Information Processing and Management 46 (2010) 762–772

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

# Division of labour and sharing of knowledge for synchronous collaborative information retrieval

Colum Foley *, Alan F. Smeaton

*Clarity: Centre for Sensor Web Technologies, Dublin City University, Dublin 9, Ireland*

ARTICLE INFO

*Article history:*
Received 15 May 2009
Received in revised form 15 September 2009
Accepted 30 October 2009
Available online 16 December 2009

*Keywords:*
Information retrieval
Collaborative information retrieval
Collaborative search

ABSTRACT

Synchronous collaborative information retrieval (SCIR) is concerned with supporting two or more users who search together at the same time in order to satisfy a shared information need. SCIR systems represent a paradigmatic shift in the way we view information retrieval, moving from an individual to a group process and as such the development of novel IR techniques is needed to support this. In this article we present what we believe are two key concepts for the development of effective SCIR namely *division of labour* (DoL) and *sharing of knowledge* (SoK). Together these concepts enable coordinated SCIR such that redundancy across group members is reduced whilst enabling each group member to benefit from the discoveries of their collaborators. In this article we outline techniques from state-of-the-art SCIR systems which support these two concepts, primarily through the provision of awareness widgets. We then outline some of our own work into system-mediated techniques for division of labour and sharing of knowledge in SCIR. Finally we conclude with a discussion on some possible future trends for these two coordination techniques.



## 1. Introduction

Collaborative information retrieval refers to systems and technologies which support collaboration between people during the information retrieval (IR) process. This research domain is placed at the convergence of the information retrieval and computer-supported cooperative work (CSCW) fields. CSCW systems are often categorised based on their position in the two dimensional space of time and place (Ellis, Gibbs, & Rein, 1991). Using this taxonomy, most of the work to date in collaborative information retrieval refers to systems and techniques which support an asynchronous, passive, implicit form of collaboration between remote users. The aim of these asynchronous collaborative IR systems is to improve the IR process for an individual by using the past experiences of others. Collaborative filtering systems (Greenberg & Roseman, 1996) attempt to provide serendipitous recommendations of items (documents, books, DVDs, etc.) to users based on the opinions of others, where opinion is expressed as implicit or explicit ratings. Collaborative re-ranking systems use click-through data associated with previous users' queries in order to re-rank a ranked list for a user (Smyth et al., 2005). Collaborative footprinting (Ahn, Brusilovsky, & Farzan, 2005) systems attempt to contextualise a user's search process by showing the *trails* left by previous users through the information space, such as showing a popularity icon indicating those links most often followed, or the comments associated with results on a ranked list. Recently with the ever growing emphasis on the *sociable* web we have begun to see the emergence of commercial collaborative or *social* search systems on the web. Collaborative bookmarking websites such as Del.icio.us. (2009) allow users to tag and share webpages with others. Heystaks (2009) enables groups of users to arrange similar searches into *staks* so that group members can see those pages found by their friends.

* Corresponding author.
*E-mail address:* Colum.Foley@computing.dcu.ie (C. Foley).

0306-4573/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ipm.2009.10.010

Although these systems are termed "collaborative", quite often the users involved are unaware that they are collaborating whatsoever. By and large the users of these systems do not know the people they are collaborating with. In essence, the goal of these systems is equivalent to any traditional IR system – to provide effective retrieval of information for an *individual*.

In our work we are interested in exploring the potential for synchronous, explicit, real-time collaborative search in which two or more users come-together in a focussed group, in either a co-located or remote setting, in order to satisfy a shared information need. We refer to these systems and technologies as synchronous collaborative information retrieval (SCIR) to emphasise the nature of the collaboration in this setting. SCIR systems represent a significant paradigmatic shift in how we view information retrieval, from an individual to a *group* process. As a result, how we approach IR in terms of information needs, notions of relevance, and ranking for example needs to be re-visited in this setting.

Despite claims that IR has always been a social process (Romano, Roussinov, Nunamaker, & Chen, 1999; Root, 1988; Wilson, 1981) and several studies showing that SCIR is common among groups of people in either a work or social scenario (Foster, 2006; Hansen & Järvelin, 2005; Morris, 2008; Poltrock et al., 2003; Twidale, Nichols, & Paice, 1997), SCIR has received little attention from the broader IR community to-date.

In (Foley, Smeaton, & Lee, 2006) we proposed the concepts of division of labour (DoL) and sharing of knowledge (SoK) for SCIR. Both DoL and SoK have long been recognised as important items in successful CSCW systems (Ellis et al., 1991; Kuutti, 1991; Poltrock et al., 2003; Yao, Neches, Ko, Eleish, & Abhinkar, 1999). In our work we are interested in exploring their application to an information retrieval environment involving two or more users. DoL and SoK operating in such an SCIR environment have the potential to allow for a more effective group search by reducing the redundancy across collaborating searchers whilst enabling users to benefit from the experiences and discoveries of their collaborators.

The majority of published work in SCIR has focussed on improving the *awareness* across groups of remote collaborating searchers through the provision of various awareness cues. This is not surprising considering that awareness has been *the* central issue in designing systems to support CSCW for many years (Benyon, Turner, & Turner, 2004; Cockburn & Greenberg, 1993; Dourish & Bly, 1992; Gutwin et al., 2008; Root, 1988; Want, Hopper, Falcao, & Gibbons, 1992). By providing users with a perception of their collaborator's activities during the search, the idea is that the users themselves can then co-ordinate the search task. An important design consideration when designing CSCW systems, however, is the trade-off between providing adequate awareness and disrupting individuals work (Benyon et al., 2004). Indeed as noted by Adcock et al. (2007) requiring users to both search and coordinate the group activity may cause searchers to suffer from cognitive overload.

In this article we will look at how DoL and SoK can be realised in SCIR systems. Firstly, in Section 2 we will describe field studies which have observed SCIR occurring *in the wild* and highlight how DoL and SoK occurs often despite the lack of provision of tools to enable effective collaboration. In Section 3 we will outline how a DoL and SoK can be realised in systems to support synchronous collaborative search, here we will draw on previous work from the general CSCW literature before outlining what they mean for the domain of SCIR. In Section 4 we will outline attempts made to enable a user-driven coordination of SCIR via the provision of awareness widgets, and show how such techniques can enable a DoL and SoK. Following that, in Section 5 we will outline our own work in implementing *system-mediated* techniques to support DoL and SoK in SCIR, where we attempt to move the burden for coordination of the search from the users to the back-end search system in order to reduce the cognitive load on collaborating searchers. We will present our implementation of system-mediated techniques to realise both DoL and SoK for search. We will conclude the article with a discussion on some possible future trends for both concepts.

## 2. Information retrieval: a social process

Researchers claim that IR has always been a social process (Romano et al., 1999; Root, 1988; Wilson, 1981) and studies have shown how collaboration is commonplace throughout the information retrieval process, despite little provision being made for collaborative tools within IR systems (Morris, 2007). The collaborative nature of the information retrieval process has been investigated in a number of domains including: academia, industry, medicine, and the military (Foster, 2006).

Twidale et al. (1997) observed collaboration in activities around 11 OPAC computer terminals in a university library. In the study they found that 10% of all uses of the terminals were of a collaborative nature including: multiple users working around the same terminal in a problem solving task and pointing and gesturing at the screen, users working on adjacent terminals and coordinating their actions, and users asking others questions such as "How do you do that?". They noted that this rich collaboration occurs despite the fact that talking in the library was discouraged and that terminals and information systems were designed for single usage.

In a study of information sharing behaviour across four academic disciplines, Talja (2002) demonstrated that collaborative information retrieval is as commonplace as individual or solo information retrieval.

In a long term study of the collaborative information retrieval activities in design teams at Boeing and Microsoft, Poltrock et al. (2003) observed collaboration occurring at all stages in the information retrieval process (identifying an information need, query formulation, retrieving information). Based on interviews, observations and records of meetings and emails, they concluded that a division of labour was vital to the coordination of the team's work, enabling individuals to perform parts of the work alone, and that a shared information space, where users could share ideas and ascertain information, was a central element of any collaboration system.

In a study of the activities involved in resolving a patent application, Hansen and Järvelin (2005) observed that collaboration amongst patent engineers was common in most phases of the patent task with information seeking the most common. They noted that overlap was common across patent searches and that this led to sharing of information such as documents, queries, and opinions. Hansen and Järvelin also observed patent engineers cooperating on work tasks, dividing the task amongst colleagues and sharing search strategies.

In a survey of web search usage amongst workers at a large corporation, Morris (2007) found that collaboration was commonplace in web search despite it not being supported in search systems. In the survey, over 97% of all users reported having used some form of collaboration when searching the web. This included: 87.7% of people saying that they had looked over someone's shoulder while they searched and suggested query terms, 30.4% of people who said that they had used instant messaging (IM) to coordinate a real-time search, and 23.5% who had used a large display to perform a web search during a group meeting. For typical group size for collaboration: 80.7% of respondents had collaborated in groups of two, with 19.3% reporting a group size of three or four (no larger group sizes were reported). When multiple users collaboratively searched an information space together using separate devices, Morris (2007) identified two common task division strategies: (1) *Divide and conquer*, where the coordination of the task involved assigning separate subtasks to each individual, which was used by 19.3% of respondents, and (2) the *brute-force* approach where no coordination took place and users searched separately was used by 24.8% of respondents. The common types of task for which collaboration took place were purchasing items, booking travel, and researching property.

From these different studies, and others reported in the literature (Foster, 2006), we can conclude that collaboration is common in information retrieval, this collaboration can occur at any stage of the IR process, and that collaboration occurs *in spite of* a lack of support for these activities in most IR systems.

## 3. Division of labour and sharing of knowledge for SCIR

Both the division of labour and the sharing of knowledge have long been identified as important aspects of any successful CSCW system (Ellis et al., 1991; Kuutti, 1991; Poltrock et al., 2003; Sharples, 1993; Yao et al., 1999). Both concepts form part of the broader *coordination theory* for CSCW (Malone & Crowston, 1990), a set of principles which define how collaborative work can be supported such that collaborators can work together harmoniously.

### 3.1. Division of labour

Division of labour refers to the process of dividing up the group task across collaborators in order to share the workload across the group. An effective DoL policy should enable the group as a whole to work more efficiently by reducing the redundancy across collaborators thus ensuring that no two users spend time completing the same task unnecessarily.

Some form of DoL can be seen across most examples of CSCW applications in the literature. Furthermore the division of labour across workers has long been studied in workplace. Poltrock and Handel (2009) used both ostensive and performative modelling techniques in order to examine the subdivision of processing in organisations such that it could inform the development of collaborative technologies.

From examining the CSCW literature we find several ways in which a DoL can be achieved. Sharples (1993) describes three strategies for collaborative writing: *parallel*, where the work is divided into subtasks and all collaborators work simultaneously; *sequential*, where a task is divided into stages such that the output from one stage is handed onto the next user; and *reciprocal*, where group members work together mutually adjusting their activities. Roles have also been investigated as a means of dividing work across users working together, where each user assumes responsibility for certain jobs according to their role. Within an organisation, the role of a manager is often to divide the activities needed to complete a project amongst a team of workers, where the roles of individuals in the organisation in terms of their skills and expertise determine the job they are assigned (Poltrock et al., 2003). *Protocols* either technological or social can also be used to divide the task (Ellis et al., 1991). A technological protocol in a co-authoring CSCW application could, for example, lock a section of a document that was being edited by one user such that other users could not change it. Social protocols do not enforce any system-level restrictions but instead rely on etiquette and user–user coordination. The Grove collaborative writing system proposed by Ellis et al. (1991) operated such a social protocol, in which collaborating users had free reign to edit any part of a document at any time. The result was not chaotic but rather, after a learning period, the social protocol was found to be successful at mediating collaboration.

It is clear that a strict division of broader tasks into subtasks to be allocated to each user is more suitable when there is a clear boundaries between the different stages of the task. An information retrieval process is by its nature a very flexible and iterative process: users can rapidly query, view results and examine the contents of documents. As a result the boundaries between the different stages can become difficult to distinguish. Implementing a stage-by-stage division of the task is therefore difficult in modern IR systems (Smeaton, Foley, Gurrin, Lee, & Mc Givney, 2006). As a result, as we will see in Section 4, the vast majority of work to date in SCIR have relied on social-protocols and awareness cues as a means to divide the search task. A DoL in SCIR means dividing the search task in such a way that each member of a collaborating group completes a piece of the search task. When users search to satisfy the same information need, as is the case in SCIR, they often use the same or very similar query terms (Foley & Smeaton, 2009). If searching within the same corpus of documents, this similarity across queries will result in highly similar ranked lists being returned to users. Therefore, without some form of DoL policy in SCIR, users may spend a lot of time replicating the work of their collaborators. There are numerous ways in which a DoL can be achieved in SCIR. For example each user may explore a subset of the information space during the search, either by searching through different collections or searching using different query terms for example. If done effectively, DoL should enable the group to gain a greater coverage of the information space by reducing the redundancy across the group and allowing a greater throughput of document assessment.

### 3.2. Sharing of knowledge

Sharing of knowledge refers to the passing of ideas and information between collaborators during a group activity. The ability to effectively share information has been recognised as one of the foundations of any group activity (Yao et al., 1999). In any group work setting each user will bring their own experiences and expertise to the shared task. Effective SoK policies should allow collaborators to exchange information and ideas seamlessly across the group to enable users to benefit from the expertise of their collaborators.

CSCW applications have attempted to support the sharing of ideas and knowledge through the provision of shared workspaces (Gutwin et al., 2008; Lai, Malone, & Yu, 1988; Poltrock et al., 2003; Tang, 1991; Whittaker, Geelhoed, & Robinson, 1993). Workflows in organisations such as those modelled in Poltrock and Handel (2009) have implicit knowledge sharing between the different stages in the processes as one worker or department passes information on in the chain and often several people or departments need to collaborate in order to produce an output. Informal and spontaneous sharing of knowledge between knowledge workers is seen as critical to the way many organisations operate (Bellotti & Bly, 1996; Gutwin et al., 2008; Poltrock et al., 2003). Knowledge sharing is particularly evident across the intensive patent application domain (Hansen & Järvelin, 2005). Large multi-nationals often have internal systems whereby potential patents applications can be posted on a shared workgroup system for critiquing by colleagues (Dilley, 2008). Such a system can enable fellow patent engineers to discuss the application and to provide information on conflicting patent applications which they may have uncovered in the course of their work. Given the high-cost associated with patent applications, such a corporate-wide knowledge sharing is extremely important in the patent domain.

Gutwin et al. provide an overview of shared workspace technologies (Gutwin et al., 2008). The advantage of using an online medium for communicating ideas and discussions is that it can be accessible across the entire organisation, even to those whom it was not originally intended. Activity Explorer (Muller, Geyer, Brownholtz, Wilcox, & Millen, 2004) allows a shared access to repositories of documents or artifacts in organisations. People can access the shared repository, manipulate the shared object and communicate with each other via an instant messaging style communication where they can chat and send snapshots of their work. Lai et al. (1988) describe a collaborative spreadsheet based system for sharing knowledge between collaborators by sending messages and transferring files. Notecards (Trigg, Suchman, & Halasz, 1986) was an early hyper-text based system for idea generation and structuring. A notecard was an electronic hypertext placeholder containing text and images. Users could share information and critique notecards using annotations. Notecards could be passed between users and each notecard had a history associated with it which contained details of the changes made by participants.

In our work we are interested in exploring the application of sharing of knowledge policies in the context of synchronous collaborative information retrieval. When people come together to search for information, each user will have different knowledge and expertise, for instance, some may be more familiar with the search topic, others may be more familiar with the search tools. Based on their knowledge, users may approach resolving an information need differently, either through the type of search engine they use, the queries they enter, the documents they view and so on. By bringing people together to search at the same time, and enabling a SoK across collaborators, we can enable group members to benefit from the diversity across the group.

An interesting aspect of IR is that it is often a learning process (Ingwersen, 1992). Users can approach an IR engine with a vague notion of their information need and develop this notion through interactions with a search engine. Therefore in SCIR an SoK policy needs to enable users to benefit from the discoveries of their collaborators during the search, in order to ensure that each group member can develop their perceptions of the shared task.

DoL and SoK policies can be achieved either through a user-driven coordination of the search, where users themselves divide the task and share information, or from a system driven coordination of the search task, where the back-end system attempts to coordinate the task across multiple users. To date the vast majority of work in SCIR has focussed on designing for awareness across collaborating searchers in the hope that the users themselves can coordinate the search. In the next section we will review some of this work.

## 4. Designing for awareness in SCIR

SCIR "in the wild" has been observed in co-located settings, where groups of people either arrange themselves around a single computer terminal with one user acting as the driver and other users contributing, or in remote settings in which users work independently on their own browsers and communicate via instant messaging or emailing.

When collaborating over a search task in a co-located setting, users are often working in close proximity, awareness of the activities of others is naturally high but sometimes at the expense of group productivity as users can be distracted by the actions of others. In a remote setting, users can search more independently through their own browser, however in such a setting awareness amongst group members often comes at a high-cost, such as having to email relevant links to colleagues (Morris & Amershi, 2008).

SCIR systems to-date have attempted to improve group productivity in both settings through the provision of awareness cues. The majority of work to-date has been in supporting remote SCIR.

### 4.1. Remote SCIR

GroupWeb (Gianoutsos & Grundy, 1996) is an early example of a remote SCIR system. In Groupweb the web browser was used as a "group presentation tool" in which one master browser selected a webpage and this page was displayed across all collaborators' browsers. Users could attach annotations to webpages and these annotations were shared across all collaborators. GroupWeb represented a tightly coupled collaborative browsing experience, awareness was high across collaborators, however such tight coupling did not allow for a DoL across collaborators as all users are required to view the same information at the same time. The system did provide for a SoK across users via shared annotations and users could use this tool to discuss the contents of a document and share their opinions of it with others.

Other remote SCIR systems were developed to enable a more relaxed form of synchronous search (Cabri, Leonardi, & Zambonelli, 1999; Krishnappa, 2005). The W4 browser (Gianoutsos & Grundy, 1996) contained many of the awareness widgets seen in state-of-the-art SCIR systems including:

- Session history – which showed the URLs of documents seen by group members.
- Shared bookmarks – which showed the URLs of documents marked as relevant to the search by group members.
- Chat window – for communicating ideas.
- Shared whiteboard – for group brainstorming.

Users could also embed chat-sessions, links and annotations directly on any webpage. The relaxed searching experience and awareness rich environment of the W4 browser can enable collaborators to coordinate their search task while searching independently. The group could use W4 to achieve a DoL by assigning certain tasks to each collaborator using the chat mechanism. Furthermore, recording those documents seen by group members in the session history widget can allow users to skip over those documents already viewed by their collaborators. SoK can be achieved through brainstorming ideas across the shared whiteboard, or viewing the annotations made by others on webpages. Users could also benefit from the discoveries of their collaborators through the shared bookmarked area where they could see those documents deemed relevant to the search task. This bookmark widget coupled with the session history is a good example of how DoL and SoK policies can work together in tandem: users can choose to ignore viewing those documents contained in the history conscious that such documents that were seen by others *and* deemed relevant to the search will be saved in the bookmarked area.

SearchTogether (Morris & Horvitz, 2007; SearchTogether, 2009) is an Internet Explorer plugin that allows for remote SCIR across groups of searchers. The system was developed to explicitly support awareness and division of labour across collaborators. *Query awareness* is achieved by showing each group member's name and photo alongside their query history. Page-specific awareness allows users to see visitation counts, ratings and comments associated with web pages. On a ranked list of search results, users can see those webpages that have been viewed by their collaborators. Both the query history and the page specific awareness widgets can achieve a DoL and SoK across searchers. As noted by the authors, users can avoid unnecessary duplication of effort by being aware of those query terms entered by collaborators, furthermore users can learn from seeing those query terms entered by others. Showing visitation information directly in the search results pane, rather than in a separate history widget, requires less effort on the part of collaborators to avoid unnecessary duplication of effort. A similar approach known as *group member URL traversal awareness* (GMUTA) was proposed by Diamadis and Polyzos (2004). SearchTogether also enables users to recommend webpages to their collaborators. This facility is similar in purpose to the bookmarked area of systems like W4, but through a recommendation device items of relevance are "pushed" towards users rather than being "pulled" by the users themselves. Furthermore, the recommendation mechanism enables documents to be pushed to particular users rather than to the group as a whole, such a mechanism could support specialisation in the group search where each user is assigned particular aspects of the search task to investigate.

### 4.2. Co-located SCIR

Recently the development of both ubiquitous computing devices and shared displays has facilitated the development of co-located collaborative information retrieval systems (Blackwell, Stringer, Toye, & Rode, 2004; Han, Perret, & Naghshineh, 2000; Maekawa, Hara, & Nishio, 2006; Morris, Paepcke, & Winograd, 2006; Smeaton, Lee, Foley, & Mc Givney, 2006). By bringing users together in a face-to-face environment, these systems improve the awareness across collaborating searchers inherently. A co-located collaborative environment is an awareness-rich setting, collaborators can converse easily while they search, furthermore this communication is enhanced by non-verbal communication such as the facial-expressions, postures and gestures of their search partners. If utilised effectively, this increased awareness has the potential to enable a more effective division of labour and a greater sharing of knowledge across the collaborating group. Design decisions through the choice of interfaces and widgets on these systems can further improve awareness across users, such as including sounds associated with actions or exaggerated gestures to complete different tasks (Smeaton, Foley, et al., 2006).

Han et al. (2000) proposed a system for co-located collaborative browsing on mobile devices. The system divided a webpage into distinct regions and distributed these sections across user's mobile devices. Maekawa et al. (2006) extended this approach to allow for a more intelligent division of the webpage, by considering factors such as the device the users was working on (e.g. mobile phone, laptop) and the user's interests represented in a set of keywords, when dividing the page.

In Smeaton, Lee, et al. (2006) we presented Físchlár-DiamondTouch, a system for co-located collaborative search on a DiamondTouch multi-user touch-sensitive electronic tabletop (Dietz & Leigh, 2001). The system enabled two users to sit opposite each other and search for video shots by querying the corpus with either text or content-based visual queries and interacting with the search results using point and drag touch-gestures. Two variants of the system were developed, one where the focus of the design was on supporting individual efficiency through the search by providing rapid menu-items and another which attempted to increase the awareness across collaborating users. Awareness was supported through the provision of a single set of search widgets for users (i.e. a single text query box and a single area to move saved shots to etc.) this required users to drag shots to different parts of the table, thereby making their collaborators aware of their activities in their peripheral vision. Each significant action also had an associated sound cue: for example saving a shot caused a "stamping" noise to be heard along with a visual "saved" stamp appearing on the shot. Results from our experiments showed that users both preferred and performed better using the awareness based system.

Físchlár-DiamondTouch supported division of search results implicitly: results from queries were orientated towards the nearest users, users were therefore inclined to interact with those shots closest to them. Due to the provision of only one query box, users collaborated on formulating a shared query with each user suggesting terms to enter. Users also collaborated over relevance judging whereby if one user was unsure if a particular video shot was relevant they consulted their search partner.

Collaboration in Físchlár-DiamondTouch was very much front-loaded, although users could co-ordinate the DoL and SoK of the search to a certain extent, the group search was impeded by the fact that the system was using a standard, single-user, IR backend. For example, the provision of only one query box inhibited a parallel search, rather users were tightly coupled around a single shared group query. We felt that such a scenario could be improved through the development of IR algorithms and methods that were designed specifically to support multi-user search. In the next section we will discuss our work in developing system-mediated DoL and SoK techniques.

## 5. System-mediated techniques for SCIR

In our work we have explored the potential for system-mediated DoL and SoK in SCIR search (Foley & Smeaton, 2009; Foley, Smeaton, & Jones, 2008; Foley et al., 2006; Smeaton, Foley, Byrne, & Jones, 2008). Our motivation for this is based on the belief that an SCIR system can be improved with a back-end search engine which is specifically designed to support collaboration. This belief is based on both our own experiences with Físchlár-DiamondTouch and from our review of the state-of-the-art awareness based systems for SCIR. In essence, in order to achieve an effective group search using a purely awareness-based system, collaborators need to both search *and* coordinate a shared group task. For example, in order to achieve an effective division of labour and ensure duplication of effort is removed, collaborating users would need to coordinate their actions using an instant messaging client for example or check if the link they are about to follow has already been followed by their collaborators. For sharing of knowledge, users need to read those documents saved in the bookmarked folder, or recommended to them. As observed by Adcock et al. (2007) requiring users to move their attention between the search task and the coordination of the group effort may require too much of a user's cognitive load or at best be an unwelcome distraction from the search task.

Through implementing system-mediated techniques, the burden for coordination is moved from the users to the back-end, and this allows users to concentrate on the search task itself without concern for organising the group effort. To this end we developed system-mediated techniques for both DoL and SoK and in this section we will summarise our work to-date.

### *5.1. System-mediated division of labour*

In Foley and Smeaton (2009) we presented some light-weight back-end division of labour techniques for adhoc SCIR. These techniques were based upon two assumptions. Firstly, that users searching together to satisfy the same information need will enter the same or similar queries and as a result, have returned to them ranked lists which contain many of the same documents, and this has been shown to be the case (Foley & Smeaton, 2009). Secondly that we could remove those documents from all subsequent ranked lists given that at least one member of the group had already seen it. Considering that all users are searching to satisfy the same information need, this approach seems reasonable.

In our experiments with simulated search we investigated the effects of different levels of division of labour on an SCIR session. In particular, we examined the effects of a full DoL environment in which no two users were presented with the same documents. This was implemented by removing from a user's ranked lists those documents that had been viewed by others previously, and also those that were contained on another collaborator's ranked list that they were examining.

Our experimental results showed that an SCIR system with a system-mediated DoL policy was better than an SCIR system without any such coordination. The experiments also highlighted the importance of a DoL for SCIR, as those SCIR system variants without any DoL were shown to perform significantly worse than two independent search baselines where users were simulated searching without any collaboration.

These results were very encouraging, considering the simplicity of the methods involved and showed the potential and importance for system-mediated DoL. The advantage of removing redundant documents, rather than just highlighting as is the case in an awareness-based SCIR system, is that redundant documents in the ranked list are replaced with new unseen documents. This enables the group to cover a greater amount of the collection over the course of the search. A potential downside of such an approach is the possibility than in removing relevant documents from a user's ranked list we could impede a user's natural development of their understanding of the information need. One way to overcome such an issue could be the inclusion of a separate awareness-based bookmark window so that they could review these relevant documents that were found by others. Another issue with a system-mediated DoL technique could occur when a user who is not familiar with the search topic skips over a potentially relevant document, causing it to be removed from the subsequent ranked lists of the group. One solution here could be to down-weight the scores of documents that have been seen by others rather than using the brute-force approach of removing them altogether.

### *5.2. Sharing of knowledge*

In Foley et al. (2008) and Foley and Smeaton (2009) we presented our techniques for system-mediated sharing of knowledge for SCIR. Our approach was to develop methods that enabled users to benefit from the discoveries of their collaborators during the search and in this way attempt to improve the performance of the group activity. This work was motivated by the observation that while many state-of-the-art SCIR systems encouraged users to make explicit judgments in the form of bookmarks, no attempt was made to use these relevance judgements in the search process itself. Relevance feedback (RF) is an IR method proven to improve the quality of ranked lists by reformulating a user's query automatically in the light of relevance judgments (Harman, 1992). This is achieved through expanding the query with new terms from these relevance judgments (query expansion) and reweighting query terms to take into account their distribution in these relevant documents (term reweighting).

If we have an environment where two or more users are providing relevance judgments to an SCIR system, then, when performing relevance feedback for a user, the SCIR system has an opportunity to incorporate each user's relevance judgments into a the relevance feedback process, which has the potential to improve the quality of the ranked lists returned to each user.

In our work we investigated two strategies for sharing relevance information across collaborating searchers: *collaborative* and *complementary* relevance feedback. By sharing judgments in this way, users can benefit from the efforts of their collaborators in finding these relevant documents.

#### *5.2.1. Collaborative relevance feedback*

In Foley et al. (2008) we outlined our approach for combining relevance information in a collaborative relevance feedback method. In particular we extended the probabilistic relevance feedback method (Robertson & Spärck Jones, 1976), in order to allow for a weighted combination of each collaborator's relevance judgments (bookmarks) in the process. In this way the reformulated query contains relevance information from multiple collaborating users thus allowing for an implicit sharing of knowledge across users.

In our experiments we investigated if sharing relevance judgments across users in an SCIR environment improved the performance of the group search. We performed simulated experiments using the same procedure as used for our DoL experiments. Our results showed that a collaborative relevance feedback method provided small improvements in the group's performance over the entire search, with greater improvements achieved at the early part of the search.

Synchronous collaborative information retrieval systems, by their very nature, bring together multiple collaborating users, each with a certain level of expertise and experience. Some users may be more familiar with a topic than others and this may be reflected in the quality of their relevance judgments during the group search activity For example, a novice user may not understand the search topic entirely and therefore may be mistaken in their relevance assessments. As our proposed collaborative relevance feedback algorithm operates by combining the relevance judgments of all collaborators, poor relevance assessments unless recognised and dealt with may degrade the quality of results being presented to users.

In Foley and Smeaton (2009) we investigated the effects of poor relevance assessments on the performance of a collaborative relevance feedback process. Unsurprisingly our results showed that poor relevance judgments decreased the performance of the collaborative relevance feedback process significantly. Such a result represents a significant concern for any process which attempts to combine relevance information from many users collaborating together.

In Foley and Smeaton (2009) we proposed an authority weighting mechanism which used the capacity for weighted combination in the collaborative relevance feedback formula to bias the process in favour of the more authoritative user. Our experimental results showed that by biasing the collaborative relevance feedback process towards more authoritative users the performance of the collaborative relevance feedback mechanism is improved significantly.

*5.2.2. Complementary relevance feedback*

Our experiments in Foley et al. (2008) showed that a collaborative relevance feedback method provided only modest increases in performance over an entire SCIR search. Our analysis of the data revealed that the collaborative relevance feedback process was causing the ranked lists returned to each collaborating user to become very similar. This result is intuitive given that the process allowed relevance information from one user to affect the process for their search partners. We felt that this loss of diversity reduced the performance of our collaborative relevance feedback method.

In Foley and Smeaton (2009) we investigated another use of relevance judgments, in particular through a complementary relevance feedback method. The idea behind complementary relevance feedback is to perform relevance feedback in such a way that diversity is maintained or increased across users' ranked lists. We experimented with two methods here: complementary query expansion and clustering of the relevance judgment set. Complementary query expansion ensured that when performing relevance feedback, each user's reformulated relevance feedback query contained unique terms. The idea here that these distinct queries should produce ranked lists with increased diversity.

Our clustering approach was to partition the set of relevance judgments into $k$ partitions where k was the number of users collaborating in the search, the idea being that each cluster would contain a set of relevance judgments that were as diverse as possible from each other.

Although both complementary query expansion and clustering improved the diversity across user's ranked lists, neither improved the performance of an SCIR search. This result was disappointing, however we believe there is still potential for a complementary feedback method.

### *5.3. iBingo: mobile collaborative video search*

In Smeaton et al. (2008) we implemented our system-mediated concepts for division of labour and sharing of knowledge in "iBingo" a collaborative video search system developed for Apple iPhones. Through iBingo we explored the concepts of DoL and SoK through *information trails*. Using simple touch gestures users could perform rapid content-based searching, which allowed users to build an information trail through the collection. The back-end search engine implemented a DoL by ensuring that once a video shot had been examined by one user it was removed from the trails of their collaborators. SoK was achieved through the sharing of user-trails and was performed when a user had exhausted their own information trail. In this way, when both users were pursuing their own trails a DoL ensured no duplication of effort, whilst a SoK policy ensured that users could help each other when needed. iBingo also represented our first attempt at combining system-mediated and awareness coordination, the users were working in close proximity and were free to discuss the search topic while the back-end coordinated their actions.

## 6. Discussion

In this paper we discussed the application of two CSCW concepts: division of labour and sharing of knowledge, to a synchronous collaborative information retrieval environment. We presented two approaches by which these concepts can be realised in SCIR: either through the provision of awareness cues or through system-mediated techniques. Both approaches have advantages and disadvantages associated with them.

### *6.1. Awareness versus system-mediated coordination*

For certain search topics, a clear division of the task across group members may not be possible using user–user awareness coordination. If users are searching in the same information space then they will invariably be returned documents that have already been viewed by their collaborators. Our system-mediated DoL techniques would remove such documents from a user's ranked list thereby ensuring that no two users spend time viewing the same documents and ensure a complete DoL. The downside of removing relevant documents from a user's ranked list before returning it to them is that we may disrupt the development of a user's understanding of the information need. Users may think that the queries they have entered are poor because they are seeing no relevant documents returned, when in fact the query could have returned many documents but they were removed because they had already been seen by a collaborator. In such a scenario an awareness widget could inform a user that a certain number of documents have been removed, with the option to show these if needed.

The bookmarks and recommendation widgets of awareness-based SCIR systems allow for a sharing of relevant documents across collaborating users. If users wish to view these documents they can follow the links and view these documents. In our system-mediated collaborative and complementary relevance feedback approaches we attempt to enable users to benefit from the discoveries of their collaborators without having to view the documents themselves and mark them as relevant. The benefit here is in the form of improved ranked lists, however what these system-mediated techniques fail to capture is the benefit, in terms of their understanding of the information need, a user gains from seeing the relevant document. A combination of both approaches would allow users to benefit from the relevance judgments of their collaborators whilst having the option to view these relevant documents if they wish. Furthermore, upon reviewing a document marked by another collaborator, a user may elect to agree or disagree with the user's relevance judgment, and this could reinforce a collaborative or complementary relevance feedback algorithm.

To summarise, although users can attempt to coordinate the search tasks amongst themselves through an awareness-based system, such techniques could impede the group search performance, while a pure system-mediated approach may cause users to feel lost in the search. Our conclusion is that the inherent benefits and shortcomings of each would suggest that a state-of-the-art SCIR system should endeavour to support both approaches.

### *6.2. Future work in system-mediated techniques*

In our work we have explored the notion of authority in relation to the relevance feedback process. For future work it would be interesting to extend this notion of authority, and model cases where users can skip relevant material. Modelling such an environment would enable us to investigate how a division of labour policy would operate in an environment where searchers may read a document but fail to recognise it as relevant. At present, the DoL techniques would exclude this document from all ranked lists returned to all users for the rest of the search. This would cause the group to miss relevant material due to the actions of a poor searcher. A potential solution to this issue would be to extend the division of labour policy to allow for an approach which reduced the rank of seen documents rather than excluding them, as is currently the case. This dampening effect could be weighted by the perceived expertise of a user so that, if a more expert user reads a document without providing a relevance judgment on it, then we can be more confident that the document is non-relevant than if a poor searcher performed the action.

We proposed techniques which allowed for a user-biased weighting of relevance judgments, and we investigated one application of a user-biasing through authority weighting. However we believe there are many more applications of a user-biased collaborative relevance feedback process. For example, in a real search system, a user could use a user-biased relevance feedback process to favour their own relevance judgements over their search partners', thereby allowing the results to be tailored to them. Or a user may decide to use an inverted approach and bias their results in the favour of their search partner if they feel they cannot locate any relevance documents on their own. We believe there are many more applications of our proposed collaborative relevance feedback process, which we have not even considered.

We introduced the notion of imperfect relevance judgements to synchronous collaborative search. In traditional, single-user information retrieval, the notion of imperfect relevance judgments is less of an issue as, if a user has made a relevance judgment, then it should be considered as relevant for that user. When we move to a synchronous collaborative information retrieval domain, in which a group of users are searching together, the issue of non-relevance or misunderstanding of the search topic could have a major effect of the performance of an SCIR search. We believe that this also represents an interesting area from future research. We investigated the application of authority weighting, and we believe there is scope for future research here and, in particular, in future work we would like to apply some of the various query performance predictions techniques to the area of SCIR authority weighting. Although the results of our evaluations of our proposed complementary relevance feedback techniques were not favourable, we believe there is still scope for further research here. In particular we feel that the clustering of documents and terms in a SCIR search may prove useful. In our work we attempted to cluster documents and terms into $k$ clusters, one for each user. An interesting avenue for further research would be to try to use clustering to discover unique concepts in the search topic. For example a search topic: "Wildlife Extinction" may have concepts related to zoos, poachers, animals, etc. By recognising the underlying concepts present in the relevance judgments, we may be able to make a more intelligent division of the clusters across users, whereby each user is assigned a unique concept to search.

Division of labour and sharing of knowledge are two closely related concepts in SCIR, we therefore believe that both should be realised in all systems. For example, a division of labour policy can remove documents it deems to be redundant to the search given that they have been seen by other users, as a sharing of knowledge policy ensures that such documents that are also marked as relevant, through bookmarks, are brought to the attention of the group.

## 7. Conclusion

In this paper we introduced and discussed two concepts which we believe are vital to the development of effective SCIR systems and on which we carried out experimental investigations: division of labour and sharing of knowledge. We outlined how both techniques had been recognised as important in the CSCW domain and discussed how they applied to the emerging research field of SCIR. We discussed how both can be achieved either through the provision of awareness or by our system-mediated techniques and showed how they can improve the performance of SCIR. Our suggestion is that future SCIR systems should endeavour to support both techniques through the provision of awareness cues and system-mediated techniques. Through iBingo we have implemented such a system, and in future work we will pursue further integration of awareness and system-mediated DoL and SoK techniques.

## References

Adcock, J., Pickens, J., Cooper, M., Anthony, L., Chen, F., & Qvarfordt, P. (2007). FXPAL Interactive search experiments for TRECVID 2007. In *TRECVid2007 – Text REtrieval conference TRECVID workshop* Gaithersburg, MD, USA, November 2007.

Ahn, J.-W., Brusilovsky, P., & Farzan, R. (2005). Investigating users' needs and behaviour for social search. In *Proceedings of the workshop on new technologies for personalized information access (part of the 10th international conference on user modelling (UM '05))* Edinburgh, Scotland, UK, July 2005 (pp. 1–12).

Bellotti, V., & Bly, S. (1996). Walking away from the desktop computer: Distributed collaboration and mobility in a product design team. In *CSCW '96: Proceedings of the 1996 ACM conference on computer supported cooperative work* (pp. 209–218). Boston, MA, USA: ACM.
Benyon, D., Turner, P., & Turner, S. (2004). *Designing interactive systems – People, activities, contexts, technologies*. Addison-Wesley.
Blackwell, A. F., Stringer, M., Toye, E. F., & Rode, J. A. (2004). Tangible interface for collaborative information retrieval. In *CHI '04: Extended abstracts on human factors in computing systems* (pp. 1473–1476). Vienna, Austria: ACM.
Cabri, G., Leonardi, L., & Zambonelli, F. (1999). Supporting cooperative WWW browsing: A proxy-based approach. In *7th Euromicro workshop on parallel and distributed processing* (pp. 138–145). Funchal, Portugal: University of Maderia, IEEE Press.
Cockburn, A., & Greenberg, S. (1993). *Making contact: Getting the group communicating with groupware*. ACM Press. pp. 31–41.
Del.icio.us. (2009). <http://www.del.icio.us/> Accessed 05.09.
Diamadis, E. T., & Polyzos, G. C. (2004). Efficient cooperative searching on the web: system design and evaluation. *International Journal of Human–Computer Studies, 61*(5), 699–724.
Dietz, P., & Leigh, D. (2001). DiamondTouch: A multi-user touch technology. In *UIST '01: Proceedings of the 14th annual ACM symposium on user interface software and technology* (pp. 219–226). Orlando, Florida: ACM.
Dilley, Keith J. (2008). Informatics & analytics – The team approach. In: IRFS2008, the information retrieval foundation symposium, November 2008.
Dourish, P., & Bly, S. (1992). Portholes: Supporting awareness in a distributed work group. In *CHI '92: Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 541–547). Monterey, California, United States: ACM.
Ellis, C. A., Gibbs, S. J., & Rein, G. (1991). Groupware: Some issues and experiences. *Communications of the ACM, 34*(1), 39–58.
Foley, C., & Smeaton, A.F. (2009). Synchronous collaborative information retrieval: Techniques and evaluation. In *ECIR 2009 – 31st European conference on information retrieval* (pp. 42–53). Toulouse, France.
Foley, C., Smeaton, A. F., & Jones, G. J. F. (2008). Combining relevance information in a synchronous collaborative information retrieval environment. *Collaborative and Social Information Retrieval and Access Techniques for Improved User Modelling*.
Foley, C., Smeaton, A.F., & Lee, H. (2006). Synchronous collaborative information retrieval with relevance feedback. In *CollaborateCom 2006 – 2nd international conference on collaborative computing: networking, applications and worksharing* (pp. 1–4). Adelaide, Australia, 2006.
Foster, J. (2006). Collaborative information seeking and retrieval. *Annual Review of Information Science and Technology, 40*(1), 329–356.
Gianoutsos, S., & Grundy, J. (1996). Collaborative work with the World Wide Web: Adding CSCW support to a Web browser. In *Proceedingss of Oz-CSCW'96, DSTC technical workshop series. University of Queensland*. Brisbane, Australia: University of Queensland.
Greenberg, S., & Roseman, M. (1996). GroupWeb: A WWW browser as real time groupware. In *CHI '96: Conference companion on human factors in computing systems* (pp. 271–272). Vancouver, British Columbia, Canada: ACM Press.
Gutwin, C., Greenberg, S., Blum, R., Dyck, J., Tee, K., & McEwan, G. (2008). Supporting informal collaboration in shared-workspace groupware. *Journal of Universal Computer Science, 14*(9), 1411–1434.
Han, R., Perret, V., & Naghshineh, M. (2000). WebSplitter: A unified XML framework for multi-device collaborative web browsing. In *CSCW '00: Proceedings of the 2000 ACM conference on computer supported cooperative work* (pp. 221–230). Philadelphia, Pennsylvania, United States: ACM.
Hansen, P., & Järvelin, K. (2005). Collaborative information retrieval in an information-intensive domain. *Information Processing and Management, 41*(5), 1101–1119.
Harman, D. (1992). Relevance feedback revisited. In *SIGIR '92: Proceedings of the 15th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 1–10). Copenhagen, Denmark: ACM Press.
Heystaks. (2009). <http://www.heystaks.com/> Accessed 05.09.
Ingwersen, P. (1992). *Information retrieval interaction*. London: Taylor Graham.
Krishnappa, R. (2005). Multi-user search engine (MUSE): Supporting collaborative information seeking and retrieval. Master's thesis, University of Missouri-Rolla, University of Missouri-Rolla, Rolla.
Kuutti, K. (1991). The concept of activity as a basic unit of analysis for cscw research. In *ECSCW'91: Proceedings of the second conference on European conference on computer-supported cooperative work* (pp. 249–264). Norwell, MA, USA: Kluwer Academic Publishers.
Lai, K.-Y., Malone, T. W., & Yu, K.-C. (1988). Object lens: A "spreadsheet" for cooperative work. *ACM Transactions on Information Systems, 6*(4), 332–353.
Maekawa, T., Hara, T., & Nishio, S. (2006). A collaborative web browsing system for multiple mobile users. In *PERCOM '06: Proceedings of the fourth annual IEEE international conference on pervasive computing and communications (PERCOM'06)* (pp. 22–35). Washington, DC, USA: IEEE Computer Society.
Malone, Thomas W., Crowston, Kevin. (1990). What is coordination theory and how can it help design cooperative work systems (pp. 357–370).
Morris, M.R. (2007). Collaborating alone and together: Investigating persistent and multi-user web search activities. Technical Report, Microsoft Research, January 2007.
Morris, M. R. (2008). A survey of collaborative web search practices. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on human factors in computing systems* (pp. 1657–1660). New York, NY, USA: ACM.
Morris, M. R., & Amershi, S. (2008). Shared sensemaking: Enhancing the value of collaborative web search tools. In *CHI 2008 workshop on sensemaking*. Florence, Italy.
Morris, M. R., & Horvitz, E. (2007). SearchTogether: An interface for collaborative web search. In *UIST '07: Proceedings of the 20th annual ACM symposium on user interface software and technology* (pp. 3–12). Newport, Rhode Island, USA: ACM Press.
Morris, M. R., Paepcke, A., & Winograd, T. (2006). TeamSearch: Comparing techniques for co-present collaborative search of digital media. In *TABLETOP '06: Proceedings of the first IEEE international workshop on horizontal interactive human–computer systems* (pp. 97–104). Washington, DC, USA: IEEE Computer Society.
Muller, M. J., Geyer, W., Brownholtz, B., Wilcox, E., & Millen, D. R. (2004). One-hundred days in an activity-centric collaboration environment based on shared objects. In *CHI '04: Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 375–382). New York, NY, USA: ACM.
Poltrock, S., Grudin, J., Dumais, S., Fidel, R., Bruce, H., & Pejtersen, A. M. (2003). Information seeking and sharing in design teams. In *GROUP '03: Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work* (pp. 239–247). Sanibel Island, Florida, USA: ACM.
Poltrock, S., & Handel, M. (2009). Modeling collaborative behavior: Foundations for collaboration technologies. *Hawaii International Conference on System Sciences, 0*, 1–10.
Robertson, S. E., & Spärck Jones, K. (1976). Relevance weighting of search terms. *Journal of the American Society for Information Science, 27*(3), 129–146.
Romano, N. C., Jr., Roussinov, D., Nunamaker, J. F., & Chen, H. (1999). Collaborative information retrieval environment: Integration of information retrieval with group support systems. *HICSS '99: Proceedings of the thirty-second annual hawaii international conference on system sciences* (Vol. 1). Maui, Hawaii: IEEE Computer Society.
Root, R. W. (1988). Design of a multi-media vehicle for social browsing. In *CSCW '88: Proceedings of the 1988 ACM conference on computer-supported cooperative work* (pp. 25–38). Portland, Oregon, United States: ACM Press.
SearchTogether. (2009). <http://www.research.microsoft.com/en-us/um/redmond/projects/searchtogether/> Accessed 05.09.
Sharples, M. (1993). Adding a little structure to collaborative writing. *CSCW in practise: An introduction and case studies*.
Smeaton, A. F., Foley, C., Byrne, D., & Jones, G. J. F. (2008). iBingo mobile collaborative search. In *CIVR 2008 - ACM international conference on image and video retrieval*. VideOlympics @ CIVR, 2008.
Smeaton, A. F., Foley, C., Gurrin, C., Lee, H., Mc Givney, S. (2006). Collaborative searching for video using the Físchlár system and a DiamondTouch table. In *TableTop2006 – The 1st IEEE international workshop on horizontal interactive human–computer systems* (pp. 149–156). Adelaide, Australia.
Smeaton, A. F., Lee, H., Foley, C., & Mc Givney, S. (2006). Collaborative video searching on a tabletop. *Multimedia Systems Journal, 12*(4), 375–391.
Smyth, B., Balfe, E., Freyne, J., Briggs, P., Coyle, M., & Boydell, O. (2005). Exploiting query repetition and regularity in an adaptive community-based web search engine. *User Modeling and User-Adapted Interaction, 14*(5), 383–423.

Talja, S. (2002). Information sharing in academic communities: Types and levels of collaboration in information seeking and use. *The New Review of Information Behaviour Research, 3*, 143–159.
Tang, J. C. (1991). Findings from observational studies of collaborative work. *International Journal of Man-Machine Studies, 34*(2), 143–160.
Trigg, R. H., Suchman, L. A., & Halasz, F. G. (1986). Supporting collaboration in notecards. In *CSCW '86: Proceedings of the 1986 ACM conference on computer-supported cooperative work* (pp. 153–162). New York, NY, USA: ACM.
Twidale, M. B., Nichols, D. M., & Paice, C. D. (1997). Browsing is a collaborative process. *Information Processing and Management, 33*(6), 761–783.
Want, R., Hopper, A., Falcao, V., & Gibbons, J. (1992). The active badge location system. *ACM Transactions on Information Systems, 10*(1), 91–102.
Whittaker, S., Geelhoed, E., & Robinson, E. (1993). Shared workspaces: How do they work and when are they useful. *International Journal of Man–Machine Studies, 39*(5), 813–842.
Wilson, T. D. (1981). On user studies and information needs. *Journal of Documentation, 37*(1), 3–15.
Yao, K.-T., Neches, R., Ko, I.-Y., Eleish, R., & Abhinkar, S. (1999). Synchronous and asynchronous collaborative information space analysis tools. In *ICPP '99: Proceedings of the 1999 international workshops on parallel processing* (pp. 74). Washington, DC, USA: IEEE Computer Society.