Collaborative Information Retrieval (CIR)

Raya Fidel and Harry Bruce
School of Library and Information Science, University of Washington, Box 352930, Seattle, WA 98195, USA,
e-mail: {fidelr, harryb} @u.washington.edu

Annelise Mark Pejtersen
Risoen National Laboratory, DK-400 Roskilde, Denmark
e-mail: Annelise.m.pejtersen@risoe.dk

Susan Dumais and Jonathan Grudin
Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA
e-mail: {sdumais, jgrudin}@microsoft.com

Steven Poltrock
The Boeing Company, PO Box 3707, MC 7L-49, Seattle, WA 98124, USA
e-mail: steven.poltrock@boeing.com

The Collaborative Information Retrieval (CIR) Project is designed to study situations where members of a work-team are seeking, searching, and using information collaboratively. Although no research project has addressed this situation before, it is clear from the literature that CIR has already manifested itself in various settings. The Cognitive Work Analysis framework will guide this field study, which will employ observation and interviews with four work-teams from the knowledge industry. The framework examines social, organizational, cognitive, and individual characteristics of information seekers, and will be focused to address collaborative situations. The study considers information seeking as a social, dynamic process and has the potential to bring new insights and ideas to the area of information behaviour.

WHAT IS CIR?

To date, the study of information seeking behaviour has focused on people seeking information on their own to solve their individual problems. Although various social and contextual factors that might bear upon this behaviour have been examined, and processes in which help is provided have been studied, no research has focused on information seeking as a collective process. This is surprising, given the fact that seeking information in a collective and collaborative way is becoming increasingly common, especially in the workplace.

In today's work environment, there is a growing emphasis on collaborative teamwork. The assumption is that a carefully selected and managed team will focus on a problem with a collective expertise that is greater than the
sum of its parts. Each member of the team has a repertoire of individual skills and perspectives but the power of the team comes from catalyzing these individual qualities in the way a magnifying glass captures and intensifies the heat generated by individual rays of light from the sun. Traditionally, working in teams meant collocation, but recent innovation in communication and information technology now means that the workplace and work-teams may be distributed across regions and nations. Collaborative teamwork is a management challenge, but it is also an intriguing information problem.

The flow of information, as well as information sharing, is essential to a teamwork's success. Indeed, research in computer support for cooperative work (CSCW) and in collaborative filtering has focused on this aspect. Various studies in CSCW addressed information transfer in collaborative situations. For example, years of research work have examined formal meetings and the sharing of information during these meetings, in order to design technology that can support these group interactions (1). Studies have also investigated desktop conferencing among distributed team members (2), and other unplanned interactions to get and to give information (3). Similarly, collaborative filtering and recommender systems have been developed to allow users to share information with each other. Here, the behaviours of individuals in the same group are used to filter out or to recommend items to other members of the group (4). This way, new members, or those who are new to a certain issue, benefit from the experiences of others (e.g., 5, 6). Recommender systems are already active in electronic commerce, recommending books, movies, or restaurants to customers.

While this research has investigated information transfer in collaborative settings, it has not examined information seeking. Moreover, it was rarely assumed that the process of information seeking was carried out collectively. Collaborative Information Retrieval (CIR) focuses on situations where team members collaborate during various processes of information retrieval. We understand CIR to be any activity that collectively resolves an information problem taken by members of a work-team, regardless of the nature of the actual retrieval of information. Information retrieval is construed in the broadest sense and includes processes such as problem identification, analysis of information need, query formulation, retrieval interactions, evaluation and presentation of results, and applying results to resolve an information problem. Members of a work-team are those who were brought together to solve a problem, which is the goal of the team.
Researchers in various areas are just starting to address explicitly the need to study CIR (e.g., 7, 8). Manifestations of it, however, have already become apparent.

**MANIFESTATIONS OF CIR**

There are several ways in which team members may be engaged in CIR. A team may formally plan to seek information, allocating different search tasks to different people. For example, a team preparing a recommendation regarding the launch of a new aeroplane programme may assign people to approach different airlines regarding their potential interest in the aeroplane capabilities. Other members of the team may evaluate the readiness of new materials that could reduce the weight of the plane, increasing its range and fuel efficiency. Teams often track progress on such information acquisition tasks.

Teams may acquire information in an ad hoc, unplanned way. For example, a team developing a new software product may have extended discussions about the product’s intended behaviour and market. Such discussions often raise questions that cannot be immediately answered by anyone on the team. Often people will seek missing information on their own initiative, then bring this information back to the team. This new information may raise new questions that they or others will investigate later.

People may work alone or together at various stages of a team project. Harper and Sellen (9), for example, found that teams at the International Monetary Fund changed their CIR process during mission projects. At the beginning, the process was divided among the economists on a team so that each one collected information about a unique facet of the project. At this stage, the information collected was somewhat private, because it required interpretation and evaluation before other economists could incorporate it into the project. Once an information gatherer evaluated and interpreted the information she or he had collected, the team could continue to seek information collectively.

Team discussions are a common means of identifying goals and the information needed to achieve them, but a single individual may be assigned to prepare an analysis of information requirements. The team may also discuss where information may be found, identifying other people who have critical experience or knowledge. Two or more people may work together to locate and extract information. Two people may conduct a joint interview with an airline executive, for example, with the expectation that two people will understand and remember better than one. Teams often
work together to integrate the resulting information into an expanding knowledge of their shared problem domain and its solutions.

Despite the richness of anecdotal evidence, to our knowledge, the research literature offers no report about a systematic study of CIR. Nevertheless, some studies in environments such as industry, medicine, education, and public policy are relevant to this discussion.

The earliest detection of CIR occurred in studies of engineers by Thomas Allen (10) who uncovered the gatekeeper phenomenon. A gatekeeper takes the responsibility to look for information and to channel it to other team members who are engaged in solving work-related problems. This way, the recipient of the information and the gatekeeper collaborate to find information that is useful to the work of the team. Later studies identified gatekeepers in other environments such as among office workers (11). The gatekeeper phenomenon has extended beyond the workplace and was found among members of ethnolinguistic communities (12), and among customers of electronic shopping (13).

Although coordination of information is essential to the success of a design process, Fisher (14) explained that engineering is a creative task that presents challenges to project management. Drawing from his experience in the design of complex electronic systems, he observed that, in particular, it was difficult to guarantee that engineers remembered to share information even if a common problem was to be solved. This was an alarming situation in particular when the information and the decisions made extended to other parts of the design.

A more promising context for CIR is that of teams working on product development through Concurrent Engineering and Integrated Product Development. Because of the complexity of new products and with the aim of reducing development time, as well as increasing quality and productivity, industry turned to this approach that brings together in a team all who are involved with a certain product, across disciplines and organizational hierarchy. The aim is to integrate and concurrently apply all necessary processes to produce a product that satisfies the customer’s needs. Such a development process requires much cooperation and coordination and is, therefore, highly information intensive (15). Of particular interest are situations in which a team is solving a common problem. While these have not been addressed, facilitating information sharing and transfer among the team members has been considered a major challenge.

From an organizational viewpoint, Jassawalla and Sashittal (16) identified
various impediments to collaborative work in concurrent engineering, after they had interviewed 40 managers of successful teams, as well as unsuccessful ones. They found that teams that were successful in collaborating sought and gathered information collectively, while in teams that were not able to work collaboratively, the management tightly controlled information. Other impediments reported in their study, such as lack of a shared understanding of the problem the team is gathered to solve, are relevant for the success of CIR as well. Several models and approaches have been proposed to handle information in concurrent engineering (e.g., 17, 18, 19). When these models are tested and implemented they might prove useful for CIR as well.

Another collaborative, information intensive process is the task of medical diagnosis and treatment planning, particularly in teaching hospitals. It involves players with various expertise who gather information to support this task, including physicians, residents, students, nurses, and laboratory technicians. Studies have shown that this is a complex social process in which information gathering is naturally distributed according to the status and area of expertise of each player (20). For example, students at times record a patient’s history. While essential to the students’ training, the information they collect is evaluated by an experienced resident or a physician before it is integrated into the team’s knowledge and considered in decision making. Further, for this task all players use both formal and informal information from a multitude of areas in addition to medicine, such as hospital procedures, past experience in following the procedures, and stereotypes about types of patients (21).

Medical students who participated in Problem-Based Learning are actually introduced to CIR during their studies. In a self-directed session with a coach, one group examined a case and identified learning issues (LIs) for the group as a whole. The processes involved then included the collective activities of (a) recognizing a topic as an LI that fulfills certain requirements; (b) researching the topic by jointly constructing a search strategy, assembling input from multiple sources, and collaboratively interpreting the findings; (c) reporting individually to the other members of the groups about the findings; and (d) reflecting about the LI processing (22).

Other studies in educational settings spotted CIR behaviour. Examining user interactions at terminals set in a university library, Twidale, Nichols, and Paice (23) observed that groups of students were working together on group-based problem solving tasks, even though, typically, the library was designed for individuals to work by themselves. These students performed
either a Joint Search, gathering around a single terminal, or a Coordinated Search, working in adjacent terminals and discussing what they were doing. Such collaboration was also observed among high school students. In a study about Web searching behaviour, Fidel et al. (24) found that students exchanged information about search strategies and promising Web sites while they all searched for the same homework assignment.

More than in any other environment, CIR is essential to the process of public policy making in a democratic society. For example, Tabor et al. (25) demonstrate the importance of collaboratively collecting information in the area of natural resources management and planning. In addition to the various agencies, advocacy coalitions will try to participate in this process. The success of groups attempting to influence policy may be contingent on access to suitable information. Ecological and environmental policy making, by its very nature, requires data sets that are compiled by many agencies at the local, state, and federal levels. The need for information from such diverse sources necessitated interagency collaboration in information gathering (26).

With this rich evidence to the existence of CIR in the workplace, the CIR Project is the first endeavour designed to systematically investigate this collaborative activity.

THE CIR PROJECT
The purpose of the Project is to conduct the first systematic study of CIR as it occurs in natural settings. The goals are threefold:

- to analyze the structure of collaboration in selected information-intensive, actual work scenarios and to describe the context that will shape collaborative information retrieval patterns;
- to explore the nature and occurrence of collaborative information retrieval in work settings and to propose technological innovation and organizational change that could support, facilitate, or improve it;
- to construct a work-centered conceptual framework that includes collaborative information retrieval.

Cognitive Work Analysis
A work-centered conceptual framework called Cognitive Work Analysis is guiding our investigation. This framework was developed as a general approach to help information system designers analyze and understand the complex interaction between (a) the activities and organizational relationships and constraints of work domains, and (b) users' cognitive and
social activities and their subjective preferences during task performance. The framework's theoretical roots are in Adaptive Control Systems and Gibson's Ecological Psychology and are the result of the generalization of experiences from field studies which led to the design of support systems for a variety of modern work domains, such as process plants, manufacturing, hospitals, and libraries. Like Cognitive Systems Engineering, it is based on the assumption that system design for work in dynamic environments should be based on the analysis of the factors that shape behaviour, rather than on the description of the procedures followed (27, 28).

This work-centered approach to the evaluation and design of information systems assumes that information interaction is determined by a number of dimensions. The Cognitive Work Analysis is carried out first. It provides an in-depth analysis of information seeking and its context, and it can be also used as a basis for evaluation. It is both descriptive and prescriptive in nature, because its purpose is not only to understand the current work but also to go beyond the work practice. This Analysis addresses the following dimensions:

- the environment within which the workplace is operating;
- the work domain;
- activity analysis: (a) the task in terms of work domain, (b) decision making that is required for the task, and (c) mental strategies that can be used for the task;

![Diagram](image)

**FIG. 1: Framework for Work Analysis. Based on (28).**
• the organization in terms of division of work and social organization;
• user characteristics, resources, and values.

Each dimension is analyzed according to four abstraction levels: goals and constraints, priorities, work process, and physical resources.

The framework for system evaluation, which is based on the Cognitive Work Analysis, answers questions such as: Does the system support cooperative work and coordination? Does the system support the task repertoire of a work situation? Does the system support relevant decision tasks? Are all relevant strategies supported? Does presentation match sensory characteristics? The answers to these questions are based on the data collected during the Cognitive Work Analysis.

With this framework as a guide, the study will answer specific questions such as:

**What are the manifestations of CIR in work settings?**

• How and when do work-teams establish a CIR path? How much is made explicit?
• How are the roles of team members in information retrieval shaped by the work content?
• What formal and informal processes do teams use to coordinate their IR activities?
• What difficulties do work-team members encounter in retrieving information collaboratively?
• What is the role of CIR in a team’s work?
• What technologies do work-team members use for CIR?

**Are there teamwork situations where CIR is not used?**

• What factors motivate work-teams to use CIR?
• What factors dissuade work-teams from retrieving information interactively?
• Can we identify situations in which CIR would be effective and others in which it would not be desirable?
• What technological innovations can improve, support, and facilitate CIR?
• What organizational changes can improve, support, and facilitate work groups?
Although the Cognitive Work Analysis framework has been applied both to individual and to co-operative work, most research and experience in work analysis, and in design and evaluation of information systems, has been gathered in work domains where teamwork has not been predominant. Part of the objectives of the Project is to expand the framework to such analyses in the area of information systems.

Research plan

The first phases of this research involve observations of four work-teams and the analysis of their collaborative information retrieval activities. The teams selected for this study are engaged in knowledge work that requires access and integration of diverse information. Their work is not principally information retrieval; it is only a supporting activity of their work.

The study uses a variety of instruments depending on the specific phase of the research activity. The first phase includes observations and interviews in the actual workplace settings where the work-team is functioning. Before going to the field, the researchers conducted a general study of work scenarios from available documents such as annual company reports, organizational charts, work manuals, procedures and instructions, and published work analyses. Members of the research team then accompany work-team members during their workday. The researchers sit in team meetings taking notes about the activities of the team and the interactions that occur between work-team members. Where possible, data has been collected on the interchanges that occur on e-mail between work-team members. Work-team members have been observed as they seek information individually or in pairs or groups using a think-aloud protocol. Finally, the researchers have interviewed each work-team member to complete the work analysis for the work-team and the individual involved. All verbal protocols, including think-aloud and interview protocols, are transcribed and made available to all members of the research team.

To analyze the data collected in the observation of the first work-team, the research team will analyze the work and the present retrieval tools and practices. This analysis will continue until the research questions are answered for that team. After the observation of the second work-team, the researchers will get together again to begin the work analysis of the second work-team and integrate it with that of the first. This process will continue until an analysis that pertains to both teams is completed. After the fourth iteration, this work analysis process will have covered all four work-teams. The research team will then develop a survey instrument and interview schedule to validate some of the findings and to consider possible generalizations.

An example

This research is exploratory in nature. The researchers cannot be certain about what they will find when they observe work-teams using the techniques of work analysis. But the aim is to find out how CIR is a part of the functioning of work-teams in organizations, and, if so, what the dimensions and manifestations of CIR might be. In spite of the exploratory nature of the research, it is possible to illustrate the methods and procedures that the researchers intend to use with a scenario.

Let us suppose, for example, that during the observation period, an engineer in a work-team indicates that when he wants to consult with someone about information seeking he always turns to one of two work-team members. During an interview, the engineer may explain that he feels comfortable with these two members, and he does not like to bother other people. The researcher might also observe that the engineer's consultations do not always help his information seeking to progress on a fruitful path. This scenario is recorded during the work analysis stage of the investigation when the researcher describes the task situation in term of mental strategies.

The system evaluation stage of the investigation focuses on a different question: Does the system support all relevant strategies for information retrieval? In the case of this engineer, the answer would be no. His strategy of consulting with other work-team members is not supported because only two members of the team are available to him for consultation.

A member in the second work-team to be observed by the researchers may display a similar behaviour; only she may explain that she is consulting only her friends, because she does not want to appear to be stupid. This behaviour is repeated in other teams and possibly for other reasons. What is clear in the evaluation is that some team members do not receive all the support they need for CIR. In particular, they receive no support for consultation with other team members.

The survey instrument may reveal the extent of this problem, and may also provide additional data about the factors that affect this behaviour. At the same time, several solutions can be considered. For example, suggestions for a more supportive environment within a work-team for CIR can be developed, say, by weekly short meetings to discuss issues relating to information seeking. Technological solutions might also be recommended. For example, a database of information strategies that have been used by work-team members could be constructed to make these strategies available to all work-team members. In addition, an electronic
communication system could be established for work-team members to pose questions to one another and to suggest answers anonymously or otherwise.

Field observation in the CIR Project started in March 2000.

CONCLUSIONS
The CIR Project examines information seeking as a social and dynamic process. It requires not only an in-depth understanding of information seeking and of the context in which it takes place, but also an understanding of the constantly changing relationships between the information seekers and their environment.

Studying a newly charted territory is especially challenging. The CIR Project addresses a phenomenon that has been first recognized only a few years ago. In addition, it is motivated by a real-life problem and guided by an established and well-tested conceptual framework. These particular attributes place the Project in a distinct position.

To our knowledge, no attempt has been made before to systematically study CIR. While various researchers have detected CIR behaviours in studies of other processes and phenomena, no scientific investigation has focused on this process. Therefore, the Project cannot build on previous research or on experience in methods used to study the phenomenon. On the other hand, because the study is problem-driven, rather than discipline-driven, it is interdisciplinary. As such, it can integrate experience gained in studies in other areas such as computer support of cooperative work, cognitive systems engineering, and human computer interaction. This will require an analysis and interpretation of methods and approaches so they could be easily implemented in the study of information behaviour. The Cognitive Work Analysis framework will guide this process, but at the same time it will itself grow to address more fully collaborative situations.

Because of this distinct position, the Project will continuously contribute new insights and ideas to the study of information behavior in context. The results of the Project will have both practical and theoretical implications. On the practical side, understanding how information is acquired, shared among team members, and used in work situations will greatly strengthen our ability to (a) address issues that relate to knowledge management in organizations, and (b) develop technologies that support, facilitate, and enhance teamwork. The Project will also extend an existing work-centered conceptual framework to include teamwork. This framework will become a new analytical tool that will inform the methods used in other studies about
CIR. In addition, this new framework could enrich other conceptual constructs both in information seeking and in collaborative work.

REFERENCES


ACKNOWLEDGEMENT
The Collaborative Information Retrieval Project is supported by a grant from the National Science Foundation.