Synchronous Collaborative Information Retrieval with Relevance Feedback

Colum Foley, Alan F. Smeaton and Hyowon Lee
Adaptive Information Cluster and Centre for Digital Video Processing,
Dublin City University, Glasnevin, Dublin 9, Ireland.
Colum.Foley@computing.dcu.ie

Abstract— **Collaboration has been identified as an important aspect in information seeking. People meet to discuss and share ideas and through this interaction an information need is quite often identified. However the process of resolving this information need, through interacting with a search engine and performing a search task, is still an individual activity. We propose an environment which allows users to collaborate to satisfy a shared information need. We discuss ways to divide the search task amongst collaborators and propose the use of Relevance Feedback, a common Information Retrieval process, to enable the transfer of knowledge across collaborators during a search session. We describe the process by which co-searchers can collaborate effectively with little redundancy and how we can combine Relevance Judgements from multiple searchers into a coherent model for Synchronous Collaborative Information Retrieval.**

I. Introduction

Most of us collaborate frequently every day and we have developed the ability to meet and share ideas with each other naturally. Finding information also sees users collaborating frequently. Studies by Hansen and Jarvëlin [4] and others [9] have shown how collaboration is common in information seeking tasks. This collaboration quite frequently happens separately to the information searching process (i.e. querying databases and viewing results) which is still generally viewed as a single user process.

In this paper we propose a Synchronous Collaborative Information Retrieval environment in which multiple users can satisfy a shared information need together. Collaborative searching enables the division of labour and sharing of knowledge within a collaborating group [10]. However two outstanding questions which remain, and that we address, are how to effectively divide the search task and how to enable the sharing of knowledge amongst co-searchers.

The division of labour in a collaborative search task is vital to remove the redundancy associated with multiple users searching for the same information. In a collaborative search task, each searcher will bring their own knowledge and expertise to the search process, for example some users may be more familiar with a search topic than their co-searchers or some searchers may be more familiar with using specific search tools. Therefore it is important that we enable co-searchers to learn from each other's experiences and benefit from their knowledge. By supporting multiple people searching together on a single search task we have in essence a large and diverse knowledge base distributed amongst all collaborators. What is needed is a way to enable each user to 'tap into' this knowledge base, i.e. a way to support the transfer of knowledge amongst co-searchers. In our system this transfer of knowledge is achieved through Collaborative Relevance Feedback.

II. The Information Retrieval Process

Information Retrieval (IR) is concerned with the development of techniques for the retrieval of documents from a collection. Today most users' interactions with Information Retrieval systems are through an internet search engine like Google. A user enters a search query and the search engine returns a list of documents according to some ranking model. These models attach weights to terms from both queries and documents in order to generate a score for each document and return a list of documents ranked in descending order of matching score to the user's query.

Despite the commercial success of search engines on the internet like Google, not all user information needs are satisfied within the initial returned document set. It is easy to forget that initially, the only information available to the search engine regarding a user's information need is the search query, typically 1-2 words, which often does not adequately describe the user's information need.

Relevance Feedback is a process by which a searcher can indicate the relevance of a search result to the IR engine, enabling the system to re-rank its document list by moving more relevant documents up the list and pushing irrelevant documents down the list, in order to better satisfy a user's information need. Having received feedback an IR system may attempt to modify the initial query in order to improve the retrieval results by adding significant search terms from these relevant documents to the query or by reweighting terms in the query. Over several iterations of Relevance Feedback the system will gain a greater understanding of the user's information need which will improve the search.

III. Synchronous Collaborative Information Retrieval

Computer Supported Cooperative Work (CSCW) is concerned with the development and understanding of systems that allow people to collaborate effectively. Often CSCW systems are classified according to the two dimensions of

1-4244-0429-0/06/$20.00 ©2006 IEEE

place and time [2]. Here we are interested in Synchronous Collaborative Information Retrieval which supports 'same-time different-place' collaboration. Our eventual goal is to incorporate techniques developed here into a co-located collaborative search system ('same-time same-place') called Físchlár-DiamondTouch, which we have developed and described elsewhere [8]. In that work we demonstrated a system for collaborative searching through video where users shared a touch-sensitive tabletop interface to a search engine and users communicated and collaborated in a face-to-face manner in order to solve a shared information need. Through developing this system we now appreciate the need to support collaboration *within* the underlying IR system itself and not just as part of the interface. The CATS [7] collaborative group recommender system is another CSCW system supporting 'same-time same-place' collaboration. Here ski-holiday critiques from multiple users are leveraged and destinations are recommended based on both an individual's preferences and the group's preferences. The key objective within the CATS system is allowing each user to see which ski-packages suit both their own preferences and those of the groups.

Information Retrieval has been transformed since the arrival of the internet and as ubiquitous computing becomes ever more commonplace the IR process will become much more spontaneous and, as a result, collaborative. Thus there is a real need to develop techniques to support Synchronous Collaborative Information Retrieval. At present if two users want to search for information together they are inhibited by the lack of support for collaboration in current IR systems. Both Físchlár-DiamondTouch and the CATS systems described above allow for collaboration at the interface level and whilst the CATS system does allow some system level collaboration the focus of the system is on combining multiple individual's preferences into a collaborative recommendation framework whilst our focus is on developing a system to support a dynamic collaborative IR session. The following is a hypothetical example of a collaborative search session whereby two users attempt to find information to satisfy a shared information need.

A. Use case for Synchronous Collaborative IR

Suppose two users, Tanya and Avena, decide to search together for information on Prague, the destination of a forthcoming vacation. Each opens up a search engine in their browser and proceeds to look for information on famous sights, places to stay and eat in the city. Both begin by typing a query; Tanya enters 'Prague' and Avena enters 'Prague Tourist Information'; each receives a list of documents returned by the search engine which are quite similar and they both decide to visit a website belonging to the Prague tourist board which appears in each of their ranked lists. After 5 minutes Tanya decides that this website has been particularly useful and decides to find 'more like this' using the search engine's Relevance Feedback system. She receives an updated ranked list with several new and relevant results. In the meantime Avena has navigated back to the ranked list and decides to visit several other websites before reentering another query and receiving a new ranked list. The session proceeds with both users entering queries, viewing results and providing feedback until they have satisfied their information need. Then they gather and discuss their findings and eventually agree on their schedule.

The above is a typical example of a Synchronous Collaborative IR activity that happens regularly in modern internet life, and from looking at this simple hypothetical example we can recognise obvious inefficiencies which result when trying to use a single-user system, the search engine, for a highly collaborative activity.

B. Division of Labour

Firstly as both users' information needs are identical (both are searching for information on Prague) the results returned by both search sessions are likely to contain many of the same documents. As a result both users spent time viewing the same web sites and this lack of co-ordination between the two users results in much redundancy in the search process. An appropriate division of labour would enable each group member to search for a particular subset of information, making the process much more efficient [10]. One approach would be for the collaborators to divide up the search task themselves whereby one user, Tanya, decides to search for 'Places to stay', and Avena finds information on 'Places to eat'. CSCW3 [3] is a collaborative browsing system which supports both shared and independent browsing and would enable multiple users to divide and coordinate their actions using embedded chat facilities. There are some potential problems with this solution however. In the above example deciding on how to split the task is quite straightforward. However, for some tasks it may be difficult to decide on how to adequately divide the work and therefore relying on the users themselves to divide the task may not be possible. Another problem is a potential lack of topic coverage, by having each user search a particular aspect of the information need some potentially relevant material may not be found. In our example above Tanya may not be able to find information on places to stay in Prague by searching alone and this problem would be exacerbated if Tanya is a novice searcher and not familiar with the topic.

In our solution there is no need to split the task as above and instead we allow the search engine itself to implicitly divide the task by returning those documents that have not been seen by *any* co-searcher during the search task. Using our technique we can ensure that each entry in the result list has not been seen before by any co-searcher. This will not simply mean discounting duplicate documents but also handling the notion of novelty in the search task, i.e. how much extra information is being provided by a document. By allowing each user to see only new information we are of course limiting the amount of information any one searcher can accumulate but we are enabling the group to have a much greater throughput and hopefully find more relevant information as a whole. This is one of the major differences between conventional Collaborative IR techniques as in [6],

where the goal is to better satisfy an individual's information need and Synchronous Collaborative IR where the goal is to better satisfy a group's shared information need.

C. Knowledge Sharing

Another problem recognisable from the above scenario is the lack of information sharing among co-searchers. In any collaborative setting there will be a large and diverse knowledge base shared amongst group members. Each will bring their own experiences, expertise and topic knowledge to a particular search task. What is needed is a way to enable the sharing of knowledge within the group. To address this our approach can be summarised as follows: as one user finds relevant information the group as a whole should benefit and this benefit can be realised in the form of improved ranked lists for each searcher within the group. How a searcher indicates this relevant information to the search engine is through Relevance Feedback.

From the example above,when Tanya and Avena were searching for information on Prague, we can see that Tanya provided Relevance Feedback to the system by selecting more pages like the Prague Tourist Board website, which resulted in a revised result list being presented to her. However there was no benefit to Avena from this feedback. If Tanya has indicated to the system that a particular document is relevant by making a Relevance Judgement, and all group members are searching to satisfy a shared information need, this should benefit the group as a whole by updating the search results for each searcher in the group. As the search progresses through iterations, the number of these Relevance Judgments will increase, and if we expand the number of co-searchers we will gain even more Relevance Judgments.

Combining multiple Relevance Judgments from many people raises a number of interesting research questions. For example, how do we know when to combine relevance judgments and how do we combine these judgments effectively ? Knowing when to combine relevance judgments has been investigated in the context of Collaborative Information Filtering before [6], where some options include comparing query-query similarity, or comparing the results of two separate searches in order to find similar searchers. For the purpose of a Synchronous Collaborative IR this is less of an issue. If users come together to search then it is reasonable to assume that they are searching in order to satisfy a shared and relatively well-defined information need. Obviously their notion of relevance will be different and this should be taken into account, but these differences are relatively minor when compared with a Collaborative Information Filtering system.

In order to combine judgements effectively we have identified three influencing factors described as follows.

1) Authority of Relevance Judgments: Several characteristics of 'good' searchers have been identified in the literature, such as search experience, training, cognitive characteristics and intelligence and personality traits, for an overview see [5]. Collaborative searching allows groups to build up expertise quickly and especially when the collaboration involves experts from different domains [10]. In order to enable the group to build up expertise quickly what is needed is a way to assign appropriate weights to each Relevance Judgement depending on the expertise or *authority* of the searcher. Our proposal is to consider Relevance Judgments made by expert searchers to be more important than those made by novice searchers, this will enable the performance of the group to improve as a whole.

The notion of authority and expert users may change during the course of a synchronous search task as a user with little topic experience may learn quickly as the search progresses and users may become more familiar with the use of certain search tools. In accordance, the authority value should not remain static for the duration of the search task but instead should change during the course of a search task in order to effectively encapsulate a user's gain in experience with the system and with the topic as the search task progresses. Therefore we propose a dynamic authority weighting scheme to encapsulate a user's expertise during the course of a search task.

2) Decay of Relevance Judgements over Time: The Ostensive Relevance Feedback model [1] recognises that the most recent Relevance Judgments are a better reflection of a user's information need as the notion of relevance in a search task can change as users learn more about a topic. Often at the beginning of a search task a user may only have a vague notion of their information need. As the search progresses, the user will gain a greater understanding of the topic and be in a better position to make Relevance Judgments. In Synchronous Collaborative IR users come together to satisfy a shared and relatively well defined information need, however there should still be scope for the group's information need to develop over the course of the search task. From the scenario earlier, Tanya and Avena may decide that after searching for information together on 'Places to Eat' they might decide to progress onto searching for information on 'Places to Stay'. If we were to consider all Relevance Judgements equally regardless of when they were made then any Relevance Feedback provided on 'Places to Eat' may pollute the results for 'Places to Stay'. Therefore we need to incorporate this notion of search progression into our Synchronous Collaborative IR system.

3) Personal Relevance Judgments: In a collaborative search session different users may approach the search task from different directions. The benefit of having multiple people searching together is that each searcher will approach the task in different ways. In a single user IR system after some time a user will have exhausted all search tactics, in a Synchronous Collaborative IR system we should benefit from multiple users' diverse search strategies. However if we were to consider all judgements made by group members as one we will lose an individual's ability for innovation and new discovery in the search task. [11] also noted how users prefer their own judgments compared to those of other users. We therefore propose a weighting scheme whereby an individual's own Relevance Judgments will contribute more to their own search results than those made by other co-searchers.

IV. Conclusion and Plans

In this paper we have articulated the case for Synchronous Collaborative Information Retrieval based on our previous work where we built a shared tabletop interface to a video retrieval and browsing system. The previous work evaluated the effectiveness of co-located collaborative search where the system offered support for collaboration through the interface only. That previous work convinced us of the need for support for collaboration in the search system itself and in this paper we have described that need and indicated how it should be realised. We have attempted to address two outstanding issues in Synchronous Collaborative Information Retrieval by proposing our approaches to dividing the search task amongst all co-searchers and enabling the sharing of knowledge across co-searchers.

In order to support effective Synchronous Collaborative IR it is important to allow a search task be divided amongst co-searchers and enable the sharing of knowledge across group members. In our work we have proposed an environment whereby the search engine decides on how to divide the task amongst collaborators by showing only novel information to each co-searcher. The sharing of expertise and transfer of knowledge is achieved through the IR process of Relevance Feedback. By providing support for both task division and knowledge transfer within the framework of our underlying IR system side we can develop a more effective Synchronous Collaborative Information Retrieval environment. We are presently implementing Synchronous Collaborative IR for document retrieval and will evaluate it in a search scenario.

Acknowledgments

This work is part-funded by the Irish Research Council for Science Engineering and Technology and partially supported by Science Foundation Ireland under grant 03/IN.3/I361.

References

[1] M. Bates. The Design of browsing and berrypicking techniques for the online search interface. In *Online Review 13:5*, pages 407–424, 1989.

[2] C. A. Ellis, S. J. Gibbs, and G. L. Rein. Groupware:some issues and experiences. *Communications of the ACM*, 34(1), 1991.

[3] T. Gross. The CSCW3 Prototype - Supporting Collaboration in Global Information Systems. In *Conference Supplement of the Fifth European Conference on Computer-Supported Cooperative Work - EC-CSCW'97*, 1997.

[4] P. Hansen and K. Jarvëlin. Collaborative information retrieval in an information-intensive domain. *IP&M*, 41(5):1101–1119, September 2005.

[5] I. Hsieh-Yee. Effects of Search Experience and Subject Knowledge on the search tactics of novice and experienced searchers. *Journal of the American Society for Information Science 44(3)*, pages 161–174, 1993.

[6] A. Hust, S. Klink, M. Junker, and A. Dengel. Towards Collaborative Information Retrieval: Three Approaches. In *Franke, Nakhaeizade, Renz (eds.): Text Mining - Theoretical Aspects and Applications, Springer-Verlag*, 2002.

[7] K. McCarthy, L. McGinty, B. Smyth, and M. Salamo. Social Interaction in the CATS Group Recommender. In *Workshop on the Social Navigation and Community based Adaptation Technologies*, 2006.

[8] A. F. Smeaton, H. Lee, C. Foley, and S. M. Givney. Collaborative Video Searching on a Tabletop. *Multimedia Systems Journal*, 2006.

[9] M. Twidale, D. Nichols, and C. Paice. Browsing is a collaborative process. *IP&M*, 33(6):761–783, 1997.

[10] G. S. Zeballos. Tools for efficient collaborative web browsing. In *Proceedings of CSCW'98 workshop on Collaborative and co-operative information seeking in digital information environment*, 1998.

[11] X. Zhang and Y. Li. An exploratory study on knowledge sharing in information retrieval. *hicss*, 08:245c, 2005.