Designing for Collaboration in Information Seeking

Gene Golovchinsky
FX Palo Alto Laboratory, Inc.
3400 Hillview Ave, Bldg 4
Palo Alto, CA 94304
gene@fxpal.com

Abdigani Diriye
University College London
London, UK, WC1E 6BT
a.diriye@ucl.ac.uk

Jeremy Pickens
Catalyst Repository Systems
1860 Blake Street, 7th Floor
Denver, Colorado 80202
jpickens@catalystsecure.com

ABSTRACT
Information seeking is often a collaborative activity that can take can take many forms; in this paper we focus on explicit, intentional collaboration of small and explore a range of design decisions that should be considered when building Human-Computer Information Retrieval (HCIR) tools that support collaboration. In particular, we are interested in exploring the interplay between algorithmic mediation of collaboration and the mediated communication among team members. We argue that certain characteristics of the group's information need call for different design decisions.

Categories and Subject Descriptors
H.5.3 [**Group and Organization Interfaces**]: Computer-supported cooperative work; H.3.m [**Information storage and retrieval]**: Miscellaneous

General Terms
Design, Human Factors

Keywords
HCIR, collaborative information seeking, CSCW.

1. INTRODUCTION
There is ample empirical evidence that information seeking is often a collaborative activity. In the context of this paper, we use the term '*collaborative search*' to characterize the activities of a small group of people working towards a common, shared goal, which is otherwise known as explicit, intentional collaboration [4]. This can be contrasted this with the kinds of implicit collaboration typical of social search such as recommendation systems [8], social Q&A [1], etc.

Collaborative information seeking has been studied in the medical [12], patent law [6], military and intelligence [15], software development [3] and academic [7] domains, among others. This ethnographic work has identified broad patterns of group and individual behavior related to information seeking, but did not provide significant guidance to inform the design systems that support collaboration in search explicitly. In fact, much of the work stopped at the system level, assuming that even though the group was engaged in collaborative activity, the mechanics of search would be handled by group members individually.

Some recent work (e.g., [10], [11], [13], [3]) has explored various aspects of mediated collaboration for information seeking. SearchTogether [10] provided an interface through which people could see others' actions (running a query, saving a document, etc.) and do a rudimentary division of their efforts in examining search results. Cerchiamo [11] took this further, by introducing asymmetric roles and algorithmic mediation that combined inputs from collaborators to produce novel results rankings and visualizations based on these combinations. Coagmento [13] and CIRLab [3] focused on supporting awareness among group participants of others' activity.

These tools all focused on specific aspects of a complex problem. In this paper, we start by considering the entire human-computer system and using its characteristics in conjunction with specific use cases to illustrate possible points in the design space. We expect that an approach that combines people's needs with system capabilities will produce more effective designs compared with efforts based primarily on people's behavior or on software system design.

2. THE HUMAN-COMPUTER SYSTEM
We approach this analysis from a human factors perspective that considers people and the technology they use simultaneously, rather than in isolation. We therefore look at collaborative search as a system composed of the following actors: two or more people engaged in collaborative search, and two distinct software components that they use to perform their information seeking tasks.

As the examples from the previous section illustrate, the role of the system is two-fold: in the traditional Computer Supported Collaborative Work (CSCW) sense, it provides a means for group members to communicate and to be aware of others' activity; in the traditional information retrieval sense, it provides a means of identifying and displaying information that may satisfy users' information needs. This is illustrated graphically in Figure 1.

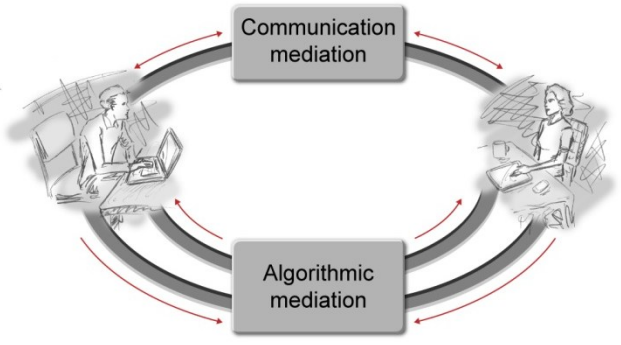

Figure 1. Actors in a collaborative search system

Communication and awareness then represent an exchange of information between the people engaged in collaborative search. The role of the software component is merely to carry this information between the individuals and to present it in an appropriate manner. Information retrieval represents communication between a person and the software, where the

HCIR 2011, October 19, 2011, Mountain View, CA, USA.

software is instructed to perform some task related to identifying, retrieving and displaying information. Thus we can envision the entire system as consisting of some number of people and two software components, one that mediates communication, and the other that performs information retrieval functions.

We can now revisit some of the systems we described previously to see how they fit into this model. The goal here is to describe existing functionality in terms of this model to suggest missed opportunities or other areas of interaction to explore.

2.1 SearchTogether

SearchTogether implements a range of tools to support person-to-person communication, including chat, recommendations, and split search. Split search is an interface feature that allows collaborators to examine search results from a single query in parallel. It provides only a limited channel for communication with the search software: people can enter keyword queries that cause the system to retrieve document references. The search component is not really aware of any collaborative activity among the people using it, and each action of the search system is triggered by an explicit request of some person.

2.2 Cerchiamo

Cerchiamo took a different approach by introducing roles and algorithmic mediation. A role in this case represents a different view on the data being identified through the information seeking session. Cerchiamo had two roles—Prospector and Miner—which were responsible, respectively, for exploring and exploiting the information landscape. The Prospector ran many queries and made a few relevance judgments to assess the utility of the returned document, whereas the Miner was responsible for making many relevance judgments, but could also suggest queries that the system would run. While the Prospector saw results one query at a time, the Miner was presented with a queue of documents aggregated from all previously-run queries. This aggregation of results into a priority queue was one of the key aspects to algorithmic mediation implemented in Cerchiamo. It allowed participants to communicate different kinds of information including the procedural (run this query) and the declarative (this document is useful, this one not so much) to the software system, and the software system responded with more documents. There was very little overt communication among the participants, and awareness of the other's actions was shown indirectly in a shared display that summarized the state of the search session in terms of queries and documents without attributing any particular aspect of that display to individuals.

2.3 Coagmento

Coagmento, like SearchTogether bundles a number of tools to improve awareness and communication among collaborators engaged in web search. It provides means to comment on documents, has an integrated chat facility, a query history, etc. Thus it provides a range of options to help people to communicate with each other, but has only a rudimentary search capability consisting of running keyword web searches.

2.4 CIRLab

CIRLab is a framework for constructing collaborative search systems that has been instantiated to create a collaborative search tool for searching over collections of source code examples. It provides a number of tools that allow searchers to share and make sense of information, including instant messaging, comments, and the ability to recommend (share) documents o other collaborators. The search component implements a split feature similar to that of SearchTogether, but no other algorithmic mediation is available. Thus the bulk of interaction with this system focuses on running individual searches and on communication activity.

This analysis of existing systems reveals clear differences in the philosophy that underpins the designs of the systems we considered. While SearchTogether, Coagmento, and CIRLab focus on communication among searchers, Cerchiamo is geared more toward algorithmic mediation.

Applying our model to these systems highlights the differences among them, but the real value of our model lies in its use as a design tool when constructing search systems.

3. THE ROLE OF MEDIATION

In the systems described above, a communication channel was used either to send messages among the collaborators, or to control the information retrieval system in some manner. But a message sent to another person can also be acted upon by the information retrieval component, and an information retrieval act can also generate a message to other people.

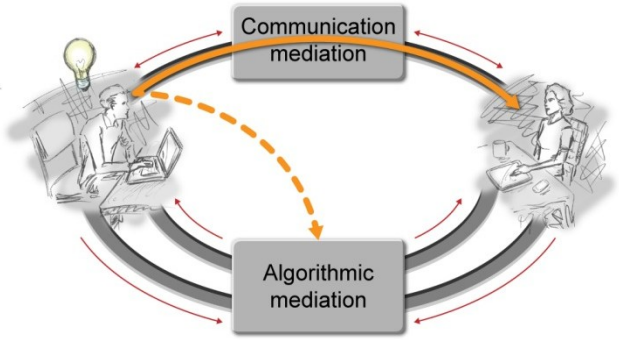

Figure 2. Relevance feedback as a side effect of communication.

For example, sharing document between collaborators can be taken as a form of relevance feedback to the system (Figure 2). Conversely, a system can keep track of relevance feedback operations made by one person for the purpose of refining a query, and communicate that to a collaborator to help him understand what his partner is doing (Figure 3). Thus, rather than having one-to-one communication as was shown in Figure 1, we can consider some additional paths.

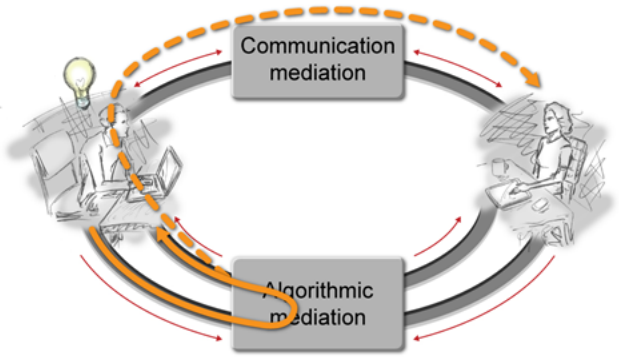

Figure 3. Communication as a side effect of search activity.

These two scenarios in which the communication mediation software component and the algorithmic mediation software component exert influence on each other give rise to four possible combinations of influence.

1. The first (degenerate) case is that no influence is propagated from an interaction with either component to the other component.
2. The second case is that interaction with the algorithmic mediation component causes the communication mediation component to notify other collaborators of a person's actions. An example of this might be a relevance feedback operation that generates some notifications that particular documents were deemed interesting or useful by a collaborator.
3. The third case is that a communication act, such as sharing a document or a query, causes the algorithmic mediation component to infer something about the utility of the shared object. This inference can then affect subsequent ranking, query expansion, or other information retrieval operations. Note that this is distinct from the saving or sharing operations as implemented by SearchTogether or Coagmento, for example, because those are acts of pure communication: they have no side effects that affect the subsequent behavior of the algorithmic mediation component.
4. Finally, the fourth possibility is that the software system makes both kinds of inferences: it reflects interactions with the search engine as communication acts, and makes inferences about the value of information objects based on patterns of communication that reference them.

4. IMPLICATIONS FOR DESIGN

This model has implications for design, the root of which is identifying and demarcating the algorithmic and communicative boundaries. Given the nature of communication and algorithmic feedback during information seeking, when is it safe to assume, for example, that a saved document should be used for relevance feedback automatically? When is it safe to assume that a shared document should be used for relevance feedback automatically? How much of a person's activity in a collaborative search application should be communicated to collaborators to promote awareness? What forms of communication during a search session constitute reliable sources of relevance feedback, and what forms should be ignored by the system? The existing literature in the Information Science & Retrieval field tells us that the role of algorithmic and communicative mediation is strongly influence by task stage [14], state of knowledge [2] and emotive factors [9].

While definitive answers to these questions would require empirical evaluation and will certainly be affected by a variety of contextual factors, we can nonetheless, make some generalizations that should guide the designer in deciding which strategies to implement when. In the following, we will discuss the two paths of influence separately, under the assumption that the effects can be combined trivially.

4.1 From search to communication

Let's consider case two, where a person's search behavior is reflected as communication to his or her collaborators. Here it is useful to distinguish between explicit communication acts and general awareness of others' activity. A person engages in explicit communication through comments, chat conversations, or "sharing" actions; a software system maintains awareness by updating lists of queries that were run or saved documents.

Since we assume that explicit communication carries meaning that helps collaborators solve their shared information need, some care must be taken to avoid cluttering that channel with automatically-generated messages that can obscure person-to-person communication. Thus it may be inappropriate to treat every query that is run or every document that is read or used for relevance feedback as a significant event that should be brought to the attention of one's collaborators. If heuristics can be found that predict reliably the value of some action such that it would otherwise be lost in the aggregation of ongoing activity, then it may be useful to flag it explicitly.

For example, if one person judges a document to be pertinent, while a collaborator dismisses it, the algorithmic mediation component should probably flag the discrepancy to draw searchers' attention to the potential disagreement. By the same token, if one person judges a document to be pertinent, and a collaborator dismisses a different, but objectively very similar document, the algorithmic mediation component should flag this discrepancy as well. User feedback on the discrepancy can then be used to better train one of the mediation components. For example, if the two users maintain their "disagreement" on the relevance of two algorithmically-similar documents, the algorithmic mediation component can modify (retrain) its similarity function. If, on the other hand, one user switches his or her assessment, then communication component can be retrained to bring other types of dissimilar judgments to the users' attention.

Another possible strategy is to elevate unlikely events or series of events: if a person who tends not to make many positive relevance judgments changes that pattern of behavior, it may be useful to notify collaborators that something unusual is going on. If a query retrieves an unusually high number of relevant or useful documents, perhaps that query should be highlighted so that all collaborators can understand why (or if) it is significant.

4.2 From communication to search

Conversely, it is possible to infer the value of particular information objects for subsequent information retrieval calculations based on the quantity and quality of communication about that object. The danger here is that not all communication is intended in the same way. A chat between two people in the context of a document may indicate the utility of that document, but it may also mean that the document is not in fact useful, or it may not mean anything at all with respect to that document.

A study of communication patterns of collaborating searchers found that collaborating teams with poor performance also exhibited the highest chat rates [5]. This suggests that simply counting the numbers of messages associated with a particular information object may not reliably identify pertinent objects. It is an open issue whether automated sentiment analysis on the stream of comments related (in some way) to an information object could be used to assess the pertinence or utility of that object with sufficient reliability to improve system effectiveness.

On the other hand, some interface actions such as explicitly sharing a document with collaborators may be a useful source of information for algorithmic mediation, assuming that participants agree on the definition of pertinence or utility of documents. In such cases, a shared document may well serve as a useful source of query expansion terms or facet values. On the other hand, if there is poor agreement about what constitutes a useful document among participants either because the topic is still insufficiently well understood by all collaborators, or because (as Morris and Horwitz found [10]) some people are not effective at judging pertinence, the act of sharing a document will have less value for algorithmic mediation.

Explicit representations of roles may make it possible to make more reliable inferences about information object utility. For example, if a reference librarian and a domain expert collaborate on a search task, the algorithmic mediation component should

give more weight to documents saved by the domain expert than those saved by the librarian. Similarly, if people are able to formulate accurate relevance judgments, such as in the case of fact-finding tasks where the information object sought is well-defined and a clear objective utility of the document exists, these signals should be leveraged by the algorithmic mediation component.

5. QUERIUM

We have built a system, called Querium, which is designed to help us test these design hypotheses. Querium is a session-based collaborative search tool that implements both algorithmic mediation and communication mediation components, rather than either one or the other. Querium allows two or more people to collaborate on an information seeking task, and includes a variety of communication tools, including a chat and note-taking facility, the ability to comment on documents, and the ability to explicitly share documents and queries. It also includes several algorithmic mediation tools, including query fusion and relevance feedback operations that operate on queries and documents regardless of which collaborator created or identified them. It also has views for maintaining awareness of overall progress in a session, and of contributions by individual searchers.

We have begun collecting data on how people use these tools to satisfy their information needs in a naturalistic setting with the goal of answering the following research questions:

1. Do searchers distinguish between saving documents (explicit relevance feedback) and sharing documents?
2. Does the effectiveness of relevance feedback depend on whether shared or saved documents are used?
3. What is the value of itemized vs. aggregated information for fostering awareness of others' actions?
4. Can we predict which documents or queries will be shared based on how they are used by those collaborators who find or create them?
5. Does explicit sharing of information during a search session lead to its use for tasks beyond the search session?
6. Does the value of document judgments extend beyond the relevance feedback queries, or are such judgments likely to be ephemeral given an evolving information need and exploratory behaviors of searchers?
7. What is the usage and roles of the different search tools across the duration of a task?

6. CONCLUSIONS

Collaborative information seeking is a complex activity that involves the interplay of multiple actors, both human and computer. We can model two classes of exchanges among these actors, person-to-person communication acts and person-to-computer information retrieval exchanges. Most existing software tools for explicit collaborative search implement either one or the other class of message exchange.

In this work, we propose that coupling the two kinds of messages, with due attention to the context of use, can lead to more interesting and richer interactions within the entire human-computer system. To test these conjectures, we have built and deployed a collaborative search system through which we are collecting patterns of behavior and system performance that will help us begin to answer some of these questions.

7. ACKNOWLEDGMENTS

We thank Tony Dunnigan for his excellent artwork, and Larry Rowe and Lynn Wilcox for supporting this research.

8. REFERENCES

[1] Adamic LA, Zhang J, Bakshy E, Ackerman MS. Knowledge sharing and Yahoo Answers: everyone knows something. In: Proc. WWW 2008. pp. 665–667 .

[2] Belkin, N., Oddy, R., & Brooks, H. (1982). ASK for information retrieval: Part I. background and theory. *Journal of Documentation*, 38(2), 61–71.

[3] Fernández-Luna, J. M., Huete, J.F. (2010) CIRLab: A Groupware Framework for Collaborative Information Retrieval Research. *Information Processing & Management* 46(6) pp. 749-761.

[4] Golovchinsky, G., Pickens, J. and Back, M. (2008) A Taxonomy of Collaboration in Online Information Seeking. In *Proc. 1st International Workshop on Collaborative Information Retrieval*. Held in conjunction with JCDL 2008. Available online at http://www.fxpal.com/?p=abstract&abstractID=454

[5] González-Ibáñez, R.I, and Shah, C. (2010) Group's affective relevance: a proposal for studying affective relevance in collaborative information seeking. In *Proc. GROUP '10*. ACM, New York, NY, USA, pp. 317-318

[6] Hansen, P. and Järvelin, K. (2005) Collaborative Information Retrieval in an Information-intensive Domain. *Information Processing & Management*, vol. 41. pp. 1101-1119.

[7] Hyldegård, J. (2006) Collaborative information behavior—exploring Kuhltau's Information Search Process model in a group-based educational setting. *Information Processing & Management*, 42, pp. 276-298.

[8] Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R., Riedl, J., (1997) GroupLens: Applying collaborative filtering to Usenet news. *CACM* 40: (3) 77-87.

[9] Kuhlthau, C.C. (2004). Seeking Meaning: a process approach to library and information services. Norwood, NJ: Ablex.

[10] Morris, M.R. and Horvitz, E. (2007) SearchTogether: an interface for collaborative web search. In *Proc. UIST 2007*, pp. 3-12.

[11] Pickens, J., Golovchinsky, G., Shah, C., Qvarfordt, P., and Back, M. (2008) Algorithmic Mediation for Collaborative Exploratory Search. In Proc. *SIGIR 2008*. ACM, New York, NY, pp. 315-322.

[12] Reddy, M. and Jansen, B.J. (2008) A Model for Understanding Collaborative Information Behavior in Context: A Study of Two Healthcare Teams. *Information Processing and Management* 44(1): pp. 256-273.

[13] Shah, C., Marchionini, G., and Kelly, D. (2009) Learning design principles for a collaborative information seeking system. In *Proc. CHI 2009 Extended Abstracts*. ACM, New York, NY, USA, 3419-3424.

[14] Vakkari, P., Pennanen, M., Serola, S. (2003). Changes in search terms and tactics while writing a research proposal: A longitudinal case study. *Information Processing & Management, 39*, 445-463.

[15] Sonnenwald, D.H. and Pierce, L.G. (2000) Information behavior in dynamic group work contexts: interwoven situational awareness, dense social networks and contested collaboration in command and control. *Information Processing & Management*, 36, pp. 461-479.