnecessary to have explicit communication, this feedback demonstrates the diversity of information that was shared and exchanged during the concurrent search.

To summarise, the search strategy employed by participants appeared to affect the query formulation and the level of redundancy of viewed documents. In particular, partic-

	C_1		C_2		C_3	
	Ind.	Team	Ind.	Team	Ind.	Team
Bookmark	23.2	37.6	18.6	36.7	16.7	32.3
Relevant	15.7	23.3	11.6	22.8	**10.3**	20.1
High Rel	11.4	16.4	**4.9**	9.7	8.1	15.8
Precision	0.66	0.63	0.61	0.60	0.60	0.61
Recall	0.16	0.23	0.12	0.23	0.10	0.20
Quality	0.75	0.73	**0.44**	0.46	0.67	0.69

Table 3: Search performance

ipants tended to submit more queries with diverse search vocabulary in C_2 and C_3. Most teams explicitly communicated at least once or twice on a range of issues during the tasks. The next section examines how this difference in search behaviour affected the search performance.

3.2 Search performance

RQ2 looked at the search performance of the three concurrent search conditions. Table 3 shows the average number of documents bookmarked by participants (same data as Table 2), of those that were relevant (including relevant and highly relevant defined by TREC's qrels), and of those that were highly relevant. The bottom three rows show the precision (proportion of relevant docs in bookmarked docs), recall (proportion of bookmarked relevant docs in all relevant docs for a topic), and finally, a quality measure (proportion of highly relevant docs in all bookmarked relevant docs). As can be seen, there was no clear evidence suggesting that the collaborative conditions, C_2 and C_3, improved the search performance over the independent condition, C_1. This was true for all measures shown in the table. We noted a low number of highly relevant documents retrieved (and subsequent quality measure) in C_2, and found that there were several topics in the Robust Track which had no highly relevant judgements. Participants selected more of those topics in C_2 than the other two conditions. This was something overlooked during the selection process of candidate topics. Other measures were not affected by this.

As mentioned in the previous section, there were duplicated documents viewed by the team members in C_1. The same effect was found in the performance measure, and summarised in Table 4. The redundancy was measured by the following way:

$$Redundancy = \frac{(\sum d_{ind}) - d_{team}}{\sum d_{ind}}$$

where d_{ind} is the number of unique documents found by individual members and d_{team} is the number of unique documents found by the team. As can be seen, approximately 18% of redundant documents were found in C_1 while this was reduced to 2 to 4% in the other two conditions. The difference between the independent and collaborative conditions was found to be significant in all three measures ($.005 \geq p \leq .04$). The low duplication in C_2 and C_3 confirmed the browsing strategy discussed in Section 3.1 had an effect in the task.

To summarise, the increased number of queries and diverse search vocabulary did not have a positive effect on the search performance in our study. However, the collaborative conditions can reduce the redundancy of relevant documents found among the team.

	C_1		C_2		C_3	
	Mean	σ	Mean	σ	Mean	σ
Bookmarked	0.17	0.13	0.02	0.03	0.03	0.04
Relevant	0.18	0.16	0.04	0.06	0.02	0.04
High Rel	0.18	0.19	0.01	0.02	0.03	0.04

Table 4: Redundancy of bookmarked documents

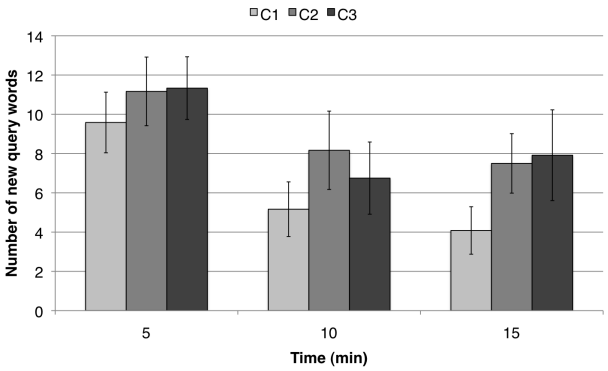

Figure 3: Number of unique query words submitted by teams within the three search stage bins.

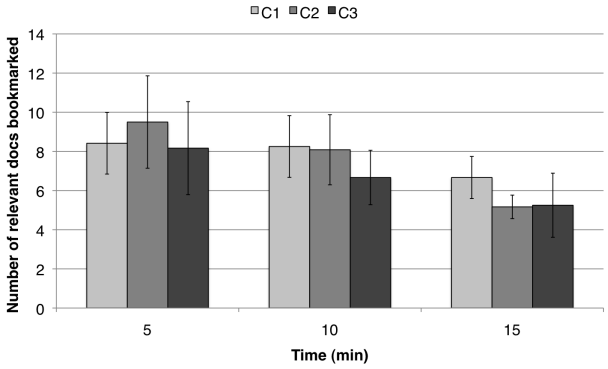

Figure 4: Number of relevant documents bookmarked by teams within the three search stage bins.

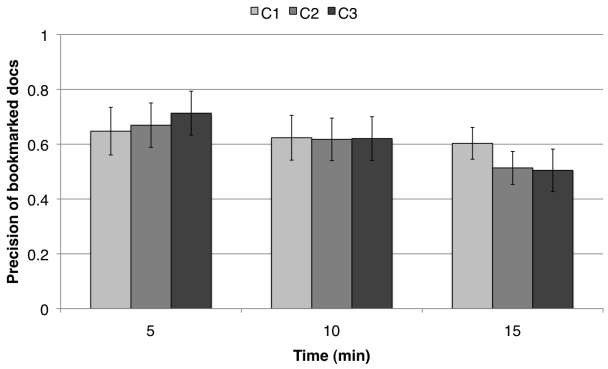

Figure 5: Precision of documents bookmarked by teams within the three search stage bins.

We also investigated the effect of search stages on the performance of the concurrent search. Here, we present the results of the size of search vocabulary, number of relevant

documents bookmarked, and precision of bookmarked documents. The data was analysed for the three stages of task (i.e., the first, second, and last five minutes) independently based on the team's performance. The results are plotted in Figure 3, 4, 5, respectively.

In Figure 3, we can see that the number of unique words was generally decreasing over the search stage, suggesting that a fatigue effect on the query reformulation might have occurred in all conditions. It also shows the collaborative conditions had a consistently higher number of unique words than the independent condition across the search stages, contributing to the significant difference shown in Table 2. We ran a two-way repeated measures ANOVA with the condition and search stage as independent variables. The tests showed that both the condition and search stage had a significant effect on the number of unique words (Condition: $F = 3.50$, $p = .02$, Search stage: $F = 12.42$, $p = .000$). However, the interaction between the two factors was not found to have a significant effect. Therefore, while participants employed more unique words in the collaborative conditions than the independent condition, there was no sufficient evidence to suggest that the effect was affected by the search stage. We repeated the same procedure of significant tests for the data shown in Figure 4 (Number of relevant documents bookmarked) and 5 (Precision of bookmarked documents). As for the number of relevant documents bookmarked, the figure seems to suggest that there is an effect of search stage. However, the two-way repeated measures ANOVA did not show a statistical significance for the search stage nor the conditions. As for the precision, the graph seems to show a small difference among the search stages for all conditions. A two-way repeated measures ANOVA showed that the precision was affected by the search stage ($F = 3.66$, $p = .04$) but not by the conditions. There was no significant interaction effect, either.

In summary, the search stage appeared to affect the vocabulary size of queries and precision of bookmarked documents. However, there was no sufficient evidence to suggest that these effects were affected by the collaborative conditions.

3.3 Searchers' perceptions and feedback

RQ3 looked at the effect of three conditions on participants' perceptions of search tasks. The subjective assessments of search tasks were captured by a 7-point Likert scale of agreement with each statement. For example, the statement "the task I performed was simple" was used to capture the perception of task complexity, and the answers ranged from 1 (disagree) to 4 (neither) to 7 (agree). Table 5 shows the results of participants subjective assessments. For each question, it shows the average score along with the standard deviation (σ) for the three conditions.

The results basically show that there was no significant effect of the conditions on participants perceptions of tasks. A relatively high score of Task Interest appears to support our design of allowing participants to select search topics based on their interest. The next three rows shows the result of the satisfaction of search outcomes, the confidence in the accuracy of bookmarked documents, and finally the sense of finding all relevant documents for a topic. A low level of satisfaction in C_3 suggests that participants had a correlated level of satisfaction to the actual performance. On the other hand, there was no significant effect on the perception of the accuracy of bookmarked documents. However, participants appeared to feel that more relevant documents were found when they worked as a team, which could be a misleading perception given that the recall was very similar across the three conditions. Overall, however, there was no significant effect of the collaborative conditions on participants' perceptions of tasks.

	C_1		C_2		C_3	
	Mean	σ	Mean	σ	Mean	σ
Task Clear	6.0	1.6	6.2	1.2	6.2	0.9
Task Simple	5.3	1.6	4.7	1.6	4.7	1.7
Task Familiar	4.5	1.8	5.1	1.7	4.5	1.9
Task Interest	5.4	1.3	5.0	1.4	5.3	1.4
Task Relax	4.9	1.4	4.4	1.5	4.6	1.6
Satisfaction	5.5	1.3	5.1	1.3	4.7	1.6
Accuracy Conf	5.7	1.0	5.5	1.0	5.8	1.0
Found All	4.0	1.7	4.6	1.3	4.5	1.5
Preference	2.8	0.4	**1.7**	0.6	**1.5**	0.7

Table 5: Perception of search tasks

	C_2		C_3	
	Mean	σ	Mean	σ
Query	5.1	1.5	5.8	1.2
Viewed docs	4.4	1.7	4.8	2.1
Bookmarked docs	5.0	1.9	5.5	1.7
Text chats			3.5	2.0
Verbal chats			4.8	2.1

Table 6: Perception of shared search history and communication channels.

In the exit questionnaire, we asked participants to order the three conditions based on their preference, using Score 1 to be most prefered and Score 3 be least preferred. The bottom row of Table 5 shows the result of the question. As can be seen, participants showed a strong preference for collaborative conditions. The Friedman test showed that the difference among the three conditions was significant, and post-hoc repeated tests by the Wilcoxon signed rank test showed that the difference between C_1 and C_2 and between C_1 and C_3 was significant (with Bonferroni correction, $p \leq .0167$). This could be an artifact of the conditions' novelty since the satisfaction of search tended to be lower in C_2 and C_3. While we do not completely dismiss such an effect, some participants' comments on the preference informed us of the impact of sharing information and communication to their search experience. For example, one stated "Communication advances our search. Search history gives you confidence that you are on the right way. Independent was not giving any of the above".

Table 6 shows participants perceptions of the usefulness of shared search history and communication channels. As we expected, participants found it particularly useful to share queries. A good score was also given to the bookmarked documents. Participants frequently commented that showing the viewed documents was useful on the search results but not on the sidebar as this can bury other information. As for the communication channels, the text messaging was less popular than the verbal communication, as participants found it easier to speak than type. An exception was spelling questions.

Finally, we asked participants their likelihood of using a search interface with a concurrent search support if it was available on the web. Five participants (21%) gave either a negative or neutral score, while nineteen (79%) gave a positive score to the question. The average score was 5.4 ($\sigma = 1.6$). This was encouraging given that participants were experienced searchers who have long been performing search alone. Many participants commented that they would use a collaborative search system for work-related tasks where search would involve a wide range of topics. One participant suggested it's use in learning environments, and another mentioned search tasks on unfamiliar topics. Searching for tourist information was also mentioned as a task. This feedback supports our motivation for investigating concurrent search with a recall-oriented task, but also it suggests a potential application in other domains.

4. DISCUSSION

This section first summarises the main findings of our experiment and discusses the implications for concurrent search environments. Ideas for improving current interface design are also discussed.

The first finding of our study was that the size of search vocabulary can be diversified by creating a collaborative search condition compared to an independent condition. The visual feedback of members' activities, a status of retrieved documents, and different forms of communication appeared to stimulate query reformulation process. Another finding was the level of redundancy occurred in the independent condition. When the team consisted of two searchers, we found that approximately 18% of redundant relevant documents were bookmarked. The collaborative conditions were able to reduce the redundancy to 3-4%. The third finding was that the benefits of collaboration such as the increased size of vocabulary and reduced redundancy were insufficient for improving the retrieval effectiveness. Adding a collaborative element in the environments did not have an effect on participants' perceptions except their preference.

Our speculation on the reasons for the lack of improvement in the collaborative conditions is as follows. In our experimental systems, a collaborative action can be as simple as looking at the sidebar for the partner's query to sending a text message to discuss something during the tasks. Every collaborative action had a potential benefit and cost. Our results seem to suggest that the cost of collaboration was larger than the benefits in our study. For example, two teams had very frequent verbal communication to share their findings during the tasks and to discuss the relevance of documents before bookmarking. These teams ended up spending less time to find relevant documents. A comparable performance between C_2 and C_3 suggests that the communication channels devised in our experimental systems need to be refined. There are two other potential factors. One is learning process of search topics. In the collaborative conditions, participants tended to avoid viewing those documents that have already been seen by the team member. In other words, participants did not learn about a topic from those documents bookmarked by the partner. This might cause them to develop a partial view of the topic space in the collection. Implicit or explicit means of sharing topic knowledge among the team might need to be devised for this factor. Another factor is peer pressure. Since the experimental systems allowed the team to see the real-time activity of the partner, there might be an element of peer pressure which caused them to bookmark documents without a sufficient level of attention to the detail of documents. These are speculations and further study need to be carried out to verify them.

The following is the implications for the support of concurrent search based on our study. When search performance is absolutely essential, we suggest letting a team work independently. While there is likely to be some level of redundancy in this process, the overall outcome is still likely to be comparable to a collaborative condition. If the level of redundancy is found to be very high, sharing the document status (i.e., viewed and/or bookmarked by team members) on the search results will help reduce it. When an information need is known to require a range of search vocabulary, we suggest to share the query terms among the team as a complement to other knowledge resources such as thesauruses. Sharing query terms during the search task is likely to increase the overall vocabulary size compared to an independent environment. This might also help when individual team members have a varying level of familiarity with the information need. A less experienced member can learn from the other members in query reformulation. Another case for collaboration might be at an early stage of search where the team is exploring the topic area and document collection. In such a problem-formulation stage [4], the communication tools are likely to facilitate the discussion and exchange of findings, hints, and potential problems.

This was our first study looking at the information seeking behaviour and search performance of concurrent search. As such, there were limitations. First, due to the resource available, we were unable to recruit a team bigger than two people. Therefore, our findings might be different for a bigger team. For most participants, this was the first time to perform concurrent search. Second, our experiment only captured the first 15 minutes of a recall-oriented task. As the recall level in the results suggest, more relevant documents can be found, and thus, a study with a longer task time is of our future interest. In particular, we are interested in observing the level of redundancy in a longitude study. The experimental systems developed for this study were only one possibility of supporting concurrent search. Different approaches should also be evaluated to improve our understanding of concurrent search support. Finally, the back-end retrieval system used in our study was not designed for concurrent search.

Finally, we discuss some ideas for improving concurrent search support based on the feedback from participants. There were two main categories of suggestion: the first category was concerned with the structure of information presented in the sidebar. While participants commented that the current format was simple and useful, they also wanted an option to view a search history summary. As opposed to showing the real-time search activity, a list of query terms and/or bookmarked documents can be shown with an indication of team members. A couple of participants also wanted a way to *flag* queries which they found particularly useful. The second category was concerned with the feedback shown in the search results. The current interface was designed to display document status icons, and participants wanted to see a feedback of query terms, too. For example, when a query was submitted, the result page can show whether or not the same query has been submitted by the team member, and how many documents were bookmarked

by the query. This could offer further support to participants' query reformulation. Another request was a real-time update of the document status icons since the current interface only showed the icons for subsequent queries. This feedback will be considered for the next version of our interface development.

5. CONCLUSION AND FUTURE WORK

This paper investigated the search performance and information seeking behaviour of concurrent search. Based on recall-oriented search tasks, a user study was carried out to compare three conditions: Independent, shared search history, and communication. The collaborative conditions helped the searchers to diversify search vocabulary and reduce the redundancy in finding relevant documents. However, the effect appeared to be insufficient to improve the retrieval effectiveness. The questions that we would like to address in future work are as follows. The first question is the effect of concurrent search in different tasks. Although this paper was motivated by a recall-oriented task, other tasks such as a decision-making tasks are also of interest. The second question is the effect of team size. This paper investigated a pair of searchers due to limited our resource, but how would the performance change when the size of team increase to three and more? The third question is the utility of relevance feedback and clustering in concurrent search scenario. Both technologies have been extensively studied in single searcher environments, but we have limited understanding in concurrent search.

6. ACKNOWLEDGEMENTS

The authors thank to participants of our study. Thanks also go to anonymous reviewers for their constructive feedback. This work was supported by MIAUCE Project (IST-033715). Any opinions, findings, and conclusions described here are the authors and do not necessarily reflect those of the sponsor.

7. REFERENCES

[1] J. Allan. Hard track overview in trec 2005 high accuracy retrieval from documents. In *NIST Special Publication: SP 500-266 The Fourteenth TREC*, 2005.

[2] R. M. Baecker. *Readings in Human-Computer Interaction: Toward the Year 2000*. Morgan Kaufmann, 1995.

[3] P. Borlund. Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 56(1):71–90, 2000.

[4] K. Byström and K. Järvelin. Task complexity affects information seeking and use. *Information Processing & Management*, 31(2):191–213, 1995.

[5] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry. Using collaborative filtering to weave an information tapestry. *Commun. ACM*, 35(12):61–70, 1992.

[6] P. Hansen and K. Järvelin. Collaborative information retrieval in an information-intensive domain. *Information Processing & Management*, 41(5):1101–1119, 2004.

[7] IRF, editor. *Proceedings of the First Information Retrieval Facility Symposium (IRFS)*, Vienna, Austria, 2007. Matrixware.

[8] M. R. Morris and E. Horvitz. Searchtogether: an interface for collaborative web search. In *UIST '07: Proceedings of the 20th annual ACM symposium on User interface software and technology*, pages 3–12, New York, NY, USA, 2007. ACM.

[9] I. Ounis, C. Lioma, C. Macdonald, and V. Plachouras. Research directions in terrier: a search engine for advanced retrieval on the web. *Novatica/UPGRADE Special Issue on Web Information Access*, 8(1):49–56, 2007.

[10] I. Ruthven, M. Baillie, and D. Elsweiler. The relative effects of knowledge, interest and confidence in assessing relevance. *Journal of Documentation*, 63(4):482 – 504, 2007.

[11] A. F. Smeaton, H. Lee, C. Foley, and S. McGivney. Collaborative video searching on a tabletop. *Multimedia Syst.*, 12(4-5):375–391, 2007.

[12] A. Spink. Term relevance feedback and query expansion: Relation to design. In B. W. Croft and C. J. Van Rijsbergen, editors, *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 81–90, Berlin, Germany, 1994. ACM.

[13] E. Voorhees. Overview of the trec 2005 robust retrieval track. In *NIST Special Publication: SP 500-266 The Fourteenth TREC*, 2005.

APPENDIX

Scenario for simulated work-task situation (C_3)

You are a member of an information broker company who specialises in finding relevant information based on a client's information need. Today, you are asked to find as many relevant documents as possible from a database of newswires.

A small team, consisting of you and your partner, is assigned to carry out today's task. Our search interface allows the team to share the following information during the search:

- Query terms submitted
- Documents viewed
- Documents bookmarked

In addition, you can send a text message to your partner to discuss anything during the task. If necessarily, you can have a verbal communication with your partner. At the end of the task, the bookmarked documents will be sent to the client.

The information need required by today's client is as follows. Please read it carefully and bookmark as many relevant documents as possible, while avoiding bookmarking non-relevant documents.

```
<num> Number:  401
<title>foreign minorities, Germany
<desc>Description:  What language and cultural differences impede the integration of foreign minorities in Germany?
<narr>Narrative:  A relevant document will focus on the causes of the lack of integration in a significant way; that is, the mere mention of immigration difficulties is not relevant.  Documents that discuss immigration problems unrelated to Germany are also not relevant.
```