How Can We Better Support Users with Non-Uniform Information Access in Collaborative Information Retrieval?

Nyi Nyi Htun
SEBE
Glasgow Caledonian University
Glasgow, G4 0BA, Scotland, UK
nyinyi.htun@gcu.ac.uk

Martin Halvey
Department of CIS
University of Strathclyde
Glasgow, G1 1XQ, Scotland, UK
martin.halvey@strath.ac.uk

Lynne Baillie
Department of CS
Heriot-Watt University
Edinburgh, EH14 4AS, Scotland, UK
l.baillie@hw.ac.uk

ABSTRACT

The majority of research in Collaborative Information Retrieval (CIR) has assumed that collaborating team members have uniform information access. However, practice and research has shown that there may not always be uniform information access among team members, e.g. in healthcare, government, etc. To the best of our knowledge, there has not been a controlled user evaluation to measure the impact of non-uniform information access on CIR outcomes. To address this shortcoming, we conducted a controlled user evaluation using 2 non-uniform access scenarios (document removal and term blacklisting) and 1 full and uniform access scenario. Following this, a design interview was undertaken to provide interface design suggestions. Evaluation results show that neither of the 2 non-uniform access scenarios had a significant negative impact on collaborative and individual search outcomes. Design interview results suggested that awareness of team's query history and intersecting viewed/judged documents could potentially help users share their expertise without disclosing sensitive information. Based on our results we provide important design recommendations to better support users with non-uniform information access in CIR.

CCS Concepts

• **Information systems~Collaborative search** • *Information systems~Search interfaces* • Information systems~Retrieval effectiveness • Information systems~Presentation of retrieval results

Keywords

collaborative information retrieval; multi-level collaboration; information access; interface design; non-uniform access

1. INTRODUCTION

Although search is often considered a solitary activity, there are many situations that call for people with shared information needs to work together to search for information and judge documents [10]. This is known as Collaborative Information Retrieval (CIR). A great deal of research in CIR [2, 11, 12, 22, 23] has utilised concepts like awareness, division of labour and persistence to enhance the collaboration experience and outcomes for users. However, much of this research has assumed that all members of a collaborative search team have equal access to the underlying information, and that they can share any information with each other without restriction, e.g. [2, 22, 23]. In reality, there are a number of societal, legal or security reasons that may prevent a searcher from sharing information within or outside of a group. An example might be a finance manager and a production manager in a manufacturing company who are working together to identify a problem. Regardless of their differing access to underlying information such as employee records, production-line etc. the two managers must collaborate somehow to achieve a successful outcome. This type of scenario is referred to as Multi-Level Collaborative Information Retrieval (MLCIR) [13]. Whilst MLCIR shares a large number of characteristics with CIR, unlike traditional CIR, MLCIR is also concerned with information flow, security and shareability as well as supporting collaboration.

CHIIR '17, March 07-11, 2017, Oslo, Norway
 ISBN 978-1-4503-4677-1/17/03…$15.00
DOI: http://dx.doi.org/10.1145/3020165.3020171

Some researchers have observed real life MLCIR scenarios and begun to examine MLCIR in areas like legal search [3], crisis management [4] and healthcare [19]. From their observations, these researchers provide insights into how non-uniform information access can negatively impact CIR outcomes. Handel and Wang [13] also presented a number of case studies in healthcare, business and government domains, and explicitly discussed the challenges in MLCIR scenarios. However, this previous work [3, 4, 13, 19] was observational and did not provide empirically based solutions to support users in MLCIR scenarios. As an early attempt to quantify the impact of non-uniform information access in CIR, Htun et al. [15] conducted a simulated evaluation. However, this work did not go as far as a user evaluation. To the best of our knowledge, there has not yet been a controlled user evaluation to assess the impact of non-uniform information access on CIR outcomes. To address this shortcoming, we conducted a user study where we investigated the impact of different information access scenarios; namely: document removal, term blacklisting and full access. The document removal and term blacklisting scenarios are based on real-life MLCIR examples [13]. After being exposed to different information access scenarios in a controlled experiment, we conducted a design interview with our participants to obtain feedback to assist with future development of MLCIR systems. The research questions we attempt to address in this study are:

RQ1: What are the effects of non-uniform information access scenarios on collaborative search outcomes?

RQ2: What are the effects of non-uniform information access scenarios on individual search outcomes?

RQ3: How can MLCIR interfaces be designed so that they are usable and useful for the end users?

The remainder of the paper is organised as follows: in Section 2, we discuss related work regarding CIR and MLCIR. In Section 3, we describe our experimental setup. In Section 4, we present the results of our study. In Section 5, we discuss findings and design recommendations based on the evaluation results and feedback from the design interview. Finally, we provide some conclusions and discuss possible future work in Section 6.

2. RELATED WORK

2.1 Collaborative Information Retrieval

A large number of people often engage in collaborative search activities [21]. Working in collaboration can not only be beneficial for recall-oriented search tasks [25], but also reduces the workload of team members [7]. However, collaborative search activities that involve information sharing via conventional communication channels e.g. email, telephone, etc. are often inefficient [21]. This is mainly because team members do not have an efficient way to obtain awareness, communicate and divide labour for the search task. Also, it is often essential for team members to be able to retain discovered content for later use [21]; this is referred to as "persistence of data" [12, 22, 23]. As such, a great deal of research has been conducted in CIR in order to enhance communication and collaboration capabilities of users [2, 8, 11, 12, 22, 23].

According to Golovchinsky et al. [10], CIR systems range from purely user interface level to algorithmic mediation. In UI-only mediated systems, collaboration is supported only at the user interface level whereas algorithmic mediation search systems employ an algorithmic layer to present results differently to different users. There are also systems that allow collaboration via communication only e.g. instant messaging, voice chat, video conferencing, etc. However, previous research has argued that CIR systems must provide more than just communication [11, 22]. Example systems for UI-level mediation include SearchTogether [22], CoSearch [2], Coagmento [11], WeSearch [23], ViGOR [12] and Físchlár-DiamondTouch [27]. SearchTogether [22] allows asynchronous collaboration for web search by enabling storage of all objects and actions performed at the end of a search session. SearchTogether also allows remote collaboration by providing features like instant messaging and split-screen search. CoSearch [2] is designed for synchronous and co-located web search. It is aimed at allowing collaboration over multiple devices e.g. shared computers and Bluetooth enabled mobile devices. Coagmento [11] was developed as a Firefox plug-in and an Android app to allow asynchronous and remote collaboration on both computers and mobile devices. WeSearch [23] is a system specifically designed for tabletops, and allows co-located web search for up to four people. Whilst SearchTogether, CoSearch, Coagmento and WeSearch support text retrieval e.g. webpages, there has also been research conducted into multimedia CIR. For example, Halvey et al. [12] developed a collaborative video retrieval system called ViGOR, which allows asynchronous and distributed collaboration. Smeaton et al. [27] developed a synchronous and co-located video retrieval system for a multi-user, touch sensitive tabletops named Físchlár.

Algorithmic mediation is widely used in recommender systems (e.g. Amazon shopping recommendations [20]). Such systems have an algorithmic layer for processing users' search behaviour to reorder new search results. In a study conducted by Pickens et al. [24], it was shown that algorithmic mediation can yield better results for recall oriented tasks compared to the UI-only mediated CIR systems. Other examples of algorithmic mediated CIR systems include I-SPY [28] and Cerchiamo [9]. I-SPY [28] is a community-based web search engine which takes advantage of past search behaviour to re-rank future search results in a way that recognises the implicit preferences of communities of searchers. Cerchiamo [9] is an algorithmically mediated synchronous collaborative search system that can allocate roles to users and then split up work based on the roles. Algorithmic mediation often attempts to leverage different roles within a search team. A common assumption within CIR is that different roles can have a positive impact on search outcomes, e.g. Golovchinsky et al. [9] and Pickens et al. [24]. This may not be the case in MLCIR; in some cases, users with differing roles may not be able to share certain information with each other [13], which could have a negative impact on CIR outcomes. More recently, Tamine and Soulier [30] conducted a user study to investigate how role assignment could impact CIR and found that assigning roles to the users could limit the precision of search results, demonstrating that in normal CIR scenarios that an assignment of roles is not always beneficial. Whilst they investigated roles, Tamine and Soulier did not consider how non-uniform information access might exist between the different roles within a team.

2.2 Multi-Level Collaborative Information Retrieval

MLCIR is a complex problem and has multifaceted constraints with regards to information accessibility and shareability [13]. MLCIR in organisations can be difficult to manage as sometimes organisations do not want any unnecessary information contamination or disclosure of sensitive information within or outside of the organisation. In an attempt to gain a better understanding of these problems and difficulties with collaboration in information-sensitive environments, some researchers have begun to investigate a range of real life scenarios. Attfield et al. [3] studied a large London law firm and discussed difficulties and complexities that may arise in a group-based awareness system, and also provided design recommendations for future developments. Karunakaran and Reddy [18] gathered critical-incident self-reports of 307 employees working in various organisations, and studied barriers that exist in collaborative information seeking. Their study found that these barriers exist due to factors like organisational setting, lack of technologies, individual perceptions, and lack of efficient communication in teams. Karunakaran and Reddy [19] also presented case studies in the healthcare domain, and discussed the frequent occurrence of non-uniform knowledge distribution and miscommunication. Bjurling and Hansen [4] looked at a Swedish crisis management system and described how different interpretations and sharing of information could lead to inefficient outcomes in a collaborative network.

In order to investigate the impact of non-uniform information access scenarios in collaborative search, Htun et al. [15] conducted a simulated evaluation using a number of established MLCIR scenarios and found that there is a level of tolerance to removing access to a document collection. Although in general there was a negative impact, Htun et al. argued that non-uniform information access may not always result in a negative impact on performance. However, their findings were based on simulated users and did not go as far as a user evaluation. Handel and Wang [13] put forward a number of design considerations for MLCIR systems, but their suggestions were not based on an empirical evaluation but were rather based on Handel and Wang's experience in Boeing and use cases from different domains. Besides, a fundamental difference in assumption between CIR and MLCIR means that not every CIR component and concept may be directly applicable to MLCIR, e.g. most components in CIR systems [12, 22] may allow users to share search results within their team without restrictions, which for MLCIR systems should be modified. To the best of our knowledge, no user evaluations have been conducted to examine the impact of MLCIR and to provide user-centred design recommendations for future development of MLCIR systems. In order to address this shortcoming, we conducted a controlled user evaluation using three different information access scenarios and pairs of participants. To be able to provide user-centred design recommendations, we undertook a design interview post evaluation using the same pairs of participants.

3. EXPERIMENTAL SETUP

3.1 Document Collection and Search Tasks

For information sources for our evaluation, there were 2 alternative options available: using the web (similar to [2, 22, 25]) or test collections (similar to [16, 26]). As our intention in this study was to remove access to documents and terms from the information source, utilising a test collection was more practical. Besides, this approach lets us precisely calculate traditional Information Retrieval (IR) evaluation metrics such as *precision*, *recall* and *f-measure*. We used the TREC HARD 2005 [1] test collection (which uses AQUAINT corpus), which has also been utilised by a number of other researchers for CIR evaluations [6, 15-17]. The AQUAINT corpus contains a total of 1,033,461 documents (about 3 GB) of newswire text data written in English which were acquired from three news services: the Xinhua News Service (People's Republic of China), the New York Times News Service and the Associated Press Worldstream News Service.

Table 1. Summary of our 10 candidate topics

Topic ID	Title
303	Hubble telescope achievements
363	transportation tunnel disasters
383	mental illness drugs
393	mercy killing
397	automobile recalls
448	ship losses
625	arrests bombing WTC
651	U.S. ethnic population
658	teenage pregnancy
689	family-planning aid

In term of search tasks, there are two alternative options available: using a task that is of mutual interest to the participants (similar to [2, 22]) or using tasks that are selected by the researchers (similar to [12, 25]). The TREC HARD 2005 has a total of 50 test topics[1], from which we selected 10 topics; the same topics have been used by Joho et al. [16] and Htun et al. [15] for their respective simulated evaluations of collaborative search performance (the 10 topics are presented in Table 1). Joho et al. [17] also generated a query pool which contains a list of query terms that are likely to be submitted by users for the 10 topics. We were provided with this query pool and utilised this to blacklist search terms for users (see Section 3.2 for details); this was another reason for selecting these 10 topics. Also by using 10 topics, rather than 3 for example, as we have 3 information access scenarios (see Section 3.2), we had a broad selection of topics for users which cover different parts of the TREC HARD 2005 collection. During the evaluation, we randomly selected from the 10 candidate topics and presented these to the participants for each scenario (see Section 3.4 for the study procedure).

[1] http://trec.nist.gov/data/hard/05/05.50.topics.txt

3.2 Information Access Scenarios and Combinations

For their simulated evaluation, Htun et al. [15] devised 4 non-uniform information access scenarios based on Handel and Wang's [13] observations of MLCIR. These 4 scenarios were 1) document removal, 2) random term blacklisting, 3) term blacklisting based on the most frequent terms in a query pool, and 4) term blacklisting based on the most frequent terms in the collection. The document removal scenario represents the case where access to documents in the collection is removed for some team members. For instance, such a scenario may occur when team members have access to different databases [13]. The term blacklisting scenarios represent the cases where individual team members do not find results if they search using certain blacklisted search terms. For instance, such a scenario may occur when documents are classified at the paragraph level [13]. For this study, we selected 2 out of the 4 scenarios from Htun et al. [15] which we believe are the most practical in real life. These 2 scenarios are: 1) document removal and 2) term blacklisting based on the most frequent terms in a query pool (i.e. the most likely search terms) (see Table 2). We believe these 2 scenarios are practical in real life because the MLCIR cases that involve users with access to different databases and blacklisting the most likely search terms are highly probable [13]. In addition to these 2 non-uniform access scenarios, we also included a full access scenario which represents the case where both team members have equal and unrestricted access to the collection. This scenario was considered as an upper bound to compare search performance with the 2 non-uniform access scenarios and is the norm for most CIR evaluation [16, 17, 25]. During the study, pairs of participants were presented with all 3 information access scenarios. To avoid order effects, the order in which the scenarios were presented was rotated using a Latin Square counterbalancing measure.

Table 2. Information access scenarios and access combinations

Code	Information access scenario	Access combination
DR	Document removal: remove access to documents from collection	100%-60%
TR	Term blacklisting: remove access to most frequent terms in query pool	100%-70%
FA	Full access	100%-100%

For the 2 non-uniform access scenarios, Htun et al. [15] also formulated 55 combinations of access for a pair of simulated users with access level to the collection ranging from 10% up to 100% e.g. 10%-10%, 20%-10%, 20%-20%, 30%-10%..., and 100%-100%. Based on Htun et al.'s [15] simulated evaluation, we selected an access combination for each of the non-uniform access scenarios, where search performance began to deteriorate significantly in comparison to the full access combination; these access combinations are: 100%-60% for document removal scenario and 100%-70% for the term blacklisting scenario. Thus, for the document removal scenario (i.e. DR in Table 2), while full access to the collection (i.e. 100%) is available for one team member, 40% of the documents in the collection are randomly removed for the other. In other words, for the second team member, the index was created using 60% of randomly selected documents from the collection (i.e. 620077 documents), at which point search performance significantly deteriorated according to the previous simulation [15]. For the term blacklisting scenario (i.e. TR in Table 2), while full access (i.e. 100%) is available for one team member, 30% of the most frequent terms in the query pool is blacklisted for the other (thus represented as 70% in Table 2) which means the latter team member does not find results if (s)he submits those blacklisted terms. Please note that although it is represented as 70%, team members in the TR scenario had access to any terms from the collection except for the blacklisted ones. One average, there were about 4 blacklisted terms per topic. For the full access scenario (i.e. FA in Table 2), full access is available for both participants. To avoid order effects, full access and non-full access between a pair of team members were rotated for the non-uniform access scenarios. In other words, each team member had a chance to experience having full access and non-full access once each in the 2 non-uniform access scenarios.

3.3 Participants

A total of twenty participants (ten pairs) were recruited for this study from personal contacts, research group and university contacts. This participant pool size is similar to previous experiments in CIR [8, 17, 22, 24]. Participants were assigned partners for the study based on their available timeslots. There were 9 females and 11 males. The average age of participants was 28 ($\sigma = 4.6$), ranging from 22 to 44. Of the twenty participants, 3 used search systems for more than 20 hours per week, 5 between 16 to 20 hours, 5 between 11 to 15 hours, 2 between 6 to 10 hours and 5 less than 6 hours. 14 of the participants reported that they had been involved in a collaborative search at least once. The majority of their collaborative search experiences consisted of working together either with friends, colleagues or a partner mainly using Google, Google Scholar or Google Maps. Each participant received a £10 Amazon voucher for participation.

3.4 Study Procedure

Figure 1 highlights the steps involved in our study. Before the study began, an email containing an information sheet and a link to a demographic questionnaire was sent to individual participants. They were instructed to complete the questionnaire prior to arriving for the study.

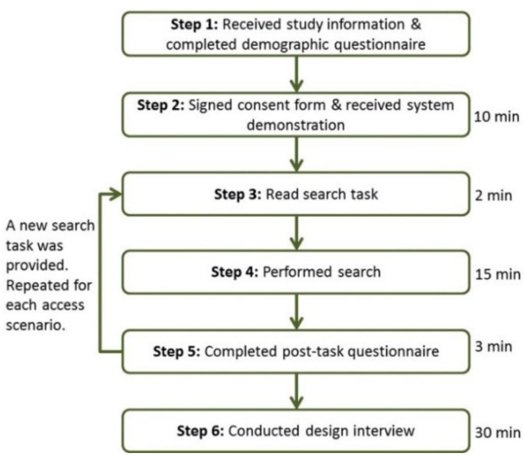

Figure 1. Study procedure

When a pair of participants arrived for the study, they were welcomed, asked to sign a consent form and given a demonstration of the search system. Participants were informed about the information access scenarios but were not told which team member had more (or less) access, nor the order of the information access scenarios presented. They were also instructed not to communicate with each other during the study and to pretend that they were in different rooms. This was to restrain discussion of results and search strategies which could violate the evaluation of MLCIR scenarios. In practice users in MLCIR may not be able to discuss search strategies, documents they can/cannot access or even the results found [13]. A similar approach was used in previous CIR evaluations [17, 22, 25]. Participants were also informed that their goal for each task was to find as many relevant documents as possible for their team. Since team members in MLCIR examples are unaware of their access, limitation and information shareability [13], we believed this strategy was appropriate. They were then given a few minutes to practice searching with the search system.

After practicing with the system, to begin a search session for the first information access scenario, both participants were provided with the same search task randomly selected from the 10 candidate topics (see Table 1). They were allocated 2 minutes to read and familiarise themselves with the task. Participants then performed the search synchronously for a maximum of 15 minutes using different computers in the same room facing opposite directions. This step was followed by a post-task questionnaire. The questionnaire was designed to assess individual participants' perception of the tasks, search performance and interface components (see Table 6). The questionnaire was in the form of 5-point Likert scales and the answers ranged from 1 (strongly disagree) to 3 (neither) to 5 (strongly agree). Once the participants completed the post-task questionnaire, the scenario was changed and the participants were provided with a new randomly selected search task. The same steps were repeated for the remaining scenarios (see Figure 1). A number of previous research in CIR, e.g. [17, 22, 24, 25], used a similar procedure successfully. When all 3 scenarios were completed, a design interview was conducted with both participants. They were briefly reminded about the non-uniform information access scenarios before asking interview questions. A detailed explanation of the design interview is presented in Section 3.6.

3.5 System Description

Although there has been a great deal of research into CIR, there are no user-centred design recommendations relating to MLCIR that we are aware of. Thus, we designed a relatively simple collaborative search system that allows a pair of users to search for and judge documents synchronously. We took our system as a starting point to enable us to empirically evaluate non-uniform information access in CIR which has not been evaluated before. Shah et al. [26] suggested that in order to ensure successful collaboration, the system must have an effective method of communication, an ability to see everyone's actions, a way to distribute tasks and aggregate information, and a mechanism to record user interactions, processes, and results. However, as highlighted by Handel and Wang [13], users in MLCIR scenarios may not be able to discuss search strategies, documents they can/cannot access, or even the results found. To that end, for this study we implemented a system that would not allow direct communication and sharing of results, but that offers an ability to see the query history of the team and documents viewed or judged by the team (see Figure 2). The system was implemented using the Google Web toolkit[2] and the Terrier toolkit[3]. The system consists of 3 main components:

Search component (Figure 2(1)): The search component contains: query-box (Figure 2(a)), search button (Figure 2(b)), result-list (Figure 2(c)) and result-detail (Figure 2(d)). The search

[2] http://gwtproject.org

[3] http://terrier.org

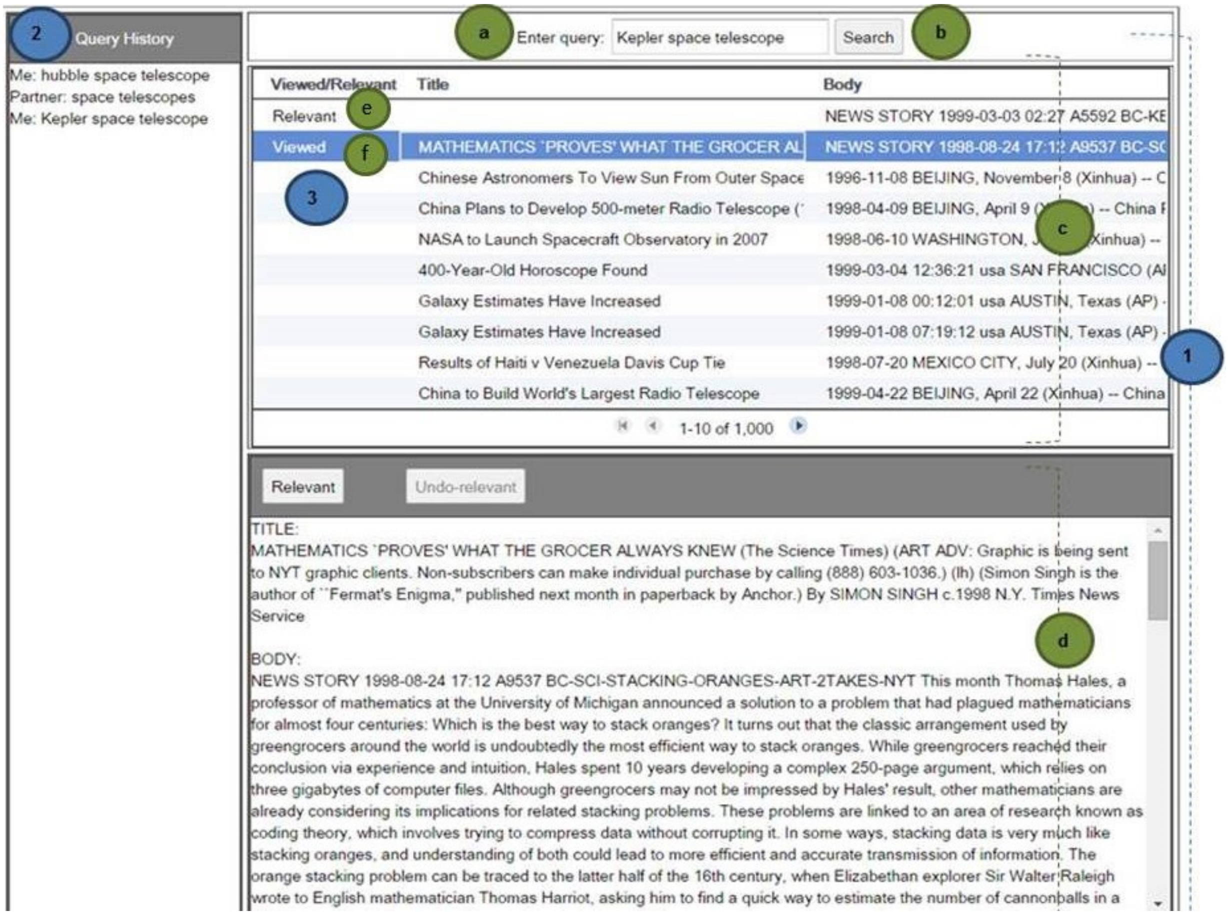

Figure 2. Search system utilised in this study: 1) search component, 2) query history component, 3) viewed/relevant component, a) query-box, b) search-button, c) result-list pane, d) result-detail panel, e) relevant label and f) viewed label

component is the main part of the search system; it allows users to enter queries and results are displayed in the result panels. A number of sample search results are presented in the result-list panel (Figure 2(c)). Clicking on any row in the result-list panel displays full text of the respective document in the result-detail panel below (Figure 2(d)) where users can mark the document as relevant, or un-do the relevant marking. The system ensured that only the documents that are not blocked or the documents that do not contain blacklisted terms appeared in the search results.

Query history component (Figure 2(2))**:** This is provided to enable the participants to gain an awareness of the queries issued by the team. We considered using Integrated Messaging in order to allow communication between the team members. However, in this study, since we wanted to control any discussion of results during the search, we decided to exclude it. Also it has been found that query history is an important component for effective collaborative search [26] and is preferred more by users than text and verbal communication channels [17]. We believed that a query history component may be suitable for non-uniform access scenarios since it provides users with query awareness without the need for direct communication. A list of sample queries entered by the team is shown in Figure 2(2). Queries are displayed in the same way to both team members. There is no information contamination because if the person with lower access submits a blacklisted term, (s)he will not receive any results, and it will not be clear whether this is due to blacklisting or because there are no relevant documents in the collection for that particular query. This is in line with an MLCIR case that was highlighted by Handel and Wang [13]. Some previous CIR systems, e.g. [17, 22, 26], also utilised a query history component successfully.

Viewed/Relevant component (Figure 2(3))**:** This is provided to allow the participants to gain awareness of the documents in their search results that are judged as relevant and/or viewed by the team. Similar to the query history component, this component allows users to obtain awareness of the documents viewed and judged by the team without the need for direct communication. Once a team member views a document (i.e. by clicking a row in the result-list (Figure 2(c))), it is marked as viewed and denoted with the "Viewed" label in the result-list (Figure 2(f)). If the document is marked as relevant by the team member, it is then denoted with the "Relevant" label (Figure 2(e)). Thus, both team members are able to see whether certain documents in their respective result-lists have been viewed or judged as relevant by the team. Previous CIR systems, e.g. [17, 26], also utilised similar components successfully.

3.6 Design Interview

Although a number of design considerations were proposed for MLCIR based on observations [13], none are based on a user-centred design evaluation. Therefore, in order to obtain the participants' feedback about the current search system, and to provide design recommendations for future development of systems that support users in MLCIR scenarios, we conducted a design interview with pairs of participants after the search sessions. The design interview was divided into 2 sections. In the first section, the pairs were asked a series of questions related to their experience with the current search system. We asked the pairs in what way each component (e.g. search component) was easy and difficult to interact with, in what way each component increased or decreased their collaboration capabilities, and how each component could be improved (see Figure 2 for the components). In the second section, the pairs were asked to suggest new components and/or functions that could be added to the current search system in order to improve their collaboration capabilities under non-uniform access. They were also provided with screenshots of the current system and blank sheets of paper to annotate and sketch their design ideas. Throughout the interview, the researcher took notes of participants' responses, and also recorded their responses on an audio recorder, which was later transcribed for analysis.

3.7 Evaluation Metrics

The search system recorded a log of the participants' interactions with the system. This log was used to calculate a number of evaluation metrics in order to compare teams' search outcomes between the information access scenarios (i.e. *DR*, *TR* and *FA*), and to compare individuals' search outcomes between full access and non-full access within the *DR* and *TR* scenarios. We used traditional IR evaluation metrics: *recall*, *precision* and *f-measure*, and also adopted some of the CIR evaluation metrics proposed by Shah and González-Ibáñez [25]: *coverage*, *relevant coverage*, *unique coverage* and *unique relevant coverage*. *Recall* is the number of true positive documents amongst all the documents judged by users divided by the number of relevant documents in the TREC HARD 2005 qrel[4]. *Precision* is the number of true positive documents amongst all the documents judged by the users divided by the number of all of the documents judged by users as relevant. *F-measure* is a harmonic mean of *recall* and *precision* represented by the formula: (2*precision*recall)/(precision+recall). *Coverage* is the number of distinct documents viewed by participants. *Relevant coverage* is the number of documents in the *coverage* that are actually relevant (i.e. true positive). *Unique coverage* is the number of distinct documents viewed by participants only in a given access scenario (e.g. *FA*) and not in any others, or only at a given access level (e.g. full access) and not in the other. *Unique relevant coverage* is the number of documents in the *unique coverage* that are actually relevant. Other CIR evaluation metrics we adopted for this study include those proposed by Soulier et al. [29]: *number of queries* and *query length*. *Number of queries* is the total number of queries submitted by participants. *Query length* represents the average number of terms within the total queries submitted.

In addition, post-task questionnaires were administered to assess individual participants' perception of the tasks, search performance and interface components (see Table 6 for questions). The questionnaire was in the form of 5-point Likert scales and the answers ranged from 1 (strongly disagree) to 3 (neither) to 5 (strongly agree).

4. RESULTS

We compared collaborative search outcomes between the information access scenarios (i.e. *DR*, *TR* & *FA*) and individual search outcomes between full access and non-full access within the *DR* and *TR* scenarios. To compare between the information access scenarios, we used one-way repeated measures ANOVA for normally distributed data, and Friedman and Wilcoxon Sign-Rank tests for non-normally distributed data since the participants within each group were subjected to all three information access scenarios (i.e. data is paired). To compare between full access and non-full access, as data is unpaired we used one-way between groups ANOVA for normally distributed data, and Mann-Whitney U test for non-normally distributed data.

4.1 Search Performance

Recall, *precision* and *f-measure* were not significantly different between teams within 3 the information access scenarios (RQ1) or between individuals with within the DR and TR scenarios (RQ2). A summary is given in Table 3. It appears that the non-uniform information access scenarios we utilised did not have a negative impact on search performance. This was not expected since Htun et al.'s [15] simulated evaluation showed a negative impact on search performance for the same access combinations we used. We discuss this further in Section 5. We also found that despite the low *recall*, participants were able to judge the documents accurately hence the *precision* was high in most cases (see Table 3). To understand the *best possible recall* that the participants could have achieved, we further analysed a new metric calculated as *relevant coverage* divided by the number of relevant documents in the qrel for a given task. As shown in Table 3, the distribution between *recall* and *best possible recall* was close to each other, demonstrating the participants' high judgement accuracy again.

Table 3. Search performance metrics between individuals within the DR and TR scenarios. S.D = standard deviation.

		DR		TR	
		Full	Non-full	Full	Non-full
Recall	Mean	0.013	0.036	0.044	0.043
	S.D	0.019	0.037	0.057	0.072
Best possible recall	Mean	0.018	0.054	0.047	0.055
	S.D	0.029	0.062	0.059	0.082
Precision	Mean	0.344	0.540	0.458	0.328
	S.D	0.410	0.473	0.431	0.304
F-measure	Mean	0.024	0.066	0.077	0.072
	S.D	0.035	0.067	0.097	0.113

4.2 Query Submission

Number of queries and *query length* [29] showed that in terms of query submission there was no significant difference between teams within the information access scenarios (RQ1) or between individuals within the *DR* and *TR* scenarios (RQ2). This suggests that the participants submitted a similar number of queries and average query words across all search sessions (see Table 4). To understand this further, we also analysed *term diversity* and *term overlap* between the individuals within the DR and TR scenarios, and *average blacklisted terms* between the individuals within the TR scenario. *Term diversity* is the number of terms submitted by individuals in a particular access (e.g. full access within the TR scenario). In contrast, *term overlap* is the number of terms submitted by individuals in both full access and non-full access. *Average blacklisted terms* is the average number of terms submitted by individuals that intersect with the blacklisted terms. In general, the participants submitted twice as many diverse terms as the overlapping ones (see Table 4) suggesting that the participants actively tried to avoid duplicated work. Looking at *query length* and *average blacklisted terms*, we found that roughly 20% of the submitted terms were blacklisted.

[4] http://trec.nist.gov/data/hard/05/TREC2005.qrels.txt

Table 4. Query submission metrics between individuals within the DR and TR scenarios. S.D = standard deviation.

		DR		TR	
		Full	Non-full	Full	Non-full
Number of queries	Mean	13.800	11.100	11.900	10.000
	S.D	9.102	6.790	7.141	5.600
Query length	Mean	3.736	3.785	3.500	3.873
	S.D	1.485	1.494	1.976	1.909
Term diversity	Mean	12.600	11.400	11.600	8.300
	S.D	9.766	8.435	10.035	6.447
Term overlap	Mean	6.300	6.300	5.700	5.700
	S.D	2.710	2.710	2.263	2.263
Average blacklisted terms	Mean	-	-	-	0.789
	S.D	-	-	-	0.601

4.3 Collection Coverage

For *coverage*, *relevant coverage*, *unique coverage* and *unique relevant coverage* [25], no significant difference was found between the information access scenarios (RQ1) and between individuals with full access and non-full access within *DR* and *TR* (RQ2) (see Table 5 for result summary). This again suggests that the non-uniform information access scenarios we utilised did not have a negative impact on collection coverage.

Table 5. Collection coverage metrics between individuals within the DR and TR scenarios. S.D = standard deviation.

		DR		TR	
		Full	Non-full	Full	Non-full
Coverage	Mean	23.400	22.400	18.400	28.700
	S.D	13.277	13.040	11.078	16.627
Relevant coverage	Mean	2.000	5.400	4.800	5.600
	S.D	3.400	5.700	5.453	7.516
Unique coverage	Mean	22.600	21.600	17.500	27.800
	S.D	13.866	12.668	10.659	17.165
Unique relevant coverage	Mean	1.800	5.200	4.400	5.200
	S.D	3.155	5.534	5.254	7.495
Relevant documents marked	Mean	4.200	4.200	5.900	9.400
	S.D	5.116	4.467	7.141	10.091

Coverage and *unique coverage* were similar in all cases suggesting that the participants mostly looked at the new documents that were not already viewed. *Relevant coverage* and *unique relevant coverage* were also similar because of the way they were calculated. Looking at the ratio between *coverage* and *relevant coverage*, we found that only about 18% of all the documents viewed by participants were actually relevant. We further looked at the number of *relevant documents marked* by the participants and found that it was very similar to the number of *relevant coverage* (see Table 5). Since the *precision* was high (see Table 3), we can conclude that the participants judged the documents quite accurately.

4.4 Participants' Perceptions

The participants' perceptions of search tasks, search performance and interface components were captured by post-task questionnaires (see Table 6 for questions, and Tables 7 and 8 for result summary). Questions Q1-Q3 measured participants' perceptions of search tasks. Q4 - Q8 measured participants' perceptions of search performance. Questions Q9-Q11 measured participants' perceptions of certain interface components of our current system. Statistical analysis results showed that none of the questions had significantly different scores between the information access scenarios (RQ1) and between individuals with full access and non-full access within the DR and TR scenarios (RQ2). Looking at the results from Tables 7 and 8, we found that participants generally thought the search tasks were easy to understand (Q1) and interesting (Q2). Participants neither agreed nor disagreed with the questions regarding task familiarity (Q3), result satisfaction (Q4) and perception of judging more documents than partner (Q7). Question regarding relevant documents found (Q5) had relatively low score. This may partly be because the candidate topics we utilised only have 86 to 152 relevant documents out of over a million documents in the AQUAINT corpus [16]. Participants were generally confident with the documents they judged (Q6) and thought that their teams performed well in judging documents (Q8). The scores for query history (Q9) and viewed/relevant components (Q10 & Q11) were generally high.

Table 6. Post-task questionnaire questions. Q1-Q3 = assessment of search tasks. Q4-Q8 = assessment of search performance. Q9-Q11 = assessment of interface components.

Q1	The description and narrative of this task were easy to understand.
Q2	The topic of this task is very interesting.
Q3	I was quite familiar with the topic of this task.
Q4	I am satisfied with the documents obtained for my queries for this task.
Q5	I found a lot of relevant documents in the result for this task.
Q6	I am confident with the documents I judged for this task.
Q7	I think I judged more documents than my partner for this task.
Q8	I think my team did a great job in judging documents for this task.
Q9	I think the 'query history' component increased my performance for this task.
Q10	I think being able to know the documents that have already been viewed increased my performance for this task.
Q11	I think being able to know the documents that have already been judged increased my performance for this task.

4.5 Design Interview

A thematic analysis [5] of the transcribed data resulted in 4 main themes: search component, team member awareness, query awareness and result awareness. We present these themes in detail in the following sub-sections. Overall, we obtained 15 design suggestions from the participants: 2 suggestions for the search component, 4 suggestions regarding team member awareness, 5 suggestions regarding query awareness and 4 suggestions regarding result awareness.

4.5.1 Search Component

With regards to the search component, 6 (30%) of the participants suggested adding options to be able to narrow down search results and/or to manipulate the queries. For example, one participant said, "*I'd also like a function to narrow down the results, like from 1000 documents to 500 documents.*" (P11). Another participant mentioned, "*I think it'd be great if you could specify exclusion/inclusion criteria in a query, like what keyword to include and what keyword to exclude.*" (P12). In order to narrow down the search results, we could consider adding a filter function

Table 7. Mean, standard deviation and median of post-task questionnaire result between the 3 information access scenarios (see Table 6 for the questions). 1 = strongly disagree, 3 = neither, 5 = strongly agree. Q1 to Q3 = assessment of search tasks. Q4 to Q8 = assessment of search performance. Q9 to Q11 = assessment of interface components.

		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
DR	Mean	4.3	3.6	2.8	3	2.3	3.7	3	3.4	3.7	4	4
	S.D	0.9	0.9	1.4	1.3	1.2	0.9	1.5	0.9	0.8	0.9	0.8
	Median	4	4	3	3	2	3.5	3	3	4	4	4
TR	Mean	4.6	4	3.4	3.1	3	3.6	3.1	3.6	3.7	4	4.1
	S.D	0.5	1	1.4	1.4	1.7	1.4	1.1	0.9	1.2	0.8	0.9
	Median	5	4	3.5	4	3	4	3	4	4	4	4
FA	Mean	4.3	3.5	3	2.8	2.7	3.8	2.9	3.4	3.5	4.1	4.2
	S.D	1.1	1.2	1.5	1.3	1.3	1.2	1	1	0.8	0.8	0.9
	Median	5	4	3.5	3	2.5	4	3	4	4	4	4

Table 8. Mean, standard deviation (S.D) and median of post-task questionnaire result between individuals with full access and non-full access (see Table 6 for the questions). 1 = strongly disagree, 3 = neither, 5 = strongly agree. Q1 to Q3 = assessment of search tasks. Q4 to Q8 = assessment of search performance. Q9 to Q11 = assessment of interface components.

			Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
DR	Full	Mean	4.5	3.4	3	3.2	2.4	3.8	3.3	3.1	3.8	4.1	3.9
		S.D	0.5	1.1	1.4	1.3	1.4	0.8	1.6	1	0.6	0.7	0.6
		Median	4.5	3.5	3.5	3	2	4	3	3	4	4	4
	Non-full	Mean	4	3.7	2.6	2.7	2.2	3.5	2.6	3.6	3.5	3.9	4
		S.D	1.1	0.8	1.4	1.3	1.1	1	1.3	0.7	1	1	0.9
		Median	4	4	2.5	2.5	2	3	3	3.5	4	4	4
TR	Full	Mean	4.5	3.9	3.3	2.5	2.5	3.3	2.7	3.5	3.3	3.9	4.1
		S.D	0.5	1	1.3	1.6	1.7	1.7	1.3	0.9	1.5	1	1.3
		Median	4.5	4	4	2	2	4	2.5	3.5	2.5	4	4.5
	Non-full	Mean	4.7	4.1	3.4	3.6	3.5	3.8	3.4	3.7	4	4	4
		S.D	0.5	1	1.5	1	1.7	1	1	1	0.7	0.5	0.5
		Median	5	4	3	4	4	4	3.5	4	4	4	4

such as filter by date, etc. For the latter, we could consider allowing the use of inclusion/exclusion keywords such as AND, OR, etc. Although these components are not new for IR and CIR systems, they could be as effectively used in MLCIR systems too. However, to be able to take advantage of these functions, we believe that users may require certain expertise or training.

4.5.2 Team Member Awareness

It appears that users in MLCIR could benefit from being able to easily identify team members and their actions. For the query history component, 12 (60%) of the participants suggested differentiating queries submitted by each person either with different colours (7 participants) or by separating them into a different list (5 participants). For the viewed/relevant component, 13 (65%) of the participants suggested differentiating documents viewed/judged by each person. Amongst them, 5 participants suggested the use of a different colour for each person and 2 participants suggested the use of text such as "by A" or "by B" whilst 4 participants suggested the use of both approaches and 2 suggested no particular approach. One of the participants who suggested both approaches commented: "*When there's more than two people, using names instead of colour would be much better.*" (P19). A previous collaborative search system, CoSearch [2] used both colours and names to identify each person. SearchTogether [22] used profile pictures and names for the same purpose. We wish to suggest here that MLCIR systems could also adopt a similar approach to provide team member awareness.

4.5.3 Query Awareness

11 (55%) participants mentioned that being aware of their partners' queries had given them an idea of what might be a good new query. One participant explained: "*I didn't know much about [a topic], so I looked at [partner's] search history and tried to search for similar keywords.*" (P17). This means that query awareness could help users share their expertise without the necessity to disclose sensitive information. Participants also suggested a number of improvements regarding query awareness. 5 (25%) of the participants suggested that the query history component should display the number of viewed and relevant documents found for a particular query. In addition to this, 2 (10%) of the participants added that effectiveness of a query could be displayed using "hot and cold icons" next to the queries in the query history component (see Figure 3 for an illustration by a participant).

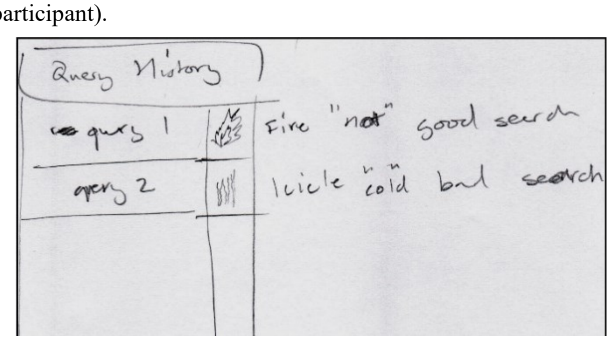

Figure 3. Illustration of hot and cold icons in the query history component (P1)

Such icons may help to identify the effectiveness of a query quickly. Moreover, displaying the time spent on an individual query was suggested by 8 (40%) of the participants. For example, one participant explained: "*The time you spent on the query is an indication of how successful your query is.*" (P13). While this explanation is correct to some extent, the time spent alone cannot always indicate the success of a query. In practice, a combination of different query properties should be used. Displaying query popularity (i.e. the number of times a query has been used) was also suggested by 2 (10%) of the participants. In addition, one of the participants suggested adding a function to sort the query history according to time issued and also according to the effectiveness. Overall, query properties such as query effectiveness, query popularity and time spent on query, as well as a query sorting function maybe important for users to able to obtain better query awareness and to work effectively in non-uniform access scenarios.

4.5.4 Result Awareness

3 (15%) of the participants mentioned that being aware of their partners' judged documents could help them double-check whether the documents are actually relevant. In addition, 2 (10%) other participants noted that being aware of their partners' judged documents could give them an idea of what might be a related new query. One participant explained: "*...if I know that this document is relevant, then I can read it and can go on to another keyword which is related to that document.*" (P18). Therefore, similar to query awareness, being aware of a partner's judged documents could help users share their expertise without the necessity to disclose sensitive information. Participants also suggested a number of improvements with regards to results awareness. 9 (45%) of the participants thought that their viewed and/or judged documents should be easily accessible. 3 (15%) of the participants suggested displaying a list of relevant documents whereas 6 (30%) other participants suggested displaying two different lists for both viewed and relevant documents. Moreover, it was suggested by 6 (30%) of the participants that the results should also be sortable based on their viewed and relevant properties. In addition, 4 (20%) of the participants suggested adding a favourite function that would allow them to keep the documents they want to read later. A favourite function has been implemented previously by Shah et al. [26] and Freyne et al. [8] in their respective systems. MLCIR systems could adopt a similar function to provide users with an option to bookmark documents.

5. DISCUSSION

In relation to our first research question (RQ1): "What are the effects of non-uniform information access scenarios on collaborative search outcomes?" it was found that neither of the two non-uniform access scenarios we investigated had a significant negative impact on search performance or collection coverage. While one might expect a decrease in search performance and collection coverage for non-uniform access scenarios compared to full access scenario, the performance was unchanged. This is different from the findings of Htun et al.'s [15] simulated evaluation where search performance deteriorated significantly at the access combinations 100%-60% for the document removal scenario (*DR*) and 100%-70% for the term blacklisting scenario (*TR*). This difference could be because we used human participants in our study, and humans can compensate for each other in ways that cannot be easily simulated.

In relation to our second research question (RQ2): "What are the effects of non-uniform information access scenarios on individual search outcomes?" we found that the individuals with non-full access performed as well as those with full access for both DR and TR scenarios. It appears that even removing 40% of the documents from the collection (i.e. for DR) and removing 30% of the most likely search terms (i.e. for TR) did not have a negative impact on individuals' search outcomes. In terms of query submission, individual participants submitted a similar number of queries and average words. It appears that since the participants were unaware of their access, they carried on regardless.

In relation to our third research question (RQ3): "How can MLCIR interfaces be designed so that they are usable and useful for the end users?" we discuss our findings from the design interview. In general, our findings 1) confirm the potential of the application of some existing CIR design suggestions to MLCIR, and 2) provide important new design suggestions for MLCIR systems. Awareness seems to play a potentially large role in MLCIR systems, perhaps because being aware of a team member's queries and judged documents could help users share their expertise without disclosing sensitive information. This confirms that whilst awareness has been an important concept in CIR, it may be applicable in MLCIR too. However, MLCIR systems must ensure that providing awareness for users does not compromise any sensitive information. With this in mind, we discuss potential interface components for MLCIR systems. MLCIR systems could provide query properties such as effectiveness, popularity and time spent, and also allow for sorting of queries according to their properties. We believe that these query properties could be implemented by adding different icons and a tooltip function in the query history component. A similar approach has been utilised by Freyne et al. [8] in an attempt to integrate search and browsing functionality. It has also been shown that looking at high-quality query examples can help users create queries that are highly effective [14]. Whilst icons could provide quick access to query properties, the tooltip function could provide further information such as number of viewed/relevant documents, number of times submitted, etc. Handel and Wang [13] highlighted that MLCIR could impact negatively on collaboration since team members are unaware of individual's access limitation. To address this, we believe MLCIR systems could display each team member's blacklisted keywords to those with higher access where appropriate. Thus, team members with higher access can compensate for those with lower access. In addition, like certain CIR systems [11, 22], we believe MLCIR systems could provide easy access to viewed/judged documents of team members. This can be implemented by adding separate lists for viewed and judged documents. However, the system must ensure that these lists do not accidently disclose unnecessary information by regularly checking for blacklisted terms and/or documents for every team member. The systems could also allow sorting of search results by means of their viewed and relevant properties. Moreover, during the search, it is important for users to be able to easily identify each team member's role and their actions. Thus, individual team members could be differentiated using different colours and/or names similar to [2, 22].

5.1 Limitation

Whilst utilising the TREC HARD 2005 collection [1] and its 10 test topics (see Table 1) allowed us to perform a controlled user-evaluation, perhaps our evaluation results could be slightly different for another document collection. However, the TREC HARD 2005 collection [1] is well known and has been utilised widely for various CIR evaluations [6, 15-17].

6. CONCLUSION AND FUTURE WORK

In this paper, we presented a user study of non-uniform information access in CIR, the first such study to the best of our knowledge. For our evaluation, we utilised 2 non-uniform access scenarios (document access removing and search term blacklisting) and 1 full access scenario, together with a collaborative search interface. We compared collaborative search outcomes between the information access scenarios and individual search outcomes between full access and non-full access. Findings from this evaluation suggests that removing 40% of the documents from the collection or removing 30% of the most likely search terms did not have a negative impact on search outcomes. Finally, we discussed design recommendations for the development of new interface components that support MLCIR, based on feedback from the design interview. Findings from the design interview suggested that being aware of team members' submitted queries and as well as viewed and judged documents (i.e. of allowed documents) could help users share their expertise without disclosing sensitive information. Also, we suggested that MLCIR systems must be aware of each user's role, action, access and limitation, and must ensure that providing awareness for users does not compromise any sensitive information. In the future, we intend to implement and test the elicited design recommendations through a number of user studies, in particular focusing on query awareness, result awareness and team awareness. Whilst these concepts have been examined for CIR, the nuanced differences with MLCIR mean more research is required for these concepts to be applicable in all CIR scenarios, not just when access to information is equal amongst a team of searchers. Overall, we believe that our paper has contributed to both HCI and IR communities by providing important design recommendations to assist with the future development of systems that better support users in MLCIR scenarios. We also believe that our paper has laid the groundwork for more research into MLCIR, which was initially laid out by Handel and Wang [13], by 1) confirming the potential for application of some existing CIR principles, an important contribution in itself; and 2) by providing new suggestions for developing MLCIR systems based on user feedback and empirical evaluation.

7. ACKNOWLEDGEMENTS

Thank you to Leif Azzopardi for his helpful comments on multiple versions of this paper.

8. REFERENCES

[1] Allan, J.: HARD Track overview in TREC 2005 high accuracy retrieval from documents. *NIST TREC 2005,* 1-17.

[2] Amershi, S. and Morris, M. R.: CoSearch: a system for co-located collaborative web search. *ACM SIGCHI 2008*, 1647-1656.

[3] Attfield, S., Blandford, A., Makri, S.: Social and interactional practices for disseminating current awareness information in an organisational setting. *IPM*, 46(6), 2010, 632-645.

[4] Bjurling, B., Hansen, P.: Contracts for Information Sharing in Collaborative Networks. *ISCRAM 2010*, 1-5.

[5] Braun, V. and Clarke, V.: Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 2006, 77-101.

[6] Capra, R., Chen, A. T., Hawthorne, K., Arguello, J., Shaw, L. and Marchionini, G.: Design and evaluation of a system to support collaborative search. *ASIS&T*, 49(1), 2012, 1-10.

[7] Foley, C. and Smeaton, A. F.: Division of labour and sharing of knowledge for synchronous collaborative information retrieval. *IPM*, 46(6), 2010, 762-772.

[8] Freyne, J., Farzan, R., Brusilovsky, P., Smyth, B. and Coyle, M.: Collecting community wisdom: integrating social search & social navigation. *ACM IUI 2007*, 52-61.

[9] Golovchinsky, G., Adcock, J., Pickens, J., Qvarfordt, P. and Back, M.: Cerchiamo: a collaborative exploratory search tool. *ACM CSCW 2008*, 8-12.

[10] Golovchinsky, G., Pickens, J. and Back, M.: A taxonomy of collaboration in online information seeking. *arXiv preprint*, arXiv:0908.0704, 2009.

[11] González-Ibáñez, R. and Shah, C.: Coagmento: A system for supporting collaborative information seeking. *ASIS&T*, 48(1), 2011, 1-4.

[12] Halvey, M., Vallet, D., Hannah, D., Feng, Y., Jose, J. M.: An asynchronous collaborative search system for online video search. *IPM*, 46(6), 2010, 733-748.

[13] Handel, M. J., Wang, E. Y.: I can't tell you what i found: problems in multi-level collaborative information retrieval. *ACM CIR 2011*, 1-6.

[14] Harvey, M., Hauff, C. and Elsweiler, D.: Learning by Example: training users with high-quality query suggestions. *ACM SIGIR 2015*, 133-142.

[15] Htun, N. N., Halvey, M. and Baillie, L.: Towards Quantifying the Impact of Non-Uniform Information Access in Collaborative Information Retrieval. *ACM SIGIR 2015*, 843-846.

[16] Joho, H., Hannah, D., Jose, J. M.: Revisiting IR techniques for collaborative search strategies. *ECIR 2009*, 66-77.

[17] Joho, H., Hannah, D., Jose, J. M.: Comparing collaborative and independent search in a recall-oriented task. *ACM IIiX 2008*, 89-96.

[18] Karunakaran, A., Reddy, M.: Barriers to collaborative information seeking in organizations. *ASIS&T*, 49(1), 2012, 1-10.

[19] Karunakaran, A., Reddy, M.: The Role of Narratives in Collaborative Information Seeking. *ACM GROUP 2012*, 273–276.

[20] Linden, G., Smith, B. and York, J.: Amazon. com recommendations: Item-to-item collaborative filtering. Internet Computing. *IEEE*, 7(1), 2003, 76-80.

[21] Morris, M. R.: Collaborative search revisited. *ACM CSCW 2013*, pp. 1181-1192.

[22] Morris, M. R., Horvitz, E.: SearchTogether: an interface for collaborative web search. *ACM UIST 2007*, 3-12.

[23] Morris, M. R., Lombardo, J. and Wigdor, D.: WeSearch: supporting collaborative search and sensemaking on a tabletop display. *ACM CSCW 2010*, 401-410.

[24] Pickens, J., Golovchinsky, G., Shah, C., Qvarfordt, P., Back, M.: Algorithmic mediation for collaborative exploratory search. *ACM SIGIR 2008*, 315-322.

[25] Shah, C., González-Ibáñez, R.: Evaluating the synergic effect of collaboration in information seeking. *ACM SIGIR 2011*, 913-922.

[26] Shah, C., Marchionini, G. and Kelly, D.: Learning design principles for a collaborative information seeking system. *ACM CHI EA 2009*, 3419-3424.

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

[28] Smyth, B., Balfe, E., Freyne, J., Briggs, P., Coyle, M. and Boydell, O.: Exploiting query repetition and regularity in an adaptive community-based web search engine. *UMUAI*, 14(5), 2004, 383-423.

[29] Soulier, L., Shah, C. and Tamine, L.: User-driven system-mediated collaborative information retrieval. *ACM SIGIR 2014*, 485-494.

[30] Tamine, L. and Soulier, L.: Understanding the Impact of the Role Factor in Collaborative Information Retrieval. *ACM CIKM 2015*, 43-52.